MMagic

MMagic Authors

Jan 31, 2024

COMMUNITY

Languages: English |

MMagic (**M**ultimodal **A**dvanced, **G**enerative, and **I**ntelligent **C**reation) is an open-source AIGC toolbox for professional AI researchers and machine learning engineers to explore image and video processing, editing and generation.

MMagic supports various foundamental generative models, including:

- Unconditional Generative Adversarial Networks (GANs)
- Conditional Generative Adversarial Networks (GANs)
- Internal Learning
- Diffusion Models
- And many other generative models are coming soon!

MMagic supports various applications, including:

- Text-to-Image
- Image-to-image translation
- 3D-aware generation
- Image super-resolution
- Video super-resolution
- Video frame interpolation
- Image inpainting
- Image matting
- Image restoration
- Image colorization
- Image generation
- And many other applications are coming soon!

MMagic is based on PyTorch and is a part of the OpenMMLab project. Codes are available on GitHub.

CHAPTER

ONE

DOCUMENTATION

1.1 Contributing guidance

Welcome to the MMagic community, we are committed to building a Multimodal Advanced, Generative, and Intelligent Creation Toolbox.

This section introduces following contents:

- *Contributing guidance*
  - *Pull Request Workflow*
    - 1. Fork and clone
    - 2. Configure pre-commit
    - 3. Create a development branch
    - 4. Commit the code and pass the unit test
    - 5. Push the code to remote
    - 6. Create a Pull Request
    - 7. Resolve conflicts
  - *Guidance*
    - Unit test
    - Document rendering
  - *Code style*
    - Python
    - C++ and CUDA
  - *PR Specs*

All kinds of contributions are welcomed, including but not limited to

Fix bug

You can directly post a Pull Request to fix typo in code or documents

The steps to fix the bug of code implementation are as follows.

1. If the modification involve significant changes, you should create an issue first and describe the error information and how to trigger the bug. Other developers will discuss with you and propose an proper solution.
2. Posting a pull request after fixing the bug and adding corresponding unit test.

New Feature or Enhancement

1. If the modification involve significant changes, you should create an issue to discuss with our developers to propose an proper design.
2. Post a Pull Request after implementing the new feature or enhancement and add corresponding unit test.

Document

You can directly post a pull request to fix documents. If you want to add a document, you should first create an issue to check if it is reasonable.

1.1.1 Pull Request Workflow

If you're not familiar with Pull Request, don't worry! The following guidance will tell you how to create a Pull Request step by step. If you want to dive into the develop mode of Pull Request, you can refer to the official documents

1. Fork and clone

If you are posting a pull request for the first time, you should fork the OpenMMLab repositories by clicking the **Fork** button in the top right corner of the GitHub page, and the forked repositories will appear under your GitHub profile.

Then, you can clone the repositories to local:

```
git clone git@github.com:{username}/mmagic.git
```

After that, you should ddd official repository as the upstream repository

```
git remote add upstream git@github.com:open-mmlab/mmagic
```

Check whether remote repository has been added successfully by `git remote -v`

```
origin        git@github.com:{username}/mmagic.git (fetch)
origin        git@github.com:{username}/mmagic.git (push)
upstream        git@github.com:open-mmlab/mmagic (fetch)
upstream        git@github.com:open-mmlab/mmagic (push)
```

Note: Here's a brief introduction to origin and upstream. When we use "git clone", we create an "origin" remote by default, which points to the repository cloned from. As for "upstream", we add it ourselves to point to the target repository. Of course, if you don't like the name "upstream", you could name it as you wish. Usually, we'll push the code to "origin". If the pushed code conflicts with the latest code in official("upstream"), we should pull the latest code from upstream to resolve the conflicts, and then push to "origin" again. The posted Pull Request will be updated automatically.

2. Configure pre-commit

You should configure pre-commit in the local development environment to make sure the code style matches that of OpenMMLab. **Note**: The following code should be executed under the mmagic directory.

```
pip install -U pre-commit
pre-commit install
```

Check that pre-commit is configured successfully, and install the hooks defined in `.pre-commit-config.yaml`.

```
pre-commit run --all-files
```

Note: Chinese users may fail to download the pre-commit hooks due to the network issue. In this case, you could download these hooks from gitee by setting the .pre-commit-config-zh-cn.yaml

pre-commit install -c .pre-commit-config-zh-cn.yaml pre-commit run –all-files -c .pre-commit-config-zh-cn.yaml

---

If the installation process is interrupted, you can repeatedly run `pre-commit run ...` to continue the installation.

If the code does not conform to the code style specification, pre-commit will raise a warning and fixes some of the errors automatically.

If we want to commit our code bypassing the pre-commit hook, we can use the `--no-verify` option(**only for temporarily commit**.

```
git commit -m "xxx" --no-verify
```

3. Create a development branch

After configuring the pre-commit, we should create a branch based on the main branch to develop the new feature or fix the bug. The proposed branch name is `username/pr_name`

```
git checkout -b yhc/refactor_contributing_doc
```

In subsequent development, if the main branch of the local repository is behind the main branch of "upstream", we need to pull the upstream for synchronization, and then execute the above command:

```
git pull upstream main
```

4. Commit the code and pass the unit test

- mmagic introduces mypy to do static type checking to increase the robustness of the code. Therefore, we need to add Type Hints to our code and pass the mypy check. If you are not familiar with Type Hints, you can refer to this tutorial.
- The committed code should pass through the unit test

```
# Pass all unit tests
pytest tests

# Pass the unit test of runner
pytest tests/test_runner/test_runner.py
```

If the unit test fails for lack of dependencies, you can install the dependencies referring to the guidance

- If the documents are modified/added, we should check the rendering result referring to guidance

5. Push the code to remote

We could push the local commits to remote after passing through the check of unit test and pre-commit. You can associate the local branch with remote branch by adding `-u` option.

```
git push -u origin {branch_name}
```

This will allow you to use the `git push` command to push code directly next time, without having to specify a branch or the remote repository.

6. Create a Pull Request

(1) Create a pull request in GitHub's Pull request interface

(2) Modify the PR description according to the guidelines so that other developers can better understand your changes

Find more details about Pull Request description in *pull request guidelines*.

note

(a) The Pull Request description should contain the reason for the change, the content of the change, and the impact of the change, and be associated with the relevant Issue (see documentation

(b) If it is your first contribution, please sign the CLA

(c) Check whether the Pull Request pass through the CI

mmagic will run unit test for the posted Pull Request on different platforms (Linux, Window, Mac), based on different versions of Python, PyTorch, CUDA to make sure the code is correct. We can see the specific test information by clicking `Details` in the above image so that we can modify the code.

(3) If the Pull Request passes the CI, then you can wait for the review from other developers. You'll modify the code based on the reviewer's comments, and repeat the steps 4-5 until all reviewers approve it. Then, we will merge it ASAP.

7. Resolve conflicts

If your local branch conflicts with the latest main branch of "upstream", you'll need to resolove them. There are two ways to do this:

```
git fetch --all --prune
git rebase upstream/main
```

or

```
git fetch --all --prune
git merge upstream/main
```

If you are very good at handling conflicts, then you can use rebase to resolve conflicts, as this will keep your commit logs tidy. If you are not familiar with `rebase`, then you can use `merge` to resolve conflicts.

1.1.2 Guidance

Unit test

We should make sure the committed code will not decrease the coverage of unit test, we could run the following command to check the coverage of unit test:

```
python -m coverage run -m pytest /path/to/test_file
python -m coverage html
# check file in htmlcov/index.html
```

Document rendering

If the documents are modified/added, we should check the rendering result. We could install the dependencies and run the following command to render the documents and check the results:

```
pip install -r requirements/docs.txt
cd docs/zh_cn/
# or docs/en
make html
# check file in ./docs/zh_cn/_build/html/index.html
```

1.1.3 Code style

Python

We adopt PEP8 as the preferred code style.

We use the following tools for linting and formatting:

- flake8: A wrapper around some linter tools.
- isort: A Python utility to sort imports.
- yapf: A formatter for Python files.
- codespell: A Python utility to fix common misspellings in text files.
- mdformat: Mdformat is an opinionated Markdown formatter that can be used to enforce a consistent style in Markdown files.
- docformatter: A formatter to format docstring.

Style configurations of yapf and isort can be found in setup.cfg.

We use pre-commit hook that checks and formats for `flake8`, `yapf`, `isort`, `trailing whitespaces`, `markdown files`, fixes `end-of-files`, `double-quoted-strings`, `python-encoding-pragma`, `mixed-line-ending`, sorts `requirments.txt` automatically on every commit. The config for a pre-commit hook is stored in .pre-commit-config.

C++ and CUDA

We follow the Google C++ Style Guide.

1.1.4 PR Specs

1. Use pre-commit hook to avoid issues of code style
2. One short-time branch should be matched with only one PR
3. Accomplish a detailed change in one PR. Avoid large PR
   - Bad: Support Faster R-CNN
   - Acceptable: Add a box head to Faster R-CNN
   - Good: Add a parameter to box head to support custom conv-layer number
4. Provide clear and significant commit message
5. Provide clear and meaningful PR description
   - Task name should be clarified in title. The general format is: [Prefix] Short description of the PR (Suffix)
   - Prefix: add new feature [Feature], fix bug [Fix], related to documents [Docs], in developing [WIP] (which will not be reviewed temporarily)
   - Introduce main changes, results and influences on other modules in short description
   - Associate related issues and pull requests with a milestone

1.2 MMagic projects

Welcome to the MMagic community! The MMagic ecosystem consists of tutorials, libraries, and projects from a broad set of researchers in academia and industry, ML and application engineers. The goal of this ecosystem is to support, accelerate, and aid in your exploration with MMagic for AIGC such as image, video, 3D content generation, editing and processing.

Here are a few projects that are built upon MMagic. They are examples of how to use MMagic as a library, to make your projects more maintainable. Please find more projects in MMagic Ecosystem.

1.2.1 Show your projects on OpenMMLab Ecosystem

You can submit your project so that it can be shown on the homepage of OpenMMLab.

1.2.2 Add example projects to MMagic

Here is an example project about how to add your projects to MMagic. You can copy and create your own project from the example project.

We also provide some documentation listed below for your reference:

- Contribution Guide

  The guides for new contributors about how to add your projects to MMagic.

- New Model Guide

  The documentation of adding new models.

- Discussions

  Welcome to start a discussion!

1.2.3 Projects of libraries and toolboxes

- PowerVQE: Open framework for quality enhancement of compressed videos based on PyTorch and MMagic.
- VR-Baseline: Video Restoration Toolbox.
- Derain-Toolbox: Single Image Deraining Toolbox and Benchmark

1.2.4 Projects of research papers

- Towards Interpretable Video Super-Resolution via Alternating Optimization, ECCV 2022[github]
- SepLUT:Separable Image-adaptive Lookup Tables for Real-time Image Enhancement, ECCV 2022[github]
- TTVSR: Learning Trajectory-Aware Transformer for Video Super-Resolution, CVPR 2022[github]
- Arbitrary-Scale Image Synthesis, CVPR 2022[github]
- Investigating Tradeoffs in Real-World Video Super-Resolution(RealBasicVSR), CVPR 2022[github]
- BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment, CVPR 2022[github]
- Multi-Scale Memory-Based Video Deblurring, CVPR 2022[github]
- AdaInt:Learning Adaptive Intervals for 3D Lookup Tables on Real-time Image Enhancement, CVPR 2022[github]
- A New Dataset and Transformer for Stereoscopic Video Super-Resolution, CVPRW 2022[github]
- Liquid warping GAN with attention: A unified framework for human image synthesis, TPAMI 2021[github]
- BasicVSR:The Search for Essential Components in Video Super-Resolution and Beyond, CVPR 2021[github]
- GLEAN:Generative Latent Bank for Large-Factor Image Super-Resolution, CVPR 2021[github]
- DAN:Unfolding the Alternating Optimization for Blind Super Resolution, NeurIPS 2020[github]

1.3 Overview

Welcome to MMagic! In this section, you will know about

- *Overview*
  - *What is MMagic?*
  - *Why should I use MMagic?*
  - *Get started*
  - *User guides*
    - *Advanced guides*
    - *How to*

1.3.1 What is MMagic?

MMagic (**M**ultimodal **A**dvanced, **G**enerative, and **I**ntelligent **C**reation) is an open-source AIGC toolbox for professional AI researchers and machine learning engineers to explore image and video processing, editing and generation.

MMagic allows researchers and engineers to use pre-trained state-of-the-art models, train and develop new customized models easily.

MMagic supports various foundamental generative models, including:

- Unconditional Generative Adversarial Networks (GANs)
- Conditional Generative Adversarial Networks (GANs)
- Internal Learning
- Diffusion Models
- And many other generative models are coming soon!

MMagic supports various applications, including:

- Text-to-Image
- Image-to-image translation
- 3D-aware generation
- Image super-resolution
- Video super-resolution
- Video frame interpolation
- Image inpainting
- Image matting
- Image restoration
- Image colorization
- Image generation
- And many other applications are coming soon!

1.3.2 Why should I use MMagic?

- **State of the Art Models**

  MMagic provides state-of-the-art generative models to process, edit and synthesize images and videos.

- **Powerful and Popular Applications**

  MMagic supports popular and contemporary image restoration, text-to-image, 3D-aware generation, inpainting, matting, super-resolution and generation applications. Specifically, MMagic supports fine-tuning for stable diffusion and many exciting diffusion's application such as ControlNet Animation with SAM. MMagic also supports GAN interpolation, GAN projection, GAN manipulations and many other popular GAN's applications. It's time to begin your AIGC exploration journey!

- **Efficient Framework**

  By using MMEngine and MMCV of OpenMMLab 2.0 framework, MMagic decompose the editing framework into different modules and one can easily construct a customized editor framework by combining different modules. We can define the training process just like playing with Legos and provide rich components and strategies.

In MMagic, you can complete controls on the training process with different levels of APIs. With the support of MMSeparateDistributedDataParallel, distributed training for dynamic architectures can be easily implemented.

1.3.3 Get started

For installation instructions, please see *Installation*.

1.3.4 User guides

For beginners, we suggest learning the basic usage of MMagic from *user_guides*.

Advanced guides

For users who are familiar with MMagic, you may want to learn the design of MMagic, as well as how to extend the repo, how to use multiple repos and other advanced usages, please refer to *advanced_guides*.

How to

For users who want to use MMagic to do something, please refer to *How to*.

1.4 Installation

In this section, you will know about:

1.4.1 Installation

We recommend that users follow our *Best practices* to install MMagic. However, the whole process is highly customizable. See *Customize installation* section for more information.

Prerequisites

In this section, we demonstrate how to prepare an environment with PyTorch.

MMagic works on Linux, Windows, and macOS. It requires:

- Python >= 3.7
- PyTorch >= 1.8
- MMCV >= 2.0.0

If you are experienced with PyTorch and have already installed it, just skip this part and jump to the *next section*. Otherwise, you can follow these steps for the preparation.

Step 0. Download and install Miniconda from official website.

Step 1. Create a conda environment and activate it

```
conda create --name mmagic python=3.8 -y
conda activate mmagic
```

Step 2. Install PyTorch following official instructions, e.g.

- On GPU platforms:

  ```
  conda install pytorch torchvision cudatoolkit=11.3 -c pytorch
  ```

- On CPU platforms:

  ```
  conda install pytorch=1.10 torchvision cpuonly -c pytorch
  ```

Best practices

Step 0. Install MMCV using MIM.

```
pip install -U openmim
mim install 'mmcv>=2.0.0'
```

Step 1. Install MMEngine.

```
mim install 'mmengine'
```

Or

```
pip install mmengine
```

Or

```
pip install git+https://github.com/open-mmlab/mmengine.git
```

Step 2. Install MMagic.

```
mim install 'mmagic'
```

Or

```
pip install mmagic
```

Or install MMagic from the source code.

```
git clone https://github.com/open-mmlab/mmagic.git
cd mmagic
pip3 install -e . -v
```

Step 5. Verify MMagic has been successfully installed.

```
cd ~
python -c "import mmagic; print(mmagic.__version__)"
# Example output: 1.0.0
```

The installation is successful if the version number is output correctly.

Note: You may be curious about what `-e .` means when supplied with `pip install`. Here is the description:

- `-e` means editable mode. When `import mmagic`, modules under the cloned directory are imported. If `pip install` without `-e`, pip will copy cloned codes to somewhere like `lib/python/site-package`. Consequently, modified code under the cloned directory takes no effect unless `pip install` again. Thus, `pip install` with `-e` is particularly convenient for developers. If some codes are modified, new codes will be imported next time without reinstallation.
- `.` means code in this directory

You can also use `pip install -e .[all]`, which will install more dependencies, especially for pre-commit hooks and unittests.

Customize installation

CUDA Version

When installing PyTorch, you need to specify the version of CUDA. If you are not clear on which to choose, follow our recommendations:

- For Ampere-based NVIDIA GPUs, such as GeForce 30 series and NVIDIA A100, CUDA 11 is a must.
- For older NVIDIA GPUs, CUDA 11 is backward compatible, but CUDA 10.2 offers better compatibility and is more lightweight.

Please make sure the GPU driver satisfies the minimum version requirements. See this table for more information.

note Installing CUDA runtime libraries is enough if you follow our best practices, because no CUDA code will be compiled locally. However, if you hope to compile MMCV from source or develop other CUDA operators, you need to install the complete CUDA toolkit from NVIDIA's website, and its version should match the CUDA version of PyTorch. i.e., the specified version of cudatoolkit in `conda install` command.

Install MMCV without MIM

MMCV contains C++ and CUDA extensions, thus depending on PyTorch in a complex way. MIM solves such dependencies automatically and makes the installation easier. However, it is not a must.

To install MMCV with pip instead of MIM, please follow MMCV installation guides. This requires manually specifying a find-url based on PyTorch version and its CUDA version.

For example, the following command install mmcv-full built for PyTorch 1.10.x and CUDA 11.3.

```
pip install 'mmcv>=2.0.0' -f https://download.openmmlab.com/mmcv/dist/cu113/torch1.10/
↪index.html
```

Using MMagic with Docker

We provide a Dockerfile to build an image. Ensure that your docker version >=19.03.

```
# build an image with PyTorch 1.8, CUDA 11.1
# If you prefer other versions, just modified the Dockerfile
docker build -t mmagic docker/
```

Run it with

```
docker run --gpus all --shm-size=8g -it -v {DATA_DIR}:/mmagic/data mmagic
```

Trouble shooting

If you have some issues during the installation, please first view the *FAQ* page. You may open an issue on GitHub if no solution is found.

Developing with multiple MMagic versions

The train and test scripts already modify the `PYTHONPATH` to ensure the script uses the `MMagic` in the current directory.

To use the default MMagic installed in the environment rather than that you are working with, you can remove the following line in those scripts

```
PYTHONPATH="$(dirname $0)/..":$PYTHONPATH
```

1.5 Quick run

After installing MMagic successfully, now you are able to play with MMagic! To generate an image from text, you only need several lines of codes by MMagic!

```
from mmagic.apis import MMagicInferencer
sd_inferencer = MMagicInferencer(model_name='stable_diffusion')
text_prompts = 'A panda is having dinner at KFC'
result_out_dir = 'output/sd_res.png'
sd_inferencer.infer(text=text_prompts, result_out_dir=result_out_dir)
```

Or you can just run the following command.

```
python demo/mmagic_inference_demo.py \
    --model-name stable_diffusion \
    --text "A panda is having dinner at KFC" \
    --result-out-dir ./output/sd_res.png
```

You will see a new image `sd_res.png` in folder `output/`, which contained generated samples.

What's more, if you want to make these photos much more clear, you only need several lines of codes for image super-resolution by MMagic!

```
from mmagic.apis import MMagicInferencer
config = 'configs/esrgan/esrgan_x4c64b23g32_1xb16-400k_div2k.py'
checkpoint = 'https://download.openmmlab.com/mmediting/restorers/esrgan/esrgan_
↪x4c64b23g32_1x16_400k_div2k_20200508-f8ccaf3b.pth'
img_path = 'tests/data/image/lq/baboon_x4.png'
editor = MMagicInferencer('esrgan', model_config=config, model_ckpt=checkpoint)
output = editor.infer(img=img_path,result_out_dir='output.png')
```

Now, you can check your fancy photos in `output.png`.

1.6 Tutorial 1: Learn about Configs in MMagic

We incorporate modular and inheritance design into our config system, which is convenient to conduct various experiments. If you wish to inspect the config file, you may run `python tools/misc/print_config.py /PATH/TO/CONFIG` to see the complete config.

You can learn about the usage of our config system according to the following tutorials.

1.6.1 Modify config through script arguments

When submitting jobs using `tools/train.py` or `tools/test.py`, you may specify `--cfg-options` to in-place modify the config.

- Update config keys of dict chains.

  The config options can be specified following the order of the dict keys in the original config. For example, `--cfg-options test_cfg.use_ema=False` changes the default sampling model to the original generator, and `--cfg-options train_dataloader.batch_size=8` changes the batch size of train dataloader.

- Update keys inside a list of configs.

  Some config dicts are composed as a list in your config. For example, the training pipeline `train_dataloader.dataset.pipeline` is normally a list e.g. `[dict(type='LoadImageFromFile'), ...]`. If you want to change `'LoadImageFromFile'` to `'LoadImageFromWebcam'` in the pipeline, you may specify `--cfg-options train_dataloader.dataset.pipeline.0.type=LoadImageFromWebcam`. The training pipeline `train_pipeline` is normally a list e.g. `[dict(type='LoadImageFromFile'), ...]`. If you want to change `'LoadImageFromFile'` to `'LoadMask'` in the pipeline, you may specify `--cfg-options train_pipeline.0.type=LoadMask`.

- Update values of list/tuples.

  If the value to be updated is a list or a tuple. You can set `--cfg-options key="[a,b]"` or `--cfg-options key=a,b`. It also allows nested list/tuple values, e.g., `--cfg-options key="[(a,b),(c,d)]"`. Note that the quotation mark " is necessary to support list/tuple data types, and that **NO** white space is allowed inside the quotation marks in the specified value.

1.6.2 Config file structure

There are 3 basic component types under `config/_base_`: datasets, models and default_runtime. Many methods could be easily constructed with one of each like AOT-GAN, EDVR, GLEAN, StyleGAN2, CycleGAN, SinGAN, etc. Configs consisting of components from `_base_` are called *primitive*.

For all configs under the same folder, it is recommended to have only **one** *primitive* config. All other configs should inherit from the *primitive* config. In this way, the maximum of inheritance level is 3.

For easy understanding, we recommend contributors to inherit from existing methods. For example, if some modification is made base on BasicVSR, user may first inherit the basic BasicVSR structure by specifying `_base_ = ../basicvsr/basicvsr_reds4.py`, then modify the necessary fields in the config files. If some modification is made base on StyleGAN2, user may first inherit the basic StyleGAN2 structure by specifying `_base_ = ../styleganv2/stylegan2_c2_ffhq_256_b4x8_800k.py`, then modify the necessary fields in the config files.

If you are building an entirely new method that does not share the structure with any of the existing methods, you may create a folder `xxx` under `configs`,

Please refer to MMEngine for detailed documentation.

1.6.3 Config name style

```
{model}_[module setting]_{training schedule}_{dataset}
```

`{xxx}` is required field and `[yyy]` is optional.

- `{model}`: model type like `stylegan`, `dcgan`, `basicvsr`, `dim`, etc. Settings referred in the original paper are included in this field as well (e.g., `Stylegan2-config-f`, `edvrm` of `edvrm_8xb4-600k_reds`.)
- `[module setting]`: specific setting for some modules, including Encoder, Decoder, Generator, Discriminator, Normalization, loss, Activation, etc. E.g. `c64n7` of `basicvsr-pp_c64n7_8xb1-600k_reds4`, learning rate `Glr4e-4_Dlr1e-4` for dcgan, `gamma32.8` for stylegan3, `woReLUInplace` in sagan. In this section, information from different submodules (e.g., generator and discriminator) are connected with `_`.
- `{training_scheduler}`: specific setting for training, including batch_size, schedule, etc. For example, learning rate (e.g., `lr1e-3`), number of gpu and batch size is used (e.g., `8xb32`), and total iterations (e.g., `160kiter`) or number of images shown in the discriminator (e.g., `12Mimgs`).
- `{dataset}`: dataset name and data size info like `celeba-256x256` of `deepfillv1_4xb4_celeba-256x256`, `reds4` of `basicvsr_2xb4_reds4`, `ffhq`, `lsun-car`, `celeba-hq`.

1.6.4 An example of EDSR

To help the users have a basic idea of a complete config, we make a brief comments on the config of the EDSR model we implemented as the following. For more detailed usage and the corresponding alternative for each modules, please refer to the API documentation and the tutorial in MMEngine.

Model config

In MMagic's config, we use model fields to set up a model.

```
model = dict(
    type='BaseEditModel',  # Name of the model
    generator=dict(  # Config of the generator
        type='EDSRNet',  # Type of the generator
        in_channels=3,  # Channel number of inputs
        out_channels=3,  # Channel number of outputs
        mid_channels=64,  # Channel number of intermediate features
        num_blocks=16,  # Block number in the trunk network
        upscale_factor=scale, # Upsampling factor
        res_scale=1,  # Used to scale the residual in residual block
        rgb_mean=(0.4488, 0.4371, 0.4040),  # Image mean in RGB orders
        rgb_std=(1.0, 1.0, 1.0)),  # Image std in RGB orders
    pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean')  # Config for␣
↪pixel loss
    train_cfg=dict(),  # Config of training model.
    test_cfg=dict(),  # Config of testing model.
    data_preprocessor=dict(  # The Config to build data preprocessor
```

(continues on next page)

(continued from previous page)

```
        type='DataPreprocessor', mean=[0., 0., 0.], std=[255., 255.,
                                                         255.]))
```

Data config

Dataloaders are required for the training, validation, and testing of the runner. Dataset and data pipeline need to be set to build the dataloader. Due to the complexity of this part, we use intermediate variables to simplify the writing of dataloader configs.

Data pipeline

```
train_pipeline = [  # Training data processing pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find the corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find the corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='SetValues', dictionary=dict(scale=scale)),  # Set value to destination␣
↪keys
    dict(type='PairedRandomCrop', gt_patch_size=96),  # Paired random crop
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='horizontal'),  # Flip direction
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='vertical'),  # Flip direction
    dict(type='RandomTransposeHW',  # Random transpose h and w for images
        keys=['lq', 'gt'],  # Images to be transposed
        transpose_ratio=0.5  # Transpose ratio
        ),
    dict(type='PackInputs')  # The config of collecting data from the current pipeline
]
test_pipeline = [  # Test pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
```

(continues on next page)

(continued from previous page)

```
        imdecode_backend='cv2'),  # decode backend
    dict(type='PackInputs')  # The config of collecting data from the current pipeline
]
```

Dataloader

```
dataset_type = 'BasicImageDataset'  # The type of dataset
data_root = 'data'  # Root path of data
train_dataloader = dict(
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    sampler=dict(type='InfiniteSampler', shuffle=True),  # The type of data sampler
    dataset=dict(  # Train dataset config
        type=dataset_type,  # Type of dataset
        ann_file='meta_info_DIV2K800sub_GT.txt',  # Path of annotation file
        metainfo=dict(dataset_type='div2k', task_name='sisr'),
        data_root=data_root + '/DIV2K',  # Root path of data
        data_prefix=dict(  # Prefix of image path
            img='DIV2K_train_LR_bicubic/X2_sub', gt='DIV2K_train_HR_sub'),
        filename_tmpl=dict(img='{}', gt='{}'),  # Filename template
        pipeline=train_pipeline))
val_dataloader = dict(
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    drop_last=False,  # Whether drop the last incomplete batch
    sampler=dict(type='DefaultSampler', shuffle=False),  # The type of data sampler
    dataset=dict(  # Validation dataset config
        type=dataset_type,  # Type of dataset
        metainfo=dict(dataset_type='set5', task_name='sisr'),
        data_root=data_root + '/Set5',  # Root path of data
        data_prefix=dict(img='LRbicx2', gt='GTmod12'),  # Prefix of image path
        pipeline=test_pipeline))
test_dataloader = val_dataloader
```

Evaluation config

Evaluators are used to compute the metrics of the trained model on the validation and testing datasets. The config of evaluators consists of one or a list of metric configs:

```
val_evaluator = [
    dict(type='MAE'),  # The name of metrics to evaluate
    dict(type='PSNR', crop_border=scale),  # The name of metrics to evaluate
    dict(type='SSIM', crop_border=scale),  # The name of metrics to evaluate
]
test_evaluator = val_evaluator # The config for testing evaluator
```

Training and testing config

MMEngine's runner uses Loop to control the training, validation, and testing processes. Users can set the maximum training iteration and validation intervals with these fields.

```
train_cfg = dict(
    type='IterBasedTrainLoop',  # The name of train loop type
    max_iters=300000,  # The number of total iterations
    val_interval=5000,  # The number of validation interval iterations
)
val_cfg = dict(type='ValLoop')  # The name of validation loop type
test_cfg = dict(type='TestLoop')  # The name of test loop type
```

Optimization config

`optim_wrapper` is the field to configure optimization related settings. The optimizer wrapper not only provides the functions of the optimizer, but also supports functions such as gradient clipping, mixed precision training, etc. Find more in optimizer wrapper tutorial.

```
optim_wrapper = dict(
    dict(
        type='OptimWrapper',
        optimizer=dict(type='Adam', lr=0.00001),
    )
)  # Config used to build optimizer, support all the optimizers in PyTorch whose␣
↪arguments are also the same as those in PyTorch.
```

`param_scheduler` is a field that configures methods of adjusting optimization hyper-parameters such as learning rate and momentum. Users can combine multiple schedulers to create a desired parameter adjustment strategy. Find more in parameter scheduler tutorial.

```
param_scheduler = dict(  # Config of learning policy
    type='MultiStepLR', by_epoch=False, milestones=[200000], gamma=0.5)
```

Hook config

Users can attach hooks to training, validation, and testing loops to insert some operations during running. There are two different hook fields, one is `default_hooks` and the other is `custom_hooks`.

`default_hooks` is a dict of hook configs. `default_hooks` are the hooks must required at runtime. They have default priority which should not be modified. If not set, runner will use the default values. To disable a default hook, users can set its config to `None`.

```
default_hooks = dict(  # Used to build default hooks
    checkpoint=dict(  # Config to set the checkpoint hook
        type='CheckpointHook',
        interval=5000,  # The save interval is 5000 iterations
        save_optimizer=True,
        by_epoch=False,  # Count by iterations
        out_dir=save_dir,
    ),
    timer=dict(type='IterTimerHook'),
```

(continues on next page)

(continued from previous page)

```
    logger=dict(type='LoggerHook', interval=100),  # Config to register logger hook
    param_scheduler=dict(type='ParamSchedulerHook'),
    sampler_seed=dict(type='DistSamplerSeedHook'),
)
```

custom_hooks is a list of hook configs. Users can develop there own hooks and insert them in this field.

```
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)] # Config of␣
↪visualization hook
```

Runtime config

```
default_scope = 'mmagic' # Used to set registries location
env_cfg = dict(  # Parameters to setup distributed training, the port can also be set
    cudnn_benchmark=False,
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=4),
    dist_cfg=dict(backend='nccl'),
)
log_level = 'INFO'  # The level of logging
log_processor = dict(type='LogProcessor', window_size=100, by_epoch=False)  # Used to␣
↪build log processor
load_from = None  # load models as a pre-trained model from a given path. This will not␣
↪resume training.
resume = False  # Resume checkpoints from a given path, the training will be resumed␣
↪from the epoch when the checkpoint's is saved.
```

1.6.5 An example of StyleGAN2

Taking Stylegan2 at 1024x1024 scale as an example, we introduce each field in the config according to different function modules.

Model config

In addition to neural network components such as generator, discriminator etc, it also requires data_preprocessor, loss_config, and some of them contains ema_config. data_preprocessor is responsible for processing a batch of data output by dataloader. loss_config is responsible for weight of loss terms. ema_config is responsible for exponential moving average (EMA) operation for generator.

```
model = dict(
    type='StyleGAN2',  # The name of the model
    data_preprocessor=dict(type='DataPreprocessor'),  # The config of data preprocessor,␣
↪usually includs image normalization and padding
    generator=dict(  # The config for generator
        type='StyleGANv2Generator',  # The name of the generator
        out_size=1024,  # The output resolution of the generator
        style_channels=512),  # The number of style channels of the generator
    discriminator=dict(  # The config for discriminator
        type='StyleGAN2Discriminator',  # The name of the discriminator
        in_size=1024),  # The input resolution of the discriminator
```

(continues on next page)

(continued from previous page)

```
    ema_config=dict(  # The config for EMA
        type='ExponentialMovingAverage',  # Specific the type of Average model
        interval=1,  # The interval of EMA operation
        momentum=0.9977843871238888),  # The momentum of EMA operation
    loss_config=dict(  # The config for loss terms
        r1_loss_weight=80.0,  # The weight for r1 gradient penalty
        r1_interval=16,  # The interval of r1 gradient penalty
        norm_mode='HWC',  # The normalization mode for r1 gradient penalty
        g_reg_interval=4,  # The interval for generator's regularization
        g_reg_weight=8.0,  # The weight for generator's regularization
        pl_batch_shrink=2))  # The factor of shrinking the batch size in path length␣
↪regularization
```

Dataset and evaluator config

Dataloaders are required for the training, validation, and testing of the runner. Dataset and data pipeline need to be set to build the dataloader. Due to the complexity of this part, we use intermediate variables to simplify the writing of dataloader configs.

```
dataset_type = 'BasicImageDataset'  # Dataset type, this will be used to define the␣
↪dataset
data_root = './data/ffhq/'  # Root path of data

train_pipeline = [  # Training data process pipeline
    dict(type='LoadImageFromFile', key='img'),  # First pipeline to load images from␣
↪file path
    dict(type='Flip', keys=['img'], direction='horizontal'),  # Argumentation pipeline␣
↪that flip the images
    dict(type='PackInputs', keys=['img'])  # The last pipeline that formats the␣
↪annotation data (if have) and decides which keys in the data should be packed into␣
↪data_samples
]
val_pipeline = [
    dict(type='LoadImageFromFile', key='img'),  # First pipeline to load images from␣
↪file path
    dict(type='PackInputs', keys=['img'])  # The last pipeline that formats the␣
↪annotation data (if have) and decides which keys in the data should be packed into␣
↪data_samples
]
train_dataloader = dict(  # The config of train dataloader
    batch_size=4,  # Batch size of a single GPU
    num_workers=8,  # Worker to pre-fetch data for each single GPU
    persistent_workers=True,  # If ``True``, the dataloader will not shutdown the worker␣
↪processes after an epoch end, which can accelerate training speed.
    sampler=dict(  # The config of training data sampler
        type='InfiniteSampler',  # InfiniteSampler for iteratiion-based training. Refers␣
↪to https://github.com/open-mmlab/mmengine/blob/
↪fe0eb0a5bbc8bf816d5649bfdd34908c258eb245/mmengine/dataset/sampler.py#L107
        shuffle=True),  # Whether randomly shuffle the training data
    dataset=dict(  # The config of the training dataset
        type=dataset_type,
```

(continues on next page)

(continued from previous page)

```
        data_root=data_root,
        pipeline=train_pipeline))
val_dataloader = dict(  # The config of validation dataloader
    batch_size=4,  # Batch size of a single GPU
    num_workers=8,  # Worker to pre-fetch data for each single GPU
    dataset=dict(  # The config of the validation dataset
        type=dataset_type,
        data_root=data_root,
        pipeline=val_pipeline),
    sampler=dict(  # The config of validatioin data sampler
        type='DefaultSampler',  # DefaultSampler which supports both distributed and non-
↪distributed training. Refer to https://github.com/open-mmlab/mmengine/blob/
↪fe0eb0a5bbc8bf816d5649bfdd34908c258eb245/mmengine/dataset/sampler.py#L14
        shuffle=False),  # Whether randomly shuffle the validation data
    persistent_workers=True)
test_dataloader = val_dataloader  # The config of the testing dataloader
```

Evaluators are used to compute the metrics of the trained model on the validation and testing datasets. The config of evaluators consists of one or a list of metric configs:

```
val_evaluator = dict(  # The config for validation evaluator
    type='Evaluator',  # The type of evaluation
    metrics=[  # The config for metrics
        dict(
            type='FrechetInceptionDistance',
            prefix='FID-Full-50k',
            fake_nums=50000,
            inception_style='StyleGAN',
            sample_model='ema'),
        dict(type='PrecisionAndRecall', fake_nums=50000, prefix='PR-50K'),
        dict(type='PerceptualPathLength', fake_nums=50000, prefix='ppl-w')
    ])
test_evaluator = val_evaluator  # The config for testing evaluator
```

Training and testing config

MMEngine's runner uses Loop to control the training, validation, and testing processes. Users can set the maximum training iteration and validation intervals with these fields.

```
train_cfg = dict(  # The config for training
    by_epoch=False,  # Set `by_epoch` as False to use iteration-based training
    val_begin=1,  # Which iteration to start the validation
    val_interval=10000,  # Validation intervals
    max_iters=800002)  # Maximum training iterations
val_cfg = dict(type='MultiValLoop')  # The validation loop type
test_cfg = dict(type='MultiTestLoop')  # The testing loop type
```

Optimization config

`optim_wrapper` is the field to configure optimization related settings. The optimizer wrapper not only provides the functions of the optimizer, but also supports functions such as gradient clipping, mixed precision training, etc. Find more in optimizer wrapper tutorial.

```
optim_wrapper = dict(
    constructor='MultiOptimWrapperConstructor',
    generator=dict(
        optimizer=dict(type='Adam', lr=0.0016, betas=(0, 0.9919919678228657))),
    discriminator=dict(
        optimizer=dict(
            type='Adam',
            lr=0.0018823529411764706,
            betas=(0, 0.9905854573074332))))
```

`param_scheduler` is a field that configures methods of adjusting optimization hyperparameters such as learning rate and momentum. Users can combine multiple schedulers to create a desired parameter adjustment strategy. Find more in parameter scheduler tutorial. Since StyleGAN2 do not use parameter scheduler, we use config in CycleGAN as an example:

```
# parameter scheduler in CycleGAN config
param_scheduler = dict(
    type='LinearLrInterval',  # The type of scheduler
    interval=400,  # The interval to update the learning rate
    by_epoch=False,  # The scheduler is called by iteration
    start_factor=0.0002,  # The number we multiply parameter value in the first iteration
    end_factor=0,  # The number we multiply parameter value at the end of linear␣
↪changing process.
    begin=40000,  # The start iteration of the scheduler
    end=80000)  # The end iteration of the scheduler
```

Hook config

Users can attach hooks to training, validation, and testing loops to insert some operations during running. There are two different hook fields, one is `default_hooks` and the other is `custom_hooks`.

`default_hooks` is a dict of hook configs. `default_hooks` are the hooks must required at runtime. They have default priority which should not be modified. If not set, runner will use the default values. To disable a default hook, users can set its config to `None`.

```
default_hooks = dict(
    timer=dict(type='IterTimerHook'),
    logger=dict(type='LoggerHook', interval=100, log_metric_by_epoch=False),
    checkpoint=dict(
        type='CheckpointHook',
        interval=10000,
        by_epoch=False,
        less_keys=['FID-Full-50k/fid'],
        greater_keys=['IS-50k/is'],
        save_optimizer=True,
        save_best='FID-Full-50k/fid'))
```

`custom_hooks` is a list of hook configs. Users can develop there own hooks and insert them in this field.

```
custom_hooks = [
    dict(
        type='VisualizationHook',
        interval=5000,
        fixed_input=True,
        vis_kwargs_list=dict(type='GAN', name='fake_img'))
]
```

Runtime config

```
default_scope = 'mmagic'  # The default registry scope to find modules. Refer to https://
↪mmengine.readthedocs.io/en/latest/advanced_tutorials/registry.html

# config for environment
env_cfg = dict(
    cudnn_benchmark=True,  # whether to enable cudnn benchmark.
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=0),  # set multi process␣
↪parameters.
    dist_cfg=dict(backend='nccl'),  # set distributed parameters.
)

log_level = 'INFO'  # The level of logging
log_processor = dict(
    type='LogProcessor',  # log processor to process runtime logs
    by_epoch=False)  # print log by iteration
load_from = None  # load model checkpoint as a pre-trained model for a given path
resume = False  # Whether to resume from the checkpoint define in `load_from`. If `load_
↪from` is `None`, it will resume the latest checkpoint in `work_dir`
```

1.6.6 Other examples

An example of config system for inpainting

To help the users have a basic idea of a complete config and the modules in a inpainting system, we make brief comments on the config of Global&Local as the following. For more detailed usage and the corresponding alternative for each modules, please refer to the API documentation.

```
model = dict(
    type='GLInpaintor', # The name of inpaintor
    data_preprocessor=dict(
        type='DataPreprocessor', # The name of data preprocessor
        mean=[127.5], # Mean value used in data normalization
        std=[127.5], # Std value used in data normalization
    ),
    encdec=dict(
        type='GLEncoderDecoder', # The name of encoder-decoder
        encoder=dict(type='GLEncoder', norm_cfg=dict(type='SyncBN')), # The config of␣
↪encoder
        decoder=dict(type='GLDecoder', norm_cfg=dict(type='SyncBN')), # The config of␣
↪decoder
```

(continues on next page)

(continued from previous page)

```
        dilation_neck=dict(
            type='GLDilationNeck', norm_cfg=dict(type='SyncBN'))), # The config of␣
↪dilation neck
    disc=dict(
        type='GLDiscs', # The name of discriminator
        global_disc_cfg=dict(
            in_channels=3, # The input channel of discriminator
            max_channels=512, # The maximum middle channel in discriminator
            fc_in_channels=512 * 4 * 4, # The input channel of last fc layer
            fc_out_channels=1024, # The output channel of last fc channel
            num_convs=6, # The number of convs used in discriminator
            norm_cfg=dict(type='SyncBN') # The config of norm layer
        ),
        local_disc_cfg=dict(
            in_channels=3, # The input channel of discriminator
            max_channels=512, # The maximum middle channel in discriminator
            fc_in_channels=512 * 4 * 4, # The input channel of last fc layer
            fc_out_channels=1024, # The output channel of last fc channel
            num_convs=5, # The number of convs used in discriminator
            norm_cfg=dict(type='SyncBN') # The config of norm layer
        ),
    ),
    loss_gan=dict(
        type='GANLoss', # The name of GAN loss
        gan_type='vanilla', # The type of GAN loss
        loss_weight=0.001 # The weight of GAN loss
    ),
    loss_l1_hole=dict(
        type='L1Loss', # The type of l1 loss
        loss_weight=1.0 # The weight of l1 loss
    ))

train_cfg = dict(
    type='IterBasedTrainLoop',# The name of train loop type
    max_iters=500002, # The number of total iterations
    val_interval=50000, # The number of validation interval iterations
)
val_cfg = dict(type='ValLoop') # The name of validation loop type
test_cfg = dict(type='TestLoop') # The name of test loop type

val_evaluator = [
    dict(type='MAE', mask_key='mask', scaling=100), # The name of metrics to evaluate
    dict(type='PSNR'), # The name of metrics to evaluate
    dict(type='SSIM'), # The name of metrics to evaluate
]
test_evaluator = val_evaluator

input_shape = (256, 256) # The shape of input image

train_pipeline = [
    dict(type='LoadImageFromFile', key='gt'), # The config of loading image
    dict(
```

(continues on next page)

(continued from previous page)

```
        type='LoadMask', # The type of loading mask pipeline
        mask_mode='bbox', # The type of mask
        mask_config=dict(
            max_bbox_shape=(128, 128), # The shape of bbox
            max_bbox_delta=40, # The changing delta of bbox height and width
            min_margin=20,  # The minimum margin from bbox to the image border
            img_shape=input_shape)),  # The input image shape
    dict(
        type='Crop', # The type of crop pipeline
        keys=['gt'],  # The keys of images to be cropped
        crop_size=(384, 384),  # The size of cropped patch
        random_crop=True,  # Whether to use random crop
    ),
    dict(
        type='Resize',  # The type of resizing pipeline
        keys=['gt'],  # They keys of images to be resized
        scale=input_shape,  # The scale of resizing function
        keep_ratio=False,  # Whether to keep ratio during resizing
    ),
    dict(
        type='Normalize',  # The type of normalizing pipeline
        keys=['gt_img'],  # The keys of images to be normed
        mean=[127.5] * 3,  # Mean value used in normalization
        std=[127.5] * 3,  # Std value used in normalization
        to_rgb=False),  # Whether to transfer image channels to rgb
    dict(type='GetMaskedImage'), # The config of getting masked image pipeline
    dict(type='PackInputs'), # The config of collecting data from the current pipeline
]

test_pipeline = train_pipeline  # Constructing testing/validation pipeline

dataset_type = 'BasicImageDataset' # The type of dataset
data_root = 'data/places'  # Root path of data

train_dataloader = dict(
    batch_size=12, # Batch size of a single GPU
    num_workers=4, # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False, # Whether maintain the workers Dataset instances alive
    sampler=dict(type='InfiniteSampler', shuffle=False), # The type of data sampler
    dataset=dict(  # Train dataset config
        type=dataset_type, # Type of dataset
        data_root=data_root, # Root path of data
        data_prefix=dict(gt='data_large'), # Prefix of image path
        ann_file='meta/places365_train_challenge.txt', # Path of annotation file
        test_mode=False,
        pipeline=train_pipeline,
    ))

val_dataloader = dict(
    batch_size=1, # Batch size of a single GPU
    num_workers=4, # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False, # Whether maintain the workers Dataset instances alive
```

(continues on next page)

(continued from previous page)

```
    drop_last=False, # Whether drop the last incomplete batch
    sampler=dict(type='DefaultSampler', shuffle=False), # The type of data sampler
    dataset=dict( # Validation dataset config
        type=dataset_type, # Type of dataset
        data_root=data_root, # Root path of data
        data_prefix=dict(gt='val_large'), # Prefix of image path
        ann_file='meta/places365_val.txt', # Path of annotation file
        test_mode=True,
        pipeline=test_pipeline,
    ))

test_dataloader = val_dataloader

model_wrapper_cfg = dict(type='MMSeparateDistributedDataParallel') # The name of model␣
↪wrapper

optim_wrapper = dict( # Config used to build optimizer, support all the optimizers in␣
↪PyTorch whose arguments are also the same as those in PyTorch
    constructor='MultiOptimWrapperConstructor',
    generator=dict(
        type='OptimWrapper', optimizer=dict(type='Adam', lr=0.0004)),
    disc=dict(type='OptimWrapper', optimizer=dict(type='Adam', lr=0.0004)))

default_scope = 'mmagic' # Used to set registries location
save_dir = './work_dirs' # Directory to save the model checkpoints and logs for the␣
↪current experiments
exp_name = 'gl_places'  # The experiment name

default_hooks = dict( # Used to build default hooks
    timer=dict(type='IterTimerHook'),
    logger=dict(type='LoggerHook', interval=100), # Config to register logger hook
    param_scheduler=dict(type='ParamSchedulerHook'),
    checkpoint=dict( # Config to set the checkpoint hook
        type='CheckpointHook',
        interval=50000,
        by_epoch=False,
        out_dir=save_dir),
    sampler_seed=dict(type='DistSamplerSeedHook'),
)

env_cfg = dict( # Parameters to setup distributed training, the port can also be set
    cudnn_benchmark=False,
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=0),
    dist_cfg=dict(backend='nccl'),
)

vis_backends = [dict(type='LocalVisBackend')] # The name of visualization backend
visualizer = dict( # Config used to build visualizer
    type='ConcatImageVisualizer',
    vis_backends=vis_backends,
    fn_key='gt_path',
    img_keys=['gt_img', 'input', 'pred_img'],
```

(continues on next page)

(continued from previous page)

```
    bgr2rgb=True)
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)] # Used to build custom␣
↪hooks

log_level = 'INFO' # The level of logging
log_processor = dict(type='LogProcessor', by_epoch=False) # Used to build log processor

load_from = None # load models as a pre-trained model from a given path. This will not␣
↪resume training.
resume = False # Resume checkpoints from a given path, the training will be resumed from␣
↪the epoch when the checkpoint's is saved.

find_unused_parameters = False  # Whether to set find unused parameters in ddp
```

An example of config system for matting

To help the users have a basic idea of a complete config, we make a brief comments on the config of the original DIM model we implemented as the following. For more detailed usage and the corresponding alternative for each modules, please refer to the API documentation.

```
# model settings
model = dict(
    type='DIM',  # The name of model (we call mattor).
    data_preprocessor=dict(  # The Config to build data preprocessor
        type='DataPreprocessor',
        mean=[123.675, 116.28, 103.53],
        std=[58.395, 57.12, 57.375],
        bgr_to_rgb=True,
        proc_inputs='normalize',
        proc_trimap='rescale_to_zero_one',
        proc_gt='rescale_to_zero_one',
    ),
    backbone=dict(  # The config of the backbone.
        type='SimpleEncoderDecoder',  # The type of the backbone.
        encoder=dict(  # The config of the encoder.
            type='VGG16'),  # The type of the encoder.
        decoder=dict(  # The config of the decoder.
            type='PlainDecoder')),  # The type of the decoder.
    pretrained='./weights/vgg_state_dict.pth',  # The pretrained weight of the encoder␣
↪to be loaded.
    loss_alpha=dict(  # The config of the alpha loss.
        type='CharbonnierLoss',  # The type of the loss for predicted alpha matte.
        loss_weight=0.5),  # The weight of the alpha loss.
    loss_comp=dict(  # The config of the composition loss.
        type='CharbonnierCompLoss',  # The type of the composition loss.
        loss_weight=0.5), # The weight of the composition loss.
    train_cfg=dict(  # Config of training DIM model.
        train_backbone=True,  # In DIM stage1, backbone is trained.
        train_refiner=False),  # In DIM stage1, refiner is not trained.
    test_cfg=dict(  # Config of testing DIM model.
        refine=False,  # Whether use refiner output as output, in stage1, we don't use it.
```

(continues on next page)

(continued from previous page)

```
        resize_method='pad',
        resize_mode='reflect',
        size_divisor=32,
    ),
)

# data settings
dataset_type = 'AdobeComp1kDataset'  # Dataset type, this will be used to define the␣
↪dataset.
data_root = 'data/adobe_composition-1k'  # Root path of data.

train_pipeline = [  # Training data processing pipeline.
    dict(
        type='LoadImageFromFile',  # Load alpha matte from file.
        key='alpha',  # Key of alpha matte in annotation file. The pipeline will read␣
↪alpha matte from path `alpha_path`.
        color_type='grayscale'),  # Load as grayscale image which has shape (height,␣
↪width).
    dict(
        type='LoadImageFromFile',  # Load image from file.
        key='fg'),  # Key of image to load. The pipeline will read fg from path `fg_path`.
    dict(
        type='LoadImageFromFile',  # Load image from file.
        key='bg'),  # Key of image to load. The pipeline will read bg from path `bg_path`.
    dict(
        type='LoadImageFromFile',  # Load image from file.
        key='merged'),  # Key of image to load. The pipeline will read merged from path␣
↪`merged_path`.
    dict(
        type='CropAroundUnknown',  # Crop images around unknown area (semi-transparent␣
↪area).
        keys=['alpha', 'merged', 'fg', 'bg'],  # Images to crop.
        crop_sizes=[320, 480, 640]),  # Candidate crop size.
    dict(
        type='Flip',  # Augmentation pipeline that flips the images.
        keys=['alpha', 'merged', 'fg', 'bg']),  # Images to be flipped.
    dict(
        type='Resize',  # Augmentation pipeline that resizes the images.
        keys=['alpha', 'merged', 'fg', 'bg'],  # Images to be resized.
        scale=(320, 320),  # Target size.
        keep_ratio=False),  # Whether to keep the ratio between height and width.
    dict(
        type='GenerateTrimap',  # Generate trimap from alpha matte.
        kernel_size=(1, 30)),  # Kernel size range of the erode/dilate kernel.
    dict(type='PackInputs'),  # The config of collecting data from the current pipeline
]
test_pipeline = [
    dict(
        type='LoadImageFromFile',  # Load alpha matte.
        key='alpha',  # Key of alpha matte in annotation file. The pipeline will read␣
↪alpha matte from path `alpha_path`.
        color_type='grayscale',
```

(continues on next page)

(continued from previous page)

```
        save_original_img=True),
    dict(
        type='LoadImageFromFile',  # Load image from file
        key='trimap',  # Key of image to load. The pipeline will read trimap from path␣
↪`trimap_path`.
        color_type='grayscale',  # Load as grayscale image which has shape (height,␣
↪width).
        save_original_img=True),  # Save a copy of trimap for calculating metrics. It␣
↪will be saved with key `ori_trimap`
    dict(
        type='LoadImageFromFile',  # Load image from file
        key='merged'),  # Key of image to load. The pipeline will read merged from path␣
↪`merged_path`.
    dict(type='PackInputs'),  # The config of collecting data from the current pipeline
]

train_dataloader = dict(
    batch_size=1,  # Batch size of a single GPU
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    sampler=dict(type='InfiniteSampler', shuffle=True),  # The type of data sampler
    dataset=dict(  # Train dataset config
        type=dataset_type,  # Type of dataset
        data_root=data_root,  # Root path of data
        ann_file='training_list.json',  # Path of annotation file
        test_mode=False,
        pipeline=train_pipeline,
    ))

val_dataloader = dict(
    batch_size=1,  # Batch size of a single GPU
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    drop_last=False,  # Whether drop the last incomplete batch
    sampler=dict(type='DefaultSampler', shuffle=False),  # The type of data sampler
    dataset=dict(  # Validation dataset config
        type=dataset_type,  # Type of dataset
        data_root=data_root,  # Root path of data
        ann_file='test_list.json',  # Path of annotation file
        test_mode=True,
        pipeline=test_pipeline,
    ))

test_dataloader = val_dataloader

val_evaluator = [
    dict(type='SAD'),  # The name of metrics to evaluate
    dict(type='MattingMSE'),  # The name of metrics to evaluate
    dict(type='GradientError'),  # The name of metrics to evaluate
    dict(type='ConnectivityError'),  # The name of metrics to evaluate
]
test_evaluator = val_evaluator
```

(continues on next page)

(continued from previous page)

```
train_cfg = dict(
    type='IterBasedTrainLoop',  # The name of train loop type
    max_iters=1_000_000,  # The number of total iterations
    val_interval=40000,  # The number of validation interval iterations
)
val_cfg = dict(type='ValLoop')  # The name of validation loop type
test_cfg = dict(type='TestLoop')  # The name of test loop type

# optimizer
optim_wrapper = dict(
    dict(
        type='OptimWrapper',
        optimizer=dict(type='Adam', lr=0.00001),
    )
)  # Config used to build optimizer, support all the optimizers in PyTorch whose␣
↪arguments are also the same as those in PyTorch.

default_scope = 'mmagic'  # Used to set registries location
save_dir = './work_dirs'  # Directory to save the model checkpoints and logs for the␣
↪current experiments.

default_hooks = dict(  # Used to build default hooks
    timer=dict(type='IterTimerHook'),
    logger=dict(type='LoggerHook', interval=100),  # Config to register logger hook
    param_scheduler=dict(type='ParamSchedulerHook'),
    checkpoint=dict(  # Config to set the checkpoint hook
        type='CheckpointHook',
        interval=40000,  # The save interval is 40000 iterations.
        by_epoch=False,  # Count by iterations.
        out_dir=save_dir),
    sampler_seed=dict(type='DistSamplerSeedHook'),
)

env_cfg = dict(  # Parameters to setup distributed training, the port can also be set
    cudnn_benchmark=False,
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=4),
    dist_cfg=dict(backend='nccl'),
)

log_level = 'INFO'  # The level of logging
log_processor = dict(type='LogProcessor', by_epoch=False)  # Used to build log processor

load_from = None  # load models as a pre-trained model from a given path. This will not␣
↪resume training.
resume = False  # Resume checkpoints from a given path, the training will be resumed␣
↪from the epoch when the checkpoint's is saved.
```

An example of config system for restoration

To help the users have a basic idea of a complete config, we make a brief comments on the config of the EDSR model we implemented as the following. For more detailed usage and the corresponding alternative for each modules, please refer to the API documentation.

```
exp_name = 'edsr_x2c64b16_1x16_300k_div2k'  # The experiment name
work_dir = f'./work_dirs/{experiment_name}'
save_dir = './work_dirs/'

load_from = None  # based on pre-trained x2 model

scale = 2  # Scale factor for upsampling
# model settings
model = dict(
    type='BaseEditModel',  # Name of the model
    generator=dict(  # Config of the generator
        type='EDSRNet',  # Type of the generator
        in_channels=3,  # Channel number of inputs
        out_channels=3,  # Channel number of outputs
        mid_channels=64,  # Channel number of intermediate features
        num_blocks=16,  # Block number in the trunk network
        upscale_factor=scale, # Upsampling factor
        res_scale=1,  # Used to scale the residual in residual block
        rgb_mean=(0.4488, 0.4371, 0.4040),  # Image mean in RGB orders
        rgb_std=(1.0, 1.0, 1.0)),  # Image std in RGB orders
    pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean')  # Config for␣
↪pixel loss
    train_cfg=dict(),  # Config of training model.
    test_cfg=dict(),  # Config of testing model.
    data_preprocessor=dict(  # The Config to build data preprocessor
        type='DataPreprocessor', mean=[0., 0., 0.], std=[255., 255.,
                                                         255.]))

train_pipeline = [  # Training data processing pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='SetValues', dictionary=dict(scale=scale)),  # Set value to destination␣
↪keys
    dict(type='PairedRandomCrop', gt_patch_size=96),  # Paired random crop
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='horizontal'),  # Flip direction
    dict(type='Flip',  # Flip images
```

(continues on next page)

(continued from previous page)

```
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='vertical'),  # Flip direction
    dict(type='RandomTransposeHW',  # Random transpose h and w for images
        keys=['lq', 'gt'],  # Images to be transposed
        transpose_ratio=0.5  # Transpose ratio
        ),
    dict(type='PackInputs')  # The config of collecting data from the current pipeline
]
test_pipeline = [  # Test pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='ToTensor', keys=['img', 'gt']),  # Convert images to tensor
    dict(type='PackInputs')  # The config of collecting data from the current pipeline
]

# dataset settings
dataset_type = 'BasicImageDataset'  # The type of dataset
data_root = 'data'  # Root path of data

train_dataloader = dict(
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    sampler=dict(type='InfiniteSampler', shuffle=True),  # The type of data sampler
    dataset=dict(  # Train dataset config
        type=dataset_type,  # Type of dataset
        ann_file='meta_info_DIV2K800sub_GT.txt',  # Path of annotation file
        metainfo=dict(dataset_type='div2k', task_name='sisr'),
        data_root=data_root + '/DIV2K',  # Root path of data
        data_prefix=dict(  # Prefix of image path
            img='DIV2K_train_LR_bicubic/X2_sub', gt='DIV2K_train_HR_sub'),
        filename_tmpl=dict(img='{}', gt='{}'),  # Filename template
        pipeline=train_pipeline))
val_dataloader = dict(
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    drop_last=False,  # Whether drop the last incomplete batch
    sampler=dict(type='DefaultSampler', shuffle=False),  # The type of data sampler
    dataset=dict(  # Validation dataset config
        type=dataset_type,  # Type of dataset
        metainfo=dict(dataset_type='set5', task_name='sisr'),
        data_root=data_root + '/Set5',  # Root path of data
        data_prefix=dict(img='LRbicx2', gt='GTmod12'),  # Prefix of image path
        pipeline=test_pipeline))
```

(continues on next page)

(continued from previous page)

```
test_dataloader = val_dataloader

val_evaluator = [
    dict(type='MAE'),  # The name of metrics to evaluate
    dict(type='PSNR', crop_border=scale),  # The name of metrics to evaluate
    dict(type='SSIM', crop_border=scale),  # The name of metrics to evaluate
]
test_evaluator = val_evaluator

train_cfg = dict(
    type='IterBasedTrainLoop', max_iters=300000, val_interval=5000)  # Config of train␣
↪loop type
val_cfg = dict(type='ValLoop')  # The name of validation loop type
test_cfg = dict(type='TestLoop')  # The name of test loop type

# optimizer
optim_wrapper = dict(
    dict(
        type='OptimWrapper',
        optimizer=dict(type='Adam', lr=0.00001),
    )
)  # Config used to build optimizer, support all the optimizers in PyTorch whose␣
↪arguments are also the same as those in PyTorch.

param_scheduler = dict(  # Config of learning policy
    type='MultiStepLR', by_epoch=False, milestones=[200000], gamma=0.5)

default_hooks = dict(  # Used to build default hooks
    checkpoint=dict(  # Config to set the checkpoint hook
        type='CheckpointHook',
        interval=5000,  # The save interval is 5000 iterations
        save_optimizer=True,
        by_epoch=False,  # Count by iterations
        out_dir=save_dir,
    ),
    timer=dict(type='IterTimerHook'),
    logger=dict(type='LoggerHook', interval=100),  # Config to register logger hook
    param_scheduler=dict(type='ParamSchedulerHook'),
    sampler_seed=dict(type='DistSamplerSeedHook'),
)

default_scope = 'mmagic'  # Used to set registries location
save_dir = './work_dirs'  # Directory to save the model checkpoints and logs for the␣
↪current experiments.

env_cfg = dict(  # Parameters to setup distributed training, the port can also be set
    cudnn_benchmark=False,
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=4),
    dist_cfg=dict(backend='nccl'),
)

log_level = 'INFO'  # The level of logging
```

(continues on next page)

(continued from previous page)

```
log_processor = dict(type='LogProcessor', window_size=100, by_epoch=False)  # Used to␣
↪build log processor

load_from = None  # load models as a pre-trained model from a given path. This will not␣
↪resume training.
resume = False  # Resume checkpoints from a given path, the training will be resumed␣
↪from the epoch when the checkpoint's is saved.
```

1.7 Tutorial 2: Prepare datasets

In this section, we will detail how to prepare data and adopt the proper dataset in our repo for different methods.

We support multiple datasets of different tasks. There are two ways to use datasets for training and testing models in MMagic:

1. Using downloaded datasets directly
2. Preprocessing downloaded datasets before using them.

The structure of this guide is as follows:

1.7.1 Download datasets

You are supposed to download datasets from their homepage first. Most datasets are available after downloaded, so you only need to make sure the folder structure is correct and further preparation is not necessary. For example, you can simply prepare Vimeo90K-triplet datasets by downloading datasets from homepage.

1.7.2 Prepare datasets

Some datasets need to be preprocessed before training or testing. We support many scripts to prepare datasets in tools/dataset_converters. And you can follow the tutorials of every dataset to run scripts. For example, we recommend cropping the DIV2K images to sub-images. We provide a script to prepare cropped DIV2K dataset. You can run the following command:

```
python tools/dataset_converters/div2k/preprocess_div2k_dataset.py --data-root ./data/
↪DIV2K
```

1.7.3 The overview of the datasets in MMagic

We support detailed tutorials and split them according to different tasks.

Please check our dataset zoo for data preparation of different tasks.

If you're interested in more details of datasets in MMagic, please check the *advanced guides*.

1.8 Tutorial 3: Inference with pre-trained models

MMagic provides Hign-level APIs for you to easily play with state-of-the-art models on your own images or videos.

In the new API, only two lines of code are needed to implement inference:

```
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance
editor = MMagicInferencer('pix2pix')
# Infer a image. Input image path and output image path is needed.
results = editor.infer(img='../resources/input/translation/gt_mask_0.png', result_out_
↪dir='../resources/output/translation/tutorial_translation_pix2pix_res.jpg')
```

MMagic supports various fundamental generative models, including:

unconditional Generative Adversarial Networks (GANs), conditional GANs, diffusion models, etc.

MMagic also supports various applications, including:

text-to-image, image-to-image translation, 3D-aware generation, image super-resolution, video super-resolution, video frame interpolation, image inpainting, image matting, image restoration, image colorization, image generation, etc.

In this section, we will specify how to play with our pre-trained models.

- *Tutorial 3: Inference with Pre-trained Models*
  - Prepare some images or videos for inference
  - Generative Models
    - Unconditional Generative Adversarial Networks (GANs)
    - Conditional Generative Adversarial Networks (GANs)
    - Diffusion Models
  - Applications
    - Text-to-Image
    - Image-to-image translation
    - 3D-aware generation
    - Image super-resolution
    - Video super-resolution
    - Video frame interpolation
    - Image inpainting
    - Image matting
    - Image restoration

    * Image colorization

- Previous Versions

1.8.1 Prepare some images or videos for inference

Please refer to our tutorials for details.

1.8.2 Generative Models

Unconditional Generative Adversarial Networks (GANs)

MMagic provides high-level APIs for sampling images with unconditional GANs. Unconditional GAN models do not need input, and output a image. We take 'styleganv1' as an example.

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
result_out_dir = './resources/output/unconditional/tutorial_unconditional_styleganv1_res.
↪png'
editor = MMagicInferencer('styleganv1')
results = editor.infer(result_out_dir=result_out_dir)
```

Indeed, we have already provided a more friendly demo script to users. You can use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name styleganv1 \
        --result-out-dir demo_unconditional_styleganv1_res.jpg
```

Conditional Generative Adversarial Networks (GANs)

MMagic provides high-level APIs for sampling images with conditional GANs. Conditional GAN models take a label as input and output a image. We take 'biggan' as an example..

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
result_out_dir = './resources/output/conditional/tutorial_conditinal_biggan_res.jpg'
editor = MMagicInferencer('biggan', model_setting=1)
results = editor.infer(label=1, result_out_dir=result_out_dir)
```

Indeed, we have already provided a more friendly demo script to users. You can use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name biggan \
```

(continues on next page)

(continued from previous page)

```
        --model-setting 1 \
        --label 1 \
        --result-out-dir demo_conditional_biggan_res.jpg
```

Diffusion Models

MMagic provides high-level APIs for sampling images with diffusion models. f

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
editor = MMagicInferencer(model_name='stable_diffusion')
text_prompts = 'A panda is having dinner at KFC'
result_out_dir = './resources/output/text2image/tutorial_text2image_sd_res.png'
editor.infer(text=text_prompts, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name stable_diffusion \
        --text "A panda is having dinner at KFC" \
        --result-out-dir demo_text2image_stable_diffusion_res.png
```

1.8.3 Applications

Text-to-Image

Text-to-image models take text as input, and output a image. We take 'controlnet-canny' as an example.

```
import cv2
import numpy as np
import mmcv
from mmengine import Config
from PIL import Image

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

register_all_modules()

cfg = Config.fromfile('configs/controlnet/controlnet-canny.py')
controlnet = MODELS.build(cfg.model).cuda()

control_url = 'https://user-images.githubusercontent.com/28132635/230288866-99603172-
↪04cb-47b3-8adb-d1aa532d1d2c.jpg'
control_img = mmcv.imread(control_url)
control = cv2.Canny(control_img, 100, 200)
control = control[:, :, None]
```

(continues on next page)

(continued from previous page)

```
control = np.concatenate([control] * 3, axis=2)
control = Image.fromarray(control)

prompt = 'Room with blue walls and a yellow ceiling.'

output_dict = controlnet.infer(prompt, control=control)
samples = output_dict['samples']
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name controlnet \
        --model-setting 1 \
        --text "Room with blue walls and a yellow ceiling." \
        --control 'https://user-images.githubusercontent.com/28132635/230297033-4f5c32df-
↪365c-4cf4-8e4f-1b76a4cbb0b7.png' \
        --result-out-dir demo_text2image_controlnet_canny_res.png
```

Image-to-image translation

MMagic provides high-level APIs for translating images by using image translation models. Here is an example of building Pix2Pix and obtaining the translated images.

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
editor = MMagicInferencer('pix2pix')
results = editor.infer(img=img_path, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name pix2pix \
        --img ${IMAGE_PATH} \
        --result-out-dir ${SAVE_PATH}
```

3D-aware generation

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
result_out_dir = './resources/output/eg3d-output'
editor = MMagicInferencer('eg3d')
results = editor.infer(result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
    --model-name eg3d \
    --result-out-dir ./resources/output/eg3d-output
```

Image super-resolution

Image super resolution models take a image as input, and output a high resolution image. We take 'esrgan' as an example.

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
img = './resources/input/restoration/0901x2.png'
result_out_dir = './resources/output/restoration/tutorial_restoration_esrgan_res.png'
editor = MMagicInferencer('esrgan')
results = editor.infer(img=img, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name esrgan \
        --img ${IMAGE_PATH} \
        --result-out-dir ${SAVE_PATH}
```

Video super-resolution

```
import os
from mmagic.apis import MMagicInferencer
from mmengine import mkdir_or_exist

# Create a MMagicInferencer instance and infer
video = './resources/input/video_interpolation/b-3LLDhc4EU_000000_000010.mp4'
result_out_dir = './resources/output/video_super_resolution/tutorial_video_super_
↪resolution_basicvsr_res.mp4'
mkdir_or_exist(os.path.dirname(result_out_dir))
editor = MMagicInferencer('basicvsr')
results = editor.infer(video=video, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name basicvsr \
        --video ./resources/input/video_restoration/QUuC4vJs_000084_000094_400x320.mp4 \
        --result-out-dir ./resources/output/video_restoration/demo_video_restoration_
↪basicvsr_res.mp4
```

Video frame interpolation

Video interpolation models take a video as input, and output a interpolated video. We take 'flavr' as an example.

```
import os
from mmagic.apis import MMagicInferencer
from mmengine import mkdir_or_exist

# Create a MMagicInferencer instance and infer
video = './resources/input/video_interpolation/b-3LLDhc4EU_000000_000010.mp4'
result_out_dir = './resources/output/video_interpolation/tutorial_video_interpolation_
↪flavr_res.mp4'
mkdir_or_exist(os.path.dirname(result_out_dir))
editor = MMagicInferencer('flavr')
results = editor.infer(video=video, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name flavr \
        --video ${VIDEO_PATH} \
        --result-out-dir ${SAVE_PATH}
```

Image inpainting

Inpaiting models take a masked image and mask pair as input, and output a inpainted image. We take 'global_local' as an example.

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

img = './resources/input/inpainting/celeba_test.png'
mask = './resources/input/inpainting/bbox_mask.png'

# Create a MMagicInferencer instance and infer
result_out_dir = './resources/output/inpainting/tutorial_inpainting_global_local_res.jpg'
editor = MMagicInferencer('global_local', model_setting=1)
results = editor.infer(img=img, mask=mask, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name global_local  \
        --img ./resources/input/inpainting/celeba_test.png \
        --mask ./resources/input/inpainting/bbox_mask.png \
        --result-out-dir ./resources/output/inpainting/demo_inpainting_global_local_res.
↪jpg
```

Image matting

Inpaiting models take a image and trimap pair as input, and output a alpha image. We take 'gca' as an example.

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

img = './resources/input/matting/GT05.jpg'
trimap = './resources/input/matting/GT05_trimap.jpg'

# Create a MMagicInferencer instance and infer
result_out_dir = './resources/output/matting/tutorial_matting_gca_res.png'
editor = MMagicInferencer('gca')
results = editor.infer(img=img, trimap=trimap, result_out_dir=result_out_dir)
```

Use demo/mmagic_inference_demo.py with the following commands:

```
python demo/mmagic_inference_demo.py \
        --model-name gca  \
        --img ./resources/input/matting/GT05.jpg \
        --trimap ./resources/input/matting/GT05_trimap.jpg \
        --result-out-dir ./resources/output/matting/demo_matting_gca_res.png
```

Image restoration

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
img = './resources/input/restoration/0901x2.png'
result_out_dir = './resources/output/restoration/tutorial_restoration_nafnet_res.png'
editor = MMagicInferencer('nafnet')
results = editor.infer(img=img, result_out_dir=result_out_dir)
```

```
python demo/mmagic_inference_demo.py \
        --model-name nafnet \
        --img ./resources/input/restoration/0901x2.png \
        --result-out-dir ./resources/output/restoration/demo_restoration_nafnet_res.png
```

Image colorization

```
import mmcv
import matplotlib.pyplot as plt
from mmagic.apis import MMagicInferencer

# Create a MMagicInferencer instance and infer
img = 'https://github-production-user-asset-6210df.s3.amazonaws.com/49083766/245713512-
↪de973677-2be8-4915-911f-fab90bb17c40.jpg'
```

(continues on next page)

(continued from previous page)

```
result_out_dir = './resources/output/colorization/tutorial_colorization_res.png'
editor = MMagicInferencer('inst_colorization')
results = editor.infer(img=img, result_out_dir=result_out_dir)
```

```
python demo/mmagic_inference_demo.py \
        --model-name inst_colorization \
        --img https://github-production-user-asset-6210df.s3.amazonaws.com/49083766/
↪245713512-de973677-2be8-4915-911f-fab90bb17c40.jpg \
        --result-out-dir demo_colorization_res.png
```

1.8.4 Previous Versions

If you want to use deprecated demos, please use MMagic v1.0.0rc7 and reference the old tutorial.

1.9 Tutorial 4: Train and test in MMagic

In this section, we introduce how to test and train models in MMagic.

In this section, we provide the following guides:

1.9.1 Prerequisite

Users need to *prepare dataset* first to enable training and testing models in MMagic.

1.9.2 Test a model in MMagic

Test with a single GPUs

You can use the following commands to test a pre-trained model with single GPUs.

```
python tools/test.py ${CONFIG_FILE} ${CHECKPOINT_FILE}
```

For example,

```
python tools/test.py configs/example_config.py work_dirs/example_exp/example_model_
↪20200202.pth
```

Test with multiple GPUs

MMagic supports testing with multiple GPUs, which can largely save your time in testing models. You can use the following commands to test a pre-trained model with multiple GPUs.

```
./tools/dist_test.sh ${CONFIG_FILE} ${CHECKPOINT_FILE} ${GPU_NUM}
```

For example,

```
./tools/dist_test.sh configs/example_config.py work_dirs/example_exp/example_model_
↪20200202.pth
```

Test with Slurm

If you run MMagic on a cluster managed with slurm, you can use the script `slurm_test.sh`. (This script also supports single machine testing.)

```
[GPUS=${GPUS}] ./tools/slurm_test.sh ${PARTITION} ${JOB_NAME} ${CONFIG_FILE} $
↪{CHECKPOINT_FILE}
```

Here is an example of using 8 GPUs to test an example model on the 'dev' partition with the job name 'test'.

```
GPUS=8 ./tools/slurm_test.sh dev test configs/example_config.py work_dirs/example_exp/
↪example_model_20200202.pth
```

You can check slurm_test.sh for full arguments and environment variables.

Test with specific metrics

MMagic provides various **evaluation metrics**, i.e., MS-SSIM, SWD, IS, FID, Precision&Recall, PPL, Equivarience, TransFID, TransIS, etc. We have provided unified evaluation scripts in tools/test.py for all models. If users want to evaluate their models with some metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/styleganv2/stylegan2_c2_ffhq_256_b4x8_800k.py
metrics = [
    dict(
        type='FrechetInceptionDistance',
        prefix='FID-Full-50k',
        fake_nums=50000,
        inception_style='StyleGAN',
        sample_model='ema'),
    dict(type='PrecisionAndRecall', fake_nums=50000, prefix='PR-50K'),
    dict(type='PerceptualPathLength', fake_nums=50000, prefix='ppl-w')
]
```

As above, `metrics` consist of multiple metric dictionaries. Each metric will contain `type` to indicate the category of the metric. `fake_nums` denotes the number of images generated by the model. Some metrics will output a dictionary of results, you can also set `prefix` to specify the prefix of the results. If you set the prefix of FID as `FID-Full-50k`, then an example of output may be

```
FID-Full-50k/fid: 3.6561  FID-Full-50k/mean: 0.4263  FID-Full-50k/cov: 3.2298
```

Then users can test models with the command below:

```
bash tools/dist_test.sh ${CONFIG_FILE} ${CKPT_FILE}
```

If you are in slurm environment, please switch to the tools/slurm_test.sh by using the following commands:

```
sh slurm_test.sh ${PLATFORM} ${JOBNAME} ${CONFIG_FILE} ${CKPT_FILE}
```

1.9.3 Train a model in MMagic

MMagic supports multiple ways of training:

1. *Train with a single GPU*
2. *Train with multiple GPUs*
3. *Train with multiple nodes*
4. *Train with Slurm*

Specifically, all outputs (log files and checkpoints) will be saved to the working directory, which is specified by `work_dir` in the config file.

Train with a single GPU

```
CUDA_VISIBLE=0 python tools/train.py configs/example_config.py --work-dir work_dirs/
↪example
```

Train with multiple nodes

To launch distributed training on multiple machines, which can be accessed via IPs, run the following commands:

On the first machine:

```
NNODES=2 NODE_RANK=0 PORT=$MASTER_PORT MASTER_ADDR=$MASTER_ADDR tools/dist_train.sh
↪$CONFIG $GPUS
```

On the second machine:

```
NNODES=2 NODE_RANK=1 PORT=$MASTER_PORT MASTER_ADDR=$MASTER_ADDR tools/dist_train.sh
↪$CONFIG $GPUS
```

To speed up network communication, high speed network hardware, such as Infiniband, is recommended. Please refer to PyTorch docs for more information.

Train with multiple GPUs

```
./tools/dist_train.sh ${CONFIG_FILE} ${GPU_NUM} [optional arguments]
```

Train with Slurm

If you run MMagic on a cluster managed with slurm, you can use the script `slurm_train.sh`. (This script also supports single machine training.)

```
[GPUS=${GPUS}] ./tools/slurm_train.sh ${PARTITION} ${JOB_NAME} ${CONFIG_FILE} ${WORK_DIR}
```

Here is an example of using 8 GPUs to train an inpainting model on the dev partition.

```
GPUS=8 ./tools/slurm_train.sh dev configs/inpainting/gl_places.py /nfs/xxxx/gl_places_256
```

You can check slurm_train.sh for full arguments and environment variables.

Optional arguments

- `--amp`: This argument is used for fixed-precision training.
- `--resume`: This argument is used for auto resume if the training is aborted.

1.9.4 Train with specific evaluation metrics

Benefit from the `mmengine`'s `Runner`. We can evaluate model during training in a simple way as below.

```
# define metrics
metrics = [
    dict(
        type='FrechetInceptionDistance',
        prefix='FID-Full-50k',
        fake_nums=50000,
        inception_style='StyleGAN')
]

# define dataloader
val_dataloader = dict(
    batch_size=128,
    num_workers=8,
    dataset=dict(
        type='BasicImageDataset',
        data_root='data/celeba-cropped/',
        pipeline=[
            dict(type='LoadImageFromFile', key='img'),
            dict(type='Resize', scale=(64, 64)),
            dict(type='PackInputs')
        ]),
    sampler=dict(type='DefaultSampler', shuffle=False),
    persistent_workers=True)

# define val interval
train_cfg = dict(by_epoch=False, val_begin=1, val_interval=10000)

# define val loop and evaluator
val_cfg = dict(type='MultiValLoop')
val_evaluator = dict(type='Evaluator', metrics=metrics)
```

You can set `val_begin` and `val_interval` to adjust when to begin validation and interval of validation.

For details of metrics, refer to *metrics' guide*.

1.10 Tutorial 5: Using metrics in MMagic

MMagic supports **17 metrics** to assess the quality of models.

Please refer to *Train and Test in MMagic* for usages.

Here, we will specify the details of different metrics one by one.

The structure of this guide are as follows:

- *Tutorial 5: Using metrics in MMagic*
  - *MAE*
  - *MSE*
  - *PSNR*

- *SNR*
- *SSIM*
- *NIQE*
- *SAD*
- *MattingMSE*
- *GradientError*
- *ConnectivityError*
- *FID and TransFID*
- *IS and TransIS*
- *Precision and Recall*
- *PPL*
- *SWD*
- *MS-SSIM*
- *Equivarience*

1.10.1 MAE

MAE is Mean Absolute Error metric for image. To evaluate with MAE, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='MAE'),
]
```

1.10.2 MSE

MSE is Mean Squared Error metric for image. To evaluate with MSE, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='MSE'),
]
```

1.10.3 PSNR

PSNR is Peak Signal-to-Noise Ratio. Our implement refers to https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio. To evaluate with PSNR, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='PSNR'),
]
```

1.10.4 SNR

SNR is Signal-to-Noise Ratio. Our implementation refers to https://en.wikipedia.org/wiki/Signal-to-noise_ratio. To evaluate with SNR, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='SNR'),
]
```

1.10.5 SSIM

SSIM is structural similarity for image, proposed in Image quality assessment: from error visibility to structural similarity. The results of our implementation are the same as that of the official released MATLAB code in https://ece.uwaterloo.ca/~z70wang/research/ssim/. To evaluate with SSIM, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='SSIM'),
]
```

1.10.6 NIQE

NIQE is Natural Image Quality Evaluator metric, proposed in Making a "Completely Blind" Image Quality Analyzer. Our implementation could produce almost the same results as the official MATLAB codes: http://live.ece.utexas.edu/research/quality/niqe_release.zip.

To evaluate with NIQE, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='NIQE'),
]
```

1.10.7 SAD

SAD is Sum of Absolute Differences metric for image matting. This metric compute per-pixel absolute difference and sum across all pixels. To evaluate with SAD, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='SAD'),
]
```

1.10.8 MattingMSE

MattingMSE is Mean Squared Error metric for image matting. To evaluate with MattingMSE, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='MattingMSE'),
]
```

1.10.9 GradientError

GradientError is Gradient error for evaluating alpha matte prediction. To evaluate with GradientError, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='GradientError'),
]
```

1.10.10 ConnectivityError

ConnectivityError is Connectivity error for evaluating alpha matte prediction. To evaluate with ConnectivityError, please add the following configuration in the config file:

```
val_evaluator = [
    dict(type='ConnectivityError'),
]
```

1.10.11 FID and TransFID

Fréchet Inception Distance is a measure of similarity between two datasets of images. It was shown to correlate well with the human judgment of visual quality and is most often used to evaluate the quality of samples of Generative Adversarial Networks. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network.

In `MMagic`, we provide two versions for FID calculation. One is the commonly used PyTorch version and the other one is used in StyleGAN paper. Meanwhile, we have compared the difference between these two implementations in the StyleGAN2-FFHQ1024 model (the details can be found here). Fortunately, there is a marginal difference in the final results. Thus, we recommend users adopt the more convenient PyTorch version.

About PyTorch version and Tero's version: The commonly used PyTorch version adopts the modified InceptionV3 network to extract features for real and fake images. However, Tero's FID requires a script module for Tensorflow InceptionV3. Note that applying this script module needs `PyTorch >= 1.6.0`.

About extracting real inception data: For the users' convenience, the real features will be automatically extracted at test time and saved locally, and the stored features will be automatically read at the next test. Specifically, we will calculate a hash value based on the parameters used to calculate the real features, and use the hash value to mark the feature file, and when testing, if the `inception_pkl` is not set, we will look for the feature in `MMAGIC_CACHE_DIR` (~/.cache/openmmlab/mmagic/). If cached inception pkl is not found, then extracting will be performed.

To use the FID metric, you should add the metric in a config file like this:

```
metrics = [
    dict(
        type='FrechetInceptionDistance',
        prefix='FID-Full-50k',
        fake_nums=50000,
        inception_style='StyleGAN',
        sample_model='ema')
]
```

If you work on an new machine, then you can copy the `pkl` files in `MMAGIC_CACHE_DIR` and copy them to new machine and set `inception_pkl` field.

```
metrics = [
    dict(
        type='FrechetInceptionDistance',
        prefix='FID-Full-50k',
        fake_nums=50000,
        inception_style='StyleGAN',
        inception_pkl=
        'work_dirs/inception_pkl/inception_state-capture_mean_cov-full-
↪33ad4546f8c9152e4b3bdb1b0c08dbaf.pkl',  # copied from old machine
        sample_model='ema')
]
```

`TransFID` has same usage as `FID`, but it's designed for translation models like `Pix2Pix` and `CycleGAN`, which is adapted for our evaluator. You can refer to *evaluation* for details.

1.10.12 IS and TransIS

Inception score is an objective metric for evaluating the quality of generated images, proposed in Improved Techniques for Training GANs. It uses an InceptionV3 model to predict the class of the generated images, and suppose that 1) If an image is of high quality, it will be categorized into a specific class. 2) If images are of high diversity, the range of images' classes will be wide. So the KL-divergence of the conditional probability and marginal probability can indicate the quality and diversity of generated images. You can see the complete implementation in `metrics.py`, which refers to https://github.com/sbarratt/inception-score-pytorch/blob/master/inception_score.py. If you want to evaluate models with `IS` metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/biggan/biggan_2xb25-500kiters_cifar10-32x32.py
metrics = [
    xxx,
    dict(
        type='IS',
        prefix='IS-50k',
        fake_nums=50000,
        inception_style='StyleGAN',
        sample_model='ema')
]
```

To be noted that, the selection of Inception V3 and image resize method can significantly influence the final IS score. Therefore, we strongly recommend users may download the Tero's script model of Inception V3 (load this script model need torch >= 1.6) and use `Bicubic` interpolation with `Pillow` backend.

Corresponding to config, you can set `resize_method` and `use_pillow_resize` for image resizing. You can also set `inception_style` as `StyleGAN` for recommended tero's inception model, or `PyTorch` for torchvision's implementa-

tion. For environment without internet, you can download the inception's weights, and set `inception_path` to your inception model.

We also perform a survey on the influence of data loading pipeline and the version of pretrained Inception V3 on the IS result. All IS are evaluated on the same group of images which are randomly selected from the ImageNet dataset.

`TransIS` has same usage as `IS`, but it's designed for translation models like `Pix2Pix` and `CycleGAN`, which is adapted for our evaluator. You can refer to *evaluation* for details.

1.10.13 Precision and Recall

Our `Precision and Recall` implementation follows the version used in StyleGAN2. In this metric, a VGG network will be adopted to extract the features for images. Unfortunately, we have not found a PyTorch VGG implementation leading to similar results with Tero's version used in StyleGAN2. (About the differences, please see this file.) Thus, in our implementation, we adopt Teor's VGG network by default. Importantly, applying this script module needs `PyTorch >= 1.6.0`. If with a lower PyTorch version, we will use the PyTorch official VGG network for feature extraction.

To evaluate with `P&R`, please add the following configuration in the config file:

```
metrics = [
    dict(type='PrecisionAndRecall', fake_nums=50000, prefix='PR-50K')
]
```

1.10.14 PPL

Perceptual path length measures the difference between consecutive images (their VGG16 embeddings) when interpolating between two random inputs. Drastic changes mean that multiple features have changed together and that they might be entangled. Thus, a smaller PPL score appears to indicate higher overall image quality by experiments.
As a basis for our metric, we use a perceptually-based pairwise image distance that is calculated as a weighted difference between two VGG16 embeddings, where the weights are fit so that the metric agrees with human perceptual similarity judgments. If we subdivide a latent space interpolation path into linear segments, we can define the total perceptual length of this segmented path as the sum of perceptual differences over each segment, and a natural definition for the perceptual path length would be the limit of this sum under infinitely fine subdivision, but in practice we approximate it using a small subdivision `$` `\epsilon=10^{-4}` `$`. The average perceptual path length in latent `space` Z, over all possible endpoints, is therefore

`$$` `L_Z = E[\frac{1}{\epsilon^2}d(G(slerp(z_1,z_2;t))), G(slerp(z_1,z_2;t+\epsilon)))]` `$$`

Computing the average perceptual path length in latent `space` W is carried out in a similar fashion:

`$$` `L_Z = E[\frac{1}{\epsilon^2}d(G(slerp(z_1,z_2;t))), G(slerp(z_1,z_2;t+\epsilon)))]` `$$`

Where `$` `z_1, z_2 \sim P(z)` `$`, and `$` ` t \sim U(0,1)` `$` if we set `sampling` to full, `$` ` t \in \{0,1\}` `$` if we set `sampling` to end. `$` ` G` `$` is the generator(i.e. `$` ` g \circ f` `$` for style-based networks), and `$` ` d(., .)` `$` evaluates the perceptual distance between the resulting images.We compute the expectation by taking 100,000 samples (set `num_images` to 50,000 in our code).

You can find the complete implementation in `metrics.py`, which refers to https://github.com/rosinality/stylegan2-pytorch/blob/master/ppl.py. If you want to evaluate models with `PPL` metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/styleganv2/stylegan2_c2_ffhq_1024_b4x8.py
metrics = [
    xxx,
    dict(type='PerceptualPathLength', fake_nums=50000, prefix='ppl-w')
]
```

1.10.15 SWD

Sliced Wasserstein distance is a discrepancy measure for probability distributions, and smaller distance indicates generated images look like the real ones. We obtain the Laplacian pyramids of every image and extract patches from the Laplacian pyramids as descriptors, then SWD can be calculated by taking the sliced Wasserstein distance of the real and fake descriptors. You can see the complete implementation in `metrics.py`, which refers to https://github.com/tkarras/progressive_growing_of_gans/blob/master/metrics/sliced_wasserstein.py. If you want to evaluate models with `SWD` metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/dcgan/dcgan_1xb128-5epoches_lsun-bedroom-64x64.py
metrics = [
    dict(
        type='SWD',
        prefix='swd',
        fake_nums=16384,
        sample_model='orig',
        image_shape=(3, 64, 64))
]
```

1.10.16 MS-SSIM

Multi-scale structural similarity is used to measure the similarity of two images. We use MS-SSIM here to measure the diversity of generated images, and a low MS-SSIM score indicates the high diversity of generated images. You can see the complete implementation in `metrics.py`, which refers to https://github.com/tkarras/progressive_growing_of_gans/blob/master/metrics/ms_ssim.py. If you want to evaluate models with `MS-SSIM` metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/dcgan/dcgan_1xb128-5epoches_lsun-bedroom-64x64.py
metrics = [
    dict(
        type='MS_SSIM', prefix='ms-ssim', fake_nums=10000,
        sample_model='orig')
]
```

1.10.17 Equivarience

Equivarience of generative models refer to the exchangeability of model forward and geometric transformations. Currently this metric is only calculated for StyleGANv3, you can see the complete implementation in `metrics.py`, which refers to https://github.com/NVlabs/stylegan3/blob/main/metrics/equivariance.py. If you want to evaluate models with `Equivarience` metrics, you can add the `metrics` into your config file like this:

```
# at the end of the configs/styleganv3/stylegan3-t_gamma2.0_8xb4-fp16-noaug_ffhq-256x256.
↪py
metrics = [
    dict(
        type='Equivariance',
        fake_nums=50000,
        sample_mode='ema',
        prefix='EQ',
        eq_cfg=dict(
```

(continues on next page)

(continued from previous page)

```
            compute_eqt_int=True, compute_eqt_frac=True, compute_eqr=True))
]
```

1.11 Tutorial 6: Visualization

The visualization of images is an important way to measure the quality of image processing, editing and synthesis. Using `visualizer` in config file can save visual results when training or testing. You can follow MMEngine Documents to learn the usage of visualization. MMagic provides a rich set of visualization functions. In this tutorial, we introduce the usage of the visualization functions provided by MMagic.

1.11.1 Overview

It is recommended to learn the basic concept of visualization in design documentation.

In MMagic, the visualization of the training or testing process requires the configuration of three components: `VisualizationHook`, `Visualizer`, and `VisBackend`, The diagram below shows the relationship between Visualizer and VisBackend,

VisualizationHook fetches the visualization results of the model output in fixed intervals during training and passes them to Visualizer. **Visualizer** is responsible for converting the original visualization results into the desired type (png, gif, etc.) and then transferring them to **VisBackend** for storage or display.

Visualization configuration of GANs

For GAN models, such as StyleGAN and SAGAN, a usual configuration is shown below:

```
# VisualizationHook
custom_hooks = [
    dict(
        type='VisualizationHook',
        interval=5000,  # visualization interval
```

(continues on next page)

(continued from previous page)

```
        fixed_input=True,  # whether use fixed noise input to generate images
        vis_kwargs_list=dict(type='GAN', name='fake_img')  # pre-defined visualization␣
↪arguments for GAN models
    )
]
# VisBackend
vis_backends = [
    dict(type='VisBackend'),  # vis_backend for saving images to file system
    dict(type='WandbVisBackend',  # vis_backend for uploading images to Wandb
        init_kwargs=dict(
            project='MMagic',   # project name for Wandb
            name='GAN-Visualization-Demo'  # name of the experiment for Wandb
        ))
]
# Visualizer
visualizer = dict(type='Visualizer', vis_backends=vis_backends)
```

If you apply Exponential Moving Average (EMA) to a generator and want to visualize the EMA model, you can modify config of `VisualizationHook` as below:

```
custom_hooks = [
    dict(
        type='VisualizationHook',
        interval=5000,
        fixed_input=True,
        # vis ema and orig in `fake_img` at the same time
        vis_kwargs_list=dict(
            type='Noise',
            name='fake_img',  # save images with prefix `fake_img`
            sample_model='ema/orig',  # specified kwargs for `NoiseSampler`
            target_keys=['ema.fake_img', 'orig.fake_img']  # specific key to␣
↪visualization
        ))
]
```

Visualization configuration of image translation models

For Translation models, such as CycleGAN and Pix2Pix, visualization configs can be formed as below:

```
# VisualizationHook
custom_hooks = [
    dict(
        type='VisualizationHook',
        interval=5000,
        fixed_input=True,
        vis_kwargs_list=[
            dict(
                type='Translation',  # Visualize results on the training set
                name='trans'),  #  save images with prefix `trans`
            dict(
                type='Translationval',  # Visualize results on the validation set
```

(continues on next page)

(continued from previous page)

```
                name='trans_val'),  #  save images with prefix `trans_val`
        ])
]
# VisBackend
vis_backends = [
    dict(type='VisBackend'),  # vis_backend for saving images to file system
    dict(type='WandbVisBackend',  # vis_backend for uploading images to Wandb
        init_kwargs=dict(
            project='MMagic',   # project name for Wandb
            name='Translation-Visualization-Demo'  # name of the experiment for Wandb
        ))
]
# Visualizer
visualizer = dict(type='Visualizer', vis_backends=vis_backends)
```

Visualization configuration of diffusion models

For Diffusion models, such as Improved-DDPM, we can use the following configuration to visualize the denoising process through a gif:

```
# VisualizationHook
custom_hooks = [
    dict(
        type='VisualizationHook',
        interval=5000,
        fixed_input=True,
        vis_kwargs_list=dict(type='DDPMDenoising'))  # pre-defined visualization␣
↪argument for DDPM models
]
# VisBackend
vis_backends = [
    dict(type='VisBackend'),  # vis_backend for saving images to file system
    dict(type='WandbVisBackend',  # vis_backend for uploading images to Wandb
        init_kwargs=dict(
            project='MMagic',   # project name for Wandb
            name='Diffusion-Visualization-Demo'  # name of the experiment for Wandb
        ))
]
# Visualizer
visualizer = dict(type='Visualizer', vis_backends=vis_backends)
```

Visualization configuration of inpainting models

For inpainting models, such as AOT-GAN and Global&Local, a usual configuration is shown below:

```
# VisBackend
vis_backends = [dict(type='LocalVisBackend')]
# Visualizer
visualizer = dict(
    type='ConcatImageVisualizer',
    vis_backends=vis_backends,
    fn_key='gt_path',
    img_keys=['gt_img', 'input', 'pred_img'],
    bgr2rgb=True)
# VisualizationHook
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)]
```

Visualization configuration of matting models

For matting models, such as DIM and GCA, a usual configuration is shown below:

```
# VisBackend
vis_backends = [dict(type='LocalVisBackend')]
# Visualizer
visualizer = dict(
    type='ConcatImageVisualizer',
    vis_backends=vis_backends,
    fn_key='trimap_path',
    img_keys=['pred_alpha', 'trimap', 'gt_merged', 'gt_alpha'],
    bgr2rgb=True)
# VisualizationHook
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)]
```

Visualization configuration of SISR/VSR/VFI models

For SISR/VSR/VFI models, such as EDSR, EDVR and CAIN, a usual configuration is shown below:

```
# VisBackend
vis_backends = [dict(type='LocalVisBackend')]
# Visualizer
visualizer = dict(
    type='ConcatImageVisualizer',
    vis_backends=vis_backends,
    fn_key='gt_path',
    img_keys=['gt_img', 'input', 'pred_img'],
    bgr2rgb=False)
# VisualizationHook
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)]
```

The specific configuration of the `VisualizationHook`, `Visualizer` and `VisBackend` components are described below

1.11.2 Visualization Hook

In MMagic, we use `BasicVisualizationHook` and `VisualizationHook` as `VisualizationHook`. `VisualizationHook` supports three following cases.

(1) Modify `vis_kwargs_list` to visualize the output of the model under specific inputs , which is suitable for visualization of the generated results of GAN and translation results of Image-to-Image-Translation models under specific data input, etc. Below are two typical examples:

```
# input as dict
vis_kwargs_list = dict(
    type='Noise',  # use 'Noise' sampler to generate model input
    name='fake_img',  # define prefix of saved images
)

# input as list of dict
vis_kwargs_list = [
    dict(type='Arguments',  # use `Arguments` sampler to generate model input
         name='arg_output',  # define prefix of saved images
         vis_mode='gif',  # specific visualization mode as GIF
         forward_kwargs=dict(forward_mode='sampling', sample_kwargs=dict(show_
↪pbar=True))  # specific kwargs for `Arguments` sampler
    ),
    dict(type='Data',  # use `Data` sampler to feed data in dataloader to model as input
         n_samples=36,  # specific how many samples want to generate
         fixed_input=False,  # specific do not use fixed input for each visualization␣
↪process
    )
]
```

`vis_kwargs_list` takes dict or list of dict as input. Each of dict must contain a `type` field indicating the **type of sampler** used to generate the model input, and each of the dict must also contain the keyword fields necessary for the sampler (e.g. `ArgumentSampler` requires that the argument dictionary contain `forward_kwargs`).

> To be noted that, this content is checked by the corresponding sampler and is not restricted by `BasicVisualizationHook`.

In addition, the other fields are generic fields (e.g. `n_samples`, `n_row`, `name`, `fixed_input`, etc.). If not passed in, the default values from the BasicVisualizationHook initialization will be used.

For the convenience of users, MMagic has pre-defined visualization parameters for **GAN**, **Translation models**, **SinGAN** and **Diffusion models**, and users can directly use the predefined visualization methods by using the following configuration:

```
vis_kwargs_list = dict(type='GAN')
vis_kwargs_list = dict(type='SinGAN')
vis_kwargs_list = dict(type='Translation')
vis_kwargs_list = dict(type='TranslationVal')
vis_kwargs_list = dict(type='TranslationTest')
vis_kwargs_list = dict(type='DDPMDenoising')
```

1.11.3 Visualizer

In MMagic, we implement `ConcatImageVisualizer` and `Visualizer`, which inherit from `mmengine.Visualizer`. The base class of `Visualizer` is `ManagerMixin` and this makes `Visualizer` a globally unique object. After being instantiated, `Visualizer` can be called at anywhere of the code by `Visualizer.get_current_instance()`, as shown below:

```
# configs
vis_backends = [dict(type='VisBackend')]
visualizer = dict(
    type='Visualizer', vis_backends=vis_backends, name='visualizer')
```

```
# `get_instance()` is called for globally unique instantiation
VISUALIZERS.build(cfg.visualizer)

# Once instantiated by the above code, you can call the `get_current_instance` method at␣
↪any location to get the visualizer
visualizer = Visualizer.get_current_instance()
```

The core interface of `Visualizer` is `add_datasample`. Through this interface, This interface will call the corresponding drawing function according to the corresponding `vis_mode` to obtain the visualization result in `np.ndarray` type. Then `show` or `add_image` will be called to directly show the results or pass the visualization result to the predefined vis_backend.

1.11.4 VisBackend

In general, users do not need to manipulate `VisBackend` objects, only when the current visualization storage can not meet the needs, users will want to manipulate the storage backend directly. MMagic supports a variety of different visualization backends, including:

- Basic VisBackend of MMEngine: including LocalVisBackend, TensorboardVisBackend and WandbVisBackend. You can follow MMEngine Documents to learn more about them
- VisBackend: Backend for **File System**. Save the visualization results to the corresponding position.
- TensorboardVisBackend: Backend for **Tensorboard**. Send the visualization results to Tensorboard.
- WandbVisBackend: Backend for **Wandb**. Send the visualization results to Tensorboard.

One `Visualizer` object can have access to any number of VisBackends and users can access to the backend by their class name in their code.

```
# configs
vis_backends = [dict(type='Visualizer'), dict(type='WandbVisBackend')]
visualizer = dict(
    type='Visualizer', vis_backends=vis_backends, name='visualizer')
```

```
# code
VISUALIZERS.build(cfg.visualizer)
visualizer = Visualizer.get_current_instance()

# access to the backend by class name
gen_vis_backend = visualizer.get_backend('VisBackend')
gen_wandb_vis_backend = visualizer.get_backend('GenWandbVisBackend')
```

When there are multiply VisBackend with the same class name, user must specific name for each VisBackend.

```
# configs
vis_backends = [
    dict(type='VisBackend', name='gen_vis_backend_1'),
    dict(type='VisBackend', name='gen_vis_backend_2')
]
visualizer = dict(
    type='Visualizer', vis_backends=vis_backends, name='visualizer')
```

```
# code
VISUALIZERS.build(cfg.visualizer)
visualizer = Visualizer.get_current_instance()

local_vis_backend_1 = visualizer.get_backend('gen_vis_backend_1')
local_vis_backend_2 = visualizer.get_backend('gen_vis_backend_2')
```

Visualize by Different Storage Backends

If you want to use a different backend (Wandb, Tensorboard, or a custom backend with a remote window), just change the `vis_backends` in the config, as follows:

Local

```
vis_backends = [dict(type='LocalVisBackend')]
```

Tensorboard

```
vis_backends = [dict(type='TensorboardVisBackend')]
visualizer = dict(
    type='ConcatImageVisualizer', vis_backends=vis_backends, name='visualizer')
```

```
vis_backends = [dict(type='WandbVisBackend')]
visualizer = dict(
    type='ConcatImageVisualizer', vis_backends=vis_backends, name='visualizer')
```

1.12 Tutorial 7: Useful tools

We provide lots of useful tools under `tools/` directory.

The structure of this guide is as follows:

1.12.1 Get the FLOPs and params

We provide a script adapted from flops-counter.pytorch to compute the FLOPs and params of a given model.

```
python tools/analysis_tools/get_flops.py ${CONFIG_FILE} [--shape ${INPUT_SHAPE}]
```

For example,

```
python tools/analysis_tools/get_flops.py configs/resotorer/srresnet.py --shape 40 40
```

You will get the result like this.

```
==============================
Input shape: (3, 40, 40)
Flops: 4.07 GMac
Params: 1.52 M
==============================
```

Note: This tool is still experimental and we do not guarantee that the number is correct. You may well use the result for simple comparisons, but double check it before you adopt it in technical reports or papers.

(1) FLOPs are related to the input shape while parameters are not. The default input shape is (1, 3, 250, 250). (2) Some operators are not counted in FLOPs like GN and custom operators. You can add support for new operators by modifying `mmcv/cnn/utils/flops_counter.py`.

1.12.2 Publish a model

Before you upload a model to AWS, you may want to

1. convert model weights to CPU tensors
2. delete the optimizer states and
3. compute the hash of the checkpoint file and append time and the hash id to the filename.

```
python tools/model_converters/publish_model.py ${INPUT_FILENAME} ${OUTPUT_FILENAME}
```

E.g.,

```
python tools/model_converters/publish_model.py work_dirs/stylegan2/latest.pth stylegan2_
↪c2_8xb4_ffhq-1024x1024.pth
```

The final output filename will be `stylegan2_c2_8xb4_ffhq-1024x1024_{time}-{hash id}.pth`.

1.12.3 Print full config

MMGeneration incorporates config mechanism to set parameters used for training and testing models. With our *config* mechanism, users can easily conduct extensive experiments without hard coding. If you wish to inspect the config file, you may run `python tools/misc/print_config.py /PATH/TO/CONFIG` to see the complete config.

An Example:

```
python tools/misc/print_config.py configs/styleganv2/stylegan2_c2-PL_8xb4-fp16-partial-
↪GD-no-scaler-800kiters_ffhq-256x256.py
```

1.13 Tutorial 8: Deploy models in MMagic

The deployment of OpenMMLab codebases, including MMClassification, MMDetection, MMagic and so on are supported by MMDeploy. The latest deployment guide for MMagic can be found from here.

This tutorial is organized as follows:

1.13.1 Installation

Please follow the *guide* to install mmagic. And then install mmdeploy from source by following this guide.

Note: If you install mmdeploy prebuilt package, please also clone its repository by 'git clone https://github.com/open-mmlab/mmdeploy.git –depth=1' to get the deployment config files.

1.13.2 Convert model

Suppose MMagic and mmdeploy repositories are in the same directory, and the working directory is the root path of MMagic.

Take ESRGAN model as an example. You can download its checkpoint from here, and then convert it to onnx model as follows:

```
from mmdeploy.apis import torch2onnx
from mmdeploy.backend.sdk.export_info import export2SDK

img = 'tests/data/image/face/000001.png'
work_dir = 'mmdeploy_models/mmagic/onnx'
save_file = 'end2end.onnx'
deploy_cfg = '../mmdeploy/configs/mmagic/super-resolution/super-resolution_onnxruntime_
↪dynamic.py'
model_cfg = 'configs/esrgan/esrgan_psnr-x4c64b23g32_1xb16-1000k_div2k.py'
model_checkpoint = 'esrgan_psnr_x4c64b23g32_1x16_1000k_div2k_20200420-bf5c993c.pth'
device = 'cpu'

# 1. convert model to onnx
torch2onnx(img, work_dir, save_file, deploy_cfg, model_cfg,
  model_checkpoint, device)
```

(continues on next page)

(continued from previous page)

```
# 2. extract pipeline info for inference by MMDeploy SDK
export2SDK(deploy_cfg, model_cfg, work_dir, pth=model_checkpoint, device=device)
```

It is crucial to specify the correct deployment config during model conversion.MMDeploy has already provided builtin deployment config files of all supported backends for mmagic, under which the config file path follows the pattern:

```
{task}/{task}_{backend}-{precision}_{static | dynamic}_{shape}.py
```

- **{task}:** task in mmagic.
- **{backend}:** inference backend, such as onnxruntime, tensorrt, pplnn, ncnn, openvino, coreml etc.
- **{precision}:** fp16, int8. When it's empty, it means fp32
- **{static | dynamic}:** static shape or dynamic shape
- **{shape}:** input shape or shape range of a model

Therefore, in the above example, you can also convert `ESRGAN` to other backend models by changing the deployment config file, e.g., converting to tensorrt-fp16 model by `super-resolution_tensorrt-fp16_dynamic-32x32-512x512.py`.

Tip: When converting mmagic models to tensorrt models, –device should be set to "cuda"

1.13.3 Model specification

Before moving on to model inference chapter, let's know more about the converted model structure which is very important for model inference.

The converted model locates in the working directory like `mmdeploy_models/mmagic/onnx` in the previous example. It includes:

```
mmdeploy_models/mmagic/onnx
├── deploy.json
├── detail.json
├── end2end.onnx
└── pipeline.json
```

in which,

- **end2end.onnx**: backend model which can be inferred by ONNX Runtime
- ***xxx*.json**: the necessary information for mmdeploy SDK

The whole package **mmdeploy_models/mmagic/onnx** is defined as **mmdeploy SDK model**, i.e., **mmdeploy SDK model** includes both backend model and inference meta information.

1.13.4 Model inference

Backend model inference

Take the previous converted `end2end.onnx` model as an example, you can use the following code to inference the model.

```
from mmdeploy.apis.utils import build_task_processor
from mmdeploy.utils import get_input_shape, load_config
import torch

deploy_cfg = '../mmdeploy/configs/mmagic/super-resolution/super-resolution_onnxruntime_
↪dynamic.py'
model_cfg = 'configs/esrgan/esrgan_psnr-x4c64b23g32_1xb16-1000k_div2k.py'
device = 'cpu'
backend_model = ['mmdeploy_models/mmagic/onnx/end2end.onnx']
image = 'tests/data/image/lq/baboon_x4.png'

# read deploy_cfg and model_cfg
deploy_cfg, model_cfg = load_config(deploy_cfg, model_cfg)

# build task and backend model
task_processor = build_task_processor(model_cfg, deploy_cfg, device)
model = task_processor.build_backend_model(backend_model)

# process input image
input_shape = get_input_shape(deploy_cfg)
model_inputs, _ = task_processor.create_input(image, input_shape)

# do model inference
with torch.no_grad():
    result = model.test_step(model_inputs)

# visualize results
task_processor.visualize(
    image=image,
    model=model,
    result=result[0],
    window_name='visualize',
    output_file='output_restorer.bmp')
```

SDK model inference

You can also perform SDK model inference like following,

```
from mmdeploy_python import Restorer
import cv2

img = cv2.imread('tests/data/image/lq/baboon_x4.png')

# create a predictor
restorer = Restorer(model_path='mmdeploy_models/mmagic/onnx', device_name='cpu', device_
↪id=0)
```

(continues on next page)

(continued from previous page)

```
# perform inference
result = restorer(img)

# visualize inference result
cv2.imwrite('output_restorer.bmp', result)
```

Besides python API, MMDeploy SDK also provides other FFI (Foreign Function Interface), such as C, C++, C#, Java and so on. You can learn their usage from demos.

1.13.5 Supported models

Please refer to here for the supported model list.

1.14 Evaluator

1.14.1 Evaluation Metrics and Evaluators

In model validation and testing, it is usually necessary to quantitatively evaluate the accuracy of the model. In mmagic, the evaluation metrics and evaluators are implemented to accomplish this functionality.

- Evaluation metrics are used to calculate specific model accuracy indicators based on test data and model prediction results. mmagic provides a variety of built-in metrics, which can be found in the metrics documentation. Additionally, metrics are decoupled from datasets and can be used for multiple datasets.
- The evaluator is the top-level module for evaluation metrics and usually contains one or more metrics. The purpose of the evaluator is to perform necessary data format conversion and call evaluation metrics to calculate the model accuracy during model evaluation. The evaluator is typically built by a `Runner` or a testing script, which are used for online evaluation and offline evaluation, respectively.

The evaluator in MMagic inherits from that in MMEngine and has a similar basic usage. For specific information, you can refer to Model Accuracy Evaluation. However, different from other high-level vision tasks, the evaluation metrics for generative models often have multiple inputs. For example, the input for the Inception Score (IS) metric is only fake images and any number of real images, while the Perceptual Path Length (PPL) requires sampling from the latent space. To accommodate different evaluation metrics, mmagic introduces two important methods, `prepare_metrics` and `prepare_samplers` to meet the above requirements.

1.14.2 prepare_metrics

```
class Evaluator(Evaluator):
        ...
    def prepare_metrics(self, module: BaseModel, dataloader: DataLoader):
        """Prepare for metrics before evaluation starts. Some metrics use
        pretrained model to extract feature. Some metrics use pretrained model
        to extract feature and input channel order may vary among those models.
        Therefore, we first parse the output color order from data
        preprocessor and set the color order for each metric. Then we pass the
        dataloader to each metrics to prepare pre-calculated items. (e.g.
        inception feature of the real images). If metric has no pre-calculated
        items, :meth:`metric.prepare` will be ignored. Once the function has
```

(continues on next page)

(continued from previous page)

```
        been called, :attr:`self.is_ready` will be set as `True`. If
        :attr:`self.is_ready` is `True`, this function will directly return to
        avoid duplicate computation.

        Args:
            module (BaseModel): Model to evaluate.
            dataloader (DataLoader): The dataloader for real images.
        """
        if self.metrics is None:
            self.is_ready = True
            return

        if self.is_ready:
            return

        # prepare metrics
        for metric in self.metrics:
            metric.prepare(module, dataloader)
        self.is_ready = True
```

The `prepare_metrics` method needs to be called before the evaluation starts. It is used to preprocess before evaluating each metric, and will sequentially call the prepare method of each metric in the evaluator to prepare any pre-calculated elements needed for that metric (such as features from hidden layers). Additionally, to avoid repeated calls, the `evaluator.is_ready` flag will be set to True after preprocessing for all metrics is completed.

```
class GenMetric(BaseMetric):
        ...
    def prepare(self, module: nn.Module, dataloader: DataLoader) -> None:
        """Prepare for the pre-calculating items of the metric. Defaults to do
        nothing.

        Args:
            module (nn.Module): Model to evaluate.
            dataloader (DataLoader): Dataloader for the real images.
        """
        if is_model_wrapper(module):
            module = module.module
        self.data_preprocessor = module.data_preprocessor
```

1.14.3 prepare_samplers

Different metrics require different inputs for generative models. For example, FID, KID, and IS only need the generated fake images, while PPL requires vectors from the latent space. Therefore, mmagic groups different evaluation metrics based on the type of input. One or more evaluation metrics in the same group share a data sampler. The sampler mode for each evaluation metric is determined by the `SAMPLER_MODE` attribute of that metric.

```
class GenMetric(BaseMetric):
        ...
    SAMPLER_MODE = 'normal'

class GenerativeMetric(GenMetric):
```

(continues on next page)

(continued from previous page)

```
        ...
    SAMPLER_MODE = 'Generative'
```

The `prepare_samplers` method of the evaluator is responsible for preparing the data samplers based on the sampler mode of all evaluation metrics.

```
class Evaluator(Evaluator):
        ...
    def prepare_samplers(self, module: BaseModel, dataloader: DataLoader
                         ) -> List[Tuple[List[BaseMetric], Iterator]]:
        """Prepare for the sampler for metrics whose sampling mode are
        different. For generative models, different metric need image
        generated with different inputs. For example, FID, KID and IS need
        images generated with random noise, and PPL need paired images on the
        specific noise interpolation path. Therefore, we first group metrics
        with respect to their sampler's mode (refers to
        :attr:~`GenMetrics.SAMPLER_MODE`), and build a shared sampler for each
        metric group. To be noted that, the length of the shared sampler
        depends on the metric of the most images required in each group.

        Args:
            module (BaseModel): Model to evaluate. Some metrics (e.g. PPL)
                require `module` in their sampler.
            dataloader (DataLoader): The dataloader for real image.

        Returns:
            List[Tuple[List[BaseMetric], Iterator]]: A list of "metrics-shared
                sampler" pair.
        """
        if self.metrics is None:
            return [[[None], []]]

        # grouping metrics based on `SAMPLER_MODE` and `sample_mode`
        metric_mode_dict = defaultdict(list)
        for metric in self.metrics:  # Specify a sampler group for each metric.
            metric_md5 = self._cal_metric_hash(metric)
            metric_mode_dict[metric_md5].append(metric)

        metrics_sampler_list = []
        for metrics in metric_mode_dict.values(): # Generate a sampler for each group.
            first_metric = metrics[0]
            metrics_sampler_list.append([
                metrics,
                first_metric.get_metric_sampler(module, dataloader, metrics)
            ])

        return metrics_sampler_list
```

The method will first check if it has any evaluation metrics to calculate: if not, it will return directly. If there are metrics to calculate, it will iterate through all the evaluation metrics and group them based on the sampler_mode and sample_model. The specific implementation is as follows: it calculates a hash code based on the sampler_mode and sample_model, and puts the evaluation metrics with the same hash code into the same list.

```
class Evaluator(Evaluator):
        ...
    @staticmethod
    def _cal_metric_hash(metric: GenMetric):
        """Calculate a unique hash value based on the `SAMPLER_MODE` and
        `sample_model`."""
        sampler_mode = metric.SAMPLER_MODE
        sample_model = metric.sample_model
        metric_dict = {
            'SAMPLER_MODE': sampler_mode,
            'sample_model': sample_model
        }
        if hasattr(metric, 'need_cond_input'):
            metric_dict['need_cond_input'] = metric.need_cond_input
        md5 = hashlib.md5(repr(metric_dict).encode('utf-8')).hexdigest()
        return md5
```

Finally, this method will generate a sampler for each evaluation metric group and add it to a list to return.

1.14.4 Evaluation process of an evaluator

The implementation of evaluation process can be found in `mmagic.engine.runner.MultiValLoop.run` and `mmagic.engine.runner.MultiTestLoop.run`. Here we take `mmagic.engine.runner.MultiValLoop.run` as example.

```
class MultiValLoop(BaseLoop):
        ...
    def run(self):
        ...
        # 1. prepare all metrics and get the total length
        metrics_sampler_lists = []
        meta_info_list = []
        dataset_name_list = []
        for evaluator, dataloader in zip(self.evaluators, self.dataloaders):
            # 1.1 prepare for metrics
            evaluator.prepare_metrics(module, dataloader)
            # 1.2 prepare for metric-sampler pair
            metrics_sampler_list = evaluator.prepare_samplers(
                module, dataloader)
            metrics_sampler_lists.append(metrics_sampler_list)
            # 1.3 update total length
            self._total_length += sum([
                len(metrics_sampler[1])
                for metrics_sampler in metrics_sampler_list
            ])
            # 1.4 save metainfo and dataset's name
            meta_info_list.append(
                getattr(dataloader.dataset, 'metainfo', None))
            dataset_name_list.append(dataloader.dataset.__class__.__name__)
```

First, the runner will perform preprocessing and obtain the necessary data samplers for evaluation using the `evaluator.prepare_metric` and `evaluator.prepare_samplers` methods. It will also update the total length of

samples obtained using the samplers. As the evaluation metrics and dataset in mmagic are separated, some meta_info required for evaluation also needs to be saved and passed to the evaluator.

```
class MultiValLoop(BaseLoop):
        ...
    def run(self):
        ...
        # 2. run evaluation
        for idx in range(len(self.evaluators)):
            # 2.1 set self.evaluator for run_iter
            self.evaluator = self.evaluators[idx]
            self.dataloader = self.dataloaders[idx]

            # 2.2 update metainfo for evaluator and visualizer
            meta_info = meta_info_list[idx]
            dataset_name = dataset_name_list[idx]
            if meta_info:
                self.evaluator.dataset_meta = meta_info
                self._runner.visualizer.dataset_meta = meta_info
            else:
                warnings.warn(
                    f'Dataset {dataset_name} has no metainfo. `dataset_meta` '
                    'in evaluator, metric and visualizer will be None.')

            # 2.3 generate images
            metrics_sampler_list = metrics_sampler_lists[idx]
            for metrics, sampler in metrics_sampler_list:
                for data in sampler:
                    self.run_iter(idx_counter, data, metrics)
                    idx_counter += 1

            # 2.4 evaluate metrics and update multi_metric
            metrics = self.evaluator.evaluate()
            if multi_metric and metrics.keys() & multi_metric.keys():
                raise ValueError('Please set different prefix for different'
                                 ' datasets in `val_evaluator`')
            else:
                multi_metric.update(metrics)
        # 3. finish evaluation and call hooks
        self._runner.call_hook('after_val_epoch', metrics=multi_metric)
        self._runner.call_hook('after_val')
```

After the preparation for evaluation is completed, the runner will iterate through all the evaluators and perform the evaluation one by one. Each evaluator needs to correspond to a data loader to complete the evaluation work for a dataset. Specifically, during the evaluation process for each evaluator, it is necessary to pass the required meta_info to the evaluator, then iterate through all the metrics_samplers of this evaluator to generate the images needed for evaluation, and finally complete the evaluation.

1.15 Data Structure

`DataSample` , the data structure interface of MMagic, inherits from `BaseDataElement`. The base class has implemented basic add/delete/update/check functions and supports data migration between different devices, as well as dictionary-like and tensor-like operations, which also allows the interfaces of different algorithms to be unified.

Specifically, an instance of BaseDataElement consists of two components:

- `metainfo`, which contains some meta information, e.g., `img_shape`, `img_id`, `color_order`, etc.
- `data`, which contains the data used in the loop.

Thanks to `DataSample` , the data flow between each module in the algorithm libraries, such as `visualizer`, `evaluator`, `model`, is greatly simplified.

The attributes in `DataSample` are divided into several parts:

```
- ``gt_img``: Ground truth image(s).
- ``pred_img``: Image(s) of model predictions.
- ``ref_img``: Reference image(s).
- ``mask``: Mask in Inpainting.
- ``trimap``: Trimap in Matting.
- ``gt_alpha``: Ground truth alpha image in Matting.
- ``pred_alpha``: Predicted alpha image in Matting.
- ``gt_fg``: Ground truth foreground image in Matting.
- ``pred_fg``: Predicted foreground image in Matting.
- ``gt_bg``: Ground truth background image in Matting.
- ``pred_bg``: Predicted background image in Matting.
- ``gt_merged``: Ground truth merged image in Matting.
```

The following sample code demonstrates the components of `DataSample`:

```
    >>> import torch
    >>> import numpy as np
    >>> from mmagic.structures import DataSample
    >>> img_meta = dict(img_shape=(800, 1196, 3))
    >>> img = torch.rand((3, 800, 1196))
    >>> data_sample = DataSample(gt_img=img, metainfo=img_meta)
    >>> assert 'img_shape' in data_sample.metainfo_keys()
    >>> data_sample
        >>># metainfo and data of DataSample
<DataSample(

    META INFORMATION
    img_shape: (800, 1196, 3)

    DATA FIELDS
    gt_img: tensor(3, 800, 1196)
) at 0x1f6a5a99a00>
```

We also support `stack` and `split` operation to handle a batch of data samples.

1. Stack

Stack a list of data samples to one. All tensor fields will be stacked at first dimension. Otherwise the values will be saved in a list.

```
    Args:
        data_samples (Sequence['DataSample']): A sequence of `DataSample` to stack.

    Returns:
        DataSample: The stacked data sample.
```

2. Split

Split a sequence of data sample in the first dimension.

```
        Args:
         allow_nonseq_value (bool): Whether allow non-sequential data in
         split operation. If True, non-sequential data will be copied
         for all split data samples. Otherwise, an error will be
         raised. Defaults to False.

    Returns:
         Sequence[DataSample]: The list of data samples after splitting.
```

The following sample code demonstrates the use of `stack` and `split`:

```
import torch
import numpy as np
from mmagic.structures import DataSample
img_meta1 = img_meta2 = dict(img_shape=(800, 1196, 3))
img1 = torch.rand((3, 800, 1196))
img2 = torch.rand((3, 800, 1196))
data_sample1 = DataSample(gt_img=img1, metainfo=img_meta1)
data_sample2 = DataSample(gt_img=img2, metainfo=img_meta1)
```

```
# stack them and then use as batched-tensor!
data_sample = DataSample.stack([data_sample1, data_sample2])
print(data_sample.gt_img.shape)
    torch.Size([2, 3, 800, 1196])
print(data_sample.metainfo)
    {'img_shape': [(800, 1196, 3), (800, 1196, 3)]}

# split them if you want
data_sample1_, data_sample2_ = data_sample.split()
assert (data_sample1_.gt_img == img1).all()
assert (data_sample2_.gt_img == img2).all()
```

1.16 Data pre-processor

1.16.1 The position of the data preprocessor in the training pipeline.

During the model training process, image data undergoes data augmentation using the transforms provided by mmcv. The augmented data is then loaded into a dataloader. Subsequently, a preprocessor is used to move the data from the CPU to CUDA (GPU), perform padding, and normalize the data.

Below is an example of the `train_pipeline` in the complete configuration file using `configs/_base_/datasets/unpaired_imgs_256x256.py`. The train_pipeline typically defines a sequence of transformations applied to training

images using the mmcv library. This pipeline is designed to prevent redundancy in the transformation functions across different downstream algorithm libraries.

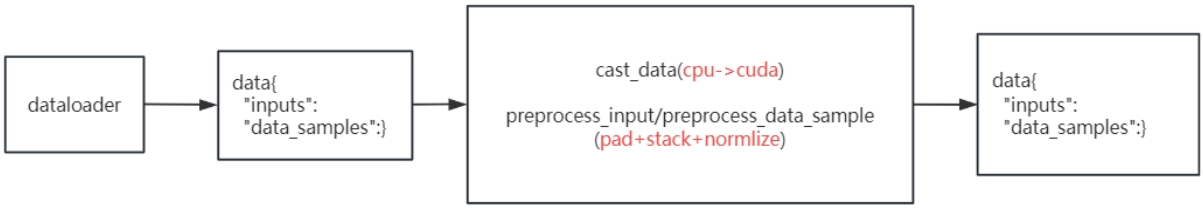
```
...
train_pipeline = [
    dict(color_type='color', key='img_A', type='LoadImageFromFile'),
    dict(color_type='color', key='img_B', type='LoadImageFromFile'),
    dict(auto_remap=True, mapping=dict(img=['img_A', 'img_B',]),
        share_random_params=True,
        transforms=[dict(interpolation='bicubic', scale=(286, 286,), type='Resize'),
                    dict(crop_size=(256, 256,), keys=['img',], random_crop=True, type=
↪'Crop'),],
        type='TransformBroadcaster'),
    dict(direction='horizontal', keys=['img_A', ], type='Flip'),
    dict(direction='horizontal', keys=['img_B', ], type='Flip'),
    dict(mapping=dict(img_mask='img_B', img_photo='img_A'),
        remapping=dict(img_mask='img_mask', img_photo='img_photo'),
        type='KeyMapper'),
    dict(data_keys=['img_photo', 'img_mask',],
        keys=['img_photo', 'img_mask',], type='PackInputs'),
]
...
```

In the `train_step` function in the `mmagic/models/editors/cyclegan/cyclegan.py` script, the data preprocessing steps involve moving, concatenating, and normalizing the transformed data before feeding it into the neural network. Below is an example of the relevant code logic:

```
...
message_hub = MessageHub.get_current_instance()
curr_iter = message_hub.get_info('iter')
data = self.data_preprocessor(data, True)
disc_optimizer_wrapper = optim_wrapper['discriminators']

inputs_dict = data['inputs']
outputs, log_vars = dict(), dict()
...
```

In mmagic, the code implementation for the data processor is located at `mmagic/models/data_preprocessors/data_preprocessor.py`. The data processing workflow is as follows:



1.17 Data flow

1.17.1 Overview of dataflow

The Runner is an "integrator" in MMEngine. It covers all aspects of the framework and shoulders the responsibility of organizing and scheduling nearly all modules, that means the dataflow between all modules also controlled by the `Runner`. As illustrated in the Runner document of MMEngine, the following diagram shows the basic dataflow. In this chapter, we will introduce the dataflow and data format convention between the internal modules managed by the Runner.

In the above diagram, at each training iteration, dataloader loads images from storage and transfer to data preprocessor, data preprocessor would put images to the specific device and stack data to batch, then model accepts the batch data as inputs, finally the outputs of the model would be compute the loss. Since model parameters are freezed when doing evaluation, the model output would be transferred to Evaluator to compute metrics or seed the data to visualize in *Visualizer*.

1.17.2 Data flow between dataset and model

In this section, we will introduce the data flow passing in the dataset in MMagic. About dataset and [transforms] pipeline (https://mmagic.readthedocs.io/en/latest/howto/transforms.html) related tutorials can be found in the development of guidelines.The data flow between dataloader and model can be generally split into four parts:

1. Read the original information of `XXDataset` collected datasets, and carry out data conversion processing through data transform pipeline;
2. use `PackInputs` to pack data from previous transformations into a dictionar;
3. use `collate_fn` to stack a list of tensors into a batched tensor;
4. use `data preprocessor` to move all these data to target device, e.g. GPUS, and unzip the dictionary from the dataloader into a tuple, containing the input images and meta info (`DataSample`).

Data from transform pipeline

In MMagic, different types of 'XXDataset' load the data (LQ) and label (GT), and perform data transformation in different data preprocessing pipelines, and finally package the processed data into a dictionary through `PackInputs`, which contains all the data required for training and testing iterations.

```
@MODELS.register_module()
class BaseEditModel(BaseModel):
    """Base model for image and video editing.
    """
    def forward(self,
```

(continues on next page)

(continued from previous page)

```
                inputs: torch.Tensor,
                data_samples: Optional[List[DataSample]] = None,
                mode: str = 'tensor',
                **kwargs) -> Union[torch.Tensor, List[DataSample], dict]:
        if isinstance(inputs, dict):
            inputs = inputs['img']
        if mode == 'tensor':
            return self.forward_tensor(inputs, data_samples, **kwargs)

        elif mode == 'predict':
            predictions = self.forward_inference(inputs, data_samples,
                                                 **kwargs)
            predictions = self.convert_to_datasample(predictions, data_samples,
                                                     inputs)
            return predictions

        elif mode == 'loss':
            return self.forward_train(inputs, data_samples, **kwargs)
```

```
@MODELS.register_module()
class BaseConditionalGAN(BaseGAN):
    """Base class for Conditional GAM models.
    """
    def forward(self,
                inputs: ForwardInputs,
                data_samples: Optional[list] = None,
                mode: Optional[str] = None) -> List[DataSample]:
        if isinstance(inputs, Tensor):
            noise = inputs
            sample_kwargs = {}
        else:
            noise = inputs.get('noise', None)
            num_batches = get_valid_num_batches(inputs, data_samples)
            noise = self.noise_fn(noise, num_batches=num_batches)
            sample_kwargs = inputs.get('sample_kwargs', dict())
        num_batches = noise.shape[0]

        pass
        ...
```

For example, in the `BaseEditModel` and `BaseConditionalGAN` models, key input including `img` and `noise` are required. At the same time, the corresponding fields should also be exposed in the configuration file,cyclegan_lsgan-id0-resnet-in_1xb1-80kiters_facades.py as an example,

Data from dataloader

After receiving a list of dictionary from dataset, `collect_fn` in dataloader will gather `inputs` in each dict and stack them into a batched tensor. In addition, `data_sample` in each dict will be also collected in a list. Then, it will output a dict, containing the same keys with those of the dict in the received list. Finally, dataloader will output the dict from the `collect_fn`. Detailed documentation can be reference DATASET AND DATALOADER

Data from data preprocessor

Data preprocessor is the last step to process the data before feeding into the model. It will apply image normalization, convert BGR to RGB and move all data to the target device, e.g. GPUs. After above steps, it will output a tuple, containing a list of batched images, and a list of data samples. Detailed documentation can be reference *data_preprocessor*

1.17.3 Data flow between model output and visualizer

MMEngine agreed Abstract Data Element for data transfer Where *data sample* as a more advanced encapsulation can hold more categories of label data. In MMagic, `ConcatImageVisualizer` for visual comparison also controls the visual content through the `add_datasample` function. The specific configuration is as follows.

```
visualizer = dict(
    type='ConcatImageVisualizer',
    vis_backends=[dict(type='LocalVisBackend')],
    fn_key='gt_path',
    img_keys=['gt_img', 'input', 'pred_img'],
    bgr2rgb=True)
```

1.18 How to design your own models

MMagic is built upon MMEngine and MMCV, which enables users to design new models quickly, train and evaluate them easily. In this section, you will learn how to design your own models.

The structure of this guide are as follows:

1.18.1 Overview of models in MMagic

In MMagic, one algorithm can be splited two compents: **Model** and **Module**.

- **Model** are topmost wrappers and always inherint from `BaseModel` provided in MMEngine. **Model** is responsible to network forward, loss calculation and backward, parameters updating, etc. In MMagic, **Model** should be registered as `MODELS`.
- **Module** includes the neural network **architectures** to train or inference, pre-defined **loss classes**, and **data preprocessors** to preprocess the input data batch. **Module** always present as elements of **Model**. In MMagic, **Module** should be registered as **MODULES**.

Take DCGAN model as an example, generator and discriminator are the **Module**, which generate images and discriminate real or fake images. `DCGAN` is the **Model**, which take data from dataloader and train generator and discriminator alternatively.

You can find the implementation of **Model** and **Module** by the following link.

- **Model**:
  - Editors
- **Module**:
  - Layers
  - Losses
  - Data Preprocessor

1.18.2 An example of SRCNN

Here, we take the implementation of the classical image super-resolution model, SRCNN [1], as an example.

Step 1: Define the network of SRCNN

SRCNN is the first deep learning method for single image super-resolution [1]. To implement the network architecture of SRCNN, we need to create a new file `mmagic/models/editors/srgan/sr_resnet.py` and implement `class MSRResNet`.

In this step, we implement `class MSRResNet` by inheriting from `mmengine.models.BaseModule` and define the network architecture in `__init__` function. In particular, we need to use `@MODELS.register_module()` to add the implementation of `class MSRResNet` into the registration of MMagic.

```
import torch.nn as nn
from mmengine.model import BaseModule
from mmagic.registry import MODELS

from mmagic.models.utils import (PixelShufflePack, ResidualBlockNoBN,
                                 default_init_weights, make_layer)


@MODELS.register_module()
class MSRResNet(BaseModule):
    """Modified SRResNet.

    A compacted version modified from SRResNet in "Photo-Realistic Single
```

(continues on next page)

(continued from previous page)

```
    Image Super-Resolution Using a Generative Adversarial Network".

    It uses residual blocks without BN, similar to EDSR.
    Currently, it supports x2, x3 and x4 upsampling scale factor.

    Args:
        in_channels (int): Channel number of inputs.
        out_channels (int): Channel number of outputs.
        mid_channels (int): Channel number of intermediate features.
            Default: 64.
        num_blocks (int): Block number in the trunk network. Default: 16.
        upscale_factor (int): Upsampling factor. Support x2, x3 and x4.
            Default: 4.
    """
    _supported_upscale_factors = [2, 3, 4]

    def __init__(self,
                 in_channels,
                 out_channels,
                 mid_channels=64,
                 num_blocks=16,
                 upscale_factor=4):

        super().__init__()
        self.in_channels = in_channels
        self.out_channels = out_channels
        self.mid_channels = mid_channels
        self.num_blocks = num_blocks
        self.upscale_factor = upscale_factor

        self.conv_first = nn.Conv2d(
            in_channels, mid_channels, 3, 1, 1, bias=True)
        self.trunk_net = make_layer(
            ResidualBlockNoBN, num_blocks, mid_channels=mid_channels)

        # upsampling
        if self.upscale_factor in [2, 3]:
            self.upsample1 = PixelShufflePack(
                mid_channels,
                mid_channels,
                self.upscale_factor,
                upsample_kernel=3)
        elif self.upscale_factor == 4:
            self.upsample1 = PixelShufflePack(
                mid_channels, mid_channels, 2, upsample_kernel=3)
            self.upsample2 = PixelShufflePack(
                mid_channels, mid_channels, 2, upsample_kernel=3)
        else:
            raise ValueError(
                f'Unsupported scale factor {self.upscale_factor}. '
                f'Currently supported ones are '
                f'{self._supported_upscale_factors}.')
```

(continues on next page)

(continued from previous page)

```
        self.conv_hr = nn.Conv2d(
            mid_channels, mid_channels, 3, 1, 1, bias=True)
        self.conv_last = nn.Conv2d(
            mid_channels, out_channels, 3, 1, 1, bias=True)

        self.img_upsampler = nn.Upsample(
            scale_factor=self.upscale_factor,
            mode='bilinear',
            align_corners=False)

        # activation function
        self.lrelu = nn.LeakyReLU(negative_slope=0.1, inplace=True)

        self.init_weights()

    def init_weights(self):
        """Init weights for models.

        Args:
            pretrained (str, optional): Path for pretrained weights. If given
                None, pretrained weights will not be loaded. Defaults to None.
            strict (boo, optional): Whether strictly load the pretrained model.
                Defaults to True.
        """

        for m in [self.conv_first, self.conv_hr, self.conv_last]:
            default_init_weights(m, 0.1)
```

Then, we implement the `forward` function of `class MSRResNet`, which takes as input tensor and then returns the results from `MSRResNet`.

```
    def forward(self, x):
        """Forward function.

        Args:
            x (Tensor): Input tensor with shape (n, c, h, w).

        Returns:
            Tensor: Forward results.
        """

        feat = self.lrelu(self.conv_first(x))
        out = self.trunk_net(feat)

        if self.upscale_factor in [2, 3]:
            out = self.upsample1(out)
        elif self.upscale_factor == 4:
            out = self.upsample1(out)
            out = self.upsample2(out)
```

(continues on next page)

(continued from previous page)

```
        out = self.conv_last(self.lrelu(self.conv_hr(out)))
        upsampled_img = self.img_upsampler(x)
        out += upsampled_img
        return out
```

After the implementation of `class MSRResNet`, we need to update the model list in `mmagic/models/editors/__init__.py`, so that we can import and use `class MSRResNet` by `mmagic.models.editors`.

```
from .srgan.sr_resnet import MSRResNet
```

Step 2: Define the model of SRCNN

After the implementation of the network architecture, we need to define our model `class BaseEditModel` and implement the forward loop of `class BaseEditModel`.

To implement `class BaseEditModel`, we create a new file `mmagic/models/base_models/base_edit_model.py`. Specifically, `class BaseEditModel` inherits from `mmengine.model.BaseModel`. In the `__init__` function, we define the loss functions, training and testing configurations, networks of `class BaseEditModel`.

```
from typing import List, Optional

import torch
from mmengine.model import BaseModel

from mmagic.registry import MODELS
from mmagic.structures import DataSample


@MODELS.register_module()
class BaseEditModel(BaseModel):
    """Base model for image and video editing.

    It must contain a generator that takes frames as inputs and outputs an
    interpolated frame. It also has a pixel-wise loss for training.

    Args:
        generator (dict): Config for the generator structure.
        pixel_loss (dict): Config for pixel-wise loss.
        train_cfg (dict): Config for training. Default: None.
        test_cfg (dict): Config for testing. Default: None.
        init_cfg (dict, optional): The weight initialized config for
            :class:`BaseModule`.
        data_preprocessor (dict, optional): The pre-process config of
            :class:`BaseDataPreprocessor`.

    Attributes:
        init_cfg (dict, optional): Initialization config dict.
        data_preprocessor (:obj:`BaseDataPreprocessor`): Used for
            pre-processing data sampled by dataloader to the format accepted by
            :meth:`forward`. Default: None.
    """
```

(continues on next page)

(continued from previous page)

```
    def __init__(self,
                 generator,
                 pixel_loss,
                 train_cfg=None,
                 test_cfg=None,
                 init_cfg=None,
                 data_preprocessor=None):
        super().__init__(
            init_cfg=init_cfg, data_preprocessor=data_preprocessor)

        self.train_cfg = train_cfg
        self.test_cfg = test_cfg

        # generator
        self.generator = MODELS.build(generator)

        # loss
        self.pixel_loss = MODELS.build(pixel_loss)
```

Since `mmengine.model.BaseModel` provides the basic functions of the algorithmic model, such as weights initialize, batch inputs preprocess, parse losses, and update model parameters. Therefore, the subclasses inherit from BaseModel, i.e., `class BaseEditModel` in this example, only need to implement the forward method, which implements the logic to calculate loss and predictions.

Specifically, the implemented `forward` function of `class BaseEditModel` takes as input `batch_inputs` and `data_samples` and return results according to mode arguments.

```
    def forward(self,
                batch_inputs: torch.Tensor,
                data_samples: Optional[List[DataSample]] = None,
                mode: str = 'tensor',
                **kwargs):
        """Returns losses or predictions of training, validation, testing, and
        simple inference process.

        ``forward`` method of BaseModel is an abstract method, its subclasses
        must implement this method.

        Accepts ``batch_inputs`` and ``data_samples`` processed by
        :attr:`data_preprocessor`, and returns results according to mode
        arguments.

        During non-distributed training, validation, and testing process,
        ``forward`` will be called by ``BaseModel.train_step``,
        ``BaseModel.val_step`` and ``BaseModel.val_step`` directly.

        During distributed data parallel training process,
        ``MMSeparateDistributedDataParallel.train_step`` will first call
        ``DistributedDataParallel.forward`` to enable automatic
        gradient synchronization, and then call ``forward`` to get training
        loss.

        Args:
```

(continues on next page)

(continued from previous page)

```
            batch_inputs (torch.Tensor): batch input tensor collated by
                :attr:`data_preprocessor`.
            data_samples (List[BaseDataElement], optional):
                data samples collated by :attr:`data_preprocessor`.
            mode (str): mode should be one of ``loss``, ``predict`` and
                ``tensor``

                - ``loss``: Called by ``train_step`` and return loss ``dict``
                  used for logging
                - ``predict``: Called by ``val_step`` and ``test_step``
                  and return list of ``BaseDataElement`` results used for
                  computing metric.
                - ``tensor``: Called by custom use to get ``Tensor`` type
                  results.

        Returns:
            ForwardResults:

                - If ``mode == loss``, return a ``dict`` of loss tensor used
                  for backward and logging.
                - If ``mode == predict``, return a ``list`` of
                  :obj:`BaseDataElement` for computing metric
                  and getting inference result.
                - If ``mode == tensor``, return a tensor or ``tuple`` of tensor
                  or ``dict or tensor for custom use.
        """

        if mode == 'tensor':
            return self.forward_tensor(batch_inputs, data_samples, **kwargs)

        elif mode == 'predict':
            return self.forward_inference(batch_inputs, data_samples, **kwargs)

        elif mode == 'loss':
            return self.forward_train(batch_inputs, data_samples, **kwargs)
```

Specifically, in `forward_tensor`, `class BaseEditModel` returns the forward tensors of the network directly.

```
    def forward_tensor(self, batch_inputs, data_samples=None, **kwargs):
        """Forward tensor.
            Returns result of simple forward.

        Args:
            batch_inputs (torch.Tensor): batch input tensor collated by
                :attr:`data_preprocessor`.
            data_samples (List[BaseDataElement], optional):
                data samples collated by :attr:`data_preprocessor`.

        Returns:
            Tensor: result of simple forward.
        """
```

(continues on next page)

(continued from previous page)

```
        feats = self.generator(batch_inputs, **kwargs)

        return feats
```

In `forward_inference` function, `class BaseEditModel` first converts the forward tensors to images and then returns the images as output.

```
    def forward_inference(self, batch_inputs, data_samples=None, **kwargs):
        """Forward inference.
            Returns predictions of validation, testing, and simple inference.

        Args:
            batch_inputs (torch.Tensor): batch input tensor collated by
                :attr:`data_preprocessor`.
            data_samples (List[BaseDataElement], optional):
                data samples collated by :attr:`data_preprocessor`.

        Returns:
            List[DataSample]: predictions.
        """

        feats = self.forward_tensor(batch_inputs, data_samples, **kwargs)
        feats = self.data_preprocessor.destructor(feats)
        predictions = []
        for idx in range(feats.shape[0]):
            predictions.append(
                DataSample(
                    pred_img=feats[idx].to('cpu'),
                    metainfo=data_samples[idx].metainfo))

        return predictions
```

In `forward_train`, `class BaseEditModel` calculate the loss function and returns a dictionary contains the losses as output.

```
    def forward_train(self, batch_inputs, data_samples=None, **kwargs):
        """Forward training.
            Returns dict of losses of training.

        Args:
            batch_inputs (torch.Tensor): batch input tensor collated by
                :attr:`data_preprocessor`.
            data_samples (List[BaseDataElement], optional):
                data samples collated by :attr:`data_preprocessor`.

        Returns:
            dict: Dict of losses.
        """

        feats = self.forward_tensor(batch_inputs, data_samples, **kwargs)
        gt_imgs = [data_sample.gt_img.data for data_sample in data_samples]
        batch_gt_data = torch.stack(gt_imgs)
```

(continues on next page)

(continued from previous page)

```
        loss = self.pixel_loss(feats, batch_gt_data)

        return dict(loss=loss)
```

After the implementation of `class BaseEditModel`, we need to update the model list in `mmagic/models/__init__.py`, so that we can import and use `class BaseEditModel` by `mmagic.models`.

```
from .base_models.base_edit_model import BaseEditModel
```

Step 3: Start training SRCNN

After implementing the network architecture and the forward loop of SRCNN, now we can create a new file `configs/srcnn/srcnn_x4k915_g1_1000k_div2k.py` to set the configurations needed by training SRCNN.

In the configuration file, we need to specify the parameters of our model, `class BaseEditModel`, including the generator network architecture, loss function, additional training and testing configuration, and data preprocessor of input tensors. Please refer to the *Introduction to the loss in MMagic* for more details of losses in MMagic.

```
# model settings
model = dict(
    type='BaseEditModel',
    generator=dict(
        type='SRCNNNet',
        channels=(3, 64, 32, 3),
        kernel_sizes=(9, 1, 5),
        upscale_factor=scale),
    pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean'),
    data_preprocessor=dict(
        type='DataPreprocessor',
        mean=[0., 0., 0.],
        std=[255., 255., 255.],
    ))
```

We also need to specify the training dataloader and testing dataloader according to create your own dataloader. Finally we can start training our own model by

```
python tools/train.py configs/srcnn/srcnn_x4k915_g1_1000k_div2k.py
```

1.18.3 An example of DCGAN

Here, we take the implementation of the classical gan model, DCGAN [2], as an example.

Step 1: Define the network of DCGAN

DCGAN is a classical image generative adversarial network [2]. To implement the network architecture of DCGAN, we need to create tow new files `mmagic/models/editors/dcgan/dcgan_generator.py` and `mmagic/models/editors/dcgan/dcgan_discriminator.py`, and implement generator (`class DCGANGenerator`) and discriminator (`class DCGANDiscriminator`).

In this step, we implement `class DCGANGenerator`, `class DCGANDiscriminator` and define the network architecture in `__init__` function. In particular, we need to use `@MODULES.register_module()` to add the generator and discriminator into the registration of MMagic.

Take the following code as example:

```
import torch.nn as nn
from mmcv.cnn import ConvModule
from mmcv.runner import load_checkpoint
from mmcv.utils.parrots_wrapper import _BatchNorm
from mmengine.logging import MMLogger
from mmengine.model.utils import normal_init

from mmagic.models.builder import MODULES
from ..common import get_module_device


@MODULES.register_module()
class DCGANGenerator(nn.Module):
    def __init__(self,
                 output_scale,
                 out_channels=3,
                 base_channels=1024,
                 input_scale=4,
                 noise_size=100,
                 default_norm_cfg=dict(type='BN'),
                 default_act_cfg=dict(type='ReLU'),
                 out_act_cfg=dict(type='Tanh'),
                 pretrained=None):
        super().__init__()
        self.output_scale = output_scale
        self.base_channels = base_channels
        self.input_scale = input_scale
        self.noise_size = noise_size

        # the number of times for upsampling
        self.num_upsamples = int(np.log2(output_scale // input_scale))

        # output 4x4 feature map
        self.noise2feat = ConvModule(
            noise_size,
            base_channels,
            kernel_size=4,
            stride=1,
            padding=0,
            conv_cfg=dict(type='ConvTranspose2d'),
            norm_cfg=default_norm_cfg,
```

(continues on next page)

(continued from previous page)

```
            act_cfg=default_act_cfg)

        # build up upsampling backbone (excluding the output layer)
        upsampling = []
        curr_channel = base_channels
        for _ in range(self.num_upsamples - 1):
            upsampling.append(
                ConvModule(
                    curr_channel,
                    curr_channel // 2,
                    kernel_size=4,
                    stride=2,
                    padding=1,
                    conv_cfg=dict(type='ConvTranspose2d'),
                    norm_cfg=default_norm_cfg,
                    act_cfg=default_act_cfg))

            curr_channel //= 2

        self.upsampling = nn.Sequential(*upsampling)

        # output layer
        self.output_layer = ConvModule(
            curr_channel,
            out_channels,
            kernel_size=4,
            stride=2,
            padding=1,
            conv_cfg=dict(type='ConvTranspose2d'),
            norm_cfg=None,
            act_cfg=out_act_cfg)
```

Then, we implement the `forward` function of `DCGANGenerator`, which takes as `noise` tensor or `num_batches` and then returns the results from `DCGANGenerator`.

```
    def forward(self, noise, num_batches=0, return_noise=False):
        noise_batch = noise_batch.to(get_module_device(self))
        x = self.noise2feat(noise_batch)
        x = self.upsampling(x)
        x = self.output_layer(x)
        return x
```

If you want to implement specific weights initialization method for you network, you need add `init_weights` function by yourself.

```
    def init_weights(self, pretrained=None):
        if isinstance(pretrained, str):
            logger = MMLogger.get_current_instance()
            load_checkpoint(self, pretrained, strict=False, logger=logger)
        elif pretrained is None:
            for m in self.modules():
                if isinstance(m, (nn.Conv2d, nn.ConvTranspose2d)):
```

(continues on next page)

(continued from previous page)

```
                    normal_init(m, 0, 0.02)
                elif isinstance(m, _BatchNorm):
                    nn.init.normal_(m.weight.data)
                    nn.init.constant_(m.bias.data, 0)
        else:
            raise TypeError('pretrained must be a str or None but'
                            f' got {type(pretrained)} instead.')
```

After the implementation of class `DCGANGenerator`, we need to update the model list in `mmagic/models/editors/__init__.py`, so that we can import and use class `DCGANGenerator` by `mmagic.models.editors`.

Implementation of Class `DCGANDiscriminator` follows the similar logic, and you can find the implementation here.

Step 2: Design the model of DCGAN

After the implementation of the network **Module**, we need to define our **Model** class `DCGAN`.

Your **Model** should inherit from `BaseModel` provided by MMEngine and implement three functions, `train_step`, `val_step` and `test_step`.

- `train_step`: This function is responsible to update the parameters of the network and called by MMEngine's Loop (`IterBasedTrainLoop` or `EpochBasedTrainLoop`). `train_step` take data batch and `OptimWrapper` as input and return a dict of log.
- `val_step`: This function is responsible for getting output for validation during the training process. and is called by `MultiValLoop`.
- `test_step`: This function is responsible for getting output in test process and is called by `MultiTestLoop`.

  Note that, in `train_step`, `val_step` and `test_step`, `DataPreprocessor` is called to preprocess the input data batch before feed them to the neural network. To know more about `DataPreprocessor` please refer to this file and this tutorial.

For simplify using, we provide `BaseGAN` class in MMagic, which implements generic `train_step`, `val_step` and `test_step` function for GAN models. With `BaseGAN` as base class, each specific GAN algorithm only need to implement `train_generator` and `train_discriminator`.

In `train_step`, we support data preprocessing, gradient accumulation (realized by `OptimWrapper`) and expontial moving averate (EMA) realized by (`ExponentialMovingAverage`). With `BaseGAN.train_step`, each specific GAN algorithm only need to implement `train_generator` and `train_discriminator`.

```
    def train_step(self, data: dict,
                   optim_wrapper: OptimWrapperDict) -> Dict[str, Tensor]:
        message_hub = MessageHub.get_current_instance()
        curr_iter = message_hub.get_info('iter')
        data = self.data_preprocessor(data, True)
        disc_optimizer_wrapper: OptimWrapper = optim_wrapper['discriminator']
        disc_accu_iters = disc_optimizer_wrapper._accumulative_counts

        # train discriminator, use context manager provided by MMEngine
        with disc_optimizer_wrapper.optim_context(self.discriminator):
            # train_discriminator should be implemented!
            log_vars = self.train_discriminator(
                **data, optimizer_wrapper=disc_optimizer_wrapper)
```

(continues on next page)

(continued from previous page)

```
        # add 1 to `curr_iter` because iter is updated in train loop.
        # Whether to update the generator. We update generator with
        # discriminator is fully updated for `self.n_discriminator_steps`
        # iterations. And one full updating for discriminator contains
        # `disc_accu_counts` times of grad accumulations.
        if (curr_iter + 1) % (self.discriminator_steps * disc_accu_iters) == 0:
            set_requires_grad(self.discriminator, False)
            gen_optimizer_wrapper = optim_wrapper['generator']
            gen_accu_iters = gen_optimizer_wrapper._accumulative_counts

            log_vars_gen_list = []
            # init optimizer wrapper status for generator manually
            gen_optimizer_wrapper.initialize_count_status(
                self.generator, 0, self.generator_steps * gen_accu_iters)
            # update generator, use context manager provided by MMEngine
            for _ in range(self.generator_steps * gen_accu_iters):
                with gen_optimizer_wrapper.optim_context(self.generator):
                    # train_generator should be implemented!
                    log_vars_gen = self.train_generator(
                        **data, optimizer_wrapper=gen_optimizer_wrapper)

                log_vars_gen_list.append(log_vars_gen)
            log_vars_gen = gather_log_vars(log_vars_gen_list)
            log_vars_gen.pop('loss', None)  # remove 'loss' from gen logs

            set_requires_grad(self.discriminator, True)

            # only do ema after generator update
            if self.with_ema_gen and (curr_iter + 1) >= (
                    self.ema_start * self.discriminator_steps *
                    disc_accu_iters):
                self.generator_ema.update_parameters(
                    self.generator.module
                    if is_model_wrapper(self.generator) else self.generator)

            log_vars.update(log_vars_gen)

        # return the log dict
        return log_vars
```

In `val_step` and `test_step`, we call data preprocessing and `BaseGAN.forward` progressively.

```
    def val_step(self, data: dict) -> SampleList:
        data = self.data_preprocessor(data)
        # call `forward`
        outputs = self(**data)
        return outputs

    def test_step(self, data: dict) -> SampleList:
        data = self.data_preprocessor(data)
        # call `orward`
        outputs = self(**data)
```

(continues on next page)

(continued from previous page)

```
        return outputs
```

Then, we implement `train_generator` and `train_discriminator` in `DCGAN` class.

```
from typing import Dict, Tuple

import torch
import torch.nn.functional as F
from mmengine.optim import OptimWrapper
from torch import Tensor

from mmagic.registry import MODELS
from .base_gan import BaseGAN


@MODELS.register_module()
class DCGAN(BaseGAN):
    def disc_loss(self, disc_pred_fake: Tensor,
                  disc_pred_real: Tensor) -> Tuple:
        losses_dict = dict()
        losses_dict['loss_disc_fake'] = F.binary_cross_entropy_with_logits(
            disc_pred_fake, 0. * torch.ones_like(disc_pred_fake))
        losses_dict['loss_disc_real'] = F.binary_cross_entropy_with_logits(
            disc_pred_real, 1. * torch.ones_like(disc_pred_real))

        loss, log_var = self.parse_losses(losses_dict)
        return loss, log_var

    def gen_loss(self, disc_pred_fake: Tensor) -> Tuple:
        losses_dict = dict()
        losses_dict['loss_gen'] = F.binary_cross_entropy_with_logits(
            disc_pred_fake, 1. * torch.ones_like(disc_pred_fake))
        loss, log_var = self.parse_losses(losses_dict)
        return loss, log_var

    def train_discriminator(
            self, inputs, data_sample,
            optimizer_wrapper: OptimWrapper) -> Dict[str, Tensor]:
        real_imgs = inputs['img']

        num_batches = real_imgs.shape[0]

        noise_batch = self.noise_fn(num_batches=num_batches)
        with torch.no_grad():
            fake_imgs = self.generator(noise=noise_batch, return_noise=False)

        disc_pred_fake = self.discriminator(fake_imgs)
        disc_pred_real = self.discriminator(real_imgs)

        parsed_losses, log_vars = self.disc_loss(disc_pred_fake,
                                                 disc_pred_real)
        optimizer_wrapper.update_params(parsed_losses)
```

(continues on next page)

(continued from previous page)

```
        return log_vars

    def train_generator(self, inputs, data_sample,
                        optimizer_wrapper: OptimWrapper) -> Dict[str, Tensor]:
        num_batches = inputs['img'].shape[0]

        noise = self.noise_fn(num_batches=num_batches)
        fake_imgs = self.generator(noise=noise, return_noise=False)

        disc_pred_fake = self.discriminator(fake_imgs)
        parsed_loss, log_vars = self.gen_loss(disc_pred_fake)

        optimizer_wrapper.update_params(parsed_loss)
        return log_vars
```

After the implementation of `class DCGAN`, we need to update the model list in `mmagic/models/__init__.py`, so that we can import and use `class DCGAN` by `mmagic.models`.

Step 3: Start training DCGAN

After implementing the network **Module** and the **Model** of DCGAN, now we can create a new file `configs/dcgan/dcgan_1xb128-5epoches_lsun-bedroom-64x64.py` to set the configurations needed by training DCGAN.

In the configuration file, we need to specify the parameters of our model, `class DCGAN`, including the generator network architecture and data preprocessor of input tensors.

```
# model settings
model = dict(
    type='DCGAN',
    noise_size=100,
    data_preprocessor=dict(type='GANDataPreprocessor'),
    generator=dict(type='DCGANGenerator', output_scale=64, base_channels=1024),
    discriminator=dict(
        type='DCGANDiscriminator',
        input_scale=64,
        output_scale=4,
        out_channels=1))
```

We also need to specify the training dataloader and testing dataloader according to *create your own dataloader*. Finally we can start training our own model by

```
python tools/train.py configs/dcgan/dcgan_1xb128-5epoches_lsun-bedroom-64x64.py
```

1.18.4 References

1. Dong, Chao and Loy, Chen Change and He, Kaiming and Tang, Xiaoou. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015.
2. Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

1.19 How to prepare your own datasets

In this document, we will introduce the design of each datasets in MMagic and how users can design their own dataset.

1.19.1 Supported Data Format

In MMagic, all datasets are inherited from `BaseDataset`. Each dataset load the list of data info (e.g., data path) by `load_data_list`. In `__getitem__`, `prepare_data` is called to get the preprocessed data. In `prepare_data`, data loading pipeline consists of the following steps:

1. fetch the data info by passed index, implemented by `get_data_info`
2. apply data transforms to the data, implemented by `pipeline`

BasicImageDataset

BasicImageDataset `mmagic.datasets.BasicImageDataset` General image dataset designed for low-level vision tasks with image, such as image super-resolution, inpainting and unconditional image generation. The annotation file is optional.

If use annotation file, the annotation format can be shown as follows.

```
   Case 1 (CelebA-HQ):

       000001.png
       000002.png

   Case 2 (DIV2K):

       0001_s001.png (480,480,3)
       0001_s002.png (480,480,3)
       0001_s003.png (480,480,3)
       0002_s001.png (480,480,3)
       0002_s002.png (480,480,3)

   Case 3 (Vimeo90k):

       00001/0266 (256, 448, 3)
       00001/0268 (256, 448, 3)
```

Here we give several examples showing how to use `BasicImageDataset`. Assume the file structure as the following:

```
mmagic (root)
├── mmagic
├── tools
├── configs
├── data
│   ├── DIV2K
│   │   ├── DIV2K_train_HR
│   │   │   ├── image.png
│   │   ├── DIV2K_train_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   │   │   │   ├── image_x4.png
│   │   ├── DIV2K_valid_HR
│   │   ├── DIV2K_valid_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   ├── places
│   │   ├── test_set
│   │   ├── train_set
│   │   ├── meta
│   │   │   ├── Places365_train.txt
│   │   │   ├── Places365_val.txt
│   ├── celebahq
│   │   ├── imgs_1024
```

Case 1: Loading DIV2K dataset for training a SISR model.

```
dataset = BasicImageDataset(
    ann_file='',
    metainfo=dict(
        dataset_type='div2k',
        task_name='sisr'),
    data_root='data/DIV2K',
    data_prefix=dict(
        gt='DIV2K_train_HR', img='DIV2K_train_LR_bicubic/X4'),
    filename_tmpl=dict(img='{}_x4', gt='{}'),
    pipeline=[])
```

Case 2: Loading places dataset for training an inpainting model.

```
dataset = BasicImageDataset(
    ann_file='meta/Places365_train.txt',
    metainfo=dict(
        dataset_type='places365',
        task_name='inpainting'),
    data_root='data/places',
    data_prefix=dict(gt='train_set'),
    pipeline=[])
```

Case 3: Loading CelebA-HQ dataset for training an PGGAN.

```
dataset = BasicImageDataset(
        pipeline=[],
        data_root='./data/celebahq/imgs_1024')
```

BasicFramesDataset

BasicFramesDataset `mmagic.datasets.BasicFramesDataset` General frames dataset designed for low-level vision tasks with frames, such as video super-resolution and video frame interpolation. The annotation file is optional.

If use annotation file, the annotation format can be shown as follows.

```
Case 1 (Vid4):

   calendar 41
   city 34
   foliage 49
   walk 47

Case 2 (REDS):

   000/00000000.png (720, 1280, 3)
   000/00000001.png (720, 1280, 3)

Case 3 (Vimeo90k):

   00001/0266 (256, 448, 3)
   00001/0268 (256, 448, 3)
```

Assume the file structure as the following:

```
mmagic (root)
├── mmagic
├── tools
├── configs
├── data
│   ├── Vid4
│   │   ├── BIx4
│   │   │   ├── city
│   │   │   │   ├── img1.png
│   │   ├── GT
│   │   │   ├── city
│   │   │   │   ├── img1.png
│   │   ├── meta_info_Vid4_GT.txt
│   ├── vimeo-triplet
│   │   ├── sequences
│   │   │   ├── 00001
│   │   │   │   ├── 0389
│   │   │   │   │   ├── img1.png
│   │   │   │   │   ├── img2.png
│   │   │   │   │   ├── img3.png
│   │   ├── tri_trainlist.txt
```

Case 1: Loading Vid4 dataset for training a VSR model.

```
dataset = BasicFramesDataset(
    ann_file='meta_info_Vid4_GT.txt',
    metainfo=dict(dataset_type='vid4', task_name='vsr'),
    data_root='data/Vid4',
    data_prefix=dict(img='BIx4', gt='GT'),
    pipeline=[],
    depth=2,
    num_input_frames=5)
```

Case 2: Loading Vimeo90k dataset for training a VFI model.

```
dataset = BasicFramesDataset(
    ann_file='tri_trainlist.txt',
    metainfo=dict(dataset_type='vimeo90k', task_name='vfi'),
    data_root='data/vimeo-triplet',
    data_prefix=dict(img='sequences', gt='sequences'),
    pipeline=[],
    depth=2,
    load_frames_list=dict(
        img=['img1.png', 'img3.png'], gt=['img2.png']))
```

BasicConditonalDataset

BasicConditonalDataset `mmagic.datasets.BasicConditonalDataset` is designed for conditional GANs (e.g., SAGAN, BigGAN). This dataset support load label for the annotation file. `BasicConditonalDataset` support three kinds of annotation as follow:

1. Annotation file read by line (e.g., txt)

Sample files structure:

```
data_prefix/
├── folder_1
│   ├── xxx.png
│   ├── xxy.png
│   └── ...
└── folder_2
    ├── 123.png
    ├── nsdf3.png
    └── ...
```

Sample annotation file (the first column is the image path and the second column is the index of category):

```
folder_1/xxx.png 0
folder_1/xxy.png 1
folder_2/123.png 5
folder_2/nsdf3.png 3
...
```

Config example for ImageNet dataset:

```
dataset=dict(
    type='BasicConditionalDataset,
    data_root='./data/imagenet/',
    ann_file='meta/train.txt',
    data_prefix='train',
    pipeline=train_pipeline),
```

2. Dict-based annotation file (e.g., json):

Sample files structure:

```
data_prefix/
├── folder_1
│   ├── xxx.png
│   ├── xxy.png
│   └── ...
└── folder_2
    ├── 123.png
    ├── nsdf3.png
    └── ...
```

Sample annotation file (the key is the image path and the value column is the label):

```
    {
        "folder_1/xxx.png": [1, 2, 3, 4],
        "folder_1/xxy.png": [2, 4, 1, 0],
        "folder_2/123.png": [0, 9, 8, 1],
        "folder_2/nsdf3.png", [1, 0, 0, 2],
        ...
    }
```

Config example for EG3D (shapenet-car) dataset:

```
dataset = dict(
    type='BasicConditionalDataset',
    data_root='./data/eg3d/shapenet-car',
    ann_file='annotation.json',
    pipeline=train_pipeline)
```

In this kind of annotation, labels can be any type and not restricted to an index.

3. Folder-based annotation (no annotation file need):

Sample files structure:

```
    data_prefix/
    ├── class_x
    │   ├── xxx.png
    │   ├── xxy.png
    │   └── ...
    │       └── xxz.png
    └── class_y
        ├── 123.png
        ├── nsdf3.png
        ├── ...
        └── asd932_.png
```

If the annotation file is specified, the dataset will be generated by the first two ways, otherwise, try the third way.

ImageNet Dataset and CIFAR10 Dataset

ImageNet Dataset `mmagic.datasets.ImageNet` and **CIFAR10 Dataset**`mmagic.datasets.CIFAR10` are datasets specific designed for ImageNet and CIFAR10 datasets. Both two datasets are encapsulation of `BasicConditionalDataset`. You can used them to load data from ImageNet dataset and CIFAR10 dataset easily.

Config example for ImageNet:

```
pipeline = [
    dict(type='LoadImageFromFile', key='img'),
    dict(type='RandomCropLongEdge', keys=['img']),
    dict(type='Resize', scale=(128, 128), keys=['img'], backend='pillow'),
    dict(type='Flip', keys=['img'], flip_ratio=0.5, direction='horizontal'),
    dict(type='PackInputs')
]
```

(continues on next page)

(continued from previous page)

```
dataset=dict(
    type='ImageNet',
    data_root='./data/imagenet/',
    ann_file='meta/train.txt',
    data_prefix='train',
    pipeline=pipeline),
```

Config example for CIFAR10:

```
pipeline = [dict(type='PackInputs')]

dataset = dict(
    type='CIFAR10',
    data_root='./data',
    data_prefix='cifar10',
    test_mode=False,
    pipeline=pipeline)
```

AdobeComp1kDataset

AdobeComp1kDataset `mmagic.datasets.AdobeComp1kDataset` Adobe composition-1k dataset.

The dataset loads (alpha, fg, bg) data and apply specified transforms to the data. You could specify whether composite merged image online or load composited merged image in pipeline.

Example for online comp-1k dataset:

```
[
   {
       "alpha": 'alpha/000.png',
       "fg": 'fg/000.png',
       "bg": 'bg/000.png'
   },
   {
       "alpha": 'alpha/001.png',
       "fg": 'fg/001.png',
       "bg": 'bg/001.png'
   },
]
```

Example for offline comp-1k dataset:

```
[
  {
      "alpha": 'alpha/000.png',
      "merged": 'merged/000.png',
      "fg": 'fg/000.png',
      "bg": 'bg/000.png'
  },
  {
      "alpha": 'alpha/001.png',
```

(continues on next page)

(continued from previous page)

```
      "merged": 'merged/001.png',
      "fg": 'fg/001.png',
      "bg": 'bg/001.png'
  },
]
```

GrowScaleImgDataset

`GrowScaleImgDataset` is designed for dynamic GAN models (e.g., PGGAN and StyleGANv1). In this dataset, we support switching the data root during training to load training images of different resolutions. This procedure is implemented by `GrowScaleImgDataset.update_annotations` and is called by `PGGANFetchDataHook.before_train_iter` in the training process.

```
def update_annotations(self, curr_scale):
    # determine if the data root needs to be updated
    if curr_scale == self._actual_curr_scale:
        return False

    # fetch new data root by resolution (scale)
    for scale in self._img_scales:
        if curr_scale <= scale:
            self._curr_scale = scale
            break
        if scale == self._img_scales[-1]:
            assert RuntimeError(
                f'Cannot find a suitable scale for {curr_scale}')
    self._actual_curr_scale = curr_scale
    self.data_root = self.data_roots[str(self._curr_scale)]

    # reload the data list with new data root
    self.load_data_list()

    # print basic dataset information to check the validity
    print_log('Update Dataset: ' + repr(self), 'current')
    return True
```

SinGANDataset

`SinGANDataset` is designed for SinGAN's training. In SinGAN's training, we do not iterate the images in the dataset but return a consistent preprocessed image dict.

Therefore, we bypass the default data loading logic of `BaseDataset` because we do not need to load the corresponding image data based on the given index.

```
def load_data_list(self, min_size, max_size, scale_factor_init):
    # load single image
    real = mmcv.imread(self.data_root)
    self.reals, self.scale_factor, self.stop_scale = create_real_pyramid(
        real, min_size, max_size, scale_factor_init)
```

(continues on next page)

(continued from previous page)

```
    self.data_dict = {}

    # generate multi scale image
    for i, real in enumerate(self.reals):
        self.data_dict[f'real_scale{i}'] = real

    self.data_dict['input_sample'] = np.zeros_like(
        self.data_dict['real_scale0']).astype(np.float32)

def __getitem__(self, index):
    # directly return the transformed data dict
    return self.pipeline(self.data_dict)
```

PairedImageDataset

`PairedImageDataset` is designed for translation models that needs paired training data (e.g., Pix2Pix). The directory structure is shown below. Each image files are the concatenation of the image pair.

```
./data/dataset_name/
├── test
│   └── XXX.jpg
└── train
    └── XXX.jpg
```

In `PairedImageDataset`, we scan the file list in `load_data_list` and save path in `pair_path` field to fit the `LoadPairedImageFromFile` transformation.

```
def load_data_list(self):
    data_infos = []
    pair_paths = sorted(self.scan_folder(self.data_root))
    for pair_path in pair_paths:
        # save path in the specific field
        data_infos.append(dict(pair_path=pair_path))

    return data_infos
```

UnpairedImageDataset

`UnpairedImageDataset` is designed for translation models that do not need paired data (e.g., CycleGAN). The directory structure is shown below.

```
./data/dataset_name/
├── testA
│   └── XXX.jpg
├── testB
│   └── XXX.jpg
├── trainA
│   └── XXX.jpg
└── trainB
    └── XXX.jpg
```

In this dataset, we overwrite `__getitem__` function to load random image pair in the training process.

```
def __getitem__(self, idx):
    if not self.test_mode:
        return self.prepare_train_data(idx)

    return self.prepare_test_data(idx)

def prepare_train_data(self, idx):
    img_a_path = self.data_infos_a[idx % self.len_a]['path']
    idx_b = np.random.randint(0, self.len_b)
    img_b_path = self.data_infos_b[idx_b]['path']
    results = dict()
    results[f'img_{self.domain_a}_path'] = img_a_path
    results[f'img_{self.domain_b}_path'] = img_b_path
    return self.pipeline(results)

def prepare_test_data(self, idx):
    img_a_path = self.data_infos_a[idx % self.len_a]['path']
    img_b_path = self.data_infos_b[idx % self.len_b]['path']
    results = dict()
    results[f'img_{self.domain_a}_path'] = img_a_path
    results[f'img_{self.domain_b}_path'] = img_b_path
    return self.pipeline(results)
```

1.19.2 Design a new dataset

If you want to create a dataset for a new low level CV task (e.g. denoise, derain, defog, and de-reflection) or existing dataset format doesn't meet your need, you can reorganize new data formats to existing format.

Or create a new dataset in `mmagic/datasets` to load the data.

Inheriting from the base class of datasets such as `BasicImageDataset` and `BasicFramesDataset` will make it easier to create a new dataset.

And you can create a new dataset inherited from BaseDataset which is the base class of datasets in MMEngine.

Here is an example of creating a dataset for video frame interpolation:

```
from .basic_frames_dataset import BasicFramesDataset
from mmagic.registry import DATASETS


@DATASETS.register_module()
class NewVFIDataset(BasicFramesDataset):
    """Introduce the dataset

    Examples of file structure.

    Args:
        pipeline (list[dict | callable]): A sequence of data transformations.
        folder (str | :obj:`Path`): Path to the folder.
        ann_file (str | :obj:`Path`): Path to the annotation file.
        test_mode (bool): Store `True` when building test dataset.
```

(continues on next page)

(continued from previous page)

```
            Default: `False`.
    """

    def __init__(self, ann_file, metainfo, data_root, data_prefix,
                 pipeline, test_mode=False):
        super().__init__(ann_file, metainfo, data_root, data_prefix,
                         pipeline, test_mode)
        self.data_infos = self.load_annotations()

    def load_annotations(self):
        """Load annoations for the dataset.

        Returns:
            list[dict]: A list of dicts for paired paths and other information.
        """
        data_infos = []
        ...
        return data_infos
```

Welcome to submit new dataset classes to MMagic.

Repeat dataset

We use RepeatDataset as wrapper to repeat the dataset. For example, suppose the original dataset is Dataset_A, to repeat it, the config looks like the following

```
dataset_A_train = dict(
        type='RepeatDataset',
        times=N,
        dataset=dict(  # This is the original config of Dataset_A
            type='Dataset_A',
            ...
            pipeline=train_pipeline
        )
    )
```

You may refer to tutorial in MMEngine.

1.20 How to design your own data transforms

In this tutorial, we introduce the design of transforms pipeline in MMagic.

The structure of this guide are as follows:

1.20.1 Data pipelines in MMagic

Following typical conventions, we use `Dataset` and `DataLoader` for data loading with multiple workers. `Dataset` returns a dict of data items corresponding the arguments of models' forward method.

The data preparation pipeline and the dataset is decomposed. Usually a dataset defines how to process the annotations and a data pipeline defines all the steps to prepare a data dict.

A pipeline consists of a sequence of operations. Each operation takes a dict as input and also output a dict for the next transform.

The operations are categorized into data loading, pre-processing, and formatting

In MMagic, all data transformations are inherited from `BaseTransform`. The input and output types of transformations are both dict.

A simple example of data transform

```
>>> from mmagic.transforms import LoadPairedImageFromFile
>>> transforms = LoadPairedImageFromFile(
>>>     key='pair',
>>>     domain_a='horse',
>>>     domain_b='zebra',
>>>     flag='color'),
>>> data_dict = {'pair_path': './data/pix2pix/facades/train/1.png'}
>>> data_dict = transforms(data_dict)
>>> print(data_dict.keys())
dict_keys(['pair_path', 'pair', 'pair_ori_shape', 'img_mask', 'img_photo', 'img_mask_path
↪', 'img_photo_path', 'img_mask_ori_shape', 'img_photo_ori_shape'])
```

Generally, the last step of the transforms pipeline must be `PackInputs`. `PackInputs` will pack the processed data into a dict containing two fields: `inputs` and `data_samples`. `inputs` is the variable you want to use as the model's input, which can be the type of `torch.Tensor`, dict of `torch.Tensor`, or any type you want. `data_samples` is a list of `DataSample`. Each `DataSample` contains groundtruth and necessary information for corresponding input.

An example of BasicVSR

Here is a pipeline example for BasicVSR.

```
train_pipeline = [
    dict(type='LoadImageFromFile', key='img', channel_order='rgb'),
    dict(type='LoadImageFromFile', key='gt', channel_order='rgb'),
    dict(type='SetValues', dictionary=dict(scale=scale)),
    dict(type='PairedRandomCrop', gt_patch_size=256),
    dict(
        type='Flip',
        keys=['img', 'gt'],
        flip_ratio=0.5,
        direction='horizontal'),
    dict(
        type='Flip', keys=['img', 'gt'], flip_ratio=0.5, direction='vertical'),
    dict(type='RandomTransposeHW', keys=['img', 'gt'], transpose_ratio=0.5),
    dict(type='MirrorSequence', keys=['img', 'gt']),
    dict(type='PackInputs')
]

val_pipeline = [
    dict(type='GenerateSegmentIndices', interval_list=[1]),
    dict(type='LoadImageFromFile', key='img', channel_order='rgb'),
    dict(type='LoadImageFromFile', key='gt', channel_order='rgb'),
    dict(type='PackInputs')
]

test_pipeline = [
    dict(type='LoadImageFromFile', key='img', channel_order='rgb'),
    dict(type='LoadImageFromFile', key='gt', channel_order='rgb'),
    dict(type='MirrorSequence', keys=['img']),
    dict(type='PackInputs')
]
```

For each operation, we list the related dict fields that are added/updated/removed, the dict fields marked by '*' are optional.

An example of Pix2Pix

Here is a pipeline example for Pix2Pix training on aerial2maps dataset.

```
source_domain = 'aerial'
target_domain = 'map'

pipeline = [
    dict(
        type='LoadPairedImageFromFile',
        io_backend='disk',
        key='pair',
        domain_a=domain_a,
        domain_b=domain_b,
        flag='color'),
```

(continues on next page)

(continued from previous page)

```
    dict(
        type='TransformBroadcaster',
        mapping={'img': [f'img_{domain_a}', f'img_{domain_b}']},
        auto_remap=True,
        share_random_params=True,
        transforms=[
            dict(
                type='mmagic.Resize', scale=(286, 286),
                interpolation='bicubic'),
            dict(type='mmagic.FixedCrop', crop_size=(256, 256))
        ]),
    dict(
        type='Flip',
        keys=[f'img_{domain_a}', f'img_{domain_b}'],
        direction='horizontal'),
    dict(
        type='PackInputs',
        keys=[f'img_{domain_a}', f'img_{domain_b}', 'pair'])
```

1.20.2 Supported transforms in MMagic

Data loading

Pre-processing

Formatting

Albumentations

MMagic support adding custom transformations from Albumentations library. Please visit https://albumentations.ai/docs/getting_started/transforms_and_targets to get more information.

An example of Albumentations's `transforms` is as followed:

```
albu_transforms = [
   dict(
         type='Resize',
         height=100,
         width=100,
   ),
   dict(
         type='RandomFog',
         p=0.5,
   ),
   dict(
         type='RandomRain',
         p=0.5
   ),
   dict(
         type='RandomSnow',
         p=0.5,
```

(continues on next page)

(continued from previous page)

```
    ),
]
pipeline = [
    dict(
        type='LoadImageFromFile',
        key='img',
        color_type='color',
        channel_order='rgb',
        imdecode_backend='cv2'),
    dict(
        type='Albumentations',
        keys=['img'],
        transforms=albu_transforms),
    dict(type='PackInputs')
]
```

1.20.3 Extend and use custom pipelines

A simple example of MyTransform

1. Write a new pipeline in a file, e.g., in `my_pipeline.py`. It takes a dict as input and returns a dict.

```
import random
from mmcv.transforms import BaseTransform
from mmagic.registry import TRANSFORMS


@TRANSFORMS.register_module()
class MyTransform(BaseTransform):
    """Add your transform

    Args:
        p (float): Probability of shifts. Default 0.5.
    """

    def __init__(self, p=0.5):
        self.p = p

    def transform(self, results):
        if random.random() > self.p:
            results['dummy'] = True
        return results

    def __repr__(self):

        repr_str = self.__class__.__name__
        repr_str += (f'(p={self.p})')

        return repr_str
```

2. Import and use the pipeline in your config file.

Make sure the import is relative to where your train script is located.

```
train_pipeline = [
    ...
    dict(type='MyTransform', p=0.2),
    ...
]
```

An example of flipping

Here we use a simple flipping transformation as example:

```
import random
import mmcv
from mmcv.transforms import BaseTransform, TRANSFORMS

@TRANSFORMS.register_module()
class MyFlip(BaseTransform):
    def __init__(self, direction: str):
        super().__init__()
        self.direction = direction

    def transform(self, results: dict) -> dict:
        img = results['img']
        results['img'] = mmcv.imflip(img, direction=self.direction)
        return results
```

Thus, we can instantiate a `MyFlip` object and use it to process the data dict.

```
import numpy as np

transform = MyFlip(direction='horizontal')
data_dict = {'img': np.random.rand(224, 224, 3)}
data_dict = transform(data_dict)
processed_img = data_dict['img']
```

Or, we can use `MyFlip` transformation in data pipeline in our config file.

```
pipeline = [
    ...
    dict(type='MyFlip', direction='horizontal'),
    ...
]
```

Note that if you want to use `MyFlip` in config, you must ensure the file containing `MyFlip` is imported during the program run.

1.21 How to design your own loss functions

losses are registered as LOSSES in MMagic. Customizing losses is similar to customizing any other model. This section is mainly for clarifying the design of loss modules in MMagic. Importantly, when writing your own loss modules, you should follow the same design, so that the new loss module can be adopted in our framework without extra effort.

This guides includes:

1.21.1 Introduction to supported losses

For convenient usage, you can directly use default loss calculation process we set for concrete algorithms like lsgan, biggan, styleganv2 etc. Take stylegan2 as an example, we use R1 gradient penalty and generator path length regularization as configurable losses, and users can adjust related arguments like r1_loss_weight and g_reg_weight.

```
# stylegan2_base.py
loss_config = dict(
    r1_loss_weight=10. / 2. * d_reg_interval,
    r1_interval=d_reg_interval,
    norm_mode='HWC',
    g_reg_interval=g_reg_interval,
    g_reg_weight=2. * g_reg_interval,
    pl_batch_shrink=2)

model = dict(
    type='StyleGAN2',
    xxx,
    loss_config=loss_config)
```

1.21.2 Design a new loss function

An example of MSELoss

In general, to implement a loss module, we will write a function implementation and then wrap it with a class implementation. Take the MSELoss as an example:

```
@masked_loss
def mse_loss(pred, target):
    return F.mse_loss(pred, target, reduction='none')

@LOSSES.register_module()
class MSELoss(nn.Module):

    def __init__(self, loss_weight=1.0, reduction='mean', sample_wise=False):
        # codes can be found in ``mmagic/models/losses/pixelwise_loss.py``

    def forward(self, pred, target, weight=None, **kwargs):
        # codes can be found in ``mmagic/models/losses/pixelwise_loss.py``
```

Given the definition of the loss, we can now use the loss by simply defining it in the configuration file:

```
pixel_loss=dict(type='MSELoss', loss_weight=1.0, reduction='mean')
```

Note that `pixel_loss` above must be defined in the model. Please refer to `customize_models` for more details. Similar to model customization, in order to use your customized loss, you need to import the loss in `mmagic/models/losses/__init__.py` after writing it.

An example of DiscShiftLoss

In general, to implement a loss module, we will write a function implementation and then wrap it with a class implementation. However, in `MMagic`, we provide another unified interface `data_info` for users to define the mapping between the input argument and data items.

```
@weighted_loss
def disc_shift_loss(pred):
    return pred**2

@MODULES.register_module()
class DiscShiftLoss(nn.Module):

    def __init__(self, loss_weight=1.0, data_info=None):
        super(DiscShiftLoss, self).__init__()
        # codes can be found in ``mmagic/models/losses/disc_auxiliary_loss.py``

    def forward(self, *args, **kwargs):
        # codes can be found in ``mmagic/models/losses/disc_auxiliary_loss.py``
```

The goal of this design for loss modules is to allow for using it automatically in the generative models (`MODELS`), without other complex codes to define the mapping between data and keyword arguments. Thus, different from other frameworks in `OpenMMLab`, our loss modules contain a special keyword, `data_info`, which is a dictionary defining the mapping between the input arguments and data from the generative models. Taking the `DiscShiftLoss` as an example, when writing the config file, users may use this loss as follows:

```
dict(type='DiscShiftLoss',
    loss_weight=0.001 * 0.5,
    data_info=dict(pred='disc_pred_real')
```

The information in `data_info` tells the module to use the `disc_pred_real` data as the input tensor for `pred` arguments. Once the `data_info` is not `None`, our loss module will automatically build up the computational graph.

```
@MODULES.register_module()
class DiscShiftLoss(nn.Module):

    def __init__(self, loss_weight=1.0, data_info=None):
        super(DiscShiftLoss, self).__init__()
        self.loss_weight = loss_weight
        self.data_info = data_info

    def forward(self, *args, **kwargs):
        # use data_info to build computational path
        if self.data_info is not None:
            # parse the args and kwargs
            if len(args) == 1:
                assert isinstance(args[0], dict), (
                    'You should offer a dictionary containing network outputs '
                    'for building up computational graph of this loss module.')
                outputs_dict = args[0]
            elif 'outputs_dict' in kwargs:
                assert len(args) == 0, (
                    'If the outputs dict is given in keyworded arguments, no'
                    ' further non-keyworded arguments should be offered.')
                outputs_dict = kwargs.pop('outputs_dict')
            else:
                raise NotImplementedError(
                    'Cannot parsing your arguments passed to this loss module.'
                    ' Please check the usage of this module')
            # link the outputs with loss input args according to self.data_info
            loss_input_dict = {
                k: outputs_dict[v]
                for k, v in self.data_info.items()
            }
            kwargs.update(loss_input_dict)
            kwargs.update(dict(weight=self.loss_weight))
            return disc_shift_loss(**kwargs)
        else:
            # if you have not define how to build computational graph, this
            # module will just directly return the loss as usual.
            return disc_shift_loss(*args, weight=self.loss_weight, **kwargs)

    @staticmethod
    def loss_name():
        return 'loss_disc_shift'
```

As shown in this part of codes, once users set the `data_info`, the loss module will receive a dictionary containing all of the necessary data and modules, which is provided by the `MODELS` in the training procedure. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using a keyword argument,

please name it as `outputs_dict`.

An example of GANWithCustomizedLoss

To build the computational graph, the generative models have to provide a dictionary containing all kinds of data. Having a close look at any generative model, you will find that we collect all kinds of features and modules into a dictionary. We provide a customized `GANWithCustomizedLoss` here to show the process.

```
class GANWithCustomizedLoss(BaseModel):

    def __init__(self, gan_loss, disc_auxiliary_loss, gen_auxiliary_loss,
                 *args, **kwargs):
        # ...
        if gan_loss is not None:
            self.gan_loss = MODULES.build(gan_loss)
        else:
            self.gan_loss = None

        if disc_auxiliary_loss:
            self.disc_auxiliary_losses = MODULES.build(disc_auxiliary_loss)
            if not isinstance(self.disc_auxiliary_losses, nn.ModuleList):
                self.disc_auxiliary_losses = nn.ModuleList(
                    [self.disc_auxiliary_losses])
        else:
            self.disc_auxiliary_loss = None

        if gen_auxiliary_loss:
            self.gen_auxiliary_losses = MODULES.build(gen_auxiliary_loss)
            if not isinstance(self.gen_auxiliary_losses, nn.ModuleList):
                self.gen_auxiliary_losses = nn.ModuleList(
                    [self.gen_auxiliary_losses])
        else:
            self.gen_auxiliary_losses = None

    def train_step(self, data: dict,
                   optim_wrapper: OptimWrapperDict) -> Dict[str, Tensor]:
        # ...

        # get data dict to compute losses for disc
        data_dict_ = dict(
            iteration=curr_iter,
            gen=self.generator,
            disc=self.discriminator,
            disc_pred_fake=disc_pred_fake,
            disc_pred_real=disc_pred_real,
            fake_imgs=fake_imgs,
            real_imgs=real_imgs)

        loss_disc, log_vars_disc = self._get_disc_loss(data_dict_)

        # ...

    def _get_disc_loss(self, outputs_dict):
```

(continues on next page)

(continued from previous page)

```
        # Construct losses dict. If you hope some items to be included in the
        # computational graph, you have to add 'loss' in its name. Otherwise,
        # items without 'loss' in their name will just be used to print
        # information.
        losses_dict = {}
        # gan loss
        losses_dict['loss_disc_fake'] = self.gan_loss(
            outputs_dict['disc_pred_fake'], target_is_real=False, is_disc=True)
        losses_dict['loss_disc_real'] = self.gan_loss(
            outputs_dict['disc_pred_real'], target_is_real=True, is_disc=True)

        # disc auxiliary loss
        if self.with_disc_auxiliary_loss:
            for loss_module in self.disc_auxiliary_losses:
                loss_ = loss_module(outputs_dict)
                if loss_ is None:
                    continue

                # the `loss_name()` function return name as 'loss_xxx'
                if loss_module.loss_name() in losses_dict:
                    losses_dict[loss_module.loss_name(
                    )] = losses_dict[loss_module.loss_name()] + loss_
                else:
                    losses_dict[loss_module.loss_name()] = loss_
        loss, log_var = self.parse_losses(losses_dict)

        return loss, log_var
```

Here, the `_get_disc_loss` will help to combine all kinds of losses automatically.

Therefore, as long as users design the loss module with the same rules, any kind of loss can be inserted in the training of generative models, without other modifications in the code of models. What you only need to do is just defining the `data_info` in the config files.

1.21.3 Available losses

We list available losses with examples in configs as follows.

regular losses

```
# dic gan
loss_gan=dict(
    type='GANLoss',
    gan_type='vanilla',
    loss_weight=0.001,
)
```

```
# deepfillv1
loss_gan=dict(
    type='GANLoss',
    gan_type='wgan',
    loss_weight=0.0001,
)
```

```
# deepfillv2
loss_gan=dict(
    type='GANLoss',
    gan_type='hinge',
    loss_weight=0.1,
)
```

```
# aot-gan
loss_gan=dict(
    type='GANLoss',
    gan_type='smgan',
    loss_weight=0.01,
)
```

```
# deepfillv1
loss_gp=dict(type='GradientPenaltyLoss', loss_weight=10.)
```

```
# deepfillv1
loss_disc_shift=dict(type='DiscShiftLoss', loss_weight=0.001)
```

```
# dim
loss_comp=dict(type='CharbonnierCompLoss', loss_weight=0.5)
```

```
# dic gan
feature_loss=dict(
    type='LightCNNFeatureLoss',
    pretrained=pretrained_light_cnn,
    loss_weight=0.1,
    criterion='l1')
```

```
# dic gan
pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean')
```

```
# dic gan
align_loss=dict(type='MSELoss', loss_weight=0.1, reduction='mean')
```

```
# dim
loss_alpha=dict(type='CharbonnierLoss', loss_weight=0.5)
```

```
# partial conv
loss_tv=dict(
    type='MaskedTVLoss',
```

(continues on next page)

(continued from previous page)

```
    loss_weight=0.1
)
```

```
# real_basicvsr
perceptual_loss=dict(
    type='PerceptualLoss',
    layer_weights={
        '2': 0.1,
        '7': 0.1,
        '16': 1.0,
        '25': 1.0,
        '34': 1.0,
    },
    vgg_type='vgg19',
    perceptual_weight=1.0,
    style_weight=0,
    norm_img=False)
```

```
# ttsr
transferal_perceptual_loss=dict(
    type='TransferalPerceptualLoss',
    loss_weight=1e-2,
    use_attention=False,
    criterion='mse')
```

losses components

For `GANWithCustomizedLoss`, we provide several components to build customized loss.

1.22 Frequently asked questions

We list some common troubles faced by many users and their corresponding solutions here. Feel free to enrich the list if you find any frequent issues and have ways to help others to solve them. If the contents here do not cover your issue, please create an issue using the provided templates and make sure you fill in all required information in the template.

1.22.1 FAQ

Q1: "xxx: 'yyy is not in the zzz registry'".

A1: The registry mechanism will be triggered only when the file of the module is imported. So you need to import that file somewhere.

Q2: What's the folder structure of xxx dataset?

A2: You can make sure the folder structure is correct following tutorials of *dataset preparation*.

Q3: How to use LMDB data to train the model?

A3: You can use scripts in `tools/data` to make LMDB files. More details are shown in tutorials of *dataset preparation*.

Q4: Why `MMCV==xxx is used but incompatible` is raised when import I try to import `mmgen`?

A4: This is because the version of MMCV and MMGeneration are incompatible. Compatible MMGeneration and MMCV versions are shown as below. Please choose the correct version of MMCV to avoid installation issues.

Note: You need to run `pip uninstall mmcv` first if you have mmcv installed. If mmcv and mmcv-full are both installed, there will be `ModuleNotFoundError`.

Q5: How can I ignore some fields in the base configs?

A5: Sometimes, you may set `_delete_=True` to ignore some of fields in base configs. You may refer to MMEngine for simple illustration.

You may have a careful look at this tutorial for better understanding of this feature.

Q6:: How can I use intermediate variables in configs?

A6: Some intermediate variables are used in the config files, like `train_pipeline/test_pipeline` in datasets. It's worth noting that when modifying intermediate variables in the children configs, users need to pass the intermediate variables into corresponding fields again.

1.23 Overview

- Number of checkpoints: 223
- Number of configs: 251
- Number of papers: 60
  - ALGORITHM: 61
- Tasks:
  - *text2video*
  - *inpainting*
  - *video super-resolution*
  - *conditional gans*
  - *video interpolation*
  - *text2image*
  - *controlnet_animation*
  - *image2image*
  - *unconditional gans*
  - *deblurring*
  - *image super-resolution*
  - *diffusers pipeline*
  - *matting*
  - *draggan*
  - *3d-aware generation*

- *image generation*
- *colorization*
- *image restoration*
- *deraining*
- *denoising*
- *internal learning*
- *image denoising*
- *jpeg compression artifact reduction*

1.24 text2video

1.24.1 Summary

- Number of checkpoints: 7
- Number of configs: 7
- Number of papers: 1
  - ALGORITHM: 1

1.24.2 AnimateDiff (2023)

AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning

Task: Text2Video

Abstract

With the advance of text-to-image models (e.g., Stable Diffusion) and corresponding personalization techniques such as DreamBooth and LoRA, everyone can manifest their imagination into high-quality images at an affordable cost. Subsequently, there is a great demand for image animation techniques to further combine generated static images with motion dynamics. In this report, we propose a practical framework to animate most of the existing personalized text-to-image models once and for all, saving efforts in model-specific tuning. At the core of the proposed framework is to insert a newly initialized motion modeling module into the frozen text-to-image model and train it on video clips to distill reasonable motion priors. Once trained, by simply injecting this motion modeling module, all personalized versions derived from the same base T2I readily become text-driven models that produce diverse and personalized animated images. We conduct our evaluation on several public representative personalized text-to-image models across anime pictures and realistic photographs, and demonstrate that our proposed framework helps these models generate temporally smooth animation clips while preserving the domain and diversity of their outputs.

Pretrained models

We use Stable Diffusion's weights provided by HuggingFace Diffusers. You do not have to download the weights manually. If you use Diffusers wrapper, the weights will be downloaded automatically.

This model has several weights including vae, unet and clip. You should download the weights from stable-diffusion-1.5 and change the 'pretrained_model_path' in config to the weights dir.

Latest models could be looked up on OpenXLab_AnimateDiff.

Quick Start

Running the following codes, you can get a text-generated image.

Reccomendation

It's highly recommended to install xformers. It would save about 20G memory for 512*512 resolution generation.

Steps

1. Download ToonYou and MotionModule checkpoint

```
#!/bin/bash

mkdir models && cd models
mkdir Motion_Module && mkdir DreamBooth_LoRA
gdown 1RqkQuGPaC05sGZ6V6KZ-jUWmsRu48Kdq -O Motion_Module/
gdown 1ql0g_Ys4UCz2RnokYlBjyOYPbttbIpbu -O models/Motion_Module/
wget https://civitai.com/api/download/models/78775 -P DreamBooth_LoRA/ --content-
↪disposition --no-check-certificate
```

2. Modify the config file in `configs/animatediff/animatediff_ToonYou.py`

```
    models_path = {Your Checkpoints Path}
    motion_module_cfg=dict(
        path={Your MotionModule Path}
    ),
    dream_booth_lora_cfg=dict(
        type='ToonYou',
        path={Your Dreambooth_Lora Path},
        steps=25,
        guidance_scale=7.5)
```

3. Enjoy Text2Video world

```
from mmengine import Config

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

import os
import torch
```

(continues on next page)

(continued from previous page)

```
from pathlib import Path
import datetime
from mmagic.models.editors.animatediff import save_videos_grid



register_all_modules()

cfg = Config.fromfile('configs/animatediff/animatediff_ToonYou.py')
animatediff = MODELS.build(cfg.model).cuda()
prompts = [
    "best quality, masterpiece, 1girl, looking at viewer, blurry background, upper body,␣
↪contemporary, dress",

    "masterpiece, best quality, 1girl, solo, cherry blossoms, hanami, pink flower, white␣
↪flower, spring season, wisteria, petals, flower, plum blossoms, outdoors, falling␣
↪petals, white hair, black eyes,",

    "best quality, masterpiece, 1boy, formal, abstract, looking at viewer, masculine,␣
↪marble pattern",

    "best quality, masterpiece, 1girl, cloudy sky, dandelion, contrapposto, alternate␣
↪hairstyle,"
]

negative_prompts = [
    "",
    "badhandv4,easynegative,ng_deepnegative_v1_75t,verybadimagenegative_v1.3, bad-artist,
↪ bad_prompt_version2-neg, teeth",
    "",
    "",
]

sample_idx = 0
random_seeds = cfg.randomness['seed']
random_seeds = [random_seeds] if isinstance(random_seeds, int) else list(random_seeds)
samples = []
time_str = datetime.datetime.now().strftime("%Y-%m-%dT%H-%M-%S")
savedir = f"samples/{Path(cfg.model['dream_booth_lora_cfg']['type']).stem}-{time_str}"
os.makedirs(savedir)
for prompt_idx, (prompt, n_prompt, random_seed) in enumerate(zip(prompts, negative_
↪prompts, random_seeds)):
    output_dict = animatediff.infer(prompt,negative_prompt=n_prompt, video_length=16,␣
↪height=256, width=256, seed=random_seed,num_inference_steps=cfg.model['dream_booth_
↪lora_cfg']['steps'])
    sample = output_dict['samples']
    prompt = "-".join((prompt.replace("/", "").split(" ")[:10]))
    save_videos_grid(sample, f"{savedir}/sample/{sample_idx}-{prompt}.gif")
    print(f"save to {savedir}/sample/{prompt}.gif")
    samples.append(sample)
    sample_idx += 1
```

(continues on next page)

(continued from previous page)

```
samples = torch.concat(samples)
save_videos_grid(samples, f"{savedir}/sample.gif", n_rows=4)
```

Prompts for other config

- Lyriel

```
  prompt:
    - "dark shot, epic realistic, portrait of halo, sunglasses, blue eyes, tartan scarf,␣
↪white hair by atey ghailan, by greg rutkowski, by greg tocchini, by james gilleard, by␣
↪joe fenton, by kaethe butcher, gradient yellow, black, brown and magenta color scheme,␣
↪grunge aesthetic!!! graffiti tag wall background, art by greg rutkowski and artgerm,␣
↪soft cinematic light, adobe lightroom, photolab, hdr, intricate, highly detailed,␣
↪depth of field, faded, neutral colors, hdr, muted colors, hyperdetailed, artstation,␣
↪cinematic, warm lights, dramatic light, intricate details, complex background,␣
↪rutkowski, teal and orange"
    - "A forbidden castle high up in the mountains, pixel art, intricate details2, hdr,␣
↪intricate details, hyperdetailed5, natural skin texture, hyperrealism, soft light,␣
↪sharp, game art, key visual, surreal"
    - "dark theme, medieval portrait of a man sharp features, grim, cold stare, dark␣
↪colors, Volumetric lighting, baroque oil painting by Greg Rutkowski, Artgerm, WLOP,␣
↪Alphonse Mucha dynamic lighting hyperdetailed intricately detailed, hdr, muted colors,␣
↪complex background, hyperrealism, hyperdetailed, amandine van ray"
    - "As I have gone alone in there and with my treasures bold, I can keep my secret␣
↪where and hint of riches new and old. Begin it where warm waters halt and take it in a␣
↪canyon down, not far but too far to walk, put in below the home of brown."

  n_prompt:
    - "3d, cartoon, lowres, bad anatomy, bad hands, text, error, missing fingers, extra␣
↪digit, fewer digits, cropped, worst quality, low quality, normal quality, jpeg␣
↪artifacts, signature, watermark, username, blurry, artist name, young, loli, elf, 3d,␣
↪illustration"
    - "3d, cartoon, anime, sketches, worst quality, low quality, normal quality, lowres,␣
↪normal quality, monochrome, grayscale, skin spots, acnes, skin blemishes, bad anatomy,␣
↪girl, loli, young, large breasts, red eyes, muscular"
    - "dof, grayscale, black and white, bw, 3d, cartoon, anime, sketches, worst quality,␣
↪low quality, normal quality, lowres, normal quality, monochrome, grayscale, skin spots,
↪ acnes, skin blemishes, bad anatomy, girl, loli, young, large breasts, red eyes,␣
↪muscular,badhandsv5-neg, By bad artist -neg 1, monochrome"
    - "holding an item, cowboy, hat, cartoon, 3d, disfigured, bad art, deformed,extra␣
↪limbs,close up,b&w, weird colors, blurry, duplicate, morbid, mutilated, [out of frame],
↪ extra fingers, mutated hands, poorly drawn hands, poorly drawn face, mutation,␣
↪deformed, ugly, blurry, bad anatomy, bad proportions, extra limbs, cloned face,␣
↪disfigured, out of frame, ugly, extra limbs, bad anatomy, gross proportions, malformed␣
↪limbs, missing arms, missing legs, extra arms, extra legs, mutated hands, fused␣
↪fingers, too many fingers, long neck, Photoshop, video game, ugly, tiling, poorly␣
↪drawn hands, poorly drawn feet, poorly drawn face, out of frame, mutation, mutated,␣
↪extra limbs, extra legs, extra arms, disfigured, deformed, cross-eye, body out of␣
↪frame, blurry, bad art, bad anatomy, 3d render"
```

- RcnzCartoon

```
prompt:
    - "Jane Eyre with headphones, natural skin texture,4mm,k textures, soft cinematic␣
↪light, adobe lightroom, photolab, hdr, intricate, elegant, highly detailed, sharp␣
↪focus, cinematic look, soothing tones, insane details, intricate details,␣
↪hyperdetailed, low contrast, soft cinematic light, dim colors, exposure blend, hdr,␣
↪faded"
    - "close up Portrait photo of muscular bearded guy in a worn mech suit, light bokeh,␣
↪intricate, steel metal [rust], elegant, sharp focus, photo by greg rutkowski, soft␣
↪lighting, vibrant colors, masterpiece, streets, detailed face"
    - "absurdres, photorealistic, masterpiece, a 30 year old man with gold framed,␣
↪aviator reading glasses and a black hooded jacket and a beard, professional photo, a␣
↪character portrait, altermodern, detailed eyes, detailed lips, detailed face, grey eyes
↪"
    - "a golden labrador, warm vibrant colours, natural lighting, dappled lighting,␣
↪diffused lighting, absurdres, highres,k, uhd, hdr, rtx, unreal, octane render, RAW␣
↪photo, photorealistic, global illumination, subsurface scattering"

  n_prompt:
    - "deformed, distorted, disfigured, poorly drawn, bad anatomy, wrong anatomy, extra␣
↪limb, missing limb, floating limbs, mutated hands and fingers, disconnected limbs,␣
↪mutation, mutated, ugly, disgusting, blurry, amputation"
    - "nude, cross eyed, tongue, open mouth, inside, 3d, cartoon, anime, sketches, worst␣
↪quality, low quality, normal quality, lowres, normal quality, monochrome, grayscale,␣
↪skin spots, acnes, skin blemishes, bad anatomy, red eyes, muscular"
    - "easynegative, cartoon, anime, sketches, necklace, earrings worst quality, low␣
↪quality, normal quality, bad anatomy, bad hands, shiny skin, error, missing fingers,␣
↪extra digit, fewer digits, jpeg artifacts, signature, watermark, username, blurry,␣
↪chubby, anorectic, bad eyes, old, wrinkled skin, red skin, photograph By bad artist -
↪neg, big eyes, muscular face,"
    - "beard, EasyNegative, lowres, chromatic aberration, depth of field, motion blur,␣
↪blurry, bokeh, bad quality, worst quality, multiple arms, badhand"
```

- MajicMix

```
prompt:
    - "1girl, offshoulder, light smile, shiny skin best quality, masterpiece,␣
↪photorealistic"
    - "best quality, masterpiece, photorealistic, 1boy, 50 years old beard, dramatic␣
↪lighting"
    - "best quality, masterpiece, photorealistic, 1girl, light smile, shirt with collars,
↪ waist up, dramatic lighting, from below"
    - "male, man, beard, bodybuilder, skinhead,cold face, tough guy, cowboyshot, tattoo,␣
↪french windows, luxury hotel masterpiece, best quality, photorealistic"

  n_prompt:
    - "ng_deepnegative_v1_75t, badhandv4, worst quality, low quality, normal quality,␣
↪lowres, bad anatomy, bad hands, watermark, moles"
    - "nsfw, ng_deepnegative_v1_75t,badhandv4, worst quality, low quality, normal␣
↪quality, lowres,watermark, monochrome"
    - "nsfw, ng_deepnegative_v1_75t,badhandv4, worst quality, low quality, normal␣
↪quality, lowres,watermark, monochrome"
    - "nude, nsfw, ng_deepnegative_v1_75t, badhandv4, worst quality, low quality, normal␣
↪quality, lowres, bad anatomy, bad hands, monochrome, grayscale watermark, moles, people
↪"
```

(continues on next page)

(continued from previous page)

- Realistic & Realistic_v2 (same prompts with different random seed, find more details in their config files)

```
  prompt:
    - "b&w photo of 42 y.o man in black clothes, bald, face, half body, body, high␣
↪detailed skin, skin pores, coastline, overcast weather, wind, waves, 8k uhd, dslr,␣
↪soft lighting, high quality, film grain, Fujifilm XT3"
    - "close up photo of a rabbit, forest, haze, halation, bloom, dramatic atmosphere,␣
↪centred, rule of thirds, 200mm 1.4f macro shot"
    - "photo of coastline, rocks, storm weather, wind, waves, lightning, 8k uhd, dslr,␣
↪soft lighting, high quality, film grain, Fujifilm XT3"
    - "night, b&w photo of old house, post apocalypse, forest, storm weather, wind,␣
↪rocks, 8k uhd, dslr, soft lighting, high quality, film grain"

  n_prompt:
    - "semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, anime, text, close up,␣
↪cropped, out of frame, worst quality, low quality, jpeg artifacts, ugly, duplicate,␣
↪morbid, mutilated, extra fingers, mutated hands, poorly drawn hands, poorly drawn face,
↪ mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs,␣
↪cloned face, disfigured, gross proportions, malformed limbs, missing arms, missing␣
↪legs, extra arms, extra legs, fused fingers, too many fingers, long neck"
    - "semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, anime, text, close up,␣
↪cropped, out of frame, worst quality, low quality, jpeg artifacts, ugly, duplicate,␣
↪morbid, mutilated, extra fingers, mutated hands, poorly drawn hands, poorly drawn face,
↪ mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs,␣
↪cloned face, disfigured, gross proportions, malformed limbs, missing arms, missing␣
↪legs, extra arms, extra legs, fused fingers, too many fingers, long neck"
    - "blur, haze, deformed iris, deformed pupils, semi-realistic, cgi, 3d, render,␣
↪sketch, cartoon, drawing, anime, mutated hands and fingers, deformed, distorted,␣
↪disfigured, poorly drawn, bad anatomy, wrong anatomy, extra limb, missing limb,␣
↪floating limbs, disconnected limbs, mutation, mutated, ugly, disgusting, amputation"
    - "blur, haze, deformed iris, deformed pupils, semi-realistic, cgi, 3d, render,␣
↪sketch, cartoon, drawing, anime, art, mutated hands and fingers, deformed, distorted,␣
↪disfigured, poorly drawn, bad anatomy, wrong anatomy, extra limb, missing limb,␣
↪floating limbs, disconnected limbs, mutation, mutated, ugly, disgusting, amputation"
```

4. Start training motion module with the following command:

```
## 4 GPUS
bash tools/dist_train.sh configs/animatediff/animatediff.py 4
## 1 GPU
python tools/train.py configs/animatediff/animatediff.py
```

Citation

```
@article{guo2023animatediff,
  title={AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without␣
↪Specific Tuning},
  author={Guo, Yuwei and Yang, Ceyuan and Rao, Anyi and Wang, Yaohui and Qiao, Yu and␣
↪Lin, Dahua and Dai, Bo},
  journal={arXiv preprint arXiv:2307.04725},
  year={2023}
}
```

1.25 inpainting

1.25.1 Summary

- Number of checkpoints: 9
- Number of configs: 12
- Number of papers: 6
  - ALGORITHM: 6

1.25.2 Stable Diffusion (2022)

> Stable Diffusion
>
> **Task**: Text2Image, Inpainting

Abstract

Stable Diffusion is a latent diffusion model conditioned on the text embeddings of a CLIP text encoder, which allows you to create images from text inputs. This model builds upon the CVPR'22 work High-Resolution Image Synthesis with Latent Diffusion Models. The official code was released at stable-diffusion and also implemented at diffusers. We support this algorithm here to facilitate the community to learn together and compare it with other text2image methods.

Pretrained models

We use stable diffusion v1.5 weights. This model has several weights including vae, unet and clip.

You may download the weights from stable-diffusion-1.5 and change the 'from_pretrained' in config to the weights dir.

Download with git:

```
git lfs install
git clone https://huggingface.co/runwayml/stable-diffusion-v1-5
```

Quick Start

Running the following codes, you can get a text-generated image.

```
from mmengine import MODELS, Config
from torchvision import utils

from mmengine.registry import init_default_scope

init_default_scope('mmagic')

config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet.py'
config = Config.fromfile(config).copy()
## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-v1-5'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-v1-5'

StableDiffuser = MODELS.build(config.model)
prompt = 'A mecha robot in a favela in expressionist style'
StableDiffuser = StableDiffuser.to('cuda')

image = StableDiffuser.infer(prompt)['samples'][0]
image.save('robot.png')
```

To inpaint an image, you could run the following codes.

```
import mmcv
from mmengine import MODELS, Config
from mmengine.registry import init_default_scope
from PIL import Image

init_default_scope('mmagic')

config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet-inpaint.py'
config = Config.fromfile(config).copy()
## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-inpainting'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-inpainting'

StableDiffuser = MODELS.build(config.model)
prompt = 'a mecha robot sitting on a bench'

img_url = 'https://raw.githubusercontent.com/CompVis/latent-diffusion/main/data/
↪inpainting_examples/overture-creations-5sI6fQgYIuo.png'  ## noqa
mask_url = 'https://raw.githubusercontent.com/CompVis/latent-diffusion/main/data/
↪inpainting_examples/overture-creations-5sI6fQgYIuo_mask.png'  ## noqa

image = Image.fromarray(mmcv.imread(img_url, channel_order='rgb'))
mask = Image.fromarray(mmcv.imread(mask_url)).convert('L')
StableDiffuser = StableDiffuser.to('cuda')

image = StableDiffuser.infer(
    prompt,
    image,
```

(continues on next page)

(continued from previous page)

```
    mask
)['samples'][0]
image.save('inpaint.png')
```

Use ToMe to accelerate your stable diffusion model

We support **tomesd** now! It is developed based on ToMe, an efficient ViT speed-up tool based on token merging. To work on with **tomesd** in `mmagic`, you just need to add `tomesd_cfg` to `model` as shown in stable_diffusion_v1.5_tomesd. The only requirement is `torch >= 1.12.1` in order to properly support `torch.Tensor.scatter_reduce()` functionality. Please do check it before running the demo.

```
...
model = dict(
    type='StableDiffusion',
    unet=unet,
    vae=vae,
    enable_xformers=False,
    text_encoder=dict(
        type='ClipWrapper',
        clip_type='huggingface',
        pretrained_model_name_or_path=stable_diffusion_v15_url,
        subfolder='text_encoder'),
    tokenizer=stable_diffusion_v15_url,
    scheduler=diffusion_scheduler,
    test_scheduler=diffusion_scheduler,
    tomesd_cfg=dict(
        ratio=0.5))
```

The detailed settings for `tomesd_cfg` are as follows:

- `ratio (float)`: The ratio of tokens to merge. For example, 0.4 would reduce the total number of tokens by 40%.The maximum value for this is 1-(1/(`sx` * `sy`)). **By default, the max ratio is 0.75, usually <= 0.5 is recommended.** Higher values result in more speed-up, but with more visual quality loss.
- `max_downsample (int)`: Apply ToMe to layers with at most this amount of downsampling. E.g., 1 only applies to layers with no downsampling, while 8 applies to all layers. Should be chosen from 1, 2, 4, 8. **1, 2 are recommended.**
- `sx, sy (int, int)`: The stride for computing dst sets. A higher stride means you can merge more tokens, **default setting of (2, 2) works well in most cases**. `sx` and `sy` do not need to divide image size.
- `use_rand (bool)`: Whether or not to allow random perturbations when computing dst sets. By default: True, but if you're having weird artifacts you can try turning this off.
- `merge_attn (bool)`: Whether or not to merge tokens for attention **(recommended)**.
- `merge_crossattn (bool)`: Whether or not to merge tokens for cross attention **(not recommended)**.
- `merge_mlp (bool)`: Whether or not to merge tokens for the mlp layers **(especially not recommended)**.

For more details about the **tomesd** setting, please refer to Token Merging for Stable Diffusion.

Then following the code below, you can evaluate the speed-up performance on stable diffusion models or stable-diffusion-based models (DreamBooth, ControlNet).

```
import time
import numpy as np

from mmengine import MODELS, Config
from mmengine.registry import init_default_scope

init_default_scope('mmagic')

_device = 0
work_dir = '/path/to/your/work_dir'
config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet-tomesd_5e-1.py'
config = Config.fromfile(config).copy()
## ## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-v1-5'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-v1-5'

## w/o tomesd
config.model.tomesd_cfg = None
StableDiffuser = MODELS.build(config.model).to(f'cuda:{_device}')
prompt = 'A mecha robot in a favela in expressionist style'

## inference time evaluation params
size = 512
ratios = [0.5, 0.75]
samples_perprompt = 5

t = time.time()
for i in range(100//samples_perprompt):
    image = StableDiffuser.infer(prompt, height=size, width=size, num_images_per_
↪prompt=samples_perprompt)['samples'][0]
    if i == 0:
        image.save(f"{work_dir}/wo_tomesd.png")
print(f"Generating 100 images with {samples_perprompt} images per prompt, without ToMe␣
↪speed-up, time used : {time.time() - t}s")

for ratio in ratios:
    ## w/ tomesd
    config.model.tomesd_cfg = dict(ratio=ratio)
    sd_model = MODELS.build(config.model).to(f'cuda:{_device}')

    t = time.time()
    for i in range(100//samples_perprompt):
        image = sd_model.infer(prompt, height=size, width=size, num_images_per_
↪prompt=samples_perprompt)['samples'][0]
        if i == 0:
            image.save(f"{work_dir}/w_tomesd_ratio_{ratio}.png")

    print(f"Generating 100 images with {samples_perprompt} images per prompt, merging␣
↪ratio {ratio}, time used : {time.time() - t}s")
```

Here are some inference performance comparisons running on **single RTX 3090** with `torch 2.0.0+cu118` as backends. The results are reasonable, when enabling `xformers`, the speed-up ratio is a little bit lower. But `tomesd` still effectively reduces the inference time. It is especially recommended that enable `tomesd` when the `image_size` and

num_images_per_prompt are large, since the number of similar tokens are larger and tomesd can achieve better performance.

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5.

Thanks for the efforts of the community!

Citation

```
@misc{rombach2021highresolution,
      title={High-Resolution Image Synthesis with Latent Diffusion Models},
      author={Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser␣
↪and Björn Ommer},
      year={2021},
      eprint={2112.10752},
      archivePrefix={arXiv},
      primaryClass={cs.CV}
}

@article{bolya2023tomesd,
  title={Token Merging for Fast Stable Diffusion},
  author={Bolya, Daniel and Hoffman, Judy},
  journal={arXiv},
  year={2023}
}

@inproceedings{bolya2023tome,
  title={Token Merging: Your {ViT} but Faster},
  author={Bolya, Daniel and Fu, Cheng-Yang and Dai, Xiaoliang and Zhang, Peizhao and␣
↪Feichtenhofer, Christoph and Hoffman, Judy},
  booktitle={International Conference on Learning Representations},
  year={2023}
}
```

1.25.3 AOT-GAN (TVCG'2021)

AOT-GAN: Aggregated Contextual Transformations for High-Resolution Image Inpainting

Task: Inpainting

Abstract

State-of-the-art image inpainting approaches can suffer from generating distorted structures and blurry textures in high-resolution images (e.g., 512x512). The challenges mainly drive from (1) image content reasoning from distant contexts, and (2) fine-grained texture synthesis for a large missing region. To overcome these two challenges, we propose an enhanced GAN-based model, named Aggregated COntextual-Transformation GAN (AOT-GAN), for high-resolution image inpainting. Specifically, to enhance context reasoning, we construct the generator of AOT-GAN by stacking multiple layers of a proposed AOT block. The AOT blocks aggregate contextual transformations from various receptive fields, allowing to capture both informative distant image contexts and rich patterns of interest for context reasoning. For improving texture synthesis, we enhance the discriminator of AOT-GAN by training it with a tailored mask-prediction task. Such a training objective forces the discriminator to distinguish the detailed appearances of real and synthesized patches, and in turn, facilitates the generator to synthesize clear textures. Extensive comparisons on Places2, the most challenging benchmark with 1.8 million high-resolution images of 365 complex scenes, show that our model outperforms the state-of-the-art by a significant margin in terms of FID with 38.60% relative improvement. A user study including more than 30 subjects further validates the superiority of AOT-GAN. We further evaluate the proposed AOT-GAN in practical applications, e.g., logo removal, face editing, and object removal. Results show that our model achieves promising completions in the real world. We release code and models in this https URL.

Results and models

More results for different mask area:

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/aot_gan/aot-gan_smpgan_4xb4_places-
↪512x512.py

## single-gpu train
python tools/train.py configs/aot_gan/aot-gan_smpgan_4xb4_places-512x512.py

## multi-gpu train
./tools/dist_train.sh configs/aot_gan/aot-gan_smpgan_4xb4_places-512x512.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/aot_gan/aot-gan_smpgan_4xb4_places-
↪512x512.py https://download.openmmlab.com/mmediting/inpainting/aot_gan/AOT-GAN_512x512_
↪4x12_places_20220509-6641441b.pth

## single-gpu test
python tools/test.py configs/aot_gan/aot-gan_smpgan_4xb4_places-512x512.py https://
↪download.openmmlab.com/mmediting/inpainting/aot_gan/AOT-GAN_512x512_4x12_places_
↪20220509-6641441b.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/aot_gan/aot-gan_smpgan_4xb4_places-512x512.py https://
↪download.openmmlab.com/mmediting/inpainting/aot_gan/AOT-GAN_512x512_4x12_places_
↪20220509-6641441b.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{yan2021agg,
  author = {Zeng, Yanhong and Fu, Jianlong and Chao, Hongyang and Guo, Baining},
  title = {Aggregated Contextual Transformations for High-Resolution Image Inpainting},
  booktitle = {Arxiv},
  pages={-},
  year = {2020}
}
```

1.25.4 DeepFillv2 (CVPR'2019)

Free-Form Image Inpainting with Gated Convolution

Task: Inpainting

Abstract

We present a generative image inpainting system to complete images with free-form mask and guidance. The system is based on gated convolutions learned from millions of images without additional labelling efforts. The proposed gated convolution solves the issue of vanilla convolution that treats all input pixels as valid ones, generalizes partial convolution by providing a learnable dynamic feature selection mechanism for each channel at each spatial location across all layers. Moreover, as free-form masks may appear anywhere in images with any shape, global and local GANs designed for a single rectangular mask are not applicable. Thus, we also present a patch-based GAN loss, named SN-PatchGAN, by applying spectral-normalized discriminator on dense image patches. SN-PatchGAN is simple in formulation, fast and stable in training. Results on automatic image inpainting and user-guided extension demonstrate that our system generates higher-quality and more flexible results than previous methods. Our system helps user quickly remove distracting objects, modify image layouts, clear watermarks and edit faces.

Results and models

CelebA-HQ

Places365-Challenge

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/deepfillv2/deepfillv2_8xb2_places-
↪256x256.py

## single-gpu train
python tools/train.py configs/deepfillv2/deepfillv2_8xb2_places-256x256.py

## multi-gpu train
./tools/dist_train.sh configs/deepfillv2/deepfillv2_8xb2_places-256x256.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/deepfillv2/deepfillv2_8xb2_places-
↪256x256.py https://download.openmmlab.com/mmediting/inpainting/deepfillv2/deepfillv2_
↪256x256_8x2_places_20200619-10d15793.pth

## single-gpu test
python tools/test.py configs/deepfillv2/deepfillv2_8xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/deepfillv2/deepfillv2_256x256_8x2_places_
↪20200619-10d15793.pth

## multi-gpu test
./tools/dist_test.sh configs/deepfillv2/deepfillv2_8xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/deepfillv2/deepfillv2_256x256_8x2_places_
↪20200619-10d15793.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{yu2019free,
  title={Free-form image inpainting with gated convolution},
  author={Yu, Jiahui and Lin, Zhe and Yang, Jimei and Shen, Xiaohui and Lu, Xin and␣
↪Huang, Thomas S},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={4471--4480},
  year={2019}
}
```

1.25.5 DeepFillv1 (CVPR'2018)

Generative Image Inpainting with Contextual Attention

Task: Inpainting

Abstract

Recent deep learning based approaches have shown promising results for the challenging task of inpainting large missing regions in an image. These methods can generate visually plausible image structures and textures, but often create distorted structures or blurry textures inconsistent with surrounding areas. This is mainly due to ineffectiveness of convolutional neural networks in explicitly borrowing or copying information from distant spatial locations. On the other hand, traditional texture and patch synthesis approaches are particularly suitable when it needs to borrow textures from the surrounding regions. Motivated by these observations, we propose a new deep generative model-based approach which can not only synthesize novel image structures but also explicitly utilize surrounding image features as references during network training to make better predictions. The model is a feed-forward, fully convolutional neural network which can process images with multiple holes at arbitrary locations and with variable sizes during the test time. Experiments on multiple datasets including faces (CelebA, CelebA-HQ), textures (DTD) and natural images (ImageNet, Places2) demonstrate that our proposed approach generates higher-quality inpainting results than existing ones.

Results and models

CelebA-HQ

Places365-Challenge

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/deepfillv1/deepfillv1_8xb2_places-
↪256x256.py

## single-gpu train
python tools/train.py configs/deepfillv1/deepfillv1_8xb2_places-256x256.py

## multi-gpu train
./tools/dist_train.sh configs/deepfillv1/deepfillv1_8xb2_places-256x256.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/deepfillv1/deepfillv1_8xb2_places-
↪256x256.py https://download.openmmlab.com/mmediting/inpainting/deepfillv1/deepfillv1_
↪256x256_8x2_places_20200619-c00a0e21.pth
```

(continues on next page)

(continued from previous page)

```
## single-gpu test
python tools/test.py configs/deepfillv1/deepfillv1_8xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/deepfillv1/deepfillv1_256x256_8x2_places_
↪20200619-c00a0e21.pth

## multi-gpu test
./tools/dist_test.sh configs/deepfillv1/deepfillv1_8xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/deepfillv1/deepfillv1_256x256_8x2_places_
↪20200619-c00a0e21.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{yu2018generative,
  title={Generative image inpainting with contextual attention},
  author={Yu, Jiahui and Lin, Zhe and Yang, Jimei and Shen, Xiaohui and Lu, Xin and␣
↪Huang, Thomas S},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={5505--5514},
  year={2018}
}
```

1.25.6 PConv (ECCV'2018)

Image Inpainting for Irregular Holes Using Partial Convolutions

Task: Inpainting

Abstract

Existing deep learning based image inpainting methods use a standard convolutional network over the corrupted image, using convolutional filter responses conditioned on both valid pixels as well as the substitute values in the masked holes (typically the mean value). This often leads to artifacts such as color discrepancy and blurriness. Post-processing is usually used to reduce such artifacts, but are expensive and may fail. We propose the use of partial convolutions, where the convolution is masked and renormalized to be conditioned on only valid pixels. We further include a mechanism to automatically generate an updated mask for the next layer as part of the forward pass. Our model outperforms other methods for irregular masks. We show qualitative and quantitative comparisons with other methods to validate our approach.

Results and models

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/partial_conv/pconv_stage2_4xb2_
↪places-256x256.py

## single-gpu train
python tools/train.py configs/partial_conv/pconv_stage2_4xb2_places-256x256.py

## multi-gpu train
./tools/dist_train.sh configs/partial_conv/pconv_stage2_4xb2_places-256x256.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/partial_conv/pconv_stage2_4xb2_
↪places-256x256.py https://download.openmmlab.com/mmediting/inpainting/pconv/pconv_
↪256x256_stage2_4x2_places_20200619-1ffed0e8.pth

## single-gpu test
python tools/test.py configs/partial_conv/pconv_stage2_4xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/pconv/pconv_256x256_stage2_4x2_places_
↪20200619-1ffed0e8.pth

## multi-gpu test
./tools/dist_test.sh configs/partial_conv/pconv_stage2_4xb2_places-256x256.py https://
↪download.openmmlab.com/mmediting/inpainting/pconv/pconv_256x256_stage2_4x2_places_
↪20200619-1ffed0e8.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{liu2018image,
  title={Image inpainting for irregular holes using partial convolutions},
  author={Liu, Guilin and Reda, Fitsum A and Shih, Kevin J and Wang, Ting-Chun and Tao,␣
↪Andrew and Catanzaro, Bryan},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  pages={85--100},
  year={2018}
}
```

1.25.7 Global&Local (ToG'2017)

Globally and Locally Consistent Image Completion

Task: Inpainting

Abstract

We present a novel approach for image completion that results in images that are both locally and globally consistent. With a fully-convolutional neural network, we can complete images of arbitrary resolutions by filling in missing regions of any shape. To train this image completion network to be consistent, we use global and local context discriminators that are trained to distinguish real images from completed ones. The global discriminator looks at the entire image to assess if it is coherent as a whole, while the local discriminator looks only at a small area centered at the completed region to ensure the local consistency of the generated patches. The image completion network is then trained to fool the both context discriminator networks, which requires it to generate images that are indistinguishable from real ones with regard to overall consistency as well as in details. We show that our approach can be used to complete a wide variety of scenes. Furthermore, in contrast with the patch-based approaches such as PatchMatch, our approach can generate fragments that do not appear elsewhere in the image, which allows us to naturally complete the image.

Results and models

Note that we do not apply the post-processing module in Global&Local for a fair comparison with current deep inpainting methods.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/global_local/gl_8xb12_places-
↪256x256.py

## single-gpu train
python tools/train.py configs/global_local/gl_8xb12_places-256x256.py

## multi-gpu train
./tools/dist_train.sh configs/global_local/gl_8xb12_places-256x256.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/global_local/gl_8xb12_places-
↪256x256.py https://download.openmmlab.com/mmediting/inpainting/global_local/gl_256x256_
↪8x12_places_20200619-52a040a8.pth

## single-gpu test
python tools/test.py configs/global_local/gl_8xb12_places-256x256.py https://download.
↪openmmlab.com/mmediting/inpainting/global_local/gl_256x256_8x12_places_20200619-
↪52a040a8.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/global_local/gl_8xb12_places-256x256.py https://download.
↪openmmlab.com/mmediting/inpainting/global_local/gl_256x256_8x12_places_20200619-
↪52a040a8.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{iizuka2017globally,
  title={Globally and locally consistent image completion},
  author={Iizuka, Satoshi and Simo-Serra, Edgar and Ishikawa, Hiroshi},
  journal={ACM Transactions on Graphics (ToG)},
  volume={36},
  number={4},
  pages={1--14},
  year={2017},
  publisher={ACM New York, NY, USA}
}
```

1.26 video super-resolution

1.26.1 Summary

- Number of checkpoints: 27
- Number of configs: 29
- Number of papers: 6
  - ALGORITHM: 7

1.26.2 BasicVSR++ (CVPR'2022)

BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment

Task: Video Super-Resolution

Abstract

A recurrent structure is a popular framework choice for the task of video super-resolution. The state-of-the-art method BasicVSR adopts bidirectional propagation with feature alignment to effectively exploit information from the entire input video. In this study, we redesign BasicVSR by proposing second-order grid propagation and flow-guided deformable alignment. We show that by empowering the recurrent framework with the enhanced propagation and alignment, one can exploit spatiotemporal information across misaligned video frames more effectively. The new components lead to an improved performance under a similar computational constraint. In particular, our model BasicVSR++ surpasses BasicVSR by 0.82 dB in PSNR with similar number of parameters. In addition to video super-resolution, BasicVSR++ generalizes well to other video restoration tasks such as compressed video enhancement. In NTIRE

2021, BasicVSR++ obtains three champions and one runner-up in the Video Super-Resolution and Compressed Video Enhancement Challenges. Codes and models will be released to MMagic.

Results and models

The pretrained weights of SPyNet can be found here.

NTIRE 2021 checkpoints

Note that the following models are finetuned from smaller models. The training schemes of these models will be released when MMagic reaches 5k stars. We provide the pre-trained models here.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-
↪600k_reds4.py

## single-gpu train
python tools/train.py configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-600k_reds4.py

## multi-gpu train
./tools/dist_train.sh configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-600k_reds4.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-
↪600k_reds4.py https://download.openmmlab.com/mmediting/restorers/basicvsr_plusplus/
↪basicvsr_plusplus_c64n7_8x1_600k_reds4_20210217-db622b2f.pth

## single-gpu test
python tools/test.py configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-600k_reds4.py https://
↪download.openmmlab.com/mmediting/restorers/basicvsr_plusplus/basicvsr_plusplus_c64n7_
↪8x1_600k_reds4_20210217-db622b2f.pth

## multi-gpu test
./tools/dist_test.sh configs/basicvsr_pp/basicvsr-pp_c64n7_8xb1-600k_reds4.py https://
↪download.openmmlab.com/mmediting/restorers/basicvsr_plusplus/basicvsr_plusplus_c64n7_
↪8x1_600k_reds4_20210217-db622b2f.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{chan2022basicvsrplusplus,
  author = {Chan, Kelvin C.K. and Zhou, Shangchen and Xu, Xiangyu and Loy, Chen Change},
  title = {BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern recognition},
  year = {2022}
}
```

1.26.3 RealBasicVSR (CVPR'2022)

RealBasicVSR: Investigating Tradeoffs in Real-World Video Super-Resolution

Task: Video Super-Resolution

Abstract

The diversity and complexity of degradations in real-world video super-resolution (VSR) pose non-trivial challenges in inference and training. First, while long-term propagation leads to improved performance in cases of mild degradations, severe in-the-wild degradations could be exaggerated through propagation, impairing output quality. To balance the tradeoff between detail synthesis and artifact suppression, we found an image pre-cleaning stage indispensable to reduce noises and artifacts prior to propagation. Equipped with a carefully designed cleaning module, our RealBasicVSR outperforms existing methods in both quality and efficiency. Second, real-world VSR models are often trained with diverse degradations to improve generalizability, requiring increased batch size to produce a stable gradient. Inevitably, the increased computational burden results in various problems, including 1) speed-performance tradeoff and 2) batch-length tradeoff. To alleviate the first tradeoff, we propose a stochastic degradation scheme that reduces up to 40% of training time without sacrificing performance. We then analyze different training settings and suggest that employing longer sequences rather than larger batches during training allows more effective uses of temporal information, leading to more stable performance during inference. To facilitate fair comparisons, we propose the new VideoLQ dataset, which contains a large variety of real-world low-quality video sequences containing rich textures and patterns. Our dataset can serve as a common ground for benchmarking. Code, models, and the dataset will be made publicly available.

Results and models

Evaluated on Y channel. The code for computing NRQM, NIQE, and PI can be found here. MATLAB official code is used to compute BRISQUE.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/real_basicvsr/realbasicvsr_c64b20-1x30x8_8xb1-lr5e-5-150k_reds.py

## single-gpu train
python tools/train.py configs/real_basicvsr/realbasicvsr_c64b20-1x30x8_8xb1-lr5e-5-150k_reds.py
```

(continues on next page)

(continued from previous page)

```
## multi-gpu train
./tools/dist_train.sh configs/real_basicvsr/realbasicvsr_c64b20-1x30x8_8xb1-lr5e-5-150k_
↪reds.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/real_basicvsr/realbasicvsr_c64b20-
↪1x30x8_8xb1-lr5e-5-150k_reds.py https://download.openmmlab.com/mmediting/restorers/
↪real_basicvsr/realbasicvsr_c64b20_1x30x8_lr5e-5_150k_reds_20211104-52f77c2c.pth

## single-gpu test
python tools/test.py python tools/test.py configs/real_basicvsr/realbasicvsr_c64b20-
↪1x30x8_8xb1-lr5e-5-150k_reds.py https://download.openmmlab.com/mmediting/restorers/
↪real_basicvsr/realbasicvsr_c64b20_1x30x8_lr5e-5_150k_reds_20211104-52f77c2c.pth

## multi-gpu test
./tools/dist_test.sh configs/real_basicvsr/realbasicvsr_c64b20-1x30x8_8xb1-lr5e-5-150k_
↪reds.py https://download.openmmlab.com/mmediting/restorers/real_basicvsr/realbasicvsr_
↪c64b20_1x30x8_lr5e-5_150k_reds_20211104-52f77c2c.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{chan2022investigating,
  author = {Chan, Kelvin C.K. and Zhou, Shangchen and Xu, Xiangyu and Loy, Chen Change},
  title = {RealBasicVSR: Investigating Tradeoffs in Real-World Video Super-Resolution},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  year = {2022}
}
```

1.26.4 BasicVSR (CVPR'2021)

BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond

Task: Video Super-Resolution

Abstract

Video super-resolution (VSR) approaches tend to have more components than the image counterparts as they need to exploit the additional temporal dimension. Complex designs are not uncommon. In this study, we wish to untangle the knots and reconsider some most essential components for VSR guided by four basic functionalities, i.e., Propagation, Alignment, Aggregation, and Upsampling. By reusing some existing components added with minimal redesigns, we show a succinct pipeline, BasicVSR, that achieves appealing improvements in terms of speed and restoration quality in comparison to many state-of-the-art algorithms. We conduct systematic analysis to explain how such gain can be obtained and discuss the pitfalls. We further show the extensibility of BasicVSR by presenting an information-refill mechanism and a coupled propagation scheme to facilitate information aggregation. The BasicVSR and its extension, IconVSR, can serve as strong baselines for future VSR approaches.

Results and models

Evaluated on RGB channels for REDS4 and Y channel for others. The metrics are `PSNR` / `SSIM` . The pretrained weights of SPyNet can be found here.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/basicvsr/basicvsr_2xb4_reds4.py

## single-gpu train
python tools/train.py configs/basicvsr/basicvsr_2xb4_reds4.py

## multi-gpu train
./tools/dist_train.sh configs/basicvsr/basicvsr_2xb4_reds4.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/basicvsr/basicvsr_2xb4_reds4.py␣
↪https://download.openmmlab.com/mmediting/restorers/basicvsr/basicvsr_reds4_20120409-
↪0e599677.pth

## single-gpu test
python tools/test.py configs/basicvsr/basicvsr_2xb4_reds4.py https://download.openmmlab.
↪com/mmediting/restorers/basicvsr/basicvsr_reds4_20120409-0e599677.pth

## multi-gpu test
./tools/dist_test.sh configs/basicvsr/basicvsr_2xb4_reds4.py https://download.openmmlab.
↪com/mmediting/restorers/basicvsr/basicvsr_reds4_20120409-0e599677.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{chan2021basicvsr,
  author = {Chan, Kelvin CK and Wang, Xintao and Yu, Ke and Dong, Chao and Loy, Chen␣
↪Change},
  title = {BasicVSR: The Search for Essential Components in Video Super-Resolution and␣
↪Beyond},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  year = {2021}
}
```

1.26.5 IconVSR (CVPR'2021)

BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond

Task: Video Super-Resolution

Abstract

Video super-resolution (VSR) approaches tend to have more components than the image counterparts as they need to exploit the additional temporal dimension. Complex designs are not uncommon. In this study, we wish to untangle the knots and reconsider some most essential components for VSR guided by four basic functionalities, i.e., Propagation, Alignment, Aggregation, and Upsampling. By reusing some existing components added with minimal redesigns, we show a succinct pipeline, BasicVSR, that achieves appealing improvements in terms of speed and restoration quality in comparison to many state-of-the-art algorithms. We conduct systematic analysis to explain how such gain can be obtained and discuss the pitfalls. We further show the extensibility of BasicVSR by presenting an information-refill mechanism and a coupled propagation scheme to facilitate information aggregation. The BasicVSR and its extension, IconVSR, can serve as strong baselines for future VSR approaches.

Results and models

Evaluated on RGB channels for REDS4 and Y channel for others. The metrics are `PSNR` / `SSIM` . The pretrained weights of the IconVSR components can be found here: SPyNet, EDVR-M for REDS, and EDVR-M for Vimeo-90K.

| Model | Dataset | PSNR (RGB) | SSIM (RGB) | PSNR (Y) | SSIM (Y) | Training Resources | Download | | :————————: | :————————: | :————————: | :————————: | :————————: | :————————: | :————————: | :————————: | | iconvsr_reds4 | REDS4 (BIx4) | **31.6926** | **0.8951** |-|-| 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_reds4 | UDM10 (BDx4) | -|-|35.3377| 0.9471| 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_reds4 | Vid4 (BIx4) | -|-|**27.4809**| **0.8354** | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_reds4 | Vid4 (BDx4) | -|-| 25.2110 | 0.7732 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_reds4 | Vimeo-90K-T (BIx4) | -|-|36.4983 | 0.9416 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_reds4 | Vimeo-90K-T (BDx4) | -|-| 34.4299 | 0.9287 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | REDS4 (BIx4) | 30.3452 |0.8659 |-|-| 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | UDM10 (BDx4) |-|-|34.2595 |0.9398 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | Vid4 (BIx4) |-|-|27.4238 |0.8297 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | Vid4 (BDx4) |-|-|24.6666|0.7491 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | Vimeo-90K-T (BIx4) |-|-|**37.3729**| **0.9467** | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bi | Vimeo-90K-T (BDx4) |-|-|34.5548 | 0.9295 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | REDS4 (BIx4) |29.0150 | 0.8465 |-|-| 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | UDM10 (BDx4) |-|-| **40.0640** | **0.9697** | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | Vid4 (BIx4) |-|-|26.3109| 0.8028 | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | Vid4 (BDx4) | -|-| **28.2464** | **0.8612** | 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | Vimeo-90K-T

(BIx4) | -|-|34.6780| 0.9339| 2 (Tesla V100-PCIE-32GB) | model | log | | iconvsr_vimeo90k_bd | Vimeo-90K-T (BDx4) |-|-|**37.7573** | **0.9517** | 2 (Tesla V100-PCIE-32GB) | model | log |

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/iconvsr/iconvsr_2xb4_reds4.py

## single-gpu train
python tools/train.py configs/iconvsr/iconvsr_2xb4_reds4.py

## multi-gpu train
./tools/dist_train.sh configs/iconvsr/iconvsr_2xb4_reds4.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/iconvsr/iconvsr_2xb4_reds4.py␣
↪https://download.openmmlab.com/mmediting/restorers/iconvsr/iconvsr_reds4_20210413-
↪9e09d621.pth

## single-gpu test
python tools/test.py configs/iconvsr/iconvsr_2xb4_reds4.py https://download.openmmlab.
↪com/mmediting/restorers/iconvsr/iconvsr_reds4_20210413-9e09d621.pth

## multi-gpu test
./tools/dist_test.sh configs/iconvsr/iconvsr_2xb4_reds4.py https://download.openmmlab.
↪com/mmediting/restorers/iconvsr/iconvsr_reds4_20210413-9e09d621.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{chan2021basicvsr,
  author = {Chan, Kelvin CK and Wang, Xintao and Yu, Ke and Dong, Chao and Loy, Chen␣
↪Change},
  title = {BasicVSR: The Search for Essential Components in Video Super-Resolution and␣
↪Beyond},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  year = {2021}
}
```

1.26.6 TDAN (CVPR'2020)

TDAN: Temporally Deformable Alignment Network for Video Super-Resolution

Task: Video Super-Resolution

Abstract

Video super-resolution (VSR) aims to restore a photo-realistic high-resolution (HR) video frame from both its corresponding low-resolution (LR) frame (reference frame) and multiple neighboring frames (supporting frames). Due to varying motion of cameras or objects, the reference frame and each support frame are not aligned. Therefore, temporal alignment is a challenging yet important problem for VSR. Previous VSR methods usually utilize optical flow between the reference frame and each supporting frame to wrap the supporting frame for temporal alignment. Therefore, the performance of these image-level wrapping-based models will highly depend on the prediction accuracy of optical flow, and inaccurate optical flow will lead to artifacts in the wrapped supporting frames, which also will be propagated into the reconstructed HR video frame. To overcome the limitation, in this paper, we propose a temporal deformable alignment network (TDAN) to adaptively align the reference frame and each supporting frame at the feature level without computing optical flow. The TDAN uses features from both the reference frame and each supporting frame to dynamically predict offsets of sampling convolution kernels. By using the corresponding kernels, TDAN transforms supporting frames to align with the reference frame. To predict the HR video frame, a reconstruction network taking aligned frames and the reference frame is utilized. Experimental results demonstrate the effectiveness of the proposed TDAN-based VSR model.

Results and models

Evaluated on Y-channel. 8 pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

TDAN is trained with two stages.

Stage 1: Train with a larger learning rate (1e-4)

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/tdan/tdan_x4_1xb16-lr1e-4-400k_
↪vimeo90k-bi.py

## single-gpu train
python tools/train.py configs/tdan/tdan_x4_1xb16-lr1e-4-400k_vimeo90k-bi.py

## multi-gpu train
./tools/dist_train.sh cconfigs/tdan/tdan_x4_1xb16-lr1e-4-400k_vimeo90k-bi.py 8
```

Stage 2: Fine-tune with a smaller learning rate (5e-5)

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_
↪vimeo90k-bi.py
```

(continues on next page)

(continued from previous page)

```
## single-gpu train
python tools/train.py configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_vimeo90k-bi.py

## multi-gpu train
./tools/dist_train.sh configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_vimeo90k-bi.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_
↪vimeo90k-bi.py https://download.openmmlab.com/mmediting/restorers/tdan/tdan_vimeo90k_
↪bix4_20210528-739979d9.pth

## single-gpu test
python tools/test.py configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_vimeo90k-bi.py https://
↪download.openmmlab.com/mmediting/restorers/tdan/tdan_vimeo90k_bix4_20210528-739979d9.
↪pth

## multi-gpu test
./tools/dist_test.sh configs/tdan/tdan_x4ft_1xb16-lr5e-5-400k_vimeo90k-bi.py https://
↪download.openmmlab.com/mmediting/restorers/tdan/tdan_vimeo90k_bix4_20210528-739979d9.
↪pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{tian2020tdan,
  title={TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution},
  author={Tian, Yapeng and Zhang, Yulun and Fu, Yun and Xu, Chenliang},
  booktitle = {Proceedings of the IEEE conference on Computer Vision and Pattern␣
↪Recognition},
  year = {2020}
}
```

1.26.7 TOFlow (IJCV'2019)

Video Enhancement with Task-Oriented Flow

Task: Video Interpolation, Video Super-Resolution

Abstract

Many video enhancement algorithms rely on optical flow to register frames in a video sequence. Precise flow estimation is however intractable; and optical flow itself is often a sub-optimal representation for particular video processing tasks. In this paper, we propose task-oriented flow (TOFlow), a motion representation learned in a self-supervised, task-specific manner. We design a neural network with a trainable motion estimation component and a video processing component, and train them jointly to learn the task-oriented flow. For evaluation, we build Vimeo-90K, a large-scale, high-quality video dataset for low-level video processing. TOFlow outperforms traditional optical flow on standard benchmarks as well as our Vimeo-90K dataset in three video processing tasks: frame interpolation, video denoising/deblocking, and video super-resolution.

Results and models

Evaluated on Vimeo90k-triplet (RGB channels). The metrics are `PSNR / SSIM` .

Note: These pretrained SPyNets don't contain BN layer since `batch_size=1`, which is consistent with `https://github.com/Coldog2333/pytoflow`.

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

TOF only supports video interpolation task for training now.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/tof/tof_spynet-chair-wobn_1xb1_
↪vimeo90k-triplet.py

## single-gpu train
python tools/train.py configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py

## multi-gpu train
./tools/dist_train.sh configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

TOF supports two tasks for testing.

Task 1: Video Interpolation

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/tof/tof_spynet-chair-wobn_1xb1_
↪vimeo90k-triplet.py https://download.openmmlab.com/mmediting/video_interpolators/
↪toflow/pretrained_spynet_chair_20220321-4d82e91b.pth

## single-gpu test
python tools/test.py configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/toflow/pretrained_spynet_chair_
↪20220321-4d82e91b.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/toflow/pretrained_spynet_chair_
↪20220321-4d82e91b.pth 8
```

Task 2: Video Super-Resolution

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/tof/tof_x4_official_vimeo90k.py␣
↪https://download.openmmlab.com/mmediting/restorers/tof/tof_x4_vimeo90k_official-
↪a569ff50.pth

## single-gpu test
python tools/test.py configs/tof/tof_x4_official_vimeo90k.py https://download.openmmlab.
↪com/mmediting/restorers/tof/tof_x4_vimeo90k_official-a569ff50.pth

## multi-gpu test
./tools/dist_test.sh configs/tof/tof_x4_official_vimeo90k.py https://download.openmmlab.
↪com/mmediting/restorers/tof/tof_x4_vimeo90k_official-a569ff50.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{xue2019video,
  title={Video enhancement with task-oriented flow},
  author={Xue, Tianfan and Chen, Baian and Wu, Jiajun and Wei, Donglai and Freeman,␣
↪William T},
  journal={International Journal of Computer Vision},
  volume={127},
  number={8},
  pages={1106--1125},
  year={2019},
  publisher={Springer}
}
```

1.26.8 EDVR (CVPRW'2019)

EDVR: Video Restoration with Enhanced Deformable Convolutional Networks

Task: Video Super-Resolution

Abstract

Video restoration tasks, including super-resolution, deblurring, etc, are drawing increasing attention in the computer vision community. A challenging benchmark named REDS is released in the NTIRE19 Challenge. This new benchmark challenges existing methods from two aspects: (1) how to align multiple frames given large motions, and (2) how to effectively fuse different frames with diverse motion and blur. In this work, we propose a novel Video Restoration framework with Enhanced Deformable networks, termed EDVR, to address these challenges. First, to handle large motions, we devise a Pyramid, Cascading and Deformable (PCD) alignment module, in which frame alignment is done at the feature level using deformable convolutions in a coarse-to-fine manner. Second, we propose a Temporal and Spatial Attention (TSA) fusion module, in which attention is applied both temporally and spatially, so as to emphasize important features for subsequent restoration. Thanks to these modules, our EDVR wins the champions and outperforms the second place by a large margin in all four tracks in the NTIRE19 video restoration and enhancement challenges. EDVR also demonstrates superior performance to state-of-the-art published methods on video super-resolution and deblurring.

Results and models

Evaluated on RGB channels. The metrics are `PSNR` and `SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/edvr/edvrm_8xb4-600k_reds.py

## single-gpu train
python tools/train.py configs/edvr/edvrm_8xb4-600k_reds.py

## multi-gpu train
./tools/dist_train.sh configs/edvr/edvrm_8xb4-600k_reds.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py python tools/test.py configs/edvr/edvrm_
↪8xb4-600k_reds.py https://download.openmmlab.com/mmediting/restorers/edvr/edvrm_x4_8x4_
↪600k_reds_20210625-e29b71b5.pth

## single-gpu test
python tools/test.py configs/edvr/edvrm_8xb4-600k_reds.py https://download.openmmlab.com/
↪mmediting/restorers/edvr/edvrm_x4_8x4_600k_reds_20210625-e29b71b5.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/edvr/edvrm_8xb4-600k_reds.py https://download.openmmlab.com/
↪mmediting/restorers/edvr/edvrm_x4_8x4_600k_reds_20210625-e29b71b5.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{wang2019edvr,
  author    = {Wang, Xintao and Chan, Kelvin C.K. and Yu, Ke and Dong, Chao and Loy,␣
↪Chen Change},
  title     = {EDVR: Video restoration with enhanced deformable convolutional networks},
  booktitle = {The IEEE Conference on Computer Vision and Pattern Recognition Workshops␣
↪(CVPRW)},
  month     = {June},
  year      = {2019},
}
```

1.27 conditional gans

1.27.1 Summary

- Number of checkpoints: 18
- Number of configs: 18
- Number of papers: 3
  - ALGORITHM: 3

1.27.2 BigGAN (ICLR'2019)

Large Scale GAN Training for High Fidelity Natural Image Synthesis

Task: Conditional GANs

Abstract

Despite recent progress in generative image modeling, successfully generating high-resolution, diverse samples from complex datasets such as ImageNet remains an elusive goal. To this end, we train Generative Adversarial Networks at the largest scale yet attempted, and study the instabilities specific to such scale. We find that applying orthogonal regularization to the generator renders it amenable to a simple "truncation trick," allowing fine control over the trade-off between sample fidelity and variety by reducing the variance of the Generator's input. Our modifications lead to models which set the new state of the art in class-conditional image synthesis. When trained on ImageNet at 128x128 resolution, our models (BigGANs) achieve an Inception Score (IS) of 166.5 and Frechet Inception Distance (FID) of 7.4, improving over the previous best IS of 52.52 and FID of 18.6.

Introduction

The `BigGAN/BigGAN-Deep` is a conditional generation model that can generate both high-resolution and high-quality images by scaling up the batch size and the number of model parameters.

We have finished training `BigGAN` in `Cifar10` (32x32) and are aligning training performance in `ImageNet1k` (128x128). Some sampled results are shown below for your reference.

Evaluation of our trained BigGAN.

Note on reproducibility

`BigGAN 128x128` model is trained with V100 GPUs and CUDA 10.1 and can hardly reproduce the result with A100 and CUDA 11.3. If you have any idea about the reproducibility, please feel free to contact with us.

Converted weights

Since we haven't finished training our models, we provide you with several pre-trained weights which have been evaluated. Here, we refer to BigGAN-PyTorch and pytorch-pretrained-BigGAN.

Evaluation results and download links are provided below.

Sampling results are shown below.

```
python demo/conditional_demo.py CONFIG_PATH CKPT_PATH --sample-cfg truncation=0.4 ## set␣
↪truncation value as you want
```

For converted weights, we provide model configs under `configs/_base_/models` listed as follows:

```
## biggan_cvt-BigGAN-PyTorch-rgb_imagenet1k-128x128.py
## biggan-deep_cvt-hugging-face-rgb_imagenet1k-128x128.py
## biggan-deep_cvt-hugging-face_rgb_imagenet1k-256x256.py
## biggan-deep_cvt-hugging-face_rgb_imagenet1k-512x512.py
```

Interpolation

To perform image Interpolation on BigGAN(or other conditional models), run

```
python apps/conditional_interpolate.py CONFIG_PATH  CKPT_PATH  --samples-path SAMPLES_
↪PATH
```

To perform image Interpolation on BigGAN with fixed noise, run

```
python apps/conditional_interpolate.py CONFIG_PATH  CKPT_PATH  --samples-path SAMPLES_
↪PATH --fix-z
```

```
python apps/conditional_interpolate.py CONFIG_PATH  CKPT_PATH  --samples-path SAMPLES_
↪PATH --fix-y
```

Citation

```
@inproceedings{
    brock2018large,
    title={Large Scale {GAN} Training for High Fidelity Natural Image Synthesis},
    author={Andrew Brock and Jeff Donahue and Karen Simonyan},
    booktitle={International Conference on Learning Representations},
    year={2019},
    url={https://openreview.net/forum?id=B1xsqj09Fm},
}
```

1.27.3 SAGAN (ICML'2019)

> Self-attention generative adversarial networks
>
> **Task**: Conditional GANs

Abstract

In this paper, we propose the Self-Attention Generative Adversarial Network (SAGAN) which allows attention-driven, long-range dependency modeling for image generation tasks. Traditional convolutional GANs generate high-resolution details as a function of only spatially local points in lower-resolution feature maps. In SAGAN, details can be generated using cues from all feature locations. Moreover, the discriminator can check that highly detailed features in distant portions of the image are consistent with each other. Furthermore, recent work has shown that generator conditioning affects GAN performance. Leveraging this insight, we apply spectral normalization to the GAN generator and find that this improves training dynamics. The proposed SAGAN performs better than prior work, boosting the best published Inception score from 36.8 to 52.52 and reducing Fréchet Inception distance from 27.62 to 18.65 on the challenging ImageNet dataset. Visualization of the attention layers shows that the generator leverages neighborhoods that correspond to object shapes rather than local regions of fixed shape.

Results and models

'*' Iteration counting rule in our implementation is different from others. If you want to align with other codebases, you can use the following conversion formula:

```
total_iters (biggan/pytorch studio gan) = our_total_iters / dist_step
```

We also provide converted pre-train models from Pytorch-StudioGAN. To be noted that, in Pytorch Studio GAN, **inplace ReLU** is used in generator and discriminator.

- `Our Pipeline` denote results evaluated with our pipeline.
- `StudioGAN` denote results released by Pytorch-StudioGAN.

For IS metric, our implementation is different from PyTorch-Studio GAN in the following aspects:

1. We use Tero's Inception for feature extraction.
2. We use bicubic interpolation with PIL backend to resize image before feed them to Inception.

For FID evaluation, we follow the pipeline of BigGAN, where the whole training set is adopted to extract inception statistics, and Pytorch Studio GAN uses 50000 randomly selected samples. Besides, we also use Tero's Inception for feature extraction.

You can download the preprocessed inception state by the following url: CIFAR10 and ImageNet1k.

You can use following commands to extract those inception states by yourself.

```
## For CIFAR10
python tools/utils/inception_stat.py --data-cfg configs/_base_/datasets/cifar10_
↪inception_stat.py --pklname cifar10.pkl --no-shuffle --inception-style stylegan --num-
↪samples -1 --subset train

## For ImageNet1k
python tools/utils/inception_stat.py --data-cfg configs/_base_/datasets/imagenet_128x128_
↪inception_stat.py --pklname imagenet.pkl --no-shuffle --inception-style stylegan --num-
↪samples -1 --subset train
```

Citation

```
@inproceedings{zhang2019self,
  title={Self-attention generative adversarial networks},
  author={Zhang, Han and Goodfellow, Ian and Metaxas, Dimitris and Odena, Augustus},
  booktitle={International conference on machine learning},
  pages={7354--7363},
  year={2019},
  organization={PMLR},
  url={https://proceedings.mlr.press/v97/zhang19d.html},
}
```

1.27.4 SNGAN (ICLR'2018)

Spectral Normalization for Generative Adversarial Networks

Task: Conditional GANs

Abstract

One of the challenges in the study of generative adversarial networks is the instability of its training. In this paper, we propose a novel weight normalization technique called spectral normalization to stabilize the training of the discriminator. Our new normalization technique is computationally light and easy to incorporate into existing implementations. We tested the efficacy of spectral normalization on CIFAR10, STL-10, and ILSVRC2012 dataset, and we experimentally confirmed that spectrally normalized GANs (SN-GANs) is capable of generating images of better or equal quality relative to the previous training stabilization techniques.

Results and models

'*' Iteration counting rule in our implementation is different from others. If you want to align with other codebases, you can use the following conversion formula:

```
total_iters (biggan/pytorch studio gan) = our_total_iters / disc_step
```

We also provide converted pre-train models from Pytorch-StudioGAN. To be noted that, in Pytorch Studio GAN, **inplace ReLU** is used in generator and discriminator.

- `Our Pipeline` denote results evaluated with our pipeline.
- `StudioGAN` denote results released by Pytorch-StudioGAN.

For IS metric, our implementation is different from PyTorch-Studio GAN in the following aspects:

1. We use Tero's Inception for feature extraction.
2. We use bicubic interpolation with PIL backend to resize image before feed them to Inception.

For FID evaluation, we follow the pipeline of BigGAN, where the whole training set is adopted to extract inception statistics, and Pytorch Studio GAN uses 50000 randomly selected samples. Besides, we also use Tero's Inception for feature extraction.

You can download the preprocessed inception state by the following url: CIFAR10 and ImageNet1k.

You can use following commands to extract those inception states by yourself.

```
## For CIFAR10
python tools/utils/inception_stat.py --data-cfg configs/_base_/datasets/cifar10_
↪inception_stat.py --pklname cifar10.pkl --no-shuffle --inception-style stylegan --num-
↪samples -1 --subset train

## For ImageNet1k
python tools/utils/inception_stat.py --data-cfg configs/_base_/datasets/imagenet_128x128_
↪inception_stat.py --pklname imagenet.pkl --no-shuffle --inception-style stylegan --num-
↪samples -1 --subset train
```

Citation

```
@inproceedings{miyato2018spectral,
  title={Spectral Normalization for Generative Adversarial Networks},
  author={Miyato, Takeru and Kataoka, Toshiki and Koyama, Masanori and Yoshida, Yuichi},
  booktitle={International Conference on Learning Representations},
  year={2018},
  url={https://openreview.net/forum?id=B1QRgziT-},
}
```

1.28 video interpolation

1.28.1 Summary

- Number of checkpoints: 7
- Number of configs: 7
- Number of papers: 3
  - ALGORITHM: 3

1.28.2 CAIN (AAAI'2020)

Channel Attention Is All You Need for Video Frame Interpolation

Task: Video Interpolation

Abstract

Prevailing video frame interpolation techniques rely heavily on optical flow estimation and require additional model complexity and computational cost; it is also susceptible to error propagation in challenging scenarios with large motion and heavy occlusion. To alleviate the limitation, we propose a simple but effective deep neural network for video frame interpolation, which is end-to-end trainable and is free from a motion estimation network component. Our algorithm employs a special feature reshaping operation, referred to as PixelShuffle, with a channel attention, which replaces the optical flow computation module. The main idea behind the design is to distribute the information in a feature map into multiple channels and extract motion information by attending the channels for pixel-level frame synthesis. The model given by this principle turns out to be effective in the presence of challenging motion and occlusion. We construct a comprehensive evaluation benchmark and demonstrate that the proposed approach achieves outstanding performance compared to the existing models with a component for optical flow computation.

Results and models

Evaluated on RGB channels. The metrics are `PSNR / SSIM` . The learning rate adjustment strategy is `Step LR scheduler with min_lr clipping`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/cain/cain_g1b32_1xb5_vimeo90k-
↪triplet.py

## single-gpu train
python tools/train.py configs/cain/cain_g1b32_1xb5_vimeo90k-triplet.py

## multi-gpu train
./tools/dist_train.sh configs/cain/cain_g1b32_1xb5_vimeo90k-triplet.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/cain/cain_g1b32_1xb5_vimeo90k-
↪triplet.py https://download.openmmlab.com/mmediting/video_interpolators/cain/cain_b5_
↪g1b32_vimeo90k_triplet_20220530-3520b00c.pth

## single-gpu test
python tools/test.py configs/cain/cain_g1b32_1xb5_vimeo90k-triplet.py https://download.
↪openmmlab.com/mmediting/video_interpolators/cain/cain_b5_g1b32_vimeo90k_triplet_
↪20220530-3520b00c.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/cain/cain_g1b32_1xb5_vimeo90k-triplet.py https://download.
↪openmmlab.com/mmediting/video_interpolators/cain/cain_b5_g1b32_vimeo90k_triplet_
↪20220530-3520b00c.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{choi2020channel,
  title={Channel attention is all you need for video frame interpolation},
  author={Choi, Myungsub and Kim, Heewon and Han, Bohyung and Xu, Ning and Lee, Kyoung␣
↪Mu},
  booktitle={Proceedings of the AAAI Conference on Artificial Intelligence},
  volume={34},
  number={07},
  pages={10663--10671},
  year={2020}
}
```

1.28.3 FLAVR (arXiv'2020)

FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation

Task: Video Interpolation

Abstract

Most modern frame interpolation approaches rely on explicit bidirectional optical flows between adjacent frames, thus are sensitive to the accuracy of underlying flow estimation in handling occlusions while additionally introducing computational bottlenecks unsuitable for efficient deployment. In this work, we propose a flow-free approach that is completely end-to-end trainable for multi-frame video interpolation. Our method, FLAVR, is designed to reason about non-linear motion trajectories and complex occlusions implicitly from unlabeled videos and greatly simplifies the process of training, testing and deploying frame interpolation models. Furthermore, FLAVR delivers up to 6× speed up compared to the current state-of-the-art methods for multi-frame interpolation while consistently demonstrating superior qualitative and quantitative results compared with prior methods on popular benchmarks including Vimeo-90K, Adobe-240FPS, and GoPro. Finally, we show that frame interpolation is a competitive self-supervised pre-training task for videos via demonstrating various novel applications of FLAVR including action recognition, optical flow estimation, motion magnification, and video object tracking. Code and trained models are provided in the supplementary material.

Results and models

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Note: FLAVR for x8 VFI task will supported in the future.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/flavr/flavr_in4out1_8xb4_vimeo90k-
↪septuplet.py

## single-gpu train
python tools/train.py configs/flavr/flavr_in4out1_8xb4_vimeo90k-septuplet.py

## multi-gpu train
./tools/dist_train.sh configs/flavr/flavr_in4out1_8xb4_vimeo90k-septuplet.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/flavr/flavr_in4out1_8xb4_vimeo90k-
↪septuplet.py https://download.openmmlab.com/mmediting/video_interpolators/flavr/flavr_
↪in4out1_g8b4_vimeo90k_septuplet_20220509-c2468995.pth

## single-gpu test
python tools/test.py configs/flavr/flavr_in4out1_8xb4_vimeo90k-septuplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/flavr/flavr_in4out1_g8b4_vimeo90k_
↪septuplet_20220509-c2468995.pth

## multi-gpu test
./tools/dist_test.sh configs/flavr/flavr_in4out1_8xb4_vimeo90k-septuplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/flavr/flavr_in4out1_g8b4_vimeo90k_
↪septuplet_20220509-c2468995.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{kalluri2020flavr,
  title={Flavr: Flow-agnostic video representations for fast frame interpolation},
  author={Kalluri, Tarun and Pathak, Deepak and Chandraker, Manmohan and Tran, Du},
  journal={arXiv preprint arXiv:2012.08512},
  year={2020}
}
```

1.28.4 TOFlow (IJCV'2019)

Video Enhancement with Task-Oriented Flow

Task: Video Interpolation, Video Super-Resolution

Abstract

Many video enhancement algorithms rely on optical flow to register frames in a video sequence. Precise flow estimation is however intractable; and optical flow itself is often a sub-optimal representation for particular video processing tasks. In this paper, we propose task-oriented flow (TOFlow), a motion representation learned in a self-supervised, task-specific manner. We design a neural network with a trainable motion estimation component and a video processing component, and train them jointly to learn the task-oriented flow. For evaluation, we build Vimeo-90K, a large-scale, high-quality video dataset for low-level video processing. TOFlow outperforms traditional optical flow on standard benchmarks as well as our Vimeo-90K dataset in three video processing tasks: frame interpolation, video denoising/deblocking, and video super-resolution.

Results and models

Evaluated on Vimeo90k-triplet (RGB channels). The metrics are `PSNR / SSIM` .

Note: These pretrained SPyNets don't contain BN layer since `batch_size=1`, which is consistent with `https://github.com/Coldog2333/pytoflow`.

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

TOF only supports video interpolation task for training now.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/tof/tof_spynet-chair-wobn_1xb1_
↪vimeo90k-triplet.py

## single-gpu train
python tools/train.py configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py

## multi-gpu train
./tools/dist_train.sh configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

TOF supports two tasks for testing.

Task 1: Video Interpolation

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/tof/tof_spynet-chair-wobn_1xb1_
↪vimeo90k-triplet.py https://download.openmmlab.com/mmediting/video_interpolators/
↪toflow/pretrained_spynet_chair_20220321-4d82e91b.pth

## single-gpu test
python tools/test.py configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/toflow/pretrained_spynet_chair_
↪20220321-4d82e91b.pth

## multi-gpu test
./tools/dist_test.sh configs/tof/tof_spynet-chair-wobn_1xb1_vimeo90k-triplet.py https://
↪download.openmmlab.com/mmediting/video_interpolators/toflow/pretrained_spynet_chair_
↪20220321-4d82e91b.pth 8
```

Task 2: Video Super-Resolution

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/tof/tof_x4_official_vimeo90k.py␣
↪https://download.openmmlab.com/mmediting/restorers/tof/tof_x4_vimeo90k_official-
↪a569ff50.pth

## single-gpu test
python tools/test.py configs/tof/tof_x4_official_vimeo90k.py https://download.openmmlab.
↪com/mmediting/restorers/tof/tof_x4_vimeo90k_official-a569ff50.pth

## multi-gpu test
./tools/dist_test.sh configs/tof/tof_x4_official_vimeo90k.py https://download.openmmlab.
↪com/mmediting/restorers/tof/tof_x4_vimeo90k_official-a569ff50.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{xue2019video,
  title={Video enhancement with task-oriented flow},
  author={Xue, Tianfan and Chen, Baian and Wu, Jiajun and Wei, Donglai and Freeman,␣
↪William T},
  journal={International Journal of Computer Vision},
  volume={127},
  number={8},
  pages={1106--1125},
  year={2019},
  publisher={Springer}
}
```

1.29 text2image

1.29.1 Summary

- Number of checkpoints: 6
- Number of configs: 18
- Number of papers: 8
  - ALGORITHM: 8

1.29.2 Control Net (2023)

> Adding Conditional Control to Text-to-Image Diffusion Models
>
> **Task**: Text2Image

Abstract

We present a neural network structure, ControlNet, to control pretrained large diffusion models to support additional input conditions. The ControlNet learns task-specific conditions in an end-to-end way, and the learning is robust even when the training dataset is small (< 50k). Moreover, training a ControlNet is as fast as fine-tuning a diffusion model, and the model can be trained on a personal devices. Alternatively, if powerful computation clusters are available, the model can scale to large amounts (millions to billions) of data. We report that large diffusion models like Stable Diffusion can be augmented with ControlNets to enable conditional inputs like edge maps, segmentation maps, keypoints, etc. This may enrich the methods to control large diffusion models and further facilitate related applications.

Pretrained models

We use ControlNet's weights provided by HuggingFace Diffusers. You do not have to download the weights manually. If you use Diffusers wrapper, the weights will be downloaded automatically.

This model has several weights including vae, unet and clip. You should download the weights from stable-diffusion-1.5 and change the 'pretrained_model_path' in config to the weights dir.

Noted that, ControlNet-Demo is a demo config to train ControlNet with toy dataset named Fill50K.

Besides above configs, ControlNet have weight with other condition inputs, such as depth, hed, mlsd, normal, scribble. You can simple change the `from_pretrained` field of ControlNet to use these weights. For example:

```
## Switch from canny....
controlnet=dict(
    type='ControlNetModel',
    from_pretrained='lllyasviel/sd-controlnet-canny')

## To normal....
controlnet=dict(
    type='ControlNetModel',
    from_pretrained='lllyasviel/sd-controlnet-normal')
```

Quick Start

Running the following codes, you can get a text-generated image.

```
import cv2
import numpy as np
import mmcv
from mmengine import Config
from PIL import Image

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

register_all_modules()

cfg = Config.fromfile('configs/controlnet/controlnet-canny.py')
controlnet = MODELS.build(cfg.model).cuda()

prompt = 'Room with blue walls and a yellow ceiling.'
control_url = 'https://user-images.githubusercontent.com/28132635/230288866-99603172-
↪04cb-47b3-8adb-d1aa532d1d2c.jpg'
control_img = mmcv.imread(control_url)
control = cv2.Canny(control_img, 100, 200)
control = control[:, :, None]
control = np.concatenate([control] * 3, axis=2)
control = Image.fromarray(control)

output_dict = controlnet.infer(prompt, control=control)
samples = output_dict['samples']
for idx, sample in enumerate(samples):
    sample.save(f'sample_{idx}.png')
controls = output_dict['controls']
for idx, control in enumerate(controls):
    control.save(f'control_{idx}.png')
```

If you want to pretrained weights rather than original Stable-Diffusion v1.5, you can refers to the following codes.

```
import mmcv
from mmengine import Config
from PIL import Image

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

register_all_modules()

cfg = Config.fromfile('configs/controlnet/controlnet-pose.py')
## convert ControlNet's weight from SD-v1.5 to Counterfeit-v2.5
cfg.model.unet.from_pretrained = 'gsdf/Counterfeit-V2.5'
cfg.model.vae.from_pretrained = 'gsdf/Counterfeit-V2.5'
cfg.model.init_cfg['type'] = 'convert_from_unet'

controlnet = MODELS.build(cfg.model).cuda()
## call init_weights manually to convert weight
```

(continues on next page)

(continued from previous page)

```
controlnet.init_weights()

prompt = 'masterpiece, best quality, sky, black hair, skirt, sailor collar, looking at␣
↪viewer, short hair, building, bangs, neckerchief, long sleeves, cloudy sky, power␣
↪lines, shirt, cityscape, pleated skirt, scenery, blunt bangs, city, night, black␣
↪sailor collar, closed mouth'

control_url = 'https://user-images.githubusercontent.com/28132635/230380893-2eae68af-
↪d610-4f7f-aa68-c2f22c2abf7e.png'
control_img = mmcv.imread(control_url)
control = Image.fromarray(control_img)
control.save('control.png')

output_dict = controlnet.infer(prompt, control=control, width=512, height=512, guidance_
↪scale=7.5)
samples = output_dict['samples']
for idx, sample in enumerate(samples):
    sample.save(f'sample_{idx}.png')
controls = output_dict['controls']
for idx, control in enumerate(controls):
    control.save(f'control_{idx}.png')
```

Using MMInferencer

You can only use several lines of codes to play controlnet by MMagic!

```
from mmagic.apis import MMagicInferencer

## controlnet-canny
controlnet_canny_inferencer = MMagicInferencer(model_name='controlnet', model_setting=1)
text_prompts = 'Room with blue walls and a yellow ceiling.'
control = 'https://user-images.githubusercontent.com/28132635/230297033-4f5c32df-365c-
↪4cf4-8e4f-1b76a4cbb0b7.png'
result_out_dir = 'controlnet_canny_res.png'
controlnet_canny_inferencer.infer(text=text_prompts, control=control, result_out_
↪dir=result_out_dir)

## controlnet-pose
controlnet_pose_inferencer = MMagicInferencer(model_name='controlnet', model_setting=2)
text_prompts = 'masterpiece, best quality, sky, black hair, skirt, sailor collar,␣
↪looking at viewer, short hair, building, bangs, neckerchief, long sleeves, cloudy sky,␣
↪power lines, shirt, cityscape, pleated skirt, scenery, blunt bangs, city, night, black␣
↪sailor collar, closed mouth'
control = 'https://user-images.githubusercontent.com/28132635/230380893-2eae68af-d610-
↪4f7f-aa68-c2f22c2abf7e.png'
result_out_dir = 'controlnet_pose_res.png'
controlnet_pose_inferencer.infer(text=text_prompts, control=control, result_out_
↪dir=result_out_dir)

## controlnet-seg
controlnet_seg_inferencer = MMagicInferencer(model_name='controlnet', model_setting=3)
```

(continues on next page)

(continued from previous page)

```
text_prompts = 'black house, blue sky'
control = 'https://github-production-user-asset-6210df.s3.amazonaws.com/49083766/
↪243599897-553a4c46-c61d-46df-b820-59a49aaf6678.png'
result_out_dir = 'controlnet_seg_res.png'
controlnet_seg_inferencer.infer(text=text_prompts, control=control, result_out_
↪dir=result_out_dir)
```

Train your own ControlNet!

You can start training your own ControlNet with the toy dataset Fill50K with the following command:

```
bash tools/dist_train.sh configs/controlnet/controlnet-1xb1-demo_dataset 1
```

If you want use gradient accumulation, you can add `accumulative_counts` field to the optimizer's config as follow:

```
## From...
optim_wrapper = dict(controlnet=dict(optimizer=dict(type='AdamW', lr=1e-5)))
## To...
optim_wrapper = dict(
    controlnet=dict(accumulative_counts=4, optimizer=dict(type='AdamW', lr=1e-5)))
```

Use ToMe to accelerate your training and inference

We support **tomesd** now! It is developed for stable-diffusion-based models referring to ToMe, an efficient ViT speed-up tool based on token merging. To work on with **tomesd** in `mmagic`, you just need to add `tomesd_cfg` to `model` in ControlNet-Canny. The only requirement is `torch >= 1.12.1` in order to properly support `torch.Tensor.scatter_reduce()` functionality. Please do check it before running the demo.

```
model = dict(
    type='ControlStableDiffusion',
    ...
    tomesd_cfg=dict(ratio=0.5),
    ...
    init_cfg=dict(type='init_from_unet'))
```

For more details, you can refer to Stable Diffusion Acceleration.

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5 and ControlNet.

Thanks for the efforts of the community!

Citation

```
@misc{zhang2023adding,
      title={Adding Conditional Control to Text-to-Image Diffusion Models},
      author={Lvmin Zhang and Maneesh Agrawala},
      year={2023},
      eprint={2302.05543},
      archivePrefix={arXiv},
      primaryClass={cs.CV}
}
```

1.29.3 Stable Diffusion XL (2023)

Stable Diffusion XL

Task: Text2Image, Inpainting

Abstract

We present SDXL, a latent diffusion model for text-to-image synthesis. Compared to previous versions of Stable Diffusion, SDXL leverages a three times larger UNet backbone: The increase of model parameters is mainly due to more attention blocks and a larger cross-attention context as SDXL uses a second text encoder. We design multiple novel conditioning schemes and train SDXL on multiple aspect ratios. We also introduce a refinement model which is used to improve the visual fidelity of samples generated by SDXL using a post-hoc image-to-image technique. We demonstrate that SDXL shows drastically improved performance compared the previous versions of Stable Diffusion and achieves results competitive with those of black-box state-of-the-art image generators.

Pretrained models

We use stable diffusion xl weights. This model has several weights including vae, unet and clip.

You may download the weights from stable-diffusion-xl and change the 'from_pretrained' in config to the weights dir.

Quick Start

Running the following codes, you can get a text-generated image.

```
from mmengine import MODELS, Config

from mmengine.registry import init_default_scope

init_default_scope('mmagic')

config = 'configs/stable_diffusion_xl/stable-diffusion_xl_ddim_denoisingunet.py'
config = Config.fromfile(config).copy()

StableDiffuser = MODELS.build(config.model)
prompt = 'A mecha robot in a favela in expressionist style'
StableDiffuser = StableDiffuser.to('cuda')
```

(continues on next page)

(continued from previous page)

```
image = StableDiffuser.infer(prompt)['samples'][0]
image.save('robot.png')
```

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-xl.

Thanks for the efforts of the community!

1.29.4 ViCo (2023)

> ViCo: Detail-Preserving Visual Condition for Personalized Text-to-Image Generation
>
> **Task**: Text2Image

Abstract

Personalized text-to-image generation using diffusion models has recently been proposed and attracted lots of attention. Given a handful of images containing a novel concept (e.g., a unique toy), we aim to tune the generative model to capture fine visual details of the novel concept and generate photorealistic images following a text condition. We present a plug-in method, named ViCo, for fast and lightweight personalized generation. Specifically, we propose an image attention module to condition the diffusion process on the patch-wise visual semantics. We introduce an attention-based object mask that comes almost at no cost from the attention module. In addition, we design a simple regularization based on the intrinsic properties of text-image attention maps to alleviate the common overfitting degradation. Unlike many existing models, our method does not finetune any parameters of the original diffusion model. This allows more flexible and transferable model deployment. With only light parameter training (~6% of the diffusion U-Net), our method achieves comparable or even better performance than all state-of-the-art models both qualitatively and quantitatively.

Configs

Quick Start

1. Download concept data and imagenet_templates_small.txt from here. and save to `data/vico/`

The file structure will be like this:

```
data
└── vico
    └──batman
       ├── 1.jpg
       ├── 2.jpg
       ├── 3.jpg
       └── 4.jpg
    └──clock
       ├── 1.jpg
       ├── 2.jpg
       ├── 3.jpg
       └── 4.jpg
    ...
    └──imagenet_templates_small.txt
```

2. Customize your config

```
## Only need to care about these

## which concept you want to customize
concept_dir = 'dog7'

## the new token to denote the concept
placeholder: str = 'S*'

## better to be the superclass of concept
initialize_token: str = 'dog'
```

3. Start training with the following command:

```
## 4 GPUS
bash tools/dist_train.sh configs/vico/vico.py 4
## 1 GPU
python tools/train.py configs/vico/vico.py
```

4. Use the pretrained checkpoins to inference

```
import torch
from mmengine import Config
from PIL import Image

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

register_all_modules()

## say you have downloaded the pretrained weights
cfg = Config.fromfile('configs/vico/dog.py')
state_dict = torch.load("./dog.pth")
vico = MODELS.build(cfg.model)
vico.load_state_dict(state_dict, strict=False)
vico = vico.cuda()

prompt = ["A photo of S*", "A photo of S* on the beach"]
reference = "data/vico/dog7/01.jpg"
image_ref = Image.open(reference)
with torch.no_grad():
    output = vico.infer(prompt=prompt, image_reference=image_ref, seed=123, num_images_
↪per_prompt=2)['samples'][0]
output.save("infer.png")
```

5. (Optional) If you want to use the weight trained by the commands at step3, here are codes to extract the trained parameters, then you can infer with it like step4

```
import torch
def extract_vico_parameters(state_dict):
    new_state_dict = dict()
    for k, v in state_dict.items():
        if 'image_cross_attention' in k or 'trainable_embeddings' in k:
```

(continues on next page)

(continued from previous page)

```
            new_k = k.replace('module.', '')
            new_state_dict[new_k] = v
    return new_state_dict

checkpoint = torch.load("work_dirs/vico/iter_400.pth")
new_checkpoint = extract_vico_parameters(checkpoint['state_dict'])
torch.save(new_checkpoint, "work_dirs/vico/dog.pth")
```

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5.

Thanks for the efforts of the community!

Citation

```
@inproceedings{Hao2023ViCo,
  title={ViCo: Detail-Preserving Visual Condition for Personalized Text-to-Image␣
↪Generation},
  author={Shaozhe Hao and Kai Han and Shihao Zhao and Kwan-Yee K. Wong},
  year={2023}
}
```

1.29.5 FastComposer (2023)

FastComposer: Tuning-Free Multi-Subject Image Generation with Localized Attention

Task: Text2Image

Abstract

Diffusion models excel at text-to-image generation, especially in subject-driven generation for personalized images. However, existing methods are inefficient due to the subject-specific fine-tuning, which is computationally intensive and hampers efficient deployment. Moreover, existing methods struggle with multi-subject generation as they often blend features among subjects. We present FastComposer which enables efficient, personalized, multi-subject text-to-image generation without fine-tuning. FastComposer uses subject embeddings extracted by an image encoder to augment the generic text conditioning in diffusion models, enabling personalized image generation based on subject images and textual instructions with only forward passes. To address the identity blending problem in the multi-subject generation, FastComposer proposes cross-attention localization supervision during training, enforcing the attention of reference subjects localized to the correct regions in the target images. Naively conditioning on subject embeddings results in subject overfitting. FastComposer proposes delayed subject conditioning in the denoising step to maintain both identity and editability in subject-driven image generation. FastComposer generates images of multiple unseen individuals with different styles, actions, and contexts. It achieves 300x-2500x speedup compared to fine-tuning-based methods and requires zero extra storage for new subjects. FastComposer paves the way for efficient, personalized, and high-quality multi-subject image creation.

Pretrained models

This model has several weights including vae, unet and clip. You should download the weights from stable-diffusion-1.5 and clipModel,and then change the 'stable_diffusion_v15_url' and 'clip_vit_url' in config to the corresponding weights path and "finetuned_model_path" to the weight path of fastcomposer.

Quick Start

You can run the demo locally by

```
python demo/gradio_fastcomposer.py
```

Or running the following codes, you can get a text-generated image.

```
import numpy as np
import mmcv
from mmengine import Config
from PIL import Image

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules
import torch, gc

gc.collect()
torch.cuda.empty_cache()

register_all_modules()

cfg_file = Config.fromfile('configs/fastcomposer/fastcomposer_8xb16_FFHQ.py')

fastcomposer = MODELS.build(cfg_file.model).cuda()

prompt = "A man img and a man img sitting in a park"
negative_prompt = "((((ugly)))), (((duplicate))), ((morbid)), ((mutilated)), [out of␣
↪frame], extra fingers, mutated hands, ((poorly drawn hands)), ((poorly drawn face)),␣
↪(((mutation))), (((deformed))), ((ugly)), blurry, ((bad anatomy)), (((bad␣
↪proportions))), ((extra limbs)), cloned face, (((disfigured))). out of frame, ugly,␣
↪extra limbs, (bad anatomy), gross proportions, (malformed limbs), ((missing arms)),␣
↪((missing legs)), (((extra arms))), (((extra legs))), mutated hands, (fused fingers),␣
↪(too many fingers), (((long neck)))"
alpha_ = 0.75
guidance_scale = 5
num_steps = 50
num_images = 1
image = []
seed = -1

image1 = mmcv.imread('https://user-images.githubusercontent.com/14927720/265911400-
↪91635451-54b6-4dc6-92a7-c1d02f88b62e.jpeg')
image2 = mmcv.imread('https://user-images.githubusercontent.com/14927720/265911502-
↪66b67f53-dff0-4d25-a9af-3330e446aa48.jpeg')

image.append(Image.fromarray(image1))
```

(continues on next page)

(continued from previous page)

```
image.append(Image.fromarray(image2))

if len(image) == 0:
    raise Exception("You need to upload at least one image.")

num_subject_in_text = (
        np.array(fastcomposer.special_tokenizer.encode(prompt))
        == fastcomposer.image_token_id
).sum()
if num_subject_in_text != len(image):
    raise Exception(f"Number of subjects in the text description doesn't match the␣
↪number of reference images, #text subjects: {num_subject_in_text} #reference image:
↪{len(image)}",
    )

if seed == -1:
    seed = np.random.randint(0, 1000000)

device = torch.device('cuda' if torch.cuda.is_available(
    ) else 'cpu')
generator = torch.Generator(device=device)
generator.manual_seed(seed)

output_dict = fastcomposer.infer(prompt,
                                 negative_prompt=negative_prompt,
                                 height=512,
                                 width=512,
                                 num_inference_steps=num_steps,
                                 guidance_scale=guidance_scale,
                                 num_images_per_prompt=num_images,
                                 generator=generator,
                                 alpha_=alpha_,
                                 reference_subject_images=image)

samples = output_dict['samples']
for idx, sample in enumerate(samples):
    sample.save(f'sample_{idx}.png')
```

Citation

```
@article{xiao2023fastcomposer,
            title={FastComposer: Tuning-Free Multi-Subject Image Generation with␣
↪Localized Attention},
            author={Xiao, Guangxuan and Yin, Tianwei and Freeman, William T. and Durand,␣
↪Frédo and Han, Song},
            journal={arXiv},
            year={2023}
          }
```

1.29.6 DreamBooth (2022)

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

Task: Text2Image

Abstract

Large text-to-image models achieved a remarkable leap in the evolution of AI, enabling high-quality and diverse synthesis of images from a given text prompt. However, these models lack the ability to mimic the appearance of subjects in a given reference set and synthesize novel renditions of them in different contexts. In this work, we present a new approach for "personalization" of text-to-image diffusion models. Given as input just a few images of a subject, we fine-tune a pretrained text-to-image model such that it learns to bind a unique identifier with that specific subject. Once the subject is embedded in the output domain of the model, the unique identifier can be used to synthesize novel photorealistic images of the subject contextualized in different scenes. By leveraging the semantic prior embedded in the model with a new autogenous class-specific prior preservation loss, our technique enables synthesizing the subject in diverse scenes, poses, views and lighting conditions that do not appear in the reference images. We apply our technique to several previously-unassailable tasks, including subject recontextualization, text-guided view synthesis, and artistic rendering, all while preserving the subject's key features. We also provide a new dataset and evaluation protocol for this new task of subject-driven generation.

Configs

Quick Start

1. Download data and save to `data/dreambooth/`

The file structure will be like this:

```
data
└── dreambooth
    └──imgs
       ├── alvan-nee-Id1DBHv4fbg-unsplash.jpeg
       ├── alvan-nee-bQaAJCbNq3g-unsplash.jpeg
       ├── alvan-nee-brFsZ7qszSY-unsplash.jpeg
       └── alvan-nee-eoqnr8ikwFE-unsplash.jpeg
```

2. Start training with the following command:

```
bash tools/dist_train.sh configs/dreambooth/dreambooth.py 1
## or
bash tools/dist_train.sh configs/dreambooth/dreambooth-lora.py 1
```

Use ToMe to accelerate your training and inference

We support **tomesd** now! It is developed for stable-diffusion-based models referring to ToMe, an efficient ViT speed-up tool based on token merging. To work on with **tomesd** in `mmagic`, you just need to add `tomesd_cfg` to `model` in DreamBooth. The only requirement is `torch >= 1.12.1` in order to properly support `torch.Tensor.scatter_reduce()` functionality. Please do check it before running the demo.

```
model = dict(
    type='DreamBooth',
    ...
    tomesd_cfg=dict(ratio=0.5),
    ...
    val_prompts=val_prompts)
```

For more details, you can refer to Stable Diffusion Acceleration.

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5.

Thanks for the efforts of the community!

Citation

```
@article{ruiz2022dreambooth,
  title={Dreambooth: Fine tuning text-to-image diffusion models for subject-driven␣
↪generation},
  author={Ruiz, Nataniel and Li, Yuanzhen and Jampani, Varun and Pritch, Yael and␣
↪Rubinstein, Michael and Aberman, Kfir},
  journal={arXiv preprint arXiv:2208.12242},
  year={2022}
}
```

1.29.7 Disco Diffusion (2022)

> Disco Diffusion
>
> **Task**: Text2Image, Image2Image

Abstract

Disco Diffusion (DD) is a Google Colab Notebook which leverages an AI Image generating technique called CLIP-Guided Diffusion to allow you to create compelling and beautiful images from text inputs.

Created by Somnai, augmented by Gandamu, and building on the work of RiversHaveWings, nshepperd, and many others. See more details in *Credits*.

Results and models

We have converted several `unet` weights and offer related configs. See more details of different `unet` in *Tutorial*.

To-do List

- [x] Text2Image
- [x] Image2Image
- [x] Imagenet, portrait diffusion models
- [ ] pixelart, watercolor, sci-fiction diffusion models
- [ ] image prompt
- [ ] video generation
- [ ] faster sampler(plms, dpm-solver etc.)

We really welcome community users supporting these items and any other interesting stuffs!

Quick Start

Running the following codes, you can get a text-generated image.

```
from mmengine import Config, MODELS
from mmengine.registry import init_default_scope
from torchvision.utils import save_image

init_default_scope('mmagic')

disco = MODELS.build(
    Config.fromfile('configs/disco_diffusion/disco-baseline.py').model).cuda().eval()
text_prompts = {
    0: [
        "A beautiful painting of a singular lighthouse, shining its light across a␣
↪tumultuous sea of blood by greg rutkowski and thomas kinkade, Trending on artstation.",
        "yellow color scheme"
    ]
}
image = disco.infer(
    height=768,
    width=1280,
    text_prompts=text_prompts,
    show_progress=True,
    num_inference_steps=250,
    eta=0.8)['samples']
save_image(image, "image.png")
```

Tutorials

Considering that `disco-diffusion` contains many adjustable parameters, we provide users with a jupyter-notebook / colab tutorial that exhibits the meaning of different parameters, and gives results corresponding to adjustment. Refer to Disco Sheet.

Credits

Since our adaptation of disco-diffusion are heavily influenced by disco colab, here we copy the credits below.

Modified by Daniel Russell (https://github.com/russelldc, https://twitter.com/danielrussruss) to include (hopefully) optimal params for quick generations in 15-100 timesteps rather than 1000, as well as more robust augmentations.

Further improvements from Dango233 and nshepperd helped improve the quality of diffusion in general, and especially so for shorter runs like this notebook aims to achieve.

Vark added code to load in multiple Clip models at once, which all prompts are evaluated against, which may greatly improve accuracy.

The latest zoom, pan, rotation, and keyframes features were taken from Chigozie Nri's VQGAN Zoom Notebook (https://github.com/chigozienri, https://twitter.com/chigozienri)

Advanced DangoCutn Cutout method is also from Dango223.

–

Disco:

Somnai (https://twitter.com/Somnai_dreams) added Diffusion Animation techniques, QoL improvements and various implementations of tech and techniques, mostly listed in the changelog below.

3D animation implementation added by Adam Letts (https://twitter.com/gandamu_ml) in collaboration with Somnai. Creation of disco.py and ongoing maintenance.

Turbo feature by Chris Allen (https://twitter.com/zippy731)

Improvements to ability to run on local systems, Windows support, and dependency installation by HostsServer (https://twitter.com/HostsServer)

VR Mode by Tom Mason (https://twitter.com/nin_artificial)

Horizontal and Vertical symmetry functionality by nshepperd. Symmetry transformation_steps by huemin (https://twitter.com/huemin_art). Symmetry integration into Disco Diffusion by Dmitrii Tochilkin (https://twitter.com/cut_pow).

Warp and custom model support by Alex Spirin (https://twitter.com/devdef).

Pixel Art Diffusion, Watercolor Diffusion, and Pulp SciFi Diffusion models from KaliYuga (https://twitter.com/KaliYuga_ai). Follow KaliYuga's Twitter for the latest models and for notebooks with specialized settings.

Integration of OpenCLIP models and initiation of integration of KaliYuga models by Palmweaver / Chris Scalf (https://twitter.com/ChrisScalf11)

Integrated portrait_generator_v001 from Felipe3DArtist (https://twitter.com/Felipe3DArtist)

Citation

```
@misc{github,
  author={alembics},
  title={disco-diffusion},
  year={2022},
  url={https://github.com/alembics/disco-diffusion},
}
```

1.29.8 Stable Diffusion (2022)

Stable Diffusion

Task: Text2Image, Inpainting

Abstract

Stable Diffusion is a latent diffusion model conditioned on the text embeddings of a CLIP text encoder, which allows you to create images from text inputs. This model builds upon the CVPR'22 work High-Resolution Image Synthesis with Latent Diffusion Models. The official code was released at stable-diffusion and also implemented at diffusers. We support this algorithm here to facilitate the community to learn together and compare it with other text2image methods.

Pretrained models

We use stable diffusion v1.5 weights. This model has several weights including vae, unet and clip.

You may download the weights from stable-diffusion-1.5 and change the 'from_pretrained' in config to the weights dir.

Download with git:

```
git lfs install
git clone https://huggingface.co/runwayml/stable-diffusion-v1-5
```

Quick Start

Running the following codes, you can get a text-generated image.

```
from mmengine import MODELS, Config
from torchvision import utils

from mmengine.registry import init_default_scope

init_default_scope('mmagic')

config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet.py'
config = Config.fromfile(config).copy()
## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-v1-5'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-v1-5'

StableDiffuser = MODELS.build(config.model)
```

(continues on next page)

(continued from previous page)

```
prompt = 'A mecha robot in a favela in expressionist style'
StableDiffuser = StableDiffuser.to('cuda')

image = StableDiffuser.infer(prompt)['samples'][0]
image.save('robot.png')
```

To inpaint an image, you could run the following codes.

```
import mmcv
from mmengine import MODELS, Config
from mmengine.registry import init_default_scope
from PIL import Image

init_default_scope('mmagic')

config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet-inpaint.py'
config = Config.fromfile(config).copy()
## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-inpainting'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-inpainting'

StableDiffuser = MODELS.build(config.model)
prompt = 'a mecha robot sitting on a bench'

img_url = 'https://raw.githubusercontent.com/CompVis/latent-diffusion/main/data/
↪inpainting_examples/overture-creations-5sI6fQgYIuo.png'  ## noqa
mask_url = 'https://raw.githubusercontent.com/CompVis/latent-diffusion/main/data/
↪inpainting_examples/overture-creations-5sI6fQgYIuo_mask.png'  ## noqa

image = Image.fromarray(mmcv.imread(img_url, channel_order='rgb'))
mask = Image.fromarray(mmcv.imread(mask_url)).convert('L')
StableDiffuser = StableDiffuser.to('cuda')

image = StableDiffuser.infer(
    prompt,
    image,
    mask
)['samples'][0]
image.save('inpaint.png')
```

Use ToMe to accelerate your stable diffusion model

We support **tomesd** now! It is developed based on ToMe, an efficient ViT speed-up tool based on token merging. To work on with **tomesd** in `mmagic`, you just need to add `tomesd_cfg` to `model` as shown in stable_diffusion_v1.5_tomesd. The only requirement is `torch >= 1.12.1` in order to properly support `torch.Tensor.scatter_reduce()` functionality. Please do check it before running the demo.

```
...
model = dict(
    type='StableDiffusion',
    unet=unet,
```

(continues on next page)

(continued from previous page)

```
    vae=vae,
    enable_xformers=False,
    text_encoder=dict(
        type='ClipWrapper',
        clip_type='huggingface',
        pretrained_model_name_or_path=stable_diffusion_v15_url,
        subfolder='text_encoder'),
    tokenizer=stable_diffusion_v15_url,
    scheduler=diffusion_scheduler,
    test_scheduler=diffusion_scheduler,
    tomesd_cfg=dict(
        ratio=0.5))
```

The detailed settings for `tomesd_cfg` are as follows:

- `ratio (float)`: The ratio of tokens to merge. For example, 0.4 would reduce the total number of tokens by 40%.The maximum value for this is 1-(1/(`sx` * `sy`)). **By default, the max ratio is 0.75, usually <= 0.5 is recommended.** Higher values result in more speed-up, but with more visual quality loss.
- `max_downsample (int)`: Apply ToMe to layers with at most this amount of downsampling. E.g., 1 only applies to layers with no downsampling, while 8 applies to all layers. Should be chosen from 1, 2, 4, 8. **1, 2 are recommended.**
- `sx, sy (int, int)`: The stride for computing dst sets. A higher stride means you can merge more tokens, **default setting of (2, 2) works well in most cases**. `sx` and `sy` do not need to divide image size.
- `use_rand (bool)`: Whether or not to allow random perturbations when computing dst sets. By default: True, but if you're having weird artifacts you can try turning this off.
- `merge_attn (bool)`: Whether or not to merge tokens for attention **(recommended)**.
- `merge_crossattn (bool)`: Whether or not to merge tokens for cross attention **(not recommended)**.
- `merge_mlp (bool)`: Whether or not to merge tokens for the mlp layers **(especially not recommended)**.

For more details about the **tomesd** setting, please refer to Token Merging for Stable Diffusion.

Then following the code below, you can evaluate the speed-up performance on stable diffusion models or stable-diffusion-based models (DreamBooth, ControlNet).

```
import time
import numpy as np

from mmengine import MODELS, Config
from mmengine.registry import init_default_scope

init_default_scope('mmagic')

_device = 0
work_dir = '/path/to/your/work_dir'
config = 'configs/stable_diffusion/stable-diffusion_ddim_denoisingunet-tomesd_5e-1.py'
config = Config.fromfile(config).copy()
## ## change the 'pretrained_model_path' if you have downloaded the weights manually
## config.model.unet.from_pretrained = '/path/to/your/stable-diffusion-v1-5'
## config.model.vae.from_pretrained = '/path/to/your/stable-diffusion-v1-5'

## w/o tomesd
```

(continues on next page)

(continued from previous page)

```
config.model.tomesd_cfg = None
StableDiffuser = MODELS.build(config.model).to(f'cuda:{_device}')
prompt = 'A mecha robot in a favela in expressionist style'

## inference time evaluation params
size = 512
ratios = [0.5, 0.75]
samples_perprompt = 5

t = time.time()
for i in range(100//samples_perprompt):
    image = StableDiffuser.infer(prompt, height=size, width=size, num_images_per_
↪prompt=samples_perprompt)['samples'][0]
    if i == 0:
        image.save(f"{work_dir}/wo_tomesd.png")
print(f"Generating 100 images with {samples_perprompt} images per prompt, without ToMe␣
↪speed-up, time used : {time.time() - t}s")

for ratio in ratios:
    ## w/ tomesd
    config.model.tomesd_cfg = dict(ratio=ratio)
    sd_model = MODELS.build(config.model).to(f'cuda:{_device}')

    t = time.time()
    for i in range(100//samples_perprompt):
        image = sd_model.infer(prompt, height=size, width=size, num_images_per_
↪prompt=samples_perprompt)['samples'][0]
        if i == 0:
            image.save(f"{work_dir}/w_tomesd_ratio_{ratio}.png")

    print(f"Generating 100 images with {samples_perprompt} images per prompt, merging␣
↪ratio {ratio}, time used : {time.time() - t}s")
```

Here are some inference performance comparisons running on **single RTX 3090** with `torch 2.0.0+cu118` as backends. The results are reasonable, when enabling `xformers`, the speed-up ratio is a little bit lower. But `tomesd` still effectively reduces the inference time. It is especially recommended that enable `tomesd` when the `image_size` and `num_images_per_prompt` are large, since the number of similar tokens are larger and `tomesd` can achieve better performance.

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5.

Thanks for the efforts of the community!

Citation

```
@misc{rombach2021highresolution,
      title={High-Resolution Image Synthesis with Latent Diffusion Models},
      author={Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser␣
↪and Björn Ommer},
      year={2021},
      eprint={2112.10752},
      archivePrefix={arXiv},
      primaryClass={cs.CV}
}

@article{bolya2023tomesd,
  title={Token Merging for Fast Stable Diffusion},
  author={Bolya, Daniel and Hoffman, Judy},
  journal={arXiv},
  year={2023}
}

@inproceedings{bolya2023tome,
  title={Token Merging: Your {ViT} but Faster},
  author={Bolya, Daniel and Fu, Cheng-Yang and Dai, Xiaoliang and Zhang, Peizhao and␣
↪Feichtenhofer, Christoph and Hoffman, Judy},
  booktitle={International Conference on Learning Representations},
  year={2023}
}
```

1.29.9 Textual Inversion (2022)

An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion

Task: Text2Image

Abstract

Text-to-image models offer unprecedented freedom to guide creation through natural language. Yet, it is unclear how such freedom can be exercised to generate images of specific unique concepts, modify their appearance, or compose them in new roles and novel scenes. In other words, we ask: how can we use language-guided models to turn our cat into a painting, or imagine a new product based on our favorite toy? Here we present a simple approach that allows such creative freedom. Using only 3-5 images of a user-provided concept, like an object or a style, we learn to represent it through new "words" in the embedding space of a frozen text-to-image model. These "words" can be composed into natural language sentences, guiding personalized creation in an intuitive way. Notably, we find evidence that a single word embedding is sufficient for capturing unique and varied concepts. We compare our approach to a wide range of baselines, and demonstrate that it can more faithfully portray the concepts across a range of applications and tasks.

Configs

Quick Start

1. Download data and template(two txt files) and save to `data`

The file structure will be like this:

```
data
└── cat_toy
    ├── 1.jpeg
    ├── 2.jpeg
    ├── 3.jpeg
    ├── 3.jpeg
    ├── 4.jpeg
    ├── 6.jpeg
    └── 7.jpeg
└── imagenet_templates_small.txt
└── imagenet_style_templates_small.txt
```

2. Start training with the following command:

```
bash tools/dist_train.sh configs/textual_inversion/textual_inversion.py 1
```

3. Inference with trained textual embedding:

```
import torch
from mmengine import Config

from mmagic.registry import MODELS
from mmagic.utils import register_all_modules

register_all_modules()


def process_state_dict(state_dict):
    new_state_dict = dict()
    for k, v in state_dict.items():
        new_k = k.replace('module.', '')
        new_state_dict[new_k] = v

    return new_state_dict


cfg = Config.fromfile('configs/textual_inversion/textual_inversion.py')
checkpoint = torch.load('work_dirs/textual_inversion/iter_3000.pth')
state_dict = process_state_dict(checkpoint['state_dict'])
model = MODELS.build(cfg.model)
model.load_state_dict(state_dict)

model = model.cuda()
with torch.no_grad():
    sample = model.infer('a <cat-toy> bag')['samples'][0]

sample.save('cat-toy-bag.png')
```

Comments

Our codebase for the stable diffusion models builds heavily on diffusers codebase and the model weights are from stable-diffusion-1.5.

Thanks for the efforts of the community!

Citation

```
@misc{gal2022textual,
      doi = {10.48550/ARXIV.2208.01618},
      url = {https://arxiv.org/abs/2208.01618},
      author = {Gal, Rinon and Alaluf, Yuval and Atzmon, Yuval and Patashnik, Or and Bermano, Amit H. and Chechik, Gal and Cohen-Or, Daniel},
      title = {An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion},
      publisher = {arXiv},
      year = {2022},
      primaryClass={cs.CV}
}
```

1.30 controlnet_animation

1.30.1 Summary

- Number of checkpoints: 1
- Number of configs: 1
- Number of papers: 1
  - ALGORITHM: 1

1.30.2 Controlnet Animation (2023)

Controlnet Application

Task: controlnet_animation

Abstract

It is difficult to keep consistency and avoid video frame flickering when using stable diffusion to generate video frame by frame. Here we reproduce two methods that effectively avoid video flickering:

Controlnet with multi-frame rendering. ControlNet is a neural network structure to control diffusion models by adding extra conditions. Multi-frame rendering is a community method to reduce flickering. We use controlnet with hed condition and stable diffusion img2img for multi-frame rendering.

Controlnet with attention injection. Attention injection is widely used to generate the current frame from a reference image. There is an implementation in sd-webui-controlnet and we use some of their code to create the animation in this repo.

You may need 40G GPU memory to run controlnet with multi-frame rendering and 10G GPU memory for controlnet with attention injection. If the config file is not changed, it defaults to using controlnet with attention injection.

Demos

prompt key words: a handsome man, silver hair, smiling, play basketball

prompt key words: a handsome man

Change prompt to get different result

prompt key words: a girl, black hair, white pants, smiling, play basketball

Pretrained models

We use pretrained model from hugging face.

Quick Start

There are two ways to try controlnet animation.

1. Use MMagic inference API.

Running the following codes, you can get an generated animation video.

```
from mmagic.apis import MMagicInferencer

## Create a MMEdit instance and infer
editor = MMagicInferencer(model_name='controlnet_animation')

prompt = 'a girl, black hair, T-shirt, smoking, best quality, extremely detailed'
negative_prompt = 'longbody, lowres, bad anatomy, bad hands, missing fingers, ' + \
                  'extra digit, fewer digits, cropped, worst quality, low quality'

## you can download the example video with this link
## https://user-images.githubusercontent.com/12782558/227418400-80ad9123-7f8e-4c1a-8e19-
↪0892ebad2a4f.mp4
video = '/path/to/your/input/video.mp4'
save_path = '/path/to/your/output/video.mp4'

## Do the inference to get result
editor.infer(video=video, prompt=prompt, negative_prompt=negative_prompt, save_path=save_
↪path)
```

2. Use controlnet animation gradio demo.

```
python demo/gradio_controlnet_animation.py
```

3. Change config to use multi-frame rendering or attention injection.

change "inference_method" in anythingv3 config

To use multi-frame rendering.

```
inference_method = 'multi-frame rendering'
```

To use attention injection.

```
inference_method = 'attention_injection'
```

Play animation with SAM

We also provide a demo to play controlnet animation with sam, for details, please see OpenMMLab PlayGround.

Citation

```
@misc{zhang2023adding,
  title={Adding Conditional Control to Text-to-Image Diffusion Models},
  author={Lvmin Zhang and Maneesh Agrawala},
  year={2023},
  eprint={2302.05543},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

1.31 image2image

1.31.1 Summary

- Number of checkpoints: 10
- Number of configs: 10
- Number of papers: 2
  - ALGORITHM: 2

1.31.2 Pix2Pix (CVPR'2017)

Pix2Pix: Image-to-Image Translation with Conditional Adversarial Networks

Task: Image2Image

Abstract

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. Moreover, since the release of the pix2pix software associated with this paper, hundreds of twitter users have posted their own artistic experiments using our system. As a community, we no longer hand-engineer our mapping functions, and this work suggests we can achieve reasonable results without handengineering our loss functions either.

Results and Models

`FID` comparison with official:

`IS` comparison with official:

Note:

1. we strictly follow the paper setting in Section 3.3: "*At inference time, we run the generator net in exactly the same manner as during the training phase. This differs from the usual protocol in that we apply dropout at test time, and we apply batch normalization using the statistics of the test batch, rather than aggregated statistics of the training batch.*" (i.e., use model.train() mode), thus may lead to slightly different inference results every time.
2. This is the training log before refactoring. Updated logs will be released soon.

Citation

```
@inproceedings{isola2017image,
  title={Image-to-image translation with conditional adversarial networks},
  author={Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={1125--1134},
  year={2017},
  url={https://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_
↪Translation_With_CVPR_2017_paper.html},
}
```

1.31.3 CycleGAN (ICCV'2017)

CycleGAN: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks

Task: Image2Image

Abstract

Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping G: X \rightarrow Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping F: Y \rightarrow X and introduce a cycle consistency loss to push F(G(X)) \approx X (and vice versa). Qualitative results are presented on several tasks where paired training data does not exist, including collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior methods demonstrate the superiority of our approach.

Results and Models

We use `FID` and `IS` metrics to evaluate the generation performance of CycleGAN.1 https://download.openmmlab.com/mmediting/cyclegan/refactor/cyclegan_lsgan_resnet_in_1x1_80k_facades_20210902_165905-5e2c0876.pth https://download.openmmlab.com/mmediting/cyclegan/refactor/cyclegan_in_1x1_80k_facades_20210902_165905-5e2c0876.pth

`FID` comparison with official:

`IS` comparison with evaluation:

Note:

1. With a larger identity loss, the image-to-image translation becomes more conservative, which makes less changes. The original authors did not say what is the best weight for identity loss. Thus, in addition to the default setting, we also set the weight of identity loss to 0 (denoting `id0`) to make a more comprehensive comparison.
2. This is the training log before refactoring. Updated logs will be released soon.

Citation

```
@inproceedings{zhu2017unpaired,
  title={Unpaired image-to-image translation using cycle-consistent adversarial networks}
↪,
  author={Zhu, Jun-Yan and Park, Taesung and Isola, Phillip and Efros, Alexei A},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={2223--2232},
  year={2017},
  url={https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_
↪Translation_ICCV_2017_paper.html},
}
```

1.32 unconditional gans

1.32.1 Summary

- Number of checkpoints: 56
- Number of configs: 57
- Number of papers: 9
  - ALGORITHM: 9

1.32.2 Positional Encoding in GANs (CVPR'2021)

Positional Encoding as Spatial Inductive Bias in GANs

Task: Unconditional GANs

Abstract

SinGAN shows impressive capability in learning internal patch distribution despite its limited effective receptive field. We are interested in knowing how such a translation-invariant convolutional generator could capture the global structure with just a spatially i.i.d. input. In this work, taking SinGAN and StyleGAN2 as examples, we show that such capability, to a large extent, is brought by the implicit positional encoding when using zero padding in the generators. Such positional encoding is indispensable for generating images with high fidelity. The same phenomenon is observed in other generative architectures such as DCGAN and PGGAN. We further show that zero padding leads to an unbalanced spatial bias with a vague relation between locations. To offer a better spatial inductive bias, we investigate alternative positional encodings and analyze their effects. Based on a more flexible positional encoding explicitly, we propose a new multi-scale training strategy and demonstrate its effectiveness in the state-of-the-art unconditional generator StyleGAN2. Besides, the explicit spatial inductive bias substantially improve SinGAN for more versatile image manipulation.

Results and models for MS-PIE

Note that we report the FID and P&R metric (FFHQ dataset) in the largest scale.

Results and Models for SinGAN

Citation

```
@article{xu2020positional,
  title={Positional Encoding as Spatial Inductive Bias in GANs},
  author={Xu, Rui and Wang, Xintao and Chen, Kai and Zhou, Bolei and Loy, Chen Change},
  journal={arXiv preprint arXiv:2012.05217},
  year={2020},
  url={https://openaccess.thecvf.com/content/CVPR2021/html/Xu_Positional_Encoding_As_
↪Spatial_Inductive_Bias_in_GANs_CVPR_2021_paper.html},
}
```

1.32.3 StyleGANv3 (NeurIPS'2021)

Alias-Free Generative Adversarial Networks

Task: Unconditional GANs

Abstract

We observe that despite their hierarchical convolutional nature, the synthesis process of typical generative adversarial networks depends on absolute pixel coordinates in an unhealthy manner. This manifests itself as, e.g., detail appearing to be glued to image coordinates instead of the surfaces of depicted objects. We trace the root cause to careless signal processing that causes aliasing in the generator network. Interpreting all signals in the network as continuous, we derive generally applicable, small architectural changes that guarantee that unwanted information cannot leak into the hierarchical synthesis process. The resulting networks match the FID of StyleGAN2 but differ dramatically in their internal representations, and they are fully equivariant to translation and rotation even at subpixel scales. Our results pave the way for generative models better suited for video and animation.

Results and Models

We perform experiments on StyleGANv3 paper settings and also experimental settings. For user convenience, we also offer the converted version of official weights.

Paper Settings

Experimental Settings

Converted Weights

Interpolation

We provide a tool to generate video by walking through GAN's latent space. Run this command to get the following video.

```
python apps/interpolate_sample.py configs/styleganv3/stylegan3_t_afhqv2_512_b4x8_
↪official.py https://download.openmmlab.com/mmediting/stylegan3/stylegan3_t_afhqv2_512_
↪b4x8_cvt_official.pkl --export-video --samples-path work_dirs/demos/ --endpoint 6 --
↪interval 60 --space z --seed 2022 --sample-cfg truncation=0.8
```

https://user-images.githubusercontent.com/22982797/151506918-83da9ee3-0d63-4c5b-ad53-a41562b92075.mp4

Equivarience Visualization && Evaluation

We also provide a tool to visualize the equivarience properties for StyleGAN3. Run these commands to get the results below.

```
python tools/utils/equivariance_viz.py configs/styleganv3/stylegan3_r_ffhqu_256_b4x8_
↪official.py https://download.openmmlab.com/mmediting/stylegan3/stylegan3_r_ffhqu_256_
↪b4x8_cvt_official.pkl --translate_max 0.5 --transform rotate --seed 5432

python tools/utils/equivariance_viz.py configs/styleganv3/stylegan3_r_ffhqu_256_b4x8_
↪official.py https://download.openmmlab.com/mmediting/stylegan3/stylegan3_r_ffhqu_256_
↪b4x8_cvt_official.pkl --translate_max 0.25 --transform x_t --seed 5432
```

(continues on next page)

(continued from previous page)

```
python tools/utils/equivariance_viz.py configs/styleganv3/stylegan3_r_ffhqu_256_b4x8_
↪official.py https://download.openmmlab.com/mmediting/stylegan3/stylegan3_r_ffhqu_256_
↪b4x8_cvt_official.pkl --translate_max 0.25 --transform y_t --seed 5432
```

https://user-images.githubusercontent.com/22982797/151504902-f3cbfef5-9014-4607-bbe1-deaf48ec6d55.mp4

https://user-images.githubusercontent.com/22982797/151504973-b96e1639-861d-434b-9d7c-411ebd4a653f.mp4

https://user-images.githubusercontent.com/22982797/151505099-cde4999e-aab1-42d4-a458-3bb069db3d32.mp4

If you want to get EQ-Metric for StyleGAN3, just add following codes into config.

```
metrics = dict(
    eqv=dict(
        type='Equivariance',
        num_images=50000,
        eq_cfg=dict(
            compute_eqt_int=True, compute_eqt_frac=True, compute_eqr=True)))
```

And we highly recommend you to use slurm_test.sh script to accelerate evaluation time.

Citation

```
@inproceedings{Karras2021,
  author = {Tero Karras and Miika Aittala and Samuli Laine and Erik H\"ark\"onen and␣
↪Janne Hellsten and Jaakko Lehtinen and Timo Aila},
  title = {Alias-Free Generative Adversarial Networks},
  booktitle = {Proc. NeurIPS},
  year = {2021}
}
```

1.32.4 StyleGANv2 (CVPR'2020)

Analyzing and Improving the Image Quality of Stylegan

Task: Unconditional GANs

Abstract

The style-based GAN architecture (StyleGAN) yields state-of-the-art results in data-driven unconditional generative image modeling. We expose and analyze several of its characteristic artifacts, and propose changes in both model architecture and training methods to address them. In particular, we redesign the generator normalization, revisit progressive growing, and regularize the generator to encourage good conditioning in the mapping from latent codes to images. In addition to improving image quality, this path length regularizer yields the additional benefit that the generator becomes significantly easier to invert. This makes it possible to reliably attribute a generated image to a particular network. We furthermore visualize how well the generator utilizes its output resolution, and identify a capacity problem, motivating us to train larger models for additional quality improvements. Overall, our improved model redefines the state of the art in unconditional image modeling, both in terms of existing distribution quality metrics as well as perceived image quality.

Results and Models

FP16 Support and Experiments

Currently, we have supported FP16 training for StyleGAN2, and here are the results for the mixed-precision training. (Experiments for FFHQ1024 will come soon.)

As shown in the figure, we provide **3** ways to do mixed-precision training for `StyleGAN2`:

- stylegan2_c2_fp16_PL-no-scaler: In this setting, we try our best to follow the official FP16 implementation in StyleGAN2-ADA. Similar to the official version, we only adopt FP16 training for the higher-resolution feature maps (the last 4 stages in G and the first 4 stages). Note that we do not adopt the `clamp` way to avoid gradient overflow used in the official implementation. We use the `autocast` function from `torch.cuda.amp` package.
- stylegan2_c2_fp16-globalG-partialD_PL-R1-no-scaler: In this config, we try to adopt mixed-precision training for the whole generator, but in partial discriminator (the first 4 higher-resolution stages). Note that we do not apply the loss scaler in the path length loss and gradient penalty loss. Because we always meet divergence after adopting the loss scaler to scale the gradient in these two losses.
- stylegan2_c2_apex_fp16_PL-R1-no-scaler: In this setting, we adopt the APEX toolkit to implement mixed-precision training with multiple loss/gradient scalers. In APEX, you can assign different loss scalers for the generator and the discriminator respectively. Note that we still ignore the gradient scaler in the path length loss and gradient penalty loss.

As shown in this table, `P&R50k_full` is the metric used in StyleGANv1 and StyleGANv2. `full` indicates that we use the whole dataset for extracting the real distribution, e.g., 70000 images in FFHQ dataset. However, adopting the VGG16 provided from Tero requires that your PyTorch version must fulfill `>=1.6.0`. Be careful about using the PyTorch's VGG16 to extract features, which will cause higher precision and recall.

Citation

```
@inproceedings{karras2020analyzing,
  title={Analyzing and improving the image quality of stylegan},
  author={Karras, Tero and Laine, Samuli and Aittala, Miika and Hellsten, Janne and Lehtinen, Jaakko and Aila, Timo},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition},
  pages={8110--8119},
  year={2020},
  url={https://openaccess.thecvf.com/content_CVPR_2020/html/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.html},
}
```

1.32.5 StyleGANv1 (CVPR'2019)

A Style-Based Generator Architecture for Generative Adversarial Networks

Task: Unconditional GANs

Abstract

We propose an alternative generator architecture for generative adversarial networks, borrowing from style transfer literature. The new architecture leads to an automatically learned, unsupervised separation of high-level attributes (e.g., pose and identity when trained on human faces) and stochastic variation in the generated images (e.g., freckles, hair), and it enables intuitive, scale-specific control of the synthesis. The new generator improves the state-of-the-art in terms of traditional distribution quality metrics, leads to demonstrably better interpolation properties, and also better disentangles the latent factors of variation. To quantify interpolation quality and disentanglement, we propose two new, automated methods that are applicable to any generator architecture. Finally, we introduce a new, highly varied and high-quality dataset of human faces.

Results and Models

Citation

```
@inproceedings{karras2019style,
  title={A style-based generator architecture for generative adversarial networks},
  author={Karras, Tero and Laine, Samuli and Aila, Timo},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern␣
↪Recognition},
  pages={4401--4410},
  year={2019},
  url={https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_
↪Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html},
}
```

1.32.6 PGGAN (ICLR'2018)

Progressive Growing of GANs for Improved Quality, Stability, and Variation

Task: Unconditional GANs

Abstract

We describe a new training methodology for generative adversarial networks. The key idea is to grow both the generator and discriminator progressively: starting from a low resolution, we add new layers that model increasingly fine details as training progresses. This both speeds the training up and greatly stabilizes it, allowing us to produce images of unprecedented quality, e.g., CelebA images at 1024^2. We also propose a simple way to increase the variation in generated images, and achieve a record inception score of 8.80 in unsupervised CIFAR10. Additionally, we describe several implementation details that are important for discouraging unhealthy competition between the generator and discriminator. Finally, we suggest a new metric for evaluating GAN results, both in terms of image quality and variation. As an additional contribution, we construct a higher-quality version of the CelebA dataset.

Results and models

Citation

```
@article{karras2017progressive,
  title={Progressive growing of gans for improved quality, stability, and variation},
  author={Karras, Tero and Aila, Timo and Laine, Samuli and Lehtinen, Jaakko},
  journal={arXiv preprint arXiv:1710.10196},
  year={2017},
  url={https://arxiv.org/abs/1710.10196},
}
```

1.32.7 WGAN-GP (NeurIPS'2017)

Improved Training of Wasserstein GANs

Task: Unconditional GANs

Abstract

Generative Adversarial Networks (GANs) are powerful generative models, but suffer from training instability. The recently proposed Wasserstein GAN (WGAN) makes progress toward stable training of GANs, but sometimes can still generate only low-quality samples or fail to converge. We find that these problems are often due to the use of weight clipping in WGAN to enforce a Lipschitz constraint on the critic, which can lead to undesired behavior. We propose an alternative to clipping weights: penalize the norm of gradient of the critic with respect to its input. Our proposed method performs better than standard WGAN and enables stable training of a wide variety of GAN architectures with almost no hyperparameter tuning, including 101-layer ResNets and language models over discrete data. We also achieve high quality generations on CIFAR-10 and LSUN bedrooms.

Results and models

Citation

```
@article{gulrajani2017improved,
  title={Improved Training of Wasserstein GANs},
  author={Gulrajani, Ishaan and Ahmed, Faruk and Arjovsky, Martin and Dumoulin, Vincent␣
↪and Courville, Aaron},
  journal={arXiv preprint arXiv:1704.00028},
  year={2017},
  url={https://arxiv.org/abs/1704.00028},
}
```

1.32.8 GGAN (ArXiv'2017)

Geometric GAN

Task: Unconditional GANs

Abstract

Generative Adversarial Nets (GANs) represent an important milestone for effective generative models, which has inspired numerous variants seemingly different from each other. One of the main contributions of this paper is to reveal a unified geometric structure in GAN and its variants. Specifically, we show that the adversarial generative model training can be decomposed into three geometric steps: separating hyperplane search, discriminator parameter update away from the separating hyperplane, and the generator update along the normal vector direction of the separating hyperplane. This geometric intuition reveals the limitations of the existing approaches and leads us to propose a new formulation called geometric GAN using SVM separating hyperplane that maximizes the margin. Our theoretical analysis shows that the geometric GAN converges to a Nash equilibrium between the discriminator and generator. In addition, extensive numerical results show that the superior performance of geometric GAN.

Results and models

Note: In the original implementation of GGAN, they set `G_iters` to 10. However our framework does not support `G_iters` currently, so we dropped the settings in the original implementation and conducted several experiments with our own settings. We have shown above the experiment results with the lowest `fid` score.
Original settings and our settings:

Citation

```
@article{lim2017geometric,
  title={Geometric gan},
  author={Lim, Jae Hyun and Ye, Jong Chul},
  journal={arXiv preprint arXiv:1705.02894},
  year={2017},
  url={https://arxiv.org/abs/1705.02894},
}
```

1.32.9 LSGAN (ICCV'2017)

Least Squares Generative Adversarial Networks

Task: Unconditional GANs

Abstract

Unsupervised learning with generative adversarial networks (GANs) has proven hugely successful. Regular GANs hypothesize the discriminator as a classifier with the sigmoid cross entropy loss function. However, we found that this loss function may lead to the vanishing gradients problem during the learning process. To overcome such a problem, we propose in this paper the Least Squares Generative Adversarial Networks (LSGANs) which adopt the least squares loss function for the discriminator. We show that minimizing the objective function of LSGAN yields minimizing the Pearson 2 divergence. There are two benefits of LSGANs over regular GANs. First, LSGANs are able to generate higher quality images than regular GANs. Second, LSGANs perform more stable during the learning process. We evaluate LSGANs on five scene datasets and the experimental results show that the images generated by LSGANs are of better quality than the ones generated by regular GANs. We also conduct two comparison experiments between LSGANs and regular GANs to illustrate the stability of LSGANs.

Results and models

Citation

```
@inproceedings{mao2017least,
  title={Least squares generative adversarial networks},
  author={Mao, Xudong and Li, Qing and Xie, Haoran and Lau, Raymond YK and Wang, Zhen␣
↪and Paul Smolley, Stephen},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={2794--2802},
  year={2017},
  url={https://openaccess.thecvf.com/content_iccv_2017/html/Mao_Least_Squares_Generative_
↪ICCV_2017_paper.html},
}
```

1.32.10 Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (ICLR'2016)

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Task: Unconditional GANs

Abstract

In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention. In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations.

Results and models

Citation

```
@article{radford2015unsupervised,
  title={Unsupervised representation learning with deep convolutional generative␣
↪adversarial networks},
  author={Radford, Alec and Metz, Luke and Chintala, Soumith},
  journal={arXiv preprint arXiv:1511.06434},
  year={2015},
  url={https://arxiv.org/abs/1511.06434},
}
```

1.33 deblurring

1.33.1 Summary

- Number of checkpoints: 5
- Number of configs: 5
- Number of papers: 2
  - ALGORITHM: 2

1.33.2 Restormer (CVPR'2022)

> Restormer: Efficient Transformer for High-Resolution Image Restoration
>
> **Task**: Denoising, Deblurring, Deraining

Abstract

Since convolutional neural networks (CNNs) perform well at learning generalizable image priors from large-scale data, these models have been extensively applied to image restoration and related tasks. Recently, another class of neural architectures, Transformers, have shown significant performance gains on natural language and high-level vision tasks. While the Transformer model mitigates the shortcomings of CNNs (i.e., limited receptive field and inadaptability to input content), its computational complexity grows quadratically with the spatial resolution, therefore making it infeasible to apply to most image restoration tasks involving high-resolution images. In this work, we propose an efficient Transformer model by making several key designs in the building blocks (multi-head attention and feed-forward network) such that it can capture long-range pixel interactions, while still remaining applicable to large images. Our model, named Restoration Transformer (Restormer), achieves state-of-the-art results on several image restoration tasks, including image deraining, single-image motion deblurring, defocus deblurring (single-image and dual-pixel data), and image denoising (Gaussian grayscale/color denoising, and real image denoising).

Results and models

Deraining

Evaluated on Y channels. The metrics are `PSNR` / `SSIM` .

Motion Deblurring

Evaluated on RGB channels for GoPro and HIDE, and Y channel for ReakBlur-J and ReakBlur-R. The metrics are `PSNR` / `SSIM` .

Defocus Deblurring

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` / `MAE` / `LPIPS`.

Gaussian Denoising

Test Grayscale Gaussian Noise

Evaluated on grayscale images. The metrics are `PSNR` / `SSIM` .

training a separate model for each noise level

Model	Dataset	Task	σ	PSNR	SSIM	Training Resources	Download	
restormer_official_dfwb-gray-sigma15	Set12	Denoising	15	34.0182	0.9160	1	model	log
restormer_official_dfwb-gray-sigma15	BSD68	Denoising	15	32.4987	0.8940	1	model	log
restormer_official_dfwb-gray-sigma15	Urban100	Denoising	15	34.4336	0.9419	1	model	log
restormer_official_dfwb-gray-sigma25	Set12	Denoising	25	31.7289	0.8811	1	model	log
restormer_official_dfwb-gray-sigma25	BSD68	Denoising	25	30.1613	0.8370	1	model	log
restormer_official_dfwb-gray-sigma25	Urban100	Denoising	25	32.1162	0.9140	1	model	log
restormer_official_dfwb-gray-sigma50	Set12	Denoising	50	28.6269	0.8188	1	model	log
restormer_official_dfwb-gray-sigma50	BSD68	Denoising	50	27.3266	0.7434	1	model	log
restormer_official_dfwb-gray-sigma50	Urban100	Denoising	50	28.9636	0.8571	1	model	log

learning a single model to handle various noise levels

Test Color Gaussian Noise

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` . **training a separate model for each noise level**

learning a single model to handle various noise levels

Model	Dataset	Task	σ	PSNR (RGB)	SSIM (RGB)	Training Resources	Download	
restormer_official_dfwb-color-sigma15	CBSD68	Denoising	15	34.3422	0.9356	1	model	log
restormer_official_dfwb-color-sigma15	Kodak24	Denoising	15	35.4544	0.9308	1	model	log
restormer_official_dfwb-color-sigma15	McMaster	Denoising	15	35.5473	0.9344	1	model	log
restormer_official_dfwb-color-sigma15	Urban100	Denoising	15	35.0754	0.9524	1	model	log
restormer_official_dfwb-color-sigma25	CBSD68	Denoising	25	31.7391	0.8945	1	model	log
restormer_official_dfwb-color-sigma25	Kodak24	Denoising	25	33.0380	0.8941	1	model	log
restormer_official_dfwb-color-sigma25	McMaster	Denoising	25	33.3040	0.9063	1	model	log
restormer_official_dfwb-color-sigma25	Urban100	Denoising	25	32.9165	0.9312	1	model	log
restormer_official_dfwb-color-sigma50	CBSD68	Denoising	50	28.5582	0.8126	1	model	log
restormer_official_dfwb-color-sigma50	Kodak24	Denoising	50	30.0074	0.8233	1	model	

log | | restormer_official_dfwb-color-sigma50| McMaster |Denoising| 50 | 30.2671 | 0.8520 | 1 | model | log | | restormer_official_dfwb-color-sigma50| Urban100 |Denoising| 50 | 30.0172 | 0.8898 | 1 | model | log |

Real Image Denoising

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` .

Quick Start

Train

You can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## Deraining
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_
↪rain13k.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪rain13k-2be7b550.pth

## Motion Deblurring
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_gopro.
↪py https://download.openmmlab.com/mmediting/restormer/restormer_official_gopro-
↪db7363a0.pth

## Defocus Deblurring
## Single
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪dual.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪single-6bc31582.pth
## Dual
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪single.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma15-da74417f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma25-08010841.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma50-ee852dfe.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## Test Color Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma15-012ceb71.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma25-e307f222.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma50-a991983d.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## single-gpu test
## Deraining
python tools/test.py configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
python tools/test.py configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
```

(continues on next page)

(continued from previous page)

```
## Single
python tools/test.py configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
python tools/test.py configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth
```

(continues on next page)

(continued from previous page)

```
## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth


## multi-gpu test
## Deraining
./tools/dist_test.sh configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
./tools/dist_test.sh configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
## Single
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{Zamir2021Restormer,
    title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
    author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat
            and Fahad Shahbaz Khan and Ming-Hsuan Yang},
    booktitle={CVPR},
    year={2022}
}
```

1.33.3 DeblurGAN-v2 (ICCV'2019)

DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better

Task: Deblurring

Abstract

We present a new end-to-end generative adversarial network (GAN) for single image motion deblurring, named DeblurGAN-v2, which considerably boosts state-of-the-art deblurring efficiency, quality, and flexibility. DeblurGAN-v2 is based on a relativistic conditional GAN with a double-scale discriminator. For the first time, we introduce the Feature Pyramid Network into deblurring, as a core building block in the generator of DeblurGAN-v2. It can flexibly work with a wide range of backbones, to navigate the balance between performance and efficiency. The plug-in of sophisticated backbones (e.g., Inception-ResNet-v2) can lead to solid state-of-the-art deblurring. Meanwhile, with light-weight backbones (e.g., MobileNet and its variants), DeblurGAN-v2 reaches 10-100 times faster than the nearest competitors, while maintaining close to state-of-the-art results, implying the option of real-time video deblurring. We demonstrate that DeblurGAN-v2 obtains very competitive performance on several popular benchmarks, in terms of deblurring quality (both objective and subjective), as well as efficiency. Besides, we show the architecture to be effective for general image restoration tasks too.

Results and models

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/deblurganv2/deblurganv2_fpn-
↪inception_1xb1_gopro.py

## single-gpu train
python tools/train.py configs/deblurganv2/deblurganv2_fpn-inception_1xb1_gopro.py

## multi-gpu train
./tools/dist_train.sh configs/deblurganv2/deblurganv2_fpn-inception_1xb1_gopro.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/deblurganv2/deblurganv2_fpn-
↪inception_1xb1_gopro.py https://download.openxlab.org.cn/models/xiaomile/DeblurGANv2/
↪weight/DeblurGANv2_fpn-inception.pth

## single-gpu test
python tools/test.py configs/deblurganv2/deblurganv2_fpn-inception_1xb1_gopro.py https://
↪download.openxlab.org.cn/models/xiaomile/DeblurGANv2/weight/DeblurGANv2_fpn-inception.
↪pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/deblurganv2/deblurganv2_fpn-inception_1xb1_gopro.py https://
↪download.openxlab.org.cn/models/xiaomile/DeblurGANv2/weight/DeblurGANv2_fpn-inception.
↪pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{Kupyn_2019_ICCV,
author = {Orest Kupyn and Tetiana Martyniuk and Junru Wu and Zhangyang Wang},
title = {DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better},
booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
month = {Oct},
year = {2019}
}
```

1.34 image super-resolution

1.34.1 Summary

- Number of checkpoints: 37
- Number of configs: 41
- Number of papers: 11
  - ALGORITHM: 11

1.34.2 LIIF (CVPR'2021)

Learning Continuous Image Representation with Local Implicit Image Function

Task: Image Super-Resolution

Abstract

How to represent an image? While the visual world is presented in a continuous manner, machines store and see the images in a discrete way with 2D arrays of pixels. In this paper, we seek to learn a continuous representation for images. Inspired by the recent progress in 3D reconstruction with implicit neural representation, we propose Local Implicit Image Function (LIIF), which takes an image coordinate and the 2D deep features around the coordinate as inputs, predicts the RGB value at a given coordinate as an output. Since the coordinates are continuous, LIIF can be presented in arbitrary resolution. To generate the continuous representation for images, we train an encoder with LIIF representation via a self-supervised task with super-resolution. The learned continuous representation can be presented in arbitrary resolution even extrapolate to x30 higher resolution, where the training tasks are not provided. We further show that LIIF representation builds a bridge between discrete and continuous representation in 2D, it naturally supports the learning tasks with size-varied image ground-truths and significantly outperforms the method with resizing the ground-truths.

Results and models

Note:

- refers to ditto.
- Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/liif/liif-edsr-norm_c64b16_1xb16-
↪1000k_div2k.py

## single-gpu train
python tools/train.py configs/liif/liif-edsr-norm_c64b16_1xb16-1000k_div2k.py

## multi-gpu train
./tools/dist_train.sh configs/liif/liif-edsr-norm_c64b16_1xb16-1000k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/liif/liif-edsr-norm_c64b16_1xb16-
↪1000k_div2k.py https://download.openmmlab.com/mmediting/restorers/liif/liif_edsr_norm_
↪c64b16_g1_1000k_div2k_20210715-ab7ce3fc.pth

## single-gpu test
python tools/test.py configs/liif/liif-edsr-norm_c64b16_1xb16-1000k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/liif/liif_edsr_norm_c64b16_g1_1000k_div2k_
↪20210715-ab7ce3fc.pth

## multi-gpu test
./tools/dist_test.sh configs/liif/liif-edsr-norm_c64b16_1xb16-1000k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/liif/liif_edsr_norm_c64b16_g1_1000k_div2k_
↪20210715-ab7ce3fc.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{chen2021learning,
  title={Learning continuous image representation with local implicit image function},
  author={Chen, Yinbo and Liu, Sifei and Wang, Xiaolong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern␣
↪Recognition},
  pages={8628--8638},
  year={2021}
}
```

1.34.3 SwinIR (ICCVW'2021)

SwinIR: Image Restoration Using Swin Transformer

Task: Image Super-Resolution, Image denoising, JPEG compression artifact reduction

Abstract

Image restoration is a long-standing low-level vision problem that aims to restore high-quality images from low-quality images (e.g., downscaled, noisy and compressed images). While state-of-the-art image restoration methods are based on convolutional neural networks, few attempts have been made with Transformers which show impressive performance on high-level vision tasks. In this paper, we propose a strong baseline model SwinIR for image restoration based on the Swin Transformer. SwinIR consists of three parts: shallow feature extraction, deep feature extraction and high-quality image reconstruction. In particular, the deep feature extraction module is composed of several residual Swin Transformer blocks (RSTB), each of which has several Swin Transformer layers together with a residual connection. We conduct experiments on three representative tasks: image super-resolution (including classical, lightweight and real-world image super-resolution), image denoising (including grayscale and color image denoising) and JPEG compression artifact reduction. Experimental results demonstrate that SwinIR outperforms state-of-the-art methods on different tasks by up to 0.14~0.45dB, while the total number of parameters can be reduced by up to 67%.

Results and models

Classical Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Lightweight Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Real-World Image Super-Resolution

Evaluated on Y channels. The metrics are `NIQE` .

Grayscale Image Deoising

Evaluated on grayscale images. The metrics are `PSNR` .

Color Image Deoising

Evaluated on RGB channels. The metrics are `PSNR` .

JPEG Compression Artifact Reduction (grayscale)

Evaluated on grayscale images. The metrics are `PSNR / SSIM

JPEG Compression Artifact Reduction (color)

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py

## color
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py


## single-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py

## 004 Grayscale Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py

## 005 Color Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
```

(continues on next page)

(continued from previous page)

```
## grayscale
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py

## color
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py


## multi-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
./tools/dist_train.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8

## 003 Real-World Image Super-Resolution
./tools/dist_train.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
```

(continues on next page)

(continued from previous page)

```
## 004 Grayscale Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py 8

## 005 Color Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py 8

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py 8

## color
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s48w8d6e180_8xb4-lr2e-4-500k_div2k-ed2d419e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s48w8d6e180_8xb4-lr2e-4-500k_div2k-926950f1.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s48w8d6e180_8xb4-lr2e-4-500k_div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-69e15fb6.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-d6982f7b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-131d3f64.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-309cb239.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-d6622d03.pth

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-c6425057.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-6f0c425f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-36960d18.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-a016a72f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-9f1599b5.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15-6782691b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25-d0d8d4da.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15-c74a2cee.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25-df2b1c0c.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10-da93c8e9.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20-d47367b1.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30-52c083cf.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40-803e8d9b.pth


## color
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## single-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth

python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth


## 003 Real-World Image Super-Resolution
python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth

python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth


## 004 Grayscale Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth


## 005 Color Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth


## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth


## color
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## multi-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

./tools/dist_test.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

./tools/dist_test.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth

## 003 Real-World Image Super-Resolution
./tools/dist_test.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth
```

(continued on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth
```

(continues on next page)

(continued from previous page)

```
## color
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{liang2021swinir,
  title={Swinir: Image restoration using swin transformer},
  author={Liang, Jingyun and Cao, Jiezhang and Sun, Guolei and Zhang, Kai and Van Gool,␣
↪Luc and Timofte, Radu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={1833--1844},
  year={2021}
}
```

1.34.4 Real-ESRGAN (ICCVW'2021)

Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data

Task: Image Super-Resolution

Abstract

Though many attempts have been made in blind super-resolution to restore low-resolution images with unknown and complex degradations, they are still far from addressing general real-world degraded images. In this work, we extend the powerful ESRGAN to a practical restoration application (namely, Real-ESRGAN), which is trained with pure synthetic data. Specifically, a high-order degradation modeling process is introduced to better simulate complex real-world degradations. We also consider the common ringing and overshoot artifacts in the synthesis process. In addition, we employ a U-Net discriminator with spectral normalization to increase discriminator capability and stabilize the training dynamics. Extensive comparisons have shown its superior visual performance than prior works on various real datasets. We also provide efficient implementations to synthesize training pairs on the fly.

Results and models

Evaluated on Set5 dataset with RGB channels. The metrics are `PSNR` and `SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/real_esrgan/realesrgan_c64b23g32_
↪4xb12-lr1e-4-400k_df2k-ost.py

## single-gpu train
python tools/train.py configs/real_esrgan/realesrgan_c64b23g32_4xb12-lr1e-4-400k_df2k-
↪ost.py

## multi-gpu train
./tools/dist_train.sh configs/real_esrgan/realesrgan_c64b23g32_4xb12-lr1e-4-400k_df2k-
↪ost.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/real_esrgan/realesrgan_c64b23g32_
↪4xb12-lr1e-4-400k_df2k-ost.py https://download.openmmlab.com/mmediting/restorers/real_
↪esrgan/realesrgan_c64b23g32_12x4_lr1e-4_400k_df2k_ost_20211010-34798885.pth

## single-gpu test
python tools/test.py configs/real_esrgan/realesrgan_c64b23g32_4xb12-lr1e-4-400k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/restorers/real_esrgan/realesrgan_c64b23g32_
↪12x4_lr1e-4_400k_df2k_ost_20211010-34798885.pth

## multi-gpu test
./tools/dist_test.sh configs/real_esrgan/realesrgan_c64b23g32_4xb12-lr1e-4-400k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/restorers/real_esrgan/realesrgan_c64b23g32_
↪12x4_lr1e-4_400k_df2k_ost_20211010-34798885.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{wang2021real,
  title={Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic␣
↪data},
  author={Wang, Xintao and Xie, Liangbin and Dong, Chao and Shan, Ying},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision␣
↪Workshop (ICCVW)},
  pages={1905--1914},
  year={2021}
}
```

1.34.5 GLEAN (CVPR'2021)

GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution

Task: Image Super-Resolution

Abstract

We show that pre-trained Generative Adversarial Networks (GANs), e.g., StyleGAN, can be used as a latent bank to improve the restoration quality of large-factor image super-resolution (SR). While most existing SR approaches attempt to generate realistic textures through learning with adversarial loss, our method, Generative LatEnt bANk (GLEAN), goes beyond existing practices by directly leveraging rich and diverse priors encapsulated in a pre-trained GAN. But unlike prevalent GAN inversion methods that require expensive image-specific optimization at runtime, our approach only needs a single forward pass to generate the upscaled image. GLEAN can be easily incorporated in a simple encoder-bank-decoder architecture with multi-resolution skip connections. Switching the bank allows the method to deal with images from diverse categories, e.g., cat, building, human face, and car. Images upscaled by GLEAN show clear improvements in terms of fidelity and texture faithfulness in comparison to existing methods.

Results and models

For the meta info used in training and test, please refer to here. The results are evaluated on RGB channels.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/glean/glean_x8_2xb8_cat.py

## single-gpu train
python tools/train.py configs/glean/glean_x8_2xb8_cat.py

## multi-gpu train
./tools/dist_train.sh configs/glean/glean_x8_2xb8_cat.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/glean/glean_x8_2xb8_cat.py https://
↪download.openmmlab.com/mmediting/restorers/glean/glean_cat_8x_20210614-d3ac8683.pth

## single-gpu test
python tools/test.py configs/glean/glean_x8_2xb8_cat.py https://download.openmmlab.com/
↪mmediting/restorers/glean/glean_cat_8x_20210614-d3ac8683.pth

## multi-gpu test
./tools/dist_test.sh configs/glean/glean_x8_2xb8_cat.py https://download.openmmlab.com/
↪mmediting/restorers/glean/glean_cat_8x_20210614-d3ac8683.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@InProceedings{chan2021glean,
  author = {Chan, Kelvin CK and Wang, Xintao and Xu, Xiangyu and Gu, Jinwei and Loy,␣
↪Chen Change},
  title = {GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  year = {2021}
}
```

1.34.6 DIC (CVPR'2020)

> Deep Face Super-Resolution with Iterative Collaboration between Attentive Recovery and Landmark Estimation
>
> **Task**: Image Super-Resolution

Abstract

Recent works based on deep learning and facial priors have succeeded in super-resolving severely degraded facial images. However, the prior knowledge is not fully exploited in existing methods, since facial priors such as landmark and component maps are always estimated by low-resolution or coarsely super-resolved images, which may be inaccurate and thus affect the recovery performance. In this paper, we propose a deep face super-resolution (FSR) method with iterative collaboration between two recurrent networks which focus on facial image recovery and landmark estimation respectively. In each recurrent step, the recovery branch utilizes the prior knowledge of landmarks to yield higher-quality images which facilitate more accurate landmark estimation in turn. Therefore, the iterative information interaction between two processes boosts the performance of each other progressively. Moreover, a new attentive fusion module is designed to strengthen the guidance of landmark maps, where facial components are generated individually and aggregated attentively for better restoration. Quantitative and qualitative experimental results show the proposed method significantly outperforms state-of-the-art FSR methods in recovering high-quality face images.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM`.

In the log data of `dic_gan_x8c48b6_g4_150k_CelebAHQ`, DICGAN is verified on the first 9 pictures of the test set of CelebA-HQ, so `PSNR` and `SSIM` shown in the follow table is different from the log data.

`Training Resources`: Training Resourcesrmation during training.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/dic/dic_gan-x8c48b6_4xb2-500k_
↪celeba-hq.py

## single-gpu train
python tools/train.py configs/dic/dic_gan-x8c48b6_4xb2-500k_celeba-hq.py

## multi-gpu train
./tools/dist_train.sh configs/dic/dic_gan-x8c48b6_4xb2-500k_celeba-hq.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/dic/dic_gan-x8c48b6_4xb2-500k_
↪celeba-hq.py https://download.openmmlab.com/mmediting/restorers/dic/dic_gan_x8c48b6_g4_
↪500k_CelebAHQ_20210625-3b89a358.pth

## single-gpu test
python tools/test.py configs/dic/dic_gan-x8c48b6_4xb2-500k_celeba-hq.py https://download.
↪openmmlab.com/mmediting/restorers/dic/dic_gan_x8c48b6_g4_500k_CelebAHQ_20210625-
↪3b89a358.pth

## multi-gpu test
./tools/dist_test.sh configs/dic/dic_gan-x8c48b6_4xb2-500k_celeba-hq.py https://download.
↪openmmlab.com/mmediting/restorers/dic/dic_gan_x8c48b6_g4_500k_CelebAHQ_20210625-
↪3b89a358.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{ma2020deep,
  title={Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation},
  author={Ma, Cheng and Jiang, Zhenyu and Rao, Yongming and Lu, Jiwen and Zhou, Jie},
  booktitle={Proceedings of the IEEE/CVF conference on computer vision and pattern recognition},
  pages={5569--5578},
  year={2020}
}
```

1.34.7 TTSR (CVPR'2020)

Learning Texture Transformer Network for Image Super-Resolution

Task: Image Super-Resolution

Abstract

We study on image super-resolution (SR), which aims to recover realistic textures from a low-resolution (LR) image. Recent progress has been made by taking high-resolution images as references (Ref), so that relevant textures can be transferred to LR images. However, existing SR approaches neglect to use attention mechanisms to transfer high-resolution (HR) textures from Ref images, which limits these approaches in challenging cases. In this paper, we propose a novel Texture Transformer Network for Image Super-Resolution (TTSR), in which the LR and Ref images are formulated as queries and keys in a transformer, respectively. TTSR consists of four closely-related modules optimized for image generation tasks, including a learnable texture extractor by DNN, a relevance embedding module, a hard-attention module for texture transfer, and a soft-attention module for texture synthesis. Such a design encourages joint feature learning across LR and Ref images, in which deep feature correspondences can be discovered by attention, and thus accurate texture features can be transferred. The proposed texture transformer can be further stacked in a cross-scale way, which enables texture recovery from different levels (e.g., from 1x to 4x magnification). Extensive experiments show that TTSR achieves significant improvements over state-of-the-art approaches on both quantitative and qualitative evaluations.

Results and models

Evaluated on CUFED dataset (RGB channels), `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR and SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_CUFED.py

## single-gpu train
python tools/train.py configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_CUFED.py
```

(continues on next page)

(continued from previous page)

```
## multi-gpu train
./tools/dist_train.sh configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_CUFED.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_
↪CUFED.py https://download.openmmlab.com/mmediting/restorers/ttsr/ttsr-gan_x4_c64b16_g1_
↪500k_CUFED_20210626-2ab28ca0.pth

## single-gpu test
python tools/test.py configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_CUFED.py https://download.
↪openmmlab.com/mmediting/restorers/ttsr/ttsr-gan_x4_c64b16_g1_500k_CUFED_20210626-
↪2ab28ca0.pth

## multi-gpu test
./tools/dist_test.sh configs/ttsr/ttsr-gan_x4c64b16_1xb9-500k_CUFED.py https://download.
↪openmmlab.com/mmediting/restorers/ttsr/ttsr-gan_x4_c64b16_g1_500k_CUFED_20210626-
↪2ab28ca0.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{yang2020learning,
  title={Learning texture transformer network for image super-resolution},
  author={Yang, Fuzhi and Yang, Huan and Fu, Jianlong and Lu, Hongtao and Guo, Baining},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern␣
↪Recognition},
  pages={5791--5800},
  year={2020}
}
```

1.34.8 RDN (CVPR'2018)

Residual Dense Network for Image Super-Resolution

Task: Image Super-Resolution

Abstract

A very deep convolutional neural network (CNN) has recently achieved great success for image super-resolution (SR) and offered hierarchical features as well. However, most deep CNN based SR models do not make full use of the hierarchical features from the original low-resolution (LR) images, thereby achieving relatively-low performance. In this paper, we propose a novel residual dense network (RDN) to address this problem in image SR. We fully exploit the hierarchical features from all the convolutional layers. Specifically, we propose residual dense block (RDB) to extract abundant local features via dense connected convolutional layers. RDB further allows direct connections from the state of preceding RDB to all the layers of current RDB, leading to a contiguous memory (CM) mechanism. Local feature fusion in RDB is then used to adaptively learn more effective features from preceding and current local features and stabilizes the training of wider network. After fully obtaining dense local features, we use global feature fusion to jointly and adaptively learn global hierarchical features in a holistic way. Extensive experiments on benchmark datasets with different degradation models show that our RDN achieves favorable performance against state-of-the-art methods.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR` and `SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.
↪py

## single-gpu train
python tools/train.py configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.py

## multi-gpu train
./tools/dist_train.sh configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.
↪py https://download.openmmlab.com/mmediting/restorers/rdn/rdn_x4c64b16_g1_1000k_div2k_
↪20210419-3577d44f.pth

## single-gpu test
python tools/test.py configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/rdn/rdn_x4c64b16_g1_1000k_div2k_20210419-3577d44f.pth

## multi-gpu test
./tools/dist_test.sh configs/rdn/rdn_x4c64b16_1xb16-1000k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/rdn/rdn_x4c64b16_g1_1000k_div2k_20210419-3577d44f.
↪pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{zhang2018residual,
  title={Residual dense network for image super-resolution},
  author={Zhang, Yulun and Tian, Yapeng and Kong, Yu and Zhong, Bineng and Fu, Yun},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={2472--2481},
  year={2018}
}
```

1.34.9 ESRGAN (ECCVW'2018)

ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks

Task: Image Super-Resolution

Abstract

The Super-Resolution Generative Adversarial Network (SRGAN) is a seminal work that is capable of generating realistic textures during single image super-resolution. However, the hallucinated details are often accompanied with unpleasant artifacts. To further enhance the visual quality, we thoroughly study three key components of SRGAN - network architecture, adversarial loss and perceptual loss, and improve each of them to derive an Enhanced SRGAN (ESRGAN). In particular, we introduce the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit. Moreover, we borrow the idea from relativistic GAN to let the discriminator predict relative realness instead of the absolute value. Finally, we improve the perceptual loss by using the features before activation, which could provide stronger supervision for brightness consistency and texture recovery. Benefiting from these improvements, the proposed ESRGAN achieves consistently better visual quality with more realistic and natural textures than SRGAN and won the first place in the PIRM2018-SR Challenge.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/esrgan/esrgan_x4c64b23g32_1xb16-
↪400k_div2k.py

## single-gpu train
python tools/train.py configs/esrgan/esrgan_x4c64b23g32_1xb16-400k_div2k.py
```

(continues on next page)

(continued from previous page)

```
## multi-gpu train
./tools/dist_train.sh configs/esrgan/esrgan_x4c64b23g32_1xb16-400k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/esrgan/esrgan_x4c64b23g32_1xb16-
↪400k_div2k.py https://download.openmmlab.com/mmediting/restorers/esrgan/esrgan_
↪x4c64b23g32_1x16_400k_div2k_20200508-f8ccaf3b.pth

## single-gpu test
python tools/test.py configs/esrgan/esrgan_x4c64b23g32_1xb16-400k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/esrgan/esrgan_x4c64b23g32_1x16_400k_div2k_
↪20200508-f8ccaf3b.pth

## multi-gpu test
./tools/dist_test.sh configs/esrgan/esrgan_x4c64b23g32_1xb16-400k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/esrgan/esrgan_x4c64b23g32_1x16_400k_div2k_
↪20200508-f8ccaf3b.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{wang2018esrgan,
  title={Esrgan: Enhanced super-resolution generative adversarial networks},
  author={Wang, Xintao and Yu, Ke and Wu, Shixiang and Gu, Jinjin and Liu, Yihao and␣
↪Dong, Chao and Qiao, Yu and Change Loy, Chen},
  booktitle={Proceedings of the European Conference on Computer Vision Workshops(ECCVW)},
  pages={0--0},
  year={2018}
}
```

1.34.10 EDSR (CVPR'2017)

> Enhanced Deep Residual Networks for Single Image Super-Resolution
>
> **Task**: Image Super-Resolution

Abstract

Recent research on super-resolution has progressed with the development of deep convolutional neural networks (DCNN). In particular, residual learning techniques exhibit improved performance. In this paper, we develop an enhanced deep super-resolution network (EDSR) with performance exceeding those of current state-of-the-art SR methods. The significant performance improvement of our model is due to optimization by removing unnecessary modules in conventional residual networks. The performance is further improved by expanding the model size while we stabilize the training procedure. We also propose a new multi-scale deep super-resolution system (MDSR) and training method, which can reconstruct high-resolution images of different upscaling factors in a single model. The proposed methods show superior performance over the state-of-the-art methods on benchmark datasets and prove its excellence by winning the NTIRE2017 Super-Resolution Challenge.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM`
.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/edsr/edsr_x4c64b16_1xb16-300k_
↪div2k.py

## single-gpu train
python tools/train.py configs/edsr/edsr_x4c64b16_1xb16-300k_div2k.py

## multi-gpu train
./tools/dist_train.sh configs/edsr/edsr_x4c64b16_1xb16-300k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/edsr/edsr_x4c64b16_1xb16-300k_div2k.
↪py https://download.openmmlab.com/mmediting/restorers/edsr/edsr_x4c64b16_1x16_300k_
↪div2k_20200608-3c2af8a3.pth

## single-gpu test
python tools/test.py configs/edsr/edsr_x4c64b16_1xb16-300k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/edsr/edsr_x4c64b16_1x16_300k_div2k_20200608-3c2af8a3.
↪pth

## multi-gpu test
./tools/dist_test.sh configs/edsr/edsr_x4c64b16_1xb16-300k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/edsr/edsr_x4c64b16_1x16_300k_div2k_20200608-3c2af8a3.
↪pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{lim2017enhanced,
  title={Enhanced deep residual networks for single image super-resolution},
  author={Lim, Bee and Son, Sanghyun and Kim, Heewon and Nah, Seungjun and Mu Lee, Kyoung},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition workshops},
  pages={136--144},
  year={2017}
}
```

1.34.11 SRGAN (CVPR'2016)

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network

Task: Image Super-Resolution

Abstract

Despite the breakthroughs in accuracy and speed of single image super-resolution using faster and deeper convolutional neural networks, one central problem remains largely unsolved: how do we recover the finer texture details when we super-resolve at large upscaling factors? The behavior of optimization-based super-resolution methods is principally driven by the choice of the objective function. Recent work has largely focused on minimizing the mean squared reconstruction error. The resulting estimates have high peak signal-to-noise ratios, but they are often lacking high-frequency details and are perceptually unsatisfying in the sense that they fail to match the fidelity expected at the higher resolution. In this paper, we present SRGAN, a generative adversarial network (GAN) for image super-resolution (SR). To our knowledge, it is the first framework capable of inferring photo-realistic natural images for 4x upscaling factors. To achieve this, we propose a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes our solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, we use a content loss motivated by perceptual similarity instead of similarity in pixel space. Our deep residual network is able to recover photo-realistic textures from heavily downsampled images on public benchmarks. An extensive mean-opinion-score (MOS) test shows hugely significant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than to those obtained with any state-of-the-art method.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation.

The metrics are `PSNR / SSIM` .

| Model |Dataset| PSNR | SSIM | Training Resources | Download | | :———-: | :———: | :———: | :———-: | :———: | :———: | :————————-: | | msrresnet_x4c64b16_1x16_300k_div2k | Set5|30.2252 | 0.8491 | 1 | model | log | | msrresnet_x4c64b16_1x16_300k_div2k | Set14|26.7762 | 0.7369 | 1 | model | log | | msrresnet_x4c64b16_1x16_300k_div2k | DIV2K|28.9748 | 0.8178 | 1 | model | log | | srgan_x4c64b16_1x16_1000k_div2k |Set5|27.9499 | 0.7846 |1 | model | log | | srgan_x4c64b16_1x16_1000k_div2k |Set14|24.7383 |0.6491 | 1 | model | log | | srgan_x4c64b16_1x16_1000k_div2k |DIV2K|26.5697 |0.7365 | 1 | model | log |

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/srgan_resnet/srgan_x4c64b16_1xb16-
↪1000k_div2k.py

## single-gpu train
python tools/train.py configs/srgan_resnet/srgan_x4c64b16_1xb16-1000k_div2k.py

## multi-gpu train
./tools/dist_train.sh configs/srgan_resnet/srgan_x4c64b16_1xb16-1000k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/srgan_resnet/srgan_x4c64b16_1xb16-
↪1000k_div2k.py https://download.openmmlab.com/mmediting/restorers/srresnet_srgan/srgan_
↪x4c64b16_1x16_1000k_div2k_20200606-a1f0810e.pth

## single-gpu test
python tools/test.py configs/srgan_resnet/srgan_x4c64b16_1xb16-1000k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/srresnet_srgan/srgan_x4c64b16_1x16_1000k_
↪div2k_20200606-a1f0810e.pth

## multi-gpu test
./tools/dist_test.sh configs/srgan_resnet/srgan_x4c64b16_1xb16-1000k_div2k.py https://
↪download.openmmlab.com/mmediting/restorers/srresnet_srgan/srgan_x4c64b16_1x16_1000k_
↪div2k_20200606-a1f0810e.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{ledig2016photo,
  title={Photo-realistic single image super-resolution using a generative adversarial␣
↪network},
  author={Ledig, Christian and Theis, Lucas and Husz{\'a}r, Ferenc and Caballero, Jose␣
↪and Cunningham, Andrew and Acosta, Alejandro and Aitken, Andrew and Tejani, Alykhan␣
↪and Totz, Johannes and Wang, Zehan},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition workshops},
  year={2016}
}
```

1.34.12 SRCNN (TPAMI'2015)

Image Super-Resolution Using Deep Convolutional Networks

Task: Image Super-Resolution

Abstract

We propose a deep learning method for single image super-resolution (SR). Our method directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one. We further show that traditional sparse-coding-based SR methods can also be viewed as a deep convolutional network. But unlike traditional methods that handle each component separately, our method jointly optimizes all layers. Our deep CNN has a lightweight structure, yet demonstrates state-of-the-art restoration quality, and achieves fast speed for practical on-line usage. We explore different network structures and parameter settings to achieve trade-offs between performance and speed. Moreover, we extend our network to cope with three color channels simultaneously, and show better overall reconstruction quality.

Results and models

Evaluated on RGB channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM`.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/srcnn/srcnn_x4k915_1xb16-1000k_
↪div2k.py

## single-gpu train
python tools/train.py configs/srcnn/srcnn_x4k915_1xb16-1000k_div2k.py

## multi-gpu train
./tools/dist_train.sh configs/srcnn/srcnn_x4k915_1xb16-1000k_div2k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/srcnn/srcnn_x4k915_1xb16-1000k_
↪div2k.py https://download.openmmlab.com/mmediting/restorers/srcnn/srcnn_x4k915_1x16_
↪1000k_div2k_20200608-4186f232.pth

## single-gpu test
python tools/test.py configs/srcnn/srcnn_x4k915_1xb16-1000k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/srcnn/srcnn_x4k915_1x16_1000k_div2k_20200608-
↪4186f232.pth
```

(continues on next page)

(continued from previous page)

```
## multi-gpu test
./tools/dist_test.sh configs/srcnn/srcnn_x4k915_1xb16-1000k_div2k.py https://download.
↪openmmlab.com/mmediting/restorers/srcnn/srcnn_x4k915_1x16_1000k_div2k_20200608-
↪4186f232.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{dong2015image,
  title={Image super-resolution using deep convolutional networks},
  author={Dong, Chao and Loy, Chen Change and He, Kaiming and Tang, Xiaoou},
  journal={IEEE transactions on pattern analysis and machine intelligence},
  volume={38},
  number={2},
  pages={295--307},
  year={2015},
  publisher={IEEE}
}
```

1.35 diffusers pipeline

1.35.1 Summary

- Number of checkpoints: 0
- Number of configs: 1
- Number of papers: 1
  - ALGORITHM: 1

1.35.2 Diffusers Pipeline (2023)

> Diffusers Pipeline
>
> **Task**: Diffusers Pipeline

Abstract

For the convenience of our community users, this inferencer supports using the pipelines from diffusers for inference to compare the results with the algorithms supported within our algorithm library.

Configs

Quick Start

sd_xl_pipeline

To run stable diffusion XL with mmagic inference API, follow these codes:

```
from mmagic.apis import MMagicInferencer

## Create a MMEdit instance and infer
editor = MMagicInferencer(model_name='diffusers_pipeline')
text_prompts = 'Japanese anime style, girl, beautiful, cute, colorful, best quality,␣
↪extremely detailed'
negative_prompt = 'bad face, bad hands'
result_out_dir = 'sd_xl_japanese.png'
editor.infer(text=text_prompts,
             negative_prompt=negative_prompt,
             result_out_dir=result_out_dir)
```

You will get this picture:

Citation

```
@misc{von-platen-etal-2022-diffusers,
  author = {Patrick von Platen and Suraj Patil and Anton Lozhkov and Pedro Cuenca and␣
↪Nathan Lambert and Kashif Rasul and Mishig Davaadorj and Thomas Wolf},
  title = {Diffusers: State-of-the-art diffusion models},
  year = {2022},
  publisher = {GitHub},
  journal = {GitHub repository},
  howpublished = {\url{https://github.com/huggingface/diffusers}}
}
```

1.36 matting

1.36.1 Summary

- Number of checkpoints: 9
- Number of configs: 10
- Number of papers: 3
  - ALGORITHM: 3

1.36.2 GCA (AAAI'2020)

Natural Image Matting via Guided Contextual Attention

Task: Matting

Abstract

Over the last few years, deep learning based approaches have achieved outstanding improvements in natural image matting. Many of these methods can generate visually plausible alpha estimations, but typically yield blurry structures or textures in the semitransparent area. This is due to the local ambiguity of transparent objects. One possible solution is to leverage the far-surrounding information to estimate the local opacity. Traditional affinity-based methods often suffer from the high computational complexity, which are not suitable for high resolution alpha estimation. Inspired by affinity-based method and the successes of contextual attention in inpainting, we develop a novel end-to-end approach for natural image matting with a guided contextual attention module, which is specifically designed for image matting. Guided contextual attention module directly propagates high-level opacity information globally based on the learned low-level affinity. The proposed method can mimic information flow of affinity-based methods and utilize rich features learned by deep neural networks simultaneously. Experiment results on Composition-1k testing set and this http URL benchmark dataset demonstrate that our method outperforms state-of-the-art approaches in natural image matting.

Results and models

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/gca/gca_r34_4xb10-200k_comp1k.py

## single-gpu train
python tools/train.py configs/gca/gca_r34_4xb10-200k_comp1k.py

## multi-gpu train
./tools/dist_train.sh configs/gca/gca_r34_4xb10-200k_comp1k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/gca/gca_r34_4xb10-200k_comp1k.py␣
↪https://download.openmmlab.com/mmediting/mattors/gca/gca_r34_4x10_200k_comp1k_SAD-33.
↪38_20220615-65595f39.pth

## single-gpu test
python tools/test.py configs/gca/gca_r34_4xb10-200k_comp1k.py https://download.openmmlab.
↪com/mmediting/mattors/gca/gca_r34_4x10_200k_comp1k_SAD-33.38_20220615-65595f39.pth

## multi-gpu test
./tools/dist_test.sh configs/gca/gca_r34_4xb10-200k_comp1k.py https://download.openmmlab.
↪com/mmediting/mattors/gca/gca_r34_4x10_200k_comp1k_SAD-33.38_20220615-65595f39.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{li2020natural,
  title={Natural Image Matting via Guided Contextual Attention},
  author={Li, Yaoyi and Lu, Hongtao},
  booktitle={Association for the Advancement of Artificial Intelligence (AAAI)},
  year={2020}
}
```

1.36.3 IndexNet (ICCV'2019)

Indices Matter: Learning to Index for Deep Image Matting

Task: Matting

Abstract

We show that existing upsampling operators can be unified with the notion of the index function. This notion is inspired by an observation in the decoding process of deep image matting where indices-guided unpooling can recover boundary details much better than other upsampling operators such as bilinear interpolation. By looking at the indices as a function of the feature map, we introduce the concept of learning to index, and present a novel index-guided encoder-decoder framework where indices are self-learned adaptively from data and are used to guide the pooling and upsampling operators, without the need of supervision. At the core of this framework is a flexible network module, termed IndexNet, which dynamically predicts indices given an input. Due to its flexibility, IndexNet can be used as a plug-in applying to any off-the-shelf convolutional networks that have coupled downsampling and upsampling stages.

Results and models

The performance of training (best performance) with different random seeds diverges in a large range. You may need to run several experiments for each setting to obtain the above performance.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/indexnet/indexnet_mobv2_1xb16-78k_
↪comp1k.py

## single-gpu train
python tools/train.py configs/indexnet/indexnet_mobv2_1xb16-78k_comp1k.py

## multi-gpu train
./tools/dist_train.sh configs/indexnet/indexnet_mobv2_1xb16-78k_comp1k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/indexnet/indexnet_mobv2_1xb16-78k_
↪comp1k.py https://download.openmmlab.com/mmediting/mattors/indexnet/indexnet_mobv2_
↪1x16_78k_comp1k_SAD-45.6_20200618_173817-26dd258d.pth

## single-gpu test
python tools/test.py configs/indexnet/indexnet_mobv2_1xb16-78k_comp1k.py https://
↪download.openmmlab.com/mmediting/mattors/indexnet/indexnet_mobv2_1x16_78k_comp1k_SAD-
↪45.6_20200618_173817-26dd258d.pth

## multi-gpu test
./tools/dist_test.sh configs/indexnet/indexnet_mobv2_1xb16-78k_comp1k.py https://
↪download.openmmlab.com/mmediting/mattors/indexnet/indexnet_mobv2_1x16_78k_comp1k_SAD-
↪45.6_20200618_173817-26dd258d.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{hao2019indexnet,
  title={Indices Matter: Learning to Index for Deep Image Matting},
  author={Lu, Hao and Dai, Yutong and Shen, Chunhua and Xu, Songcen},
  booktitle={Proc. IEEE/CVF International Conference on Computer Vision (ICCV)},
  year={2019}
}
```

1.36.4 DIM (CVPR'2017)

> Deep Image Matting
>
> **Task**: Matting

Abstract

Image matting is a fundamental computer vision problem and has many applications. Previous algorithms have poor performance when an image has similar foreground and background colors or complicated textures. The main reasons are prior methods 1) only use low-level features and 2) lack high-level context. In this paper, we propose a novel deep learning based algorithm that can tackle both these problems. Our deep model has two parts. The first part is a deep convolutional encoder-decoder network that takes an image and the corresponding trimap as inputs and predict the alpha matte of the image. The second part is a small convolutional network that refines the alpha matte predictions of the first network to have more accurate alpha values and sharper edges. In addition, we also create a large-scale image matting dataset including 49300 training images and 1000 testing images. We evaluate our algorithm on the image matting benchmark, our testing set, and a wide variety of real images. Experimental results clearly demonstrate the superiority of our algorithm over previous methods.

Results and models

NOTE

- stage1: train the encoder-decoder part without the refinement part.
- stage2: fix the encoder-decoder part and train the refinement part.
- stage3: fine-tune the whole network.

  The performance of the model is not stable during the training. Thus, the reported performance is not from the last checkpoint. Instead, it is the best performance of all validations during training.

  The performance of training (best performance) with different random seeds diverges in a large range. You may need to run several experiments for each setting to obtain the above performance.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

DIM is trained with three stages.

Stage 1: train the encoder-decoder part without the refinement part.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/dim/dim_stage1-v16_1xb1-1000k_
↪comp1k.py

## single-gpu train
python tools/train.py configs/dim/dim_stage1-v16_1xb1-1000k_comp1k.py

## multi-gpu train
./tools/dist_train.sh configs/dim/dim_stage1-v16_1xb1-1000k_comp1k.py 8
```

Stage 2: fix the encoder-decoder part and train the refinement part.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/dim/dim_stage2-v16-pln_1xb1-1000k_
↪comp1k.py

## single-gpu train
python tools/train.py configs/dim/dim_stage2-v16-pln_1xb1-1000k_comp1k.py

## multi-gpu train
./tools/dist_train.sh configs/dim/dim_stage2-v16-pln_1xb1-1000k_comp1k.py 8
```

Stage 3: fine-tune the whole network.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/dim/dim_stage3-v16-pln_1xb1-1000k_
↪comp1k.py

## single-gpu train
python tools/train.py configs/dim/dim_stage3-v16-pln_1xb1-1000k_comp1k.py
```

(continues on next page)

(continued from previous page)

```
## multi-gpu train
./tools/dist_train.sh configs/dim/dim_stage3-v16-pln_1xb1-1000k_comp1k.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/dim/dim_stage3-v16-pln_1xb1-1000k_
↪comp1k.py https://download.openmmlab.com/mmediting/mattors/dim/dim_stage3_v16_pln_1x1_
↪1000k_comp1k_SAD-50.6_20200609_111851-647f24b6.pth

## single-gpu test
python tools/test.py configs/dim/dim_stage3-v16-pln_1xb1-1000k_comp1k.py https://
↪download.openmmlab.com/mmediting/mattors/dim/dim_stage3_v16_pln_1x1_1000k_comp1k_SAD-
↪50.6_20200609_111851-647f24b6.pth

## multi-gpu test
./tools/dist_test.sh configs/dim/dim_stage3-v16-pln_1xb1-1000k_comp1k.py https://
↪download.openmmlab.com/mmediting/mattors/dim/dim_stage3_v16_pln_1x1_1000k_comp1k_SAD-
↪50.6_20200609_111851-647f24b6.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{xu2017deep,
  title={Deep image matting},
  author={Xu, Ning and Price, Brian and Cohen, Scott and Huang, Thomas},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern␣
↪Recognition},
  pages={2970--2979},
  year={2017}
}
```

1.37 draggan

1.37.1 Summary

- Number of checkpoints: 3
- Number of configs: 3
- Number of papers: 1
  - ALGORITHM: 1

1.37.2 DragGAN (SIGGRAPH'2023)

Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold

Task: DragGAN

Abstract

Synthesizing visual content that meets users' needs often requires flexible and precise controllability of the pose, shape, expression, and layout of the generated objects. Existing approaches gain controllability of generative adversarial networks (GANs) via manually annotated training data or a prior 3D model, which often lack flexibility, precision, and generality. In this work, we study a powerful yet much less explored way of controlling GANs, that is, to "drag" any points of the image to precisely reach target points in a user-interactive manner, as shown in Fig.1. To achieve this, we propose DragGAN, which consists of two main components: 1) a feature-based motion supervision that drives the handle point to move towards the target position, and 2) a new point tracking approach that leverages the discriminative generator features to keep localizing the position of the handle points. Through DragGAN, anyone can deform an image with precise control over where pixels go, thus manipulating the pose, shape, expression,and layout of diverse categories such as animals, cars, humans, landscapes, etc. As these manipulations are performed on the learned generative image manifold of a GAN, they tend to produce realistic outputs even for challenging scenarios such as hallucinating occluded content and deforming shapes that consistently follow the object's rigidity. Both qualitative and quantitative comparisons demonstrate the advantage of DragGAN over prior approaches in the tasks of image manipulation and point tracking. We also showcase the manipulation of real images through GAN inversion.

Results and Models

Demo

To run DragGAN demo, please follow these two steps:

First, put your checkpoint path in `./checkpoints`, *e.g.* `./checkpoints/stylegan2_lions_512_pytorch_mmagic.pth`. To be specific,

```
mkdir checkpoints
cd checkpoints
wget -O stylegan2_lions_512_pytorch_mmagic.pth https://download.openxlab.org.cn/models/
↪qsun1/DragGAN-StyleGAN2-checkpoint/weight//StyleGAN2-Lions-internet
```

Then, try on the script:

```
python demo/gradio_draggan.py
```

Citation

```
@inproceedings{pan2023drag,
  title={Drag your gan: Interactive point-based manipulation on the generative image␣
↪manifold},
  author={Pan, Xingang and Tewari, Ayush and Leimk{\"u}hler, Thomas and Liu, Lingjie and␣
↪Meka, Abhimitra and Theobalt, Christian},
  booktitle={ACM SIGGRAPH 2023 Conference Proceedings},
  pages={1--11},
  year={2023}
}
```

1.38 3d-aware generation

1.38.1 Summary

- Number of checkpoints: 3
- Number of configs: 3
- Number of papers: 1
  - ALGORITHM: 1

1.38.2 EG3D (CVPR'2022)

Efficient geometry-aware 3D generative adversarial networks

Task: 3D-aware Generation

Abstract

Unsupervised generation of high-quality multi-view-consistent images and 3D shapes using only collections of single-view 2D photographs has been a long-standing challenge. Existing 3D GANs are either compute-intensive or make approximations that are not 3D-consistent; the former limits quality and resolution of the generated images and the latter adversely affects multi-view consistency and shape quality. In this work, we improve the computational efficiency and image quality of 3D GANs without overly relying on these approximations. We introduce an expressive hybrid explicit-implicit network architecture that, together with other design choices, synthesizes not only high-resolution multi-view-consistent images in real time but also produces high-quality 3D geometry. By decoupling feature generation and neural rendering, our framework is able to leverage state-of-the-art 2D CNN generators, such as StyleGAN2, and inherit their efficiency and expressiveness. We demonstrate state-of-the-art 3D-aware synthesis with FFHQ and AFHQ Cats, among other experiments.

Results and Models

- `FID50k-Camera` denotes image generated with random sampled camera position.
- `FID50k` denotes image generated with camera position randomly sampled from the original dataset.

Influence of FP16

All metrics are evaluated under FP32, and it's hard to determine how they will change if we use FP16. For example, if we use FP16 at the super resolution module in FFHQ model, the output images will be slightly blurrier than the ones generated under FP32, but FID (**4.03**) will be better than FP32 ones.

About generate images and videos with High-Level API

You can use the following command to generate sequence images with continuous changed camera position as input.

```
python demo/mmagic_inference_demo.py --model-name eg3d \
    --model-config configs/eg3d/eg3d_cvt-official-rgb_afhq-512x512.py \
    --model-ckpt https://download.openmmlab.com/mmediting/eg3d/eg3d_cvt-official-rgb_
↪afhq-512x512-ca1dd7c9.pth \
    --result-out-dir eg3d_output \  ## save images and videos to `eg3d_output`
    --interpolation camera \  ## interpolation camera position only
    --num-images 100  ## generate 100 images during interpolation
```

The the following video will be saved to `eg3d_output`.

To interpolate the camera position and style code at the same time, you can use the following command.

```
python demo/mmagic_inference_demo.py --model-name eg3d \
    --model-config configs/eg3d/eg3d_cvt-official-rgb_ffhq-512x512.py \
    --model-ckpt https://download.openmmlab.com/mmediting/eg3d/eg3d_cvt-official-rgb_
↪ffhq-512x512-5a0ddcb6.pth \
    --result-out-dir eg3d_output \  ## save images and videos to `eg3d_output`
    --interpolation both \  ## interpolation camera and conditioning both
    --num-images 100  ## generate 100 images during interpolation
    --seed 233  ## set random seed as 233
```

If you only want to save video of depth map, you can use the following command:

```
python demo/mmagic_inference_demo.py --model-name eg3d \
    --model-config configs/eg3d/eg3d_cvt-official-rgb_shapenet-128x128.py \
    --model-ckpt https://download.openmmlab.com/mmediting/eg3d/eg3d_cvt-official-rgb_
↪shapenet-128x128-85757f4d.pth \
    --result-out-dir eg3d_output \  ## save images and videos to `eg3d_output`
    --interpolation camera \  ## interpolation camera position only
    --num-images 100 \  ## generate 100 images during interpolation
    --vis-mode depth  ## only visualize depth image
```

How to prepare dataset

You should prepare your dataset follow the official repo. Then preprocess the `dataset.json` with the following script:

```
import json
from argparse import ArgumentParser

from mmengine.fileio.io import load


def main():

    parser = ArgumentParser()
    parser.add_argument(
        'in-anno', type=str, help='Path to the official annotation file.')
    parser.add_argument(
        'out-anno', type=str, help='Path to MMagicing\'s annotation file.')
```

(continues on next page)

(continued from previous page)

```
    args = parser.parse_args()

    anno = load(args.in_anno)
    label = anno['labels']

    anno_dict = {}
    for line in label:
        name, label = line
        anno_dict[name] = label

    with open(args.out_anno, 'w') as file:
        json.dump(anno_dict, file)


if __name__ == '__main__':
    main()
```

Citation

```
@InProceedings{Chan_2022_CVPR,
    author    = {Chan, Eric R. and Lin, Connor Z. and Chan, Matthew A. and Nagano, Koki␣
↪and Pan, Boxiao and De Mello, Shalini and Gallo, Orazio and Guibas, Leonidas J. and␣
↪Tremblay, Jonathan and Khamis, Sameh and Karras, Tero and Wetzstein, Gordon},
    title     = {Efficient Geometry-Aware 3D Generative Adversarial Networks},
    booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern␣
↪Recognition (CVPR)},
    month     = {June},
    year      = {2022},
    pages     = {16123-16133}
}
```

1.39 image generation

1.39.1 Summary

- Number of checkpoints: 6
- Number of configs: 6
- Number of papers: 1
  - ALGORITHM: 1

1.39.2 Guided Diffusion (NeurIPS'2021)

Diffusion Models Beat GANs on Image Synthesis

Task: Image Generation

Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128x128, 4.59 on ImageNet 256x256, and 7.72 on ImageNet 512x512, and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet 256x256 and 3.85 on ImageNet 512x512.

Results and models

ImageNet

Quick Start

infer

You can run adm as follows:

```
from mmengine import Config, MODELS
from mmengine.registry import init_default_scope
from torchvision.utils import save_image

init_default_scope('mmagic')

## sampling without classifier guidance, CGS=1.0
config = 'configs/guided_diffusion/adm-g_ddim25_8xb32_imagenet-64x64.py'
ckpt_path = 'https://download.openmmlab.com/mmediting/guided_diffusion/adm-g_8xb32_
↪imagenet-64x64-2c0fbeda.pth'  ## noqa

model_cfg = Config.fromfile(config).model
model_cfg.pretrained_cfgs = dict(unet=dict(ckpt_path=ckpt_path, prefix='unet'),
                                 classifier=dict(ckpt_path=ckpt_path, prefix='classifier
↪'))
model = MODELS.build(model_cfg).cuda().eval()

samples = model.infer(
            init_image=None,
            batch_size=4,
            num_inference_steps=25,
            labels=333,
            classifier_scale=1.0,
            show_progress=True)['samples']
```

(continues on next page)

(continued from previous page)

```
## sampling without classifier guidance
config = 'configs/guided_diffusion/adm_ddim250_8xb32_imagenet-64x64.py'
ckpt_path = 'https://download.openmmlab.com/mmediting/guided_diffusion/adm-u-cvt-rgb_
↪8xb32_imagenet-64x64-7ff0080b.pth'  ## noqa

model_cfg = Config.fromfile(config).model
model_cfg.pretrained_cfgs = dict(unet=dict(ckpt_path=ckpt_path, prefix='unet'))
model = MODELS.build(model_cfg).cuda().eval()

samples = model.infer(
            init_image=None,
            batch_size=4,
            num_inference_steps=250,
            labels=None,
            classifier_scale=0.0,
            show_progress=True)['samples']
```

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/guided_diffusion/adm-u_ddim250_
↪8xb32_imagenet-64x64.py https://download.openmmlab.com/mmgen/guided_diffusion/adm-u-
↪cvt-rgb_8xb32_imagenet-64x64-7ff0080b.pth

## single-gpu test
python tools/test.py configs/guided_diffusion/adm-u_ddim250_8xb32_imagenet-64x64.py␣
↪https://download.openmmlab.com/mmgen/guided_diffusion/adm-u-cvt-rgb_8xb32_imagenet-
↪64x64-7ff0080b.pth

## multi-gpu test
./tools/dist_test.sh configs/guided_diffusion/adm-u_ddim250_8xb32_imagenet-64x64.py␣
↪https://download.openmmlab.com/mmgen/guided_diffusion/adm-u-cvt-rgb_8xb32_imagenet-
↪64x64-7ff0080b.pth 8
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@article{PrafullaDhariwal2021DiffusionMB,
  title={Diffusion Models Beat GANs on Image Synthesis},
  author={Prafulla Dhariwal and Alex Nichol},
  journal={arXiv: Learning},
  year={2021}
}
```

1.40 colorization

1.40.1 Summary

- Number of checkpoints: 1
- Number of configs: 1
- Number of papers: 1
  - ALGORITHM: 1

1.40.2 Instance-aware Image Colorization (CVPR'2020)

Instance-Aware Image Colorization

Task: Colorization

Abstract

Image colorization is inherently an ill-posed problem with multi-modal uncertainty. Previous methods leverage the deep neural network to map input grayscale images to plausible color outputs directly. Although these learning-based methods have shown impressive performance, they usually fail on the input images that contain multiple objects. The leading cause is that existing models perform learning and colorization on the entire image. In the absence of a clear figure-ground separation, these models cannot effectively locate and learn meaningful object-level semantics. In this paper, we propose a method for achieving instance-aware colorization. Our network architecture leverages an off-the-shelf object detector to obtain cropped object images and uses an instance colorization network to extract object-level features. We use a similar network to extract the full-image features and apply a fusion module to full object-level and image-level features to predict the final colors. Both colorization networks and fusion modules are learned from a large-scale dataset. Experimental results show that our work outperforms existing methods on different quality metrics and achieves state-of-the-art performance on image colorization.

Results and models

Quick Start

You can use the following commands to colorize an image.

```
python demo/mmagic_inference_demo.py --model-name inst_colorization --img input.jpg --
↪result-out-dir output.png
```

For more demos, you can refer to Tutorial 3: inference with pre-trained models.

```
@inproceedings{Su-CVPR-2020,
  author = {Su, Jheng-Wei and Chu, Hung-Kuo and Huang, Jia-Bin},
  title = {Instance-aware Image Colorization},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2020}
}
```

1.41 image restoration

1.41.1 Summary

- Number of checkpoints: 2
- Number of configs: 2
- Number of papers: 1
  - ALGORITHM: 1

1.41.2 NAFNet (ECCV'2022)

Simple Baselines for Image Restoration

Task: Image Restoration

Abstract

Although there have been significant advances in the field of image restoration recently, the system complexity of the state-of-the-art (SOTA) methods is increasing as well, which may hinder the convenient analysis and comparison of methods. In this paper, we propose a simple baseline that exceeds the SOTA methods and is computationally efficient. To further simplify the baseline, we reveal that the nonlinear activation functions, e.g. Sigmoid, ReLU, GELU, Softmax, etc. are not necessary: they could be replaced by multiplication or removed. Thus, we derive a Nonlinear Activation Free Network, namely NAFNet, from the baseline. SOTA results are achieved on various challenging benchmarks, e.g. 33.69 dB PSNR on GoPro (for image deblurring), exceeding the previous SOTA 0.38 dB with only 8.4% of its computational costs; 40.30 dB PSNR on SIDD (for image denoising), exceeding the previous SOTA 0.28 dB with less than half of its computational costs.

Results and models

Note:

- a(b) where a denotes the value run by MMagic, b denotes the value copied from the original paper.
- PSNR is evaluated on RGB channels.
- SSIM is evaluated by averaging SSIMs on RGB channels, however the original paper uses the 3D SSIM kernel.

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/nafnet/nafnet_c64eb2248mb12db2222_
↪8xb8-lr1e-3-400k_sidd.py

## single-gpu train
python tools/train.py configs/nafnet/nafnet_c64eb2248mb12db2222_8xb8-lr1e-3-400k_sidd.py
```

(continues on next page)

(continued from previous page)

```
## multi-gpu train
./tools/dist_train.sh configs/nafnet/nafnet_c64eb2248mb12db2222_8xb8-lr1e-3-400k_sidd.py␣
↪8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/nafnet/nafnet_c64eb2248mb12db2222_
↪8xb8-lr1e-3-400k_sidd.py /path/to/checkpoint

## single-gpu test
python tools/test.py configs/nafnet/nafnet_c64eb2248mb12db2222_8xb8-lr1e-3-400k_sidd.py /
↪path/to/checkpoint

## multi-gpu test
./tools/dist_test.sh configs/nafnet/nafnet_c64eb2248mb12db2222_8xb8-lr1e-3-400k_sidd.py /
↪path/to/checkpoint 8
```

Pretrained checkpoints will come soon.

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```bibtex
@article{chen2022simple,
  title={Simple Baselines for Image Restoration},
  author={Chen, Liangyu and Chu, Xiaojie and Zhang, Xiangyu and Sun, Jian},
  journal={arXiv preprint arXiv:2204.04676},
  year={2022}
}
```

1.42 deraining

1.42.1 Summary

- Number of checkpoints: 1
- Number of configs: 1
- Number of papers: 1
  - ALGORITHM: 1

1.42.2 Restormer (CVPR'2022)

Restormer: Efficient Transformer for High-Resolution Image Restoration

Task: Denoising, Deblurring, Deraining

Abstract

Since convolutional neural networks (CNNs) perform well at learning generalizable image priors from large-scale data, these models have been extensively applied to image restoration and related tasks. Recently, another class of neural architectures, Transformers, have shown significant performance gains on natural language and high-level vision tasks. While the Transformer model mitigates the shortcomings of CNNs (i.e., limited receptive field and inadaptability to input content), its computational complexity grows quadratically with the spatial resolution, therefore making it infeasible to apply to most image restoration tasks involving high-resolution images. In this work, we propose an efficient Transformer model by making several key designs in the building blocks (multi-head attention and feed-forward network) such that it can capture long-range pixel interactions, while still remaining applicable to large images. Our model, named Restoration Transformer (Restormer), achieves state-of-the-art results on several image restoration tasks, including image deraining, single-image motion deblurring, defocus deblurring (single-image and dual-pixel data), and image denoising (Gaussian grayscale/color denoising, and real image denoising).

Results and models

Deraining

Evaluated on Y channels. The metrics are `PSNR` / `SSIM` .

Motion Deblurring

Evaluated on RGB channels for GoPro and HIDE, and Y channel for ReakBlur-J and ReakBlur-R. The metrics are `PSNR` / `SSIM` .

Defocus Deblurring

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` / `MAE` / `LPIPS`.

Gaussian Denoising

Test Grayscale Gaussian Noise

Evaluated on grayscale images. The metrics are `PSNR` / `SSIM` .

training a separate model for each noise level

| Model | Dataset | Task | σ | PSNR | SSIM | Training Resources | Download | | :————————————————————————————: | :——:|:-:| :———-: | :——: | :—-: | :————————————————————————————-: | | restormer_official_dfwb-gray-sigma15 | Set12 |Denoising| 15 | 34.0182 | 0.9160 | 1 | model | log | | restormer_official_dfwb-gray-sigma15 | BSD68 |Denoising| 15 | 32.4987 | 0.8940 | 1 | model | log | | restormer_official_dfwb-gray-sigma15 | Urban100 |Denoising| 15 | 34.4336 | 0.9419 | 1 | model | log | | restormer_official_dfwb-gray-sigma25 | Set12 |Denoising| 25 | 31.7289 | 0.8811 | 1 | model | log | | restormer_official_dfwb-gray-sigma25 | BSD68 |Denoising| 25 | 30.1613 | 0.8370 | 1 | model | log | | restormer_official_dfwb-gray-sigma25 | Urban100 |Denoising| 25 | 32.1162 | 0.9140

| 1 | model | log | | restormer_official_dfwb-gray-sigma50 | Set12 |Denoising| 50 | 28.6269 | 0.8188 | 1 | model | log | | restormer_official_dfwb-gray-sigma50 | BSD68 |Denoising| 50 | 27.3266 | 0.7434 | 1 | model | log | | restormer_official_dfwb-gray-sigma50 | Urban100 |Denoising| 50 | 28.9636 | 0.8571 | 1 | model | log |

learning a single model to handle various noise levels

Test Color Gaussian Noise

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` . **training a separate model for each noise level**

learning a single model to handle various noise levels

| Model | Dataset |Task| σ | PSNR (RGB) | SSIM (RGB) | Training Resources | Download | | :——————————————————————————: | :——: | :———: | :———: | :———: | :———————————————————————————————: | :——: | | restormer_official_dfwb-color-sigma15| CBSD68 |Denoising| 15 | 34.3422 | 0.9356 | 1 | model | log | | restormer_official_dfwb-color-sigma15| Kodak24 |Denoising| 15 | 35.4544 | 0.9308 | 1 | model | log | | restormer_official_dfwb-color-sigma15| McMaster |Denoising| 15 | 35.5473 | 0.9344 | 1 | model | log | | restormer_official_dfwb-color-sigma15| Urban100 |Denoising| 15 | 35.0754 | 0.9524 | 1 | model | log | | restormer_official_dfwb-color-sigma25| CBSD68 |Denoising| 25 | 31.7391 | 0.8945 | 1 | model | log | | restormer_official_dfwb-color-sigma25| Kodak24 |Denoising| 25 | 33.0380 | 0.8941 | 1 | model | log | | restormer_official_dfwb-color-sigma25| McMaster |Denoising| 25 | 33.3040 | 0.9063 | 1 | model | log | | restormer_official_dfwb-color-sigma25| Urban100 |Denoising| 25 | 32.9165 | 0.9312 | 1 | model | log | | restormer_official_dfwb-color-sigma50| CBSD68 |Denoising| 50 | 28.5582 | 0.8126 | 1 | model | log | | restormer_official_dfwb-color-sigma50| Kodak24 |Denoising| 50 | 30.0074 | 0.8233 | 1 | model | log | | restormer_official_dfwb-color-sigma50| McMaster |Denoising| 50 | 30.2671 | 0.8520 | 1 | model | log | | restormer_official_dfwb-color-sigma50| Urban100 |Denoising| 50 | 30.0172 | 0.8898 | 1 | model | log |

Real Image Denoising

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` .

Quick Start

Train

You can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## Deraining
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_
↪rain13k.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪rain13k-2be7b550.pth

## Motion Deblurring
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_gopro.
↪py https://download.openmmlab.com/mmediting/restormer/restormer_official_gopro-
↪db7363a0.pth

## Defocus Deblurring
## Single
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪dual.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪single-6bc31582.pth
```

(continues on next page)

(continued from previous page)

```
## Dual
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪single.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma15-da74417f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma25-08010841.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma50-ee852dfe.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## Test Color Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma15-012ceb71.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma25-e307f222.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth
```

(continues on next page)

(continued from previous page)

```
## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma50-a991983d.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## single-gpu test
## Deraining
python tools/test.py configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
python tools/test.py configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
## Single
python tools/test.py configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
python tools/test.py configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth


## multi-gpu test
## Deraining
./tools/dist_test.sh configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
./tools/dist_test.sh configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
## Single
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
```

(continues on next page)

(continued from previous page)

```
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth
```

(continues on next page)

(continued from previous page)

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{Zamir2021Restormer,
    title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
    author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat
            and Fahad Shahbaz Khan and Ming-Hsuan Yang},
    booktitle={CVPR},
    year={2022}
}
```

1.43 denoising

1.43.1 Summary

- Number of checkpoints: 3
- Number of configs: 7
- Number of papers: 1
  - ALGORITHM: 1

1.43.2 Restormer (CVPR'2022)

Restormer: Efficient Transformer for High-Resolution Image Restoration

Task: Denoising, Deblurring, Deraining

Abstract

Since convolutional neural networks (CNNs) perform well at learning generalizable image priors from large-scale data, these models have been extensively applied to image restoration and related tasks. Recently, another class of neural architectures, Transformers, have shown significant performance gains on natural language and high-level vision tasks. While the Transformer model mitigates the shortcomings of CNNs (i.e., limited receptive field and inadaptability to input content), its computational complexity grows quadratically with the spatial resolution, therefore making it infeasible to apply to most image restoration tasks involving high-resolution images. In this work, we propose an efficient Transformer model by making several key designs in the building blocks (multi-head attention and feed-forward network) such that it can capture long-range pixel interactions, while still remaining applicable to large images. Our model, named Restoration Transformer (Restormer), achieves state-of-the-art results on several image restoration tasks, including image deraining, single-image motion deblurring, defocus deblurring (single-image and dual-pixel data), and image denoising (Gaussian grayscale/color denoising, and real image denoising).

Results and models

Deraining

Evaluated on Y channels. The metrics are `PSNR` / `SSIM` .

Motion Deblurring

Evaluated on RGB channels for GoPro and HIDE, and Y channel for ReakBlur-J and ReakBlur-R. The metrics are `PSNR` / `SSIM` .

Defocus Deblurring

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` / `MAE` / `LPIPS`.

Gaussian Denoising

Test Grayscale Gaussian Noise

Evaluated on grayscale images. The metrics are `PSNR` / `SSIM` .

training a separate model for each noise level

Model	Dataset	Task	σ	PSNR	SSIM	Training Resources	Download	
restormer_official_dfwb-gray-sigma15	Set12	Denoising	15	34.0182	0.9160	1	model	log
restormer_official_dfwb-gray-sigma15	BSD68	Denoising	15	32.4987	0.8940	1	model	log
restormer_official_dfwb-gray-sigma15	Urban100	Denoising	15	34.4336	0.9419	1	model	log
restormer_official_dfwb-gray-sigma25	Set12	Denoising	25	31.7289	0.8811	1	model	log
restormer_official_dfwb-gray-sigma25	BSD68	Denoising	25	30.1613	0.8370	1	model	log
restormer_official_dfwb-gray-sigma25	Urban100	Denoising	25	32.1162	0.9140	1	model	log
restormer_official_dfwb-gray-sigma50	Set12	Denoising	50	28.6269	0.8188	1	model	log
restormer_official_dfwb-gray-sigma50	BSD68	Denoising	50	27.3266	0.7434	1	model	log
restormer_official_dfwb-gray-sigma50	Urban100	Denoising	50	28.9636	0.8571	1	model	log

learning a single model to handle various noise levels

Test Color Gaussian Noise

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` . **training a separate model for each noise level**

learning a single model to handle various noise levels

Model	Dataset	Task	σ	PSNR (RGB)	SSIM (RGB)	Training Resources	Download	
restormer_official_dfwb-color-sigma15	CBSD68	Denoising	15	34.3422	0.9356	1	model	log
restormer_official_dfwb-color-sigma15	Kodak24	Denoising	15	35.4544	0.9308	1	model	log
restormer_official_dfwb-color-sigma15	McMaster	Denoising	15	35.5473	0.9344	1	model	log
restormer_official_dfwb-color-sigma15	Urban100	Denoising	15	35.0754	0.9524	1	model	log
restormer_official_dfwb-color-sigma25	CBSD68	Denoising	25	31.7391	0.8945	1	model	log
restormer_official_dfwb-color-sigma25	Kodak24	Denoising	25	33.0380	0.8941	1	model	log
restormer_official_dfwb-color-sigma25	McMaster	Denoising	25	33.3040	0.9063	1	model	log
restormer_official_dfwb-color-sigma25	Urban100	Denoising	25	32.9165	0.9312	1	model	log
restormer_official_dfwb-color-sigma50	CBSD68	Denoising	50	28.5582	0.8126	1	model	log
restormer_official_dfwb-color-sigma50	Kodak24	Denoising	50	30.0074	0.8233	1	model	

log | | restormer_official_dfwb-color-sigma50| McMaster |Denoising| 50 | 30.2671 | 0.8520 | 1 | model | log | | restormer_official_dfwb-color-sigma50| Urban100 |Denoising| 50 | 30.0172 | 0.8898 | 1 | model | log |

Real Image Denoising

Evaluated on RGB channels. The metrics are `PSNR` / `SSIM` .

Quick Start

Train

You can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## Deraining
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_
↪rain13k.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪rain13k-2be7b550.pth

## Motion Deblurring
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_gopro.
↪py https://download.openmmlab.com/mmediting/restormer/restormer_official_gopro-
↪db7363a0.pth

## Defocus Deblurring
## Single
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪dual.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪single-6bc31582.pth
## Dual
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dpdd-
↪single.py https://download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-
↪dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma15-da74417f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma25-08010841.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-sigma50-ee852dfe.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪gray-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-gray-blind-5f094bcc.pth

## Test Color Gaussian Noise
## sigma15
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma15-012ceb71.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma15.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## sigma25
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma25-e307f222.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma25.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## sigma50
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-sigma50-a991983d.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/restormer/restormer_official_dfwb-
↪color-sigma50.py https://download.openmmlab.com/mmediting/restormer/restormer_official_
↪dfwb-color-blind-dfd03c9f.pth

## single-gpu test
## Deraining
python tools/test.py configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
python tools/test.py configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
```

(continues on next page)

(continued from previous page)

```
## Single
python tools/test.py configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
python tools/test.py configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth

python tools/test.py configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth
```

(continues on next page)

(continued from previous page)

```
## sigma50
python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

python tools/test.py configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth


## multi-gpu test
## Deraining
./tools/dist_test.sh configs/restormer/restormer_official_rain13k.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_rain13k-2be7b550.pth

## Motion Deblurring
./tools/dist_test.sh configs/restormer/restormer_official_gopro.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_gopro-db7363a0.pth

## Defocus Deblurring
## Single
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-dual.py https://download.
↪openmmlab.com/mmediting/restormer/restormer_official_dpdd-single-6bc31582.pth
## Dual
./tools/dist_test.sh configs/restormer/restormer_official_dpdd-single.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dpdd-dual-52c94c00.pth

## Gaussian Denoising
## Test Grayscale Gaussian Noise
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma15-
↪da74417f.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma25-
↪08010841.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-sigma50-
↪ee852dfe.pth
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-gray-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-gray-blind-5f094bcc.
↪pth

## Test Color Gaussian Noise
## sigma15
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma15-
↪012ceb71.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma15.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma25
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma25-
↪e307f222.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma25.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth

## sigma50
./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-sigma50-
↪a991983d.pth

./tools/dist_test.sh configs/restormer/restormer_official_dfwb-color-sigma50.py https://
↪download.openmmlab.com/mmediting/restormer/restormer_official_dfwb-color-blind-
↪dfd03c9f.pth
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{Zamir2021Restormer,
    title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
    author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat
            and Fahad Shahbaz Khan and Ming-Hsuan Yang},
    booktitle={CVPR},
    year={2022}
}
```

1.44 internal learning

1.44.1 Summary

- Number of checkpoints: 3
- Number of configs: 3
- Number of papers: 1
  - ALGORITHM: 1

1.44.2 SinGAN (ICCV'2019)

Singan: Learning a Generative Model from a Single Natural Image

Task: Internal Learning

Abstract

We introduce SinGAN, an unconditional generative model that can be learned from a single natural image. Our model is trained to capture the internal distribution of patches within the image, and is then able to generate high quality, diverse samples that carry the same visual content as the image. SinGAN contains a pyramid of fully convolutional GANs, each responsible for learning the patch distribution at a different scale of the image. This allows generating new samples of arbitrary size and aspect ratio, that have significant variability, yet maintain both the global structure and the fine textures of the training image. In contrast to previous single image GAN schemes, our approach is not limited to texture images, and is not conditional (i.e. it generates samples from noise). User studies confirm that the generated samples are commonly confused to be real images. We illustrate the utility of SinGAN in a wide range of image manipulation tasks.

Results and Models

Notes for using SinGAN

When training SinGAN models, users may obtain the number of scales (stages) in advance via the following commands. This number is important for constructing config file, which is related to the generator, discriminator, the training iterations and so on.

```
>>> from mmgen.datasets.singan_dataset import create_real_pyramid
>>> import mmcv
>>> real = mmcv.imread('real_img_path')
>>> _, _, num_scales = create_real_pyramid(real, min_size=25, max_size=300, scale_factor_
↪init=0.75)
```

When testing SinGAN models, users have to modify the config file to add the `test_cfg`. As shown in `configs/singan/singan_balloons.py`, the only thing you need to do is add the path for `pkl` data. There are some important data containing in the pickle files which you can download from our website.

```
test_cfg = dict(
    _delete_ = True
    pkl_data = 'path to pkl data'
)
```

Citation

```
@inproceedings{shaham2019singan,
  title={Singan: Learning a generative model from a single natural image},
  author={Shaham, Tamar Rott and Dekel, Tali and Michaeli, Tomer},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={4570--4580},
  year={2019},
  url={https://openaccess.thecvf.com/content_ICCV_2019/html/Shaham_SinGAN_Learning_a_
↪Generative_Model_From_a_Single_Natural_Image_ICCV_2019_paper.html},
}
```

1.45 image denoising

1.45.1 Summary

- Number of checkpoints: 6
- Number of configs: 6
- Number of papers: 1
  - ALGORITHM: 1

1.45.2 SwinIR (ICCVW'2021)

SwinIR: Image Restoration Using Swin Transformer

Task: Image Super-Resolution, Image denoising, JPEG compression artifact reduction

Abstract

Image restoration is a long-standing low-level vision problem that aims to restore high-quality images from low-quality images (e.g., downscaled, noisy and compressed images). While state-of-the-art image restoration methods are based on convolutional neural networks, few attempts have been made with Transformers which show impressive performance on high-level vision tasks. In this paper, we propose a strong baseline model SwinIR for image restoration based on the Swin Transformer. SwinIR consists of three parts: shallow feature extraction, deep feature extraction and high-quality image reconstruction. In particular, the deep feature extraction module is composed of several residual Swin Transformer blocks (RSTB), each of which has several Swin Transformer layers together with a residual connection. We conduct experiments on three representative tasks: image super-resolution (including classical, lightweight and real-world image super-resolution), image denoising (including grayscale and color image denoising) and JPEG compression artifact reduction. Experimental results demonstrate that SwinIR outperforms state-of-the-art methods on different tasks by up to 0.14~0.45dB, while the total number of parameters can be reduced by up to 67%.

Results and models

Classical Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Lightweight Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Real-World Image Super-Resolution

Evaluated on Y channels. The metrics are `NIQE` .

Grayscale Image Deoising

Evaluated on grayscale images. The metrics are `PSNR` .

Color Image Deoising

Evaluated on RGB channels. The metrics are `PSNR` .

JPEG Compression Artifact Reduction (grayscale)

Evaluated on grayscale images. The metrics are `PSNR / SSIM

JPEG Compression Artifact Reduction (color)

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py

## color
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py


## single-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py

## 004 Grayscale Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py
```

(continues on next page)

(continued from previous page)

```
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py

## 005 Color Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py

## color
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py


## multi-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
./tools/dist_train.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8

## 003 Real-World Image Super-Resolution
```

(continues on next page)

(continued from previous page)

```
./tools/dist_train.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8

## 004 Grayscale Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py 8

## 005 Color Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py 8

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py 8

## color
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s48w8d6e180_8xb4-lr2e-4-500k_div2k-ed2d419e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s48w8d6e180_8xb4-lr2e-4-500k_div2k-926950f1.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s48w8d6e180_8xb4-lr2e-4-500k_div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-69e15fb6.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-d6982f7b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-131d3f64.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-309cb239.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-d6622d03.pth

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-c6425057.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-6f0c425f.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-36960d18.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-a016a72f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-9f1599b5.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15-6782691b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25-d0d8d4da.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15-c74a2cee.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25-df2b1c0c.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10-da93c8e9.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20-d47367b1.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30-52c083cf.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40-803e8d9b.pth


## color
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## single-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth

python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth


## 003 Real-World Image Super-Resolution
python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth

python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth


## 004 Grayscale Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth


## 005 Color Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth


## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth


## color
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## multi-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
./tools/dist_test.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

./tools/dist_test.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

./tools/dist_test.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth

## 003 Real-World Image Super-Resolution
./tools/dist_test.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth

./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth

## color
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```bibtex
@inproceedings{liang2021swinir,
  title={Swinir: Image restoration using swin transformer},
  author={Liang, Jingyun and Cao, Jiezhang and Sun, Guolei and Zhang, Kai and Van Gool,␣
↪Luc and Timofte, Radu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={1833--1844},
  year={2021}
}
```

1.46 jpeg compression artifact reduction

1.46.1 Summary

- Number of checkpoints: 8
- Number of configs: 8
- Number of papers: 1
  - ALGORITHM: 1

1.46.2 SwinIR (ICCVW'2021)

SwinIR: Image Restoration Using Swin Transformer

Task: Image Super-Resolution, Image denoising, JPEG compression artifact reduction

Abstract

Image restoration is a long-standing low-level vision problem that aims to restore high-quality images from low-quality images (e.g., downscaled, noisy and compressed images). While state-of-the-art image restoration methods are based on convolutional neural networks, few attempts have been made with Transformers which show impressive performance on high-level vision tasks. In this paper, we propose a strong baseline model SwinIR for image restoration based on the Swin Transformer. SwinIR consists of three parts: shallow feature extraction, deep feature extraction and high-quality image reconstruction. In particular, the deep feature extraction module is composed of several residual Swin Transformer blocks (RSTB), each of which has several Swin Transformer layers together with a residual connection. We conduct experiments on three representative tasks: image super-resolution (including classical, lightweight and real-world image super-resolution), image denoising (including grayscale and color image denoising) and JPEG compression artifact reduction. Experimental results demonstrate that SwinIR outperforms state-of-the-art methods on different tasks by up to 0.14~0.45dB, while the total number of parameters can be reduced by up to 67%.

Results and models

Classical Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Lightweight Image Super-Resolution

Evaluated on Y channels, `scale` pixels in each border are cropped before evaluation. The metrics are `PSNR / SSIM` .

Real-World Image Super-Resolution

Evaluated on Y channels. The metrics are `NIQE` .

Grayscale Image Deoising

Evaluated on grayscale images. The metrics are `PSNR` .

Color Image Deoising

Evaluated on RGB channels. The metrics are `PSNR` .

JPEG Compression Artifact Reduction (grayscale)

Evaluated on grayscale images. The metrics are `PSNR / SSIM

JPEG Compression Artifact Reduction (color)

Evaluated on RGB channels. The metrics are `PSNR / SSIM` .

Quick Start

Train

You can use the following commands to train a model with cpu or single/multiple GPUs.

```
## cpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py

## color
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py


## single-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/train.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/train.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py

## 002 Lightweight Image Super-Resolution (small size)
python tools/train.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py
python tools/train.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py

## 003 Real-World Image Super-Resolution
python tools/train.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py
python tools/train.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py

## 004 Grayscale Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py

## 005 Color Image Deoising (middle size)
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py
python tools/train.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
```

(continues on next page)

(continued from previous page)

```
## grayscale
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py

## color
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py
python tools/train.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py


## multi-gpu train
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
./tools/dist_train.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py 8

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py 8

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_train.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8
./tools/dist_train.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py 8

## 003 Real-World Image Super-Resolution
./tools/dist_train.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py 8
```

(continues on next page)

(continued from previous page)

```
## 004 Grayscale Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py 8

## 005 Color Image Deoising (middle size)
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN15.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN25.py 8
./tools/dist_train.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorDN50.py 8

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪grayCAR40.py 8

## color
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py 8
./tools/dist_train.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py 8
```

For more details, you can refer to **Train a model** part in train_test.md.

Test

You can use the following commands to test a model with cpu or single/multiple GPUs.

```
## cpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s48w8d6e180_8xb4-lr2e-4-500k_div2k-ed2d419e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s48w8d6e180_8xb4-lr2e-4-500k_div2k-926950f1.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s48w8d6e180_8xb4-lr2e-4-500k_div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-69e15fb6.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-d6982f7b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-
↪lr2e-4-500k_df2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-131d3f64.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-309cb239.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-
↪lr2e-4-500k_div2k.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-d6622d03.pth

## 003 Real-World Image Super-Resolution
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-c6425057.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-6f0c425f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-36960d18.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-os-a016a72f.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-9f1599b5.pth
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_
↪8xb4-lr1e-4-600k_df2k-ost.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15-6782691b.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25-d0d8d4da.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN15.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15-c74a2cee.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN25.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25-df2b1c0c.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorDN50.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10-da93c8e9.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20-d47367b1.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30-52c083cf.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-grayCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40-803e8d9b.pth


## color
```

(continues on next page)

(continued from previous page)

```
CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

CUDA_VISIBLE_DEVICES=-1 python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_
↪s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## single-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
python tools/test.py configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

python tools/test.py configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

python tools/test.py configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
python tools/test.py configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

python tools/test.py configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth

python tools/test.py configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth


## 002 Lightweight Image Super-Resolution (small size)
python tools/test.py configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

python tools/test.py configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth


## 003 Real-World Image Super-Resolution
python tools/test.py configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth

python tools/test.py configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth

python tools/test.py configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

python tools/test.py configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth


## 004 Grayscale Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth


## 005 Color Image Deoising (middle size)
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth
```

(continues on next page)

(continued from previous page)

```
python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

python tools/test.py configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth


## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding usesx8 blocks)
## grayscale
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth


## color
python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

python tools/test.py configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR40-5b77a6e6.pth



## multi-gpu test
## 001 Classical Image Super-Resolution (middle size)
## (setting1: when model is trained on DIV2K and with training_patch_size=48)
```

(continues on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-ed2d419e.pth

./tools/dist_test.sh configs/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x3s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-926950f1.pth

./tools/dist_test.sh configs/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_div2k.py␣
↪https://download.openmmlab.com/mmediting/swinir/swinir_x4s48w8d6e180_8xb4-lr2e-4-500k_
↪div2k-88e4903d.pth

## (setting2: when model is trained on DIV2K+Flickr2K and with training_patch_size=64)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪69e15fb6.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪d6982f7b.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d6e180_8xb4-lr2e-4-500k_df2k-
↪0502d775.pth

## 002 Lightweight Image Super-Resolution (small size)
./tools/dist_test.sh configs/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x2s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪131d3f64.pth

./tools/dist_test.sh configs/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x3s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪309cb239.pth

./tools/dist_test.sh configs/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k.py https:/
↪/download.openmmlab.com/mmediting/swinir/swinir_x4s64w8d4e60_8xb4-lr2e-4-500k_div2k-
↪d6622d03.pth

## 003 Real-World Image Super-Resolution
./tools/dist_test.sh configs/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-c6425057.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x2s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-6f0c425f.pth

./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-36960d18.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d6e180_8xb4-lr1e-
↪4-600k_df2k-os-a016a72f.pth
```

(continued on next page)

(continued from previous page)

```
./tools/dist_test.sh configs/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_gan-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-9f1599b5.pth

./tools/dist_test.sh configs/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-4-600k_df2k-ost.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_psnr-x4s64w8d9e240_8xb4-lr1e-
↪4-600k_df2k-os-25f1722a.pth

## 004 Grayscale Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN15-6782691b.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN25-d0d8d4da.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-grayDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayDN50-54c9968a.pth

## 005 Color Image Deoising (middle size)
./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN15.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN15-c74a2cee.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN25.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN25-df2b1c0c.pth

./tools/dist_test.sh configs/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-1600k_dfwb-colorDN50.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s128w8d6e180_8xb1-lr2e-4-
↪1600k_dfwb-colorDN50-e369874c.pth

## 006 JPEG Compression Artifact Reduction (middle size, using window_size=7 because␣
↪JPEG encoding uses 8x8 blocks)
## grayscale
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR10.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR10-da93c8e9.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR20.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR20-d47367b1.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR30.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR30-52c083cf.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-grayCAR40.
↪py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-
↪1600k_dfwb-grayCAR40-803e8d9b.pth
```

(continues on next page)

(continued from previous page)

```
## color
./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR10.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR10-09aafadc.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR20.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR20-b8a42b5e.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR30.py https://download.openmmlab.com/mmediting/swinir/swinir_s126w7d6e180_8xb1-
↪lr2e-4-1600k_dfwb-colorCAR30-e9fe6859.pth

./tools/dist_test.sh configs/swinir/swinir_s126w7d6e180_8xb1-lr2e-4-1600k_dfwb-
↪colorCAR40.py https://download.openmmlab.com/mmediting/swinir/
```

For more details, you can refer to **Test a pre-trained model** part in train_test.md.

Citation

```
@inproceedings{liang2021swinir,
  title={Swinir: Image restoration using swin transformer},
  author={Liang, Jingyun and Cao, Jiezhang and Sun, Guolei and Zhang, Kai and Van Gool,␣
↪Luc and Timofte, Radu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={1833--1844},
  year={2021}
}
```

1.47 Overview

- *df2k_ost*
- *realsrset*
- *videolq*
- *glean*
- *celeba-hq*
- *gopro*
- *vimeo90k*
- *paired-pix2pix*
- *sidd*
- *dpdd*
- *udm10*
- *unpaired-cyclegan*

- *hide*
- *paris-street-view*
- *vid4*
- *ntire21_decompression*
- *deraining*
- *reds*
- *denoising*
- *comp1k*
- *realblur*
- *unconditional_gans*
- *spmcs*
- *places365*
- *live1*
- *vimeo90k-triplet*
- *classic5*
- *div2k*

1.48 Preparing DF2K_OST Dataset

```
@inproceedings{wang2021real,
  title={Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic␣
↪Data},
  author={Wang, Xintao and Xie, Liangbin and Dong, Chao and Shan, Ying},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={1905--1914},
  year={2021}
}
```

- The DIV2K dataset can be downloaded from here (We use the training set only).
- The Flickr2K dataset can be downloaded here (We use the training set only).
- The OST dataset can be downloaded here (We use the training set only).

Please first put all the images into the `GT` folder (naming does not need to be in order):

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── df2k_ost
│   │   ├── GT
│   │   │   ├── 0001.png
│   │   │   ├── 0002.png
```

(continues on next page)

(continued from previous page)

```
│   │   │   ├── ...
...
```

1.48.1 Crop sub-images

For faster IO, we recommend to crop the images to sub-images. We provide such a script:

```
python tools/dataset_converters/df2k_ost/preprocess_df2k_ost_dataset.py --data-root ./
↪data/df2k_ost
```

The generated data is stored under `df2k_ost` and the data structure is as follows, where `_sub` indicates the sub-images.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── df2k_ost
│   │   ├── GT
│   │   ├── GT_sub
│   │   ├── meta_info_df2k_ost.txt
...
```

1.48.2 Prepare annotation list

If you use the annotation mode for the dataset, you first need to prepare a specific `txt` file.

Each line in the annotation file contains the image names and image shape (usually for the ground-truth images), separated by a white space.

Example of an annotation file:

```
0001_s001.png (480,480,3)
0001_s002.png (480,480,3)
```

Note that `preprocess_df2k_ost_dataset.py` will generate default annotation files.

1.48.3 Prepare LMDB dataset for DF2K_OST

If you want to use LMDB datasets for faster IO speed, you can make LMDB files by:

```
python tools/dataset_converters/df2k_ost/preprocess_df2k_ost_dataset.py --data-root ./
↪data/df2k_ost --make-lmdb
```

1.49 Preparing RealSRSet Dataset

```bibtex
@inproceedings{zhang2021designing,
    title={Designing a Practical Degradation Model for Deep Blind Image Super-Resolution}
↪,
    author={Zhang, Kai and Liang, Jingyun and Van Gool, Luc and Timofte, Radu},
    booktitle={IEEE International Conference on Computer Vision},
    pages={4791--4800},
    year={2021}
}
```

The datasets RealSRSet can be download from here.

The datasets RealSRSet+5images can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── realsrset
│├── RealSRSet+5images
```

1.50 Preparing VideoLQ Dataset

```bibtex
@inproceedings{chan2022investigating,
  author = {Chan, Kelvin C.K. and Zhou, Shangchen and Xu, Xiangyu and Loy, Chen Change},
  title = {Investigating Tradeoffs in Real-World Video Super-Resolution},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition},
  year = {2022}
}
```

You can download the dataset using Dropbox / Google Drive / OneDrive.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── VideoLQ
│   │   ├── 000
│   │   │   ├── 00000000.png
│   │   │   ├── 00000001.png
│   │   │   ├── ...
│   │   ├── 001
│   │   ├── 002
│   │   ├── ...
```

1.51 Preparing GLEAN Dataset

```
@InProceedings{chan2021glean,
  author = {Chan, Kelvin CK and Wang, Xintao and Xu, Xiangyu and Gu, Jinwei and Loy, Chen Change},
  title = {GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution},
  booktitle = {Proceedings of the IEEE conference on computer vision and pattern recognition},
  year = {2021}
}
```

1.51.1 Preparing cat_train dataset

1. Download cat dataset from LSUN homepage
2. Download cat_train/meta_info_LSUNcat_GT.txt from GLEAN homepage
3. Export and downsample images

Export images from lmdb file and resize the input images to the designated size. We provide such a script:

```
python tools/dataset_converters/glean/preprocess_cat_train_dataset.py --lmdb-path .data/cat --meta-file-path ./data/cat_train/meta_info_LSUNcat_GT.txt --out-dir ./data/cat_train
```

The generated data is stored under `cat_train` and the folder structure is as follows.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── cat_train
│   │   ├── GT
│   │   ├── BIx8_down
│   │   ├── BIx16_down
│   │   ├── meta_info_LSUNcat_GT.txt
...
```

1.51.2 Preparing cat_test dataset

1. Download CAT dataset from here.
2. Download cat_test/meta_info_Cat100_GT.txt from GLEAN homepage
3. Downsample images

Resize the input images to the designated size. We provide such a script:

```
python tools/dataset_converters/glean/preprocess_cat_test_dataset.py --data-path ./data/CAT_03 --meta-file-path ./data/cat_test/meta_info_Cat100_GT.txt --out-dir ./data/cat_test
```

The generated data is stored under `cat_test` and the folder structure is as follows.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── cat_test
│   │   ├── GT
│   │   ├── BIx8_down
│   │   ├── BIx16_down
│   │   ├── meta_info_Cat100_GT.txt
...
```

1.51.3 Preparing FFHQ dataset

1. Download FFHQ dataset (images1024x1024) from it's homepage

Then you can refactor the folder structure looks like:

```
ffhq
├── images
│├── 00000.png
│├── 00001.png
│├── ...
│├── 69999.png
```

2. Download ffhq/meta_info_FFHQ_GT.txt from GLEAN homepage
3. Downsample images

Resize the input images to the designated size. We provide such a script:

```
python tools/dataset_converters/glean/preprocess_ffhq_celebahq_dataset.py --data-root ./
↪data/ffhq/images
```

The generated data is stored under `ffhq` and the folder structure is as follows.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
|   ├── ffhq
|   |   ├── images
|   |   ├── BIx8_down
|   |   ├── BIx16_down
|   |   ├── meta_info_FFHQ_GT.txt
...
```

1.51.4 Preparing CelebA-HQ dataset

1. Preparing datasets following it's homepage

Then you can refactor the folder structure looks like:

```
CelebA-HQ
├── GT
|├── 00000.png
|├── 00001.png
|├── ...
|├── 30000.png
```

2. Download CelebA-HQ/meta_info_CelebAHQ_val100_GT.txt from GLEAN homepage
3. Downsample images

Resize the input images to the designated size. We provide such a script:

```
python tools/dataset_converters/glean/preprocess_ffhq_celebahq_dataset.py --data-root ./
↪data/CelebA-HQ/GT
```

The generated data is stored under `CelebA-HQ` and the folder structure is as follows.

```
mmagic
├── mmagic
├── tools
├── configsdata
├── data
|   ├── CelebA-HQ
|   |   ├── GT
|   |   ├── BIx8_down
|   |   ├── BIx16_down
|   |   ├── meta_info_CelebAHQ_val100_GT.txt
...
```

1.51.5 Preparing FFHQ_CelebAHQ dataset

We merge FFHQ(`ffhq/images`) and CelebA-HQ(`CelebA-HQ/GT`) to generate FFHQ_CelebAHQ dataset.

The folder structure should looks like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
|   ├── FFHQ_CelebAHQ
|   |   ├── GT
...
```

1.52 Preparing CelebA-HQ Dataset

```
@article{karras2017progressive,
  title={Progressive growing of gans for improved quality, stability, and variation},
  author={Karras, Tero and Aila, Timo and Laine, Samuli and Lehtinen, Jaakko},
  journal={arXiv preprint arXiv:1710.10196},
  year={2017}
}
```

Follow the instructions here to prepare the dataset.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── CelebA-HQ
│   │   ├── train_256
|   |   ├── test_256
|   |   ├── train_celeba_img_list.txt
|   |   ├── val_celeba_img_list.txt
```

1.53 Preparing GoPro Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The train datasets can be download from here. The test datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
|├── GoPro
||├── train
|||   ├── blur
|||   ├── sharp
||├── test
|||   ├── blur
|||   ├── sharp
```

1.54 Preparing Vimeo90K Dataset

```
@article{xue2019video,
  title={Video Enhancement with Task-Oriented Flow},
  author={Xue, Tianfan and Chen, Baian and Wu, Jiajun and Wei, Donglai and Freeman,␣
↪William T},
  journal={International Journal of Computer Vision (IJCV)},
  volume={127},
  number={8},
  pages={1106--1125},
  year={2019},
  publisher={Springer}
}
```

The training and test datasets can be downloaded from here.

Then you can rename the directory `vimeo_septuplet/sequences` to `vimeo90k/GT`. The Vimeo90K dataset has a `clip/sequence/img` folder structure:

```
vimeo90k
├── GT
│   ├── 00001
│   │   ├── 0001
│   │   │   ├── im1.png
│   │   │   ├── im2.png
│   │   │   ├── ...
│   │   ├── 0002
│   │   ├── 0003
│   │   ├── ...
│   ├── 00002
│   ├── ...
├── sep_trainlist.txt
├── sep_testlist.txt
```

To generate the downsampling images BIx4 and BDx4 and prepare the annotation file, you need to run the following command:

```
python tools/dataset_converters/vimeo90k/preprocess_vimeo90k_dataset.py --data-root ./
↪data/vimeo90k
```

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── vimeo_triplet
│   │   ├── GT
│   │   │   ├── 00001
│   │   │   │   ├── 0001
│   │   │   │   │   ├── im1.png
│   │   │   │   │   ├── im2.png
```

(continues on next page)

(continued from previous page)

```
│   │   │   │   │   ├── ...
│   │   │   │   ├── 0002
│   │   │   │   ├── 0003
│   │   │   │   ├── ...
│   │   │   ├── 00002
│   │   │   ├── ...
│   │   ├── BIx4
│   │   ├── BDx4
│   │   ├── meta_info_Vimeo90K_test_GT.txt
│   │   ├── meta_info_Vimeo90K_train_GT.txt
```

1.54.1 Prepare LMDB dataset for Vimeo90K

If you want to use LMDB datasets for faster IO speed, you can make LMDB files by:

```
python tools/dataset_converters/vimeo90k/preprocess_vimeo90k_dataset.py --data-root ./
↪data/vimeo90k --train_list ./data/vimeo90k/sep_trainlist.txt --gt-path ./data/vimeo90k/
↪GT --lq-path ./data/Vimeo90k/BIx4  --make-lmdb
```

1.55 Preparing Paired Dataset for Pix2pix

```bibtex
@inproceedings{isola2017image,
  title={Image-to-image translation with conditional adversarial networks},
  author={Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={1125--1134},
  year={2017}
}
```

You can download paired datasets from here. Then, you need to unzip and move corresponding datasets to follow the folder structure shown above. The datasets have been well-prepared by the original authors.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
    ├── paired
    │   ├── facades
    │   ├── maps
|   |   ├── edges2shoes
|   |   |   ├── train
|   |   |   ├── test
```

1.56 Preparing SIDD Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The train datasets can be download from here. The validation datasets can be download from here.

For validation datasets, we need to export images from mat file. We provide such a script:

```
python tools/dataset_converters/sidd/preprocess_sidd_test_dataset.py --data-root ./data/
↪SIDD/val --out-dir ./data/SIDD/val
```

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── SIDD
││├── train
│││   ├── gt
│││   ├── noisy
││├── val
│││   ├── gt
│││   ├── noisy
│││   ├── ValidationNoisyBlocksSrgb.mat
│││   ├── ValidationGtBlocksSrgb.mat
```

1.57 Preparing DPDD Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The test datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
```

(continues on next page)

(continued from previous page)

```
├── data
│├── DPDD
│└── inputL
│└── inputR
│└── inputC
│└── target
│└── indoor_labels.npy
│└── indoor_labels.txt
│└── outdoor_labels.npy
│└── outdoor_labels.txt
```

1.58 Preparing UDM10 Dataset

```
@inproceedings{PFNL,
    title={Progressive Fusion Video Super-Resolution Network via Exploiting Non-Local␣
↪Spatio-Temporal Correlations},
    author={Yi, Peng and Wang, Zhongyuan and Jiang, Kui and Jiang, Junjun and Ma, Jiayi},
    booktitle={IEEE International Conference on Computer Vision (ICCV)},
    pages={3106-3115},
    year={2019},
}
```

The datasets can be downloaded from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── UDM10
│└── GT
│└── BIx4
│└── BDx4
```

1.59 Preparing Unpaired Dataset for CycleGAN

```
@inproceedings{zhu2017unpaired,
  title={Unpaired image-to-image translation using cycle-consistent adversarial networks}
↪,
  author={Zhu, Jun-Yan and Park, Taesung and Isola, Phillip and Efros, Alexei A},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={2223--2232},
  year={2017}
}
```

You can download unpaired datasets from here. Then, you need to unzip and move corresponding datasets to follow the folder structure shown above. The datasets have been well-prepared by the original authors.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
    ├── unpaired
    |   ├── facades
|   |   ├── horse2zebra
|   |   ├── summer2winter_yosemite
|   |   |   ├── trainA
|   |   |   ├── trainB
|   |   |   ├── testA
|   |   |   ├── testB
```

1.60 Preparing HIDE Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The test datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
|├── HIDE
||├── input
||├── target
```

1.61 Preparing Paris Street View Dataset

```
@inproceedings{pathak2016context,
  title={Context encoders: Feature learning by inpainting},
  author={Pathak, Deepak and Krahenbuhl, Philipp and Donahue, Jeff and Darrell, Trevor␣
↪and Efros, Alexei A},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={2536--2544},
  year={2016}
}
```

Obtain the dataset here.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── paris_street_view
│   │   ├── train
│   │   ├── val
```

1.62 Preparing Vid4 Dataset

```bibtex
@article{xue2019video,
  title={On Bayesian adaptive video super resolution},
  author={Liu, Ce and Sun, Deqing},
  journal={IEEE Transactions on Pattern Analysis and Machine Intelligence},
  volume={36},
  number={2},
  pages={346--360},
  year={2013},
  publisher={IEEE}
}
```

The Vid4 dataset can be downloaded from here. There are two degradations in the dataset.

1. BIx4 contains images downsampled by bicubic interpolation
2. BDx4 contains images blurred by Gaussian kernel with =1.6, followed by a subsampling every four pixels.

Note that we should prepare a annotation file (such as meta_info_Vid4_GT.txt) for Vid4 dataset as follows.

```
calendar 41 (576,720,3)
city 34 (576,704,3)
foliage 49 (480,720,3)
walk 47 (480,720,3)
```

For ToFlow, we should prepare directly upsampling dataset. We provide such a script:

```
python tools/dataset_converters/vid4/preprocess_vid4_dataset.py --data-root ./data/Vid4/
↪BIx4
```

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── Vid4
│   │   ├── GT
│   │   │   ├── calendar
│   │   │   ├── city
```

(continues on next page)

(continued from previous page)

```
│   │   │   ├── foliage
│   │   │   ├── walk
│   │   ├── BDx4
│   │   ├── BIx4
│   │   ├── BIx4up_direct
│   │   ├── meta_info_Vid4_GT.txt
```

1.63 Preparing NTIRE21 decompression Dataset

```
@inproceedings{yang2021dataset,
  title={{NTIRE 2021} Challenge on Quality Enhancement of Compressed Video: Dataset and
↪Study},
  author={Ren Yang and Radu Timofte},
  booktitle={IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops},
  year={2021}
}
```

The test datasets can be download from it's Homepage.

Please follows the tutorials of the Homepage to generate datasets.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── NTIRE21_decompression_track1
││├── GT
│││   ├── 001
│││   │   ├── 001.png
│││   │   ├── ...
│││   ├── ...
│││   ├── 010
││├── LQ
│││   ├── 001
│││   │   ├── 001.png
│││   │   ├── ...
│││   ├── ...
│││   ├── 010
│├── NTIRE21_decompression_track2
││├── GT
││├── LQ
│├── NTIRE21_decompression_track3
││├── GT
││├── LQ
```

1.64 Preparing Deraining Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The test datasets (Rain100H, Rain100L, Test100, Test1200, Test2800) can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── Rain100H
│└── input
│└── target
│├── Rain100L
│└── input
│└── target
│├── Test100
│└── input
│└── target
│├── Test1200
│└── input
│└── target
│├── Test2800
│└── input
│└── target
```

1.65 Preparing REDS Dataset

```
@InProceedings{Nah_2019_CVPR_Workshops_REDS,
  author = {Nah, Seungjun and Baik, Sungyong and Hong, Seokil and Moon, Gyeongsik and␣
↪Son, Sanghyun and Timofte, Radu and Lee, Kyoung Mu},
  title = {NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and␣
↪Study},
  booktitle = {The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)␣
↪Workshops},
  month = {June},
  year = {2019}
}
```

- Training dataset: REDS dataset.
- Validation dataset: REDS dataset and Vid4.

Note that we merge train and val datasets in REDS for easy switching between REDS4 partition (used in EDVR) and the official validation partition. The original val dataset (clip names from 000 to 029) are modified to avoid conflicts with training dataset (total 240 clips). Specifically, the clip names are changed to 240, 241, … 269.

You can prepare the REDS dataset by running:

```
python tools/dataset_converters/reds/preprocess_reds_dataset.py --root-path ./data/REDS
```

```
mmagic
├── mmagic
├── tools
├── configs
├── data
    ├── REDS
        ├── train_sharp
            ├── 000
            ├── 001
            ├── ...
        ├── train_sharp_bicubic
            ├── X4
            |   ├── 000
            |   ├── 001
            |   ├── ...
        ├── meta_info_reds4_train.txt
        ├── meta_info_reds4_val.txt
        ├── meta_info_official_train.txt
        ├── meta_info_official_val.txt
        ├── meta_info_REDS_GT.txt
    ├── REDS4
        ├── GT
        ├── sharp_bicubic
```

1.65.1 Prepare LMDB dataset for REDS

If you want to use LMDB datasets for faster IO speed, you can make LMDB files by:

```
python tools/dataset_converters/reds/preprocess_reds_dataset.py --root-path ./data/REDS -
↪-make-lmdb
```

1.65.2 Crop to sub-images

MMagic also supports cropping REDS images to sub-images for faster IO. We provide such a script:

```
python tools/dataset_converters/reds/crop_sub_images.py --data-root ./data/REDS  -scales␣
↪4
```

The generated data is stored under `REDS` and the data structure is as follows, where `_sub` indicates the sub-images.

```
mmagic
├── mmagic
├── tools
```

(continues on next page)

(continued from previous page)

```
├── configs
├── data
│   ├── REDS
│   │   ├── train_sharp
│   │   │   ├── 000
│   │   │   ├── 001
│   │   │   ├── ...
│   │   ├── train_sharp_sub
│   │   │   ├── 000_s001
│   │   │   ├── 000_s002
│   │   │   ├── ...
│   │   │   ├── 001_s001
│   │   │   ├── ...
│   │   ├── train_sharp_bicubic
│   │   │   ├── X4
│   │   │   │   ├── 000
│   │   │   │   ├── 001
│   │   │   │   ├── ...
│   │   │   ├── X4_sub
│   │   │   ├── 000_s001
│   │   │   ├── 000_s002
│   │   │   ├── ...
│   │   │   ├── 001_s001
│   │   │   ├── ...
```

Note that by default `preprocess_reds_dataset.py` does not make lmdb and annotation file for the cropped dataset. You may need to modify the scripts a little bit for such operations.

1.66 Preparing Denoising Dataset

```bibtex
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The test datasets (Set12, BSD68, CBSD68, Kodak, McMaster, Urban100) can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── denoising_gaussian_test
││├── Set12
││├── BSD68
││├── CBSD68
```

(continues on next page)

(continued from previous page)

```
||├── Kodak
||├── McMaster
||├── Urban100
```

1.67 Preparing Composition-1k Dataset

1.67.1 Introduction

```bibtex
@inproceedings{xu2017deep,
  title={Deep image matting},
  author={Xu, Ning and Price, Brian and Cohen, Scott and Huang, Thomas},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern␣
↪recognition},
  pages={2970--2979},
  year={2017}
}
```

The Adobe Composition-1k dataset consists of foreground images and their corresponding alpha images. To get the full dataset, one need to composite the foregrounds with selected backgrounds from the COCO dataset and the Pascal VOC dataset.

1.67.2 Obtain and Extract

Please follow the instructions of paper authors to obtain the Composition-1k (comp1k) dataset.

1.67.3 Composite the full dataset

The Adobe composition-1k dataset contains only `alpha` and `fg` (and `trimap` in test set). It is needed to merge `fg` with COCO data (training) or VOC data (test) before training or evaluation. Use the following script to perform image composition and generate annotation files for training or testing:

```
# The script is run under the root folder of MMagic
python tools/dataset_converters/matting/comp1k/preprocess_comp1k_dataset.py data/adobe_
↪composition-1k data/coco data/VOCdevkit --composite
```

The generated data is stored under `adobe_composition-1k/Training_set` and `adobe_composition-1k/Test_set` respectively. If you only want to composite test data (since compositing training data is time-consuming), you can skip compositing the training set by removing the `--composite` option:

```
# skip compositing training set
python tools/dataset_converters/matting/comp1k/preprocess_comp1k_dataset.py data/adobe_
↪composition-1k data/coco data/VOCdevkit
```

If you only want to preprocess test data, i.e. for FBA, you can skip the train set by adding the `--skip-train` option:

```
# skip preprocessing training set
python tools/data/matting/comp1k/preprocess_comp1k_dataset.py data/adobe_composition-1k␣
↪data/coco data/VOCdevkit --skip-train
```

Currently, GCA and FBA support online composition of training data. But you can modify the data pipeline of other models to perform online composition instead of loading composited images (we called it merged in our data pipeline).

1.67.4 Check Directory Structure for DIM

The result folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── adobe_composition-1k
│   │   ├── Test_set
│   │   │   ├── Adobe-licensed images
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   │   ├── trimaps
│   │   │   ├── merged  (generated by tools/dataset_converters/matting/comp1k/preprocess_
↪comp1k_dataset.py)
│   │   │   ├── bg      (generated by tools/dataset_converters/matting/comp1k/preprocess_
↪comp1k_dataset.py)
│   │   ├── Training_set
│   │   │   ├── Adobe-licensed images
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   ├── Other
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   ├── merged  (generated by tools/dataset_converters/matting/comp1k/preprocess_
↪comp1k_dataset.py)
│   │   │   ├── bg      (generated by tools/dataset_converters/matting/comp1k/preprocess_
↪comp1k_dataset.py)
│   │   ├── test_list.json     (generated by tools/dataset_converters/matting/comp1k/
↪preprocess_comp1k_dataset.py)
│   │   ├── training_list.json (generated by tools/dataset_converters/matting/comp1k/
↪preprocess_comp1k_dataset.py)
│   ├── coco
│   │   ├── train2014   (or train2017)
│   ├── VOCdevkit
│   │   ├── VOC2012
```

1.67.5 Prepare the dataset for FBA

FBA adopts dynamic dataset augmentation proposed in Learning-base Sampling for Natural Image Matting. In addition, to reduce artifacts during augmentation, it uses the extended version of foreground as foreground. We provide scripts to estimate foregrounds.

Prepare the test set as follows:

```
# skip preprocessing training set, as it composites online during training
python tools/dataset_converters/matting/comp1k/preprocess_comp1k_dataset.py data/adobe_
↪composition-1k data/coco data/VOCdevkit --skip-train
```

Extend the foreground of training set as follows:

```
python tools/dataset_converters/matting/comp1k/extend_fg.py data/adobe_composition-1k
```

1.67.6 Check Directory Structure for DIM

The final folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── adobe_composition-1k
│   │   ├── Test_set
│   │   │   ├── Adobe-licensed images
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   │   ├── trimaps
│   │   │   ├── merged  (generated by tools/data/matting/comp1k/preprocess_comp1k_
↪dataset.py)
│   │   │   ├── bg      (generated by tools/data/matting/comp1k/preprocess_comp1k_
↪dataset.py)
│   │   ├── Training_set
│   │   │   ├── Adobe-licensed images
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   │   ├── fg_extended (generated by tools/data/matting/comp1k/extend_fg.py)
│   │   │   ├── Other
│   │   │   │   ├── alpha
│   │   │   │   ├── fg
│   │   │   │   ├── fg_extended (generated by tools/data/matting/comp1k/extend_fg.py)
│   │   ├── test_list.json         (generated by tools/data/matting/comp1k/preprocess_
↪comp1k_dataset.py)
│   │   ├── training_list_fba.json (generated by tools/data/matting/comp1k/extend_fg.py)
│   ├── coco
│   │   ├── train2014   (or train2017)
│   ├── VOCdevkit
│   │   ├── VOC2012
```

1.68 Preparing RealBlur Dataset

```
@inproceedings{Zamir2021Restormer,
  title={Restormer: Efficient Transformer for High-Resolution Image Restoration},
  author={Syed Waqas Zamir and Aditya Arora and Salman Khan and Munawar Hayat and Fahad␣
↪Shahbaz Khan and Ming-Hsuan Yang},
  booktitle={CVPR},
  year={2022}
}
```

The test datasets RealBlurR can be download from here. The test datasets RealBlurJ can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── RealBlur_R
││├── input
││├── target
│├── RealBlur_J
││├── input
││├── target
```

1.69 Unconditional GANs Datasets

Data preparation for unconditional model is simple. What you need to do is downloading the images and put them into a directory. Next, you should set a symlink in the `data` directory. For standard unconditional gans with static architectures, like DCGAN and StyleGAN2, `UnconditionalImageDataset` is designed to train such unconditional models. Here is an example config for FFHQ dataset:

```
dataset_type = 'BasicImageDataset'

train_pipeline = [
    dict(type='LoadImageFromFile', key='img'),
    dict(type='Flip', keys=['img'], direction='horizontal'),
    dict(type='PackInputs', keys=['img'], meta_keys=['img_path'])
]

# `batch_size` and `data_root` need to be set.
train_dataloader = dict(
    batch_size=4,
    num_workers=8,
    persistent_workers=True,
    sampler=dict(type='InfiniteSampler', shuffle=True),
    dataset=dict(
        type=dataset_type,
        data_root=None,  # set by user
        pipeline=train_pipeline))
```

Here, we adopt `InfinitySampler` to avoid frequent dataloader reloading, which will accelerate the training procedure. As shown in the example, `pipeline` provides important data pipeline to process images, including loading from file system, resizing, cropping, transferring to `torch.Tensor` and packing to `DataSample`. All of supported data pipelines can be found in `mmagic/datasets/transforms`.

For unconditional GANs with dynamic architectures like PGGAN and StyleGANv1, `GrowScaleImgDataset` is recommended to use for training. Since such dynamic architectures need real images in different scales, directly adopting `UnconditionalImageDataset` will bring heavy I/O cost for loading multiple high-resolution images. Here is an example we use for training PGGAN in CelebA-HQ dataset:

```
dataset_type = 'GrowScaleImgDataset'

pipeline = [
    dict(type='LoadImageFromFile', key='img'),
    dict(type='Flip', keys=['img'], direction='horizontal'),
    dict(type='PackInputs')
]

# `samples_per_gpu` and `imgs_root` need to be set.
train_dataloader = dict(
    num_workers=4,
    batch_size=64,
    dataset=dict(
        type='GrowScaleImgDataset',
        data_roots={
            '1024': './data/ffhq/images',
            '256': './data/ffhq/ffhq_imgs/ffhq_256',
            '64': './data/ffhq/ffhq_imgs/ffhq_64'
        },
        gpu_samples_base=4,
        # note that this should be changed with total gpu number
        gpu_samples_per_scale={
            '4': 64,
            '8': 32,
            '16': 16,
            '32': 8,
            '64': 4,
            '128': 4,
            '256': 4,
            '512': 4,
            '1024': 4
        },
        len_per_stage=300000,
        pipeline=pipeline),
    sampler=dict(type='InfiniteSampler', shuffle=True))
```

In this dataset, you should provide a dictionary of image paths to the `data_roots`. Thus, you should resize the images in the dataset in advance. For the resizing methods in the data pre-processing, we adopt bilinear interpolation methods in all of the experiments studied in MMagic.

Note that this dataset should be used with `PGGANFetchDataHook`. In this config file, this hook should be added in the customized hooks, as shown below.

```
custom_hooks = [
    dict(
```

(continues on next page)

(continued from previous page)

```
        type='VisualizationHook',
        interval=5000,
        fixed_input=True,
        # vis ema and orig at the same time
        vis_kwargs_list=dict(
            type='Noise',
            name='fake_img',
            sample_model='ema/orig',
            target_keys=['ema', 'orig'])),
    dict(type='PGGANFetchDataHook')
]
```

This fetching data hook helps the dataloader update the status of dataset to change the data source and batch size during training.

Here, we provide several download links of datasets frequently used in unconditional models: LSUN, CelebA, CelebA-HQ, FFHQ.

1.70 Preparing SPMCS Dataset

```
@InProceedings{tao2017spmc,
  author={Xin Tao and Hongyun Gao and Renjie Liao and Jue Wang and Jiaya Jia},
  title = {Detail-Revealing Deep Video Super-Resolution},
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
  month = {Oct},
  year = {2017}
}
```

The datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── SPMCS
│└── GT
│└── BIx4
│└── BDx4
│└── meta_info_SPMCS_GT.txt
```

1.71 Preparing Places365 Dataset

```
@article{zhou2017places,
  title={Places: A 10 million Image Database for Scene Recognition},
  author={Zhou, Bolei and Lapedriza, Agata and Khosla, Aditya and Oliva, Aude and␣
↪Torralba, Antonio},
  journal={IEEE Transactions on Pattern Analysis and Machine Intelligence},
  year={2017},
  publisher={IEEE}
}
```

Prepare the data from Places365.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── Places
│   │   ├── data_large
│   │   ├── val_large
│   │   ├── meta
│   │   │   ├── places365_train_challenge.txt
│   │   │   ├── places365_val.txt
```

1.72 Preparing LIVE1 Dataset

```
@article{zhang2017beyond,
  title={Beyond a {Gaussian} denoiser: Residual learning of deep {CNN} for image␣
↪denoising},
  author={Zhang, Kai and Zuo, Wangmeng and Chen, Yunjin and Meng, Deyu and Zhang, Lei},
  journal={IEEE Transactions on Image Processing},
  year={2017},
  volume={26},
  number={7},
  pages={3142-3155},
}
```

The test datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── LIVE1
```

1.73 Preparing Vimeo90K-triplet Dataset

```
@article{xue2019video,
  title={Video Enhancement with Task-Oriented Flow},
  author={Xue, Tianfan and Chen, Baian and Wu, Jiajun and Wei, Donglai and Freeman,␣
↪William T},
  journal={International Journal of Computer Vision (IJCV)},
  volume={127},
  number={8},
  pages={1106--1125},
  year={2019},
  publisher={Springer}
}
```

The training and test datasets can be download from here.

The Vimeo90K-triplet dataset has a `clip/sequence/img` folder structure:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── vimeo_triplet
│   │   ├── tri_testlist.txt
│   │   ├── tri_trainlist.txt
│   │   ├── sequences
│   │   │   ├── 00001
│   │   │   │   ├── 0001
│   │   │   │   │   ├── im1.png
│   │   │   │   │   ├── im2.png
│   │   │   │   │   └── im3.png
│   │   │   │   ├── 0002
│   │   │   │   ├── 0003
│   │   │   │   ├── ...
│   │   │   ├── 00002
│   │   │   ├── ...
```

1.74 Preparing Classic5 Dataset

```
@article{zhang2017beyond,
  title={Beyond a {Gaussian} denoiser: Residual learning of deep {CNN} for image␣
↪denoising},
  author={Zhang, Kai and Zuo, Wangmeng and Chen, Yunjin and Meng, Deyu and Zhang, Lei},
  journal={IEEE Transactions on Image Processing},
  year={2017},
  volume={26},
  number={7},
  pages={3142-3155},
}
```

The test datasets can be download from here.

The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│├── Classic5
```

1.75 Preparing DIV2K Dataset

```
@InProceedings{Agustsson_2017_CVPR_Workshops,
    author = {Agustsson, Eirikur and Timofte, Radu},
    title = {NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study},
    booktitle = {The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)␣
↪Workshops},
    month = {July},
    year = {2017}
}
```

- Training dataset: DIV2K dataset.
- Validation dataset: Set5 and Set14.

Note that we merge the original val dataset (image names from 0801 to 0900) to the original train dataset (image names from 0001 to 0800). The folder structure should look like:

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── DIV2K
│   │   ├── DIV2K_train_HR
│   │   │   ├── 0001.png
│   │   │   ├── 0002.png
│   │   │   ├── ...
│   │   │   ├── 0800.png
│   │   │   ├── 0801.png
│   │   │   ├── ...
│   │   │   ├── 0900.png
│   │   ├── DIV2K_train_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   │   ├── DIV2K_valid_HR
│   │   ├── DIV2K_valid_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   ├── Set5
│   │   ├── GTmod12
```

(continues on next page)

(continued from previous page)

```
│   │   │   ├── LRbicx2
│   │   │   ├── LRbicx3
│   │   │   ├── LRbicx4
│   │   ├── Set14
│   │   │   ├── GTmod12
│   │   │   ├── LRbicx2
│   │   │   ├── LRbicx3
│   │   │   ├── LRbicx4
```

1.75.1 Crop sub-images

For faster IO, we recommend to crop the DIV2K images to sub-images. We provide such a script:

```
python tools/dataset_converters/div2k/preprocess_div2k_dataset.py --data-root ./data/
↪DIV2K
```

The generated data is stored under `DIV2K` and the data structure is as follows, where `_sub` indicates the sub-images.

```
mmagic
├── mmagic
├── tools
├── configs
├── data
│   ├── DIV2K
│   │   ├── DIV2K_train_HR
│   │   ├── DIV2K_train_HR_sub
│   │   ├── DIV2K_train_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   │   │   ├── X2_sub
│   │   │   ├── X3_sub
│   │   │   ├── X4_sub
│   │   ├── DIV2K_valid_HR
│   │   ├── ...
│   │   ├── meta_info_DIV2K800sub_GT.txt
│   │   ├── meta_info_DIV2K100sub_GT.txt
...
```

1.75.2 Prepare annotation list

If you use the annotation mode for the dataset, you first need to prepare a specific `txt` file.

Each line in the annotation file contains the image names and image shape (usually for the ground-truth images), separated by a white space.

Example of an annotation file:

```
0001_s001.png (480,480,3)
0001_s002.png (480,480,3)
```

Note that `preprocess_div2k_dataset` will generate default annotation files.

1.75.3 Prepare LMDB dataset for DIV2K

If you want to use LMDB datasets for faster IO speed, you can make LMDB files by:

```
python tools/dataset_converters/div2k/preprocess_div2k_dataset.py --data-root ./data/
↪DIV2K --make-lmdb
```

1.76 Changelog

Highlights

- An advanced and powerful inpainting algorithm named PowerPaint is released in our repository. Click to View

New Features & Improvements

- [Release] Post release for v1.1.0 by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2043
- [CodeCamp2023-645]Add dreambooth new cfg by @YanxingLiu in https://github.com/open-mmlab/mmagic/pull/2042
- [Enhance] add new config for *base* dir by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2053
- [Enhance] support using from_pretrained for instance_crop by @zengyh1900 in https://github.com/open-mmlab/mmagic/pull/2066
- [Enhance] update support for latest diffusers with lora by @zengyh1900 in https://github.com/open-mmlab/mmagic/pull/2067
- [Feature] PowerPaint by @zhuang2002 in https://github.com/open-mmlab/mmagic/pull/2076
- [Enhance] powerpaint improvement by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2078
- [Enhance] Improve powerpaint by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2080
- [Enhance] add outpainting to gradio_PowerPaint.py by @zhuang2002 in https://github.com/open-mmlab/mmagic/pull/2084
- [MMSIG] Add new configuration files for StyleGAN2 by @xiaomile in https://github.com/open-mmlab/mmagic/pull/2057
- [MMSIG] [Doc] Update data_preprocessor.md by @jinxianwei in https://github.com/open-mmlab/mmagic/pull/2055
- [Enhance] Enhance PowerPaint by @zhuang2002 in https://github.com/open-mmlab/mmagic/pull/2093

Bug Fixes

- [Fix] Update README.md by @eze1376 in https://github.com/open-mmlab/mmagic/pull/2048
- [Fix] Fix test tokenizer by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2050
- [Fix] fix readthedocs building by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2052
- [Fix] –local-rank for PyTorch >= 2.0.0 by @youqingxiaozhua in https://github.com/open-mmlab/mmagic/pull/2051
- [Fix] animatediff download from openxlab by @JianxinDong in https://github.com/open-mmlab/mmagic/pull/2061
- [Fix] fix best practice by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2063
- [Fix] try import expand mask from transformers by @zengyh1900 in https://github.com/open-mmlab/mmagic/pull/2064

- [Fix] Update diffusers to v0.23.0 by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2069
- [Fix] add openxlab link to powerpaint by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2082
- [Fix] Update swinir_x2s48w8d6e180_8xb4-lr2e-4-500k_div2k.py, use MultiValLoop. by @ashutoshsingh0223 in https://github.com/open-mmlab/mmagic/pull/2085
- [Fix] Fix a test expression that has a logical short circuit. by @munahaf in https://github.com/open-mmlab/mmagic/pull/2046
- [Fix] Powerpaint to load safetensors by @sdbds in https://github.com/open-mmlab/mmagic/pull/2088

New Contributors

- @eze1376 made their first contribution in https://github.com/open-mmlab/mmagic/pull/2048
- @youqingxiaozhua made their first contribution in https://github.com/open-mmlab/mmagic/pull/2051
- @JianxinDong made their first contribution in https://github.com/open-mmlab/mmagic/pull/2061
- @zhuang2002 made their first contribution in https://github.com/open-mmlab/mmagic/pull/2076
- @ashutoshsingh0223 made their first contribution in https://github.com/open-mmlab/mmagic/pull/2085
- @jinxianwei made their first contribution in https://github.com/open-mmlab/mmagic/pull/2055
- @munahaf made their first contribution in https://github.com/open-mmlab/mmagic/pull/2046
- @sdbds made their first contribution in https://github.com/open-mmlab/mmagic/pull/2088

Full Changelog: https://github.com/open-mmlab/mmagic/compare/v1.1.0...v1.2.0

1.76.1 v1.1.0 (22/09/2023)

Highlights

In this new version of MMagic, we have added support for the following five new algorithms.

- Support ViCo, a new SD personalization method. Click to View
- Support AnimateDiff, a popular text2animation method. Click to View

- Support SDXL. Click to View
- Support DragGAN implementation with MMagic. Click to View
- Support for FastComposer. Click to View

New Features & Improvements

- [Feature] Support inference with diffusers pipeline, sd_xl first. by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2023
- [Enhance] add negative prompt for sd inferencer by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2021
- [Enhance] Update flake8 checking config in setup.cfg by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/2007
- [Enhance] Add 'config_name' as a supplement to the 'model_setting' by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2027
- [Enhance] faster test by @okotaku in https://github.com/open-mmlab/mmagic/pull/2034
- [Enhance] Add OpenXLab Badge by @ZhaoQiiii in https://github.com/open-mmlab/mmagic/pull/2037

CodeCamp Contributions

- [CodeCamp2023-643] Add new configs of BigGAN by @limafang in https://github.com/open-mmlab/mmagic/pull/2003
- [CodeCamp2023-648] MMagic new config GuidedDiffusion by @ooooo-create in https://github.com/open-mmlab/mmagic/pull/2005
- [CodeCamp2023-649] MMagic new config Instance Colorization by @ooooo-create in https://github.com/open-mmlab/mmagic/pull/2010
- [CodeCamp2023-652] MMagic new config StyleGAN3 by @hhy150 in https://github.com/open-mmlab/mmagic/pull/2018
- [CodeCamp2023-653] Add new configs of Real BasicVSR by @RangeKing in https://github.com/open-mmlab/mmagic/pull/2030

Bug Fixes

- [Fix] Fix best practice and back to contents on mainpage, add new models to model zoo by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2001
- [Fix] Check CI error and remove main stream gpu test by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2013
- [Fix] Check circle ci memory by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2016
- [Fix] remove code and fix clip loss ut test by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2017
- [Fix] mock infer in diffusers pipeline inferencer ut. by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2026
- [Fix] Fix bug caused by merging draggan by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2029
- [Fix] Update QRcode by @crazysteeaam in https://github.com/open-mmlab/mmagic/pull/2009
- [Fix] Replace the download links in README with OpenXLab version by @FerryHuang in https://github.com/open-mmlab/mmagic/pull/2038
- [Fix] Increase docstring coverage by @liuwenran in https://github.com/open-mmlab/mmagic/pull/2039

New Contributors

- @limafang made their first contribution in https://github.com/open-mmlab/mmagic/pull/2003
- @ooooo-create made their first contribution in https://github.com/open-mmlab/mmagic/pull/2005
- @hhy150 made their first contribution in https://github.com/open-mmlab/mmagic/pull/2018
- @ZhaoQiiii made their first contribution in https://github.com/open-mmlab/mmagic/pull/2037
- @ElliotQi made their first contribution in https://github.com/open-mmlab/mmagic/pull/1980
- @Beaconsyh08 made their first contribution in https://github.com/open-mmlab/mmagic/pull/2012

Full Changelog: https://github.com/open-mmlab/mmagic/compare/v1.0.2…v1.0.3

1.76.2 v1.0.2 (24/08/2023)

Highlights

1. More detailed documentation

Thank you to the community contributors for helping us improve the documentation. We have improved many documents, including both Chinese and English versions. Please refer to the documentation for more details.

2. New algorithms

- Support Prompt-to-prompt, DDIM Inversion and Null-text Inversion. Click to View.

From right to left: origin image, DDIM inversion, Null-text inversion

Prompt-to-prompt Editing

- Support Textual Inversion. Click to view.
- Support Attention Injection for more stable video generation with controlnet. Click to view.
- Support Stable Diffusion Inpainting. Click to view.

New Features & Improvements

- [Enhancement] Support noise offset in stable diffusion training by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1880
- [Community] Support Glide Upsampler by @Taited in https://github.com/open-mmlab/mmagic/pull/1663
- [Enhance] support controlnet inferencer by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1891
- [Feature] support Albumentations augmentation transformations and pipeline by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1894
- [Feature] Add Attention Injection for unet by @liuwenran in https://github.com/open-mmlab/mmagic/pull/1895
- [Enhance] update benchmark scripts by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1907
- [Enhancement] update mmagic docs by @crazysteeaam in https://github.com/open-mmlab/mmagic/pull/1920
- [Enhancement] Support Prompt-to-prompt, ddim inversion and null-text inversion by @FerryHuang in https://github.com/open-mmlab/mmagic/pull/1908
- [CodeCamp2023-302] Support MMagic visualization and write a user guide by @aptsunny in https://github.com/open-mmlab/mmagic/pull/1939
- [Feature] Support Textual Inversion by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1822
- [Feature] Support stable diffusion inpaint by @Taited in https://github.com/open-mmlab/mmagic/pull/1976
- [Enhancement] Adopt `BaseModule` for some models by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1543
- [MMSIG] DeblurGANv2 inference by @xiaomile in https://github.com/open-mmlab/mmagic/pull/1955
- [CodeCamp2023-647] Add new configs of EG3D by @RangeKing in https://github.com/open-mmlab/mmagic/pull/1985

Bug Fixes

- Fix dtype error in StableDiffusion and DreamBooth training by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1879
- Fix gui VideoSlider bug by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1885
- Fix init_model and glide demo by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1888

- Fix InstColorization bug when dim=3 by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1901
- Fix sd and controlnet fp16 bugs by @Z-Fran in https://github.com/open-mmlab/mmagic/pull/1914
- Fix num_images_per_prompt in controlnet by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1936
- Revise metafile for sd-inpainting to fix inferencer init by @LeoXing1996 in https://github.com/open-mmlab/mmagic/pull/1995

New Contributors

- @wyyang23 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1886
- @yehuixie made their first contribution in https://github.com/open-mmlab/mmagic/pull/1912
- @crazysteeaam made their first contribution in https://github.com/open-mmlab/mmagic/pull/1920
- @BUPT-NingXinyu made their first contribution in https://github.com/open-mmlab/mmagic/pull/1921
- @zhjunqin made their first contribution in https://github.com/open-mmlab/mmagic/pull/1918
- @xuesheng1031 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1923
- @wslgqq277g made their first contribution in https://github.com/open-mmlab/mmagic/pull/1934
- @LYMDLUT made their first contribution in https://github.com/open-mmlab/mmagic/pull/1933
- @RangeKing made their first contribution in https://github.com/open-mmlab/mmagic/pull/1930
- @xin-li-67 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1932
- @chg0901 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1931
- @aptsunny made their first contribution in https://github.com/open-mmlab/mmagic/pull/1939
- @YanxingLiu made their first contribution in https://github.com/open-mmlab/mmagic/pull/1943
- @tackhwa made their first contribution in https://github.com/open-mmlab/mmagic/pull/1937
- @Geo-Chou made their first contribution in https://github.com/open-mmlab/mmagic/pull/1940
- @qsun1 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1956
- @thththh888 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1961
- @sijiua made their first contribution in https://github.com/open-mmlab/mmagic/pull/1967
- @MING-ZCH made their first contribution in https://github.com/open-mmlab/mmagic/pull/1982
- @AllYoung made their first contribution in https://github.com/open-mmlab/mmagic/pull/1996

1.76.3 v1.0.1 (26/05/2023)

New Features & Improvements

- Support tomesd for StableDiffusion speed-up. #1801
- Support all inpainting/matting/image restoration models inferencer. #1833, #1873
- Support animated drawings at projects. #1837
- Support Style-Based Global Appearance Flow for Virtual Try-On at projects. #1786
- Support tokenizer wrapper and support EmbeddingLayerWithFixe. #1846

Bug Fixes

- Fix install requirements. #1819

- Fix inst-colorization PackInputs. #1828, #1827
- Fix inferencer in pip-install. #1875

New Contributors

- @XDUWQ made their first contribution in https://github.com/open-mmlab/mmagic/pull/1830
- @FerryHuang made their first contribution in https://github.com/open-mmlab/mmagic/pull/1786
- @bobo0810 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1851
- @jercylew made their first contribution in https://github.com/open-mmlab/mmagic/pull/1874

1.76.4 v1.0.0 (25/04/2023)

We are excited to announce the release of MMagic v1.0.0 that inherits from MMEditing and MMGeneration.

Since its inception, MMEditing has been the preferred algorithm library for many super-resolution, editing, and generation tasks, helping research teams win more than 10 top international competitions and supporting over 100 GitHub ecosystem projects. After iterative updates with OpenMMLab 2.0 framework and merged with MMGeneration, MMEditing has become a powerful tool that supports low-level algorithms based on both GAN and CNN.

Today, MMEditing embraces Generative AI and transforms into a more advanced and comprehensive AIGC toolkit: **MMagic** (**M**ultimodal **A**dvanced, **G**enerative, and **I**ntelligent **C**reation).

In MMagic, we have supported 53+ models in multiple tasks such as fine-tuning for stable diffusion, text-to-image, image and video restoration, super-resolution, editing and generation. With excellent training and experiment management support from MMEngine, MMagic will provide more agile and flexible experimental support for researchers and AIGC enthusiasts, and help you on your AIGC exploration journey. With MMagic, experience more magic in generation! Let's open a new era beyond editing together. More than Editing, Unlock the Magic!

Highlights

1. New Models

We support 11 new models in 4 new tasks.

- Text2Image / Diffusion
  - ControlNet
  - DreamBooth
  - Stable Diffusion

  - Disco Diffusion
  - GLIDE
  - Guided Diffusion
- 3D-aware Generation
  - EG3D
- Image Restoration
  - NAFNet
  - Restormer
  - SwinIR
- Image Colorization
  - InstColorization

https://user-images.githubusercontent.com/49083766/233564593-7d3d48ed-e843-4432-b610-35e3d257765c.mp4

2. Magic Diffusion Model

For the Diffusion Model, we provide the following "magic" :

- Support image generation based on Stable Diffusion and Disco Diffusion.
- Support Finetune methods such as Dreambooth and DreamBooth LoRA.

- Support controllability in text-to-image generation using ControlNet.
- Support acceleration and optimization strategies based on xFormers to improve training and inference efficiency.
- Support video generation based on MultiFrame Render. MMagic supports the generation of long videos in various styles through ControlNet and MultiFrame Render. prompt keywords: a handsome man, silver hair, smiling, play basketball

  https://user-images.githubusercontent.com/12782558/227149757-fd054d32-554f-45d5-9f09-319184866d85.mp4

  prompt keywords: a girl, black hair, white pants, smiling, play basketball

  https://user-images.githubusercontent.com/49083766/233559964-bd5127bd-52f6-44b6-a089-9d7adfbc2430.mp4

prompt keywords: a handsome man

https://user-images.githubusercontent.com/12782558/227152129-d70d5f76-a6fc-4d23-97d1-a94abd08f95a.mp4

- Support calling basic models and sampling strategies through DiffuserWrapper.
- SAM + MMagic = Generate Anything SAM (Segment Anything Model) is a popular model these days and can also provide more support for MMagic! If you want to create your own animation, you can go to OpenMMLab PlayGround.

  https://user-images.githubusercontent.com/49083766/233562228-f39fc675-326c-4ae8-986a-c942059effd0.mp4

3. Upgraded Framework

To improve your "spellcasting" efficiency, we have made the following adjustments to the "magic circuit":

- By using MMEngine and MMCV of OpenMMLab 2.0 framework, We decompose the editing framework into different modules and one can easily construct a customized editor framework by combining different modules. We can define the training process just like playing with Legos and provide rich components and strategies. In MMagic, you can complete controls on the training process with different levels of APIs.
- Support for 33+ algorithms accelerated by Pytorch 2.0.
- Refactor DataSample to support the combination and splitting of batch dimensions.
- Refactor DataPreprocessor and unify the data format for various tasks during training and inference.
- Refactor MultiValLoop and MultiTestLoop, supporting the evaluation of both generation-type metrics (e.g. FID) and reconstruction-type metrics (e.g. SSIM), and supporting the evaluation of multiple datasets at once.
- Support visualization on local files or using tensorboard and wandb.

New Features & Improvements

- Support 53+ algorithms, 232+ configs, 213+ checkpoints, 26+ loss functions, and 20+ metrics.
- Support controlnet animation and Gradio gui. Click to view.
- Support Inferencer and Demo using High-level Inference APIs. Click to view.
- Support Gradio gui of Inpainting inference. Click to view.
- Support qualitative comparison tools. Click to view.
- Enable projects. Click to view.
- Improve converters scripts and documents for datasets. Click to view.

1.76.5 v1.0.0rc7 (07/04/2023)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc7. This release supports 51+ models, 226+ configs and 212+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- Support DiffuserWrapper
- Support ControlNet (training and inference).
- Support PyTorch 2.0.

New Features & Improvements

- Support DiffuserWrapper. #1692

- Support ControlNet (training and inference). #1744
- Support PyTorch 2.0 (successfully compile 33+ models on 'inductor' backend). #1742
- Support Image Super-Resolution and Video Super-Resolution models inferencer. #1662, #1720
- Refactor tools/get_flops script. #1675
- Refactor dataset_converters and documents for datasets. #1690
- Move stylegan ops to MMCV. #1383

Bug Fixes

- Fix disco inferencer. #1673
- Fix nafnet optimizer config. #1716
- Fix tof typo. #1711

Contributors

A total of 8 developers contributed to this release. Thanks @LeoXing1996, @Z-Fran, @plyfager, @zengyh1900, @liuwenran, @ryanxingql, @HAOCHENYE, @VongolaWu

New Contributors

- @HAOCHENYE made their first contribution in https://github.com/open-mmlab/mmagic/pull/1712

1.76.6 v1.0.0rc6 (02/03/2023)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc6. This release supports 50+ models, 222+ configs and 209+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- Support Gradio gui of Inpainting inference.
- Support Colorization, Translationin and GAN models inferencer.

New Features & Improvements

- Refactor FileIO. #1572
- Refactor registry. #1621
- Refactor Random degradations. #1583
- Refactor DataSample, DataPreprocessor, Metric and Loop. #1656
- Use mmengine.basemodule instead of nn.module. #1491
- Refactor Main Page. #1609
- Support Gradio gui of Inpainting inference. #1601
- Support Colorization inferencer. #1588
- Support Translation models inferencer. #1650
- Support GAN models inferencer. #1653, #1659
- Print config tool. #1590
- Improve type hints. #1604
- Update Chinese documents of metrics and datasets. #1568, #1638
- Update Chinese documents of BigGAN and Disco-Diffusion. #1620

- Update Evaluation and README of Guided-Diffusion. #1547

Bug Fixes

- Fix the meaning of `momentum` in EMA. #1581
- Fix output dtype of RandomNoise. #1585
- Fix pytorch2onnx tool. #1629
- Fix API documents. #1641, #1642
- Fix loading RealESRGAN EMA weights. #1647
- Fix arg passing bug of dataset_converters scripts. #1648

Contributors

A total of 17 developers contributed to this release. Thanks @plyfager, @LeoXing1996, @Z-Fran, @zengyh1900, @VongolaWu, @liuwenran, @austinmw, @dienachtderwelt, @liangzelong, @i-aki-y, @xiaomile, @Li-Qingyun, @vansin, @Luo-Yihang, @ydengbi, @ruoningYu, @triple-Mu

New Contributors

- @dienachtderwelt made their first contribution in https://github.com/open-mmlab/mmagic/pull/1578
- @i-aki-y made their first contribution in https://github.com/open-mmlab/mmagic/pull/1590
- @triple-Mu made their first contribution in https://github.com/open-mmlab/mmagic/pull/1618
- @Li-Qingyun made their first contribution in https://github.com/open-mmlab/mmagic/pull/1640
- @Luo-Yihang made their first contribution in https://github.com/open-mmlab/mmagic/pull/1648
- @ydengbi made their first contribution in https://github.com/open-mmlab/mmagic/pull/1557

1.76.7 v1.0.0rc5 (04/01/2023)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc5. This release supports 49+ models, 180+ configs and 177+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- Support Restormer.
- Support GLIDE.
- Support SwinIR.
- Support Stable Diffusion.

New Features & Improvements

- Disco notebook. (#1507)
- Revise test requirements and CI. (#1514)
- Recursive generate summary and docstring. (#1517)
- Enable projects. (#1526)
- Support mscoco dataset. (#1520)
- Improve Chinese documents. (#1532)
- Type hints. (#1481)
- Update download link of checkpoints. (#1554)

- Update deployment guide. (#1551)

Bug Fixes

- Fix documentation link checker. (#1522)
- Fix ssim first channel bug. (#1515)
- Fix extract_gt_data of realesrgan. (#1542)
- Fix model index. (#1559)
- Fix config path in disco-diffusion. (#1553)
- Fix text2image inferencer. (#1523)

Contributors

A total of 16 developers contributed to this release. Thanks @plyfager, @LeoXing1996, @Z-Fran, @zengyh1900, @VongolaWu, @liuwenran, @AlexZou14, @lvhan028, @xiaomile, @ldr426, @austin273, @whu-lee, @willaty, @curiosity654, @Zdafeng, @Taited

New Contributors

- @xiaomile made their first contribution in https://github.com/open-mmlab/mmagic/pull/1481
- @ldr426 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1542
- @austin273 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1553
- @whu-lee made their first contribution in https://github.com/open-mmlab/mmagic/pull/1539
- @willaty made their first contribution in https://github.com/open-mmlab/mmagic/pull/1541
- @curiosity654 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1556
- @Zdafeng made their first contribution in https://github.com/open-mmlab/mmagic/pull/1476
- @Taited made their first contribution in https://github.com/open-mmlab/mmagic/pull/1534

1.76.8 v1.0.0rc4 (05/12/2022)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc4. This release supports 45+ models, 176+ configs and 175+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- Support High-level APIs.
- Support diffusion models.
- Support Text2Image Task.
- Support 3D-Aware Generation.

New Features & Improvements

- Refactor High-level APIs. (#1410)
- Support disco-diffusion text-2-image. (#1234, #1504)
- Support EG3D. (#1482, #1493, #1494, #1499)
- Support NAFNet model. (#1369)

Bug Fixes

- fix srgan train config. (#1441)

- fix cain config. (#1404)
- fix rdn and srcnn train configs. (#1392)

Contributors

A total of 14 developers contributed to this release. Thanks @plyfager, @LeoXing1996, @Z-Fran, @zengyh1900, @VongolaWu, @gaoyang07, @ChangjianZhao, @zxczrx123, @jackghosts, @liuwenran, @CCODING04, @RoseZhao929, @shaocongliu, @liangzelong.

New Contributors

- @gaoyang07 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1372
- @ChangjianZhao made their first contribution in https://github.com/open-mmlab/mmagic/pull/1461
- @zxczrx123 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1462
- @jackghosts made their first contribution in https://github.com/open-mmlab/mmagic/pull/1463
- @liuwenran made their first contribution in https://github.com/open-mmlab/mmagic/pull/1410
- @CCODING04 made their first contribution in https://github.com/open-mmlab/mmagic/pull/783
- @RoseZhao929 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1474
- @shaocongliu made their first contribution in https://github.com/open-mmlab/mmagic/pull/1470
- @liangzelong made their first contribution in https://github.com/open-mmlab/mmagic/pull/1488

1.76.9 v1.0.0rc3 (10/11/2022)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc3. This release supports 43+ models, 170+ configs and 169+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- convert `mmdet` and `clip` to optional requirements.

New Features & Improvements

- Support `try_import` for `mmdet`. (#1408)
- Support `try_import` for `flip`. (#1420)
- Update `.gitignore`. ($1416)
- Set `real_feat` to cpu in `inception_utils`. (#1415)
- Modify README and configs of StyleGAN2 and PEGAN (#1418)
- Improve the rendering of Docs-API (#1373)

Bug Fixes

- Revise config and pretrain model loading in ESRGAN (#1407)
- Revise config of LSGAN (#1409)
- Revise config of CAIN (#1404)

Contributors

A total of 5 developers contributed to this release. @Z-Fran, @zengyh1900, @plyfager, @LeoXing1996, @ruoningYu.

1.76.10 v1.0.0rc2 (02/11/2022)

Highlights

We are excited to announce the release of MMEditing 1.0.0rc2. This release supports 43+ models, 170+ configs and 169+ checkpoints in MMGeneration and MMEditing. We highlight the following new features

- patch-based and slider-based image and video comparison viewer.
- image colorization.

New Features & Improvements

- Support qualitative comparison tools. (#1303)
- Support instance aware colorization. (#1370)
- Support multi-metrics with different sample-model. (#1171)
- Improve the implementation
  - refactoring evaluation metrics. (#1164)
  - Save gt images in PGGAN's `forward`. (#1332)
  - Improve type and change default number of `preprocess_div2k_dataset.py`. (#1380)
  - Support pixel value clip in visualizer. (#1365)
  - Support SinGAN Dataset and SinGAN demo. (#1363)
  - Avoid cast int and float in GenDataPreprocessor. (#1385)
- Improve the documentation
  - Update a menu switcher. (#1162)
  - Fix TTSR's README. (#1325)

Bug Fixes

- Fix PPL bug. (#1172)
- Fix RDN number of channels. (#1328)
- Fix types of exceptions in demos. (#1372)
- Fix realesrgan ema. (#1341)
- Improve the assertion to ensuer `GenerateFacialHeatmap` as `np.float32`. (#1310)
- Fix sampling behavior of `unpaired_dataset.py` and urls in cyclegan's README. (#1308)
- Fix vsr models in pytorch2onnx. (#1300)
- Fix incorrect settings in configs. (#1167,#1200,#1236,#1293,#1302,#1304,#1319,#1331,#1336,#1349,#1352,#1353,#1358,#1364,

New Contributors

- @gaoyang07 made their first contribution in https://github.com/open-mmlab/mmagic/pull/1372

Contributors

A total of 7 developers contributed to this release. Thanks @LeoXing1996, @Z-Fran, @zengyh1900, @plyfager, @ryanxingql, @ruoningYu, @gaoyang07.

1.76.11 v1.0.0rc1(23/9/2022)

MMEditing 1.0.0rc1 has merged MMGeneration 1.x.

- Support 42+ algorithms, 169+ configs and 168+ checkpoints.
- Support 26+ loss functions, 20+ metrics.
- Support tensorboard, wandb.
- Support unconditional GANs, conditional GANs, image2image translation and internal learning.

1.76.12 v1.0.0rc0(31/8/2022)

MMEditing 1.0.0rc0 is the first version of MMEditing 1.x, a part of the OpenMMLab 2.0 projects.

Built upon the new training engine, MMEditing 1.x unifies the interfaces of dataset, models, evaluation, and visualization.

And there are some BC-breaking changes. Please check the migration tutorial for more details.

1.77 mmagic.apis.inferencers

1.77.1 Package Contents

Classes

ColorizationInferencer	inferencer that predicts with colorization models.
ConditionalInferencer	inferencer that predicts with conditional models.
ControlnetAnimationInferencer	Base inferencer.
DiffusersPipelineInferencer	inferencer that predicts with text2image models.
EG3DInferencer	Base inferencer.
ImageSuperResolutionInferencer	inferencer that predicts with restoration models.
InpaintingInferencer	inferencer that predicts with inpainting models.
MattingInferencer	inferencer that predicts with matting models.
Text2ImageInferencer	inferencer that predicts with text2image models.
TranslationInferencer	inferencer that predicts with translation models.
UnconditionalInferencer	inferencer that predicts with unconditional models.
VideoInterpolationInferencer	inferencer that predicts with video interpolation models.
VideoRestorationInferencer	inferencer that predicts with video restoration models.

class mmagic.apis.inferencers.**ColorizationInferencer**(*config: Union[mmagic.utils.ConfigType, str], ckpt: Optional[str], device: Optional[str] = None, extra_parameters: Optional[Dict] = None, seed: int = 2022, **kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with colorization models.

func_kwargs

preprocess(*img: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters **img** (*InputsType*) – Image to be translated by models.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**ConditionalInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with conditional models.

func_kwargs

extra_parameters

preprocess(*label: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters **label** (*InputsType*) – Input label for condition models.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.

- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

_pred2dict(*data_sample:* mmagic.structures.DataSample) → Dict

Extract elements necessary to represent a prediction into a dictionary. It's better to contain only basic data elements such as strings and numbers in order to guarantee it's json-serializable.

Parameters **data_sample** (*DataSample*) – The data sample to be converted.

Returns The output dictionary.

Return type dict

class mmagic.apis.inferencers.**ControlnetAnimationInferencer**(*config: Union[mmagic.utils.ConfigType, str], device: Optional[str] = None, extra_parameters: Optional[Dict] = None, dtype=torch.float32, **kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

Base inferencer.

Parameters

- **config** (*str or ConfigType*) – Model config or the path to it.
- **ckpt** (*str, optional*) – Path to the checkpoint.
- **device** (*str, optional*) – Device to run inference. If None, the best device will be automatically used.
- **result_out_dir** (*str*) – Output directory of images. Defaults to ''.

func_kwargs

func_order

extra_parameters

__call__(*prompt=None, video=None, negative_prompt=None, controlnet_conditioning_scale=0.7, image_width=512, image_height=512, save_path=None, strength=0.75, num_inference_steps=20, seed=1, output_fps=None, reference_img=None, **kwargs*) → Union[Dict, List[Dict]]

Call the inferencer.

Parameters **kwargs** – Keyword arguments for the inferencer.

Returns Results of inference pipeline.

Return type Union[Dict, List[Dict]]

class mmagic.apis.inferencers.**DiffusersPipelineInferencer**(*config: Union[mmagic.utils.ConfigType, str], ckpt: Optional[str], device: Optional[str] = None, extra_parameters: Optional[Dict] = None, seed: int = 2022, **kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with text2image models.

func_kwargs

preprocess(*text: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = None*, *negative_prompt: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = None*, *num_inference_steps: int = 20*, *height=None*, *width=None*) → Dict

Process the inputs into a model-feedable format.

Parameters

- **text** (*InputsType*) – text input for text-to-image model.
- **negative_prompt** (*InputsType*) – negative prompt.

Returns Results of preprocess.

Return type result(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**EG3DInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

Base inferencer.

Parameters

- **config** (*str or ConfigType*) – Model config or the path to it.
- **ckpt** (*str, optional*) – Path to the checkpoint.
- **device** (*str, optional*) – Device to run inference. If None, the best device will be automatically used.
- **extra_parameters** (*Dict, optional*) – Extra parameters for different models in inference stage.
- **seed** (*str, optional*) – Seed for inference.

func_kwargs

extra_parameters

preprocess(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = None*) → mmagic.utils.ForwardInputs

Process the inputs into a model-feedable format.

Parameters **inputs** (*List[Union[str, np.ndarray]]*) – The conditional inputs for the inferencer. Defaults to None.

Returns The preprocessed inputs and data samples.

Return type ForwardInputs

forward(*inputs: mmagic.utils.ForwardInputs*, *interpolation: Optional[str] = 'both'*, *num_images: int = 100*) → Union[dict, List[dict]]

Forward the inputs to the model.

Parameters

- **inputs** (*ForwardInputs*) – Model inputs. If data sample (the second element of *inputs*) is not passed, will generate a sequence of images corresponding to passed *interpolation* mode.
- **interpolation** (*str*) – The interpolation mode. Supported choices are 'both', 'conditioning', and 'camera'. Defaults to 'both'.
- **num_images** (*int*) – The number of frames of interpolation. Defaults to 500.

Returns

Output dict corresponds to the input condition or the list of output dict of each frame during the interpolation process.

Return type Union[dict, List[dict]]

visualize(*preds: Union[mmagic.apis.inferencers.base_mmagic_inferencer.PredType, List[mmagic.apis.inferencers.base_mmagic_inferencer.PredType]]*, *vis_mode: str = 'both'*, *save_img: bool = True*, *save_video: bool = True*, *img_suffix: str = '.png'*, *video_suffix: str = '.mp4'*, *result_out_dir: str = 'eg3d_output'*) → None

Visualize predictions.

Parameters

- **preds** (*Union[PredType, List[PredType]]*) – Prediction os model.
- **vis_mode** (*str, optional*) – Which output to visualize. Supported choices are 'both', 'depth', and 'img'. Defaults to 'all'.
- **save_img** (*bool, optional*) – Whether save images. Defaults to True.
- **save_video** (*bool, optional*) – Whether save videos. Defaults to True.
- **img_suffix** (*str, optional*) – The suffix of saved images. Defaults to '.png'.
- **video_suffix** (*str, optional*) – The suffix of saved videos. Defaults to '.mp4'.
- **result_out_dir** (*str, optional*) – The save director of image and videos. Defaults to 'eg3d_output'.

preprocess_img(*preds: List[dict]*) → torch.Tensor

Preprocess images in the predictions.

Parameters **preds** (*List[dict]*) – List of prediction dict of each frame.

Returns

Preprocessed image tensor shape like [num_frame * bz, 3, H, W].

Return type torch.Tensor

preprocess_depth(*preds: List[dict]*) → torch.Tensor

Preprocess depth in the predictions.

Parameters **preds** (*List[dict]*) – List of prediction dict of each frame.

Returns

Preprocessed depth tensor shape like [num_frame * bz, 3, H, W].

Return type torch.Tensor

postprocess(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *imgs: Optional[List[numpy.ndarray]] = None*, *is_batch: bool = False*, *get_datasample: bool = False*) → Dict[str, torch.tensor]

Postprocess predictions.

Parameters

- **preds** (*List[Dict]*) – Predictions of the model.
- **imgs** (*Optional[np.ndarray]*) – Visualized predictions.
- **is_batch** (*bool*) – Whether the inputs are in a batch. Defaults to False.
- **get_datasample** (*bool*) – Whether to use Datasample to store inference results. If False, dict will be used.

Returns Inference results as a dict.

Return type Dict[str, torch.Tensor]

class mmagic.apis.inferencers.**ImageSuperResolutionInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with restoration models.

func_kwargs

preprocess(*img: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*, *ref: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = None*) → Dict

Process the inputs into a model-feedable format.

Parameters

- **img** (*InputsType*) – Image to be restored by models.
- **ref** (*InputsType*) – Reference image for restoration models. Defaults to None.

Returns Results of preprocess.

Return type data(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**InpaintingInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with inpainting models.

func_kwargs

_init_pipeline(*cfg*) → mmengine.dataset.Compose

Initialize the test pipeline.

preprocess(*img: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*, *mask: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters

- **img** (*InputsType*) – Image to be inpainted by models.
- **mask** (*InputsType*) – Image mask for inpainting models.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Mask of input image.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**MattingInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with matting models.

func_kwargs

preprocess(*img: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*, *trimap: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters

- **img** (*InputsType*) – Image to be processed by models.
- **mask** (*InputsType*) – Mask corresponding to the input image.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

_pred2dict(*data_sample:* mmagic.structures.DataSample) → Dict

Extract elements necessary to represent a prediction into a dictionary. It's better to contain only basic data elements such as strings and numbers in order to guarantee it's json-serializable.

Parameters **data_sample** (*DataSample*) – The data sample to be converted.

Returns The output dictionary.

Return type dict

class mmagic.apis.inferencers.**Text2ImageInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with text2image models.

func_kwargs

extra_parameters

preprocess(*text: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*, *control: str = None*, *negative_prompt: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = None*) → Dict

Process the inputs into a model-feedable format.

Parameters

- **text** (*InputsType*) – text input for text-to-image model.
- **control** (*str*) – control img dir for controlnet.
- **negative_prompt** (*InputsType*) – negative prompt.

Returns Results of preprocess.

Return type result(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**TranslationInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with translation models.

func_kwargs

preprocess(*img: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters **img** (*InputsType*) – Image to be translated by models.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = None*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

class mmagic.apis.inferencers.**UnconditionalInferencer**(*config: Union[mmagic.utils.ConfigType, str]*, *ckpt: Optional[str]*, *device: Optional[str] = None*, *extra_parameters: Optional[Dict] = None*, *seed: int = 2022*, ***kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with unconditional models.

func_kwargs

extra_parameters

preprocess() → Dict

Process the inputs into a model-feedable format.

Returns Results of preprocess.

Return type results(Dict)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = ''*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

_pred2dict(*data_sample:* mmagic.structures.DataSample) → Dict

Extract elements necessary to represent a prediction into a dictionary. It's better to contain only basic data elements such as strings and numbers in order to guarantee it's json-serializable.

Parameters **data_sample** (*DataSample*) – The data sample to be converted.

Returns The output dictionary.

Return type dict

class mmagic.apis.inferencers.**VideoInterpolationInferencer**(*config: Union[mmagic.utils.ConfigType, str], ckpt: Optional[str], device: Optional[str] = None, extra_parameters: Optional[Dict] = None, seed: int = 2022, **kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with video interpolation models.

func_kwargs

extra_parameters

preprocess(*video: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters **video** (*InputsType*) – Video to be interpolated by models.

Returns Video to be interpolated by models.

Return type video(InputsType)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType, result_out_dir: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType = ''*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

Parameters

- **inputs** (*InputsType*) – Input video directory.
- **result_out_dir** (*str*) – Output directory of video. Defaults to ''.

Returns Result of forwarding

Return type PredType

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType, result_out_dir: str = ''*) → List[numpy.ndarray]

Visualize is not needed in this inferencer.

postprocess(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType, imgs: Optional[List[numpy.ndarray]] = None*) → Union[mmagic.apis.inferencers.base_mmagic_inferencer.ResType, Tuple[mmagic.apis.inferencers.base_mmagic_inferencer.ResType, numpy.ndarray]]

Postprocess is not needed in this inferencer.

class mmagic.apis.inferencers.**VideoRestorationInferencer**(*config: Union[mmagic.utils.ConfigType, str], ckpt: Optional[str], device: Optional[str] = None, extra_parameters: Optional[Dict] = None, seed: int = 2022, **kwargs*)

Bases: mmagic.apis.inferencers.base_mmagic_inferencer.BaseMMagicInferencer

inferencer that predicts with video restoration models.

func_kwargs

extra_parameters

preprocess(*video: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → Dict

Process the inputs into a model-feedable format.

Parameters **video** (*InputsType*) – Video to be restored by models.

Returns Results of preprocess.

Return type results(InputsType)

forward(*inputs: mmagic.apis.inferencers.base_mmagic_inferencer.InputsType*) → mmagic.apis.inferencers.base_mmagic_inferencer.PredType

Forward the inputs to the model.

Parameters **inputs** (*InputsType*) – Images array of input video.

Returns Results of forwarding

Return type PredType

visualize(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *result_out_dir: str = ''*) → List[numpy.ndarray]

Visualize predictions.

Parameters

- **preds** (*List[Union[str, np.ndarray]]*) – Forward results by the inferencer.
- **data** (*List[Dict]*) – Not needed by this kind of inferencer.
- **result_out_dir** (*str*) – Output directory of image. Defaults to ''.

Returns Result of visualize

Return type List[np.ndarray]

postprocess(*preds: mmagic.apis.inferencers.base_mmagic_inferencer.PredType*, *imgs: Optional[List[numpy.ndarray]] = None*) → Union[mmagic.apis.inferencers.base_mmagic_inferencer.ResType, Tuple[mmagic.apis.inferencers.base_mmagic_inferencer.ResType, numpy.ndarray]]

Postprocess is not needed in this inferencer.

1.78 mmagic.structures

1.78.1 Package Contents

Classes

DataSample	A data structure interface of MMagic. They are used as interfaces

class mmagic.structures.**DataSample**(*, *metainfo: Optional[dict] = None*, ***kwargs*)

Bases: mmengine.structures.BaseDataElement

A data structure interface of MMagic. They are used as interfaces between different components, e.g., model, visualizer, evaluator, etc. Typically, DataSample contains all the information and data from ground- truth and predictions.

***DataSample* inherits from *BaseDataElement*. See more details in:** https://mmengine.readthedocs.io/en/latest/advanced_tutorials/data_element.html Specifically, an instance of BaseDataElement consists of two components, - `metainfo`, which contains some meta information,

e.g., *img_shape*, *img_id*, *color_order*, etc.

- `data`, which contains the data used in the loop.

The attributes in `DataSample` are divided into several parts:

- `gt_img`: Ground truth image(s).
- `pred_img`: Image(s) of model predictions.
- `ref_img`: Reference image(s).
- `mask`: Mask in Inpainting.
- `trimap`: Trimap in Matting.
- `gt_alpha`: Ground truth alpha image in Matting.
- `pred_alpha`: Predicted alpha image in Matting.
- `gt_fg`: Ground truth foreground image in Matting.
- `pred_fg`: Predicted foreground image in Matting.
- `gt_bg`: Ground truth background image in Matting.
- `pred_bg`: Predicted background image in Matting.
- `gt_merged`: Ground truth merged image in Matting.

Examples:

```
>>> import torch
>>> import numpy as np
>>> from mmagic.structures import DataSample
>>> img_meta = dict(img_shape=(800, 1196, 3))
>>> img = torch.rand((3, 800, 1196))
>>> data_sample = DataSample(gt_img=img, metainfo=img_meta)
>>> assert 'img_shape' in data_sample.metainfo_keys()
>>> data_sample
<DataSample(

    META INFORMATION
    img_shape: (800, 1196, 3)

    DATA FIELDS
    gt_img: tensor(...)
) at 0x1f6a5a99a00>
```

We also support *stack* and *split* operation to handle a batch of data samples:

```
>>> import torch
>>> import numpy as np
>>> from mmagic.structures import DataSample
>>> img_meta1 = img_meta2 = dict(img_shape=(800, 1196, 3))
>>> img1 = torch.rand((3, 800, 1196))
```

(continues on next page)

(continued from previous page)

```
>>> img2 = torch.rand((3, 800, 1196))
>>> data_sample1 = DataSample(gt_img=img1, metainfo=img_meta1)
>>> data_sample2 = DataSample(gt_img=img2, metainfo=img_meta1)
```

```
>>> # stack them and then use as batched-tensor!
>>> data_sample = DataSample.stack([data_sample1, data_sample2])
>>> print(data_sample.gt_img.shape)
torch.Size([2, 3, 800, 1196])
>>> print(data_sample.metainfo)
{'img_shape': [(800, 1196, 3), (800, 1196, 3)]}
```

```
>>> # split them if you want
>>> data_sample1_, data_sample2_ = data_sample.split()
>>> assert (data_sample1_.gt_img == img1).all()
>>> assert (data_sample2_.gt_img == img2).all()
```

property gt_label

This the function to fetch gt label.

Returns gt label.

Return type LabelData

META_KEYS

DATA_KEYS

set_predefined_data(*data: dict*) → None

set or change pre-defined key-value pairs in `data_field` by parameter `data`.

Parameters **data** (*dict*) – A dict contains annotations of image or model predictions.

set_tensor_data(*data: dict*) → None

convert input data to tensor, and then set or change key-value pairs in `data_field` by parameter `data`.

Parameters **data** (*dict*) – A dict contains annotations of image or model predictions.

set_gt_label(*value: Union[numpy.ndarray, torch.Tensor, Sequence[numbers.Number], numbers.Number]*) → *DataSample*

Set label of `gt_label`.

gt_label()

Delete gt label.

classmethod stack(*data_samples: Sequence[*DataSample*]*) → *DataSample*

Stack a list of data samples to one. All tensor fields will be stacked at first dimension. Otherwise the values will be saved in a list.

Parameters **data_samples** (*Sequence['DataSample']*) – A sequence of *DataSample* to stack.

Returns The stacked data sample.

Return type *DataSample*

split(*allow_nonseq_value: bool = False*) → Sequence[*DataSample*]

Split a sequence of data sample in the first dimension.

Parameters **allow_nonseq_value** (*bool*) – Whether allow non-sequential data in split operation. If True, non-sequential data will be copied for all split data samples. Otherwise, an error will be raised. Defaults to False.

Returns The list of data samples after splitting.

Return type Sequence[*DataSample*]

__len__()

Get the length of the data sample.

1.79 mmagic.datasets

1.79.1 Package Contents

Classes

BasicConditionalDataset	Custom dataset for conditional GAN. This class is based on the
BasicFramesDataset	BasicFramesDataset for open source projects in OpenMMLab/MMagic.
BasicImageDataset	BasicImageDataset for open source projects in OpenMMLab/MMagic.
CIFAR10	CIFAR10 Dataset.
AdobeComp1kDataset	Adobe composition-1k dataset.
ControlNetDataset	Demo dataset to test ControlNet. Modified from https://github.com/lllyas
DreamBoothDataset	Dataset for DreamBooth.
GrowScaleImgDataset	Grow Scale Unconditional Image Dataset.
ImageNet	ImageNet Dataset.
MSCoCoDataset	MSCoCo 2014 dataset.
PairedImageDataset	General paired image folder dataset for image generation.
SinGANDataset	SinGAN Dataset.
TextualInversionDataset	Dataset for Textual Inversion and ViCo.
UnpairedImageDataset	General unpaired image folder dataset for image generation.

class mmagic.datasets.**BasicConditionalDataset**(*ann_file: str = '', metainfo: Optional[dict] = None, data_root: str = '', data_prefix: Union[str, dict] = '', extensions: Sequence[str] = ('.jpg', '.jpeg', '.png', '.ppm', '.bmp', '.pgm', '.tif'), lazy_init: bool = False, classes: Union[str, Sequence[str], None] = None, **kwargs*)

Bases: mmengine.dataset.BaseDataset

Custom dataset for conditional GAN. This class is based on the combination of *BaseDataset* (https://github.com/open- mmlab/mmclassification/blob/main/mmcls/datasets/base_dataset.py) # noqa and *CustomDataset* (https://github.com/open- mmlab/mmclassification/blob/main/mmcls/datasets/custom.py). # noqa.

The dataset supports two kinds of annotation format.

1. A annotation file read by line (e.g., txt) is provided, and each line indicates a sample:

   The sample files:

```
data_prefix/
├── folder_1
│   ├── xxx.png
│   ├── xxy.png
│   └── ...
└── folder_2
    ├── 123.png
    ├── nsdf3.png
    └── ...
```

The annotation file (the first column is the image path and the second column is the index of category):

```
folder_1/xxx.png 0
folder_1/xxy.png 1
folder_2/123.png 5
folder_2/nsdf3.png 3
...
```

Please specify the name of categories by the argument `classes` or `metainfo`.

2. A dict-based annotation file (e.g., json) is provided, key and value indicate the path and label of the sample:

   The sample files:

```
data_prefix/
├── folder_1
│   ├── xxx.png
│   ├── xxy.png
│   └── ...
└── folder_2
    ├── 123.png
    ├── nsdf3.png
    └── ...
```

   The annotation file (the key is the image path and the value column is the label):

```
{
    "folder_1/xxx.png": [1, 2, 3, 4],
    "folder_1/xxy.png": [2, 4, 1, 0],
    "folder_2/123.png": [0, 9, 8, 1],
    "folder_2/nsdf3.png", [1, 0, 0, 2],
    ...
}
```

   In this kind of annotation, labels can be any type and not restricted to an index.

3. The samples are arranged in the specific way:

```
data_prefix/
├── class_x
│   ├── xxx.png
│   ├── xxy.png
│   └── ...
│       └── xxz.png
└── class_y
```

(continues on next page)

(continued from previous page)

```
    ├── 123.png
    ├── nsdf3.png
    ├── ...
    └── asd932_.png
```

If the `ann_file` is specified, the dataset will be generated by the first two ways, otherwise, try the third way.

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as class information. Defaults to None.
- **data_root** (*str*) – The root directory for `data_prefix` and `ann_file`. Defaults to ''.
- **data_prefix** (*str | dict*) – Prefix for the data. Defaults to ''.
- **extensions** (*Sequence[str]*) – A sequence of allowed extensions. Defaults to ('.jpg', '.jpeg', '.png', '.ppm', '.bmp', '.pgm', '.tif').
- **lazy_init** (*bool*) – Whether to load annotation during instantiation. In some cases, such as visualization, only the meta information of the dataset is needed, which is not necessary to load annotation file. `Basedataset` can skip load annotations to save time by set `lazy_init=False`. Defaults to False.
- ****kwargs** – Other keyword arguments in `BaseDataset`.

property img_prefix

The prefix of images.

property CLASSES

Return all categories names.

property class_to_idx

Map mapping class name to class index.

Returns mapping from class name to class index.

Return type dict

_find_samples(*file_backend*)

find samples from `data_prefix`.

load_data_list()

Load image paths and gt_labels.

is_valid_file(*filename: str*) → bool

Check if a file is a valid sample.

get_gt_labels()

Get all ground-truth labels (categories).

Returns categories for all images.

Return type np.ndarray

get_cat_ids(*idx: int*) → List[int]

Get category id by index.

Parameters **idx** (*int*) – Index of data.

Returns Image category of specified index.

Return type cat_ids (List[int])

_compat_classes(*metainfo*, *classes*)

Merge the old style `classes` arguments to `metainfo`.

full_init()

Load annotation file and set `BaseDataset._fully_initialized` to True.

__repr__()

Print the basic information of the dataset.

Returns Formatted string.

Return type str

extra_repr() → List[str]

The extra repr information of the dataset.

class mmagic.datasets.**BasicFramesDataset**(*ann_file: str = ''*, *metainfo: Optional[dict] = None*, *data_root: Optional[str] = None*, *data_prefix: dict = dict(img='')*, *pipeline: List[Union[dict, Callable]] = []*, *test_mode: bool = False*, *filename_tmpl: dict = dict()*, *search_key: Optional[str] = None*, *backend_args: Optional[dict] = None*, *depth: int = 1*, *num_input_frames: Optional[int] = None*, *num_output_frames: Optional[int] = None*, *fixed_seq_len: Optional[int] = None*, *load_frames_list: dict = dict()*, ***kwargs*)

Bases: `mmengine.dataset.BaseDataset`

BasicFramesDataset for open source projects in OpenMMLab/MMagic.

This dataset is designed for low-level vision tasks with frames, such as video super-resolution and video frame interpolation.

The annotation file is optional.

If use annotation file, the annotation format can be shown as follows.

```
Case 1 (Vid4):

    calendar 41
    city 34
    foliage 49
    walk 47

Case 2 (REDS):

    000/00000000.png (720, 1280, 3)
    000/00000001.png (720, 1280, 3)

Case 3 (Vimeo90k):

    00001/0266 (256, 448, 3)
    00001/0268 (256, 448, 3)
```

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as class information. Defaults to None.
- **data_root** (*str, optional*) – The root directory for `data_prefix` and `ann_file`. Defaults to None.
- **data_prefix** (*dict, optional*) – Prefix for training data. Defaults to dict(img='', gt='').
- **pipeline** (*list, optional*) – Processing pipeline. Defaults to [].
- **test_mode** (*bool, optional*) – `test_mode=True` means in test phase. Defaults to False.
- **filename_tmpl** (*str*) – Template for each filename. Note that the template excludes the file extension. Default: '{}'.
- **search_key** (*str*) – The key used for searching the folder to get data_list. Default: 'gt'.
- **backend_args** (*dict, optional*) – Arguments to instantiate the prefix of uri corresponding backend. Defaults to None.
- **depth** (*int*) – The depth of path. Default: 1
- **num_input_frames** (*None | int*) – Number of input frames. Default: None.
- **num_output_frames** (*None | int*) – Number of output frames. Default: None.
- **fixed_seq_len** (*None | int*) – The fixed sequence length. If None, BasicFramesDataset will obtain the length of each sequence. Default: None.
- **load_frames_list** (*dict*) – Load frames list for each key. Default: dict().

Examples

Assume the file structure as the following:

mmagic (root) ├── mmagic ├── tools ├── configs ├── data │ ├── Vid4 │ │ ├── BIx4 │ │ │ ├── city │ │ │ │ ├── img1.png │ │ ├── GT │ │ │ ├── city │ │ │ │ ├── img1.png │ │ ├── meta_info_Vid4_GT.txt │ ├── places │ │ ├── sequences │ │ │ ├── 00001 │ │ │ │ ├── 0389 │ │ │ │ │ ├── img1.png │ │ │ │ │ ├── img2.png │ │ │ │ │ ├── img3.png │ │ ├── tri_trainlist.txt

Case 1: Loading Vid4 dataset for training a VSR model.

```
dataset = BasicFramesDataset(
    ann_file='meta_info_Vid4_GT.txt',
    metainfo=dict(dataset_type='vid4', task_name='vsr'),
    data_root='data/Vid4',
    data_prefix=dict(img='BIx4', gt='GT'),
    pipeline=[],
    depth=2,
    num_input_frames=5)
```

Case 2: Loading Vimeo90k dataset for training a VFI model.

```
dataset = BasicFramesDataset(
    ann_file='tri_trainlist.txt',
    metainfo=dict(dataset_type='vimeo90k', task_name='vfi'),
    data_root='data/vimeo-triplet',
    data_prefix=dict(img='sequences', gt='sequences'),
```

(continues on next page)

(continued from previous page)

```
    pipeline=[],
    depth=2,
    load_frames_list=dict(
        img=['img1.png', 'img3.png'], gt=['img2.png']))
```

See more details in unittest

tests/test_datasets/test_base_frames_dataset.py TestFramesDatasets().test_version_1_method()

METAINFO

load_data_list() → List[dict]

Load data list from folder or annotation file.

Returns A list of annotation.

Return type list[dict]

_get_path_list()

Get list of paths from annotation file or folder of dataset.

Returns A list of paths.

Return type list[str]

_get_path_list_from_ann()

Get list of paths from annotation file.

Returns A list of paths.

Return type list[str]

_get_path_list_from_folder(*sub_folder=None*, *need_ext=True*, *depth=1*)

Get list of paths from folder.

Parameters

- **sub_folder** (*None | str*) – The path of sub_folder. Default: None.
- **need_ext** (*bool*) – Whether need ext. Default: True.
- **depth** (*int*) – Residual depth of path, recursively called to `depth == 1`. Default: 1

Returns A list of paths.

Return type list[str]

_set_seq_lens()

Get sequence lengths.

_get_frames_list(*key*, *folder*)

Obtain list of frames.

Parameters

- **key** (*str*) – The key of frames list, e.g. `img`, `gt`.
- **folder** (*str*) – Folder of frames.

Returns The paths list of frames.

Return type list[str]

class mmagic.datasets.**BasicImageDataset**(*ann_file: str = ''*, *metainfo: Optional[dict] = None*, *data_root: Optional[str] = None*, *data_prefix: dict = dict(img='')*, *pipeline: List[Union[dict, Callable]] = []*, *test_mode: bool = False*, *filename_tmpl: dict = dict()*, *search_key: Optional[str] = None*, *backend_args: Optional[dict] = None*, *img_suffix: Optional[Union[str, Tuple[str]]] = IMG_EXTENSIONS*, *recursive: bool = False*, ***kwards*)

Bases: mmengine.dataset.BaseDataset

BasicImageDataset for open source projects in OpenMMLab/MMagic.

This dataset is designed for low-level vision tasks with image, such as super-resolution and inpainting.

The annotation file is optional.

If use annotation file, the annotation format can be shown as follows.

```
Case 1 (CelebA-HQ):

    000001.png
    000002.png

Case 2 (DIV2K):

    0001_s001.png (480,480,3)
    0001_s002.png (480,480,3)
    0001_s003.png (480,480,3)
    0002_s001.png (480,480,3)
    0002_s002.png (480,480,3)

Case 3 (Vimeo90k):

    00001/0266 (256, 448, 3)
    00001/0268 (256, 448, 3)
```

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as class information. Defaults to None.
- **data_root** (*str, optional*) – The root directory for data_prefix and ann_file. Defaults to None.
- **data_prefix** (*dict, optional*) – Prefix for training data. Defaults to dict(img=None, ann=None).
- **pipeline** (*list, optional*) – Processing pipeline. Defaults to [].
- **test_mode** (*bool, optional*) – test_mode=True means in test phase. Defaults to False.
- **filename_tmpl** (*dict*) – Template for each filename. Note that the template excludes the file extension. Default: dict().
- **search_key** (*str*) – The key used for searching the folder to get data_list. Default: 'gt'.
- **backend_args** (*dict, optional*) – Arguments to instantiate the prefix of uri corresponding backend. Defaults to None.

- **suffix** (*str or tuple[str], optional*) – File suffix that we are interested in. Default: None.
- **recursive** (*bool*) – If set to True, recursively scan the directory. Default: False.

Note: Assume the file structure as the following:

```
mmagic (root)
├── mmagic
├── tools
├── configs
├── data
│   ├── DIV2K
│   │   ├── DIV2K_train_HR
│   │   │   ├── image.png
│   │   ├── DIV2K_train_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   │   │   │   ├── image_x4.png
│   │   ├── DIV2K_valid_HR
│   │   ├── DIV2K_valid_LR_bicubic
│   │   │   ├── X2
│   │   │   ├── X3
│   │   │   ├── X4
│   ├── places
│   │   ├── test_set
│   │   ├── train_set
│   │   ├── meta
│   │   │   ├── Places365_train.txt
│   │   │   ├── Places365_val.txt
```

Examples

Case 1: Loading DIV2K dataset for training a SISR model.

```
dataset = BasicImageDataset(
    ann_file='',
    metainfo=dict(
        dataset_type='div2k',
        task_name='sisr'),
    data_root='data/DIV2K',
    data_prefix=dict(
        gt='DIV2K_train_HR', img='DIV2K_train_LR_bicubic/X4'),
    filename_tmpl=dict(img='{}_x4', gt='{}'),
    pipeline=[])
```

Case 2: Loading places dataset for training an inpainting model.

```
dataset = BasicImageDataset(
    ann_file='meta/Places365_train.txt',
```

(continues on next page)

(continued from previous page)

```
    metainfo=dict(
        dataset_type='places365',
        task_name='inpainting'),
    data_root='data/places',
    data_prefix=dict(gt='train_set'),
    pipeline=[])
```

METAINFO

load_data_list() → List[dict]

Load data list from folder or annotation file.

Returns A list of annotation.

Return type list[dict]

_get_path_list()

Get list of paths from annotation file or folder of dataset.

Returns A list of paths.

Return type list[dict]

_get_path_list_from_ann()

Get list of paths from annotation file.

Returns List of paths.

Return type List

_get_path_list_from_folder()

Get list of paths from folder.

Returns List of paths.

Return type List

class mmagic.datasets.**CIFAR10**(*data_prefix: str*, *test_mode: bool*, *metainfo: Optional[dict] = None*, *data_root: str = ''*, *download: bool = True*, ***kwargs*)

Bases: mmagic.datasets.basic_conditional_dataset.BasicConditionalDataset

CIFAR10 Dataset.

This implementation is modified from https://github.com/pytorch/vision/blob/master/torchvision/datasets/cifar.py

Parameters

- **data_prefix** (*str*) – Prefix for data.
- **test_mode** (*bool*) – test_mode=True means in test phase. It determines to use the training set or test set.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as categories information. Defaults to None.
- **data_root** (*str*) – The root directory for data_prefix. Defaults to ''.
- **download** (*bool*) – Whether to download the dataset if not exists. Defaults to True.
- ****kwargs** – Other keyword arguments in BaseDataset.

base_folder = cifar-10-batches-py

url = https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz

filename = cifar-10-python.tar.gz

tgz_md5 = c58f30108f718f92721af3b95e74349a

train_list = [['data_batch_1', 'c99cafc152244af753f735de768cd75f'], ['data_batch_2',...

test_list = [['test_batch', '40351d587109b95175f43aff81a1287e']]

meta

METAINFO

load_data_list()

Load images and ground truth labels.

_load_meta()

Load categories information from metafile.

_check_integrity()

Check the integrity of data files.

extra_repr() → List[str]

The extra repr information of the dataset.

class mmagic.datasets.**AdobeComp1kDataset**(*ann_file: Optional[str] = ''*, *metainfo: Union[collections.abc.Mapping, mmengine.config.Config, None] = None*, *data_root: Optional[str] = ''*, *data_prefix: dict = dict(img_path='')*, *filter_cfg: Optional[dict] = None*, *indices: Optional[Union[int, Sequence[int]]] = None*, *serialize_data: bool = True*, *pipeline: List[Union[dict, Callable]] = []*, *test_mode: bool = False*, *lazy_init: bool = False*, *max_refetch: int = 1000*)

Bases: mmengine.dataset.BaseDataset

Adobe composition-1k dataset.

The dataset loads (alpha, fg, bg) data and apply specified transforms to the data. You could specify whether composite merged image online or load composited merged image in pipeline.

Example for online comp-1k dataset:

```
[
    {
        "alpha": 'alpha/000.png',
        "fg": 'fg/000.png',
        "bg": 'bg/000.png'
    },
    {
        "alpha": 'alpha/001.png',
        "fg": 'fg/001.png',
        "bg": 'bg/001.png'
    },
]
```

Example for offline comp-1k dataset:

```
[
    {
        "alpha": 'alpha/000.png',
        "merged": 'merged/000.png',
        "fg": 'fg/000.png',
        "bg": 'bg/000.png'
    },
    {
        "alpha": 'alpha/001.png',
        "merged": 'merged/001.png',
        "fg": 'fg/001.png',
        "bg": 'bg/001.png'
    },
]
```

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **data_root** (*str, optional*) – The root directory for `data_prefix` and `ann_file`. Defaults to None.
- **pipeline** (*list, optional*) – Processing pipeline. Defaults to [].
- **test_mode** (*bool, optional*) – `test_mode=True` means in test phase. Defaults to False.
- ****kwargs** – Other arguments passed to `mmengine.dataset.BaseDataset`.

Examples

See unit-tests TODO: Move some codes in unittest here

METAINFO

load_data_list() → List[dict]

Load annotations from an annotation file named as `self.ann_file`

In order to be compatible to both new and old annotation format, we copy implementations from mmengine and do some modifications.

Returns A list of annotation.

Return type list[dict]

parse_data_info(*raw_data_info: dict*) → Union[dict, List[dict]]

Join data_root to each path in data_info.

class mmagic.datasets.**ControlNetDataset**(*ann_file: str = 'prompt.json', data_root: str = './data/fill50k', control_key='source', image_key='target', pipeline: List[Union[dict, Callable]] = []*)

Bases: `mmengine.dataset.BaseDataset`

Demo dataset to test ControlNet. Modified from https://github.com/lllyasviel/ControlNet/blob/16ea3b5379c1e78a4bc8e3fc9cae8d65c42511b1/tutorial_data set.py # noqa.

You can download the demo data from https://huggingface.co/lllyasviel/ControlNet/blob/main/training/fill50k.zip # noqa and then unzip the file to the `data` folder.

Parameters

- **ann_file** (*str*) – Path to the annotation file. Defaults to 'prompt.json' as ControlNet's default.
- **data_root** (*str*) – Path to the data root. Defaults to './data/fill50k'.
- **pipeline** (*list[dict | callable]*) – A sequence of data transforms.

load_data_list() → List[dict]

Load annotations from an annotation file named as `self.ann_file`

Returns A list of annotation.

Return type list[dict]

class mmagic.datasets.**DreamBoothDataset**(*data_root: str, concept_dir: str, prompt: str, pipeline: List[Union[dict, Callable]] = []*)

Bases: `mmengine.dataset.BaseDataset`

Dataset for DreamBooth.

Parameters

- **data_root** (*str*) – Path to the data root.
- **concept_dir** (*str*) – Path to the concept images.
- **prompt** (*str*) – Prompt of the concept.
- **pipeline** (*list[dict | callable]*) – A sequence of data transforms.

load_data_list() → list

Load data list from concept_dir and class_dir.

class mmagic.datasets.**GrowScaleImgDataset**(*data_roots: dict, pipeline, len_per_stage=int(1000000.0), gpu_samples_per_scale=None, gpu_samples_base=32, io_backend: Optional[str] = None, file_lists: Optional[Union[str, dict]] = None, test_mode=False*)

Bases: `mmengine.dataset.BaseDataset`

Grow Scale Unconditional Image Dataset.

This dataset is similar with `UnconditionalImageDataset`, but offer more dynamic functionalities for the supporting complex algorithms, like PGGAN.

Highlight functionalities:

1. Support growing scale dataset. The motivation is to decrease data pre-processing load in CPU. In this dataset, you can provide `imgs_roots` like:

```
{'64': 'path_to_64x64_imgs',
 '512': 'path_to_512x512_imgs'}
```

   Then, in training scales lower than 64x64, this dataset will set `self.imgs_root` as 'path_to_64x64_imgs';

2. Offer `samples_per_gpu` according to different scales. In this dataset, `self.samples_per_gpu` will help runner to know the updated batch size.

Basically, This dataset contains raw images for training unconditional GANs. Given a root dir, we will recursively find all images in this root. The transformation on data is defined by the pipeline.

Parameters

- **imgs_root** (*str*) – Root path for unconditional images.

- **pipeline** (*list[dict | callable]*) – A sequence of data transforms.
- **len_per_stage** (*int, optional*) – The length of dataset for each scale. This args change the length dataset by concatenating or extracting subset. If given a value less than 0., the original length will be kept. Defaults to 1e6.
- **gpu_samples_per_scale** (*dict | None, optional*) – Dict contains samples_per_gpu for each scale. For example, {'32': 4} will set the scale of 32 with samples_per_gpu=4, despite other scale with samples_per_gpu=self.gpu_samples_base.
- **gpu_samples_base** (*int, optional*) – Set default samples_per_gpu for each scale. Defaults to 32.
- **io_backend** (*str, optional*) – The storage backend type. Options are "disk", "ceph", "memcached", "lmdb", "http" and "petrel". Default: None.
- **test_mode** (*bool, optional*) – If True, the dataset will work in test mode. Otherwise, in train mode. Default to False.

_VALID_IMG_SUFFIX = ('.jpg', '.png', '.jpeg', '.JPEG')

load_data_list()

Load annotations.

update_annotations(*curr_scale*)

Update annotations.

Parameters **curr_scale** (*int*) – Current image scale.

Returns Whether to update.

Return type bool

concat_imgs_list_to(*num*)

Concat image list to specified length.

Parameters **num** (*int*) – The length of the concatenated image list.

prepare_train_data(*idx*)

Prepare training data.

Parameters **idx** (*int*) – Index of current batch.

Returns Prepared training data batch.

Return type dict

prepare_test_data(*idx*)

Prepare testing data.

Parameters **idx** (*int*) – Index of current batch.

Returns Prepared training data batch.

Return type dict

__getitem__(*idx*)

Get the idx-th image and data information of dataset after self.pipeline, and full_init will be called if the dataset has not been fully initialized.

During training phase, if self.pipeline get None, self._rand_another will be called until a valid image is fetched or

the maximum limit of refetch is reached.

Parameters **idx** (*int*) – The index of self.data_list.

Returns The idx-th image and data information of dataset after `self.pipeline`.

Return type dict

__repr__()

Print `self.transforms` in sequence.

Returns Formatted string.

Return type str

class mmagic.datasets.**ImageNet**(*ann_file: str = '', metainfo: Optional[dict] = None, data_root: str = '', data_prefix: Union[str, dict] = '', **kwargs*)

Bases: `mmagic.datasets.basic_conditional_dataset.BasicConditionalDataset`

ImageNet Dataset.

The dataset supports two kinds of annotation format. More details can be found in `CustomDataset`.

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as class information. Defaults to None.
- **data_root** (*str*) – The root directory for `data_prefix` and `ann_file`. Defaults to ''.
- **data_prefix** (*str | dict*) – Prefix for training data. Defaults to ''.
- ****kwargs** – Other keyword arguments in `CustomDataset` and `BaseDataset`.

IMG_EXTENSIONS = ('.jpg', '.jpeg', '.png', '.ppm', '.bmp', '.pgm', '.tif')

METAINFO

class mmagic.datasets.**MSCoCoDataset**(*ann_file: str = '', metainfo: Optional[dict] = None, data_root: str = '', drop_caption_rate=0.0, phase='train', year=2014, data_prefix: Union[str, dict] = '', extensions: Sequence[str] = ('.jpg', '.jpeg', '.png', '.ppm', '.bmp', '.pgm', '.tif'), lazy_init: bool = False, classes: Union[str, Sequence[str], None] = None, caption_style: str = '', **kwargs*)

Bases: `mmagic.datasets.basic_conditional_dataset.BasicConditionalDataset`

MSCoCo 2014 dataset.

Parameters

- **ann_file** (*str*) – Annotation file path. Defaults to ''.
- **metainfo** (*dict, optional*) – Meta information for dataset, such as class information. Defaults to None.
- **data_root** (*str*) – The root directory for `data_prefix` and `ann_file`. Defaults to ''.
- **drop_caption_rate** (*float, optional*) – Rate of dropping caption, used for training. Defaults to 0.0.
- **phase** (*str, optional*) – Subdataset used for certain phase, can be set to *train*, *test* and *val*. Defaults to 'train'.

- **year** (*int, optional*) – Version of CoCo dataset, can be set to 2014 and 2017. Defaults to 2014.
- **data_prefix** (*str | dict*) – Prefix for the data. Defaults to ''.
- **extensions** (*Sequence[str]*) – A sequence of allowed extensions. Defaults to ('.jpg', '.jpeg', '.png', '.ppm', '.bmp', '.pgm', '.tif').
- **lazy_init** (*bool*) – Whether to load annotation during instantiation. In some cases, such as visualization, only the meta information of the dataset is needed, which is not necessary to load annotation file. `Basedataset` can skip load annotations to save time by set `lazy_init=False`. Defaults to False.
- **caption_style** (*str*) – If you want to add a style description for each caption, you can set caption_style to your style prompt. For example, 'realistic style'. Defaults to empty str.
- ***kwargs** – Other keyword arguments in `BaseDataset`.

METAINFO

load_data_list()

Load image paths and gt_labels.

class mmagic.datasets.**PairedImageDataset**(*data_root*, *pipeline*, *io_backend: Optional[str] = None*, *test_mode=False*, *test_dir='test'*)

Bases: `mmengine.dataset.BaseDataset`

General paired image folder dataset for image generation.

It assumes that the training directory is '/path/to/data/train'. During test time, the directory is '/path/to/data/test'. '/path/to/data' can be initialized by args 'dataroot'. Each sample contains a pair of images concatenated in the w dimension (A|B).

Parameters

- **dataroot** (str | `Path`) – Path to the folder root of paired images.
- **pipeline** (*List[dict | callable]*) – A sequence of data transformations.
- **test_mode** (*bool*) – Store *True* when building test dataset. Default: *False*.
- **test_dir** (*str*) – Subfolder of dataroot which contain test images. Default: 'test'.

load_data_list()

Load paired image paths.

Returns List that contains paired image paths.

Return type list[dict]

scan_folder(*path*)

Obtain image path list (including sub-folders) from a given folder.

Parameters **path** (str | `Path`) – Folder path.

Returns Image list obtained from the given folder.

Return type list[str]

class mmagic.datasets.**SinGANDataset**(*data_root*, *min_size*, *max_size*, *scale_factor_init*, *pipeline*, *num_samples=- 1*)

Bases: `mmengine.dataset.BaseDataset`

SinGAN Dataset.

In this dataset, we create an image pyramid and save it in the cache.

Parameters

- **img_path** (*str*) – Path to the single image file.
- **min_size** (*int*) – Min size of the image pyramid. Here, the number will be set to the `min(H, W)`.
- **max_size** (*int*) – Max size of the image pyramid. Here, the number will be set to the `max(H, W)`.
- **scale_factor_init** (*float*) – Rescale factor. Note that the actual factor we use may be a little bit different from this value.
- **num_samples** (*int, optional*) – The number of samples (length) in this dataset. Defaults to -1.

full_init()

Skip the full init process for SinGANDataset.

load_data_list(*min_size*, *max_size*, *scale_factor_init*)

Load annotations for SinGAN Dataset.

Parameters

- **min_size** (*int*) – The minimum size for the image pyramid.
- **max_size** (*int*) – The maximum size for the image pyramid.
- **scale_factor_init** (*float*) – The initial scale factor.

__getitem__(*index*)

Get *:attr:self.data_dict*. For SinGAN, we use single image with different resolution to train the model.

Parameters **idx** (*int*) – This will be ignored in *SinGANDataset*.

Returns Dict contains input image in different resolution. `self.pipeline`.

Return type dict

__len__()

Get the length of filtered dataset and automatically call `full_init` if the dataset has not been fully init.

Returns The length of filtered dataset.

Return type int

class mmagic.datasets.**TextualInversionDataset**(*data_root: str*, *concept_dir: str*, *placeholder: str*, *template: str*, *with_image_reference: bool = False*, *pipeline: List[Union[dict, Callable]] = []*)

Bases: `mmengine.dataset.BaseDataset`

Dataset for Textual Inversion and ViCo.

Parameters

- **data_root** (*str*) – Path to the data root.
- **concept_dir** (*str*) – Path to the concept images.
- **placeholder** (*str*) – A string to denote the concept.
- **template** (*list[str]*) – A list of strings like 'A photo of {}'.
- **with_image_reference** (*bool*) – Is used for vico training.

- **pipeline** (*list[dict | callable]*) – A sequence of data transforms.

load_data_list() → list

Load data list from concept_dir and class_dir.

prepare_data(*idx*)

Get data processed by `self.pipeline`.

Parameters **idx** (*int*) – The index of `data_info`.

Returns Depends on `self.pipeline`.

Return type Any

class mmagic.datasets.**UnpairedImageDataset**(*data_root*, *pipeline*, *io_backend: Optional[str] = None*, *test_mode=False*, *domain_a='A'*, *domain_b='B'*)

Bases: `mmengine.dataset.BaseDataset`

General unpaired image folder dataset for image generation.

It assumes that the training directory of images from domain A is '/path/to/data/trainA', and that from domain B is '/path/to/data/trainB', respectively. '/path/to/data' can be initialized by args 'dataroot'. During test time, the directory is '/path/to/data/testA' and '/path/to/data/testB', respectively.

Parameters

- **dataroot** (str | `Path`) – Path to the folder root of unpaired images.
- **pipeline** (*List[dict | callable]*) – A sequence of data transformations.
- **io_backend** (*str, optional*) – The storage backend type. Options are "disk", "ceph", "memcached", "lmdb", "http" and "petrel". Default: None.
- **test_mode** (*bool*) – Store *True* when building test dataset. Default: *False*.
- **domain_a** (*str, optional*) – Domain of images in trainA / testA. Defaults to 'A'.
- **domain_b** (*str, optional*) – Domain of images in trainB / testB. Defaults to 'B'.

load_data_list()

Load the data list.

Returns The data info list of source and target domain.

Return type list

_load_domain_data_list(*dataroot*)

Load unpaired image paths of one domain.

Parameters **dataroot** (*str*) – Path to the folder root for unpaired images of one domain.

Returns List that contains unpaired image paths of one domain.

Return type list[dict]

get_data_info(*idx*) → dict

Get annotation by index and automatically call `full_init` if the dataset has not been fully initialized.

Parameters **idx** (*int*) – The index of data.

Returns The idx-th annotation of the dataset.

Return type dict

__len__()

The length of the dataset.

scan_folder(*path*)

Obtain image path list (including sub-folders) from a given folder.

Parameters **path** (str | Path) – Folder path.

Returns Image list obtained from the given folder.

Return type list[str]

1.80 mmagic.datasets.transforms

1.80.1 Package Contents

Classes

AlbuCorruptFunction	AlbuCorruptFunction augmentation.
PairedAlbuTransForms	PairedAlbuTransForms augmentation.
Albumentations	Albumentation augmentation.
GenerateSeg	Generate segmentation mask from alpha matte.
GenerateSoftSeg	Generate soft segmentation mask from input segmentation mask.
MirrorSequence	Extend short sequences (e.g. Vimeo-90K) by mirroring the sequences.
TemporalReverse	Reverse frame lists for temporal augmentation.
BinarizeImage	Binarize image.
Clip	Clip the pixels.
ColorJitter	An interface for torch color jitter so that it can be invoked in mmagic
RandomAffine	Apply random affine to input images.
RandomMaskDilation	Randomly dilate binary masks.
UnsharpMasking	Apply unsharp masking to an image or a sequence of images.
Flip	Flip the input data with a probability.
NumpyPad	Numpy Padding.
RandomRotation	Rotate the image by a randomly-chosen angle, measured in degree.
RandomTransposeHW	Randomly transpose images in H and W dimensions with a probability.
Resize	Resize data to a specific size for training or resize the images to fit
CenterCropLongEdge	Center crop the given image by the long edge.
Crop	Crop data to specific size for training.
CropAroundCenter	Randomly crop the images around unknown area in the center 1/4 images.
CropAroundFg	Crop around the whole foreground in the segmentation mask.
CropAroundUnknown	Crop around unknown area with a randomly selected scale.

continues on next page

Table 1 – continued from previous page

CropLike	Crop/pad the image in the target_key according to the size of image in
FixedCrop	Crop paired data (at a specific position) to specific size for training.
InstanceCrop	Use maskrcnn to detect instances on image.
ModCrop	Mod crop images, used during testing.
PairedRandomCrop	Paired random crop.
RandomCropLongEdge	Random crop the given image by the long edge.
RandomResizedCrop	Crop data to random size and aspect ratio.
CompositeFg	Composite foreground with a random foreground.
MergeFgAndBg	Composite foreground image and background image with alpha.
PerturbBg	Randomly add gaussian noise or gamma change to background image.
RandomJitter	Randomly jitter the foreground in hsv space.
RandomLoadResizeBg	Randomly load a background image and resize it.
PackInputs	Pack data into DataSample for training, evaluation and testing.
GenerateCoordinateAndCell	Generate coordinate and cell. Generate coordinate from the desired size
GenerateFacialHeatmap	Generate heatmap from keypoint.
GenerateFrameIndices	Generate frame index for REDS datasets. It also performs temporal
GenerateFrameIndiceswithPadding	Generate frame index with padding for REDS dataset and Vid4 dataset
GenerateSegmentIndices	Generate frame indices for a segment. It also performs temporal
GetMaskedImage	Get masked image.
GetSpatialDiscountMask	Get spatial discounting mask constant.
LoadImageFromFile	Load a single image or image frames from corresponding paths. Required
LoadMask	Load Mask for multiple types.
LoadPairedImageFromFile	Load a pair of images from file.
MATLABLikeResize	Resize the input image using MATLAB-like downsampling.
Normalize	Normalize images with the given mean and std value.
RescaleToZeroOne	Transform the images into a range between 0 and 1.
DegradationsWithShuffle	Apply random degradations to input, with degradations being shuffled.
RandomBlur	Apply random blur to the input.
RandomJPEGCompression	Apply random JPEG compression to the input.
RandomNoise	Apply random noise to the input.
RandomResize	Randomly resize the input.
RandomVideoCompression	Apply random video compression to the input.
RandomDownSampling	Generate LQ image from GT (and crop), which will randomly pick a scale.
FormatTrimap	Convert trimap (tensor) to one-hot representation.
GenerateTrimap	Using random erode/dilate to generate trimap from alpha matte.
GenerateTrimapWithDistTransform	Generate trimap with distance transform function.
TransformTrimap	Transform trimap into two-channel and six-channel.

continues on next page

Table 1 – continued from previous page

CopyValues	Copy the value of source keys to destination keys.
SetValues	Set value to destination keys.

class mmagic.datasets.transforms.**AlbuCorruptFunction**(*keys: List[str], config: List[dict], p: float = 1.0*)

Bases: mmcv.transforms.BaseTransform

AlbuCorruptFunction augmentation.

Apply the same AlbuCorruptFunction augmentation to the input images.

transform(*results*)

processing input results according to *self.augs*.

Parameters

- **results** (*dict*) – contains the processed data
- **pipeline.** (*through the transform*) –

Returns the processed data.

Return type results

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**PairedAlbuTransForms**(*size: int, lq_key: str = 'img', gt_key: str = 'gt', scope: str = 'geometric', crop: str = 'random', p: float = 0.5*)

Bases: mmcv.transforms.BaseTransform

PairedAlbuTransForms augmentation.

Apply the same AlbuTransforms augmentation to paired images.

transform(*results*)

processing input results according to *self.pipeline*.

Parameters

- **results** (*dict*) – contains the processed data
- **pipeline.** (*through the transform*) –

Returns the processed data.

Return type results

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Albumentations**(*keys: List[str], transforms: List[dict]*)

Bases: mmcv.transforms.BaseTransform

Albumentation augmentation.

Adds custom transformations from Albumentations library. Please, visit *https://github.com/albumentations-team/albumentations* and *https://albumentations.ai/docs/getting_started/transforms_and_targets* to get more information.

An example of transforms is as followed:

```
albu_transforms = [
    dict(
        type='Resize',
        height=100,
        width=100,
    ),
    dict(
        type='RandomFog',
        p=0.5,
    ),
    dict(
        type='RandomRain',
        p=0.5
    ),
    dict(
        type='RandomSnow',
        p=0.5,
    ),
]
pipeline = [
    dict(
        type='LoadImageFromFile',
        key='img',
        color_type='color',
        channel_order='rgb',
        imdecode_backend='cv2'),
    dict(
        type='Albumentations',
        keys=['img'],
        transforms=albu_transforms),
    dict(type='PackInputs')
]
```

Parameters

- **keys** (*list[str]*) – A list specifying the keys whose values are modified.
- **transforms** (*list[dict]*) – A list of albu transformations.

albu_builder(*cfg: dict*) → albumentations

Import a module from albumentations.

It inherits some of `build_from_cfg()` logic.

Parameters **cfg** (*dict*) – Config dict. It should at least contain the key "type".

Returns The constructed object.

Return type obj

_apply_albu(*imgs*)

transform(*results*)

Transform function of Albumentations.

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateSeg**(*kernel_size=5*, *erode_iter_range=(10, 20)*, *dilate_iter_range=(15, 30)*, *num_holes_range=(0, 3)*, *hole_sizes=[(15, 15), (25, 25), (35, 35), (45, 45)]*, *blur_ksizes=[(21, 21), (31, 31), (41, 41)]*)

Bases: mmcv.transforms.BaseTransform

Generate segmentation mask from alpha matte.

Parameters

- **kernel_size** (*int, optional*) – Kernel size for both erosion and dilation. The kernel will have the same height and width. Defaults to 5.
- **erode_iter_range** (*tuple, optional*) – Iteration of erosion. Defaults to (10, 20).
- **dilate_iter_range** (*tuple, optional*) – Iteration of dilation. Defaults to (15, 30).
- **num_holes_range** (*tuple, optional*) – Range of number of holes to randomly select from. Defaults to (0, 3).
- **hole_sizes** (*list, optional*) – List of (h, w) to be selected as the size of the rectangle hole. Defaults to [(15, 15), (25, 25), (35, 35), (45, 45)].
- **blur_ksizes** (*list, optional*) – List of (h, w) to be selected as the kernel_size of the gaussian blur. Defaults to [(21, 21), (31, 31), (41, 41)].

static _crop_hole(*img*, *start_point*, *hole_size*)

Create a all-zero rectangle hole in the image.

Parameters

- **img** (*np.ndarray*) – Source image.
- **start_point** (*tuple[int]*) – The top-left point of the rectangle.
- **hole_size** (*tuple[int]*) – The height and width of the rectangle hole.

Returns The cropped image.

Return type np.ndarray

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateSoftSeg**(*fg_thr=0.2*, *border_width=25*, *erode_ksize=3*, *dilate_ksize=5*, *erode_iter_range=(10, 20)*, *dilate_iter_range=(3, 7)*, *blur_ksizes=[(21, 21), (31, 31), (41, 41)]*)

Bases: mmcv.transforms.BaseTransform

Generate soft segmentation mask from input segmentation mask.

Required key is "seg", added key is "soft_seg".

Parameters

- **fg_thr** (*float, optional*) – Threshold of the foreground in the normalized input segmentation mask. Defaults to 0.2.
- **border_width** (*int, optional*) – Width of border to be padded to the bottom of the mask. Defaults to 25.
- **erode_ksize** (*int, optional*) – Fixed kernel size of the erosion. Defaults to 5.
- **dilate_ksize** (*int, optional*) – Fixed kernel size of the dilation. Defaults to 5.
- **erode_iter_range** (*tuple, optional*) – Iteration of erosion. Defaults to (10, 20).
- **dilate_iter_range** (*tuple, optional*) – Iteration of dilation. Defaults to (3, 7).
- **blur_ksizes** (*list, optional*) – List of (h, w) to be selected as the kernel_size of the gaussian blur. Defaults to [(21, 21), (31, 31), (41, 41)].

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**MirrorSequence**(*keys*)

Bases: mmcv.transforms.BaseTransform

Extend short sequences (e.g. Vimeo-90K) by mirroring the sequences.

Given a sequence with N frames (x1, …, xN), extend the sequence to (x1, …, xN, xN, …, x1).

Required Keys:

- [KEYS]

Modified Keys:

- [KEYS]

Parameters **keys** (*list[str]*) – The frame lists to be extended.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**TemporalReverse**(*keys*, *reverse_ratio=0.5*)

Bases: mmcv.transforms.BaseTransform

Reverse frame lists for temporal augmentation.

Required keys are the keys in attributes "lq" and "gt", added or modified keys are "lq", "gt" and "reverse".

Parameters

- **keys** (*list[str]*) – The frame lists to be reversed.
- **reverse_ratio** (*float*) – The probability to reverse the frame lists. Default: 0.5.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**BinarizeImage**(*keys*, *binary_thr*, *a_min=0*, *a_max=1*, *dtype=np.uint8*)

Bases: mmcv.transforms.BaseTransform

Binarize image.

Parameters

- **keys** (*Sequence[str]*) – The images to be binarized.
- **binary_thr** (*float*) – Threshold for binarization.
- **a_min** (*int*) – Lower limits of pixel value.
- **a_max** (*int*) – Upper limits of pixel value.
- **dtype** (*np.dtype*) – Set the data type of the output. Default: np.uint8

_binarize(*img*)

Binarize image.

Parameters **img** (*np.ndarray*) – Input image.

Returns Output image.

Return type img (np.ndarray)

transform(*results*)

The transform function of BinarizeImage.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Clip**(*keys*, *a_min=0*, *a_max=255*)

Bases: mmcv.transforms.BaseTransform

Clip the pixels.

Modified keys are the attributes specified in "keys".

Parameters

- **keys** (*list[str]*) – The keys whose values are clipped.
- **a_min** (*int*) – Lower limits of pixel value.
- **a_max** (*int*) – Upper limits of pixel value.

_clip(*input_*)

Clip the pixels.

Parameters **input** (*Union[List, np.ndarray]*) – Pixels to clip.

Returns Clipped pixels.

Return type Union[List, np.ndarray]

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns

A dict with the values of the specified keys are rounded and clipped.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**ColorJitter**(*keys*, *channel_order='rgb'*, ***kwargs*)

Bases: mmcv.transforms.BaseTransform

An interface for torch color jitter so that it can be invoked in mmagic pipeline.

Randomly change the brightness, contrast and saturation of an image. Modified keys are the attributes specified in "keys".

Required Keys:

- [KEYS]

Modified Keys:

- [KEYS]

Parameters

- **keys** (*list[str]*) – The images to be resized.
- **channel_order** (*str*) – Order of channel, candidates are 'bgr' and 'rgb'. Default: 'rgb'.

Notes

`**kwards` follows the args list of `torchvision.transforms.ColorJitter`.

brightness (float or tuple of float (min, max)): How much to jitter brightness. brightness_factor is chosen uniformly from [max(0, 1 - brightness), 1 + brightness] or the given [min, max]. Should be non negative numbers.

contrast (float or tuple of float (min, max)): How much to jitter contrast. contrast_factor is chosen uniformly from [max(0, 1 - contrast), 1 + contrast] or the given [min, max]. Should be non negative numbers.

saturation (float or tuple of float (min, max)): How much to jitter saturation. saturation_factor is chosen uniformly from [max(0, 1 - saturation), 1 + saturation] or the given [min, max]. Should be non negative numbers.

hue (float or tuple of float (min, max)): How much to jitter hue. hue_factor is chosen uniformly from [-hue, hue] or the given [min, max]. Should have 0<= hue <= 0.5 or -0.5 <= min <= max <= 0.5.

`_color_jitter`(*image*, *this_seed*)

Color Jitter Function.

Parameters

- **image** (*np.ndarray*) – Image.
- **this_seed** (*int*) – Seed of torch.

Returns The output image.

Return type image (np.ndarray)

`transform`(*results: Dict*) → Dict

The transform function of ColorJitter.

Parameters **results** (*dict*) – The result dict.

Returns The result dict.

Return type dict

`__repr__`()

Return repr(self).

class `mmagic.datasets.transforms.`**`RandomAffine`**(*keys*, *degrees*, *translate=None*, *scale=None*, *shear=None*, *flip_ratio=None*)

Bases: `mmcv.transforms.BaseTransform`

Apply random affine to input images.

This class is adopted from https://github.com/pytorch/vision/blob/v0.5.0/torchvision/transforms/ transforms.py#L1015 It should be noted that in https://github.com/Yaoyi-Li/GCA-Matting/blob/master/dataloader/data_generator.py#L70 random flip is added. See explanation of *flip_ratio* below. Required keys are the keys in attribute "keys", modified keys are keys in attribute "keys".

Parameters

- **keys** (*Sequence[str]*) – The images to be affined.
- **degrees** (*float | tuple[float]*) – Range of degrees to select from. If it is a float instead of a tuple like (min, max), the range of degrees will be (-degrees, +degrees). Set to 0 to deactivate rotations.

- **translate** (*tuple, optional*) – Tuple of maximum absolute fraction for horizontal and vertical translations. For example translate=(a, b), then horizontal shift is randomly sampled in the range -img_width * a < dx < img_width * a and vertical shift is randomly sampled in the range -img_height * b < dy < img_height * b. Default: None.
- **scale** (*tuple, optional*) – Scaling factor interval, e.g (a, b), then scale is randomly sampled from the range a <= scale <= b. Default: None.
- **shear** (*float | tuple[float], optional*) – Range of shear degrees to select from. If shear is a float, a shear parallel to the x axis and a shear parallel to the y axis in the range (-shear, +shear) will be applied. Else if shear is a tuple of 2 values, a x-axis shear and a y-axis shear in (shear[0], shear[1]) will be applied. Default: None.
- **flip_ratio** (*float, optional*) – Probability of the image being flipped. The flips in horizontal direction and vertical direction are independent. The image may be flipped in both directions. Default: None.

static _get_params(*degrees*, *translate*, *scale_ranges*, *shears*, *flip_ratio*, *img_size*)

Get parameters for affine transformation.

Returns Params to be passed to the affine transformation.

Return type paras (tuple)

static _get_inverse_affine_matrix(*center*, *angle*, *translate*, *scale*, *shear*, *flip*)

Helper method to compute inverse matrix for affine transformation.

As it is explained in PIL.Image.rotate, we need compute INVERSE of affine transformation matrix: M = T * C * RSS * C^-1 where T is translation matrix:

[1, 0, tx | 0, 1, ty | 0, 0, 1];

C is translation matrix to keep center: [1, 0, cx | 0, 1, cy | 0, 0, 1];

RSS is rotation with scale and shear matrix.

It is different from the original function in torchvision. 1. The order are changed to flip -> scale -> rotation -> shear. 2. x and y have different scale factors. RSS(shear, a, scale, f) =

[cos(a + shear)*scale_x*f -sin(a + shear)*scale_y 0] [sin(a)*scale_x*f cos(a)*scale_y 0] [0 0 1]

Thus, the inverse is M^-1 = C * RSS^-1 * C^-1 * T^-1.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomMaskDilation**(*keys*, *binary_thr=0.0*, *kernel_min=9*, *kernel_max=49*)

Bases: mmcv.transforms.BaseTransform

Randomly dilate binary masks.

Parameters

- **keys** (*Sequence[str]*) – The images to be resized.
- **binary_thr** (*float*) – Threshold for obtaining binary mask. Default: 0.
- **kernel_min** (*int*) – Min size of dilation kernel. Default: 9.
- **kernel_max** (*int*) – Max size of dilation kernel. Default: 49.

_random_dilate(*img*)

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**UnsharpMasking**(*kernel_size*, *sigma*, *weight*, *threshold*, *keys*)

Bases: mmcv.transforms.BaseTransform

Apply unsharp masking to an image or a sequence of images.

Parameters

- **kernel_size** (*int*) – The kernel_size of the Gaussian kernel.
- **sigma** (*float*) – The standard deviation of the Gaussian.
- **weight** (*float*) – The weight of the "details" in the final output.
- **threshold** (*float*) – Pixel differences larger than this value are regarded as "details".
- **keys** (*list[str]*) – The keys whose values are processed.

Added keys are "xxx_unsharp", where "xxx" are the attributes specified in "keys".

_unsharp_masking(*imgs*)

Unsharp masking function.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Flip**(*keys*, *flip_ratio=0.5*, *direction='horizontal'*)

Bases: mmcv.transforms.BaseTransform

Flip the input data with a probability.

Reverse the order of elements in the given data with a specific direction. The shape of the data is preserved, but the elements are reordered. Required keys are the keys in attributes "keys", added or modified keys are "flip", "flip_direction" and the keys in attributes "keys". It also supports flipping a list of images with the same flip.

Required Keys:

- [KEYS]

Modified Keys:

- [KEYS]

Parameters

- **keys** (*Union[str, List[str]]*) – The images to be flipped.
- **flip_ratio** (*float*) – The probability to flip the images. Default: 0.5.
- **direction** (*str*) – Flip images horizontally or vertically. Options are "horizontal" | "vertical". Default: "horizontal".

_directions = ['horizontal', 'vertical']

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**NumpyPad**(*keys*, *padding*, ***kwargs*)

Bases: mmcv.transforms.BaseTransform

Numpy Padding.

In this augmentation, numpy padding is adopted to customize padding augmentation. Please carefully read the numpy manual in: https://numpy.org/doc/stable/reference/generated/numpy.pad.html

If you just hope a single dimension to be padded, you must set padding like this:

```
padding = ((2, 2), (0, 0), (0, 0))
```

In this case, if you adopt an input with three dimension, only the first dimension will be padded.

Parameters

- **keys** (*Union[str, List[str]]*) – The images to be padded.
- **padding** (*int | tuple(int)*) – Please refer to the args pad_width in numpy.pad.

transform(*results*)

Call function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__() → str

Return repr(self).

class mmagic.datasets.transforms.**RandomRotation**(*keys*, *degrees*)

Bases: mmcv.transforms.BaseTransform

Rotate the image by a randomly-chosen angle, measured in degree.

Parameters

- **keys** (*list[str]*) – The images to be rotated.
- **degrees** (*tuple[float] | tuple[int] | float | int*) – If it is a tuple, it represents a range (min, max). If it is a float or int, the range is constructed as (-degrees, degrees).

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomTransposeHW**(*keys*, *transpose_ratio=0.5*)

Bases: mmcv.transforms.BaseTransform

Randomly transpose images in H and W dimensions with a probability.

(TransposeHW = horizontal flip + anti-clockwise rotation by 90 degrees) When used with horizontal/vertical flips, it serves as a way of rotation augmentation. It also supports randomly transposing a list of images.

Required keys are the keys in attributes "keys", added or modified keys are "transpose" and the keys in attributes "keys".

Parameters

- **keys** (*list[str]*) – The images to be transposed.
- **transpose_ratio** (*float*) – The probability to transpose the images. Default: 0.5.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Resize**(*keys: Union[str, List[str]] = 'img'*, *scale=None*, *keep_ratio=False*, *size_factor=None*, *max_size=None*, *interpolation='bilinear'*, *backend=None*, *output_keys=None*)

Bases: mmcv.transforms.BaseTransform

Resize data to a specific size for training or resize the images to fit the network input regulation for testing.

When used for resizing images to fit network input regulation, the case is that a network may have several downsample and then upsample operation, then the input height and width should be divisible by the downsample

factor of the network. For example, the network would downsample the input for 5 times with stride 2, then the downsample factor is 2^5 = 32 and the height and width should be divisible by 32.

Required keys are the keys in attribute "keys", added or modified keys are "keep_ratio", "scale_factor", "interpolation" and the keys in attribute "keys".

Required Keys:

- Required keys are the keys in attribute "keys"

Modified Keys:

- Modified the keys in attribute "keys" or save as new key ([OUT_KEY])

Added Keys:

- [OUT_KEY]_shape
- keep_ratio
- scale_factor
- interpolation

All keys in "keys" should have the same shape. "test_trans" is used to record the test transformation to align the input's shape.

Parameters

- **keys** (*str | list[str]*) – The image(s) to be resized.
- **scale** (*float | tuple[int]*) – If scale is tuple[int], target spatial size (h, w). Otherwise, target spatial size is scaled by input size. Note that when it is used, *size_factor* and *max_size* are useless. Default: None
- **keep_ratio** (*bool*) – If set to True, images will be resized without changing the aspect ratio. Otherwise, it will resize images to a given size. Default: False. Note that it is used together with *scale*.
- **size_factor** (*int*) – Let the output shape be a multiple of size_factor. Default:None. Note that when it is used, *scale* should be set to None and *keep_ratio* should be set to False.
- **max_size** (*int*) – The maximum size of the longest side of the output. Default:None. Note that it is used together with *size_factor*.
- **interpolation** (*str*) – Algorithm used for interpolation: "nearest" | "bilinear" | "bicubic" | "area" | "lanczos". Default: "bilinear".
- **backend** (*str | None*) – The image resize backend type. Options are *cv2*, *pillow*, *None*. If backend is None, the global imread_backend specified by `mmcv.use_backend()` will be used. Default: None.
- **output_keys** (*list[str] | None*) – The resized images. Default: None Note that if it is not *None*, its length should be equal to keys.

_resize(*img*)

Resize function.

Parameters **img** (*np.ndarray*) – Image.

Returns Resized image.

Return type img (np.ndarray)

transform(*results: Dict*) → Dict

Transform function to resize images.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CenterCropLongEdge**(*keys='img'*)

Bases: mmcv.transforms.BaseTransform

Center crop the given image by the long edge.

Parameters **keys** (*list[str]*) – The images to be cropped.

transform(*results*)

Call function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Crop**(*keys*, *crop_size*, *random_crop=True*, *is_pad_zeros=False*)

Bases: mmcv.transforms.BaseTransform

Crop data to specific size for training.

Parameters

- **keys** (*Sequence[str]*) – The images to be cropped.
- **crop_size** (*Tuple[int]*) – Target spatial size (h, w).
- **random_crop** (*bool*) – If set to True, it will random crop image. Otherwise, it will work as center crop. Default: True.
- **is_pad_zeros** (*bool, optional*) – Whether to pad the image with 0 if crop_size is greater than image size. Default: False.

_crop(*data*)

Crop the data.

Parameters **data** (*Union[List, np.ndarray]*) – Input data to crop.

Returns cropped data and corresponding crop box.

Return type tuple

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CropAroundCenter**(*crop_size*)

Bases: mmcv.transforms.BaseTransform

Randomly crop the images around unknown area in the center 1/4 images.

This cropping strategy is adopted in GCA matting. The *unknown area* is the same as *semi-transparent area*. https://arxiv.org/pdf/2001.04069.pdf

It retains the center 1/4 images and resizes the images to 'crop_size'. Required keys are "fg", "bg", "trimap" and "alpha", added or modified keys are "crop_bbox", "fg", "bg", "trimap" and "alpha".

Parameters **crop_size** (*int | tuple*) – Desired output size. If int, square crop is applied.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CropAroundFg**(*keys*, *bd_ratio_range=(0.1, 0.4)*, *test_mode=False*)

Bases: mmcv.transforms.BaseTransform

Crop around the whole foreground in the segmentation mask.

Required keys are "seg" and the keys in argument *keys*. Meanwhile, "seg" must be in argument *keys*. Added or modified keys are "crop_bbox" and the keys in argument *keys*.

Parameters

- **keys** (*Sequence[str]*) – The images to be cropped. It must contain 'seg'.
- **bd_ratio_range** (*tuple, optional*) – The range of the boundary (bd) ratio to select from. The boundary ratio is the ratio of the boundary to the minimal bbox that contains the whole foreground given by segmentation. Default to (0.1, 0.4).
- **test_mode** (*bool*) – Whether use test mode. In test mode, the tight crop area of foreground will be extended to the a square. Default to False.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

class mmagic.datasets.transforms.**CropAroundUnknown**(*keys*, *crop_sizes*, *unknown_source='alpha'*, *interpolations='bilinear'*)

Bases: mmcv.transforms.BaseTransform

Crop around unknown area with a randomly selected scale.

Randomly select the w and h from a list of (w, h). Required keys are the keys in argument *keys*, added or modified keys are "crop_bbox" and the keys in argument *keys*. This class assumes value of "alpha" ranges from 0 to 255.

Parameters

- **keys** (*Sequence[str]*) – The images to be cropped. It must contain 'alpha'. If unknown_source is set to 'trimap', then it must also contain 'trimap'.
- **crop_sizes** (*list[int | tuple[int]]*) – List of (w, h) to be selected.
- **unknown_source** (*str, optional*) – Unknown area to select from. It must be 'alpha' or 'trimap'. Default to 'alpha'.
- **interpolations** (*str | list[str], optional*) – Interpolation method of mmcv.imresize. The interpolation operation will be applied when image size is smaller than the crop_size. If given as a list of str, it should have the same length as *keys*. Or if given as a str all the keys will be resized with the same method. Default to 'bilinear'.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CropLike**(*target_key*, *reference_key=None*)

Bases: mmcv.transforms.BaseTransform

Crop/pad the image in the target_key according to the size of image in the reference_key .

Parameters

- **target_key** (*str*) – The key needs to be cropped.
- **reference_key** (*str | None*) – The reference key, need its size. Default: None.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation. Require self.target_key and self.reference_key.

Returns

A dict containing the processed data and information. Modify self.target_key.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**FixedCrop**(*keys*, *crop_size*, *crop_pos=None*)

Bases: mmcv.transforms.BaseTransform

Crop paired data (at a specific position) to specific size for training.

Parameters

- **keys** (*Sequence[str]*) – The images to be cropped.
- **crop_size** (*Tuple[int]*) – Target spatial size (h, w).
- **crop_pos** (*Tuple[int]*) – Specific position (x, y). If set to None, random initialize the position to crop paired data batch. Default: None.

_crop(*data*, *x_offset*, *y_offset*, *crop_w*, *crop_h*)

Crop the data.

Parameters

- **data** (*Union[List, np.ndarray]*) – Input data to crop.
- **x_offset** (*int*) – The offset of x axis.
- **y_offset** (*int*) – The offset of y axis.
- **crop_w** (*int*) – The width of crop bbox.
- **crop_h** (*int*) – The height of crop bbox.

Returns cropped data and corresponding crop box.

Return type tuple

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**InstanceCrop**(*config_file*, *from_pretrained=None*, *key='img'*, *box_num_upbound=- 1*, *finesize=256*)

Bases: mmcv.transforms.BaseTransform

Use maskrcnn to detect instances on image.

Mask R-CNN is used to detect the instance on the image pred_bbox is used to segment the instance on the image

Parameters

- **config_file** (*str*) – config file name relative to detectron2's "configs/"
- **key** (*str*) – Unused
- **box_num_upbound** (*int*) – The upper limit on the number of instances in the figure

transform(*results: dict*) → dict

The transform function of InstanceCrop.

Parameters **results** (*dict*) – A dict containing the necessary information and data for Conversion

Returns

A dict containing the processed data and information.

Return type results (dict)

predict_bbox(*image*)

class mmagic.datasets.transforms.**ModCrop**(*key='gt'*)

Bases: mmcv.transforms.BaseTransform

Mod crop images, used during testing.

Required keys are "scale" and "KEY", added or modified keys are "KEY".

Parameters **key** (*str*) – The key of image. Default: 'gt'

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**PairedRandomCrop**(*gt_patch_size*, *lq_key='img'*, *gt_key='gt'*)

Bases: mmcv.transforms.BaseTransform

Paired random crop.

It crops a pair of img and gt images with corresponding locations. It also supports accepting img list and gt list. Required keys are "scale", "lq_key", and "gt_key", added or modified keys are "lq_key" and "gt_key".

Parameters

- **gt_patch_size** (*int*) – cropped gt patch size.
- **lq_key** (*str*) – Key of LQ img. Default: 'img'.
- **gt_key** (*str*) – Key of GT img. Default: 'gt'.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomCropLongEdge**(*keys='img'*)

Bases: mmcv.transforms.BaseTransform

Random crop the given image by the long edge.

Parameters **keys** (*list[str]*) – The images to be cropped.

transform(*results*)

Call function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomResizedCrop**(*keys*, *crop_size*, *scale=(0.08, 1.0)*, *ratio=(3.0 / 4.0, 4.0 / 3.0)*, *interpolation='bilinear'*)

Bases: mmcv.transforms.BaseTransform

Crop data to random size and aspect ratio.

A crop of a random proportion of the original image and a random aspect ratio of the original aspect ratio is made. The cropped image is finally resized to a given size specified by 'crop_size'. Modified keys are the attributes specified in "keys".

This code is partially adopted from torchvision.transforms.RandomResizedCrop: [https://pytorch.org/vision/stable/_modules/torchvision/transforms/ transforms.html#RandomResizedCrop].

Parameters

- **keys** (*list[str]*) – The images to be resized and random-cropped.
- **crop_size** (*int | tuple[int]*) – Target spatial size (h, w).
- **scale** (*tuple[float], optional*) – Range of the proportion of the original image to be cropped. Default: (0.08, 1.0).
- **ratio** (*tuple[float], optional*) – Range of aspect ratio of the crop. Default: (3. / 4., 4. / 3.).
- **interpolation** (*str, optional*) – Algorithm used for interpolation. It can be only either one of the following: "nearest" | "bilinear" | "bicubic" | "area" | "lanczos". Default: "bilinear".

get_params(*data*)

Get parameters for a random sized crop.

Parameters **data** (*np.ndarray*) – Image of type numpy array to be cropped.

Returns A tuple containing the coordinates of the top left corner and the chosen crop size.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CompositeFg**(*fg_dirs*, *alpha_dirs*, *interpolation='nearest'*)

Bases: mmcv.transforms.BaseTransform

Composite foreground with a random foreground.

This class composites the current training sample with additional data randomly (could be from the same dataset). With probability 0.5, the sample will be composited with a random sample from the specified directory. The composition is performed as:

$$fg_{new} = \alpha_1 * fg_1 + (1 - \alpha_1) * fg_2$$
$$\alpha_{new} = 1 - (1 - \alpha_1) * (1 - \alpha_2)$$

where (fg_1, α_1) is from the current sample and (fg_2, α_2) is the randomly loaded sample. With the above composition, α_{new} is still in *[0, 1]*.

Required keys are "alpha" and "fg". Modified keys are "alpha" and "fg".

Parameters

- **fg_dirs** (*str | list[str]*) – Path of directories to load foreground images from.
- **alpha_dirs** (*str | list[str]*) – Path of directories to load alpha mattes from.
- **interpolation** (*str*) – Interpolation method of *mmcv.imresize* to resize the randomly loaded images. Default: 'nearest'.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

_get_file_list(*fg_dirs*, *alpha_dirs*)

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**MergeFgAndBg**

Bases: mmcv.transforms.BaseTransform

Composite foreground image and background image with alpha.

Required keys are "alpha", "fg" and "bg", added key is "merged".

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__() → str

Return repr(self).

class mmagic.datasets.transforms.**PerturbBg**(*gamma_ratio=0.6*)

Bases: mmcv.transforms.BaseTransform

Randomly add gaussian noise or gamma change to background image.

Required key is "bg", added key is "noisy_bg".

Parameters **gamma_ratio** (*float, optional*) – The probability to use gamma correction instead of gaussian noise. Defaults to 0.6.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomJitter**(*hue_range=40*)

Bases: mmcv.transforms.BaseTransform

Randomly jitter the foreground in hsv space.

The jitter range of hue is adjustable while the jitter ranges of saturation and value are adaptive to the images. Side effect: the "fg" image will be converted to *np.float32*. Required keys are "fg" and "alpha", modified key is "fg".

Parameters **hue_range** (*float | tuple[float]*) – Range of hue jittering. If it is a float instead of a tuple like (min, max), the range of hue jittering will be (-hue_range, +hue_range). Default: 40.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RandomLoadResizeBg**(*bg_dir, flag='color', channel_order='bgr'*)

Bases: mmcv.transforms.BaseTransform

Randomly load a background image and resize it.

Required key is "fg", added key is "bg".

Parameters

- **bg_dir** (*str*) – Path of directory to load background images from.
- **flag** (*str*) – Loading flag for images. Default: 'color'.
- **channel_order** (*str*) – Order of channel, candidates are 'bgr' and 'rgb'. Default: 'bgr'.
- **kwargs** (*dict*) – Args for file client.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**PackInputs**(*keys: Tuple[List[str], str] = ['merged', 'img'], meta_keys: Tuple[List[str], str] = [], data_keys: Tuple[List[str], str] = []*)

Bases: mmcv.transforms.base.BaseTransform

Pack data into DataSample for training, evaluation and testing.

MMagic follows the design of data structure from MMEngine. Data from the loader will be packed into data field of DataSample. More details of DataSample refer to the documentation of MMEngine: https://mmengine.readthedocs.io/en/latest/advanced_tutorials/data_element.html

Parameters

- **Tuple[List[str]** (*meta_keys*) – The keys to saved in returned inputs, which are used as the input of models, default to ['img', 'noise', 'merged'].
- **str** – The keys to saved in returned inputs, which are used as the input of models, default to ['img', 'noise', 'merged'].
- **None]** – The keys to saved in returned inputs, which are used as the input of models, default to ['img', 'noise', 'merged'].
- **Tuple[List[str]** – The keys to saved in *data_field* of the *data_samples*.
- **str** – The keys to saved in *data_field* of the *data_samples*.
- **None]** – The keys to saved in *data_field* of the *data_samples*.
- **Tuple[List[str]** – The meta keys to saved in *metainfo* of the *data_samples*. All the other data will be packed into the data of the *data_samples*
- **str** – The meta keys to saved in *metainfo* of the *data_samples*. All the other data will be packed into the data of the *data_samples*
- **None]** – The meta keys to saved in *metainfo* of the *data_samples*. All the other data will be packed into the data of the *data_samples*

transform(*results: dict*) → dict

Method to pack the input data.

Parameters **results** (*dict*) – Result dict from the data pipeline.

Returns

A dict contains

- 'inputs' (obj:*dict*): The forward data of models. According to different tasks, the *inputs* may contain images, videos, labels, text, etc.
- **'data_samples' (obj:*DataSample*): The annotation info of the** sample.

Return type dict

__repr__() → str

Return repr(self).

class mmagic.datasets.transforms.**GenerateCoordinateAndCell**(*sample_quantity=None*, *scale=None*, *target_size=None*, *reshape_gt=True*)

Bases: mmcv.transforms.base.BaseTransform

Generate coordinate and cell. Generate coordinate from the desired size of SR image.

Train or val:

1. Generate coordinate from GT.

#. Reshape GT image to (HgWg, 3) and transpose to (3, HgWg). where *Hg* and *Wg* represent the height and width of GT.

Test:

1. Generate coordinate from LQ and scale or target_size.
2. Then generate cell from coordinate.

Parameters

- **sample_quantity** (*int | None*) – The quantity of samples in coordinates. To ensure that the GT tensors in a batch have the same dimensions. Default: None.
- **scale** (*float*) – Scale of upsampling. Default: None.
- **target_size** (*tuple[int]*) – Size of target image. Default: None.
- **reshape_gt** (*bool*) – Whether reshape gt to (-1, 3). Default: True If sample_quantity is not None, reshape_gt = True.

The priority of getting 'size of target image' is:

1. results['gt'].shape[-2:]
2. results['lq'].shape[-2:] * scale
3. target_size

transform(*results*)

Call function.

Parameters

- **results** (*Require either in*) – A dict containing the necessary information
- **augmentation.** (*and data for*) –
- **results** –
- **'lq'** (*1.*) –
- **'gt'** (*2.*) –
- **None** (*3.*) –
- **and** (*the premise is self.target_size*) –
- **len** (*self.target_size*) –

Returns A dict containing the processed data and information. Reshape 'gt' to (-1, 3) and transpose to (3, -1) if 'gt' in results. Add 'coord' and 'cell'.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateFacialHeatmap**(*image_key*, *ori_size*, *target_size*, *sigma=1.0*, *use_cache=True*)

Bases: mmcv.transforms.base.BaseTransform

Generate heatmap from keypoint.

Parameters

- **image_key** (*str*) – Key of facial image in dict.
- **ori_size** (*int | Tuple[int]*) – Original image size of keypoint.
- **target_size** (*int | Tuple[int]*) – Target size of heatmap.
- **sigma** (*float*) – Sigma parameter of heatmap. Default: 1.0
- **use_cache** (*bool*) – If True, load all heatmap at once. Default: True.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation. Require keypoint.

Returns

A dict containing the processed data and information. Add 'heatmap'.

Return type dict

generate_heatmap_from_img(*image*)

Generate heatmap from img.

Parameters **image** (*np.ndarray*) – Face image.

results: heatmap (np.ndarray): Heatmap the face image.

_face_alignment_detector(*image*)

Generate face landmark by face_alignment.

Parameters **image** (*np.ndarray*) – Face image.

Returns Location of landmark.

Return type landmark (Tuple[float])

_generate_one_heatmap(*keypoint*)

Generate One Heatmap.

Parameters **keypoint** (*Tuple[float]*) – Location of a landmark.

results: heatmap (np.ndarray): A heatmap of landmark.

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateFrameIndices**(*interval_list*, *frames_per_clip=99*)

Bases: mmcv.transforms.BaseTransform

Generate frame index for REDS datasets. It also performs temporal augmentation with random interval.

Required Keys:

- img_path
- gt_path
- key
- num_input_frames

Modified Keys:

- img_path
- gt_path

Added Keys:

- interval
- reverse

Parameters

- **interval_list** (*list[int]*) – Interval list for temporal augmentation. It will randomly pick an interval from interval_list and sample frame index with the interval.
- **frames_per_clip** (*int*) – Number of frames per clips. Default: 99 for REDS dataset.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateFrameIndiceswithPadding**(*padding*, *filename_tmpl='{:08d}'*)

Bases: mmcv.transforms.BaseTransform

Generate frame index with padding for REDS dataset and Vid4 dataset during testing.

Required Keys:

- img_path
- gt_path
- key
- num_input_frames
- sequence_length

Modified Keys:

- img_path
- gt_path

Parameters **padding** – padding mode, one of 'replicate' | 'reflection' | 'reflection_circle' | 'circle'.

Examples: current_idx = 0, num_input_frames = 5 The generated frame indices under different padding mode:

replicate: [0, 0, 0, 1, 2] reflection: [2, 1, 0, 1, 2] reflection_circle: [4, 3, 0, 1, 2] circle: [3, 4, 0, 1, 2]

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateSegmentIndices**(*interval_list*, *start_idx=0*, *filename_tmpl='{:08d}.png'*)

Bases: mmcv.transforms.BaseTransform

Generate frame indices for a segment. It also performs temporal augmentation with random interval.

Required Keys:

- img_path
- gt_path
- key
- num_input_frames
- sequence_length

Modified Keys:

- img_path
- gt_path

Added Keys:

- interval
- reverse

Parameters

- **interval_list** (*list[int]*) – Interval list for temporal augmentation. It will randomly pick an interval from interval_list and sample frame index with the interval.
- **start_idx** (*int*) – The index corresponds to the first frame in the sequence. Default: 0.
- **filename_tmpl** (*str*) – Template for file name. Default: '{:08d}.png'.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GetMaskedImage**(*img_key='gt'*, *mask_key='mask'*, *out_key='img'*, *zero_value=127.5*)

Bases: mmcv.transforms.base.BaseTransform

Get masked image.

Parameters

- **img_key** (*str*) – Key for clean image. Default: 'gt'.
- **mask_key** (*str*) – Key for mask image. The mask shape should be (h, w, 1) while '1' indicate holes and '0' indicate valid regions. Default: 'mask'.
- **img_key** – Key for output image. Default: 'img'.
- **zero_value** (*float*) – Pixel value of masked area.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GetSpatialDiscountMask**(*gamma=0.99*, *beta=1.5*)

Bases: mmcv.transforms.BaseTransform

Get spatial discounting mask constant.

Spatial discounting mask is first introduced in: Generative Image Inpainting with Contextual Attention.

Parameters

- **gamma** (*float, optional*) – Gamma for computing spatial discounting. Defaults to 0.99.
- **beta** (*float, optional*) – Beta for computing spatial discounting. Defaults to 1.5.

spatial_discount_mask(*mask_width*, *mask_height*)

Generate spatial discounting mask constant.

Parameters

- **mask_width** (*int*) – The width of bbox hole.
- **mask_height** (*int*) – The height of bbox height.

Returns Spatial discounting mask.

Return type np.ndarray

`transform`(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

`__repr__`()

Return repr(self).

class `mmagic.datasets.transforms.`**`LoadImageFromFile`**(*key: str, color_type: str = 'color', channel_order: str = 'bgr', imdecode_backend: Optional[str] = None, use_cache: bool = False, to_float32: bool = False, to_y_channel: bool = False, save_original_img: bool = False, backend_args: Optional[dict] = None*)

Bases: `mmcv.transforms.BaseTransform`

Load a single image or image frames from corresponding paths. Required Keys: - [Key]_path

New Keys: - [KEY] - ori_[KEY]_shape - ori_[KEY]

Parameters

- **key** (*str*) – Keys in results to find corresponding path.
- **color_type** (*str*) – The flag argument for :func:`mmcv.imfrombytes`. Defaults to 'color'.
- **channel_order** (*str*) – Order of channel, candidates are 'bgr' and 'rgb'. Default: 'bgr'.
- **imdecode_backend** (*str*) – The image decoding backend type. The backend argument for :func:`mmcv.imfrombytes`. See :func:`mmcv.imfrombytes` for details. candidates are 'cv2', 'turbojpeg', 'pillow', and 'tifffile'. Defaults to None.
- **use_cache** (*bool*) – If True, load all images at once. Default: False.
- **to_float32** (*bool*) – Whether to convert the loaded image to a float32 numpy array. If set to False, the loaded image is an uint8 array. Defaults to False.
- **to_y_channel** (*bool*) – Whether to convert the loaded image to y channel. Only support 'rgb2ycbcr' and 'rgb2ycbcr' Defaults to False.
- **backend_args** (*dict, optional*) – Arguments to instantiate the prefix of uri corresponding backend. Defaults to None.

`transform`(*results: dict*) → dict

Functions to load image or frames.

Parameters **results** (*dict*) – Result dict from :obj:`mmcv.BaseDataset`.

Returns The dict contains loaded image and meta information.

Return type dict

`_load_image`(*filename*)

Load an image from file.

Parameters **filename** (*str*) – Path of image file.

Returns Image.

Return type np.ndarray

_convert(*img: numpy.ndarray*)

Convert an image to the require format.

Parameters **img** (*np.ndarray*) – The original image.

Returns The converted image.

Return type np.ndarray

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**LoadMask**(*mask_mode='bbox'*, *mask_config=None*)

Bases: mmcv.transforms.BaseTransform

Load Mask for multiple types.

For different types of mask, users need to provide the corresponding config dict.

Example config for bbox:

```
config = dict(img_shape=(256, 256), max_bbox_shape=128)
```

Example config for irregular:

```
config = dict(
    img_shape=(256, 256),
    num_vertices=(4, 12),
    max_angle=4.,
    length_range=(10, 100),
    brush_width=(10, 40),
    area_ratio_range=(0.15, 0.5))
```

Example config for ff:

```
config = dict(
    img_shape=(256, 256),
    num_vertices=(4, 12),
    mean_angle=1.2,
    angle_range=0.4,
    brush_width=(12, 40))
```

Example config for set:

```
config = dict(
    mask_list_file='xxx/xxx/ooxx.txt',
    prefix='/xxx/xxx/ooxx/',
    io_backend='local',
    color_type='unchanged',
    file_client_kwargs=dict()
)

The mask_list_file contains the list of mask file name like this:
    test1.jpeg
    test2.jpeg
    ...
```

(continues on next page)

(continued from previous page)

```
    ...

The prefix gives the data path.
```

Parameters

- **mask_mode** (*str*) – Mask mode in ['bbox', 'irregular', 'ff', 'set', 'file']. Default: 'bbox'. * bbox: square bounding box masks. * irregular: irregular holes. * ff: free-form holes from DeepFillv2. * set: randomly get a mask from a mask set. * file: get mask from 'mask_path' in results.
- **mask_config** (*dict*) – Params for creating masks. Each type of mask needs different configs. Default: None.

_init_info()

_get_random_mask_from_set()

_get_mask_from_file(*path*)

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**LoadPairedImageFromFile**(*key: str, domain_a: str = 'A', domain_b: str = 'B', color_type: str = 'color', channel_order: str = 'bgr', imdecode_backend: Optional[str] = None, use_cache: bool = False, to_float32: bool = False, to_y_channel: bool = False, save_original_img: bool = False, backend_args: Optional[dict] = None*)

Bases: *LoadImageFromFile*

Load a pair of images from file.

Each sample contains a pair of images, which are concatenated in the w dimension (a|b). This is a special loading class for generation paired dataset. It loads a pair of images as the common loader does and crops it into two images with the same shape in different domains.

Required key is "pair_path". Added or modified keys are "pair", "pair_ori_shape", "ori_pair", "img_{domain_a}", "img_{domain_b}", "img_{domain_a}_path", "img_{domain_b}_path", "img_{domain_a}_ori_shape", "img_{domain_b}_ori_shape", "ori_img_{domain_a}" and "ori_img_{domain_b}".

Parameters

- **key** (*str*) – Keys in results to find corresponding path.

- **domain_a** (*str, Optional*) – One of the paired image domain. Defaults to 'A'.
- **domain_b** (*str, Optional*) – The other of the paired image domain. Defaults to 'B'.
- **color_type** (*str*) – The flag argument for :func:mmcv.imfrombytes. Defaults to 'color'.
- **channel_order** (*str*) – Order of channel, candidates are 'bgr' and 'rgb'. Default: 'bgr'.
- **imdecode_backend** (*str*) – The image decoding backend type. The backend argument for :func:mmcv.imfrombytes. See :func:mmcv.imfrombytes for details. candidates are 'cv2', 'turbojpeg', 'pillow', and 'tifffile'. Defaults to None.
- **use_cache** (*bool*) – If True, load all images at once. Default: False.
- **to_float32** (*bool*) – Whether to convert the loaded image to a float32 numpy array. If set to False, the loaded image is an uint8 array. Defaults to False.
- **to_y_channel** (*bool*) – Whether to convert the loaded image to y channel. Only support 'rgb2ycbcr' and 'rgb2ycbcr' Defaults to False.
- **backend_args** (*dict, optional*) – Arguments to instantiate the prefix of uri corresponding backend. Defaults to None.
- **io_backend** (*str, optional*) – io backend where images are store. Defaults to None.

transform(*results: dict*) → dict

Functions to load paired images.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

class mmagic.datasets.transforms.**MATLABLikeResize**(*keys*, *scale=None*, *output_shape=None*, *kernel='bicubic'*, *kernel_width=4.0*)

Bases: mmcv.transforms.BaseTransform

Resize the input image using MATLAB-like downsampling.

Currently support bicubic interpolation only. Note that the output of this function is slightly different from the official MATLAB function.

Required keys are the keys in attribute "keys". Added or modified keys are "scale" and "output_shape", and the keys in attribute "keys".

Parameters

- **keys** (*list[str]*) – A list of keys whose values are modified.
- **scale** (*float | None, optional*) – The scale factor of the resize operation. If None, it will be determined by output_shape. Default: None.
- **output_shape** (*tuple(int) | None, optional*) – The size of the output image. If None, it will be determined by scale. Note that if scale is provided, output_shape will not be used. Default: None.
- **kernel** (*str, optional*) – The kernel for the resize operation. Currently support 'bicubic' only. Default: 'bicubic'.
- **kernel_width** (*float*) – The kernel width. Currently support 4.0 only. Default: 4.0.

_resize(*img*)

resize an image to the require size.

Parameters **img** (*np.ndarray*) – The original image.

Returns The resized image.

Return type output (np.ndarray)

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**Normalize**(*keys*, *mean*, *std*, *to_rgb=False*, *save_original=False*)

Bases: mmcv.transforms.BaseTransform

Normalize images with the given mean and std value.

Required keys are the keys in attribute "keys", added or modified keys are the keys in attribute "keys" and these keys with postfix '_norm_cfg'. It also supports normalizing a list of images.

Parameters

- **keys** (*Sequence[str]*) – The images to be normalized.
- **mean** (*np.ndarray*) – Mean values of different channels.
- **std** (*np.ndarray*) – Std values of different channels.
- **to_rgb** (*bool*) – Whether to convert channels from BGR to RGB. Default: False.
- **save_original** (*bool*) – Whether to save original images. Default: False.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**RescaleToZeroOne**(*keys*)

Bases: mmcv.transforms.BaseTransform

Transform the images into a range between 0 and 1.

Required keys are the keys in attribute "keys", added or modified keys are the keys in attribute "keys". It also supports rescaling a list of images.

Parameters **keys** (*Sequence[str]*) – The images to be transformed.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**DegradationsWithShuffle**(*degradations*, *keys*, *shuffle_idx=None*)

Apply random degradations to input, with degradations being shuffled.

Degradation groups are supported. The order of degradations within the same group is preserved. For example, if we have degradations = [a, b, [c, d]] and shuffle_idx = None, then the possible orders are

```
[a, b, [c, d]]
[a, [c, d], b]
[b, a, [c, d]]
[b, [c, d], a]
[[c, d], a, b]
[[c, d], b, a]
```

Modified keys are the attributed specified in "keys".

Parameters

- **degradations** (*list[dict]*) – The list of degradations.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.
- **shuffle_idx** (*list | None, optional*) – The degradations corresponding to these indices are shuffled. If None, all degradations are shuffled. Default: None.

_build_degradations(*degradations*)

__call__(*results*)

Call this transform.

__repr__()

Print the basic information of the transform.

class mmagic.datasets.transforms.**RandomBlur**(*params*, *keys*)

Apply random blur to the input.

Modified keys are the attributed specified in "keys".

Parameters

- **params** (*dict*) – A dictionary specifying the degradation settings.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.

get_kernel(*num_kernels: int*)

This is the function to create kernel.

Parameters **num_kernels** (*int*) – the number of kernels

Returns _description_

Return type _type_

`_apply_random_blur`(*imgs*)

This is the function to apply blur operation on images.

Parameters **imgs** (*Tensor*) – images

Returns Images applied blur

Return type Tensor

`__call__`(*results*)

Call this transform.

`__repr__`()

Print the basic information of the transform.

class `mmagic.datasets.transforms.`**RandomJPEGCompression**(*params*, *keys*, *color_type='color'*, *bgr2rgb=False*)

Apply random JPEG compression to the input.

Modified keys are the attributed specified in "keys".

Parameters

- **params** (*dict*) – A dictionary specifying the degradation settings.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.
- **bgr2rgb** (*str*) – Whether change channel order. Default: False.

`_apply_random_compression`(*imgs*)

`__call__`(*results*)

Call this transform.

`__repr__`()

Print the basic information of the transform.

class `mmagic.datasets.transforms.`**RandomNoise**(*params*, *keys*)

Apply random noise to the input.

Currently support Gaussian noise and Poisson noise.

Modified keys are the attributed specified in "keys".

Parameters

- **params** (*dict*) – A dictionary specifying the degradation settings.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.

`_apply_gaussian_noise`(*imgs*)

This is the function used to apply gaussian noise on images.

Parameters **imgs** (*Tensor*) – images

Returns images applied gaussian noise

Return type Tensor

`_apply_poisson_noise`(*imgs*)

_apply_random_noise(*imgs*)

This is the function used to apply random noise on images.

Parameters **imgs** (*Tensor*) – training images

Returns _description_

Return type _type_

__call__(*results*)

Call this transform.

__repr__()

Print the basic information of the transform.

class mmagic.datasets.transforms.**RandomResize**(*params*, *keys*)

Randomly resize the input.

Modified keys are the attributed specified in "keys".

Parameters

- **params** (*dict*) – A dictionary specifying the degradation settings.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.

_random_resize(*imgs*)

This is the function used to randomly resize images for training augmentation.

Parameters **imgs** (*Tensor*) – training images.

Returns images after randomly resized

Return type Tensor

__call__(*results*)

Call this transform.

__repr__()

Print the basic information of the transform.

class mmagic.datasets.transforms.**RandomVideoCompression**(*params*, *keys*)

Apply random video compression to the input.

Modified keys are the attributed specified in "keys".

Parameters

- **params** (*dict*) – A dictionary specifying the degradation settings.
- **keys** (*list[str]*) – A list specifying the keys whose values are modified.

_apply_random_compression(*imgs*)

This is the function to apply random compression on images.

Parameters **imgs** (*Tensor*) – training images

Returns images after randomly compressed

Return type Tensor

__call__(*results*)

Call this transform.

__repr__()

Print the basic information of the transform.

class mmagic.datasets.transforms.**RandomDownSampling**(*scale_min=1.0*, *scale_max=4.0*, *patch_size=None*, *interpolation='bicubic'*, *backend='pillow'*)

Bases: mmcv.transforms.BaseTransform

Generate LQ image from GT (and crop), which will randomly pick a scale.

Parameters

- **scale_min** (*float*) – The minimum of upsampling scale, inclusive. Default: 1.0.
- **scale_max** (*float*) – The maximum of upsampling scale, exclusive. Default: 4.0.
- **patch_size** (*int*) – The cropped lr patch size. Default: None, means no crop.
- **interpolation** (*str*) – Interpolation method, accepted values are "nearest", "bilinear", "bicubic", "area", "lanczos" for 'cv2' backend, "nearest", "bilinear", "bicubic", "box", "lanczos", "hamming" for 'pillow' backend. Default: "bicubic".
- **backend** (*str | None*) – The image resize backend type. Options are *cv2*, *pillow*, *None*. If backend is None, the global imread_backend specified by mmcv.use_backend() will be used. Default: "pillow".
- **[scale_min** (*Scale will be picked in the range of*) –
- **scale_max).** –

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation. 'gt' is required.

Returns

A dict containing the processed data and information. modified 'gt', supplement 'lq' and 'scale' to keys.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**FormatTrimap**(*to_onehot=False*)

Bases: mmcv.transforms.BaseTransform

Convert trimap (tensor) to one-hot representation.

It transforms the trimap label from (0, 128, 255) to (0, 1, 2). If to_onehot is set to True, the trimap will convert to one-hot tensor of shape (3, H, W). Required key is "trimap", added or modified key are "trimap" and "format_trimap_to_onehot".

Parameters **to_onehot** (*bool*) – whether convert trimap to one-hot tensor. Default: False.

transform(*results*)

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateTrimap**(*kernel_size*, *iterations=1*, *random=True*)

Bases: mmcv.transforms.BaseTransform

Using random erode/dilate to generate trimap from alpha matte.

Required key is "alpha", added key is "trimap".

Parameters

- **kernel_size** (*int | tuple[int]*) – The range of random kernel_size of erode/dilate; int indicates a fixed kernel_size. If *random* is set to False and kernel_size is a tuple of length 2, then it will be interpreted as (erode kernel_size, dilate kernel_size). It should be noted that the kernel of the erosion and dilation has the same height and width.
- **iterations** (*int | tuple[int], optional*) – The range of random iterations of erode/dilate; int indicates a fixed iterations. If *random* is set to False and iterations is a tuple of length 2, then it will be interpreted as (erode iterations, dilate iterations). Default to 1.
- **random** (*bool, optional*) – Whether use random kernel_size and iterations when generating trimap. See *kernel_size* and *iterations* for more information. Default to True.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**GenerateTrimapWithDistTransform**(*dist_thr=20*, *random=True*)

Bases: mmcv.transforms.BaseTransform

Generate trimap with distance transform function.

Parameters

- **dist_thr** (*int, optional*) – Distance threshold. Area with alpha value between (0, 255) will be considered as initial unknown area. Then area with distance to unknown area smaller than the distance threshold will also be consider as unknown area. Defaults to 20.
- **random** (*bool, optional*) – If True, use random distance threshold from [1, dist_thr). If False, use *dist_thr* as the distance threshold directly. Defaults to True.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**TransformTrimap**

Bases: mmcv.transforms.BaseTransform

Transform trimap into two-channel and six-channel.

This class will generate a two-channel trimap composed of definite foreground and background masks and encode it into a six-channel trimap using Gaussian blurs of the generated two-channel trimap at three different scales. The transformed trimap has 6 channels.

Required key is "trimap", added key is "transformed_trimap" and "two_channel_trimap".

Adopted from the following repository: https://github.com/MarcoForte/FBA_Matting/blob/master/networks/transforms.py.

transform(*results: dict*) → dict

Transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict containing the processed data and information.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**CopyValues**(*src_keys*, *dst_keys*)

Bases: mmcv.transforms.BaseTransform

Copy the value of source keys to destination keys.

TODO Change to dict(dst=src)

It does the following: results[dst_key] = results[src_key] for (src_key, dst_key) in zip(src_keys, dst_keys).

Added keys are the keys in the attribute "dst_keys".

Required Keys:

- [SRC_KEYS]

Added Keys:

- [DST_KEYS]

Parameters

- **src_keys** (*list[str]*) – The source keys.
- **dst_keys** (*list[str]*) – The destination keys.

transform(*results*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict with a key added/modified.

Return type dict

__repr__()

Return repr(self).

class mmagic.datasets.transforms.**SetValues**(*dictionary*)

Bases: mmcv.transforms.BaseTransform

Set value to destination keys.

It does the following: results[key] = value

Added keys are the keys in the dictionary.

Required Keys:

- None

Added or Modified Keys:

- keys in the dictionary

Parameters **dictionary** (*dict*) – The dictionary to update.

transform(*results: Dict*)

transform function.

Parameters **results** (*dict*) – A dict containing the necessary information and data for augmentation.

Returns A dict with a key added/modified.

Return type dict

__repr__()

Return repr(self).

1.81 mmagic.evaluation

1.81.1 Package Contents

Classes

Evaluator	Evaluator for generative models. Unlike high-level vision tasks, metrics
MAE	Mean Absolute Error metric for image.
MSE	Mean Squared Error metric for image.
NIQE	Calculate NIQE (Natural Image Quality Evaluator) metric.
PSNR	Peak Signal-to-Noise Ratio.
SAD	Sum of Absolute Differences metric for image matting.
SNR	Signal-to-Noise Ratio.
SSIM	Calculate SSIM (structural similarity).
ConnectivityError	Connectivity error for evaluating alpha matte prediction.
Equivariance	Metric for generative metrics. Except for the preparation phase
FrechetInceptionDistance	FID metric. In this metric, we calculate the distance between real
GradientError	Gradient error for evaluating alpha matte prediction.
InceptionScore	IS (Inception Score) metric. The images are split into groups, and the
MattingMSE	Mean Squared Error metric for image matting.
MultiScaleStructureSimilarity	MS-SSIM (Multi-Scale Structure Similarity) metric.
PerceptualPathLength	Perceptual path length.
PrecisionAndRecall	Improved Precision and recall metric.
SlicedWassersteinDistance	SWD (Sliced Wasserstein distance) metric. We calculate the SWD of two
TransFID	FID metric. In this metric, we calculate the distance between real
TransIS	IS (Inception Score) metric. The images are split into groups, and the

Functions

gauss_gradient(img, sigma)	Gaussian gradient.

class mmagic.evaluation.**Evaluator**(*metrics: Union[dict, mmengine.evaluator.BaseMetric, Sequence]*)

Bases: mmengine.evaluator.Evaluator

Evaluator for generative models. Unlike high-level vision tasks, metrics for generative models have various input types. For example, Inception Score (IS, *InceptionScore*) only needs to take fake images as input. However, Frechet Inception Distance (FID, *FrechetInceptionDistance*) needs to take both real images and fake images as input, and the numbers of real images and fake images can be set arbitrarily. For Perceptual path length (PPL, *PerceptualPathLength*), generator need to sample images along a latent path.

In order to be compatible with different metrics, we designed two critical functions, *prepare_metrics()* and *prepare_samplers()* to support those requirements.

- *prepare_metrics()* set the image images' color order and pass the dataloader to all metrics. Therefore metrics need pre-processing to prepare the corresponding feature.
- *prepare_samplers()* pass the dataloader and model to the metrics, and get the corresponding sampler of each kind of metrics. Metrics with same sample mode can share the sampler.

The whole evaluation process can be found in *mmagic.engine.runner.MultiValLoop.run()* and *mmagic.engine.runner.MultiTestLoop.run()*.

Parameters **metrics** (*dict or BaseMetric or Sequence*) – The config of metrics.

prepare_metrics(*module: mmengine.model.BaseModel*, *dataloader: torch.utils.data.dataloader.DataLoader*)

Prepare for metrics before evaluation starts. Some metrics use pretrained model to extract feature. Some metrics use pretrained model to extract feature and input channel order may vary among those models. Therefore, we first parse the output color order from data preprocessor and set the color order for each metric. Then we pass the dataloader to each metrics to prepare pre-calculated items. (e.g. inception feature of the real images). If metric has no pre-calculated items, `metric.prepare()` will be ignored. Once the function has been called, `self.is_ready` will be set as *True*. If `self.is_ready` is *True*, this function will directly return to avoid duplicate computation.

Parameters

- **module** (*BaseModel*) – Model to evaluate.
- **dataloader** (*DataLoader*) – The dataloader for real images.

static _cal_metric_hash(*metric: mmagic.evaluation.metrics.base_gen_metric.GenMetric*)

Calculate a unique hash value based on the *SAMPLER_MODE* and *sample_model*.

prepare_samplers(*module: mmengine.model.BaseModel*, *dataloader: torch.utils.data.dataloader.DataLoader*) → List[Tuple[List[mmengine.evaluator.BaseMetric], Iterator]]

Prepare for the sampler for metrics whose sampling mode are different. For generative models, different metric need image generated with different inputs. For example, FID, KID and IS need images generated with random noise, and PPL need paired images on the specific noise interpolation path. Therefore, we first group metrics with respect to their sampler's mode (refers to :attr:~`GenMetrics.SAMPLER_MODE`), and build a shared sampler for each metric group. To be noted that, the length of the shared sampler depends on the metric of the most images required in each group.

Parameters

- **module** (*BaseModel*) – Model to evaluate. Some metrics (e.g. PPL) require *module* in their sampler.
- **dataloader** (*DataLoader*) – The dataloader for real image.

Returns

A list of "metrics-shared sampler" pair.

Return type List[Tuple[List[BaseMetric], Iterator]]

process(*data_samples: Sequence[*mmagic.structures.DataSample*]*, *data_batch: Optional[Any]*, *metrics: Sequence[mmengine.evaluator.BaseMetric]*) → None

Pass *data_batch* from dataloader and *predictions* (generated results) to corresponding *metrics*.

Parameters

- **data_samples** (*Sequence[DataSample]*) – A batch of generated results from model.

- **data_batch** (*Optional[Any]*) – A batch of data from the metrics specific sampler or the dataloader.
- **metrics** (*Optional[Sequence[BaseMetric]]*) – Metrics to evaluate.

evaluate() → dict

Invoke `evaluate` method of each metric and collect the metrics dictionary. Different from *Evaluator.evaluate*, this function does not take *size* as input, and elements in *self.metrics* will call their own *evaluate* method to calculate the metric.

Returns

Evaluation results of all metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type dict

`mmagic.evaluation.`**gauss_gradient**(*img*, *sigma*)

Gaussian gradient.

From https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/ submissions/8060/versions/2/previews/gaussgradient/gaussgradient.m/ index.html

Parameters

- **img** (*np.ndarray*) – Input image.
- **sigma** (*float*) – Standard deviation of the gaussian kernel.

Returns Gaussian gradient of input *img*.

Return type np.ndarray

class `mmagic.evaluation.`**MAE**(*gt_key: str = 'gt_img'*, *pred_key: str = 'pred_img'*, *mask_key: Optional[str] = None*, *scaling=1*, *device='cpu'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*)

Bases: `mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric`

Mean Absolute Error metric for image.

mean(abs(a-b))

Parameters

- **gt_key** (*str*) – Key of ground-truth. Default: 'gt_img'
- **pred_key** (*str*) – Key of prediction. Default: 'pred_img'
- **mask_key** (*str, optional*) – Key of mask, if mask_key is None, calculate all regions. Default: None
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None

Metrics:

- MAE (float): Mean of Absolute Error

metric = MAE

process_image(*gt*, *pred*, *mask*)

Process an image.

Parameters

- **gt** (*Tensor | np.ndarray*) – GT image.
- **pred** (*Tensor | np.ndarray*) – Pred image.
- **mask** (*Tensor | np.ndarray*) – Mask of evaluation.

Returns MAE result.

Return type result (np.ndarray)

class mmagic.evaluation.**MSE**(*gt_key: str = 'gt_img'*, *pred_key: str = 'pred_img'*, *mask_key: Optional[str] = None*, *scaling=1*, *device='cpu'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Mean Squared Error metric for image.

mean((a-b)^2)

Parameters

- **gt_key** (*str*) – Key of ground-truth. Default: 'gt_img'
- **pred_key** (*str*) – Key of prediction. Default: 'pred_img'
- **mask_key** (*str, optional*) – Key of mask, if mask_key is None, calculate all regions. Default: None
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None

Metrics:

- MSE (float): Mean of Squared Error

metric = MSE

process_image(*gt*, *pred*, *mask*)

Process an image.

Parameters

- **gt** (*Torch | np.ndarray*) – GT image.
- **pred** (*Torch | np.ndarray*) – Pred image.
- **mask** (*Torch | np.ndarray*) – Mask of evaluation.

Returns MSE result.

Return type result (np.ndarray)

class mmagic.evaluation.**NIQE**(*key: str = 'pred_img'*, *is_predicted: bool = True*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *crop_border=0*, *input_order='HWC'*, *convert_to='gray'*)

Bases: `mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric`

Calculate NIQE (Natural Image Quality Evaluator) metric.

Ref: Making a "Completely Blind" Image Quality Analyzer. This implementation could produce almost the same results as the official MATLAB codes: http://live.ece.utexas.edu/research/quality/niqe_release.zip

We use the official params estimated from the pristine dataset. We use the recommended block size (96, 96) without overlaps.

Parameters

- **key** (*str*) – Key of image. Default: 'pred_img'
- **is_predicted** (*bool*) – If the image is predicted, it will be picked from predictions; otherwise, it will be picked from data_batch. Default: True
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None
- **crop_border** (*int*) – Cropped pixels in each edges of an image. These pixels are not involved in the PSNR calculation. Default: 0.
- **input_order** (*str*) – Whether the input order is 'HWC' or 'CHW'. Default: 'HWC'.
- **convert_to** (*str*) – Whether to convert the images to other color models. If None, the images are not altered. When computing for 'Y', the images are assumed to be in BGR order. Options are 'Y' and None. Default: 'gray'.

Metrics:

- NIQE (float): Natural Image Quality Evaluator

metric = NIQE

process_image(*gt*, *pred*, *mask*) → None

Process an image.

Parameters

- **gt** (*np.ndarray*) – GT image.
- **pred** (*np.ndarray*) – Pred image.
- **mask** (*np.ndarray*) – Mask of evaluation.

Returns NIQE result.

Return type result (np.ndarray)

class `mmagic.evaluation.`**PSNR**(*gt_key: str = 'gt_img'*, *pred_key: str = 'pred_img'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *crop_border=0*, *input_order='CHW'*, *convert_to=None*)

Bases: `mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric`

Peak Signal-to-Noise Ratio.

Ref: https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

Parameters

- **gt_key** (*str*) – Key of ground-truth. Default: 'gt_img'
- **pred_key** (*str*) – Key of prediction. Default: 'pred_img'
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None
- **crop_border** (*int*) – Cropped pixels in each edges of an image. These pixels are not involved in the PSNR calculation. Default: 0.
- **input_order** (*str*) – Whether the input order is 'HWC' or 'CHW'. Default: 'CHW'.
- **convert_to** (*str*) – Whether to convert the images to other color models. If None, the images are not altered. When computing for 'Y', the images are assumed to be in BGR order. Options are 'Y' and None. Default: None.

Metrics:

- PSNR (float): Peak Signal-to-Noise Ratio

metric = PSNR

process_image(*gt*, *pred*, *mask*)

Process an image.

Parameters

- **gt** (*Torch | np.ndarray*) – GT image.
- **pred** (*Torch | np.ndarray*) – Pred image.
- **mask** (*Torch | np.ndarray*) – Mask of evaluation.

Returns PSNR result.

Return type np.ndarray

class mmagic.evaluation.**SAD**(*norm_const=1000*, ***kwargs*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Sum of Absolute Differences metric for image matting.

This metric compute per-pixel absolute difference and sum across all pixels. i.e. sum(abs(a-b)) / norm_const

Note: Current implementation assume image / alpha / trimap array in numpy format and with pixel value ranging from 0 to 255.

Note: pred_alpha should be masked by trimap before passing into this metric

Default prefix: ''

Parameters **norm_const** (*int*) – Divide the result to reduce its magnitude. Default to 1000.

Metrics:

- SAD (float): Sum of Absolute Differences

default_prefix =

metric = SAD

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*)

process(*data_batch: Sequence[dict]*, *data_samples: Sequence[dict]*) → None

Process one batch of data and predictions.

Parameters

- **data_batch** (*Sequence[Tuple[Any, dict]]*) – A batch of data from the dataloader.
- **predictions** (*Sequence[dict]*) – A batch of outputs from the model.

compute_metrics(*results: List*)

Compute the metrics from processed results.

Parameters **results** (*dict*) – The processed results of each batch.

Returns The computed metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type Dict

class mmagic.evaluation.**SNR**(*gt_key: str = 'gt_img'*, *pred_key: str = 'pred_img'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *crop_border=0*, *input_order='CHW'*, *convert_to=None*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Signal-to-Noise Ratio.

Ref: https://en.wikipedia.org/wiki/Signal-to-noise_ratio

Parameters

- **gt_key** (*str*) – Key of ground-truth. Default: 'gt_img'
- **pred_key** (*str*) – Key of prediction. Default: 'pred_img'
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None
- **crop_border** (*int*) – Cropped pixels in each edges of an image. These pixels are not involved in the SNR calculation. Default: 0.
- **input_order** (*str*) – Whether the input order is 'HWC' or 'CHW'. Default: 'CHW'.
- **convert_to** (*str*) – Whether to convert the images to other color models. If None, the images are not altered. When computing for 'Y', the images are assumed to be in BGR order. Options are 'Y' and None. Default: None.

Metrics:

- SNR (float): Signal-to-Noise Ratio

metric = SNR

process_image(*gt*, *pred*, *mask*)

Process an image.

Parameters

- **gt** (*Torch | np.ndarray*) – GT image.
- **pred** (*Torch | np.ndarray*) – Pred image.
- **mask** (*Torch | np.ndarray*) – Mask of evaluation.

Returns SNR result.

Return type np.ndarray

class mmagic.evaluation.**SSIM**(*gt_key: str = 'gt_img'*, *pred_key: str = 'pred_img'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *crop_border=0*, *input_order='CHW'*, *convert_to=None*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Calculate SSIM (structural similarity).

Ref: Image quality assessment: From error visibility to structural similarity

The results are the same as that of the official released MATLAB code in https://ece.uwaterloo.ca/~z70wang/research/ssim/.

For three-channel images, SSIM is calculated for each channel and then averaged.

Parameters

- **gt_key** (*str*) – Key of ground-truth. Default: 'gt_img'
- **pred_key** (*str*) – Key of prediction. Default: 'pred_img'
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Default: None
- **crop_border** (*int*) – Cropped pixels in each edges of an image. These pixels are not involved in the PSNR calculation. Default: 0.
- **input_order** (*str*) – Whether the input order is 'HWC' or 'CHW'. Default: 'HWC'.
- **convert_to** (*str*) – Whether to convert the images to other color models. If None, the images are not altered. When computing for 'Y', the images are assumed to be in BGR order. Options are 'Y' and None. Default: None.

Metrics:

- SSIM (float): Structural similarity

metric = SSIM

process_image(*gt*, *pred*, *mask*)

Process an image.

Parameters

- **gt** (*Torch | np.ndarray*) – GT image.
- **pred** (*Torch | np.ndarray*) – Pred image.

- **mask** (*Torch | np.ndarray*) – Mask of evaluation.

Returns SSIM result.

Return type np.ndarray

class mmagic.evaluation.**ConnectivityError**(*step=0.1*, *norm_constant=1000*, ***kwargs*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Connectivity error for evaluating alpha matte prediction.

Note: Current implementation assume image / alpha / trimap array in numpy format and with pixel value ranging from 0 to 255.

Note: pred_alpha should be masked by trimap before passing into this metric

Parameters

- **step** (*float*) – Step of threshold when computing intersection between *alpha* and *pred_alpha*. Default to 0.1 .
- **norm_const** (*int*) – Divide the result to reduce its magnitude. Default to 1000.

Default prefix: ''

Metrics:

- ConnectivityError (float): Connectivity Error

metric = ConnectivityError

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*)

process(*data_batch: Sequence[dict]*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in self.results, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*Sequence[dict]*) – A batch of data from the dataloader.
- **predictions** (*Sequence[dict]*) – A batch of outputs from the model.

compute_metrics(*results: List*)

Compute the metrics from processed results.

Parameters **results** (*dict*) – The processed results of each batch.

Returns The computed metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type Dict

class mmagic.evaluation.**Equivariance**(*fake_nums: int*, *real_nums: int = 0*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'gt_img'*, *need_cond_input: bool = False*, *sample_mode: str = 'ema'*, *sample_kwargs: dict = dict()*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *eq_cfg=dict()*)

Bases: `mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric`

Metric for generative metrics. Except for the preparation phase (`prepare()`), generative metrics do not need extra real images.

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **real_nums** (*int*) – Numbers of the real image need for the metric. If *-1* is passed means all images from the dataset is need. Defaults to 0.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **real_key** (*Optional[str]*) – Key for get real images from the input dict. Defaults to 'img'.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'ema'.
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.
- **sample_kwargs** (*dict*) – Sampling arguments for model test.

name = Equivariance

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

get_metric_sampler(*model: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*, *metrics: List[mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric]*)

Get sampler for generative metrics. Returns a dummy iterator, whose return value of each iteration is a dict containing batch size and sample mode to generate images.

Parameters

- **model** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for real images. Used to get batch size during generate fake images.
- **metrics** (*List['GenerativeMetric']*) – Metrics with the same sampler mode.

Returns Sampler for generative metrics.

Return type dummy_iterator

compute_metrics(*results*) → dict

Compute the metrics from processed results.

Parameters **results** (*list*) – The processed results of each batch.

Returns The computed metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type dict

_collect_target_results(*target: str*) → Optional[list]

Collect function for Eq metric. This function support collect results typing as Dict[List[Tensor]]`.

Parameters **target** (*str*) – Target results to collect.

Returns The collected results.

Return type Optional[list]

class mmagic.evaluation.**FrechetInceptionDistance**(*fake_nums: int*, *real_nums: int = - 1*, *inception_style='StyleGAN'*, *inception_path: Optional[str] = None*, *inception_pkl: Optional[str] = None*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'gt_img'*, *need_cond_input: bool = False*, *sample_model: str = 'orig'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *sample_kwargs: dict = dict()*)

Bases: mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric

FID metric. In this metric, we calculate the distance between real distributions and fake distributions. The distributions are modeled by the real samples and fake samples, respectively. *Inception_v3* is adopted as the feature extractor, which is widely used in StyleGAN and BigGAN.

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **real_nums** (*int*) – Numbers of the real images need for the metric. If -1 is passed, means all real images in the dataset will be used. Defaults to -1.
- **inception_style** (*str*) – The target inception style want to load. If the given style cannot be loaded successful, will attempt to load a valid one. Defaults to 'StyleGAN'.
- **inception_path** (*str, optional*) – Path the the pretrain Inception network. Defaults to None.
- **inception_pkl** (*str, optional*) – Path to reference inception pickle file. If *None*, the statistical value of real distribution will be calculated at running time. Defaults to None.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **real_key** (*Optional[str]*) – Key for get real images from the input dict. Defaults to 'img'.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.

- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'orig'.
- **collect_device** (*str, optional*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

name = FID

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*) → None

Preparing inception feature for the real images.

Parameters

- **module** (*nn.Module*) – The model to evaluate.
- **dataloader** (*DataLoader*) – The dataloader for real images.

_load_inception(*inception_style: str*, *inception_path: Optional[str]*) → Tuple[torch.nn.Module, str]

Load inception and return the successful loaded style.

Parameters

- **inception_style** (*str*) – Target style of Inception network want to load.
- **inception_path** (*Optional[str]*) – The path to the inception.

Returns

The actually loaded inception network and corresponding style.

Return type Tuple[nn.Module, str]

forward_inception(*image: torch.Tensor*) → torch.Tensor

Feed image to inception network and get the output feature.

Parameters **data_samples** (*Sequence[dict]*) – A batch of data sample dict used to extract inception feature.

Returns Image feature extracted from inception.

Return type Tensor

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

static _calc_fid(*sample_mean: numpy.ndarray*, *sample_cov: numpy.ndarray*, *real_mean: numpy.ndarray*, *real_cov: numpy.ndarray*, *eps: float = 1e-06*) → Tuple[float]

Refer to the implementation from:

https://github.com/rosinality/stylegan2-pytorch/blob/master/fid.py#L34

compute_metrics(*fake_results: list*) → dict

Compute the result of FID metric.

Parameters **fake_results** (*list*) – List of image feature of fake images.

Returns

A dict of the computed FID metric and its mean and covariance.

Return type dict

class mmagic.evaluation.**GradientError**(*sigma=1.4*, *norm_constant=1000*, ***kwargs*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Gradient error for evaluating alpha matte prediction.

Note: Current implementation assume image / alpha / trimap array in numpy format and with pixel value ranging from 0 to 255.

Note: pred_alpha should be masked by trimap before passing into this metric

---

Parameters

- **sigma** (*float*) – Standard deviation of the gaussian kernel. Defaults to 1.4 .
- **norm_const** (*int*) – Divide the result to reduce its magnitude. Defaults to 1000 .

Default prefix: ''

Metrics:

- GradientError (float): Gradient Error

metric = GradientError

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*)

process(*data_batch: Sequence[dict]*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in self.results, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*Sequence[dict]*) – A batch of data from the dataloader.
- **predictions** (*Sequence[dict]*) – A batch of outputs from the model.

compute_metrics(*results: List*)

Compute the metrics from processed results.

Parameters **results** (*dict*) – The processed results of each batch.

Returns The computed metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type Dict

class mmagic.evaluation.**InceptionScore**(*fake_nums: int = 50000.0*, *resize: bool = True*, *splits: int = 10*, *inception_style: str = 'StyleGAN'*, *inception_path: Optional[str] = None*, *resize_method='bicubic'*, *use_pillow_resize: bool = True*, *fake_key: Optional[str] = None*, *need_cond_input: bool = False*, *sample_model='orig'*, *collect_device: str = 'cpu'*, *prefix: str = None*)

Bases: mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric

IS (Inception Score) metric. The images are split into groups, and the inception score is calculated on each group of images, then the mean and standard deviation of the score is reported. The calculation of the inception score on a group of images involves first using the inception v3 model to calculate the conditional probability for each image (p(y|x)). The marginal probability is then calculated as the average of the conditional probabilities for the images in the group (p(y)). The KL divergence is then calculated for each image as the conditional probability multiplied by the log of the conditional probability minus the log of the marginal probability. The KL divergence is then summed over all images and averaged over all classes and the exponent of the result is calculated to give the final score.

Ref: https://github.com/sbarratt/inception-score-pytorch/blob/master/inception_score.py # noqa

Note that we highly recommend that users should download the Inception V3 script module from the following address. Then, the *inception_pkl* can be set with user's local path. If not given, we will use the Inception V3 from pytorch model zoo. However, this may bring significant different in the final results.

Tero's Inception V3: https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada-pytorch/pretrained/metrics/inception-2015-12-05.pt # noqa

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **resize** (*bool, optional*) – Whether resize image to 299x299. Defaults to True.
- **splits** (*int, optional*) – The number of groups. Defaults to 10.
- **inception_style** (*str*) – The target inception style want to load. If the given style cannot be loaded successful, will attempt to load a valid one. Defaults to 'StyleGAN'.
- **inception_path** (*str, optional*) – Path the the pretrain Inception network. Defaults to None.
- **resize_method** (*str*) – Resize method. If *resize* is False, this will be ignored. Defaults to 'bicubic'.
- **use_pil_resize** (*bool*) – Whether use Bicubic interpolation with Pillow's backend. If set as True, the evaluation process may be a little bit slow, but achieve a more accurate IS result. Defaults to False.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'orig'.
- **collect_device** (*str, optional*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.

- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

name = IS

pil_resize_method_mapping

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*) → None

Prepare for the pre-calculating items of the metric. Defaults to do nothing.

Parameters

- **module** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for the real images.

_load_inception(*inception_style: str*, *inception_path: Optional[str]*) → Tuple[torch.nn.Module, str]

Load pretrain model of inception network. :param inception_style: Target style of Inception network want to

load.

Parameters **inception_path** (*Optional[str]*) – The path to the inception.

Returns

The actually loaded inception network and corresponding style.

Return type Tuple[nn.Module, str]

_preprocess(*image: torch.Tensor*) → torch.Tensor

Preprocess image before pass to the Inception. Preprocess operations contain channel conversion and resize.

Parameters **image** (*Tensor*) – Image tensor before preprocess.

Returns

Image tensor after resize and channel conversion (if need.)

Return type Tensor

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

compute_metrics(*fake_results: list*) → dict

Compute the results of Inception Score metric.

Parameters **fake_results** (*list*) – List of image feature of fake images.

Returns A dict of the computed IS metric and its standard error

Return type dict

class mmagic.evaluation.**MattingMSE**(*norm_const=1000*, ***kwargs*)

Bases: mmagic.evaluation.metrics.base_sample_wise_metric.BaseSampleWiseMetric

Mean Squared Error metric for image matting.

This metric compute per-pixel squared error average across all pixels. i.e. mean((a-b)^2) / norm_const

Note: Current implementation assume image / alpha / trimap array in numpy format and with pixel value ranging from 0 to 255.

Note: pred_alpha should be masked by trimap before passing into this metric

Default prefix: ''

Parameters **norm_const** (*int*) – Divide the result to reduce its magnitude. Default to 1000.

Metrics:

- MattingMSE (float): Mean of Squared Error

default_prefix =

metric = MattingMSE

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*)

process(*data_batch: Sequence[dict]*, *data_samples: Sequence[dict]*) → None

Process one batch of data and predictions.

Parameters

- **data_batch** (*Sequence[dict]*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

compute_metrics(*results: List*)

Compute the metrics from processed results.

Parameters **results** (*dict*) – The processed results of each batch.

Returns The computed metrics. The keys are the names of the metrics, and the values are corresponding results.

Return type Dict

class mmagic.evaluation.**MultiScaleStructureSimilarity**(*fake_nums: int*, *fake_key: Optional[str] = None*, *need_cond_input: bool = False*, *sample_model: str = 'ema'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*)

Bases: mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric

MS-SSIM (Multi-Scale Structure Similarity) metric.

Ref: https://github.com/tkarras/progressive_growing_of_gans/blob/master/metrics/ms_ssim.py # noqa

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.

- **real_key** (*Optional[str]*) – Key for get real images from the input dict. Defaults to 'img'.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'ema'.
- **collect_device** (*str, optional*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

name = MS-SSIM

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Feed data to the metric.

Parameters

- **data_batch** (*dict*) – Real images from dataloader. Do not be used in this metric.
- **data_samples** (*Sequence[dict]*) – Generated images.

_collect_target_results(*target: str*) → Optional[list]

Collected results for MS-SSIM metric. Size of *self.fake_results* in MS-SSIM does not relay on *self.fake_nums* but *self.num_pairs*.

Parameters **target** (*str*) – Target results to collect.

Returns The collected results.

Return type Optional[list]

compute_metrics(*results_fake: List*)

Computed the result of MS-SSIM.

Returns Calculated MS-SSIM result.

Return type dict

class mmagic.evaluation.**PerceptualPathLength**(*fake_nums: int*, *real_nums: int = 0*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'gt_img'*, *need_cond_input: bool = False*, *sample_model: str = 'ema'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *crop=True*, *epsilon=0.0001*, *space='W'*, *sampling='end'*, *latent_dim=512*)

Bases: mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric

Perceptual path length.

Measure the difference between consecutive images (their VGG16 embeddings) when interpolating between two random inputs. Drastic changes mean that multiple features have changed together and that they might be entangled.

Ref: https://github.com/rosinality/stylegan2-pytorch/blob/master/ppl.py # noqa

Parameters

- **num_images** (*int*) – The number of evaluated generated samples.
- **image_shape** (*tuple, optional*) – Image shape in order "CHW". Defaults to None.
- **crop** (*bool, optional*) – Whether crop images. Defaults to True.
- **epsilon** (*float, optional*) – Epsilon parameter for path sampling. Defaults to 1e-4.
- **space** (*str, optional*) – Latent space. Defaults to 'W'.
- **sampling** (*str, optional*) – Sampling mode, whether sampling in full path or endpoints. Defaults to 'end'.
- **latent_dim** (*int, optional*) – Latent dimension of input noise. Defaults to 512.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.

SAMPLER_MODE = path

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

_compute_distance(*images*)

Feed data to the metric.

Parameters **images** (*Tensor*) – Input tensor.

compute_metrics(*fake_results: list*) → dict

Summarize the results.

Returns Summarized results.

Return type dict | list

get_metric_sampler(*model: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*, *metrics: list*)

Get sampler for generative metrics. Returns a dummy iterator, whose return value of each iteration is a dict containing batch size and sample mode to generate images.

Parameters

- **model** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for real images. Used to get batch size during generate fake images.
- **metrics** (*list*) – Metrics with the same sampler mode.

Returns Sampler for generative metrics.

Return type dummy_iterator

class mmagic.evaluation.**PrecisionAndRecall**(*fake_nums*, *real_nums=- 1*, *k=3*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'gt_img'*, *need_cond_input: bool = False*, *sample_model: str = 'ema'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*, *vgg16_script='work_dirs/cache/vgg16.pt'*, *vgg16_pkl=None*, *row_batch_size=10000*, *col_batch_size=10000*, *auto_save=True*)

Bases: mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric

Improved Precision and recall metric.

In this metric, we draw real and generated samples respectively, and embed them into a high-dimensional feature space using a pre-trained classifier network. We use these features to estimate the corresponding manifold. We obtain the estimation by calculating pairwise Euclidean distances between all feature vectors in the set and, for each feature vector, construct a hypersphere with radius equal to the distance to its kth nearest neighbor. Together, these hyperspheres define a volume in the feature space that serves as an estimate of the true manifold. Precision is quantified by querying for each generated image whether the image is within the estimated manifold of real images. Symmetrically, recall is calculated by querying for each real image whether the image is within estimated manifold of generated image.

Ref: https://github.com/NVlabs/stylegan2-ada-pytorch/blob/main/metrics/precision_recall.py # noqa

Note that we highly recommend that users should download the vgg16 script module from the following address. Then, the *vgg16_script* can be set with user's local path. If not given, we will use the vgg16 from pytorch model zoo. However, this may bring significant different in the final results.

Tero's vgg16: https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada-pytorch/pretrained/metrics/vgg16.pt

Parameters

- **num_images** (*int*) – The number of evaluated generated samples.
- **image_shape** (*tuple*) – Image shape in order "CHW". Defaults to None.
- **num_real_need** (*int | None, optional*) – The number of real images. Defaults to None.
- **full_dataset** (*bool, optional*) – Whether to use full dataset for evaluation. Defaults to False.
- **k** (*int, optional*) – Kth nearest parameter. Defaults to 3.
- **bgr2rgb** (*bool, optional*) – Whether to change the order of image channel. Defaults to True.
- **vgg16_script** (*str, optional*) – Path for the Tero's vgg16 module. Defaults to 'work_dirs/cache/vgg16.pt'.
- **row_batch_size** (*int, optional*) – The batch size of row data. Defaults to 10000.
- **col_batch_size** (*int, optional*) – The batch size of col data. Defaults to 10000.
- **auto_save** (*bool, optional*) – Whether save vgg feature automatically.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.

name = PR

_load_vgg(*vgg16_script: Optional[str]*) → Tuple[torch.nn.Module, bool]

Load VGG network from the given path.

Parameters **vgg16_script** – The path of script model of VGG network. If None, will load the pytorch version.

Returns

The actually loaded VGG network and corresponding style.

Return type Tuple[nn.Module, str]

extract_features(*images: torch.Tensor*) → torch.Tensor

Extracting image features.

Parameters **images** (*torch.Tensor*) – Images tensor.

Returns Vgg16 features of input images.

Return type torch.Tensor

compute_metrics(*results_fake*) → dict

compute_metrics.

Returns Summarized results.

Return type dict

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

prepare(*module: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*) → None

Prepare for the pre-calculating items of the metric. Defaults to do nothing.

Parameters

- **module** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for the real images.

class mmagic.evaluation.**SlicedWassersteinDistance**(*fake_nums: int*, *image_shape: tuple*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'gt_img'*, *sample_model: str = 'ema'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*)

Bases: `mmagic.evaluation.metrics.base_gen_metric.GenMetric`

SWD (Sliced Wasserstein distance) metric. We calculate the SWD of two sets of images in the following way. In every 'feed', we obtain the Laplacian pyramids of every images and extract patches from the Laplacian pyramids as descriptors. In 'summary', we normalize these descriptors along channel, and reshape them so that we can use these descriptors to represent the distribution of real/fake images. And we can calculate the sliced Wasserstein distance of the real and fake descriptors as the SWD of the real and fake images.

Ref: https://github.com/tkarras/progressive_growing_of_gans/blob/master/metrics/sliced_wasserstein.py # noqa

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **image_shape** (*tuple*) – Image shape in order "CHW".
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **real_key** (*Optional[str]*) – Key for get real images from the input dict. Defaults to 'gt_img'.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'ema'.
- **collect_device** (*str*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

name = SWD

process(*data_batch: dict*, *data_samples: Sequence[dict]*) → None

Process one batch of data samples and predictions. The processed results should be stored in `self.fake_results` and `self.real_results`, which will be used to compute the metrics when all batches have been processed.

Parameters

- **data_batch** (*dict*) – A batch of data from the dataloader.
- **data_samples** (*Sequence[dict]*) – A batch of outputs from the model.

_collect_target_results(*target: str*) → Optional[list]

Collect function for SWD metric. This function support collect results typing as *List[List[Tensor]]*.

Parameters **target** (*str*) – Target results to collect.

Returns The collected results.

Return type Optional[list]

compute_metrics(*results_fake*, *results_real*) → dict

Compute the result of SWD metric.

Parameters

- **fake_results** (*list*) – List of image feature of fake images.
- **real_results** (*list*) – List of image feature of real images.

Returns A dict of the computed SWD metric.

Return type dict

class mmagic.evaluation.**TransFID**(*fake_nums: int*, *real_nums: int = - 1*, *inception_style='StyleGAN'*, *inception_path: Optional[str] = None*, *inception_pkl: Optional[str] = None*, *fake_key: Optional[str] = None*, *real_key: Optional[str] = 'img'*, *sample_model: str = 'ema'*, *collect_device: str = 'cpu'*, *prefix: Optional[str] = None*)

Bases: *FrechetInceptionDistance*

FID metric. In this metric, we calculate the distance between real distributions and fake distributions. The distributions are modeled by the real samples and fake samples, respectively. *Inception_v3* is adopted as the feature extractor, which is widely used in StyleGAN and BigGAN.

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **real_nums** (*int*) – Numbers of the real images need for the metric. If -1 is passed, means all real images in the dataset will be used. Defaults to -1.
- **inception_style** (*str*) – The target inception style want to load. If the given style cannot be loaded successful, will attempt to load a valid one. Defaults to 'StyleGAN'.
- **inception_path** (*str, optional*) – Path the the pretrain Inception network. Defaults to None.
- **inception_pkl** (*str, optional*) – Path to reference inception pickle file. If *None*, the statistical value of real distribution will be calculated at running time. Defaults to None.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **real_key** (*Optional[str]*) – Key for get real images from the input dict. Defaults to 'img'.
- **need_cond_input** (*bool*) – If true, the sampler will return the conditional input randomly sampled from the original dataset. This require the dataset implement *get_data_info* and field *gt_label* must be contained in the return value of *get_data_info*. Noted that, for unconditional models, set *need_cond_input* as True may influence the result of evaluation results since the conditional inputs are sampled from the dataset distribution; otherwise will be sampled from the uniform distribution. Defaults to False.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'orig'.
- **collect_device** (*str, optional*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

get_metric_sampler(*model: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*, *metrics: List[mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric]*) → torch.utils.data.dataloader.DataLoader

Get sampler for normal metrics. Directly returns the dataloader.

Parameters

- **model** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for real images.
- **metrics** (*List['GenMetric']*) – Metrics with the same sample mode.

Returns Default sampler for normal metrics.

Return type DataLoader

class mmagic.evaluation.**TransIS**(*fake_nums: int = 50000*, *resize: bool = True*, *splits: int = 10*, *inception_style: str = 'StyleGAN'*, *inception_path: Optional[str] = None*, *resize_method='bicubic'*, *use_pillow_resize: bool = True*, *fake_key: Optional[str] = None*, *sample_model='ema'*, *collect_device: str = 'cpu'*, *prefix: str = None*)

Bases: *InceptionScore*

IS (Inception Score) metric. The images are split into groups, and the inception score is calculated on each group of images, then the mean and standard deviation of the score is reported. The calculation of the inception score on a group of images involves first using the inception v3 model to calculate the conditional probability for each image (p(y|x)). The marginal probability is then calculated as the average of the conditional probabilities for the images in the group (p(y)). The KL divergence is then calculated for each image as the conditional probability multiplied by the log of the conditional probability minus the log of the marginal probability. The KL divergence is then summed over all images and averaged over all classes and the exponent of the result is calculated to give the final score.

Ref: https://github.com/sbarratt/inception-score-pytorch/blob/master/inception_score.py # noqa

Note that we highly recommend that users should download the Inception V3 script module from the following address. Then, the *inception_pkl* can be set with user's local path. If not given, we will use the Inception V3 from pytorch model zoo. However, this may bring significant different in the final results.

Tero's Inception V3: https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada-pytorch/pretrained/metrics/inception-2015-12-05.pt # noqa

Parameters

- **fake_nums** (*int*) – Numbers of the generated image need for the metric.
- **resize** (*bool, optional*) – Whether resize image to 299x299. Defaults to True.
- **splits** (*int, optional*) – The number of groups. Defaults to 10.
- **inception_style** (*str*) – The target inception style want to load. If the given style cannot be loaded successful, will attempt to load a valid one. Defaults to 'StyleGAN'.
- **inception_path** (*str, optional*) – Path the the pretrain Inception network. Defaults to None.
- **resize_method** (*str*) – Resize method. If *resize* is False, this will be ignored. Defaults to 'bicubic'.
- **use_pil_resize** (*bool*) – Whether use Bicubic interpolation with Pillow's backend. If set as True, the evaluation process may be a little bit slow, but achieve a more accurate IS result. Defaults to False.
- **fake_key** (*Optional[str]*) – Key for get fake images of the output dict. Defaults to None.
- **sample_model** (*str*) – Sampling mode for the generative model. Support 'orig' and 'ema'. Defaults to 'ema'.
- **collect_device** (*str, optional*) – Device name used for collecting results from different ranks during distributed training. Must be 'cpu' or 'gpu'. Defaults to 'cpu'.
- **prefix** (*str, optional*) – The prefix that will be added in the metric names to disambiguate homonymous metrics of different evaluators. If prefix is not provided in the argument, self.default_prefix will be used instead. Defaults to None.

get_metric_sampler(*model: torch.nn.Module*, *dataloader: torch.utils.data.dataloader.DataLoader*, *metrics: List[mmagic.evaluation.metrics.base_gen_metric.GenerativeMetric]*) → torch.utils.data.dataloader.DataLoader

Get sampler for normal metrics. Directly returns the dataloader.

Parameters

- **model** (*nn.Module*) – Model to evaluate.
- **dataloader** (*DataLoader*) – Dataloader for real images.
- **metrics** (*List['GenMetric']*) – Metrics with the same sample mode.

Returns Default sampler for normal metrics.

Return type DataLoader

1.82 mmagic.visualization

1.82.1 Package Contents

Classes

ConcatImageVisualizer	Visualize multiple images by concatenation.
PaviVisBackend	Visualization backend for Pavi.
TensorboardVisBackend	Tensorboard visualization backend class.
VisBackend	MMagic visualization backend class. It can write image, config, scalars,
WandbVisBackend	Wandb visualization backend for MMagic.
Visualizer	MMagic Visualizer.

class mmagic.visualization.**ConcatImageVisualizer**(*fn_key: str*, *img_keys: Sequence[str]*, *pixel_range={}*, *bgr2rgb=False*, *name: str = 'visualizer'*, **args*, ***kwargs*)

Bases: mmengine.visualization.Visualizer

Visualize multiple images by concatenation.

This visualizer will horizontally concatenate images belongs to different keys and vertically concatenate images belongs to different frames to visualize.

Image to be visualized can be:

- torch.Tensor or np.array
- Image sequences of shape (T, C, H, W)
- Multi-channel image of shape (1/3, H, W)
- Single-channel image of shape (C, H, W)

Parameters

- **fn_key** (*str*) – key used to determine file name for saving image. Usually it is the path of some input image. If the value is *dir/basename.ext*, the name used for saving will be basename.
- **img_keys** (*str*) – keys, values of which are images to visualize.
- **pixel_range** (*dict*) – min and max pixel value used to denormalize images, note that only float array or tensor will be denormalized, uint8 arrays are assumed to be unnormalized.
- **bgr2rgb** (*bool*) – whether to convert the image from BGR to RGB.
- **name** (*str*) – name of visualizer. Default: 'visualizer'.
- ****kwargs** (**args and*) – Other arguments are passed to *Visualizer*. # noqa

add_datasample(*data_sample:* mmagic.structures.DataSample, *step=0*) → None

Concatenate image and draw.

Parameters

- **input** (*torch.Tensor*) – Single input tensor from data_batch.
- **data_sample** (*DataSample*) – Single data_sample from data_batch.
- **output** (*DataSample*) – Single prediction output by model.
- **step** (*int*) – Global step value to record. Default: 0.

class mmagic.visualization.**PaviVisBackend**(*save_dir: str*, *exp_name: Optional[str] = None*, *labels: Optional[str] = None*, *project: Optional[str] = None*, *model: Optional[str] = None*, *description: Optional[str] = None*)

Bases: mmengine.visualization.BaseVisBackend

Visualization backend for Pavi.

property experiment: *VisBackend*

Return the experiment object associated with this visualization backend.

_init_env()

Init save dir.

add_image(*name: str*, *image: numpy.array*, *step: int = 0*, ***kwargs*) → None

Record the image to Pavi.

Parameters

- **name** (*str*) – The image identifier.
- **image** (*np.ndarray*) – The image to be saved. The format should be RGB. Default to None.
- **step** (*int*) – Global step value to record. Default to 0.

add_scalar(*name: str*, *value: Union[int, float, torch.Tensor, numpy.ndarray]*, *step: int = 0*, ***kwargs*) → None

Record the scalar data to Pavi.

Parameters

- **name** (*str*) – The scalar identifier.
- **value** (*int, float, torch.Tensor, np.ndarray*) – Value to save.
- **step** (*int*) – Global step value to record. Default to 0.

add_scalars(*scalar_dict: dict*, *step: int = 0*, *file_path: Optional[str] = None*, ***kwargs*) → None

Record the scalars to Pavi.

The scalar dict will be written to the default and specified files if `file_path` is specified.

Parameters

- **scalar_dict** (*dict*) – Key-value pair storing the tag and corresponding values. The value must be dumped into json format.
- **step** (*int*) – Global step value to record. Default to 0.
- **file_path** (*str, optional*) – The scalar's data will be saved to the `file_path` file at the same time if the `file_path` parameter is specified. Default to None.

class mmagic.visualization.**TensorboardVisBackend**(*save_dir: str*)

Bases: mmengine.visualization.TensorboardVisBackend

Tensorboard visualization backend class.

It can write images, config, scalars, etc. to a tensorboard file.

Examples

```
>>> from mmengine.visualization import TensorboardVisBackend
>>> import numpy as np
>>> vis_backend = TensorboardVisBackend(save_dir='temp_dir')
>>> img = np.random.randint(0, 256, size=(10, 10, 3))
>>> vis_backend.add_image('img', img)
>>> vis_backend.add_scaler('mAP', 0.6)
>>> vis_backend.add_scalars({'loss': 0.1,'acc':0.8})
>>> cfg = Config(dict(a=1, b=dict(b1=[0, 1])))
>>> vis_backend.add_config(cfg)
```

Parameters **save_dir** (*str*) – The root directory to save the files produced by the backend.

add_image(*name: str*, *image: numpy.array*, *step: int = 0*, ***kwargs*)

Record the image to Tensorboard. Additional support upload gif files.

Parameters

- **name** (*str*) – The image identifier.
- **image** (*np.ndarray*) – The image to be saved. The format should be RGB.
- **step** (*int*) – Useless parameter. Wandb does not need this parameter. Default to 0.

class mmagic.visualization.**VisBackend**(*save_dir: str*, *img_save_dir: str = 'vis_image'*, *config_save_file: str = 'config.py'*, *scalar_save_file: str = 'scalars.json'*, *ceph_path: Optional[str] = None*, *delete_local_image: bool = True*)

Bases: mmengine.visualization.BaseVisBackend

MMagic visualization backend class. It can write image, config, scalars, etc. to the local hard disk and ceph path. You can get the drawing backend through the experiment property for custom drawing.

Examples

```
>>> from mmagic.visualization import VisBackend
>>> import numpy as np
>>> vis_backend = VisBackend(save_dir='temp_dir',
>>>                          ceph_path='s3://temp-bucket')
>>> img = np.random.randint(0, 256, size=(10, 10, 3))
>>> vis_backend.add_image('img', img)
>>> vis_backend.add_scalar('mAP', 0.6)
>>> vis_backend.add_scalars({'loss': [1, 2, 3], 'acc': 0.8})
>>> cfg = Config(dict(a=1, b=dict(b1=[0, 1])))
>>> vis_backend.add_config(cfg)
```

Parameters

- **save_dir** (*str*) – The root directory to save the files produced by the visualizer.
- **img_save_dir** (*str*) – The directory to save images. Default to 'vis_image'.
- **config_save_file** (*str*) – The file name to save config. Default to 'config.py'.
- **scalar_save_file** (*str*) – The file name to save scalar values. Default to 'scalars.json'.
- **ceph_path** (*Optional[str]*) – The remote path of Ceph cloud storage. Defaults to None.
- **delete_local** (*bool*) – Whether delete local after uploading to ceph or not. If `ceph_path` is None, this will be ignored. Defaults to True.

property experiment: *VisBackend*

Return the experiment object associated with this visualization backend.

_init_env()

Setup env for VisBackend.

add_config(*config: mmengine.config.Config*, ***kwargs*) → None

Record the config to disk.

Parameters **config** (*Config*) – The Config object

add_image(*name: str*, *image: numpy.array*, *step: int = 0*, ***kwargs*) → None

Record the image to disk.

Parameters

- **name** (*str*) – The image identifier.
- **image** (*np.ndarray*) – The image to be saved. The format should be RGB. Default to None.
- **step** (*int*) – Global step value to record. Default to 0.

add_scalar(*name: str*, *value: Union[int, float, torch.Tensor, numpy.ndarray]*, *step: int = 0*, ***kwargs*) → None

Record the scalar data to disk.

Parameters

- **name** (*str*) – The scalar identifier.
- **value** (*int, float, torch.Tensor, np.ndarray*) – Value to save.
- **step** (*int*) – Global step value to record. Default to 0.

add_scalars(*scalar_dict: dict*, *step: int = 0*, *file_path: Optional[str] = None*, ***kwargs*) → None

Record the scalars to disk.

The scalar dict will be written to the default and specified files if `file_path` is specified.

Parameters

- **scalar_dict** (*dict*) – Key-value pair storing the tag and corresponding values. The value must be dumped into json format.
- **step** (*int*) – Global step value to record. Default to 0.
- **file_path** (*str, optional*) – The scalar's data will be saved to the `file_path` file at the same time if the `file_path` parameter is specified. Default to None.

_dump(*value_dict: dict*, *file_path: str*, *file_format: str*) → None

dump dict to file.

Parameters

- **value_dict** (*dict*) – The dict data to saved.
- **file_path** (*str*) – The file path to save data.
- **file_format** (*str*) – The file format to save data.

_upload(*path: str*, *delete_local=False*) → None

Upload file at path to remote.

Parameters **path** (*str*) – Path of file to upload.

class mmagic.visualization.**WandbVisBackend**(*save_dir: str*, *init_kwargs: Optional[dict] = None*, *define_metric_cfg: Union[dict, list, None] = None*, *commit: Optional[bool] = True*, *log_code_name: Optional[str] = None*, *watch_kwargs: Optional[dict] = None*)

Bases: mmengine.visualization.WandbVisBackend

Wandb visualization backend for MMagic.

_init_env()

Setup env for wandb.

add_image(*name: str*, *image: numpy.array*, *step: int = 0*, ***kwargs*)

Record the image to wandb. Additional support upload gif files.

Parameters

- **name** (*str*) – The image identifier.
- **image** (*np.ndarray*) – The image to be saved. The format should be RGB.
- **step** (*int*) – Useless parameter. Wandb does not need this parameter. Default to 0.

class mmagic.visualization.**Visualizer**(*name='visualizer'*, *vis_backends: Optional[List[Dict]] = None*, *save_dir: Optional[str] = None*)

Bases: mmengine.visualization.Visualizer

MMagic Visualizer.

Parameters

- **name** (*str*) – Name of the instance. Defaults to 'visualizer'.
- **vis_backends** (*list, optional*) – Visual backend config list. Defaults to None.
- **save_dir** (*str, optional*) – Save file dir for all storage backends. If it is None, the backend storage will not save any data.

Examples:

```
>>> # Draw image
>>> vis = Visualizer()
>>> vis.add_datasample(
>>>     'random_noise',
>>>     gen_samples=torch.rand(2, 3, 10, 10),
>>>     gt_samples=dict(imgs=torch.randn(2, 3, 10, 10)),
>>>     gt_keys='imgs',
```

(continues on next page)

(continued from previous page)

```
>>>     vis_mode='image',
>>>     n_rows=2,
>>>     step=10)
```

static _post_process_image(*image: torch.Tensor*) → torch.Tensor

Post process images.

Parameters **image** (*Tensor*) – Image to post process. The value range of image should be in [0, 255] and the channel order should be BGR.

Returns Image in RGB color order.

Return type Tensor

static _get_n_row_and_padding(*samples: Tuple[dict, torch.Tensor]*, *n_row: Optional[int] = None*) → Tuple[int, Optional[torch.Tensor]]

Get number of sample in each row and tensor for padding the empty position.

Parameters

- **samples** (*Tuple[dict, Tensor]*) – Samples to visualize.
- **n_row** (*int, optional*) – Number of images displayed in each row of. If not passed, n_row will be set as `int(sqrt(batch_size))`.

Returns

Number of sample in each row and tensor for padding the empty position.

Return type Tuple[int, Optional[int]]

_vis_gif_sample(*gen_samples: mmagic.utils.typing.SampleList*, *target_keys: Union[str, List[str], None]*, *n_row: int*) → numpy.ndarray

Visualize gif samples.

Parameters

- **gen_samples** (*SampleList*) – List of data samples to visualize
- **target_keys** (*Union[str, List[str], None]*) – Keys of the visualization target in data samples.
- **n_rows** (*int, optional*) – Number of images in one row.

Returns The visualization results.

Return type np.ndarray

_vis_image_sample(*gen_samples: mmagic.utils.typing.SampleList*, *target_keys: Union[str, List[str], None]*, *n_row: int*) → numpy.ndarray

Visualize image samples.

Parameters

- **gen_samples** (*SampleList*) – List of data samples to visualize
- **target_keys** (*Union[str, List[str], None]*) – Keys of the visualization target in data samples.
- **color_order** (*str*) – The color order of the passed images.
- **target_mean** (*Sequence[Union[float, int]]*) – The target mean of the visualization results.

- **target_std** (*Sequence[Union[float, int]]*) – The target std of the visualization results.
- **n_rows** (*int, optional*) – Number of images in one row.

Returns The visualization results.

Return type np.ndarray

_get_vis_data_by_key(*sample:* mmagic.structures.DataSample, *key: str*) → torch.Tensor

Get tensor in `DataSample` by the given key.

Parameters

- **sample** (*DataSample*) – Input data sample.
- **key** (*str*) – Name of the target tensor.

Returns Tensor from the data sample.

Return type Tensor

add_datasample(*name: str, *, gen_samples: Sequence[*mmagic.structures.DataSample*], target_keys: Optional[Tuple[str, List[str]]] = None, vis_mode: Optional[str] = None, n_row: Optional[int] = None, show: bool = False, wait_time: int = 0, step: int = 0, **kwargs*) → None

Draw datasample and save to all backends.

If GT and prediction are plotted at the same time, they are displayed in a stitched image where the left image is the ground truth and the right image is the prediction.

If `show` is True, all storage backends are ignored, and the images will be displayed in a local window.

Parameters

- **name** (*str*) – The image identifier.
- **gen_samples** (*List[DataSample]*) – Data samples to visualize.
- **vis_mode** (*str, optional*) – Visualization mode. If not passed, will visualize results as image. Defaults to None.
- **n_rows** (*int, optional*) – Number of images in one row. Defaults to None.
- **color_order** (*str*) – The color order of the passed images. Defaults to 'bgr'.
- **target_mean** (*Sequence[Union[float, int]]*) – The target mean of the visualization results. Defaults to 127.5.
- **target_std** (*Sequence[Union[float, int]]*) – The target std of the visualization results. Defaults to 127.5.
- **show** (*bool*) – Whether to display the drawn image. Default to False.
- **wait_time** (*float*) – The interval of show (s). Defaults to 0.
- **step** (*int*) – Global step value to record. Defaults to 0.

add_image(*name: str, image: numpy.ndarray, step: int = 0, **kwargs*) → None

Record the image. Support input kwargs.

Parameters

- **name** (*str*) – The image identifier.
- **image** (*np.ndarray, optional*) – The image to be saved. The format should be RGB. Default to None.

- **step** (*int*) – Global step value to record. Default to 0.

1.83 mmagic.engine.hooks

1.83.1 Package Contents

Classes

ExponentialMovingAverageHook	Exponential Moving Average Hook.
IterTimerHook	IterTimerHooks inherits from `mmengine.hooks.IterTimerHook` and
PGGANFetchDataHook	PGGAN Fetch Data Hook.
PickleDataHook	Pickle Useful Data Hook.
ReduceLRSchedulerHook	A hook to update learning rate.
BasicVisualizationHook	Basic hook that invoke visualizers during validation and test.
VisualizationHook	MMagic Visualization Hook. Used to visual output samples in training,

class mmagic.engine.hooks.**ExponentialMovingAverageHook**(*module_keys*, *interp_mode='lerp'*, *interp_cfg=None*, *interval=- 1*, *start_iter=0*)

Bases: `mmengine.hooks.Hook`

Exponential Moving Average Hook.

Exponential moving average is a trick that widely used in current GAN literature, e.g., PGGAN, StyleGAN, and BigGAN. This general idea of it is maintaining a model with the same architecture, but its parameters are updated as a moving average of the trained weights in the original model. In general, the model with moving averaged weights achieves better performance.

Parameters

- **module_keys** (*str | tuple[str]*) – The name of the ema model. Note that we require these keys are followed by '_ema' so that we can easily find the original model by discarding the last four characters.
- **interp_mode** (*str, optional*) – Mode of the interpolation method. Defaults to 'lerp'.
- **interp_cfg** (*dict | None, optional*) – Set arguments of the interpolation function. Defaults to None.
- **interval** (*int, optional*) – Evaluation interval (by iterations). Default: -1.
- **start_iter** (*int, optional*) – Start iteration for ema. If the start iteration is not reached, the weights of ema model will maintain the same as the original one. Otherwise, its parameters are updated as a moving average of the trained weights in the original model. Default: 0.

static lerp(*a*, *b*, *momentum=0.001*, *momentum_nontrainable=1.0*, *trainable=True*)

Does a linear interpolation of two parameters/ buffers.

Parameters

- **a** (*torch.Tensor*) – Interpolation start point, refer to orig state.
- **b** (*torch.Tensor*) – Interpolation end point, refer to ema state.

- **momentum** (*float, optional*) – The weight for the interpolation formula. Defaults to 0.001.
- **momentum_nontrainable** (*float, optional*) – The weight for the interpolation formula used for nontrainable parameters. Defaults to 1..
- **trainable** (*bool, optional*) – Whether input parameters is trainable. If set to False, momentum_nontrainable will be used. Defaults to True.

Returns Interpolation result.

Return type torch.Tensor

every_n_iters(*runner: mmengine.runner.Runner*, *n: int*)

This is the function to perform every n iterations.

Parameters

- **runner** (*Runner*) – runner used to drive the whole pipeline
- **n** (*int*) – the number of iterations

Returns the latest iterations

Return type int

after_train_iter(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: DATA_BATCH = None*, *outputs: Optional[dict] = None*) → None

This is the function to perform after each training iteration.

Parameters

- **runner** (*Runner*) – runner to drive the pipeline
- **batch_idx** (*int*) – the id of batch
- **data_batch** (*DATA_BATCH, optional*) – data batch. Defaults to None.
- **outputs** (*Optional[dict], optional*) – output. Defaults to None.

before_run(*runner: mmengine.runner.Runner*)

This is the function perform before each run.

Parameters **runner** (*Runner*) – runner used to drive the whole pipeline

Raises **RuntimeError** – error message

class mmagic.engine.hooks.**IterTimerHook**

Bases: mmengine.hooks.IterTimerHook

IterTimerHooks inherits from mmengine.hooks.IterTimerHook and overwrites self._after_iter().

This hooks should be used along with *mmagic.engine.runner.MultiValLoop* and *mmagic.engine.runner.MultiTestLoop*.

_after_iter(*runner*, *batch_idx: int*, *data_batch: DATA_BATCH = None*, *outputs: Optional[Union[dict, Sequence[mmengine.structures.BaseDataElement]]] = None*, *mode: str = 'train'*) → None

Calculating time for an iteration and updating "time" HistoryBuffer of runner.message_hub. If *mode* is 'train', we take *runner.max_iters* as the total iterations and calculate the rest time. If *mode* in *val* or *test*, we use *runner.val_loop.total_length* or *runner.test_loop.total_length* as total number of iterations. If you want to know how *total_length* is calculated, please refers to *mmagic.engine.runner.MultiValLoop.run()* and *mmagic.engine.runner.MultiTestLoop.run()*.

Parameters

- **runner** (*Runner*) – The runner of the training validation and testing process.
- **batch_idx** (*int*) – The index of the current batch in the loop.
- **data_batch** (*Sequence[dict], optional*) – Data from dataloader. Defaults to None.
- **outputs** (*dict or sequence, optional*) – Outputs from model. Defaults to None.
- **mode** (*str*) – Current mode of runner. Defaults to 'train'.

class mmagic.engine.hooks.**PGGANFetchDataHook**

Bases: mmengine.hooks.Hook

PGGAN Fetch Data Hook.

Parameters **interval** (*int, optional*) – The interval of calling this hook. If set to -1, the visualization hook will not be called. Defaults to 1.

before_train_iter(*runner*, *batch_idx: int*, *data_batch: DATA_BATCH = None*) → None

All subclasses should override this method, if they need any operations before each training iteration.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the train loop.
- **data_batch** (*dict or tuple or list, optional*) – Data from dataloader.

update_dataloader(*dataloader: torch.utils.data.dataloader.DataLoader*, *curr_scale: int*) → Optional[torch.utils.data.dataloader.DataLoader]

Update the data loader.

Parameters

- **dataloader** (*DataLoader*) – The dataloader to be updated.
- **curr_scale** (*int*) – The current scale of the generated image.

Returns

The updated dataloader. If the dataloader do not need to update, return None.

Return type Optional[DataLoader]

class mmagic.engine.hooks.**PickleDataHook**(*output_dir*, *data_name_list*, *interval=- 1*, *before_run=False*, *after_run=False*, *filename_tmpl='iter_{}.pkl'*)

Bases: mmengine.hooks.Hook

Pickle Useful Data Hook.

This hook will be used in SinGAN training for saving some important data that will be used in testing or inference.

Parameters

- **output_dir** (*str*) – The output path for saving pickled data.
- **data_name_list** (*list[str]*) – The list contains the name of results in outputs dict.
- **interval** (*int*) – The interval of calling this hook. If set to -1, the PickleDataHook will not be called during training. Default: -1.
- **before_run** (*bool, optional*) – Whether to save before running. Defaults to False.
- **after_run** (*bool, optional*) – Whether to save after running. Defaults to False.

- **filename_tmpl** (*str, optional*) – Format string used to save images. The output file name will be formatted as this args. Defaults to 'iter_{}.pkl'.

after_run(*runner*)

The behavior after each train iteration.

Parameters **runner** (*object*) – The runner.

before_run(*runner*)

The behavior after each train iteration.

Parameters **runner** (*object*) – The runner.

after_train_iter(*runner*, *batch_idx: int*, *data_batch: DATA_BATCH = None*, *outputs: Optional[dict] = None*)

The behavior after each train iteration.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the train loop.
- **data_batch** (*Sequence[dict], optional*) – Data from dataloader. Defaults to None.
- **outputs** (*dict, optional*) – Outputs from model. Defaults to None.

_pickle_data(*runner: mmengine.runner.Runner*)

Save target data to pickle file.

Parameters **runner** (*Runner*) – The runner of the training process.

_get_numpy_data(*data: Tuple[List[torch.Tensor], torch.Tensor, int]*) → Tuple[List[numpy.ndarray], numpy.ndarray, int]

Convert tensor or list of tensor to numpy or list of numpy.

Parameters **data** (*Tuple[List[Tensor], Tensor, int]*) – Data to be converted.

Returns Converted data.

Return type Tuple[List[np.ndarray], np.ndarray, int]

class mmagic.engine.hooks.**ReduceLRSchedulerHook**(*val_metric: str = None*, *by_epoch=True*, *interval=1*)

Bases: mmengine.hooks.ParamSchedulerHook

A hook to update learning rate.

Parameters

- **val_metric** (*str*) – The metric of validation. If val_metric is not None, we check val_metric to reduce learning. Default: None.
- **by_epoch** (*bool*) – Whether to update by epoch. Default: True.
- **interval** (*int*) – The interval of iterations to update. Default: 1.

_calculate_average_value()

after_train_epoch(*runner: mmengine.runner.Runner*)

Call step function for each scheduler after each train epoch.

Parameters **runner** (*Runner*) – The runner of the training process.

after_train_iter(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: DATA_BATCH = None*, *outputs: Optional[dict] = None*) → None

Call step function for each scheduler after each iteration.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the train loop.
- **data_batch** (*Sequence[dict], optional*) – Data from dataloader. In order to keep this interface consistent with other hooks, we keep `data_batch` here. Defaults to None.
- **outputs** (*dict, optional*) – Outputs from model. In order to keep this interface consistent with other hooks, we keep `data_batch` here. Defaults to None.

after_val_epoch(*runner*, *metrics: Optional[Dict[str, float]] = None*)

Call step function for each scheduler after each validation epoch.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **metrics** (*dict, optional*) – The metrics of validation. Default: None.

class mmagic.engine.hooks.**BasicVisualizationHook**(*interval: dict = {}*, *on_train=False*, *on_val=True*, *on_test=True*)

Bases: `mmengine.hooks.Hook`

Basic hook that invoke visualizers during validation and test.

Parameters

- **interval** (*int | dict*) – Visualization interval. Default: {}.
- **on_train** (*bool*) – Whether to call hook during train. Default to False.
- **on_val** (*bool*) – Whether to call hook during validation. Default to True.
- **on_test** (*bool*) – Whether to call hook during test. Default to True.

priority = NORMAL

_after_iter(*runner*, *batch_idx: int*, *data_batch: Optional[Sequence[dict]]*, *outputs: Optional[Sequence[mmengine.structures.BaseDataElement]]*, *mode=None*) → None

Show or Write the predicted results.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the test loop.
- **data_batch** (*Sequence[dict], optional*) – Data from dataloader. Defaults to None.
- **outputs** (*Sequence[BaseDataElement], optional*) – Outputs from model. Defaults to None.

class mmagic.engine.hooks.**VisualizationHook**(*interval: int = 1000, vis_kwargs_list: Tuple[List[dict], dict] = None, fixed_input: bool = True, n_samples: Optional[int] = 64, n_row: Optional[int] = None, message_hub_vis_kwargs: Optional[Tuple[str, dict, List[str], List[Dict]]] = None, save_at_test: bool = True, max_save_at_test: int = 100, test_vis_keys: Optional[Union[str, List[str]]] = None, show: bool = False, wait_time: float = 0*)

Bases: mmengine.hooks.Hook

MMagic Visualization Hook. Used to visual output samples in training, validation and testing. In this hook, we use a list called *sample_kwargs_list* to control how to generate samples and how to visualize them. Each element in *sample_kwargs_list*, called *sample_kwargs*, may contains the following keywords:

- **Required key words:**
  - **'type': Value must be string. Denotes what kind of sampler is used to** generate image. Refers to *get_sampler()*.
- **Optional key words (If not passed, will use the default value):**
  - 'n_row': Value must be int. The number of images in one row.
  - 'num_samples': Value must be int. The number of samples to visualize.
  - **'vis_mode': Value must be string. How to visualize the generated** samples (e.g. image, gif).
  - **'fixed_input': Value must be bool. Whether use the fixed input** during the loop.
  - 'draw_gt': Value must be bool. Whether save the real images.
  - **'target_keys': Value must be string or list of string. The keys of** the target image to visualize.
  - **'name': Value must be string. If not passed, will use** *sample_kwargs['type']* as default.

For convenience, we also define a group of alias of samplers' type for models supported in MMagic. Refers to *:attr:self.SAMPLER_TYPE_MAPPING*.

Example

```
>>> # for GAN models
>>> custom_hooks = [
>>>     dict(
>>>         type='VisualizationHook',
>>>         interval=1000,
>>>         fixed_input=True,
>>>         vis_kwargs_list=dict(type='GAN', name='fake_img'))]
>>> # for Translation models
>>> custom_hooks = [
>>>     dict(
>>>         type='VisualizationHook',
>>>         interval=10,
>>>         fixed_input=False,
>>>         vis_kwargs_list=[dict(type='Translation',
>>>                               name='translation_train',
>>>                               n_samples=6, draw_gt=True,
>>>                               n_row=3),
>>>                          dict(type='TranslationVal',
```

(continues on next page)

(continued from previous page)

```
>>>                                     name='translation_val',
>>>                                     n_samples=16, draw_gt=True,
>>>                                     n_row=4)])]
```

NOTE: user-defined vis_kwargs > vis_kwargs_mapping > hook init args

Parameters

- **interval** (*int*) – Visualization interval. Default: 1000.
- **sampler_kwargs_list** (*Tuple[List[dict], dict]*) – The list of sampling behavior to generate images.
- **fixed_input** (*bool*) – The default action of whether use fixed input to generate samples during the loop. Defaults to True.
- **n_samples** (*Optional[int]*) – The default value of number of samples to visualize. Defaults to 64.
- **n_row** (*Optional[int]*) – The default value of number of images in each row in the visualization results. Defaults to None.
- **(Optional[Tuple[str** (*message_hub_vis_kwargs*) – List[Dict]]]): Key arguments visualize images in message hub. Defaults to None.
- **dict** – List[Dict]]]): Key arguments visualize images in message hub. Defaults to None.
- **List[str]** – List[Dict]]]): Key arguments visualize images in message hub. Defaults to None.

:param [List[Dict]]]): Key arguments visualize images in message hub.] Defaults to None.

Parameters

- **save_at_test** (*bool*) – Whether save images during test. Defaults to True.
- **max_save_at_test** (*int*) – Maximum number of samples saved at test time. If None is passed, all samples will be saved. Defaults to 100.
- **show** (*bool*) – Whether to display the drawn image. Default to False.
- **wait_time** (*float*) – The interval of show (s). Defaults to 0.

priority = NORMAL

VIS_KWARGS_MAPPING

after_val_iter(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: dict*, *outputs*) → None

VisualizationHook do not support visualize during validation.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the test loop.
- **data_batch** (*Sequence[dict], optional*) – Data from dataloader. Defaults to None.
- **outputs** – outputs of the generation model

after_test_iter(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: dict*, *outputs*)

Visualize samples after test iteration.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the test loop.
- **data_batch** (*dict, optional*) – Data from dataloader. Defaults to None.
- **outputs** – outputs of the generation model Defaults to None.

after_train_iter(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: dict = None*, *outputs: Optional[dict] = None*) → None

Visualize samples after train iteration.

Parameters

- **runner** (*Runner*) – The runner of the training process.
- **batch_idx** (*int*) – The index of the current batch in the train loop.
- **data_batch** (*dict*) – Data from dataloader. Defaults to None.
- **outputs** (*dict, optional*) – Outputs from model. Defaults to None.

vis_sample(*runner: mmengine.runner.Runner*, *batch_idx: int*, *data_batch: dict*, *outputs: Optional[dict] = None*) → None

Visualize samples.

Parameters

- **runner** (*Runner*) – The runner contains model to visualize.
- **batch_idx** (*int*) – The index of the current batch in loop.
- **data_batch** (*dict*) – Data from dataloader. Defaults to None.
- **outputs** (*dict, optional*) – Outputs from model. Defaults to None.

vis_from_message_hub(*batch_idx: int*)

Visualize samples from message hub.

Parameters

- **batch_idx** (*int*) – The index of the current batch in the test loop.
- **color_order** (*str*) – The color order of generated images.
- **target_mean** (*Sequence[Union[float, int]]*) – The original mean of the image tensor before preprocessing. Image will be re-shifted to `target_mean` before visualizing.
- **target_std** (*Sequence[Union[float, int]]*) – The original std of the image tensor before preprocessing. Image will be re-scaled to `target_std` before visualizing.

1.84 mmagic.engine.optimizers

1.84.1 Package Contents

Classes

MultiOptimWrapperConstructor	OptimizerConstructor for GAN models. This class construct optimizer for
PGGANOptimWrapperConstructor	OptimizerConstructor for PGGAN models. Set optimizers for each
SinGANOptimWrapperConstructor	OptimizerConstructor for SinGAN models. Set optimizers for each

class mmagic.engine.optimizers.**MultiOptimWrapperConstructor**(*optim_wrapper_cfg: dict, paramwise_cfg=None*)

OptimizerConstructor for GAN models. This class construct optimizer for the submodules of the model separately, and return a `mmengine.optim.OptimWrapperDict` or `mmengine.optim.OptimWrapper`.

Example 1: Build multi optimizers (e.g., GANs):

```
>>> # build GAN model
>>> model = dict(
>>>     type='GANModel',
>>>     num_classes=10,
>>>     generator=dict(type='Generator'),
>>>     discriminator=dict(type='Discriminator'))
>>> gan_model = MODELS.build(model)
>>> # build constructor
>>> optim_wrapper = dict(
>>>     generator=dict(
>>>         type='OptimWrapper',
>>>         accumulative_counts=1,
>>>         optimizer=dict(type='Adam', lr=0.0002,
>>>                        betas=(0.5, 0.999))),
>>>     discriminator=dict(
>>>         type='OptimWrapper',
>>>         accumulative_counts=1,
>>>         optimizer=dict(type='Adam', lr=0.0002,
>>>                            betas=(0.5, 0.999))))
>>> optim_dict_builder = MultiOptimWrapperConstructor(optim_wrapper)
>>> # build optim wrapper dict
>>> optim_wrapper_dict = optim_dict_builder(gan_model)
```

Example 2: Build multi optimizers for specific submodules:

```
>>> # build model
>>> class GAN(nn.Module):
>>>     def __init__(self) -> None:
>>>         super().__init__()
>>>         self.generator = nn.Conv2d(3, 3, 1)
>>>         self.discriminator = nn.Conv2d(3, 3, 1)
>>> class TextEncoder(nn.Module):
```

(continues on next page)

(continued from previous page)

```
>>>     def __init__(self):
>>>         super().__init__()
>>>         self.embedding = nn.Embedding(100, 100)
>>> class ToyModel(nn.Module):
>>>     def __init__(self) -> None:
>>>         super().__init__()
>>>         self.m1 = GAN()
>>>         self.m2 = nn.Conv2d(3, 3, 1)
>>>         self.m3 = nn.Linear(2, 2)
>>>         self.text_encoder = TextEncoder()
>>> model = ToyModel()
>>> # build constructor
>>> optim_wrapper = {
>>>     '.*embedding': {
>>>         'type': 'OptimWrapper',
>>>         'optimizer': {
>>>             'type': 'Adam',
>>>             'lr': 1e-4,
>>>             'betas': (0.9, 0.99)
>>>         }
>>>     },
>>>     'm1.generator': {
>>>         'type': 'OptimWrapper',
>>>         'optimizer': {
>>>             'type': 'Adam',
>>>             'lr': 1e-5,
>>>             'betas': (0.9, 0.99)
>>>         }
>>>     },
>>>     'm2': {
>>>         'type': 'OptimWrapper',
>>>         'optimizer': {
>>>             'type': 'Adam',
>>>             'lr': 1e-5,
>>>         }
>>>     }
>>> }
>>> optim_dict_builder = MultiOptimWrapperConstructor(optim_wrapper)
>>> # build optim wrapper dict
>>> optim_wrapper_dict = optim_dict_builder(model)
```

Example 3: Build a single optimizer for multi modules (e.g., DreamBooth):

```
>>> # build StableDiffusion model
>>> model = dict(
>>>     type='StableDiffusion',
>>>     unet=dict(type='unet'),
>>>     vae=dict(type='vae'),
        text_encoder=dict(type='text_encoder'))
>>> diffusion_model = MODELS.build(model)
>>> # build constructor
>>> optim_wrapper = dict(
```

(continues on next page)

(continued from previous page)

```
>>>     modules=['unet', 'text_encoder']
>>>     optimizer=dict(type='Adam', lr=0.0002),
>>>     accumulative_counts=1)
>>> optim_dict_builder = MultiOptimWrapperConstructor(optim_wrapper)
>>> # build optim wrapper dict
>>> optim_wrapper_dict = optim_dict_builder(diffusion_model)
```

Parameters

- **optim_wrapper_cfg_dict** (*dict*) – Config of the optimizer wrapper.
- **paramwise_cfg** (*dict*) – Config of parameter-wise settings. Default: None.

__call__(*module: torch.nn.Module*) → Union[mmengine.optim.OptimWrapperDict, mmengine.optim.OptimWrapper]

Build optimizer and return a optimizer_wrapper_dict.

class mmagic.engine.optimizers.**PGGANOptimWrapperConstructor**(*optim_wrapper_cfg: dict*, *paramwise_cfg: Optional[dict] = None*)

OptimizerConstructor for PGGAN models. Set optimizers for each stage of PGGAN. All submodule must be contained in a `torch.nn.ModuleList` named 'blocks'. And we access each submodule by *MODEL.blocks[SCALE]*, where *MODEL* is generator or discriminator, and the scale is the index of the resolution scale.

More detail about the resolution scale and naming rule please refers to `PGGANGenerator` and `PGGANDiscriminator`.

Example

```
>>> # build PGGAN model
>>> model = dict(
>>>     type='ProgressiveGrowingGAN',
>>>     data_preprocessor=dict(type='GANDataPreprocessor'),
>>>     noise_size=512,
>>>     generator=dict(type='PGGANGenerator', out_scale=1024,
>>>                    noise_size=512),
>>>     discriminator=dict(type='PGGANDiscriminator', in_scale=1024),
>>>     nkimgs_per_scale={
>>>         '4': 600,
>>>         '8': 1200,
>>>         '16': 1200,
>>>         '32': 1200,
>>>         '64': 1200,
>>>         '128': 1200,
>>>         '256': 1200,
>>>         '512': 1200,
>>>         '1024': 12000,
>>>     },
>>>     transition_kimgs=600,
>>>     ema_config=dict(interval=1))
>>> pggan = MODELS.build(model)
```

(continues on next page)

(continued from previous page)

```
>>> # build constructor
>>> optim_wrapper = dict(
>>>     generator=dict(optimizer=dict(type='Adam', lr=0.001,
>>>                                   betas=(0., 0.99))),
>>>     discriminator=dict(
>>>         optimizer=dict(type='Adam', lr=0.001, betas=(0., 0.99))),
>>>     lr_schedule=dict(
>>>         generator={
>>>             '128': 0.0015,
>>>             '256': 0.002,
>>>             '512': 0.003,
>>>             '1024': 0.003
>>>         },
>>>         discriminator={
>>>             '128': 0.0015,
>>>             '256': 0.002,
>>>             '512': 0.003,
>>>             '1024': 0.003
>>>         }))
>>> optim_wrapper_dict_builder = PGGANOptimWrapperConstructor(
>>>     optim_wrapper)
>>> # build optim wrapper dict
>>> optim_wrapper_dict = optim_wrapper_dict_builder(pggan)
```

Parameters

- **optim_wrapper_cfg** (*dict*) – Config of the optimizer wrapper.
- **paramwise_cfg** (*Optional[dict]*) – Parameter-wise options.

__call__(*module: torch.nn.Module*) → mmengine.optim.OptimWrapperDict

Build optimizer and return a optimizerwrapperdict.

class mmagic.engine.optimizers.**SinGANOptimWrapperConstructor**(*optim_wrapper_cfg: dict*, *paramwise_cfg: Optional[dict] = None*)

OptimizerConstructor for SinGAN models. Set optimizers for each submodule of SinGAN. All submodule must be contained in a `torch.nn.ModuleList` named 'blocks'. And we access each submodule by *MODEL.blocks[SCALE]*, where *MODEL* is generator or discriminator, and the scale is the index of the resolution scale.

More detail about the resolution scale and naming rule please refers to `SinGANMultiScaleGenerator` and `SinGANMultiScaleDiscriminator`.

Example

```
>>> # build SinGAN model
>>> model = dict(
>>>     type='SinGAN',
>>>     data_preprocessor=dict(
>>>         type='GANDataPreprocessor',
>>>         non_image_keys=['input_sample']),
>>>     generator=dict(
>>>         type='SinGANMultiScaleGenerator',
>>>         in_channels=3,
>>>         out_channels=3,
>>>         num_scales=2),
>>>     discriminator=dict(
>>>         type='SinGANMultiScaleDiscriminator',
>>>         in_channels=3,
>>>         num_scales=3))
>>> singan = MODELS.build(model)
>>> # build constructor
>>> optim_wrapper = dict(
>>>     generator=dict(optimizer=dict(type='Adam', lr=0.0005,
>>>                                   betas=(0.5, 0.999))),
>>>     discriminator=dict(
>>>         optimizer=dict(type='Adam', lr=0.0005,
>>>                        betas=(0.5, 0.999))))
>>> optim_wrapper_dict_builder = SinGANOptimWrapperConstructor(
>>>     optim_wrapper)
>>> # build optim wrapper dict
>>> optim_wrapper_dict = optim_wrapper_dict_builder(singan)
```

Parameters

- **optim_wrapper_cfg** (*dict*) – Config of the optimizer wrapper.
- **paramwise_cfg** (*Optional[dict]*) – Parameter-wise options.

__call__(*module: torch.nn.Module*) → mmengine.optim.OptimWrapperDict

Build optimizer and return a optimizerwrapperdict.

1.85 mmagic.engine.runner

1.85.1 Package Contents

Classes

LogProcessor	LogProcessor inherits from `mmengine.runner.LogProcessor` and
MultiTestLoop	Test loop for MMagic models which support evaluate multiply dataset at
MultiValLoop	Validation loop for MMagic models which support evaluate multiply

class mmagic.engine.runner.**LogProcessor**(*window_size=10*, *by_epoch=True*, *custom_cfg: Optional[List[dict]] = None*, *num_digits: int = 4*, *log_with_hierarchy: bool = False*, *mean_pattern='.*(loss|time|data_time|grad_norm).*'*)

Bases: mmengine.runner.LogProcessor

LogProcessor inherits from mmengine.runner.LogProcessor and overwrites self.get_log_after_iter().

This log processor should be used along with *mmagic.engine.runner.MultiValLoop* and *mmagic.engine.runner.MultiTestLoop*.

_get_dataloader_size(*runner*, *mode*) → int

Get dataloader size of current loop. In *MultiValLoop* and *MultiTestLoop*, we use *total_length* instead of *len(dataloader)* to denote the total number of iterations.

Parameters

- **runner** (*Runner*) – The runner of the training/validation/testing
- **mode** (*str*) – Current mode of runner.

Returns The dataloader size of current loop.

Return type int

class mmagic.engine.runner.**MultiTestLoop**(*runner*, *dataloader*, *evaluator*, *fp16=False*)

Bases: mmengine.runner.base_loop.BaseLoop

Test loop for MMagic models which support evaluate multiply dataset at the same time. This class support evaluate:

1. Metrics (metric) on a single dataset (e.g. PSNR and SSIM on DIV2K dataset)
2. Different metrics on different datasets (e.g. PSNR on DIV2K and SSIM and PSNR on SET5)

Use cases:

Case 1: metrics on a single dataset

```
>>> # add the following lines in your config
>>> # 1. use `MultiTestLoop` instead of `TestLoop` in MMEngine
>>> val_cfg = dict(type='MultiTestLoop')
>>> # 2. specific MultiEvaluator instead of Evaluator in MMEngine
>>> test_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=[
>>>         dict(type='PSNR', crop_border=2, prefix='Set5'),
>>>         dict(type='SSIM', crop_border=2, prefix='Set5'),
>>>     ])
>>> # 3. define dataloader
>>> test_dataloader = dict(...)
```

Case 2: different metrics on different datasets

```
>>> # add the following lines in your config
>>> # 1. use `MultiTestLoop` instead of `TestLoop` in MMEngine
>>> Test_cfg = dict(type='MultiTestLoop')
>>> # 2. specific a list MultiEvaluator
>>> # do not forget to add prefix for each metric group
```

(continues on next page)

(continued from previous page)

```
>>> div2k_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=dict(type='SSIM', crop_border=2, prefix='DIV2K'))
>>> set5_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=[
>>>         dict(type='PSNR', crop_border=2, prefix='Set5'),
>>>         dict(type='SSIM', crop_border=2, prefix='Set5'),
>>>     ])
>>> # define evaluator config
>>> test_evaluator = [div2k_evaluator, set5_evaluator]
>>> # 3. specific a list dataloader for each metric groups
>>> div2k_dataloader = dict(...)
>>> set5_dataloader = dict(...)
>>> # define dataloader config
>>> test_dataloader = [div2k_dataloader, set5_dataloader]
```

Parameters

- **runner** (*Runner*) – A reference of runner.
- **dataloader** (*Dataloader or dict or list*) – A dataloader object or a dict to build a dataloader a list of dataloader object or a list of config dicts.
- **evaluator** (*Evaluator or dict or list*) – A evaluator object or a dict to build the evaluator or a list of evaluator object or a list of config dicts.

property total_length: int

_build_dataloaders(*dataloader: DATALOADER_TYPE*) → List[torch.utils.data.DataLoader]

Build dataloaders.

Parameters **dataloader** (*Dataloader or dict or list*) – A dataloader object or a dict to build a dataloader a list of dataloader object or a list of config dict.

Returns List of dataloaders for compute metrics.

Return type List[Dataloader]

_build_evaluators(*evaluator: EVALUATOR_TYPE*) → List[mmengine.evaluator.Evaluator]

Build evaluators.

Parameters **evaluator** (*Evaluator or dict or list*) – A evaluator object or a dict to build the evaluator or a list of evaluator object or a list of config dicts.

Returns List of evaluators for compute metrics.

Return type List[Evaluator]

run()

Launch validation. The evaluation process consists of four steps.

1. Prepare pre-calculated items for all metrics by calling `self.evaluator.prepare_metrics()`.
2. Get a list of metrics-sampler pair. Each pair contains a list of metrics with the same sampler mode and a shared sampler.
3. Generate images for the each metrics group. Loop for elements in each sampler and feed to the model as input by calling `self.run_iter()`.

4. Evaluate all metrics by calling `self.evaluator.evaluate()`.

run_iter(*idx*, *data_batch: dict*, *metrics: Sequence[mmengine.evaluator.BaseMetric]*)

Iterate one mini-batch and feed the output to corresponding *metrics*.

Parameters

- **idx** (*int*) – Current idx for the input data.
- **data_batch** (*dict*) – Batch of data from dataloader.
- **metrics** (*Sequence[BaseMetric]*) – Specific metrics to evaluate.

class mmagic.engine.runner.**MultiValLoop**(*runner*, *dataloader: DATALOADER_TYPE*, *evaluator: EVALUATOR_TYPE*, *fp16: bool = False*)

Bases: `mmengine.runner.base_loop.BaseLoop`

Validation loop for MMagic models which support evaluate multiply dataset at the same time. This class support evaluate:

1. Metrics (metric) on a single dataset (e.g. PSNR and SSIM on DIV2K dataset)
2. Different metrics on different datasets (e.g. PSNR on DIV2K and SSIM and PSNR on SET5)

Use cases:

Case 1: metrics on a single dataset

```
>>> # add the following lines in your config
>>> # 1. use `MultiValLoop` instead of `ValLoop` in MMEngine
>>> val_cfg = dict(type='MultiValLoop')
>>> # 2. specific MultiEvaluator instead of Evaluator in MMEngine
>>> val_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=[
>>>         dict(type='PSNR', crop_border=2, prefix='Set5'),
>>>         dict(type='SSIM', crop_border=2, prefix='Set5'),
>>>     ])
>>> # 3. define dataloader
>>> val_dataloader = dict(...)
```

Case 2: different metrics on different datasets

```
>>> # add the following lines in your config
>>> # 1. use `MultiValLoop` instead of `ValLoop` in MMEngine
>>> val_cfg = dict(type='MultiValLoop')
>>> # 2. specific a list MultiEvaluator
>>> # do not forget to add prefix for each metric group
>>> div2k_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=dict(type='SSIM', crop_border=2, prefix='DIV2K'))
>>> set5_evaluator = dict(
>>>     type='MultiEvaluator',
>>>     metrics=[
>>>         dict(type='PSNR', crop_border=2, prefix='Set5'),
>>>         dict(type='SSIM', crop_border=2, prefix='Set5'),
>>>     ])
>>> # define evaluator config
```

(continues on next page)

(continued from previous page)

```
>>> val_evaluator = [div2k_evaluator, set5_evaluator]
>>> # 3. specific a list dataloader for each metric groups
>>> div2k_dataloader = dict(...)
>>> set5_dataloader = dict(...)
>>> # define dataloader config
>>> val_dataloader = [div2k_dataloader, set5_dataloader]
```

Parameters

- **runner** (*Runner*) – A reference of runner.
- **dataloader** (*Dataloader or dict or list*) – A dataloader object or a dict to build a dataloader a list of dataloader object or a list of config dicts.
- **evaluator** (*Evaluator or dict or list*) – A evaluator object or a dict to build the evaluator or a list of evaluator object or a list of config dicts.

property total_length: int

_build_dataloaders(*dataloader: DATALOADER_TYPE*) → List[torch.utils.data.DataLoader]

Build dataloaders.

Parameters **dataloader** (*Dataloader or dict or list*) – A dataloader object or a dict to build a dataloader a list of dataloader object or a list of config dict.

Returns List of dataloaders for compute metrics.

Return type List[Dataloader]

_build_evaluators(*evaluator: EVALUATOR_TYPE*) → List[mmengine.evaluator.Evaluator]

Build evaluators.

Parameters **evaluator** (*Evaluator or dict or list*) – A evaluator object or a dict to build the evaluator or a list of evaluator object or a list of config dicts.

Returns List of evaluators for compute metrics.

Return type List[Evaluator]

run()

Launch validation. The evaluation process consists of four steps.

1. Prepare pre-calculated items for all metrics by calling `self.evaluator.prepare_metrics()`.
2. Get a list of metrics-sampler pair. Each pair contains a list of metrics with the same sampler mode and a shared sampler.
3. Generate images for the each metrics group. Loop for elements in each sampler and feed to the model as input by calling `self.run_iter()`.
4. Evaluate all metrics by calling `self.evaluator.evaluate()`.

run_iter(*idx*, *data_batch: dict*, *metrics: Sequence[mmengine.evaluator.BaseMetric]*)

Iterate one mini-batch and feed the output to corresponding *metrics*.

Parameters

- **idx** (*int*) – Current idx for the input data.
- **data_batch** (*dict*) – Batch of data from dataloader.
- **metrics** (*Sequence[BaseMetric]*) – Specific metrics to evaluate.

1.86 mmagic.engine.schedulers

1.86.1 Package Contents

Classes

LinearLrInterval	Linear learning rate scheduler for image generation.
ReduceLR	Decays the learning rate of each parameter group by linearly changing

class mmagic.engine.schedulers.**LinearLrInterval**(**args*, *interval=1*, ***kwargs*)

Bases: mmengine.optim.LinearLR

Linear learning rate scheduler for image generation.

In the beginning, the learning rate is 'start_factor' defined in mmengine. We give a target learning rate 'end_factor' and a start point 'begin'. If :attr:self.by_epoch is True, 'begin' is calculated by epoch, otherwise, calculated by iteration." Before 'begin', we fix learning rate as 'start_factor'; After 'begin', we linearly update learning rate to 'end_factor'.

Parameters **interval** (*int*) – The interval to update the learning rate. Default: 1.

_get_value()

Compute value using chainable form of the scheduler.

class mmagic.engine.schedulers.**ReduceLR**(*optimizer*, *mode: str = 'min'*, *factor: float = 0.1*, *patience: int = 10*, *threshold: float = 0.0001*, *threshold_mode: str = 'rel'*, *cooldown: int = 0*, *min_lr: float = 0.0*, *eps: float = 1e-08*, ***kwargs*)

Bases: mmengine.optim._ParamScheduler

Decays the learning rate of each parameter group by linearly changing small multiplicative factor until the number of epoch reaches a pre-defined milestone: end.

Notice that such decay can happen simultaneously with other changes to the learning rate from outside this scheduler.

Note:

The learning rate of each parameter group will be update at regular intervals.

Parameters

- **optimizer** (*Optimizer or OptimWrapper*) – Wrapped optimizer.
- **mode** (*str, optional*) – One of *min*, *max*. In *min* mode, lr will be reduced when the quantity monitored has stopped decreasing; in *max* mode it will be reduced when the quantity monitored has stopped increasing. Default: 'min'.
- **factor** (*float, optional*) – Factor by which the learning rate will be reduced. new_lr = lr * factor. Default: 0.1.
- **patience** (*int, optional*) – Number of epochs with no improvement after which learning rate will be reduced. For example, if *patience = 2*, then we will ignore the first 2 epochs with no improvement, and will only decrease the LR after the 3rd epoch if the loss still hasn't improved then. Default: 10.

- **threshold** (*float, optional*) – Threshold for measuring the new optimum, to only focus on significant changes. Default: 1e-4.
- **threshold_mode** (*str, optional*) – One of *rel*, *abs*. In *rel* mode, dynamic_threshold = best * (1 + threshold) in 'max' mode or best * (1 - threshold) in *min* mode. In *abs* mode, dynamic_threshold = best + threshold in *max* mode or best - threshold in *min* mode. Default: 'rel'.
- **cooldown** (*int, optional*) – Number of epochs to wait before resuming normal operation after lr has been reduced. Default: 0.
- **min_lr** (*float, optional*) – Minimum LR value to keep. If LR after decay is lower than *min_lr*, it will be clipped to this value. Default: 0.
- **eps** (*float, optional*) – Minimal decay applied to lr. If the difference between new and old lr is smaller than eps, the update is ignored. Default: 1e-8.
- **begin** (*int*) – Step at which to start updating the learning rate. Defaults to 0.
- **end** (*int*) – Step at which to stop updating the learning rate.
- **last_step** (*int*) – The index of last step. Used for resume without state dict. Defaults to -1.
- **by_epoch** (*bool*) – Whether the scheduled learning rate is updated by epochs. Defaults to True.

property in_cooldown

_get_value()

Compute value using chainable form of the scheduler.

_init_is_better(*mode*)

_reset()

is_better(*a*, *best*)

1.87 mmagic.models.archs

1.87.1 Package Contents

Classes

AllGatherLayer	All gather layer with backward propagation path.
ASPP	ASPP module from DeepLabV3.
AttentionInjection	Wrapper for stable diffusion unet.
SpatialTemporalEnsemble	Apply spatial and temporal ensemble and compute outputs.
SimpleGatedConvModule	Simple Gated Convolutional Module.
ImgNormalize	Normalize images with the given mean and std value.
LinearModule	A linear block that contains linear/norm/activation layers.
LoRAWrapper	Wrapper for LoRA layer.
MultiLayerDiscriminator	Multilayer Discriminator.
PatchDiscriminator	A PatchGAN discriminator.
ResNet	General ResNet.
DepthwiseSeparableConvModule	Depthwise separable convolution module.
SimpleEncoderDecoder	Simple encoder-decoder model from matting.
SoftMaskPatchDiscriminator	A Soft Mask-Guided PatchGAN discriminator.
ResidualBlockNoBN	Residual block without BN.
TokenizerWrapper	Tokenizer wrapper for CLIPTokenizer. Only support CLIPTokenizer
PixelShufflePack	Pixel Shuffle upsample layer.
VGG16	Customized VGG16 Encoder.

Functions

pixel_unshuffle(→ torch.Tensor)	Down-sample by pixel unshuffle.
set_lora(→ torch.nn.Module)	Set LoRA for module.
set_lora_disable(→ torch.nn.Module)	Disable LoRA modules.
set_lora_enable(→ torch.nn.Module)	Enable LoRA modules.
set_only_lora_trainable(→ torch.nn.Module)	Set only LoRA modules trainable.

class mmagic.models.archs.**AllGatherLayer**(**args*, ***kwargs*)

Bases: `torch.autograd.Function`

All gather layer with backward propagation path.

Indeed, this module is to make `dist.all_gather()` in the backward graph. Such kind of operation has been widely used in Moco and other contrastive learning algorithms.

static forward(*ctx*, *x*)

Forward function.

static backward(*ctx*, **grad_outputs*)

Backward function.

class mmagic.models.archs.**ASPP**(*in_channels: int*, *out_channels: int = 256*, *mid_channels: int = 256*, *dilations: Sequence[int] = (12, 24, 36)*, *conv_cfg: Optional[dict] = None*, *norm_cfg: Optional[dict] = dict(type='BN')*, *act_cfg: Optional[dict] = dict(type='ReLU')*, *separable_conv: bool = False*)

Bases: `torch.nn.Module`

ASPP module from DeepLabV3.

The code is adopted from https://github.com/pytorch/vision/blob/master/torchvision/models/ segmentation/deeplabv3.py

For more information about the module: "Rethinking Atrous Convolution for Semantic Image Segmentation".

Parameters

- **in_channels** (*int*) – Input channels of the module.
- **out_channels** (*int*) – Output channels of the module. Default: 256.
- **mid_channels** (*int*) – Output channels of the intermediate ASPP conv modules. Default: 256.
- **dilations** (*Sequence[int]*) – Dilation rate of three ASPP conv module. Default: [12, 24, 36].
- **conv_cfg** (*dict*) – Config dict for convolution layer. If "None", nn.Conv2d will be applied. Default: None.
- **norm_cfg** (*dict*) – Config dict for normalization layer. Default: dict(type='BN').
- **act_cfg** (*dict*) – Config dict for activation layer. Default: dict(type='ReLU').
- **separable_conv** (*bool*) – Whether replace normal conv with depthwise separable conv which is faster. Default: False.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function for ASPP module.

Parameters **x** (*Tensor*) – Input tensor with shape (N, C, H, W).

Returns Output tensor.

Return type Tensor

class mmagic.models.archs.**AttentionInjection**(*module: torch.nn.Module*, *injection_weight=5*)

Bases: torch.nn.Module

Wrapper for stable diffusion unet.

Parameters **module** (*nn.Module*) – The module to be wrapped.

forward(*x: torch.Tensor*, *t*, *encoder_hidden_states=None*, *down_block_additional_residuals=None*, *mid_block_additional_residual=None*, *ref_x=None*) → torch.Tensor

Forward and add LoRA mapping.

Parameters **x** (*Tensor*) – The input tensor.

Returns The output tensor.

Return type Tensor

mmagic.models.archs.**pixel_unshuffle**(*x: torch.Tensor*, *scale: int*) → torch.Tensor

Down-sample by pixel unshuffle.

Parameters

- **x** (*Tensor*) – Input tensor.
- **scale** (*int*) – Scale factor.

Returns Output tensor.

Return type Tensor

class mmagic.models.archs.**SpatialTemporalEnsemble**(*is_temporal_ensemble: Optional[bool] = False*)

Bases: torch.nn.Module

Apply spatial and temporal ensemble and compute outputs.

Parameters **is_temporal_ensemble** (*bool, optional*) – Whether to apply ensemble temporally. If True, the sequence will also be flipped temporally. If the input is an image, this argument must be set to False. Default: False.

_transform(*imgs: torch.Tensor*, *mode: str*) → torch.Tensor

Apply spatial transform (flip, rotate) to the images.

Parameters

- **imgs** (*torch.Tensor*) – The images to be transformed/
- **mode** (*str*) – The mode of transform. Supported values are 'vertical', 'horizontal', and 'transpose', corresponding to vertical flip, horizontal flip, and rotation, respectively.

Returns Output of the model with spatial ensemble applied.

Return type torch.Tensor

spatial_ensemble(*imgs: torch.Tensor*, *model: torch.nn.Module*) → torch.Tensor

Apply spatial ensemble.

Parameters

- **imgs** (*torch.Tensor*) – The images to be processed by the model. Its size should be either (n, t, c, h, w) or (n, c, h, w).
- **model** (*nn.Module*) – The model to process the images.

Returns Output of the model with spatial ensemble applied.

Return type torch.Tensor

forward(*imgs: torch.Tensor*, *model: torch.nn.Module*) → torch.Tensor

Apply spatial and temporal ensemble.

Parameters

- **imgs** (*torch.Tensor*) – The images to be processed by the model. Its size should be either (n, t, c, h, w) or (n, c, h, w).
- **model** (*nn.Module*) – The model to process the images.

Returns Output of the model with spatial ensemble applied.

Return type torch.Tensor

class mmagic.models.archs.**SimpleGatedConvModule**(*in_channels: int*, *out_channels: int*, *kernel_size: Union[int, Tuple[int, int]]*, *feat_act_cfg: Optional[dict] = dict(type='ELU')*, *gate_act_cfg: Optional[dict] = dict(type='Sigmoid')*, ***kwargs*)

Bases: torch.nn.Module

Simple Gated Convolutional Module.

This module is a simple gated convolutional module. The detailed formula is:

$$y = \phi(conv1(x)) * \sigma(conv2(x)),$$

where *phi* is the feature activation function and *sigma* is the gate activation function. In default, the gate activation function is sigmoid.

Parameters

- **in_channels** (*int*) – Same as nn.Conv2d.
- **out_channels** (*int*) – The number of channels of the output feature. Note that *out_channels* in the conv module is doubled since this module contains two convolutions for feature and gate separately.
- **kernel_size** (*int or tuple[int]*) – Same as nn.Conv2d.
- **feat_act_cfg** (*dict*) – Config dict for feature activation layer. Default: dict(type='ELU').
- **gate_act_cfg** (*dict*) – Config dict for gate activation layer. Default: dict(type='Sigmoid').
- **kwargs** (*keyword arguments*) – Same as *ConvModule*.

forward(*x: torch.Tensor*) → torch.Tensor

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.archs.**ImgNormalize**(*pixel_range: float*, *img_mean: Tuple[float, float, float]*, *img_std: Tuple[float, float, float]*, *sign: int = - 1*)

Bases: torch.nn.Conv2d

Normalize images with the given mean and std value.

Based on Conv2d layer, can work in GPU.

Parameters

- **pixel_range** (*float*) – Pixel range of feature.
- **img_mean** (*Tuple[float]*) – Image mean of each channel.
- **img_std** (*Tuple[float]*) – Image std of each channel.
- **sign** (*int*) – Sign of bias. Default -1.

class mmagic.models.archs.**LinearModule**(*in_features: int*, *out_features: int*, *bias: bool = True*, *act_cfg: Optional[dict] = dict(type='ReLU')*, *inplace: bool = True*, *with_spectral_norm: bool = False*, *order: Tuple[str, str] = ('linear', 'act')*)

Bases: torch.nn.Module

A linear block that contains linear/norm/activation layers.

For low level vision, we add spectral norm and padding layer.

Parameters

- **in_features** (*int*) – Same as nn.Linear.
- **out_features** (*int*) – Same as nn.Linear.
- **bias** (*bool*) – Same as nn.Linear. Default: True.
- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.
- **inplace** (*bool*) – Whether to use inplace mode for activation. Default: True.

- **with_spectral_norm** (*bool*) – Whether use spectral norm in linear module. Default: False.
- **order** (*tuple[str]*) – The order of linear/activation layers. It is a sequence of "linear", "norm" and "act". Examples are ("linear", "act") and ("act", "linear").

init_weights() → None

Init weights for the model.

forward(*x: torch.Tensor*, *activate: Optional[bool] = True*) → torch.Tensor

Forward Function.

Parameters

- **x** (*torch.Tensor*) – Input tensor with shape of $(n, *, c)$. Same as `torch.nn.Linear`.
- **activate** (*bool, optional*) – Whether to use activation layer. Defaults to True.

Returns Same as `torch.nn.Linear`.

Return type torch.Tensor

class mmagic.models.archs.**LoRAWrapper**(*module: torch.nn.Module*, *in_feat: int*, *out_feat: int*, *rank: int*, *scale: float = 1*, *names: Optional[Union[str, List[str]]] = None*)

Bases: `torch.nn.Module`

Wrapper for LoRA layer.

Parameters

- **module** (*nn.Module*) – The module to be wrapped.
- **in_feat** (*int*) – Number of input features.
- **out_feat** (*int*) – Number of output features.
- **rank** (*int*) – The rank of LoRA.
- **scale** (*float*) – The scale of LoRA feature.
- **names** (*Union[str, List[str]], optional*) – The name of LoRA layers. If you want to add multi LoRA for one module, names for each LoRA mapping must be defined.

add_lora(*name: str*, *rank: int*, *scale: float = 1*, *state_dict: Optional[dict] = None*)

Add LoRA mapping.

Parameters

- **name** (*str*) – The name of added LoRA.
- **rank** (*int*) – The rank of added LoRA.
- **scale** (*float, optional*) – The scale of added LoRA. Defaults to 1.
- **state_dict** (*dict, optional*) – The state dict of added LoRA. Defaults to None.

_set_value(*attr_name: str*, *value: Any*, *name: Optional[str] = None*)

Set value of attribute.

Parameters

- **attr_name** (*str*) – The name of attribute to be set value.
- **value** (*Any*) – The value to be set.
- **name** (*str, optional*) – The name of field in *attr_name*. If passed, will set value to *attr_name[name]*. Defaults to None.

set_scale(*scale: float*, *name: Optional[str] = None*)

Set LoRA scale.

Parameters

- **scale** (*float*) – The scale to be set.
- **name** (*str, optional*) – The name of LoRA to be set. Defaults to None.

set_enable(*name: Optional[str] = None*)

Enable LoRA for the current layer.

Parameters **name** (*str, optional*) – The name of LoRA to be set. Defaults to None.

set_disable(*name: Optional[str] = None*)

Disable LoRA for the current layer.

Parameters **name** (*str, optional*) – The name of LoRA to be set. Defaults to None.

forward_lora_mapping(*x: torch.Tensor*) → torch.Tensor

Forward LoRA mapping.

Parameters **x** (*Tensor*) – The input tensor.

Returns The output tensor.

Return type Tensor

forward(*x: torch.Tensor*, **args*, ***kwargs*) → torch.Tensor

Forward and add LoRA mapping.

Parameters **x** (*Tensor*) – The input tensor.

Returns The output tensor.

Return type Tensor

classmethod wrap_lora(*module*, *rank=4*, *scale=1*, *names=None*, *state_dict=None*)

Wrap LoRA.

Use case: >>> linear = nn.Linear(2, 4) >>> lora_linear = LoRAWrapper.wrap_lora(linear, 4, 1)

Parameters

- **module** (*nn.Module*) – The module to add LoRA.
- **rank** (*int*) – The rank for LoRA.
- **scale** (*float*) –

Return type *LoRAWrapper*

mmagic.models.archs.**set_lora**(*module: torch.nn.Module*, *config: dict*, *verbose: bool = True*) → torch.nn.Module

Set LoRA for module.

Use case: >>> 1. set all lora with same parameters >>> lora_config = dict(>>> rank=4, >>> scale=1, >>> target_modules=['to_q', 'to_k', 'to_v'])

```
>>> 2. set lora with different parameters
>>> lora_config = dict(
>>>     rank=4,
>>>     scale=1,
>>>     target_modules=[
```

(continues on next page)

(continued from previous page)

```
>>>         # set `to_q` the default parameters
>>>         'to_q',
>>>         # set `to_k` the defined parameters
>>>         dict(target_module='to_k', rank=8, scale=1),
>>>         # set `to_v` the defined `rank` and default `scale`
>>>         dict(target_module='to_v', rank=16)
>>>     ])
```

Parameters

- **module** (*nn.Module*) – The module to set LoRA.
- **config** (*dict*) – The config dict.
- **verbose** (*bool*) – Whether to print log. Defaults to True.

`mmagic.models.archs.`**`set_lora_disable`**(*module: torch.nn.Module*) → torch.nn.Module

Disable LoRA modules.

`mmagic.models.archs.`**`set_lora_enable`**(*module: torch.nn.Module*) → torch.nn.Module

Enable LoRA modules.

`mmagic.models.archs.`**`set_only_lora_trainable`**(*module: torch.nn.Module*) → torch.nn.Module

Set only LoRA modules trainable.

class `mmagic.models.archs.`**`MultiLayerDiscriminator`**(*in_channels: int, max_channels: int, num_convs: int = 5, fc_in_channels: Optional[int] = None, fc_out_channels: int = 1024, kernel_size: int = 5, conv_cfg: Optional[dict] = None, norm_cfg: Optional[dict] = None, act_cfg: Optional[dict] = dict(type='ReLU'), out_act_cfg: Optional[dict] = dict(type='ReLU'), with_input_norm: bool = True, with_out_convs: bool = False, with_spectral_norm: bool = False, **kwargs*)

Bases: `torch.nn.Module`

Multilayer Discriminator.

This is a commonly used structure with stacked multiply convolution layers.

Parameters

- **in_channels** (*int*) – Input channel of the first input convolution.
- **max_channels** (*int*) – The maximum channel number in this structure.
- **num_conv** (*int*) – Number of stacked intermediate convs (including input conv but excluding output conv). Default to 5.
- **fc_in_channels** (*int | None*) – Input dimension of the fully connected layer. If *fc_in_channels* is None, the fully connected layer will be removed. Default to None.
- **fc_out_channels** (*int*) – Output dimension of the fully connected layer. Default to 1024.
- **kernel_size** (*int*) – Kernel size of the conv modules. Default to 5.
- **conv_cfg** (*dict*) – Config dict to build conv layer.
- **norm_cfg** (*dict*) – Config dict to build norm layer.

- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.
- **out_act_cfg** (*dict*) – Config dict for output activation, "relu" by default.
- **with_input_norm** (*bool*) – Whether add normalization after the input conv. Default to True.
- **with_out_convs** (*bool*) – Whether add output convs to the discriminator. The output convs contain two convs. The first out conv has the same setting as the intermediate convs but a stride of 1 instead of 2. The second out conv is a conv similar to the first out conv but reduces the number of channels to 1 and has no activation layer. Default to False.
- **with_spectral_norm** (*bool*) – Whether use spectral norm after the conv layers. Default to False.
- **kwargs** (*keyword arguments*) –

forward(*x: torch.Tensor*) → torch.Tensor

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w') or (n, c).

Return type torch.Tensor

init_weights(*pretrained: Optional[str] = None*) → None

Init weights for models.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

class mmagic.models.archs.**PatchDiscriminator**(*in_channels: int*, *base_channels: int = 64*, *num_conv: int = 3*, *norm_cfg: dict = dict(type='BN')*, *init_cfg: Optional[dict] = dict(type='normal', gain=0.02)*)

Bases: mmengine.model.BaseModule

A PatchGAN discriminator.

Parameters

- **in_channels** (*int*) – Number of channels in input images.
- **base_channels** (*int*) – Number of channels at the first conv layer. Default: 64.
- **num_conv** (*int*) – Number of stacked intermediate convs (excluding input and output conv). Default: 3.
- **norm_cfg** (*dict*) – Config dict to build norm layer. Default: *dict(type='BN')*.
- **init_cfg** (*dict*) – Config dict for initialization. *type*: The name of our initialization method. Default: 'normal'. *gain*: Scaling factor for normal, xavier and orthogonal. Default: 0.02.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

init_weights() → None

Initialize weights for the model.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Default: None.

class mmagic.models.archs.**ResNet**(*depth: int, in_channels: int = 3, stem_channels: int = 64, base_channels: int = 64, num_stages: int = 4, strides: Sequence[int] = (1, 2, 2, 2), dilations: Sequence[int] = (1, 1, 2, 4), deep_stem: bool = False, avg_down: bool = False, frozen_stages: int = - 1, act_cfg: dict = dict(type='ReLU'), conv_cfg: Optional[dict] = None, norm_cfg: dict = dict(type='BN'), with_cp: bool = False, multi_grid: Optional[Sequence[int]] = None, contract_dilation: bool = False, zero_init_residual: bool = True*)

Bases: torch.nn.Module

General ResNet.

This class is adopted from https://github.com/open-mmlab/mmsegmentation/blob/master/mmseg/models/backbones/resnet.py.

Parameters

- **depth** (*int*) – Depth of resnet, from {18, 34, 50, 101, 152}.
- **in_channels** (*int*) – Number of input image channels. Default" 3.
- **stem_channels** (*int*) – Number of stem channels. Default: 64.
- **base_channels** (*int*) – Number of base channels of res layer. Default: 64.
- **num_stages** (*int*) – Resnet stages, normally 4.
- **strides** (*Sequence[int]*) – Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) – Dilation of each stage. Default: (1, 1, 2, 4).
- **deep_stem** (*bool*) – Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg_down** (*bool*) – Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen_stages** (*int*) – Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **act_cfg** (*dict*) – Dictionary to construct and config activation layer. Default: dict(type='ReLU').
- **conv_cfg** (*dict*) – Dictionary to construct and config convolution layer. Default: None.
- **norm_cfg** (*dict*) – Dictionary to construct and config norm layer. Default: dict(type='BN').
- **with_cp** (*bool*) – Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **multi_grid** (*Sequence[int]|None*) – Multi grid dilation rates of last stage. Default: None.
- **contract_dilation** (*bool*) – Whether contract first dilation of each layer Default: False.
- **zero_init_residual** (*bool*) – Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

property norm1: torch.nn.Module

normalization layer after the second convolution layer

Type nn.Module

arch_settings

_make_stem_layer(*in_channels: int*, *stem_channels: int*) → None

Make stem layer for ResNet.

_make_layer(*block: BasicBlock*, *planes: int*, *blocks: int*, *stride: int = 1*, *dilation: int = 1*) → torch.nn.Module

_nostride_dilate(*m: torch.nn.Module*, *dilate: int*) → None

init_weights(*pretrained: Optional[str] = None*) → None

Init weights for the model.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

_freeze_stages() → None

Freeze stages param and norm stats.

forward(*x: torch.Tensor*) → List[torch.Tensor]

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (N, C, H, W).

Returns Output tensor.

Return type Tensor

class mmagic.models.archs.**DepthwiseSeparableConvModule**(*in_channels: int*, *out_channels: int*, *kernel_size: Union[int, Tuple[int, int]]*, *stride: Union[int, Tuple[int, int]] = 1*, *padding: Union[int, Tuple[int, int]] = 0*, *dilation: Union[int, Tuple[int, int]] = 1*, *norm_cfg: Optional[dict] = None*, *act_cfg: Optional[dict] = dict(type='ReLU')*, *dw_norm_cfg: Union[dict, str] = 'default'*, *dw_act_cfg: Union[dict, str] = 'default'*, *pw_norm_cfg: Union[dict, str] = 'default'*, *pw_act_cfg: Union[dict, str] = 'default'*, ***kwargs*)

Bases: torch.nn.Module

Depthwise separable convolution module.

See https://arxiv.org/pdf/1704.04861.pdf for details.

This module can replace a ConvModule with the conv block replaced by two conv block: depthwise conv block and pointwise conv block. The depthwise conv block contains depthwise-conv/norm/activation layers. The pointwise conv block contains pointwise-conv/norm/activation layers. It should be noted that there will be norm/activation layer in the depthwise conv block if norm_cfg and act_cfg are specified.

Parameters

- **in_channels** (*int*) – Same as nn.Conv2d.
- **out_channels** (*int*) – Same as nn.Conv2d.

- **kernel_size** (*int or tuple[int]*) – Same as nn.Conv2d.
- **stride** (*int or tuple[int]*) – Same as nn.Conv2d. Default: 1.
- **padding** (*int or tuple[int]*) – Same as nn.Conv2d. Default: 0.
- **dilation** (*int or tuple[int]*) – Same as nn.Conv2d. Default: 1.
- **norm_cfg** (*dict*) – Default norm config for both depthwise ConvModule and pointwise ConvModule. Default: None.
- **act_cfg** (*dict*) – Default activation config for both depthwise ConvModule and pointwise ConvModule. Default: dict(type='ReLU').
- **dw_norm_cfg** (*dict*) – Norm config of depthwise ConvModule. If it is 'default', it will be the same as `norm_cfg`. Default: 'default'.
- **dw_act_cfg** (*dict*) – Activation config of depthwise ConvModule. If it is 'default', it will be the same as `act_cfg`. Default: 'default'.
- **pw_norm_cfg** (*dict*) – Norm config of pointwise ConvModule. If it is 'default', it will be the same as *norm_cfg*. Default: 'default'.
- **pw_act_cfg** (*dict*) – Activation config of pointwise ConvModule. If it is 'default', it will be the same as `act_cfg`. Default: 'default'.
- **kwargs** (*optional*) – Other shared arguments for depthwise and pointwise ConvModule. See ConvModule for ref.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (N, C, H, W).

Returns Output tensor.

Return type Tensor

class mmagic.models.archs.**SimpleEncoderDecoder**(*encoder: dict, decoder: dict, init_cfg: Optional[dict] = None*)

Bases: `mmengine.model.BaseModule`

Simple encoder-decoder model from matting.

Parameters

- **encoder** (*dict*) – Config of the encoder.
- **decoder** (*dict*) – Config of the decoder.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward(**args, **kwargs*) → torch.Tensor

Forward function.

Returns The output tensor of the decoder.

Return type Tensor

class mmagic.models.archs.**SoftMaskPatchDiscriminator**(*in_channels: int, base_channels: Optional[int] = 64, num_conv: Optional[int] = 3, norm_cfg: Optional[dict] = None, init_cfg: Optional[dict] = dict(type='normal', gain=0.02), with_spectral_norm: Optional[bool] = False*)

Bases: `mmengine.model.BaseModule`

A Soft Mask-Guided PatchGAN discriminator.

Parameters

- **in_channels** (*int*) – Number of channels in input images.
- **base_channels** (*int, optional*) – Number of channels at the first conv layer. Default: 64.
- **num_conv** (*int, optional*) – Number of stacked intermediate convs (excluding input and output conv). Default: 3.
- **norm_cfg** (*dict, optional*) – Config dict to build norm layer. Default: None.
- **init_cfg** (*dict, optional*) – Config dict for initialization. *type*: The name of our initialization method. Default: 'normal'. *gain*: Scaling factor for normal, xavier and orthogonal. Default: 0.02.
- **with_spectral_norm** (*bool, optional*) – Whether use spectral norm after the conv layers. Default: False.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

init_weights() → None

Initialize weights for the model.

class mmagic.models.archs.**ResidualBlockNoBN**(*mid_channels: int = 64, res_scale: float = 1.0*)

Bases: `torch.nn.Module`

Residual block without BN.

It has a style of:

```
---Conv-ReLU-Conv-+-
 |________________|
```

Parameters

- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64.
- **res_scale** (*float*) – Used to scale the residual before addition. Default: 1.0.

init_weights() → None

Initialize weights for ResidualBlockNoBN.

Initialization methods like *kaiming_init* are for VGG-style modules. For modules with residual paths, using smaller std is better for stability and performance. We empirically use 0.1. See more details in "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks"

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.archs.**TokenizerWrapper**(*from_pretrained: Optional[Union[str, os.PathLike]] = None*, *from_config: Optional[Union[str, os.PathLike]] = None*, **args*, ***kwargs*)

Tokenizer wrapper for CLIPTokenizer. Only support CLIPTokenizer currently. This wrapper is modified from https://github.com/huggingface/diffusers/blob/e51f19aee82c8dd874b715a09dbc521d88835d68/src/diffusers/loaders.py#L358 # noqa.

Parameters

- **from_pretrained** (*Union[str, os.PathLike], optional*) – The *model id* of a pretrained model or a path to a *directory* containing model weights and config. Defaults to None.
- **from_config** (*Union[str, os.PathLike], optional*) – The *model id* of a pretrained model or a path to a *directory* containing model weights and config. Defaults to None.
- ***args** – If *from_pretrained* is passed, *args and **kwargs will be passed to *from_pretrained* function. Otherwise, *args and **kwargs will be used to initialize the model by *self._module_cls(*args, **kwargs)*.
- ****kwargs** – If *from_pretrained* is passed, *args and **kwargs will be passed to *from_pretrained* function. Otherwise, *args and **kwargs will be used to initialize the model by *self._module_cls(*args, **kwargs)*.

__getattr__(*name: str*) → Any

try_adding_tokens(*tokens: Union[str, List[str]]*, **args*, ***kwargs*)

Attempt to add tokens to the tokenizer.

Parameters **tokens** (*Union[str, List[str]]*) – The tokens to be added.

get_token_info(*token: str*) → dict

Get the information of a token, including its start and end index in the current tokenizer.

Parameters **token** (*str*) – The token to be queried.

Returns

The information of the token, including its start and end index in current tokenizer.

Return type dict

add_placeholder_token(*placeholder_token: str*, **args*, *num_vec_per_token: int = 1*, ***kwargs*)

Add placeholder tokens to the tokenizer.

Parameters

- **placeholder_token** (*str*) – The placeholder token to be added.
- **num_vec_per_token** (*int, optional*) – The number of vectors of the added placeholder token.
- ***args** – The arguments for *self.wrapped.add_tokens*.
- ****kwargs** – The arguments for *self.wrapped.add_tokens*.

replace_placeholder_tokens_in_text(*text: Union[str, List[str]]*, *vector_shuffle: bool = False*, *prop_tokens_to_load: float = 1.0*) → Union[str, List[str]]

Replace the keywords in text with placeholder tokens. This function will be called in *self.__call__* and *self.encode*.

Parameters

- **text** (*Union[str, List[str]]*) – The text to be processed.
- **vector_shuffle** (*bool, optional*) – Whether to shuffle the vectors. Defaults to False.
- **prop_tokens_to_load** (*float, optional*) – The proportion of tokens to be loaded. If 1.0, all tokens will be loaded. Defaults to 1.0.

Returns The processed text.

Return type Union[str, List[str]]

replace_text_with_placeholder_tokens(*text: Union[str, List[str]]*) → Union[str, List[str]]

Replace the placeholder tokens in text with the original keywords. This function will be called in *self.decode*.

Parameters **text** (*Union[str, List[str]]*) – The text to be processed.

Returns The processed text.

Return type Union[str, List[str]]

__call__(*text: Union[str, List[str]]*, **args*, *vector_shuffle: bool = False*, *prop_tokens_to_load: float = 1.0*, ***kwargs*)

The call function of the wrapper.

Parameters

- **text** (*Union[str, List[str]]*) – The text to be tokenized.
- **vector_shuffle** (*bool, optional*) – Whether to shuffle the vectors. Defaults to False.
- **prop_tokens_to_load** (*float, optional*) – The proportion of tokens to be loaded. If 1.0, all tokens will be loaded. Defaults to 1.0
- ***args** – The arguments for *self.wrapped.__call__*.
- ****kwargs** – The arguments for *self.wrapped.__call__*.

encode(*text: Union[str, List[str]]*, **args*, ***kwargs*)

Encode the passed text to token index.

Parameters

- **text** (*Union[str, List[str]]*) – The text to be encode.
- ***args** – The arguments for *self.wrapped.__call__*.
- ****kwargs** – The arguments for *self.wrapped.__call__*.

decode(*token_ids*, *return_raw: bool = False*, **args*, ***kwargs*) → Union[str, List[str]]

Decode the token index to text.

Parameters

- **token_ids** – The token index to be decoded.
- **return_raw** – Whether keep the placeholder token in the text. Defaults to False.
- ***args** – The arguments for *self.wrapped.decode*.
- ****kwargs** – The arguments for *self.wrapped.decode*.

Returns The decoded text.

Return type Union[str, List[str]]

__repr__()

The representation of the wrapper.

class mmagic.models.archs.**PixelShufflePack**(*in_channels: int*, *out_channels: int*, *scale_factor: int*, *upsample_kernel: int*)

Bases: torch.nn.Module

Pixel Shuffle upsample layer.

Parameters

- **in_channels** (*int*) – Number of input channels.
- **out_channels** (*int*) – Number of output channels.
- **scale_factor** (*int*) – Upsample ratio.
- **upsample_kernel** (*int*) – Kernel size of Conv layer to expand channels.

Returns Upsampled feature map.

init_weights() → None

Initialize weights for PixelShufflePack.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function for PixelShufflePack.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.archs.**VGG16**(*in_channels: int*, *batch_norm: Optional[bool] = False*, *aspp: Optional[bool] = False*, *dilations: Optional[List[int]] = None*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModule

Customized VGG16 Encoder.

A 1x1 conv is added after the original VGG16 conv layers. The indices of max pooling layers are returned for unpooling layers in decoders.

Parameters

- **in_channels** (*int*) – Number of input channels.
- **batch_norm** (*bool, optional*) – Whether use nn.BatchNorm2d. Default to False.
- **aspp** (*bool, optional*) – Whether use ASPP module after the last conv layer. Default to False.
- **dilations** (*list[int], optional*) – Atrous rates of ASPP module. Default to None.
- **init_cfg** (*dict, optional*) – Initialization config dict.

_make_layer(*inplanes: int*, *planes: int*, *convs_layers: int*) → torch.nn.Module

init_weights() → None

Init weights for the model.

forward(*x: torch.Tensor*) → Dict[str, torch.Tensor]

Forward function for ASPP module.

Parameters **x** (*Tensor*) – Input tensor with shape (N, C, H, W).

Returns Dict containing output tensor and maxpooling indices.

Return type dict

1.88 mmagic.models.base_models

1.88.1 Package Contents

Classes

ExponentialMovingAverage	Implements the exponential moving average (EMA) of the model.
RampUpEMA	Implements the exponential moving average with ramping up momentum.
BaseConditionalGAN	Base class for Conditional GAM models.
BaseEditModel	Base model for image and video editing.
BaseGAN	Base class for GAN models.
BaseMattor	Base class for trimap-based matting models.
BaseTranslationModel	Base Translation Model.
BasicInterpolator	Basic model for video interpolation.
OneStageInpaintor	Standard one-stage inpaintor with commonly used losses.
TwoStageInpaintor	Standard two-stage inpaintor with commonly used losses. A two-stage

class mmagic.models.base_models.**ExponentialMovingAverage**(*model: torch.nn.Module*, *momentum: float = 0.0002*, *interval: int = 1*, *device: Optional[torch.device] = None*, *update_buffers: bool = False*)

Bases: mmengine.model.BaseAveragedModel

Implements the exponential moving average (EMA) of the model.

All parameters are updated by the formula as below:

$$Xema_{t+1} = (1 - momentum) * Xema_t + momentum * X_t$$

Parameters

- **model** (*nn.Module*) – The model to be averaged.
- **momentum** (*float*) – The momentum used for updating ema parameter. Defaults to 0.0002. Ema's parameter are updated with the formula $averaged\_param = (1 - momentum) * averaged\_param + momentum * source\_param$.
- **interval** (*int*) – Interval between two updates. Defaults to 1.
- **device** (*torch.device, optional*) – If provided, the averaged model will be stored on the device. Defaults to None.
- **update_buffers** (*bool*) – if True, it will compute running averages for both the parameters and the buffers of the model. Defaults to False.

avg_func(*averaged_param: torch.Tensor*, *source_param: torch.Tensor*, *steps: int*) → None

Compute the moving average of the parameters using exponential moving average.

Parameters

- **averaged_param** (*Tensor*) – The averaged parameters.
- **source_param** (*Tensor*) – The source parameters.
- **steps** (*int*) – The number of times the parameters have been updated.

_load_from_state_dict(*state_dict: dict*, *prefix: str*, *local_metadata: dict*, *strict: bool*, *missing_keys: list*, *unexpected_keys: list*, *error_msgs: List[str]*) → None

Overrides `nn.Module._load_from_state_dict` to support loading `state_dict` without wrap ema module with `BaseAveragedModel`.

In OpenMMLab 1.0, model will not wrap ema submodule with `BaseAveragedModel`, and the ema weight key in *state_dict* will miss *module* prefix. Therefore, `BaseAveragedModel` need to automatically add the `module` prefix if the corresponding key in `state_dict` misses it.

Parameters

- **state_dict** (*dict*) – A dict containing parameters and persistent buffers.
- **prefix** (*str*) – The prefix for parameters and buffers used in this module
- **local_metadata** (*dict*) – a dict containing the metadata for this module.
- **strict** (*bool*) – Whether to strictly enforce that the keys in `state_dict` with `prefix` match the names of parameters and buffers in this module
- **missing_keys** (*List[str]*) – if `strict=True`, add missing keys to this list
- **unexpected_keys** (*List[str]*) – if `strict=True`, add unexpected keys to this list
- **error_msgs** (*List[str]*) – error messages should be added to this list, and will be reported together in `load_state_dict()`.

sync_buffers(*model: torch.nn.Module*) → None

Copy buffer from model to averaged model.

Parameters **model** (*nn.Module*) – The model whose parameters will be averaged.

sync_parameters(*model: torch.nn.Module*) → None

Copy buffer and parameters from model to averaged model.

Parameters **model** (*nn.Module*) – The model whose parameters will be averaged.

class mmagic.models.base_models.**RampUpEMA**(*model: torch.nn.Module*, *interval: int = 1*, *ema_kimg: int = 10*, *ema_rampup: float = 0.05*, *batch_size: int = 32*, *eps: float = 1e-08*, *start_iter: int = 0*, *device: Optional[torch.device] = None*, *update_buffers: bool = False*)

Bases: `mmengine.model.BaseAveragedModel`

Implements the exponential moving average with ramping up momentum.

Ref: https://github.com/NVlabs/stylegan3/blob/master/training/training_loop.py # noqa

Parameters

- **model** (*nn.Module*) – The model to be averaged.
- **interval** (*int*) – Interval between two updates. Defaults to 1.
- **ema_kimg** (*int, optional*) – EMA kimgs. Defaults to 10.

- **ema_rampup** (*float, optional*) – Ramp up rate. Defaults to 0.05.
- **batch_size** (*int, optional*) – Global batch size. Defaults to 32.
- **eps** (*float, optional*) – Ramp up epsilon. Defaults to 1e-8.
- **start_iter** (*int, optional*) – EMA start iter. Defaults to 0.
- **device** (*torch.device, optional*) – If provided, the averaged model will be stored on the `device`. Defaults to None.
- **update_buffers** (*bool*) – if True, it will compute running averages for both the parameters and the buffers of the model. Defaults to False.

static rampup(*steps*, *ema_kimg=10*, *ema_rampup=0.05*, *batch_size=4*, *eps=1e-08*)

Ramp up ema momentum.

Ref: https://github.com/NVlabs/stylegan3/blob/a5a69f58294509598714d1e88c9646c3d7c6ec94/training/training_loop.py#L300-L308 # noqa

Parameters

- **steps** –
- **ema_kimg** (*int, optional*) – Half-life of the exponential moving average of generator weights. Defaults to 10.
- **ema_rampup** (*float, optional*) – EMA ramp-up coefficient.If set to None, then rampup will be disabled. Defaults to 0.05.
- **batch_size** (*int, optional*) – Total batch size for one training iteration. Defaults to 4.
- **eps** (*float, optional*) – Epsiolon to avoid `batch_size` divided by zero. Defaults to 1e-8.

Returns Updated momentum.

Return type dict

avg_func(*averaged_param: torch.Tensor*, *source_param: torch.Tensor*, *steps: int*) → None

Compute the moving average of the parameters using exponential moving average.

Parameters

- **averaged_param** (*Tensor*) – The averaged parameters.
- **source_param** (*Tensor*) – The source parameters.
- **steps** (*int*) – The number of times the parameters have been updated.

_load_from_state_dict(*state_dict: dict*, *prefix: str*, *local_metadata: dict*, *strict: bool*, *missing_keys: list*, *unexpected_keys: list*, *error_msgs: List[str]*) → None

Overrides `nn.Module._load_from_state_dict` to support loading `state_dict` without wrap ema module with `BaseAveragedModel`.

In OpenMMLab 1.0, model will not wrap ema submodule with `BaseAveragedModel`, and the ema weight key in *state_dict* will miss *module* prefix. Therefore, `BaseAveragedModel` need to automatically add the `module` prefix if the corresponding key in `state_dict` misses it.

Parameters

- **state_dict** (*dict*) – A dict containing parameters and persistent buffers.
- **prefix** (*str*) – The prefix for parameters and buffers used in this module

- **local_metadata** (*dict*) – a dict containing the metadata for this module.
- **strict** (*bool*) – Whether to strictly enforce that the keys in `state_dict` with `prefix` match the names of parameters and buffers in this module
- **missing_keys** (*List[str]*) – if `strict=True`, add missing keys to this list
- **unexpected_keys** (*List[str]*) – if `strict=True`, add unexpected keys to this list
- **error_msgs** (*List[str]*) – error messages should be added to this list, and will be reported together in `load_state_dict()`.

sync_buffers(*model: torch.nn.Module*) → None

Copy buffer from model to averaged model.

Parameters **model** (*nn.Module*) – The model whose parameters will be averaged.

sync_parameters(*model: torch.nn.Module*) → None

Copy buffer and parameters from model to averaged model.

Parameters **model** (*nn.Module*) – The model whose parameters will be averaged.

class mmagic.models.base_models.**BaseConditionalGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *num_classes: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: `mmagic.models.base_models.base_gan.BaseGAN`

Base class for Conditional GAM models.

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or `DataPreprocessor`.
- **generator_steps** (*int*) – The number of times the generator is completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – The number of times the discriminator is completely updated before the generator is updated. Defaults to 1.
- **noise_size** (*Optional[int]*) – Size of the input noise vector. Default to None.
- **num_classes** (*Optional[int]*) – The number classes you would like to generate. Defaults to None.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

label_fn(*label: mmagic.utils.typing.LabelVar = None*, *num_batches: int = 1*) → torch.Tensor

Sampling function for label. There are three scenarios in this function:

- If *label* is a callable function, sample *num_batches* of labels with passed *label*.
- If *label* is *None*, sample *num_batches* of labels in range of *[0, self.num_classes-1]* uniformly.

- If *label* is a *torch.Tensor*, check the range of the tensor is in *[0, self.num_classes-1]*. If all values are in valid range, directly return *label*.

Parameters

- **label** (*Union[Tensor, Callable, List[int], None]*) – You can directly give a batch of label through a `torch.Tensor` or offer a callable function to sample a batch of label data. Otherwise, the `None` indicates to use the default label sampler. Defaults to *None*.
- **num_batches** (*int, optional*) – The number of batches label want to sample. If *label* is a Tensor, this will be ignored. Defaults to 1.

Returns Sampled label tensor.

Return type Tensor

data_sample_to_label(*data_sample:* mmagic.structures.DataSample) → Optional[torch.Tensor]

Get labels from input *data_sample* and pack to *torch.Tensor*. If no label is found in the passed *data_sample*, *None* would be returned.

Parameters **data_sample** (*DataSample*) – Input data samples.

Returns Packed label tensor.

Return type Optional[torch.Tensor]

static _get_valid_num_classes(*num_classes: Optional[int]*, *generator: ModelType*, *discriminator: Optional[ModelType]*) → int

Try to get the value of *num_classes* from input, *generator* and *discriminator* and check the consistency of these values. If no conflict is found, return the *num_classes*.

Parameters

- **num_classes** (*Optional[int]*) – *num_classes* passed to *BaseConditionalGAN_refactor*'s initialize function.
- **generator** (*ModelType*) – The config or the model of generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of discriminator.

Returns The number of classes to be generated.

Return type int

forward(*inputs: mmagic.utils.typing.ForwardInputs*, *data_samples: Optional[list] = None*, *mode: Optional[str] = None*) → List[*mmagic.structures.DataSample*]

Sample images with the given inputs. If forward mode is 'ema' or 'orig', the image generated by corresponding generator will be returned. If forward mode is 'ema/orig', images generated by original generator and EMA generator will both be returned in a dict.

Parameters

- **inputs** (*ForwardInputs*) – Dict containing the necessary information (e.g. noise, num_batches, mode) to generate image.
- **data_samples** (*Optional[list]*) – Data samples collated by `data_preprocessor`. Defaults to None.
- **mode** (*Optional[str]*) – *mode* is not used in *BaseConditionalGAN*. Defaults to None.

Returns Generated images or image dict.

Return type List[DataSample]

train_generator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Training function for discriminator. All GANs should implement this function by themselves.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

train_discriminator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Training function for discriminator. All GANs should implement this function by themselves.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.base_models.**BaseEditModel**(*generator: dict*, *pixel_loss: dict*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *init_cfg: Optional[dict] = None*, *data_preprocessor: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Base model for image and video editing.

It must contain a generator that takes frames as inputs and outputs an interpolated frame. It also has a pixel-wise loss for training.

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.

init_cfg

Initialization config dict.

Type dict, optional

data_preprocessor

Used for pre-processing data sampled by dataloader to the format accepted by *forward()*. Default: None.

Type BaseDataPreprocessor

forward(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, *mode: str = 'tensor'*, ***kwargs*) → Union[torch.Tensor, List[*mmagic.structures.DataSample*], dict]

Returns losses or predictions of training, validation, testing, and simple inference process.

`forward` method of BaseModel is an abstract method, its subclasses must implement this method.

Accepts `inputs` and `data_samples` processed by *data_preprocessor*, and returns results according to mode arguments.

During non-distributed training, validation, and testing process, `forward` will be called by `BaseModel.train_step`, `BaseModel.val_step` and `BaseModel.val_step` directly.

During distributed data parallel training process, `MMSeparateDistributedDataParallel.train_step` will first call `DistributedDataParallel.forward` to enable automatic gradient synchronization, and then call `forward` to get training loss.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.
- **mode** (*str*) – mode should be one of `loss`, `predict` and `tensor`. Default: 'tensor'.
  - `loss`: Called by `train_step` and return loss `dict` used for logging
  - `predict`: Called by `val_step` and `test_step` and return list of `BaseDataElement` results used for computing metric.
  - `tensor`: Called by custom use to get `Tensor` type results.

Returns

- If `mode == loss`, return a `dict` of loss tensor used for backward and logging.
- If `mode == predict`, return a `list` of `BaseDataElement` for computing metric and getting inference result.
- If `mode == tensor`, return a tensor or `tuple` of tensor or `dict` or tensor for custom use.

Return type ForwardResults

convert_to_datasample(*predictions:* mmagic.structures.DataSample, *data_samples:* mmagic.structures.DataSample, *inputs: Optional[torch.Tensor]*) → List[*mmagic.structures.DataSample*]

Add predictions and destructed inputs (if passed) to data samples.

Parameters

- **predictions** (*DataSample*) – The predictions of the model.
- **data_samples** (*DataSample*) – The data samples loaded from dataloader.
- **inputs** (*Optional[torch.Tensor]*) – The input of model. Defaults to None.

Returns Modified data samples.

Return type List[DataSample]

forward_tensor(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, ***kwargs*) → torch.Tensor

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns result of simple forward.

Return type Tensor

forward_inference(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, ***kwargs*) → *mmagic.structures.DataSample*

Forward inference. Returns predictions of validation, testing, and simple inference.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns predictions.

Return type DataSample

forward_train(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, ***kwargs*) → Dict[str, torch.Tensor]

Forward training. Returns dict of losses of training.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns Dict of losses.

Return type dict

class mmagic.models.base_models.**BaseGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: mmengine.model.BaseModel

Base class for GAN models.

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or DataPreprocessor.

- **generator_steps** (*int*) – The number of times the generator is completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – The number of times the discriminator is completely updated before the generator is updated. Defaults to 1.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

property generator_steps: int

The number of times the generator is completely updated before the discriminator is updated.

Type int

property discriminator_steps: int

The number of times the discriminator is completely updated before the generator is updated.

Type int

property device: torch.device

Get current device of the model.

Returns The current device of the model.

Return type torch.device

property with_ema_gen: bool

Whether the GAN adopts exponential moving average.

Returns

If *True*, means this GAN model is adopted to exponential moving average and vice versa.

Return type bool

static gather_log_vars(*log_vars_list: List[Dict[str, torch.Tensor]]*) → Dict[str, torch.Tensor]

Gather a list of log_vars. :param log_vars_list: List[Dict[str, Tensor]]

Returns Dict[str, Tensor]

_init_loss(*loss_config: Optional[Dict] = None*) → None

Initialize customized loss modules.

If loss_config is a dict, we allow kinds of value for each field.

1. ***loss_config* is None: Users will implement all loss calculations** in their own function. Weights for each loss terms are hard coded.
2. ***loss_config* is dict of scalar or string: Users will implement all** loss calculations and use passed *loss_config* to control the weight or behavior of the loss calculation. Users will unpack and use each field in this dict by themselves.

   loss_config = dict(gp_norm_mode='HWC', gp_loss_weight=10)
3. ***loss_config* is dict of dict: Each field in *loss_config* will** used to build a corresponding loss module. And use loss calculation function predefined by `BaseGAN` to calculate the loss.

   loss_config = dict()

Example

loss_config = dict(# *BaseGAN* pre-defined fields gan_loss=dict(type='GANLoss', gan_type='wgan-logistic-ns'), disc_auxiliary_loss=dict(

type='R1GradientPenalty', loss_weight=10. / 2., interval=2, norm_mode='HWC', data_info=dict(

real_data='real_imgs', discriminator='disc')),

gen_auxiliary_loss=dict(type='GeneratorPathRegularizer', loss_weight=2, pl_batch_shrink=2, interval=g_reg_interval, data_info=dict(

generator='gen', num_batches='batch_size')),

user-defined field for loss weights or loss calculation my_loss_2=dict(weight=2, norm_mode='L1'), my_loss_3=2, my_loss_4_norm_type='L2')

Parameters **loss_config** (*Optional[Dict], optional*) – Loss config used to build loss modules or define the loss weights. Defaults to None.

noise_fn(*noise: mmagic.utils.typing.NoiseVar = None*, *num_batches: int = 1*)

Sampling function for noise. There are three scenarios in this function:

- If *noise* is a callable function, sample *num_batches* of noise with passed *noise*.
- If *noise* is *None*, sample *num_batches* of noise from gaussian distribution.
- If *noise* is a *torch.Tensor*, directly return *noise*.

Parameters

- **noise** (*Union[Tensor, Callable, List[int], None]*) – You can directly give a batch of label through a `torch.Tensor` or offer a callable function to sample a batch of label data. Otherwise, the `None` indicates to use the default noise sampler. Defaults to *None*.
- **num_batches** (*int, optional*) – The number of batches label want to sample. If *label* is a Tensor, this will be ignored. Defaults to 1.

Returns Sampled noise tensor.

Return type Tensor

_init_ema_model(*ema_config: dict*)

Initialize a EMA model corresponding to the given *ema_config*. If *ema_config* is an empty dict or *None*, EMA model will not be initialized.

Parameters **ema_config** (*dict*) – Config to initialize the EMA model.

_get_valid_model(*batch_inputs: mmagic.utils.typing.ForwardInputs*) → str

Try to get the valid forward model from inputs.

- If forward model is defined in *batch_inputs*, it will be used as forward model.
- If forward model is not defined in *batch_inputs*, 'ema' will returned if **:property:`with_ema_gen`** is true. Otherwise, 'orig' will be returned.

Parameters **batch_inputs** (*ForwardInputs*) – Inputs passed to *forward()*.

Returns

Forward model to generate image. ('orig', 'ema' or 'ema/orig').

Return type str

forward(*inputs: mmagic.utils.typing.ForwardInputs*, *data_samples: Optional[list] = None*, *mode: Optional[str] = None*) → mmagic.utils.typing.SampleList

Sample images with the given inputs. If forward mode is 'ema' or 'orig', the image generated by corresponding generator will be returned. If forward mode is 'ema/orig', images generated by original generator and EMA generator will both be returned in a dict.

Parameters

- **batch_inputs** (*ForwardInputs*) – Dict containing the necessary information (e.g. noise, num_batches, mode) to generate image.
- **data_samples** (*Optional[list]*) – Data samples collated by `data_preprocessor`. Defaults to None.
- **mode** (*Optional[str]*) – *mode* is not used in *BaseGAN*. Defaults to None.

Returns A list of `DataSample` contain generated results.

Return type SampleList

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data.

Calls `self.data_preprocessor(data)` and `self(inputs, data_sample, mode=None)` in order. Return the generated results which will be passed to evaluator.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as *val_step()*.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type List[DataSample]

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train GAN model. In the training of GAN models, generator and discriminator are updated alternatively. In MMagic's design, *self.train_step* is called with data input. Therefore we always update discriminator, whose updating is relay on real data, and then determine if the generator needs to be updated based on the current number of iterations. More details about whether to update generator can be found in `should_gen_update()`.

Parameters

- **data** (*dict*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapperDict*) – OptimWrapperDict instance contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

_get_gen_loss(*out_dict*)

_get_disc_loss(*out_dict*)

train_generator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Training function for discriminator. All GANs should implement this function by themselves.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

train_discriminator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Training function for discriminator. All GANs should implement this function by themselves.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.base_models.**BaseMattor**(*data_preprocessor: Union[dict, mmengine.config.Config]*, *backbone: dict*, *init_cfg: Optional[dict] = None*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Base class for trimap-based matting models.

A matting model must contain a backbone which produces *pred_alpha*, a dense prediction with the same height and width of input image. In some cases (such as DIM), the model has a refiner which refines the prediction of the backbone.

Subclasses should overwrite the following functions:

- _forward_train(), to return a loss
- _forward_test(), to return a prediction
- _forward(), to return raw tensors

For test, this base class provides functions to resize inputs and post-process pred_alphas to get predictions

Parameters

- **backbone** (*dict*) – Config of backbone.
- **data_preprocessor** (*dict*) – Config of data_preprocessor. See MattorPreprocessor for details.

- **init_cfg** (*dict, optional*) – Initialization config dict.
- **train_cfg** (*dict*) – Config of training. Customized by subclassesCustomized bu In train_cfg, train_backbone should be specified. If the model has a refiner, train_refiner should be specified.
- **test_cfg** (*dict*) – Config of testing. In test_cfg, If the model has a refiner, train_refiner should be specified.

resize_inputs(*batch_inputs: torch.Tensor*) → torch.Tensor

Pad or interpolate images and trimaps to multiple of given factor.

restore_size(*pred_alpha: torch.Tensor*, *data_sample:* mmagic.structures.DataSample) → torch.Tensor

Restore the predicted alpha to the original shape.

The shape of the predicted alpha may not be the same as the shape of original input image. This function restores the shape of the predicted alpha.

Parameters

- **pred_alpha** (*torch.Tensor*) – A single predicted alpha of shape (1, H, W).
- **data_sample** (*DataSample*) – Data sample containing original shape as meta data.

Returns The reshaped predicted alpha.

Return type torch.Tensor

postprocess(*batch_pred_alpha: torch.Tensor*, *data_samples:* mmagic.structures.DataSample) → List[*mmagic.structures.DataSample*]

Post-process alpha predictions.

This function contains the following steps:

1. Restore padding or interpolation
2. Mask alpha prediction with trimap
3. Clamp alpha prediction to 0-1
4. Convert alpha prediction to uint8
5. Pack alpha prediction into DataSample

Currently only batch_size 1 is actually supported.

Parameters

- **batch_pred_alpha** (*torch.Tensor*) – A batch of predicted alpha of shape (N, 1, H, W).
- **data_samples** (*List[DataSample]*) – List of data samples.

Returns

A list of predictions. Each data sample contains a pred_alpha, which is a torch.Tensor with dtype=uint8, device=cuda:0

Return type List[DataSample]

forward(*inputs: torch.Tensor*, *data_samples: DataSamples = None*, *mode: str = 'tensor'*) → List[*mmagic.structures.DataSample*]

General forward function.

Parameters

- **inputs** (*torch.Tensor*) – A batch of inputs. with image and trimap concatenated alone channel dimension.
- **data_samples** (*List[DataSample], optional*) – A list of data samples, containing: - Ground-truth alpha / foreground / background to compute loss - other meta information
- **mode** (*str*) – mode should be one of `loss`, `predict` and `tensor`. Default: 'tensor'.
  - `loss`: Called by `train_step` and return loss `dict` used for logging
  - `predict`: Called by `val_step` and `test_step` and return list of `BaseDataElement` results used for computing metric.
  - `tensor`: Called by custom use to get `Tensor` type results.

Returns Sequence of predictions packed into DataElement

Return type List[DataElement]

convert_to_datasample(*predictions: List[*mmagic.structures.DataSample*], data_samples:* mmagic.structures.DataSample) → List[*mmagic.structures.DataSample*]

Add predictions to data samples.

Parameters

- **predictions** (*List[DataSample]*) – The predictions of the model.
- **data_samples** (*DataSample*) – The data samples loaded from dataloader.

Returns Modified data samples.

Return type List[DataSample]

class mmagic.models.base_models.**BaseTranslationModel**(*generator, discriminator, default_domain: str, reachable_domains: List[str], related_domains: List[str], data_preprocessor, discriminator_steps: int = 1, disc_init_steps: int = 0, real_img_key: str = 'real_img', loss_config: Optional[dict] = None*)

Bases: `mmengine.model.BaseModel`

Base Translation Model.

Translation models can transfer images from one domain to another. Domain information like *default_domain*, *reachable_domains* are needed to initialize the class. And we also provide query functions like *is_domain_reachable*, *get_other_domains*.

You can get a specific generator based on the domain, and by specifying *target_domain* in the forward function, you can decide the domain of generated images. Considering the difference among different image translation models, we only provide the external interfaces mentioned above. When you implement image translation with a specific method, you can inherit both *BaseTranslationModel* and the method (e.g BaseGAN) and implement abstract methods.

Parameters

- **default_domain** (*str*) – Default output domain.
- **reachable_domains** (*list[str]*) – Domains that can be generated by the model.
- **related_domains** (*list[str]*) – Domains involved in training and testing. *reachable_domains* must be contained in *related_domains*. However, related_domains may contain source domains that are used to retrieve source images from data_batch but not in reachable_domains.

- **discriminator_steps** (*int*) – The number of times the discriminator is completely updated before the generator is updated. Defaults to 1.
- **disc_init_steps** (*int*) – The number of initial steps used only to train discriminators.

init_weights()

Initialize weights for the module dict.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Default: None.

get_module(*module*)

Get *nn.ModuleDict* to fit the *MMDistributedDataParallel* interface.

Parameters **module** (*MMDistributedDataParallel | nn.ModuleDict*) – The input module that needs processing.

Returns The ModuleDict of multiple networks.

Return type nn.ModuleDict

forward(*img*, *test_mode=False*, ***kwargs*)

Forward function.

Parameters

- **img** (*tensor*) – Input image tensor.
- **test_mode** (*bool*) – Whether in test mode or not. Default: False.
- **kwargs** (*dict*) – Other arguments.

forward_train(*img*, *target_domain*, ***kwargs*)

Forward function for training.

Parameters

- **img** (*tensor*) – Input image tensor.
- **target_domain** (*str*) – Target domain of output image.
- **kwargs** (*dict*) – Other arguments.

Returns Forward results.

Return type dict

forward_test(*img*, *target_domain*, ***kwargs*)

Forward function for testing.

Parameters

- **img** (*tensor*) – Input image tensor.
- **target_domain** (*str*) – Target domain of output image.
- **kwargs** (*dict*) – Other arguments.

Returns Forward results.

Return type dict

is_domain_reachable(*domain*)

Whether image of this domain can be generated.

get_other_domains(*domain*)

get other domains.

_get_target_generator(*domain*)

get target generator.

_get_target_discriminator(*domain*)

get target discriminator.

translation(*image*, *target_domain=None*, ***kwargs*)

Translation Image to target style.

Parameters

- **image** (*tensor*) – Image tensor with a shape of (N, C, H, W).
- **target_domain** (*str, optional*) – Target domain of output image. Default to None.

Returns Image tensor of target style.

Return type dict

class mmagic.models.base_models.**BasicInterpolator**(*generator: dict*, *pixel_loss: dict*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *required_frames: int = 2*, *step_frames: int = 1*, *init_cfg: Optional[dict] = None*, *data_preprocessor: Optional[dict] = None*)

Bases: mmagic.models.base_models.base_edit_model.BaseEditModel

Basic model for video interpolation.

It must contain a generator that takes frames as inputs and outputs an interpolated frame. It also has a pixel-wise loss for training.

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **required_frames** (*int*) – Required frames in each process. Default: 2
- **step_frames** (*int*) – Step size of video frame interpolation. Default: 1
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.

init_cfg

Initialization config dict.

Type dict, optional

data_preprocessor

Used for pre-processing data sampled by dataloader to the format accepted by forward().

Type BaseDataPreprocessor

split_frames(*input_tensors: torch.Tensor*) → torch.Tensor

split input tensors for inference.

Parameters **input_tensors** (*Tensor*) – Tensor of input frames with shape [1, t, c, h, w]

Returns Split tensor with shape [t-1, 2, c, h, w]

Return type Tensor

static merge_frames(*input_tensors: torch.Tensor*, *output_tensors: torch.Tensor*) → list

merge input frames and output frames.

Interpolate a frame between the given two frames.

Merged from [[in1, in2], [in2, in3], [in3, in4], . . .] [[out1], [out2], [out3], . . .]

to [in1, out1, in2, out2, in3, out3, in4, . . .]

Parameters

- **input_tensors** (*Tensor*) – The input frames with shape [n, 2, c, h, w]
- **output_tensors** (*Tensor*) – The output frames with shape [n, 1, c, h, w].

Returns The final frames.

Return type list[np.array]

class mmagic.models.base_models.**OneStageInpaintor**(*data_preprocessor: Union[dict, mmengine.config.Config]*, *encdec: dict*, *disc: Optional[dict] = None*, *loss_gan: Optional[dict] = None*, *loss_gp: Optional[dict] = None*, *loss_disc_shift: Optional[dict] = None*, *loss_composed_percep: Optional[dict] = None*, *loss_out_percep: bool = False*, *loss_l1_hole: Optional[dict] = None*, *loss_l1_valid: Optional[dict] = None*, *loss_tv: Optional[dict] = None*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Standard one-stage inpaintor with commonly used losses.

An inpaintor must contain an encoder-decoder style generator to inpaint masked regions. A discriminator will be adopted when adversarial training is needed.

In this class, we provide a common interface for inpaintors. For other inpaintors, only some funcs may be modified to fit the input style or training schedule.

Parameters

- **data_preprocessor** (*dict*) – Config of data_preprocessor.
- **encdec** (*dict*) – Config for encoder-decoder style generator.
- **disc** (*dict*) – Config for discriminator.
- **loss_gan** (*dict*) – Config for adversarial loss.
- **loss_gp** (*dict*) – Config for gradient penalty loss.
- **loss_disc_shift** (*dict*) – Config for discriminator shift loss.

- **loss_composed_percep** (*dict*) – Config for perceptual and style loss with composed image as input.
- **loss_out_percep** (*dict*) – Config for perceptual and style loss with direct output as input.
- **loss_l1_hole** (*dict*) – Config for l1 loss in the hole.
- **loss_l1_valid** (*dict*) – Config for l1 loss in the valid region.
- **loss_tv** (*dict*) – Config for total variation loss.
- **train_cfg** (*dict*) – Configs for training scheduler. *disc_step* must be contained for indicates the discriminator updating steps in each training step.
- **test_cfg** (*dict*) – Configs for testing scheduler.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward(*inputs: torch.Tensor*, *data_samples: Optional[mmagic.utils.SampleList]*, *mode: str = 'tensor'*) → FORWARD_RETURN_TYPE

Forward function.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.
- **mode** (*str*) – mode should be one of `loss`, `predict` and `tensor`. Default: 'tensor'.
  - `loss`: Called by `train_step` and return loss `dict` used for logging
  - `predict`: Called by `val_step` and `test_step` and return list of `BaseDataElement` results used for computing metric.
  - `tensor`: Called by custom use to get `Tensor` type results.

Returns

- If `mode == loss`, return a `dict` of loss tensor used for backward and logging.
- If `mode == predict`, return a `list` of `BaseDataElement` for computing metric and getting inference result.
- If `mode == tensor`, return a tensor or `tuple` of tensor or `dict` or tensor for custom use.

Return type ForwardResults

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → dict

Train step function.

In this function, the inpaintor will finish the train step following the pipeline:

1. get fake res/image
2. optimize discriminator (if have)
3. optimize generator

If *self.train_cfg.disc_step > 1*, the train step will contain multiple iterations for optimizing discriminator with different input data and only one iteration for optimizing generator after *disc_step* iterations for discriminator.

Parameters

- **data** (*List[dict]*) – Batch of data as input.

- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns

Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

abstract forward_train(**args*, ***kwargs*) → None

Forward function for training.

In this version, we do not use this interface.

forward_train_d(*data_batch: torch.Tensor*, *is_real: bool*, *is_disc: bool*) → dict

Forward function in discriminator training step.

In this function, we compute the prediction for each data batch (real or fake). Meanwhile, the standard gan loss will be computed with several proposed losses for stable training.

Parameters

- **data_batch** (*torch.Tensor*) – Batch of real data or fake data.
- **is_real** (*bool*) – If True, the gan loss will regard this batch as real data. Otherwise, the gan loss will regard this batch as fake data.
- **is_disc** (*bool*) – If True, this function is called in discriminator training step. Otherwise, this function is called in generator training step. This will help us to compute different types of adversarial loss, like LSGAN.

Returns Contains the loss items computed in this function.

Return type dict

generator_loss(*fake_res: torch.Tensor*, *fake_img: torch.Tensor*, *gt: torch.Tensor*, *mask: torch.Tensor*, *masked_img: torch.Tensor*) → Tuple[dict, dict]

Forward function in generator training step.

In this function, we mainly compute the loss items for generator with the given (fake_res, fake_img). In general, the *fake_res* is the direct output of the generator and the *fake_img* is the composition of direct output and ground-truth image.

Parameters

- **fake_res** (*torch.Tensor*) – Direct output of the generator.
- **fake_img** (*torch.Tensor*) – Composition of *fake_res* and ground-truth image.
- **gt** (*torch.Tensor*) – Ground-truth image.
- **mask** (*torch.Tensor*) – Mask image.
- **masked_img** (*torch.Tensor*) – Composition of mask image and ground-truth image.

Returns Dict contains the results computed within this function for visualization and dict contains the loss items computed in this function.

Return type tuple(dict)

forward_tensor(*inputs: torch.Tensor*, *data_samples: mmagic.utils.SampleList*) → Tuple[torch.Tensor, torch.Tensor]

Forward function in tensor mode.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_samples** (*List[dict]*) – List of data sample dict.

Returns

Direct output of the generator and composition of *fake_res* and ground-truth image.

Return type tuple

forward_test(*inputs: torch.Tensor*, *data_samples: mmagic.utils.SampleList*) → *mmagic.structures.DataSample*

Forward function for testing.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_samples** (*List[dict]*) – List of data sample dict.

Returns

List of prediction saved in DataSample.

Return type predictions (List[DataSample])

convert_to_datasample(*predictions:* mmagic.structures.DataSample, *data_samples:* mmagic.structures.DataSample, *inputs: Optional[torch.Tensor]*) → List[*mmagic.structures.DataSample*]

Add predictions and destructed inputs (if passed) to data samples.

Parameters

- **predictions** (*DataSample*) – The predictions of the model.
- **data_samples** (*DataSample*) – The data samples loaded from dataloader.
- **inputs** (*Optional[torch.Tensor]*) – The input of model. Defaults to None.

Returns Modified data samples.

Return type List[DataSample]

forward_dummy(*x: torch.Tensor*) → torch.Tensor

Forward dummy function for getting flops.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Results tensor with shape of (n, 3, h, w).

Return type torch.Tensor

class mmagic.models.base_models.**TwoStageInpaintor**(*data_preprocessor: Union[dict, mmengine.config.Config]*, *encdec: dict*, *disc: Optional[dict] = None*, *loss_gan: Optional[dict] = None*, *loss_gp: Optional[dict] = None*, *loss_disc_shift: Optional[dict] = None*, *loss_composed_percep: Optional[dict] = None*, *loss_out_percep: bool = False*, *loss_l1_hole: Optional[dict] = None*, *loss_l1_valid: Optional[dict] = None*, *loss_tv: Optional[dict] = None*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *init_cfg: Optional[dict] = None*, *stage1_loss_type: Optional[Sequence[str]] = ('loss_l1_hole',)*, *stage2_loss_type: Optional[Sequence[str]] = ('loss_l1_hole', 'loss_gan')*, *input_with_ones: bool = True*, *disc_input_with_mask: bool = False*)

Bases: mmagic.models.base_models.one_stage.OneStageInpaintor

Standard two-stage inpaintor with commonly used losses. A two-stage inpaintor contains two encoder-decoder style generators to inpaint masked regions. Currently, we support these loss types in each of two stage inpaintors:

['loss_gan', 'loss_l1_hole', 'loss_l1_valid', 'loss_composed_percep', 'loss_out_percep', 'loss_tv'] The *stage1_loss_type* and *stage2_loss_type* should be chosen from these loss types.

Parameters

- **data_preprocessor** (*dict*) – Config of data_preprocessor.
- **encdec** (*dict*) – Config for encoder-decoder style generator.
- **disc** (*dict*) – Config for discriminator.
- **loss_gan** (*dict*) – Config for adversarial loss.
- **loss_gp** (*dict*) – Config for gradient penalty loss.
- **loss_disc_shift** (*dict*) – Config for discriminator shift loss.
- **loss_composed_percep** (*dict*) – Config for perceptual and style loss with composed image as input.
- **loss_out_percep** (*dict*) – Config for perceptual and style loss with direct output as input.
- **loss_l1_hole** (*dict*) – Config for l1 loss in the hole.
- **loss_l1_valid** (*dict*) – Config for l1 loss in the valid region.
- **loss_tv** (*dict*) – Config for total variation loss.
- **train_cfg** (*dict*) – Configs for training scheduler. *disc_step* must be contained for indicates the discriminator updating steps in each training step.
- **test_cfg** (*dict*) – Configs for testing scheduler.
- **init_cfg** (*dict, optional*) – Initialization config dict.
- **stage1_loss_type** (*tuple[str]*) – Contains the loss names used in the first stage model. Default: ('loss_l1_hole').
- **stage2_loss_type** (*tuple[str]*) – Contains the loss names used in the second stage model. Default: ('loss_l1_hole', 'loss_gan').
- **input_with_ones** (*bool*) – Whether to concatenate an extra ones tensor in input. Default: True.

- **disc_input_with_mask** (*bool*) – Whether to add mask as input in discriminator. Default: False.

forward_tensor(*inputs: torch.Tensor*, *data_samples: mmagic.utils.SampleList*) → Tuple[torch.Tensor, torch.Tensor]

Forward function in tensor mode.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_samples** (*List[dict]*) – List of data sample dict.

Returns Dict contains output results.

Return type dict

two_stage_loss(*stage1_data: dict*, *stage2_data: dict*, *gt: torch.Tensor*, *mask: torch.Tensor*, *masked_img: torch.Tensor*) → Tuple[dict, dict]

Calculate two-stage loss.

Parameters

- **stage1_data** (*dict*) – Contain stage1 results.
- **stage2_data** (*dict*) – Contain stage2 results..
- **gt** (*torch.Tensor*) – Ground-truth image.
- **mask** (*torch.Tensor*) – Mask image.
- **masked_img** (*torch.Tensor*) – Composition of mask image and ground-truth image.

Returns Dict contains the results computed within this function for visualization and dict contains the loss items computed in this function.

Return type tuple(dict)

calculate_loss_with_type(*loss_type: str*, *fake_res: torch.Tensor*, *fake_img: torch.Tensor*, *gt: torch.Tensor*, *mask: torch.Tensor*, *prefix: Optional[str] = 'stage1_'*) → dict

Calculate multiple types of losses.

Parameters

- **loss_type** (*str*) – Type of the loss.
- **fake_res** (*torch.Tensor*) – Direct results from model.
- **fake_img** (*torch.Tensor*) – Composited results from model.
- **gt** (*torch.Tensor*) – Ground-truth tensor.
- **mask** (*torch.Tensor*) – Mask tensor.
- **prefix** (*str, optional*) – Prefix for loss name. Defaults to 'stage1_'. # noqa

Returns Contain loss value with its name.

Return type dict

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → dict

Train step function.

In this function, the inpaintor will finish the train step following the pipeline: 1. get fake res/image 2. optimize discriminator (if have) 3. optimize generator

If *self.train_cfg.disc_step > 1*, the train step will contain multiple iterations for optimizing discriminator with different input data and only one iteration for optimizing generator after *disc_step* iterations for discriminator.

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

1.89 mmagic.models.losses

1.89.1 Package Contents

Classes

AdvLoss	Returns the loss of discriminator with the specified type for
CLIPLoss	Clip loss. In styleclip, this loss is used to optimize the latent code
CharbonnierCompLoss	Charbonnier composition loss.
L1CompositionLoss	L1 composition loss.
MSECompositionLoss	MSE (L2) composition loss.
FaceIdLoss	Face similarity loss. Generally this loss is used to keep the id
LightCNNFeatureLoss	Feature loss of DICGAN, based on LightCNN.
DiscShiftLoss	Disc shift loss.
GANLoss	Define GAN loss.
GaussianBlur	A Gaussian filter which blurs a given tensor with a two-dimensional
GradientPenaltyLoss	Gradient penalty loss for wgan-gp.
GradientLoss	Gradient loss.
CLIPLossComps	Clip loss. In styleclip, this loss is used to optimize the latent code
DiscShiftLossComps	Disc Shift Loss.
FaceIdLossComps	Face similarity loss. Generally this loss is used to keep the id
GANLossComps	Define GAN loss.
GeneratorPathRegularizerComps	Generator Path Regularizer.
GradientPenaltyLossComps	Gradient Penalty for WGAN-GP.
R1GradientPenaltyComps	R1 gradient penalty for WGAN-GP.
PerceptualLoss	Perceptual loss with commonly used style loss.
PerceptualVGG	VGG network used in calculating perceptual loss.
TransferalPerceptualLoss	Transferal perceptual loss.
CharbonnierLoss	Charbonnier loss (one variant of Robust L1Loss, a differentiable variant
L1Loss	L1 (mean absolute error, MAE) loss.
MaskedTVLoss	Masked TV loss.
MSELoss	MSE (L2) loss.
PSNRLoss	PSNR Loss in "HINet: Half Instance Normalization Network for Image

Functions

disc_shift_loss(→ torch.Tensor)	Disc Shift loss.
gen_path_regularizer(→ Tuple[torch.Tensor])	Generator Path Regularization.
gradient_penalty_loss(→ torch.Tensor)	Calculate gradient penalty for wgan-gp.
r1_gradient_penalty_loss(→ torch.Tensor)	Calculate R1 gradient penalty for WGAN-GP.
mask_reduce_loss(→ torch.Tensor)	Apply element-wise weight and reduce loss.
reduce_loss(→ torch.Tensor)	Reduce loss as specified.
tv_loss(→ torch.Tensor)	L2 total variation loss, as in Mahendran et al.

class `mmagic.models.losses.`**`AdvLoss`**

Returns the loss of discriminator with the specified type for DeblurGanv2.

Parameters **loss_type** (*Str*) – One of value in [wgan-gp,lsgan,gan,ragan,ragan-ls].

class mmagic.models.losses.**CLIPLoss**(*loss_weight: float = 1.0*, *data_info: Optional[dict] = None*, *clip_model: dict = dict()*, *loss_name: str = 'loss_clip'*)

Bases: torch.nn.Module

Clip loss. In styleclip, this loss is used to optimize the latent code to generate image that match the text.

In this loss, we may need to provide image, text. Thus, an example of the data_info is:

```
data_info = dict(
    image='fake_imgs',
    text='descriptions')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to 1..
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If None, this module will directly pass the input data to the loss function. Defaults to None.
- **clip_model** (*dict, optional*) – Kwargs for clip loss model. Defaults to dict().
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_clip'.

forward(*image: torch.Tensor*, *text: torch.Tensor*) → torch.Tensor

Forward function.

If self.data_info is not None, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If self.data_info is None, the input argument or key-word argument will be directly passed to loss function, third_party_net_loss.

class mmagic.models.losses.**CharbonnierCompLoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*, *eps: bool = 1e-12*)

Bases: torch.nn.Module

Charbonnier composition loss.

Parameters

- **loss_weight** (*float*) – Loss weight for L1 loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.
- **eps** (*float*) – A value used to control the curvature near zero. Default: 1e-12.

forward(*pred_alpha: torch.Tensor*, *fg: torch.Tensor*, *bg: torch.Tensor*, *ori_merged: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Parameters

- **pred_alpha** (*Tensor*) – of shape (N, 1, H, W). Predicted alpha matte.
- **fg** (*Tensor*) – of shape (N, 3, H, W). Tensor of foreground object.
- **bg** (*Tensor*) – of shape (N, 3, H, W). Tensor of background object.
- **ori_merged** (*Tensor*) – of shape (N, 3, H, W). Tensor of origin merged image before normalized by ImageNet mean and std.
- **weight** (*Tensor, optional*) – of shape (N, 1, H, W). It is an indicating matrix: weight[trimap == 128] = 1. Default: None.

class mmagic.models.losses.**L1CompositionLoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*)

Bases: torch.nn.Module

L1 composition loss.

Parameters

- **loss_weight** (*float*) – Loss weight for L1 loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.

forward(*pred_alpha: torch.Tensor*, *fg: torch.Tensor*, *bg: torch.Tensor*, *ori_merged: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Parameters

- **pred_alpha** (*Tensor*) – of shape (N, 1, H, W). Predicted alpha matte.
- **fg** (*Tensor*) – of shape (N, 3, H, W). Tensor of foreground object.
- **bg** (*Tensor*) – of shape (N, 3, H, W). Tensor of background object.
- **ori_merged** (*Tensor*) – of shape (N, 3, H, W). Tensor of origin merged image before normalized by ImageNet mean and std.
- **weight** (*Tensor, optional*) – of shape (N, 1, H, W). It is an indicating matrix: weight[trimap == 128] = 1. Default: None.

class mmagic.models.losses.**MSECompositionLoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*)

Bases: torch.nn.Module

MSE (L2) composition loss.

Parameters

- **loss_weight** (*float*) – Loss weight for MSE loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.

forward(*pred_alpha: torch.Tensor*, *fg: torch.Tensor*, *bg: torch.Tensor*, *ori_merged: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Parameters

- **pred_alpha** (*Tensor*) – of shape (N, 1, H, W). Predicted alpha matte.
- **fg** (*Tensor*) – of shape (N, 3, H, W). Tensor of foreground object.
- **bg** (*Tensor*) – of shape (N, 3, H, W). Tensor of background object.
- **ori_merged** (*Tensor*) – of shape (N, 3, H, W). Tensor of origin merged image before normalized by ImageNet mean and std.
- **weight** (*Tensor, optional*) – of shape (N, 1, H, W). It is an indicating matrix: weight[trimap == 128] = 1. Default: None.

class mmagic.models.losses.**FaceIdLoss**(*loss_weight: float = 1.0*, *data_info: Optional[dict] = None*, *facenet: dict = dict(type='ArcFace', ir_se50_weights=None)*, *loss_name: str = 'loss_id'*)

Bases: torch.nn.Module

Face similarity loss. Generally this loss is used to keep the id consistency of the input face image and output face image.

In this loss, we may need to provide gt, pred and x. Thus, an example of the data_info is:

```
data_info = dict(
    gt='real_imgs',
    pred='fake_imgs')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to 1..
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If None, this module will directly pass the input data to the loss function. Defaults to None.
- **facenet** (*dict, optional*) – Config dict for facenet. Defaults to dict(type='ArcFace', ir_se50_weights=None, device='cuda').
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_id'.

forward(*pred: torch.Tensor*, *gt: torch.Tensor*) → torch.Tensor

Forward function.

class mmagic.models.losses.**LightCNNFeatureLoss**(*pretrained: str*, *loss_weight: float = 1.0*, *criterion: str = 'l1'*)

Bases: torch.nn.Module

Feature loss of DICGAN, based on LightCNN.

Parameters

- **pretrained** (*str*) – Path for pretrained weights.
- **loss_weight** (*float*) – Loss weight. Default: 1.0.
- **criterion** (*str*) – Criterion type. Options are 'l1' and 'mse'. Default: 'l1'.

forward(*pred: torch.Tensor*, *gt: torch.Tensor*) → torch.Tensor

Forward function.

Parameters

- **pred** (*Tensor*) – Predicted tensor.
- **gt** (*Tensor*) – GT tensor.

Returns Forward results.

Return type Tensor

class mmagic.models.losses.**DiscShiftLoss**(*loss_weight: float = 0.1*)

Bases: torch.nn.Module

Disc shift loss.

Parameters **loss_weight** (*float, optional*) – Loss weight. Defaults to 1.0.

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Tensor with shape (n, c, h, w)

Returns Loss.

Return type Tensor

class mmagic.models.losses.**GANLoss**(*gan_type: str*, *real_label_val: float = 1.0*, *fake_label_val: float = 0.0*, *loss_weight: float = 1.0*)

Bases: torch.nn.Module

Define GAN loss.

Parameters

- **gan_type** (*str*) – Support 'vanilla', 'lsgan', 'wgan', 'hinge', 'l1'.
- **real_label_val** (*float*) – The value for real label. Default: 1.0.
- **fake_label_val** (*float*) – The value for fake label. Default: 0.0.
- **loss_weight** (*float*) – Loss weight. Default: 1.0. Note that loss_weight is only for generators; and it is always 1.0 for discriminators.

_wgan_loss(*input: torch.Tensor*, *target: bool*) → torch.Tensor

wgan loss.

Parameters

- **input** (*Tensor*) – Input tensor.
- **target** (*bool*) – Target label.

Returns wgan loss.

Return type Tensor

get_target_label(*input: torch.Tensor*, *target_is_real: bool*) → Union[bool, torch.Tensor]

Get target label.

Parameters

- **input** (*Tensor*) – Input tensor.
- **target_is_real** (*bool*) – Whether the target is real or fake.

Returns

Target tensor. Return bool for wgan, otherwise, return Tensor.

Return type (bool | Tensor)

forward(*input: torch.Tensor*, *target_is_real: bool*, *is_disc: bool = False*, *mask: Optional[torch.Tensor] = None*) → torch.Tensor

Parameters

- **input** (*Tensor*) – The input for the loss module, i.e., the network prediction.
- **target_is_real** (*bool*) – Whether the target is real or fake.
- **is_disc** (*bool*) – Whether the loss for discriminators or not. Default: False.
- **mask** (*Tensor*) – The mask tensor. Default: None.

Returns GAN loss value.

Return type Tensor

class mmagic.models.losses.**GaussianBlur**(*kernel_size: Tuple[int, int] = (71, 71)*, *sigma: Tuple[float, float] = (10.0, 10.0)*)

Bases: torch.nn.Module

A Gaussian filter which blurs a given tensor with a two-dimensional gaussian kernel by convolving it along each channel. Batch operation is supported.

This function is modified from kornia.filters.gaussian: *<https://kornia.readthedocs.io/en/latest/_modules/kornia/filters/gaussian.ht*

Parameters

- **kernel_size** (*tuple[int]*) – The size of the kernel. Default: (71, 71).
- **sigma** (*tuple[float]*) – The standard deviation of the kernel.
- **Default** (*10.0, 10.0*) –

Returns The Gaussian-blurred tensor.

Return type Tensor

Shape:

- input: Tensor with shape of (n, c, h, w)
- output: Tensor with shape of (n, c, h, w)

static compute_zero_padding(*kernel_size: Tuple[int, int]*) → tuple

Compute zero padding tuple.

Parameters **kernel_size** (*tuple[int]*) – The size of the kernel.

Returns Padding of height and weight.

Return type tuple

get_2d_gaussian_kernel(*kernel_size: Tuple[int, int]*, *sigma: Tuple[float, float]*) → torch.Tensor

Get the two-dimensional Gaussian filter matrix coefficients.

Parameters

- **kernel_size** (*tuple[int]*) – Kernel filter size in the x and y direction. The kernel sizes should be odd and positive.

- **sigma** (*tuple[int]*) – Gaussian standard deviation in the x and y direction.

Returns

A 2D torch tensor with gaussian filter matrix coefficients.

Return type kernel_2d (Tensor)

get_1d_gaussian_kernel(*kernel_size: int*, *sigma: float*) → torch.Tensor

Get the Gaussian filter coefficients in one dimension (x or y direction).

Parameters

- **kernel_size** (*int*) – Kernel filter size in x or y direction. Should be odd and positive.
- **sigma** (*float*) – Gaussian standard deviation in x or y direction.

Returns

A 1D torch tensor with gaussian filter coefficients in x or y direction.

Return type kernel_1d (Tensor)

gaussian(*kernel_size: int*, *sigma: float*) → torch.Tensor

Gaussian function.

Parameters

- **kernel_size** (*int*) – Kernel filter size in x or y direction. Should be odd and positive.
- **sigma** (*float*) – Gaussian standard deviation in x or y direction.

Returns

A 1D torch tensor with gaussian filter coefficients in x or y direction.

Return type Tensor

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Tensor with shape (n, c, h, w)

Returns The Gaussian-blurred tensor.

Return type Tensor

class mmagic.models.losses.**GradientPenaltyLoss**(*loss_weight: float = 1.0*)

Bases: torch.nn.Module

Gradient penalty loss for wgan-gp.

Parameters **loss_weight** (*float*) – Loss weight. Default: 1.0.

forward(*discriminator: torch.nn.Module*, *real_data: torch.Tensor*, *fake_data: torch.Tensor*, *mask: Optional[torch.Tensor] = None*) → torch.Tensor

Forward function.

Parameters

- **discriminator** (*nn.Module*) – Network for the discriminator.
- **real_data** (*Tensor*) – Real input data.
- **fake_data** (*Tensor*) – Fake input data.
- **mask** (*Tensor*) – Masks for inpainting. Default: None.

Returns Loss.

Return type Tensor

`mmagic.models.losses.disc_shift_loss(pred: torch.Tensor)` → torch.Tensor

Disc Shift loss.

This loss is proposed in PGGAN as an auxiliary loss for discriminator.

Parameters **pred** (*Tensor*) – Input tensor.

Returns loss tensor.

Return type torch.Tensor

`mmagic.models.losses.gen_path_regularizer(generator: torch.nn.Module, num_batches: int, mean_path_length: torch.Tensor, pl_batch_shrink: int = 1, decay: float = 0.01, weight: float = 1.0, pl_batch_size: Optional[int] = None, sync_mean_buffer: bool = False, loss_scaler: Optional[torch.cuda.amp.grad_scaler.GradScaler] = None, use_apex_amp: bool = False)` → Tuple[torch.Tensor]

Generator Path Regularization.

Path regularization is proposed in StyleGAN2, which can help the improve the continuity of the latent space. More details can be found in: Analyzing and Improving the Image Quality of StyleGAN, CVPR2020.

Parameters

- **generator** (*nn.Module*) – The generator module. Note that this loss requires that the generator contains `return_latents` interface, with which we can get the latent code of the current sample.
- **num_batches** (*int*) – The number of samples used in calculating this loss.
- **mean_path_length** (*Tensor*) – The mean path length, calculated by moving average.
- **pl_batch_shrink** (*int, optional*) – The factor of shrinking the batch size for saving GPU memory. Defaults to 1.
- **decay** (*float, optional*) – Decay for moving average of mean path length. Defaults to 0.01.
- **weight** (*float, optional*) – Weight of this loss item. Defaults to `1.`.
- **pl_batch_size** (*int | None, optional*) – The batch size in calculating generator path. Once this argument is set, the `num_batches` will be overridden with this argument and won't be affected by `pl_batch_shrink`. Defaults to None.
- **sync_mean_buffer** (*bool, optional*) – Whether to sync mean path length across all of GPUs. Defaults to False.

Returns The penalty loss, detached mean path tensor, and current path length.

Return type tuple[Tensor]

`mmagic.models.losses.gradient_penalty_loss(discriminator: torch.nn.Module, real_data: torch.Tensor, fake_data: torch.Tensor, mask: Optional[torch.Tensor] = None, norm_mode: str = 'pixel')` → torch.Tensor

Calculate gradient penalty for wgan-gp.

Parameters

- **discriminator** (*nn.Module*) – Network for the discriminator.

- **real_data** (*Tensor*) – Real input data.
- **fake_data** (*Tensor*) – Fake input data.
- **mask** (*Tensor*) – Masks for inpainting. Default: None.

Returns A tensor for gradient penalty.

Return type Tensor

mmagic.models.losses.**r1_gradient_penalty_loss**(*discriminator: torch.nn.Module*, *real_data: torch.Tensor*, *mask: Optional[torch.Tensor] = None*, *norm_mode: str = 'pixel'*, *loss_scaler: Optional[torch.cuda.amp.grad_scaler.GradScaler] = None*, *use_apex_amp: bool = False*) → torch.Tensor

Calculate R1 gradient penalty for WGAN-GP.

R1 regularizer comes from: "Which Training Methods for GANs do actually Converge?" ICML'2018

Different from original gradient penalty, this regularizer only penalized gradient w.r.t. real data.

Parameters

- **discriminator** (*nn.Module*) – Network for the discriminator.
- **real_data** (*Tensor*) – Real input data.
- **mask** (*Tensor*) – Masks for inpainting. Default: None.
- **norm_mode** (*str*) – This argument decides along which dimension the norm of the gradients will be calculated. Currently, we support ["pixel" , "HWC"]. Defaults to "pixel".

Returns A tensor for gradient penalty.

Return type Tensor

class mmagic.models.losses.**GradientLoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*)

Bases: torch.nn.Module

Gradient loss.

Parameters

- **loss_weight** (*float*) – Loss weight for L1 loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.

forward(*pred: torch.Tensor*, *target: torch.Tensor*, *weight: Optional[torch.Tensor] = None*) → torch.Tensor

Parameters

- **pred** (*Tensor*) – of shape (N, C, H, W). Predicted tensor.
- **target** (*Tensor*) – of shape (N, C, H, W). Ground truth tensor.
- **weight** (*Tensor, optional*) – of shape (N, C, H, W). Element-wise weights. Default: None.

class mmagic.models.losses.**CLIPLossComps**(*loss_weight: float = 1.0*, *data_info: Optional[dict] = None*, *clip_model: dict = dict()*, *loss_name: str = 'loss_clip'*)

Bases: torch.nn.Module

Clip loss. In styleclip, this loss is used to optimize the latent code to generate image that match the text.

In this loss, we may need to provide image, text. Thus, an example of the data_info is:

```
data_info = dict(
    image='fake_imgs',
    text='descriptions')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to 1..
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If None, this module will directly pass the input data to the loss function. Defaults to None.
- **clip_model** (*dict, optional*) – Kwargs for clip loss model. Defaults to dict().
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_clip'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If self.data_info is not None, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If self.data_info is None, the input argument or key-word argument will be directly passed to loss function, third_party_net_loss.

static loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

class mmagic.models.losses.**DiscShiftLossComps**(*loss_weight: float = 1.0*, *data_info: Optional[dict] = None*, *loss_name: str = 'loss_disc_shift'*)

Bases: torch.nn.Module

Disc Shift Loss.

This loss is proposed in PGGAN as an auxiliary loss for discriminator.

Note for the design of ``data_info``: In MMagic, almost all of loss modules contain the argument data_info, which can be used for constructing the link between the input items (needed in loss calculation) and the data from the generative model. For example, in the training of GAN model, we will collect all of important data/modules into a dictionary:

Listing 1: Code from StaticUnconditionalGAN, train_step

```
data_dict_ = dict(
    gen=self.generator,
    disc=self.discriminator,
    disc_pred_fake=disc_pred_fake,
    disc_pred_real=disc_pred_real,
```

(continues on next page)

(continued from previous page)

```
    fake_imgs=fake_imgs,
    real_imgs=real_imgs,
    iteration=curr_iter,
    batch_size=batch_size)
```

But in this loss, we will need to provide `pred` as input. Thus, an example of the `data_info` is:

```
data_info = dict(
    pred='disc_pred_fake')
```

Then, the module will automatically construct this mapping from the input data dictionary.

In addition, in general, `disc_shift_loss` will be applied over real and fake data. In this case, users just need to add this loss module twice, but with different `data_info`. Our model will automatically add these two items.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to `1.`.
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If `None`, this module will directly pass the input data to the loss function. Defaults to None.
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_disc_shift'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If `self.data_info` is not `None`, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If `self.data_info` is `None`, the input argument or key-word argument will be directly passed to loss function, `disc_shift_loss`.

loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

class mmagic.models.losses.**FaceIdLossComps**(*loss_weight: float = 1.0*, *data_info: Optional[dict] = None*, *facenet: dict = dict(type='ArcFace', ir_se50_weights=None)*, *loss_name: str = 'loss_id'*)

Bases: `torch.nn.Module`

Face similarity loss. Generally this loss is used to keep the id consistency of the input face image and output face image.

In this loss, we may need to provide `gt`, `pred` and `x`. Thus, an example of the `data_info` is:

```
data_info = dict(
    gt='real_imgs',
    pred='fake_imgs')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to `1.`.
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If `None`, this module will directly pass the input data to the loss function. Defaults to None.
- **facenet** (*dict, optional*) – Config dict for facenet. Defaults to dict(type='ArcFace', ir_se50_weights=None).
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_id'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If `self.data_info` is not `None`, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If `self.data_info` is `None`, the input argument or key-word argument will be directly passed to loss function, `third_party_net_loss`.

loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

class mmagic.models.losses.**GANLossComps**(*gan_type: str*, *real_label_val: float = 1.0*, *fake_label_val: float = 0.0*, *loss_weight: float = 1.0*)

Bases: `torch.nn.Module`

Define GAN loss.

Parameters

- **gan_type** (*str*) – Support 'vanilla', 'lsgan', 'wgan', 'hinge', 'wgan-logistic-ns'.
- **real_label_val** (*float*) – The value for real label. Default: 1.0.
- **fake_label_val** (*float*) – The value for fake label. Default: 0.0.
- **loss_weight** (*float*) – Loss weight. Default: 1.0. Note that loss_weight is only for generators; and it is always 1.0 for discriminators.

_wgan_loss(*input: torch.Tensor*, *target: bool*) → torch.Tensor

wgan loss.

Parameters

- **input** (*Tensor*) – Input tensor.
- **target** (*bool*) – Target label.

Returns wgan loss.

Return type Tensor

_wgan_logistic_ns_loss(*input: torch.Tensor*, *target: bool*) → torch.Tensor

WGAN loss in logistically non-saturating mode.

This loss is widely used in StyleGANv2.

Parameters

- **input** (*Tensor*) – Input tensor.
- **target** (*bool*) – Target label.

Returns wgan loss.

Return type Tensor

get_target_label(*input: torch.Tensor*, *target_is_real: bool*) → Union[bool, torch.Tensor]

Get target label.

Parameters

- **input** (*Tensor*) – Input tensor.
- **target_is_real** (*bool*) – Whether the target is real or fake.

Returns Target tensor. Return bool for wgan, otherwise, return Tensor.

Return type (bool | Tensor)

forward(*input: torch.Tensor*, *target_is_real: bool*, *is_disc: bool = False*) → torch.Tensor

Parameters

- **input** (*Tensor*) – The input for the loss module, i.e., the network prediction.
- **target_is_real** (*bool*) – Whether the targe is real or fake.
- **is_disc** (*bool*) – Whether the loss for discriminators or not. Default: False.

Returns GAN loss value.

Return type Tensor

class mmagic.models.losses.**GeneratorPathRegularizerComps**(*loss_weight: float = 1.0*, *pl_batch_shrink: int = 1*, *decay: float = 0.01*, *pl_batch_size: Optional[int] = None*, *sync_mean_buffer: bool = False*, *interval: int = 1*, *data_info: Optional[dict] = None*, *use_apex_amp: bool = False*, *loss_name: str = 'loss_path_regular'*)

Bases: torch.nn.Module

Generator Path Regularizer.

Path regularization is proposed in StyleGAN2, which can help the improve the continuity of the latent space. More details can be found in: Analyzing and Improving the Image Quality of StyleGAN, CVPR2020.

Users can achieve lazy regularization by setting `interval` arguments here.

Note for the design of ``data_info``: In `MMagic`, almost all of loss modules contain the argument `data_info`, which can be used for constructing the link between the input items (needed in loss calculation) and the data from the generative model. For example, in the training of GAN model, we will collect all of important data/modules into a dictionary:

Listing 2: Code from StaticUnconditionalGAN, train_step

```
data_dict_ = dict(
    gen=self.generator,
    disc=self.discriminator,
    fake_imgs=fake_imgs,
    disc_pred_fake_g=disc_pred_fake_g,
    iteration=curr_iter,
    batch_size=batch_size)
```

But in this loss, we will need to provide `generator` and `num_batches` as input. Thus an example of the `data_info` is:

```
data_info = dict(
    generator='gen',
    num_batches='batch_size')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to `1.`.
- **pl_batch_shrink** (*int, optional*) – The factor of shrinking the batch size for saving GPU memory. Defaults to 1.
- **decay** (*float, optional*) – Decay for moving average of mean path length. Defaults to 0.01.
- **pl_batch_size** (*int | None, optional*) – The batch size in calculating generator path. Once this argument is set, the `num_batches` will be overridden with this argument and won't be affected by `pl_batch_shrink`. Defaults to None.
- **sync_mean_buffer** (*bool, optional*) – Whether to sync mean path length across all of GPUs. Defaults to False.
- **interval** (*int, optional*) – The interval of calculating this loss. This argument is used to support lazy regularization. Defaults to 1.
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If `None`, this module will directly pass the input data to the loss function. Defaults to None.
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_path_regular'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If `self.data_info` is not `None`, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If `self.data_info` is `None`, the input argument or key-word argument will be directly passed to loss function, `gen_path_regularizer`.

loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

class mmagic.models.losses.**GradientPenaltyLossComps**(*loss_weight: float = 1.0, norm_mode: str = 'pixel', data_info: Optional[dict] = None, loss_name: str = 'loss_gp'*)

Bases: `torch.nn.Module`

Gradient Penalty for WGAN-GP.

In the detailed implementation, there are two streams where one uses the pixel-wise gradient norm, but the other adopts normalization along instance (HWC) dimensions. Thus, `norm_mode` are offered to define which mode you want.

Note for the design of ``data_info``: In `MMagic`, almost all of loss modules contain the argument `data_info`, which can be used for constructing the link between the input items (needed in loss calculation) and the data from the generative model. For example, in the training of GAN model, we will collect all of important data/modules into a dictionary:

Listing 3: Code from StaticUnconditionalGAN, train_step

```
data_dict_ = dict(
    gen=self.generator,
    disc=self.discriminator,
    disc_pred_fake=disc_pred_fake,
    disc_pred_real=disc_pred_real,
    fake_imgs=fake_imgs,
    real_imgs=real_imgs,
    iteration=curr_iter,
    batch_size=batch_size)
```

But in this loss, we will need to provide `discriminator`, `real_data`, and `fake_data` as input. Thus, an example of the `data_info` is:

```
data_info = dict(
    discriminator='disc',
    real_data='real_imgs',
    fake_data='fake_imgs')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to `1.`.
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If `None`, this module will directly pass the input data to the loss function. Defaults to None.

- **norm_mode** (*str*) – This argument decides along which dimension the norm of the gradients will be calculated. Currently, we support ["pixel" , "HWC"]. Defaults to "pixel".
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_gp'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If `self.data_info` is not `None`, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If `self.data_info` is `None`, the input argument or key-word argument will be directly passed to loss function, `gradient_penalty_loss`.

loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

class mmagic.models.losses.**R1GradientPenaltyComps**(*loss_weight: float = 1.0*, *norm_mode: str = 'pixel'*, *interval: int = 1*, *data_info: Optional[dict] = None*, *use_apex_amp: bool = False*, *loss_name: str = 'loss_r1_gp'*)

Bases: `torch.nn.Module`

R1 gradient penalty for WGAN-GP.

R1 regularizer comes from: "Which Training Methods for GANs do actually Converge?" ICML'2018

Different from original gradient penalty, this regularizer only penalized gradient w.r.t. real data.

Note for the design of ``data_info``: In `MMagic`, almost all of loss modules contain the argument `data_info`, which can be used for constructing the link between the input items (needed in loss calculation) and the data from the generative model. For example, in the training of GAN model, we will collect all of important data/modules into a dictionary:

Listing 4: Code from StaticUnconditionalGAN, train_step

```
data_dict_ = dict(
    gen=self.generator,
    disc=self.discriminator,
    disc_pred_fake=disc_pred_fake,
    disc_pred_real=disc_pred_real,
    fake_imgs=fake_imgs,
    real_imgs=real_imgs,
    iteration=curr_iter,
    batch_size=batch_size)
```

But in this loss, we will need to provide `discriminator` and `real_data` as input. Thus, an example of the `data_info` is:

```
data_info = dict(
    discriminator='disc',
    real_data='real_imgs')
```

Then, the module will automatically construct this mapping from the input data dictionary.

Parameters

- **loss_weight** (*float, optional*) – Weight of this loss item. Defaults to 1..
- **data_info** (*dict, optional*) – Dictionary contains the mapping between loss input args and data dictionary. If None, this module will directly pass the input data to the loss function. Defaults to None.
- **norm_mode** (*str*) – This argument decides along which dimension the norm of the gradients will be calculated. Currently, we support ["pixel" , "HWC"]. Defaults to "pixel".
- **interval** (*int, optional*) – The interval of calculating this loss. Defaults to 1.
- **loss_name** (*str, optional*) – Name of the loss item. If you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name. Defaults to 'loss_r1_gp'.

forward(**args*, ***kwargs*) → torch.Tensor

Forward function.

If self.data_info is not None, a dictionary containing all of the data and necessary modules should be passed into this function. If this dictionary is given as a non-keyword argument, it should be offered as the first argument. If you are using keyword argument, please name it as *outputs_dict*.

If self.data_info is None, the input argument or key-word argument will be directly passed to loss function, r1_gradient_penalty_loss.

loss_name() → str

Loss Name.

This function must be implemented and will return the name of this loss function. This name will be used to combine different loss items by simple sum operation. In addition, if you want this loss item to be included into the backward graph, *loss_* must be the prefix of the name.

Returns The name of this loss item.

Return type str

mmagic.models.losses.**mask_reduce_loss**(*loss: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, *reduction: str = 'mean'*, *sample_wise: bool = False*) → torch.Tensor

Apply element-wise weight and reduce loss.

Parameters

- **loss** (*Tensor*) – Element-wise loss.
- **weight** (*Tensor*) – Element-wise weights. Default: None.
- **reduction** (*str*) – Same as built-in losses of PyTorch. Options are "none", "mean" and "sum". Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.

Returns Processed loss values.

Return type Tensor

`mmagic.models.losses.reduce_loss(loss: torch.Tensor, reduction: str)` → torch.Tensor

Reduce loss as specified.

Parameters

- **loss** (*Tensor*) – Elementwise loss tensor.
- **reduction** (*str*) – Options are "none", "mean" and "sum".

Returns Reduced loss tensor.

Return type Tensor

`class mmagic.models.losses.PerceptualLoss(layer_weights: dict, layer_weights_style: Optional[dict] = None, vgg_type: str = 'vgg19', use_input_norm: bool = True, perceptual_weight: float = 1.0, style_weight: float = 1.0, norm_img: bool = True, pretrained: str = 'torchvision://vgg19', criterion: str = 'l1')`

Bases: `torch.nn.Module`

Perceptual loss with commonly used style loss.

Parameters

- **layers_weights** (*dict*) – The weight for each layer of vgg feature for perceptual loss. Here is an example: {'4': 1., '9': 1., '18': 1.}, which means the 5th, 10th and 18th feature layer will be extracted with weight 1.0 in calculating losses.
- **layers_weights_style** (*dict*) – The weight for each layer of vgg feature for style loss. If set to 'None', the weights are set equal to the weights for perceptual loss. Default: None.
- **vgg_type** (*str*) – The type of vgg network used as feature extractor. Default: 'vgg19'.
- **use_input_norm** (*bool*) – If True, normalize the input image in vgg. Default: True.
- **perceptual_weight** (*float*) – If *perceptual_weight > 0*, the perceptual loss will be calculated and the loss will multiplied by the weight. Default: 1.0.
- **style_weight** (*float*) – If *style_weight > 0*, the style loss will be calculated and the loss will multiplied by the weight. Default: 1.0.
- **norm_img** (*bool*) – If True, the image will be normed to [0, 1]. Note that this is different from the *use_input_norm* which norm the input in in forward function of vgg according to the statistics of dataset. Importantly, the input image must be in range [-1, 1].
- **pretrained** (*str*) – Path for pretrained weights. Default: 'torchvision://vgg19'.
- **criterion** (*str*) – Criterion type. Options are 'l1' and 'mse'. Default: 'l1'.

`forward(x: torch.Tensor, gt: torch.Tensor)` → Tuple[torch.Tensor]

Forward function.

Parameters

- **x** (*Tensor*) – Input tensor with shape (n, c, h, w).
- **gt** (*Tensor*) – Ground-truth tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

_gram_mat(*x: torch.Tensor*) → torch.Tensor

Calculate Gram matrix.

Parameters **x** (*torch.Tensor*) – Tensor with shape of (n, c, h, w).

Returns Gram matrix.

Return type torch.Tensor

class mmagic.models.losses.**PerceptualVGG**(*layer_name_list: List[str]*, *vgg_type: str = 'vgg19'*, *use_input_norm: bool = True*, *pretrained: str = 'torchvision://vgg19'*)

Bases: torch.nn.Module

VGG network used in calculating perceptual loss.

In this implementation, we allow users to choose whether use normalization in the input feature and the type of vgg network. Note that the pretrained path must fit the vgg type.

Parameters

- **layer_name_list** (*list[str]*) – According to the name in this list, forward function will return the corresponding features. This list contains the name each layer in *vgg.feature*. An example of this list is ['4', '10'].
- **vgg_type** (*str*) – Set the type of vgg network. Default: 'vgg19'.
- **use_input_norm** (*bool*) – If True, normalize the input image. Importantly, the input feature must in the range [0, 1]. Default: True.
- **pretrained** (*str*) – Path for pretrained weights. Default: 'torchvision://vgg19'

forward(*x: torch.Tensor*) → torch.Tensor

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

init_weights(*model: torch.nn.Module*, *pretrained: str*) → None

Init weights.

Parameters

- **model** (*nn.Module*) – Models to be inited.
- **pretrained** (*str*) – Path for pretrained weights.

class mmagic.models.losses.**TransferalPerceptualLoss**(*loss_weight: float = 1.0*, *use_attention: bool = True*, *criterion: str = 'mse'*)

Bases: torch.nn.Module

Transferal perceptual loss.

Parameters

- **loss_weight** (*float*) – Loss weight. Default: 1.0.
- **use_attention** (*bool*) – If True, use soft-attention tensor. Default: True
- **criterion** (*str*) – Criterion type. Options are 'l1' and 'mse'. Default: 'mse'.

forward(*maps: Tuple[torch.Tensor]*, *soft_attention: torch.Tensor*, *textures: Tuple[torch.Tensor]*) → torch.Tensor

Forward function.

Parameters

- **maps** (*Tuple[Tensor]*) – Input tensors.
- **soft_attention** (*Tensor*) – Soft-attention tensor.
- **textures** (*Tuple[Tensor]*) – Ground-truth tensors.

Returns Forward results.

Return type Tensor

class mmagic.models.losses.**CharbonnierLoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*, *eps: float = 1e-12*)

Bases: torch.nn.Module

Charbonnier loss (one variant of Robust L1Loss, a differentiable variant of L1Loss).

Described in "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution".

Parameters

- **loss_weight** (*float*) – Loss weight for L1 loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.
- **eps** (*float*) – A value used to control the curvature near zero. Default: 1e-12.

forward(*pred: torch.Tensor*, *target: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Forward Function.

Parameters

- **pred** (*Tensor*) – of shape (N, C, H, W). Predicted tensor.
- **target** (*Tensor*) – of shape (N, C, H, W). Ground truth tensor.
- **weight** (*Tensor, optional*) – of shape (N, C, H, W). Element-wise weights. Default: None.

class mmagic.models.losses.**L1Loss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*)

Bases: torch.nn.Module

L1 (mean absolute error, MAE) loss.

Parameters

- **loss_weight** (*float*) – Loss weight for L1 loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.

- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduce loss with 'mean' per-sample, and then it means over all the samples. Default: False.

forward(*pred: torch.Tensor*, *target: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Forward Function.

Parameters

- **pred** (*Tensor*) – of shape (N, C, H, W). Predicted tensor.
- **target** (*Tensor*) – of shape (N, C, H, W). Ground truth tensor.
- **weight** (*Tensor, optional*) – of shape (N, C, H, W). Element-wise weights. Default: None.

class mmagic.models.losses.**MaskedTVLoss**(*loss_weight: float = 1.0*)

Bases: *L1Loss*

Masked TV loss.

Parameters **loss_weight** (*float, optional*) – Loss weight. Defaults to 1.0.

forward(*pred: torch.Tensor*, *mask: Optional[torch.Tensor] = None*) → torch.Tensor

Forward function.

Parameters

- **pred** (*torch.Tensor*) – Tensor with shape of (n, c, h, w).
- **mask** (*torch.Tensor, optional*) – Tensor with shape of (n, 1, h, w). Defaults to None.

Returns [description]

Return type [type]

class mmagic.models.losses.**MSELoss**(*loss_weight: float = 1.0*, *reduction: str = 'mean'*, *sample_wise: bool = False*)

Bases: torch.nn.Module

MSE (L2) loss.

Parameters

- **loss_weight** (*float*) – Loss weight for MSE loss. Default: 1.0.
- **reduction** (*str*) – Specifies the reduction to apply to the output. Supported choices are 'none' | 'mean' | 'sum'. Default: 'mean'.
- **sample_wise** (*bool*) – Whether calculate the loss sample-wise. This argument only takes effect when *reduction* is 'mean' and *weight* (argument of *forward()*) is not None. It will first reduces loss with 'mean' per-sample, and then it means over all the samples. Default: False.

forward(*pred: torch.Tensor*, *target: torch.Tensor*, *weight: Optional[torch.Tensor] = None*, ***kwargs*) → torch.Tensor

Forward Function.

Parameters

- **pred** (*Tensor*) – of shape (N, C, H, W). Predicted tensor.
- **target** (*Tensor*) – of shape (N, C, H, W). Ground truth tensor.

- **weight** (*Tensor, optional*) – of shape (N, C, H, W). Element-wise weights. Default: None.

class mmagic.models.losses.**PSNRLoss**(*loss_weight: float = 1.0, toY: bool = False*)

Bases: torch.nn.Module

PSNR Loss in "HINet: Half Instance Normalization Network for Image Restoration".

Parameters

- **loss_weight** (*float, optional*) – Loss weight. Defaults to 1.0.
- **reduction** – reduction for PSNR. Can only be mean here.
- **toY** – change to calculate the PSNR of Y channel in YCbCr format

forward(*pred: torch.Tensor, target: torch.Tensor*) → torch.Tensor

mmagic.models.losses.**tv_loss**(*input: torch.Tensor*) → torch.Tensor

L2 total variation loss, as in Mahendran et al.

1.90 mmagic.models.data_preprocessors

1.90.1 Package Contents

Classes

DataPreprocessor	Image pre-processor for generative models. This class provide
MattorPreprocessor	DataPreprocessor for matting models.

class mmagic.models.data_preprocessors.**DataPreprocessor**(*mean: Union[Sequence[Union[float, int]], float, int] = 127.5, std: Union[Sequence[Union[float, int]], float, int] = 127.5, pad_size_divisor: int = 1, pad_value: Union[float, int] = 0, pad_mode: str = 'constant', non_image_keys: Optional[Tuple[str, List[str]]] = None, non_concentate_keys: Optional[Tuple[str, List[str]]] = None, output_channel_order: Optional[str] = None, data_keys: Union[List[str], str] = 'gt_img', input_view: Optional[tuple] = None, output_view: Optional[tuple] = None, stack_data_sample=True*)

Bases: mmengine.model.ImgDataPreprocessor

Image pre-processor for generative models. This class provide normalization and bgr to rgb conversion for image tensor inputs. The input of this classes should be dict which keys are *inputs* and *data_samples*.

Besides to process tensor *inputs*, this class support dict as *inputs*. - If the value is *Tensor* and the corresponding key is not contained in *_NON_IMAGE_KEYS*, it will be processed as image tensor. - If the value is *Tensor* and the corresponding key belongs to *_NON_IMAGE_KEYS*, it will not remains unchanged. - If value is string or integer, it will not remains unchanged.

Parameters

- **mean** (*Sequence[float or int], float or int, optional*) – The pixel mean of image channels. Noted that normalization operation is performed *after channel order conversion*. If it is not specified, images will not be normalized. Defaults None.
- **std** (*Sequence[float or int], float or int, optional*) – The pixel standard deviation of image channels. Noted that normalization operation is performed *after channel order conversion*. If it is not specified, images will not be normalized. Defaults None.
- **pad_size_divisor** (*int*) – The size of padded image should be divisible by `pad_size_divisor`. Defaults to 1.
- **pad_value** (*float or int*) – The padded pixel value. Defaults to 0.
- **pad_mode** (*str*) – Padding mode for `torch.nn.functional.pad`. Defaults to 'constant'.
- **non_image_keys** (*List[str] or str*) – Keys for fields that not need to be processed (padding, channel conversion and normalization) as images. If not passed, the keys in *_NON_IMAGE_KEYS* will be used. This argument will only work when *inputs* is dict or list of dict. Defaults to None.
- **non_concatenate_keys** (*List[str] or str*) – Keys for fields that not need to be concatenated. If not passed, the keys in *_NON_CONCATENATE_KEYS* will be used. This argument will only work when *inputs* is dict or list of dict. Defaults to None.
- **output_channel_order** (*str, optional*) – The desired image channel order of output the data preprocessor. This is also the desired input channel order of model (and this most likely to be the output order of model). If not passed, no channel order conversion will be performed. Defaults to None.
- **data_keys** (*List[str] or str*) – Keys to preprocess in data samples. Defaults to 'gt_img'.
- **input_view** (*tuple, optional*) – The view of input tensor. This argument maybe deleted in the future. Defaults to None.
- **output_view** (*tuple, optional*) – The view of output tensor. This argument maybe deleted in the future. Defaults to None.
- **stack_data_sample** (*bool*) – Whether stack a list of data samples to one data sample. Only support with input data samples are *DataSamples*. Defaults to True.

_NON_IMAGE_KEYS = ['noise']

_NON_CONCATENATE_KEYS = ['num_batches', 'mode', 'sample_kwargs', 'eq_cfg']

cast_data(*data: CastData*) → CastData

Copying data to the target device.

Parameters **data** (*dict*) – Data returned by `DataLoader`.

Returns Inputs and data sample at target device.

Return type CollatedResult

static _parse_channel_index(*inputs*) → int

Parse channel index of inputs.

_parse_channel_order(*key: str, inputs: torch.Tensor, data_sample: Optional[*mmagic.structures.DataSample*] = None*) → str

_parse_batch_channel_order(*key: str*, *inputs: Sequence*, *data_samples: Optional[Sequence[*mmagic.structures.DataSample*]]*) → str

Parse channel order of inputs in batch.

_update_metainfo(*padding_info: torch.Tensor*, *channel_order_info: Optional[dict] = None*, *data_samples: Optional[mmagic.utils.typing.SampleList] = None*) → mmagic.utils.typing.SampleList

Update *padding_info* and *channel_order* to metainfo of.

a batch of `data_samples`. For channel order, we consider same field among data samples share the same channel order. Therefore *channel_order* is passed as a dict, which key and value are field name and corresponding channel order. For padding info, we consider padding info is same among all field of a sample, but can vary between samples. Therefore, we pass *padding_info* as Tensor shape like (B, 1, 1).

Parameters

- **padding_info** (*Tensor*) – The padding info of each sample. Shape like (B, 1, 1).
- **channel_order** (*dict, Optional*) – The channel order of target field. Key and value are field name and corresponding channel order respectively.
- **data_samples** (*List[DataSample], optional*) – The data samples to be updated. If not passed, will initialize a list of empty data samples. Defaults to None.

Returns The updated data samples.

Return type List[DataSample]

_do_conversion(*inputs: torch.Tensor*, *inputs_order: str = 'BGR'*, *target_order: Optional[str] = None*) → Tuple[torch.Tensor, str]

Conduct channel order conversion for *a batch of inputs*, and return the converted inputs and order after conversion.

inputs_order:

- RGB / RGB: Convert to target order.
- SINGLE: Do not change

_do_norm(*inputs: torch.Tensor*, *do_norm: Optional[bool] = None*) → torch.Tensor

_preprocess_image_tensor(*inputs: torch.Tensor*, *data_samples: Optional[mmagic.utils.typing.SampleList] = None*, *key: str = 'img'*) → Tuple[torch.Tensor, mmagic.utils.typing.SampleList]

Preprocess a batch of image tensor and update metainfo to corresponding data samples.

Parameters

- **inputs** (*Tensor*) – Image tensor with shape (C, H, W), (N, C, H, W) or (N, t, C, H, W) to preprocess.
- **data_samples** (*List[DataSample], optional*) – The data samples of corresponding inputs. If not passed, a list of empty data samples will be initialized to save metainfo. Defaults to None.
- **key** (*str*) – The key of image tensor in data samples. Defaults to 'img'.

Returns

The preprocessed image tensor and updated data samples.

Return type Tuple[Tensor, List[DataSample]]

_preprocess_image_list(*tensor_list: List[torch.Tensor]*, *data_samples: Optional[mmagic.utils.typing.SampleList]*, *key: str = 'img'*) → Tuple[torch.Tensor, mmagic.utils.typing.SampleList]

Preprocess a list of image tensor and update metainfo to corresponding data samples.

Parameters

- **tensor_list** (*List[Tensor]*) – Image tensor list to be preprocess.
- **data_samples** (*List[DataSample], optional*) – The data samples of corresponding inputs. If not passed, a list of empty data samples will be initialized to save metainfo. Defaults to None.
- **key** (*str*) – The key of tensor list in data samples. Defaults to 'img'.

Returns

The preprocessed image tensor and updated data samples.

Return type Tuple[Tensor, List[DataSample]]

_preprocess_dict_inputs(*batch_inputs: dict*, *data_samples: Optional[mmagic.utils.typing.SampleList] = None*) → Tuple[dict, mmagic.utils.typing.SampleList]

Preprocess dict type inputs.

Parameters

- **batch_inputs** (*dict*) – Input dict.
- **data_samples** (*List[DataSample], optional*) – The data samples of corresponding inputs. If not passed, a list of empty data samples will be initialized to save metainfo. Defaults to None.

Returns

The preprocessed dict and updated data samples.

Return type Tuple[dict, List[DataSample]]

_preprocess_data_sample(*data_samples: mmagic.utils.typing.SampleList*, *training: bool*) → *mmagic.structures.DataSample*

Preprocess data samples. When *training* is True, fields belong to `self.data_keys` will be converted to `self.output_channel_order` and then normalized by *self.mean* and *self.std*. When *training* is False, fields belongs to `self.data_keys` will be attempted to convert to 'BGR' without normalization. The corresponding metainfo related to normalization, channel order conversion will be updated to data sample as well.

Parameters

- **data_samples** (*List[DataSample]*) – A list of data samples to preprocess.
- **training** (*bool*) – Whether in training mode.

Returns The list of processed data samples.

Return type list

forward(*data: dict*, *training: bool = False*) → dict

Performs normalizationpadding and channel order conversion.

Parameters

- **data** (*dict*) – Input data to process.
- **training** (*bool*) – Whether to in training mode. Default: False.

Returns Data in the same format as the model input.

Return type dict

destruct(*outputs: torch.Tensor, data_samples: Union[mmagic.utils.typing.SampleList, mmagic.structures.DataSample, None] = None, key: str = 'img'*) → Union[list, torch.Tensor]

Destruct padding, normalization and convert channel order to BGR if could. If *data_samples* is a list, outputs will be destructed as a batch of tensor. If *data_samples* is a *DataSample*, *outputs* will be destructed as a single tensor.

Before feed model outputs to visualizer and evaluator, users should call this function for model outputs and inputs.

Use cases:

```
>>> # destruct model outputs.
>>> # model outputs share the same preprocess information with inputs
>>> # ('img') therefore use 'img' as key
>>> feats = self.forward_tensor(inputs, data_samples, **kwargs)
>>> feats = self.data_preprocessor.destruct(feats, data_samples, 'img')
```

```
>>> # destruct model inputs for visualization
>>> for idx, data_sample in enumerate(data_samples):
>>>     destructed_input = self.data_preprocessor.destruct(
>>>         inputs[idx], data_sample, key='img')
>>>     data_sample.set_data({'input': destructed_input})
```

Parameters

- **outputs** (*Tensor*) – Tensor to destruct.
- **data_samples** (*Union[SampleList, DataSample], optional*) – Data samples (or data sample) corresponding to *outputs*. Defaults to None
- **key** (*str*) – The key of field in data sample. Defaults to 'img'.

Returns Destructed outputs.

Return type Union[list, Tensor]

_destruct_norm_and_conversion(*batch_tensor: torch.Tensor, data_samples: Union[mmagic.utils.typing.SampleList, mmagic.structures.DataSample, None], key: str*) → torch.Tensor

De-norm and de-convert channel order. Noted that, we de-norm first, and then de-conversion, since mean and std used in normalization is based on channel order after conversion.

Parameters

- **batch_tensor** (*Tensor*) – Tensor to destruct.
- **data_samples** (*Union[SampleList, DataSample], optional*) – Data samples (or data sample) corresponding to *outputs*.
- **key** (*str*) – The key of field in data sample.

Returns Destructed tensor.

Return type Tensor

_destruct_padding(*batch_tensor: torch.Tensor*, *data_samples: Union[mmagic.utils.typing.SampleList, mmagic.structures.DataSample, None]*, *same_padding: bool = True*) → Union[list, torch.Tensor]

Destruct padding of the input tensor.

Parameters

- **batch_tensor** (*Tensor*) – Tensor to destruct.
- **data_samples** (*Union[SampleList, DataSample], optional*) – Data samples (or data sample) corresponding to *outputs*. If
- **same_padding** (*bool*) – Whether all samples will un-padded with the padding info of the first sample, and return a stacked un-padded tensor. Otherwise each sample will be unpadded with padding info saved in corresponding data samples, and return a list of un-padded tensor, since each un-padded tensor may have the different shape. Defaults to True.

Returns Destructed outputs.

Return type Union[list, Tensor]

class mmagic.models.data_preprocessors.**MattorPreprocessor**(*mean: MEAN_STD_TYPE = [123.675, 116.28, 103.53]*, *std: MEAN_STD_TYPE = [58.395, 57.12, 57.375]*, *output_channel_order: str = 'RGB'*, *proc_trimap: str = 'rescale_to_zero_one'*, *stack_data_sample=True*)

Bases: mmagic.models.data_preprocessors.data_preprocessor.DataPreprocessor

DataPreprocessor for matting models.

See base class DataPreprocessor for detailed information.

Workflow as follow :

- Collate and move data to the target device.
- Convert inputs from bgr to rgb if the shape of input is (3, H, W).
- Normalize image with defined std and mean.
- Stack inputs to batch_inputs.

Parameters

- **mean** (*Sequence[float or int], float or int, optional*) – The pixel mean of image channels. Noted that normalization operation is performed *after channel order conversion*. If it is not specified, images will not be normalized. Defaults None.
- **std** (*Sequence[float or int], float or int, optional*) – The pixel standard deviation of image channels. Noted that normalization operation is performed *after channel order conversion*. If it is not specified, images will not be normalized. Defaults None.
- **proc_trimap** (*str*) – Methods to process gt tensors. Default: 'rescale_to_zero_one'. Available options are rescale_to_zero_one and as-is.
- **stack_data_sample** (*bool*) – Whether stack a list of data samples to one data sample. Only support with input data samples are *DataSamples*. Defaults to True.

_proc_batch_trimap(*batch_trimaps: torch.Tensor*)

_preprocess_data_sample(*data_samples: mmagic.utils.typing.SampleList*, *training: bool*) → list

Preprocess data samples. When *training* is True, fields belong to `self.data_keys` will be converted to `self.output_channel_order` and *divided by 255*. When *training* is False, fields belongs to `self.data_keys` will be attempted to convert to 'BGR' without normalization. The corresponding metainfo related to normalization, channel order conversion will be updated to data sample as well.

Parameters

- **data_samples** (*List[DataSample]*) – A list of data samples to preprocess.
- **training** (*bool*) – Whether in training mode.

Returns The list of processed data samples.

Return type list

forward(*data: Sequence[dict]*, *training: bool = False*) → Tuple[torch.Tensor, list]

Pre-process input images, trimaps, ground-truth as configured.

Parameters

- **data** (*Sequence[dict]*) – data sampled from dataloader.
- **training** (*bool*) – Whether to enable training time augmentation. Default: False.

Returns Batched inputs and list of data samples.

Return type Tuple[torch.Tensor, list]

1.91 mmagic.models.editors

1.91.1 Package Contents

Classes

AnimateDiff	Implementation of `AnimateDiff.
UNet3DConditionMotionModel	
AOTBlockNeck	Dilation backbone used in AOT-GAN model.
AOTEncoderDecoder	Encoder-Decoder used in AOT-GAN model.
AOTInpaintor	Inpaintor for AOT-GAN method.
IDLossModel	Face id loss model.
BasicVSR	BasicVSR model for video super-resolution.
BasicVSRNet	BasicVSR network structure for video super-resolution.
BasicVSRPlusPlusNet	BasicVSR++ network structure.
BigGAN	Implementation of `Large Scale GAN Training for High Fidelity Natural
CAIN	CAIN model for Video Interpolation.
CAINNet	CAIN network structure.
ControlStableDiffusion	Implementation of `ControlNet with Stable Diffusion.
CycleGAN	CycleGAN model for unpaired image-to-image translation.
DCGAN	Implementation of `Unsupervised Representation Learning with Deep

continues on next page

Table 2 – continued from previous page

DenoisingUnet	Denoising Unet. This network receives a diffused image `x_t` and
DeblurGanV2	Base class for all algorithmic models.
DeblurGanV2Discriminator	Defines the discriminator for DeblurGanv2 with the specified arguments..
DeblurGanV2Generator	Defines the generator for DeblurGanv2 with the specified arguments..
ContextualAttentionModule	Contexture attention module.
ContextualAttentionNeck	Neck with contextual attention module.
DeepFillDecoder	Decoder used in DeepFill model.
DeepFillEncoder	Encoder used in DeepFill model.
DeepFillRefiner	Refiner used in DeepFill model.
DeepFillv1Discriminators	Discriminators used in DeepFillv1 model.
DeepFillv1Inpaintor	Inpaintor for deepfillv1 method.
DeepFillEncoderDecoder	Two-stage encoder-decoder structure used in DeepFill model.
DIC	DIC model for Face Super-Resolution.
DICNet	DIC network structure for face super-resolution.
FeedbackBlock	Feedback Block of DIC.
FeedbackBlockCustom	Custom feedback block, will be used as the first feedback block.
FeedbackBlockHeatmapAttention	Feedback block with HeatmapAttention.
LightCNN	LightCNN discriminator with input size 128 x 128.
MaxFeature	Conv2d or Linear layer with max feature selector.
DIM	Deep Image Matting model.
ClipWrapper	Clip Models wrapper.
DiscoDiffusion	Disco Diffusion (DD) is a Google Colab Notebook which leverages an AI
DreamBooth	Implementation of `DreamBooth with Stable Diffusion.
EDSRNet	EDSR network structure.
EDVR	EDVR model for video super-resolution.
EDVRNet	EDVR network structure for video super-resolution.
EG3D	Implementation of `Efficient Geometry-aware 3D Generative Adversarial
ESRGAN	Enhanced SRGAN model for single image super-resolution.
RRDBNet	Networks consisting of Residual in Residual Dense Block, which is used
FastComposer	Class for Stable Diffusion. Refers to https://github.com/Stability-
FBADecoder	Decoder for FBA matting.
FBAResnetDilated	ResNet-based encoder for FBA image matting.
FLAVR	FLAVR model for video interpolation.
FLAVRNet	PyTorch implementation of FLAVR for video frame interpolation.
GCA	Guided Contextual Attention image matting model.
GGAN	Implementation of *Geometric GAN*.
GLEANStyleGANv2	GLEAN (using StyleGANv2) architecture for super-resolution.
GLDecoder	Decoder used in Global&Local model.
GLDilationNeck	Dilation Backbone used in Global&Local model.

continues on next page

Table 2 – continued from previous page

GLEncoder	Encoder used in Global&Local model.
GLEncoderDecoder	Encoder-Decoder used in Global&Local model.
AblatedDiffusionModel	Guided diffusion Model.
IconVSRNet	IconVSR network structure for video super-resolution.
DepthwiseIndexBlock	Depthwise index block.
HolisticIndexBlock	Holistic Index Block.
IndexedUpsample	Indexed upsample module.
IndexNet	IndexNet matting model.
IndexNetDecoder	Decoder for IndexNet.
IndexNetEncoder	Encoder for IndexNet.
InstColorization	Colorization InstColorization method.
LIIF	LIIF model for single image super-resolution.
MLPRefiner	Multilayer perceptrons (MLPs), refiner used in LIIF.
LSGAN	Implementation of *Least Squares Generative Adversarial Networks*.
MSPIEStyleGAN2	MS-PIE StyleGAN2.
PESinGAN	Positional Encoding in SinGAN.
NAFBaseline	The original version of Baseline model in "Simple Baseline for Image
NAFBaselineLocal	The original version of Baseline model in "Simple Baseline for Image
NAFNet	NAFNet.
NAFNetLocal	The original version of NAFNetLocal in "Simple Baseline for Image
MaskConvModule	Mask convolution module.
PartialConv2d	Implementation for partial convolution.
PConvDecoder	Decoder with partial conv.
PConvEncoder	Encoder with partial conv.
PConvEncoderDecoder	Encoder-Decoder with partial conv module.
PConvInpaintor	Inpaintor for Partial Convolution method.
ProgressiveGrowingGAN	Progressive Growing Unconditional GAN.
Pix2Pix	Pix2Pix model for paired image-to-image translation.
PlainDecoder	Simple decoder from Deep Image Matting.
PlainRefiner	Simple refiner from Deep Image Matting.
RDNNet	RDN model for single image super-resolution.
RealBasicVSR	RealBasicVSR model for real-world video super-resolution.
RealBasicVSRNet	RealBasicVSR network structure for real-world video super-resolution.
RealESRGAN	Real-ESRGAN model for single image super-resolution.
UNetDiscriminatorWithSpectralNorm	A U-Net discriminator with spectral normalization.
Restormer	Restormer A PyTorch impl of: `Restormer: Efficient Transformer for High-
SAGAN	Implementation of *Self-Attention Generative Adversarial Networks*.
SinGAN	SinGAN.
SRCNNNet	SRCNN network structure for image super resolution.
SRGAN	SRGAN model for single image super-resolution.
ModifiedVGG	A modified VGG discriminator with input size 128 x 128.
MSRResNet	Modified SRResNet.

continues on next page

Table 2 – continued from previous page

StableDiffusion	Class for Stable Diffusion. Refers to https://github.com/Stability-
StableDiffusionInpaint	Class for Stable Diffusion. Refers to https://github.com/Stability-
StableDiffusionXL	Class for Stable Diffusion XL. Refers to https://github.com/Stability-
StyleGAN1	Implementation of `A Style-Based Generator Architecture for Generative
StyleGAN2	Implementation of `Analyzing and Improving the Image Quality of
StyleGAN3	Implementation of *Alias-Free Generative Adversarial Networks*. # noqa.
StyleGAN3Generator	StyleGAN3 Generator.
SwinIRNet	SwinIR
TDAN	TDAN model for video super-resolution.
TDANNet	TDAN network structure for video super-resolution.
TextualInversion	Implementation of `An Image is Worth One Word: Personalizing Text-to-
TOFlowVFINet	PyTorch implementation of TOFlow for video frame interpolation.
TOFlowVSRNet	PyTorch implementation of TOFlow.
ToFResBlock	ResNet architecture.
LTE	Learnable Texture Extractor.
TTSR	TTSR model for Reference-based Image Super-Resolution.
SearchTransformer	Search texture reference by transformer.
TTSRDiscriminator	A discriminator for TTSR.
TTSRNet	TTSR network structure (main-net) for reference-based super-resolution.
ViCo	Implementation of `ViCo with Stable Diffusion.
WGANGP	Implementation of *Improved Training of Wasserstein GANs*.

class mmagic.models.editors.**AnimateDiff**(*vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: str = 'fp32'*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor=dict(type='DataPreprocessor')*, *motion_module_cfg: Optional[dict] = None*, *dream_booth_lora_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Implementation of `AnimateDiff.

<https://arxiv.org/abs/2307.04725>`_ (AnimateDiff).

Parameters

- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.

- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **lora_config** (*dict, optional*) – The config for LoRA finetuning. Defaults to None.
- **val_prompts** (*Union[str, List[str]], optional*) – The prompts for validation. Defaults to None.
- **class_prior_prompt** (*str, optional*) – The prompt for class prior loss.
- **num_class_images** (*int, optional*) – The number of images for class prior. Defaults to 3.
- **prior_loss_weight** (*float, optional*) – The weight for class prior loss. Defaults to 0.
- **fine_tune_text_encoder** (*bool, optional*) – Whether to fine-tune text encoder. Defaults to False.
- **dtype** (*str, optional*) – The dtype for the model. Defaults to 'fp16'.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise # noqa Defaults to 0.
- **tomesd_cfg** (*dict, optional*) – The config for TOMESD. Please refers to https://github.com/dbolya/tomesd and https://github.com/open-mmlab/mmagic/blob/main/mmagic/models/utils/tome_utils.py for detail. # noqa Defaults to None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Defaults to

  dict(type='DataPreprocessor').
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Defaults to None/

property device

Set device for the model.

set_xformers(*module: Optional[torch.nn.Module] = None*) → torch.nn.Module

Set xformers for the model.

Returns The model with xformers.

Return type nn.Module

set_tomesd() → torch.nn.Module

Set ToMe for the stable diffusion model.

Returns The model with ToMe.

Return type nn.Module

init_motion_module(*motion_module_cfg*)

init_dreambooth_lora(*dream_booth_lora_cfg*)

_encode_prompt(*prompt*, *device*, *num_videos_per_prompt*, *do_classifier_free_guidance*, *negative_prompt*)

Encodes the prompt into text encoder hidden states.

decode_latents(*latents*)

latents decoder.

prepare_extra_step_kwargs(*generator*, *eta*)

Prepare extra kwargs for the scheduler step, since not all schedulers have the same signature eta () is only used with the DDIMScheduler, it will be ignored for other schedulers.

check_inputs(*prompt*, *height*, *width*)

Check inputs.

Raise error if not correct

convert_lora(*state_dict*, *LORA_PREFIX_UNET='lora_unet'*, *LORA_PREFIX_TEXT_ENCODER='lora_te'*, *alpha=0.6*)

Convert lora for unet and text_encoder TODO: use this function to convert lora

Parameters

- **state_dict** (*_type_*) – _description_
- **LORA_PREFIX_UNET** (*str, optional*) –
- **'lora_unet'.** (*_description_. Defaults to*) –
- **LORA_PREFIX_TEXT_ENCODER** (*str, optional*) –
- **'lora_te'.** (*_description_. Defaults to*) –
- **alpha** (*float, optional*) – _description_. Defaults to 0.6.

Returns check each output type _type_: unet && text_encoder

Return type TODO

prepare_latents(*batch_size*, *num_channels_latents*, *video_length*, *height*, *width*, *dtype*, *device*, *generator*, *latents=None*)

Prepare latent variables.

prepare_model()

Prepare model for training.

Move model to target dtype and disable gradient for some models.

set_lora()

Set LORA for model.

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

infer(*prompt: Union[str, List[str]], video_length: Optional[int] = 16, height: Optional[int] = None, width: Optional[int] = None, num_inference_steps: int = 50, guidance_scale: float = 7.5, negative_prompt: Optional[Union[str, List[str]]] = None, num_videos_per_prompt: Optional[int] = 1, eta: float = 0.0, generator: Optional[Union[torch.Generator, List[torch.Generator]]] = None, latents: Optional[torch.FloatTensor] = None, return_type: Optional[str] = 'tensor', show_progress: bool = True, seed: Optional[int] = 1007*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str or List[str]*) – The prompt or prompts to guide the video generation.
- **video_length** (*int, Option*) – The number of frames of the generated video. Defaults to 16.
- **height** (*int, Optional*) – The height in pixels of the generated image. If not passed, the height will be *self.unet_sample_size * self.vae_scale_factor* Defaults to None.
- **width** (*int, Optional*) – The width in pixels of the generated image. If not passed, the width will be *self.unet_sample_size * self.vae_scale_factor* Defaults to None.
- **num_inference_steps** (*int*) – The number of denoising steps. More denoising steps usually lead to a higher quality video at the expense of slower inference. Defaults to 50.
- **guidance_scale** (*float*) – Guidance scale as defined in Classifier- Free Diffusion Guidance (https://arxiv.org/abs/2207.12598). Defaults to 7.5
- **negative_prompt** (*str or List[str], optional*) – The prompt or prompts not to guide the video generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than 1). Defaults to None.
- **num_videos_per_prompt** (*int*) – The number of videos to generate per prompt. Defaults to 1.
- **eta** (*float*) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to DDIMScheduler, will be ignored for others. Defaults to 0.0.
- **generator** (*torch.Generator, optional*) – A torch generator to make generation deterministic. Defaults to None.
- **latents** (*torch.FloatTensor, optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for video generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*. Defaults to None.
- **return_type** (*str*) – The return type of the inference results. Supported types are 'video', 'numpy', 'tensor'. If 'video' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

#TODO :returns: A dict containing the generated video :rtype: dict

abstract forward(*inputs: torch.Tensor*, *data_samples: Optional[list] = None*, *mode: str = 'tensor'*) → Union[Dict[str, torch.Tensor], list]

forward is not implemented now.

class mmagic.models.editors.**UNet3DConditionMotionModel**(*sample_size: Optional[int] = None*, *in_channels: int = 4*, *out_channels: int = 4*, *center_input_sample: bool = False*, *flip_sin_to_cos: bool = True*, *freq_shift: int = 0*, *down_block_types: Tuple[str] = ('CrossAttnDownBlock3D', 'CrossAttnDownBlock3D', 'CrossAttnDownBlock3D', 'DownBlock3D')*, *mid_block_type: str = 'UNetMidBlock3DCrossAttn'*, *up_block_types: Tuple[str] = ('UpBlock3D', 'CrossAttnUpBlock3D', 'CrossAttnUpBlock3D', 'CrossAttnUpBlock3D')*, *only_cross_attention: Union[bool, Tuple[bool]] = False*, *block_out_channels: Tuple[int] = (320, 640, 1280, 1280)*, *layers_per_block: int = 2*, *downsample_padding: int = 1*, *mid_block_scale_factor: float = 1*, *act_fn: str = 'silu'*, *norm_num_groups: int = 32*, *norm_eps: float = 1e-05*, *cross_attention_dim: int = 768*, *attention_head_dim: Union[int, Tuple[int]] = 8*, *dual_cross_attention: bool = False*, *use_linear_projection: bool = False*, *class_embed_type: Optional[str] = None*, *num_class_embeds: Optional[int] = None*, *upcast_attention: bool = False*, *resnet_time_scale_shift: str = 'default'*, *use_inflated_groupnorm=False*, *use_motion_module=False*, *motion_module_resolutions=(1, 2, 4, 8)*, *motion_module_mid_block=False*, *motion_module_decoder_only=False*, *motion_module_type=None*, *motion_module_kwargs={}*, *unet_use_cross_frame_attention=None*, *unet_use_temporal_attention=None*, *subfolder=None*, *from_pretrained=None*, *unet_addtion_kwargs=None*)

Bases: `diffusers.models.modeling_utils.ModelMixin`, `diffusers.configuration_utils.ConfigMixin`

_supports_gradient_checkpointing = True

Implementation of UNet3DConditionMotionModel

init_weights(*subfolder=None*, *from_pretrained=None*)

Init weights for models.

We just use the initialization method proposed in the original paper.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

set_attention_slice(*slice_size*)

Enable sliced attention computation.

When this option is enabled, the attention module will split the input tensor in slices, to compute attention in several steps. This is useful to save some memory in exchange for a small speed decrease.

Parameters **slice_size** (*str* or *int* or `list(int) – defaults to *"auto"*): When *"auto"*, halves the input to the attention heads, so attention will be computed in two steps. If *"max"*, maximum amount of memory will be saved by running only one slice at a time. If a number is provided, uses as many slices as *attention_head_dim // slice_size*. In this case, *attention_head_dim' must be a multiple of `slice_size*.

_set_gradient_checkpointing(*module*, *value=False*)

set gradient checkpoint.

forward(*sample: torch.FloatTensor*, *timestep: Union[torch.Tensor, float, int]*, *encoder_hidden_states: torch.Tensor*, *class_labels: Optional[torch.Tensor] = None*, *attention_mask: Optional[torch.Tensor] = None*, *return_dict: bool = True*) → Union[UNet3DConditionOutput, Tuple]

Parameters

- **sample** (*torch.FloatTensor*) – (batch, channel, height, width)
- **tensor** (*noisy inputs*) –
- **timestep** (*torch.FloatTensor* or *float* or *int*) –
- **timesteps** (*(batch)*) –
- **encoder_hidden_states** (*torch.FloatTensor*) –
- **(batch** –
- **sequence_length** –
- **states** (*feature_dim) encoder hidden*) –
- **return_dict** (*bool*, *optional*, defaults to *True*) – Whether or not to return a [*UNet3DConditionOutput*] instead of a plain tuple.

Returns [*UNet3DConditionOutput*] if *return_dict* is True, otherwise a *tuple*. When returning a tuple, the first element is the sample tensor.

Return type [*UNet3DConditionOutput*] or *tuple*

classmethod from_pretrained_2d(*pretrained_model_path*, *subfolder=None*, *unet_additional_kwargs=None*)

a class method for initialization.

class mmagic.models.editors.**AOTBlockNeck**(*in_channels=256*, *dilation_rates=(1, 2, 4, 8)*, *num_aotblock=8*, *act_cfg=dict(type='ReLU')*, ***kwargs*)

Bases: mmengine.model.BaseModule

Dilation backbone used in AOT-GAN model.

This implementation follows: Aggregated Contextual Transformations for High-Resolution Image Inpainting

Parameters

- **in_channels** (*int, optional*) – Channel number of input feature. Default: 256.
- **dilation_rates** (*Tuple[int], optional*) – The dilation rates used

- **Default** (*for AOT block.*) – (1, 2, 4, 8).
- **num_aotblock** (*int, optional*) – Number of AOT blocks. Default: 8.
- **act_cfg** (*dict, optional*) – Config dict for activation layer, "relu" by default.
- **kwargs** (*keyword arguments*) –

forward(*x*)

class mmagic.models.editors.**AOTEncoderDecoder**(*encoder=dict(type='AOTEncoder'), decoder=dict(type='AOTDecoder'), dilation_neck=dict(type='AOTBlockNeck')*)

Bases: mmagic.models.editors.global_local.GLEncoderDecoder

Encoder-Decoder used in AOT-GAN model.

This implementation follows: Aggregated Contextual Transformations for High-Resolution Image Inpainting The architecture of the encoder-decoder is: (conv2d x 3) –> (dilated conv2d x 8) –> (conv2d or deconv2d x 3).

Parameters

- **encoder** (*dict*) – Config dict to encoder.
- **decoder** (*dict*) – Config dict to build decoder.
- **dilation_neck** (*dict*) – Config dict to build dilation neck.

class mmagic.models.editors.**AOTInpaintor**(*data_preprocessor: Union[dict, mmengine.config.Config], encdec: dict, disc: Optional[dict] = None, loss_gan: Optional[dict] = None, loss_gp: Optional[dict] = None, loss_disc_shift: Optional[dict] = None, loss_composed_percep: Optional[dict] = None, loss_out_percep: bool = False, loss_l1_hole: Optional[dict] = None, loss_l1_valid: Optional[dict] = None, loss_tv: Optional[dict] = None, train_cfg: Optional[dict] = None, test_cfg: Optional[dict] = None, init_cfg: Optional[dict] = None*)

Bases: *mmagic.models.base_models.OneStageInpaintor*

Inpaintor for AOT-GAN method.

This inpaintor is implemented according to the paper: Aggregated Contextual Transformations for High-Resolution Image Inpainting

forward_train_d(*data_batch, is_real, is_disc, mask*)

Forward function in discriminator training step.

In this function, we compute the prediction for each data batch (real or fake). Meanwhile, the standard gan loss will be computed with several proposed losses for stable training.

Parameters

- **data_batch** (*torch.Tensor*) – Batch of real data or fake data.
- **is_real** (*bool*) – If True, the gan loss will regard this batch as real data. Otherwise, the gan loss will regard this batch as fake data.
- **is_disc** (*bool*) – If True, this function is called in discriminator training step. Otherwise, this function is called in generator training step. This will help us to compute different types of adversarial loss, like LSGAN.
- **mask** (*torch.Tensor*) – Mask of data.

Returns Contains the loss items computed in this function.

Return type dict

`generator_loss`(*fake_res*, *fake_img*, *gt*, *mask*, *masked_img*)

Forward function in generator training step.

In this function, we mainly compute the loss items for generator with the given (fake_res, fake_img). In general, the *fake_res* is the direct output of the generator and the *fake_img* is the composition of direct output and ground-truth image.

Parameters

- **fake_res** (*torch.Tensor*) – Direct output of the generator.
- **fake_img** (*torch.Tensor*) – Composition of *fake_res* and ground-truth image.
- **gt** (*torch.Tensor*) – Ground-truth image.
- **mask** (*torch.Tensor*) – Mask image.
- **masked_img** (*torch.Tensor*) – Composition of mask image and ground-truth image.

Returns

Dict contains the results computed within this function for visualization and dict contains the loss items computed in this function.

Return type tuple(dict)

`forward_tensor`(*inputs*, *data_samples*)

Forward function in tensor mode.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_samples** (*List[dict]*) – List of data sample dict.

Returns

Direct output of the generator and composition of *fake_res* and ground-truth image.

Return type tuple

`train_step`(*data: List[dict]*, *optim_wrapper*)

Train step function.

In this function, the inpaintor will finish the train step following the pipeline: 1. get fake res/image 2. compute reconstruction losses for generator 3. compute adversarial loss for discriminator 4. optimize generator 5. optimize discriminator

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns

Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

class mmagic.models.editors.**IDLossModel**(*ir_se50_weights=None*)

Bases: torch.nn.Module

Face id loss model.

Parameters **ir_se50_weights** (*str, optional*) – Url of ir-se50 weights. Defaults to None.

_ir_se50_url = https://download.openxlab.org.cn/models/rangoliu/Arcface-IR-SE50/weight/Arcface-IR-SE50

extract_feats(*x*)

Extracting face features.

Parameters **x** (*torch.Tensor*) – Image tensor of faces.

Returns Face features.

Return type torch.Tensor

forward(*pred=None*, *gt=None*)

Calculate face loss.

Parameters

- **pred** (*torch.Tensor, optional*) – Predictions of face images. Defaults to None.
- **gt** (*torch.Tensor, optional*) – Ground truth of face images. Defaults to None.

Returns

A tuple contain face similarity loss and improvement.

Return type Tuple(float, float)

class mmagic.models.editors.**BasicVSR**(*generator*, *pixel_loss*, *ensemble=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.BaseEditModel

BasicVSR model for video super-resolution.

Note that this model is used for IconVSR.

Paper: BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond, CVPR, 2021

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **ensemble** (*dict*) – Config for ensemble. Default: None.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.

check_if_mirror_extended(*lrs*)

Check whether the input is a mirror-extended sequence.

If mirror-extended, the i-th (i=0, …, t-1) frame is equal to the (t-1-i)-th frame.

Parameters **lrs** (*tensor*) – Input LR images with shape (n, t, c, h, w)

forward_train(*inputs*, *data_samples=None*, ***kwargs*)

Forward training. Returns dict of losses of training.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.

Returns Dict of losses.

Return type dict

forward_inference(*inputs*, *data_samples=None*, ***kwargs*)

Forward inference. Returns predictions of validation, testing.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.

Returns predictions.

Return type DataSample

class mmagic.models.editors.**BasicVSRNet**(*mid_channels=64*, *num_blocks=30*, *spynet_pretrained=None*)

Bases: mmengine.model.BaseModule

BasicVSR network structure for video super-resolution.

Support only x4 upsampling.

Paper: BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond, CVPR, 2021

Parameters

- **mid_channels** (*int*) – Channel number of the intermediate features. Default: 64.
- **num_blocks** (*int*) – Number of residual blocks in each propagation branch. Default: 30.
- **spynet_pretrained** (*str*) – Pre-trained model path of SPyNet. Default: None.

check_if_mirror_extended(*lrs*)

Check whether the input is a mirror-extended sequence.

If mirror-extended, the i-th (i=0, …, t-1) frame is equal to the (t-1-i)-th frame.

Parameters **lrs** (*tensor*) – Input LR images with shape (n, t, c, h, w)

compute_flow(*lrs*)

Compute optical flow using SPyNet for feature warping.

Note that if the input is an mirror-extended sequence, 'flows_forward' is not needed, since it is equal to 'flows_backward.flip(1)'.

Parameters **lrs** (*tensor*) – Input LR images with shape (n, t, c, h, w)

Returns

Optical flow. 'flows_forward' corresponds to the flows used for forward-time propagation (current to previous). 'flows_backward' corresponds to the flows used for backward-time propagation (current to next).

Return type tuple(Tensor)

`forward`(*lrs*)

Forward function for BasicVSR.

Parameters **lrs** (*Tensor*) – Input LR sequence with shape (n, t, c, h, w).

Returns Output HR sequence with shape (n, t, c, 4h, 4w).

Return type Tensor

class `mmagic.models.editors.`**`BasicVSRPlusPlusNet`**(*mid_channels=64*, *num_blocks=7*, *max_residue_magnitude=10*, *is_low_res_input=True*, *spynet_pretrained=None*, *cpu_cache_length=100*)

Bases: `mmengine.model.BaseModule`

BasicVSR++ network structure.

Support either x4 upsampling or same size output.

Paper: BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment

Parameters

- **mid_channels** (*int, optional*) – Channel number of the intermediate features. Default: 64.
- **num_blocks** (*int, optional*) – The number of residual blocks in each propagation branch. Default: 7.
- **max_residue_magnitude** (*int*) – The maximum magnitude of the offset residue (Eq. 6 in paper). Default: 10.
- **is_low_res_input** (*bool, optional*) – Whether the input is low-resolution or not. If False, the output resolution is equal to the input resolution. Default: True.
- **spynet_pretrained** (*str, optional*) – Pre-trained model path of SPyNet. Default: None.
- **cpu_cache_length** (*int, optional*) – When the length of sequence is larger than this value, the intermediate features are sent to CPU. This saves GPU memory, but slows down the inference speed. You can increase this number if you have a GPU with large memory. Default: 100.

`check_if_mirror_extended`(*lqs*)

Check whether the input is a mirror-extended sequence.

If mirror-extended, the i-th (i=0, …, t-1) frame is equal to the (t-1-i)-th frame.

Parameters **lqs** (*tensor*) – Input low quality (LQ) sequence with shape (n, t, c, h, w).

`compute_flow`(*lqs*)

Compute optical flow using SPyNet for feature alignment.

Note that if the input is an mirror-extended sequence, 'flows_forward' is not needed, since it is equal to 'flows_backward.flip(1)'.

Parameters **lqs** (*tensor*) – Input low quality (LQ) sequence with shape (n, t, c, h, w).

Returns

Optical flow. 'flows_forward' corresponds to the flows used for forward-time propagation (current to previous). 'flows_backward' corresponds to the flows used for backward-time propagation (current to next).

Return type tuple(Tensor)

`propagate`(*feats*, *flows*, *module_name*)

Propagate the latent features throughout the sequence.

Parameters

- **dict** (*feats*) – Features from previous branches. Each component is a list of tensors with shape (n, c, h, w).
- **flows** (*tensor*) – Optical flows with shape (n, t - 1, 2, h, w).
- **module_name** (*str*) – The name of the propagation branches. Can either be 'backward_1', 'forward_1', 'backward_2', 'forward_2'.

Returns

A dictionary containing all the propagated features. Each key in the dictionary corresponds to a propagation branch, which is represented by a list of tensors.

Return type dict(list[tensor])

`upsample`(*lqs*, *feats*)

Compute the output image given the features.

Parameters

- **lqs** (*tensor*) – Input low quality (LQ) sequence with shape (n, t, c, h, w).
- **feats** (*dict*) – The features from the propagation branches.

Returns Output HR sequence with shape (n, t, c, 4h, 4w).

Return type Tensor

`forward`(*lqs*)

Forward function for BasicVSR++.

Parameters **lqs** (*tensor*) – Input low quality (LQ) sequence with shape (n, t, c, h, w).

Returns Output HR sequence with shape (n, t, c, 4h, 4w).

Return type Tensor

`class mmagic.models.editors.BigGAN`(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *num_classes: Optional[int] = None*, *ema_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseConditionalGAN*

Implementation of Large Scale GAN Training for High Fidelity Natural Image Synthesis (BigGAN).

Detailed architecture can be found in `BigGANGenerator` and `BigGANDiscriminator`

Parameters

- **generator** (*ModelType*) – The config or model of the generator.

- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or DataPreprocessor.
- **generator_steps** (*int*) – Number of times the generator was completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – Number of times the discriminator was completely updated before the generator is updated. Defaults to 1.
- **noise_size** (*Optional[int]*) – Size of the input noise vector. Default to 128.
- **num_classes** (*Optional[int]*) – The number classes you would like to generate. Defaults to None.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple

Get disc loss. BigGAN use hinge loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake*)

Get disc loss. BigGAN use hinge loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.

- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.editors.**CAIN**(*generator: dict*, *pixel_loss: dict*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *required_frames: int = 2*, *step_frames: int = 1*, *init_cfg: Optional[dict] = None*, *data_preprocessor: Optional[dict] = None*)

Bases: *mmagic.models.base_models.BasicInterpolator*

CAIN model for Video Interpolation.

Paper: Channel Attention Is All You Need for Video Frame Interpolation Ref repo: https://github.com/myungsub/CAIN

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **required_frames** (*int*) – Required frames in each process. Default: 2
- **step_frames** (*int*) – Step size of video frame interpolation. Default: 1
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.

init_cfg

Initialization config dict.

Type dict, optional

data_preprocessor

Used for pre-processing data sampled by dataloader to the format accepted by forward().

Type BaseDataPreprocessor

forward_inference(*inputs*, *data_samples=None*)

Forward inference. Returns predictions of validation, testing, and simple inference.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns predictions.

Return type List[DataSample]

class mmagic.models.editors.**CAINNet**(*in_channels=3*, *kernel_size=3*, *num_block_groups=5*, *num_block_layers=12*, *depth=3*, *reduction=16*, *norm=None*, *padding=7*, *act=nn.LeakyReLU(0.2, True)*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

CAIN network structure.

Paper: Channel Attention Is All You Need for Video Frame Interpolation. Ref repo: https://github.com/myungsub/CAIN

Parameters

- **in_channels** (*int*) – Channel number of inputs. Default: 3.
- **kernel_size** (*int*) – Kernel size of CAINNet. Default: 3.
- **num_block_groups** (*int*) – Number of block groups. Default: 5.
- **num_block_layers** (*int*) – Number of blocks in a group. Default: 12.
- **depth** (*int*) – Down scale depth, scale = 2**depth. Default: 3.
- **reduction** (*int*) – Channel reduction of CA. Default: 16.
- **norm** (*str | None*) – Normalization layer. If it is None, no normalization is performed. Default: None.
- **padding** (*int*) – Padding of CAINNet. Default: 7.
- **act** (*function*) – activate function. Default: nn.LeakyReLU(0.2, True).
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

forward(*imgs*, *padding_flag=False*)

Forward function.

Parameters

- **imgs** (*Tensor*) – Input tensor with shape (n, 2, c, h, w).
- **padding_flag** (*bool*) – Padding or not. Default: False.

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**ControlStableDiffusion**(*vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *controlnet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: str = 'fp32'*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor=dict(type='DataPreprocessor')*, *init_cfg: Optional[dict] = None*, *attention_injection=False*)

Bases: mmagic.models.editors.stable_diffusion.StableDiffusion

Implementation of `ControlNet with Stable Diffusion.

<https://arxiv.org/abs/2302.05543>`_ (ControlNet).

Parameters

- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.

- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **controlnet** (*Union[dict, nn.Module]*) – The config or module for ControlNet.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model. Defaults to 'fp16'.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise # noqa Defaults to 0.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Defaults to

  dict(type='DataPreprocessor').
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Defaults to None/

init_weights()

Initialize the weights. Noted that this function will only be called at train. If you want to inference with a different unet model, you can call this function manually or use *mmagic.models.editors.controlnet.controlnet_utils.change_base_model* to convert the weight of ControlNet manually.

Example: >>> 1. init controlnet from unet >>> init_cfg = dict(type='init_from_unet')

```
>>> 2. switch controlnet weight from unet
>>> # base model is not defined, use `runwayml/stable-diffusion-v1-5`
>>> # as default
>>> init_cfg = dict(type='convert_from_unet')
>>> # base model is defined
>>> init_cfg = dict(
>>>     type='convert_from_unet',
>>>     base_model=dict(
>>>         type='UNet2DConditionModel',
>>>         from_pretrained='REPO_ID',
>>>         subfolder='unet'))
```

train_step(*data: dict, optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step for ControlNet model. :param data: Data sampled from dataloader. :type data: dict :param optim_wrapper: OptimWrapperDict instance

contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

static prepare_control(*image: Tuple[PIL.Image.Image, List[PIL.Image.Image], torch.Tensor, List[torch.Tensor]]*, *width: int*, *height: int*, *batch_size: int*, *num_images_per_prompt: int*, *device: str*, *dtype: str*) → torch.Tensor

A helper function to prepare single control images.

Parameters

- **image** (*Tuple[Image.Image, List[Image.Image], Tensor, List[Tensor]]*) – # noqa The input image for control.
- **batch_size** (*int*) – The number of the prompt. The control will be repeated for *batch_size* times.
- **num_images_per_prompt** (*int*) – The number images generate for one prompt.
- **device** (*str*) – The device of the control.
- **dtype** (*str*) – The dtype of the control.

Returns The control in torch.tensor.

Return type Tensor

train(*mode: bool = True*)

Set train/eval mode.

Parameters **mode** (*bool, optional*) – Whether set train mode. Defaults to True.

infer(*prompt: Union[str, List[str]]*, *height: Optional[int] = None*, *width: Optional[int] = None*, *control: Optional[Union[str, numpy.ndarray, torch.Tensor]] = None*, *controlnet_conditioning_scale: float = 1.0*, *num_inference_steps: int = 20*, *guidance_scale: float = 7.5*, *negative_prompt: Optional[Union[str, List[str]]] = None*, *num_images_per_prompt: Optional[int] = 1*, *eta: float = 0.0*, *generator: Optional[torch.Generator] = None*, *latents: Optional[torch.FloatTensor] = None*, *return_type='image'*, *show_progress=True*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str or List[str]*) – The prompt or prompts to guide the image generation.
- **height** (*int, Optional*) – The height in pixels of the generated image. If not passed, the height will be *self.unet_sample_size * self.vae_scale_factor* Defaults to None.

- **width** (*int, Optional*) – The width in pixels of the generated image. If not passed, the width will be *self.unet_sample_size * self.vae_scale_factor* Defaults to None.
- **num_inference_steps** (*int*) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference. Defaults to 50.
- **guidance_scale** (*float*) – Guidance scale as defined in Classifier- Free Diffusion Guidance (https://arxiv.org/abs/2207.12598). Defaults to 7.5
- **negative_prompt** (*str or List[str], optional*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than 1). Defaults to None.
- **num_images_per_prompt** (*int*) – The number of images to generate per prompt. Defaults to 1.
- **eta** (*float*) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to DDIMScheduler, will be ignored for others. Defaults to 0.0.
- **generator** (*torch.Generator, optional*) – A torch generator to make generation deterministic. Defaults to None.
- **latents** (*torch.FloatTensor, optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*. Defaults to None.
- **return_type** (*str*) – The return type of the inference results. Supported types are 'image', 'numpy', 'tensor'. If 'image' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

Returns A dict containing the generated images and Control image.

Return type dict

abstract forward(**args*, ***kwargs*)

forward is not implemented now.

class mmagic.models.editors.**CycleGAN**(**args*, *buffer_size=50*, *loss_config=dict(cycle_loss_weight=10.0, id_loss_weight=0.5)*, ***kwargs*)

Bases: *mmagic.models.base_models.BaseTranslationModel*

CycleGAN model for unpaired image-to-image translation.

Ref: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

forward_test(*img*, *target_domain*, ***kwargs*)

Forward function for testing.

Parameters

- **img** (*tensor*) – Input image tensor.
- **target_domain** (*str*) – Target domain of output image.
- **kwargs** (*dict*) – Other arguments.

Returns Forward results.

Return type dict

_get_disc_loss(*outputs*)

Backward function for the discriminators.

Parameters **outputs** (*dict*) – Dict of forward results.

Returns Discriminators' loss and loss dict.

Return type dict

_get_gen_loss(*outputs*)

Backward function for the generators.

Parameters **outputs** (*dict*) – Dict of forward results.

Returns Generators' loss and loss dict.

Return type dict

_get_opposite_domain(*domain*)

Get the opposite domain respect to the input domain.

Parameters **domain** (*str*) – The input domain.

Returns The opposite domain.

Return type str

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

Training step function.

Parameters

- **data_batch** (*dict*) – Dict of the input data batch.
- **optimizer** (*dict[torch.optim.Optimizer]*) – Dict of optimizers for the generators and discriminators.
- **ddp_reducer** (Reducer | None, optional) – Reducer from ddp. It is used to prepare for `backward()` in ddp. Defaults to None.
- **running_status** (*dict | None, optional*) – Contains necessary basic information for training, e.g., iteration number. Defaults to None.

Returns Dict of loss, information for logger, the number of samples and results for visualization.

Return type dict

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as *val_step()*.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as *val_step()*.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

class mmagic.models.editors.**DCGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseGAN*

Implementation of *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*.

Paper link: <https://arxiv.org/abs/1511.06434>`_ (DCGAN).

Detailed architecture can be found in `DCGANGenerator` # noqa and `DCGANDiscriminator` # noqa

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple

Get disc loss. DCGAN use the vanilla gan loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple

Get gen loss. DCGAN use the vanilla gan loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.

- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.editors.**DenoisingUnet**(*image_size*, *in_channels=3*, *out_channels=None*, *base_channels=128*, *resblocks_per_downsample=3*, *num_timesteps=1000*, *use_rescale_timesteps=False*, *dropout=0*, *embedding_channels=- 1*, *num_classes=0*, *use_fp16=False*, *channels_cfg=None*, *output_cfg=dict(mean='eps', var='learned_range')*, *norm_cfg=dict(type='GN', num_groups=32)*, *act_cfg=dict(type='SiLU', inplace=False)*, *shortcut_kernel_size=1*, *use_scale_shift_norm=False*, *resblock_updown=False*, *num_heads=4*, *time_embedding_mode='sin'*, *time_embedding_cfg=None*, *resblock_cfg=dict(type='DenoisingResBlock')*, *attention_cfg=dict(type='MultiHeadAttention')*, *encoder_channels=None*, *downsample_conv=True*, *upsample_conv=True*, *downsample_cfg=dict(type='DenoisingDownsample')*, *upsample_cfg=dict(type='DenoisingUpsample')*, *attention_res=[16, 8]*, *pretrained=None*, *unet_type=''*, *down_block_types: Tuple[str] = ()*, *up_block_types: Tuple[str] = ()*, *cross_attention_dim=768*, *layers_per_block: int = 2*)

Bases: mmengine.model.BaseModule

Denoising Unet. This network receives a diffused image x_t and current timestep t, and returns a output_dict corresponding to the passed output_cfg.

output_cfg defines the number of channels and the meaning of the output. output_cfg mainly contains keys of mean and var, denoting how the network outputs mean and variance required for the denoising process. For mean: 1. dict(mean='EPS'): Model will predict noise added in the

diffusion process, and the output_dict will contain a key named eps_t_pred.

2. **dict(mean='START_X'): Model will direct predict the mean of the** original image *x_0*, and the output_dict will contain a key named x_0_pred.
3. **dict(mean='X_TM1_PRED'): Model will predict the mean of diffused** image at *t-1* timestep, and the output_dict will contain a key named x_tm1_pred.

For var: 1. dict(var='FIXED_SMALL') or dict(var='FIXED_LARGE'): Variance in

the denoising process is regarded as a fixed value. Therefore only 'mean' will be predicted, and the output channels will equal to the input image (e.g., three channels for RGB image.)

2. **dict(var='LEARNED'): Model will predict *log_variance* in the** denoising process, and the output_dict will contain a key named log_var.
3. **dict(var='LEARNED_RANGE'): Model will predict an interpolation** factor and the *log_variance* will be calculated as *factor * upper_bound + (1-factor) * lower_bound*. The output_dict will contain a key named factor.

If `var` is not `FIXED_SMALL` or `FIXED_LARGE`, the number of output channels will be the double of input channels, where the first half part contains predicted mean values and the other part is the predicted variance values. Otherwise, the number of output channels equals to the input channels, only containing the predicted mean values.

Parameters

- **image_size** (*int | list[int]*) – The size of image to denoise.
- **in_channels** (*int, optional*) – The input channels of the input image. Defaults as `3`.
- **out_channels** (*int, optional*) – The output channels of the output prediction. Defaults as `None` for automaticaaly assigned by `var_mode`.
- **base_channels** (*int, optional*) – The basic channel number of the generator. The other layers contain channels based on this number. Defaults to `128`.
- **resblocks_per_downsample** (*int, optional*) – Number of ResBlock used between two downsample operations. The number of ResBlock between upsample operations will be the same value to keep symmetry. Defaults to 3.
- **num_timesteps** (*int, optional*) – The total timestep of the denoising process and the diffusion process. Defaults to `1000`.
- **use_rescale_timesteps** (*bool, optional*) – Whether rescale the input timesteps in range of [0, 1000]. Defaults to `True`.
- **dropout** (*float, optional*) – The probability of dropout operation of each ResBlock. Pass `0` to do not use dropout. Defaults as 0.
- **embedding_channels** (*int, optional*) – The output channels of time embedding layer and label embedding layer. If not passed (or passed `-1`), output channels of the embedding layers will set as four times of `base_channels`. Defaults to `-1`.
- **num_classes** (*int, optional*) – The number of conditional classes. If set to 0, this model will be degraded to an unconditional model. Defaults to 0.
- **channels_cfg** (*list | dict[list], optional*) – Config for input channels of the intermediate blocks. If list is passed, each element of the list indicates the scale factor for the input channels of the current block with regard to the `base_channels`. For block `i`, the input and output channels should be `channels_cfg[i] * base_channels` and `channels_cfg[i+1] * base_channels` If dict is provided, the key of the dict should be the output scale and corresponding value should be a list to define channels. Default: Please refer to `_default_channels_cfg`.
- **output_cfg** (*dict, optional*) – Config for output variables. Defaults to `dict(mean='eps', var='learned_range')`.
- **norm_cfg** (*dict, optional*) – The config for normalization layers. Defaults to `dict(type='GN', num_groups=32)`.
- **act_cfg** (*dict, optional*) – The config for activation layers. Defaults to `dict(type='SiLU', inplace=False)`.
- **shortcut_kernel_size** (*int, optional*) – The kernel size for shortcut conv in ResBlocks. The value of this argument will overwrite the default value of *resblock_cfg*. Defaults to *3*.
- **use_scale_shift_norm** (*bool, optional*) – Whether perform scale and shift after normalization operation. Defaults to True.
- **num_heads** (*int, optional*) – The number of attention heads. Defaults to 4.

- **time_embedding_mode** (*str, optional*) – Embedding method of `time_embedding`. Defaults to 'sin'.
- **time_embedding_cfg** (*dict, optional*) – Config for `time_embedding`. Defaults to None.
- **resblock_cfg** (*dict, optional*) – Config for ResBlock. Defaults to `dict(type='DenoisingResBlock')`.
- **attention_cfg** (*dict, optional*) – Config for attention operation. Defaults to `dict(type='MultiHeadAttention')`.
- **upsample_conv** (*bool, optional*) – Whether use conv in upsample block. Defaults to `True`.
- **downsample_conv** (*bool, optional*) – Whether use conv operation in downsample block. Defaults to `True`.
- **upsample_cfg** (*dict, optional*) – Config for upsample blocks. Defaults to `dict(type='DenoisingDownsample')`.
- **downsample_cfg** (*dict, optional*) – Config for downsample blocks. Defaults to `dict(type='DenoisingUpsample')`.
- **attention_res** (*int | list[int], optional*) – Resolution of feature maps to apply attention operation. Defaults to `[16, 8]`.
- **pretrained** (*str | dict, optional*) – Path for the pretrained model or dict containing information for pretrained models whose necessary key is 'ckpt_path'. Besides, you can also provide 'prefix' to load the generator part from the whole state dict. Defaults to None.

_default_channels_cfg

forward(*x_t*, *t*, *encoder_hidden_states=None*, *label=None*, *return_noise=False*)

Forward function. :param x_t: Diffused image at timestep *t* to denoise. :type x_t: torch.Tensor :param t: Current timestep. :type t: torch.Tensor :param label: You can directly give a

batch of label through a `torch.Tensor` or offer a callable function to sample a batch of label data. Otherwise, the `None` indicates to use the default label sampler.

Parameters **return_noise** (*bool, optional*) – If True, inputted `x_t` and `t` will be returned in a dict with output desired by `output_cfg`. Defaults to False.

Returns If not `return_noise`

Return type torch.Tensor | dict

init_weights(*pretrained=None*)

Init weights for models.

We just use the initialization method proposed in the original paper.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

convert_to_fp16()

Convert the precision of the model to float16.

convert_to_fp32()

Convert the precision of the model to float32.

class mmagic.models.editors.**DeblurGanV2**(*generator: ModelType, discriminator: Optional[ModelType] = None, pixel_loss: Optional[Union[dict, str]] = None, disc_loss: Optional[Union[dict, str]] = None, adv_lambda: float = 0.001, warmup_num: int = 3, train_cfg: Optional[dict] = None, test_cfg: Optional[dict] = None, init_cfg: Optional[dict] = None, data_preprocessor: Optional[dict] = None*)

Bases: `mmengine.model.BaseModel`

Base class for all algorithmic models.

BaseModel implements the basic functions of the algorithmic model, such as weights initialize, batch inputs preprocess(see more information in `BaseDataPreprocessor`), parse losses, and update model parameters.

Subclasses inherit from BaseModel only need to implement the forward method, which implements the logic to calculate loss and predictions, then can be trained in the runner.

Examples

```
>>> @MODELS.register_module()
>>> class ToyModel(BaseModel):
>>>
>>>     def __init__(self):
>>>         super().__init__()
>>>         self.backbone = nn.Sequential()
>>>         self.backbone.add_module('conv1', nn.Conv2d(3, 6, 5))
>>>         self.backbone.add_module('pool', nn.MaxPool2d(2, 2))
>>>         self.backbone.add_module('conv2', nn.Conv2d(6, 16, 5))
>>>         self.backbone.add_module('fc1', nn.Linear(16 * 5 * 5, 120))
>>>         self.backbone.add_module('fc2', nn.Linear(120, 84))
>>>         self.backbone.add_module('fc3', nn.Linear(84, 10))
>>>
>>>         self.criterion = nn.CrossEntropyLoss()
>>>
>>>     def forward(self, batch_inputs, data_samples, mode='tensor'):
>>>         data_samples = torch.stack(data_samples)
>>>         if mode == 'tensor':
>>>             return self.backbone(batch_inputs)
>>>         elif mode == 'predict':
>>>             feats = self.backbone(batch_inputs)
>>>             predictions = torch.argmax(feats, 1)
>>>             return predictions
>>>         elif mode == 'loss':
>>>             feats = self.backbone(batch_inputs)
>>>             loss = self.criterion(feats, data_samples)
>>>             return dict(loss=loss)
```

Parameters

- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`.

data_preprocessor

Used for pre-processing data sampled by dataloader to the format accepted by *forward()*.

Type BaseDataPreprocessor

init_cfg

Initialization config dict.

Type dict, optional

forward(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, *mode: str = 'tensor'*, ***kwargs*) → Union[torch.Tensor, List[*mmagic.structures.DataSample*], dict]

Returns losses or predictions of training, validation, testing, and simple inference process.

`forward` method of BaseModel is an abstract method, its subclasses must implement this method.

Accepts `inputs` and `data_samples` processed by *data_preprocessor*, and returns results according to mode arguments.

During non-distributed training, validation, and testing process, `forward` will be called by `BaseModel.train_step`, `BaseModel.val_step` and `BaseModel.val_step` directly.

During distributed data parallel training process, `MMSeparateDistributedDataParallel.train_step` will first call `DistributedDataParallel.forward` to enable automatic gradient synchronization, and then call `forward` to get training loss.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.
- **mode** (*str*) – mode should be one of `loss`, `predict` and `tensor`. Default: 'tensor'.
  - `loss`: Called by `train_step` and return loss `dict` used for logging
  - `predict`: Called by `val_step` and `test_step` and return list of `BaseDataElement` results used for computing metric.
  - `tensor`: Called by custom use to get `Tensor` type results.

Returns

- If `mode == loss`, return a `dict` of loss tensor used for backward and logging.
- If `mode == val`, return a `list` of `BaseDataElement` for computing metric and getting inference result.
- If `mode == predict`, return a `list` of `BaseDataElement` for computing metric and getting inference result.
- If `mode == tensor`, return a tensor or `tuple` of tensor or `dict` or tensor for custom use.

Return type ForwardResults

convert_to_datasample(*predictions:* mmagic.structures.DataSample, *data_samples:* mmagic.structures.DataSample, *inputs: Optional[torch.Tensor]*) → List[*mmagic.structures.DataSample*]

Add predictions and destructed inputs (if passed) to data samples.

Parameters

- **predictions** (*DataSample*) – The predictions of the model.

- **data_samples** (*DataSample*) – The data samples loaded from dataloader.
- **inputs** (*Optional[torch.Tensor]*) – The input of model. Defaults to None.

Returns Modified data samples.

Return type List[DataSample]

forward_tensor(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, ***kwargs*) → torch.Tensor

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns result of simple forward.

Return type Tensor

forward_inference(*inputs: torch.Tensor*, *data_samples: Optional[List[*mmagic.structures.DataSample*]] = None*, ***kwargs*) → List[*mmagic.structures.DataSample*]

Forward inference. Returns predictions of validation, testing, and simple inference.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns predictions.

Return type List[EditDataSample]

forward_train(*inputs*, *data_samples=None*, ***kwargs*)

Forward training. Losses of training is calculated in train_step.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by *data_preprocessor*.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by *data_preprocessor*.

Returns Result of `forward_tensor` with `training=True`.

Return type Tensor

val_step(*data: Union[tuple, dict, list]*) → list

Gets the predictions of given data.

Calls `self.data_preprocessor(data, False)` and `self(inputs, data_sample, mode='predict')` in order. Return the predictions which will be passed to evaluator.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns The predictions of given data.

Return type list

test_step(*data: Union[dict, tuple, list]*) → list

BaseModel implements test_step the same as val_step.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns The predictions of given data.

Return type list

_run_forward(*data: Union[dict, tuple, list]*, *mode: str*) → Union[Dict[str, torch.Tensor], list]

Unpacks data for *forward()*

Parameters

- **data** (*dict or tuple or list*) – Data sampled from dataset.
- **mode** (*str*) – Mode of forward.

Returns Results of training or testing mode.

Return type dict or list

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step of GAN-based method.

Parameters

- **data** (*List[dict]*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, torch.Tensor]

g_step(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*)

G step of DobuleGAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

d_step(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*)

D step of DobuleGAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

g_step_with_optim(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

G step with optim of GAN: Calculate losses of generator and run optim.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.
- **optim_wrapper** (*OptimWrapperDict*) – Optim wrapper dict.

Returns Dict of parsed losses.

Return type dict

d_step_with_optim(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

D step with optim of GAN: Calculate losses of discriminator and run optim.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.
- **optim_wrapper** (*OptimWrapperDict*) – Optim wrapper dict.

Returns Dict of parsed losses.

Return type dict

extract_gt_data(*data_samples*)

extract gt data from data samples.

Parameters **data_samples** (*list*) – List of DataSample.

Returns Extract gt data.

Return type Tensor

class mmagic.models.editors.**DeblurGanV2Discriminator**

Defines the discriminator for DeblurGanv2 with the specified arguments..

Parameters **model** (*Str*) – Type of the discriminator model

class mmagic.models.editors.**DeblurGanV2Generator**

Defines the generator for DeblurGanv2 with the specified arguments..

Parameters **model** (*Str*) – Type of the generator model

class mmagic.models.editors.**ContextualAttentionModule**(*unfold_raw_kernel_size=4*, *unfold_raw_stride=2*, *unfold_raw_padding=1*, *unfold_corr_kernel_size=3*, *unfold_corr_stride=1*, *unfold_corr_dilation=1*, *unfold_corr_padding=1*, *scale=0.5*, *fuse_kernel_size=3*, *softmax_scale=10*, *return_attention_score=True*)

Bases: mmengine.model.BaseModule

Contexture attention module.

The details of this module can be found in: Generative Image Inpainting with Contextual Attention

Parameters

- **unfold_raw_kernel_size** (*int*) – Kernel size used in unfolding raw feature. Default: 4.
- **unfold_raw_stride** (*int*) – Stride used in unfolding raw feature. Default: 2.
- **unfold_raw_padding** (*int*) – Padding used in unfolding raw feature. Default: 1.
- **unfold_corr_kernel_size** (*int*) – Kernel size used in unfolding context for computing correlation maps. Default: 3.
- **unfold_corr_stride** (*int*) – Stride used in unfolding context for computing correlation maps. Default: 1.
- **unfold_corr_dilation** (*int*) – Dilation used in unfolding context for computing correlation maps. Default: 1.
- **unfold_corr_padding** (*int*) – Padding used in unfolding context for computing correlation maps. Default: 1.
- **scale** (*float*) – The resale factor used in resize input features. Default: 0.5.
- **fuse_kernel_size** (*int*) – The kernel size used in fusion module. Default: 3.
- **softmax_scale** (*float*) – The scale factor for softmax function. Default: 10.
- **return_attention_score** (*bool*) – If True, the attention score will be returned. Default: True.

forward(*x*, *context*, *mask=None*)

Forward Function.

Parameters

- **x** (*torch.Tensor*) – Tensor with shape (n, c, h, w).
- **context** (*torch.Tensor*) – Tensor with shape (n, c, h, w).
- **mask** (*torch.Tensor*) – Tensor with shape (n, 1, h, w). Default: None.

Returns Features after contextural attention.

Return type tuple(torch.Tensor)

patch_correlation(*x*, *kernel*)

Calculate patch correlation.

Parameters

- **x** (*torch.Tensor*) – Input tensor.
- **kernel** (*torch.Tensor*) – Kernel tensor.

Returns Tensor with shape of (n, l, h, w).

Return type torch.Tensor

patch_copy_deconv(*attention_score*, *context_filter*)

Copy patches using deconv.

Parameters

- **attention_score** (*torch.Tensor*) – Tensor with shape of (n, l , h, w).
- **context_filter** (*torch.Tensor*) – Filter kernel.

Returns Tensor with shape of (n, c, h, w).

Return type torch.Tensor

fuse_correlation_map(*correlation_map*, *h_unfold*, *w_unfold*)

Fuse correlation map.

This operation is to fuse correlation map for increasing large consistent correlation regions.

The mechanism behind this op is simple and easy to understand. A standard 'Eye' matrix will be applied as a filter on the correlation map in horizontal and vertical direction.

The shape of input correlation map is (n, h_unfold*w_unfold, h, w). When adopting fusing, we will apply convolutional filter in the reshaped feature map with shape of (n, 1, h_unfold*w_fold, h*w).

A simple specification for horizontal direction is shown below:

```
       (h, (h, (h, (h,
        0)  1)  2)  3)  ...
(h, 0)
(h, 1)      1
(h, 2)          1
(h, 3)              1
...
```

calculate_unfold_hw(*input_size*, *kernel_size=3*, *stride=1*, *dilation=1*, *padding=0*)

Calculate (h, w) after unfolding.

The official implementation of *unfold* in pytorch will put the dimension (h, w) into *L*. Thus, this function is just to calculate the (h, w) according to the equation in: https://pytorch.org/docs/stable/nn.html#torch.nn.Unfold

calculate_overlap_factor(*attention_score*)

Calculate the overlap factor after applying deconv.

Parameters **attention_score** (*torch.Tensor*) – The attention score with shape of (n, c, h, w).

Returns The overlap factor will be returned.

Return type torch.Tensor

mask_correlation_map(*correlation_map*, *mask*)

Add mask weight for correlation map.

Add a negative infinity number to the masked regions so that softmax function will result in 'zero' in those regions.

Parameters

- **correlation_map** (*torch.Tensor*) – Correlation map with shape of (n, h_unfold*w_unfold, h_map, w_map).
- **mask** (*torch.Tensor*) – Mask tensor with shape of (n, c, h, w). '1' in the mask indicates masked region while '0' indicates valid region.

Returns Updated correlation map with mask.

Return type torch.Tensor

im2col(*img*, *kernel_size*, *stride=1*, *padding=0*, *dilation=1*, *normalize=False*, *return_cols=False*)

Reshape image-style feature to columns.

This function is used for unfold feature maps to columns. The details of this function can be found in: https://pytorch.org/docs/1.1.0/nn.html?highlight=unfold#torch.nn.Unfold

Parameters

- **img** (*torch.Tensor*) – Features to be unfolded. The shape of this feature should be (n, c, h, w).
- **kernel_size** (*int*) – In this function, we only support square kernel with same height and width.
- **stride** (*int*) – Stride number in unfolding. Default: 1.
- **padding** (*int*) – Padding number in unfolding. Default: 0.
- **dilation** (*int*) – Dilation number in unfolding. Default: 1.
- **normalize** (*bool*) – If True, the unfolded feature will be normalized. Default: False.
- **return_cols** (*bool*) – The official implementation in PyTorch of unfolding will return features with shape of (n, c*$prod{kernel_size}$, L). If True, the features will be reshaped to (n, L, c, kernel_size, kernel_size). Otherwise, the results will maintain the shape as the official implementation.

Returns Unfolded columns. If *return_cols* is True, the shape of output tensor is *(n, L, c, kernel_size, kernel_size)*. Otherwise, the shape will be *(n, c*$prod{kernel_size}$, L)*.

Return type torch.Tensor

class mmagic.models.editors.**ContextualAttentionNeck**(*in_channels*, *conv_type='conv'*, *conv_cfg=None*, *norm_cfg=None*, *act_cfg=dict(type='ELU')*, *contextual_attention_args=dict(softmax_scale=10.0)*, ***kwargs*)

Bases: mmengine.model.BaseModule

Neck with contextual attention module.

Parameters

- **in_channels** (*int*) – The number of input channels.
- **conv_type** (*str*) – The type of conv module. In DeepFillv1 model, the *conv_type* should be 'conv'. In DeepFillv2 model, the *conv_type* should be 'gated_conv'.
- **conv_cfg** (*dict | None*) – Config of conv module. Default: None.
- **norm_cfg** (*dict | None*) – Config of norm module. Default: None.
- **act_cfg** (*dict | None*) – Config of activation layer. Default: dict(type='ELU').
- **contextual_attention_args** (*dict*) – Config of contextual attention module. Default: dict(softmax_scale=10.).
- **kwargs** (*keyword arguments*) –

_conv_type

forward(*x*, *mask*)

Forward Function.

Parameters

- **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).
- **mask** (*torch.Tensor*) – Input tensor with shape of (n, 1, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**DeepFillDecoder**(*in_channels*, *conv_type='conv'*, *norm_cfg=None*, *act_cfg=dict(type='ELU')*, *out_act_cfg=dict(type='clip', min=- 1.0, max=1.0)*, *channel_factor=1.0*, ***kwargs*)

Bases: mmengine.model.BaseModule

Decoder used in DeepFill model.

This implementation follows: Generative Image Inpainting with Contextual Attention

Parameters

- **in_channels** (*int*) – The number of input channels.
- **conv_type** (*str*) – The type of conv module. In DeepFillv1 model, the *conv_type* should be 'conv'. In DeepFillv2 model, the *conv_type* should be 'gated_conv'.
- **norm_cfg** (*dict*) – Config dict to build norm layer. Default: None.
- **act_cfg** (*dict*) – Config dict for activation layer, "elu" by default.
- **out_act_cfg** (*dict*) – Config dict for output activation layer. Here, we provide commonly used *clamp* or *clip* operation.
- **channel_factor** (*float*) – The scale factor for channel size. Default: 1.
- **kwargs** (*keyword arguments*) –

_conv_type

forward(*input_dict*)

Forward Function.

Parameters **input_dict** (*dict | torch.Tensor*) – Input dict with middle features or torch.Tensor.

Returns Output tensor with shape of (n, c, h, w).

Return type torch.Tensor

class mmagic.models.editors.**DeepFillEncoder**(*in_channels=5*, *conv_type='conv'*, *norm_cfg=None*, *act_cfg=dict(type='ELU')*, *encoder_type='stage1'*, *channel_factor=1.0*, ***kwargs*)

Bases: mmengine.model.BaseModule

Encoder used in DeepFill model.

This implementation follows: Generative Image Inpainting with Contextual Attention

Parameters

- **in_channels** (*int*) – The number of input channels. Default: 5.
- **conv_type** (*str*) – The type of conv module. In DeepFillv1 model, the *conv_type* should be 'conv'. In DeepFillv2 model, the *conv_type* should be 'gated_conv'.
- **norm_cfg** (*dict*) – Config dict to build norm layer. Default: None.
- **act_cfg** (*dict*) – Config dict for activation layer, "elu" by default.
- **encoder_type** (*str*) – Type of the encoder. Should be one of ['stage1', 'stage2_conv', 'stage2_attention']. Default: 'stage1'.
- **channel_factor** (*float*) – The scale factor for channel size. Default: 1.

- **kwargs** (*keyword arguments*) –

_conv_type

forward(*x*)

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**DeepFillRefiner**(*encoder_attention=dict(type='DeepFillEncoder', encoder_type='stage2_attention'), encoder_conv=dict(type='DeepFillEncoder', encoder_type='stage2_conv'), dilation_neck=dict(type='GLDilationNeck', in_channels=128, act_cfg=dict(type='ELU')), contextual_attention=dict(type='ContextualAttentionNeck', in_channels=128), decoder=dict(type='DeepFillDecoder', in_channels=256)*)

Bases: mmengine.model.BaseModule

Refiner used in DeepFill model.

This implementation follows: Generative Image Inpainting with Contextual Attention.

Parameters

- **encoder_attention** (*dict*) – Config dict for encoder used in branch with contextual attention module.
- **encoder_conv** (*dict*) – Config dict for encoder used in branch with just convolutional operation.
- **dilation_neck** (*dict*) – Config dict for dilation neck in branch with just convolutional operation.
- **contextual_attention** (*dict*) – Config dict for contextual attention neck.
- **decoder** (*dict*) – Config dict for decoder used to fuse and decode features.

forward(*x*, *mask*)

Forward Function.

Parameters

- **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).
- **mask** (*torch.Tensor*) – Input tensor with shape of (n, 1, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**DeepFillv1Discriminators**(*global_disc_cfg*, *local_disc_cfg*)

Bases: mmengine.model.BaseModule

Discriminators used in DeepFillv1 model.

In DeepFillv1 model, the discriminators are independent without any concatenation like Global&Local model. Thus, we call this model *DeepFillv1Discriminators*. There exist a global discriminator and a local discriminator with global and local input respectively.

The details can be found in: Generative Image Inpainting with Contextual Attention.

Parameters

- **global_disc_cfg** (*dict*) – Config dict for global discriminator.
- **local_disc_cfg** (*dict*) – Config dict for local discriminator.

forward(*x*)

Forward function.

Parameters **x** (*tuple[torch.Tensor]*) – Contains global image and the local image patch.

Returns Contains the prediction from discriminators in global image and local image patch.

Return type tuple[torch.Tensor]

init_weights()

Init weights for models.

class mmagic.models.editors.**DeepFillv1Inpaintor**(*data_preprocessor: dict*, *encdec: dict*, *disc=None*, *loss_gan=None*, *loss_gp=None*, *loss_disc_shift=None*, *loss_composed_percep=None*, *loss_out_percep=False*, *loss_l1_hole=None*, *loss_l1_valid=None*, *loss_tv=None*, *stage1_loss_type=None*, *stage2_loss_type=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg: Optional[dict] = None*)

Bases: *mmagic.models.base_models.TwoStageInpaintor*

Inpaintor for deepfillv1 method.

This inpaintor is implemented according to the paper: Generative image inpainting with contextual attention

Importantly, this inpaintor is an example for using custom training schedule based on *TwoStageInpaintor*.

The training pipeline of deepfillv1 is as following:

```
if cur_iter < iter_tc:
    update generator with only l1 loss
else:
    update discriminator
    if cur_iter > iter_td:
        update generator with l1 loss and adversarial loss
```

The new attribute *cur_iter* is added for recording current number of iteration. The *train_cfg* contains the setting of the training schedule:

```
train_cfg = dict(
    start_iter=0,
    disc_step=1,
    iter_tc=90000,
    iter_td=100000
)
```

iter_tc and *iter_td* correspond to the notation T_C and T_D of the original paper.

Parameters

- **generator** (*dict*) – Config for encoder-decoder style generator.
- **disc** (*dict*) – Config for discriminator.

- **loss_gan** (*dict*) – Config for adversarial loss.
- **loss_gp** (*dict*) – Config for gradient penalty loss.
- **loss_disc_shift** (*dict*) – Config for discriminator shift loss.
- **loss_composed_percep** (*dict*) – Config for perceptual and style loss with composed image as input.
- **loss_out_percep** (*dict*) – Config for perceptual and style loss with direct output as input.
- **loss_l1_hole** (*dict*) – Config for l1 loss in the hole.
- **loss_l1_valid** (*dict*) – Config for l1 loss in the valid region.
- **loss_tv** (*dict*) – Config for total variation loss.
- **train_cfg** (*dict*) – Configs for training scheduler. *disc_step* must be contained for indicates the discriminator updating steps in each training step.
- **test_cfg** (*dict*) – Configs for testing scheduler.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward_train_d(*data_batch*, *is_real*, *is_disc*)

Forward function in discriminator training step.

In this function, we modify the default implementation with only one discriminator. In DeepFillv1 model, they use two separated discriminators for global and local consistency.

Parameters

- **data_batch** (*torch.Tensor*) – Batch of real data or fake data.
- **is_real** (*bool*) – If True, the gan loss will regard this batch as real data. Otherwise, the gan loss will regard this batch as fake data.
- **is_disc** (*bool*) – If True, this function is called in discriminator training step. Otherwise, this function is called in generator training step. This will help us to compute different types of adversarial loss, like LSGAN.

Returns Contains the loss items computed in this function.

Return type dict

two_stage_loss(*stage1_data*, *stage2_data*, *gt*, *mask*, *masked_img*)

Calculate two-stage loss.

Parameters

- **stage1_data** (*dict*) – Contain stage1 results.
- **stage2_data** (*dict*) – Contain stage2 results.
- **gt** (*torch.Tensor*) – Ground-truth image.
- **mask** (*torch.Tensor*) – Mask image.
- **masked_img** (*torch.Tensor*) – Composition of mask image and ground-truth image.

Returns Dict contains the results computed within this function for visualization and dict contains the loss items computed in this function.

Return type tuple(dict)

calculate_loss_with_type(*loss_type*, *fake_res*, *fake_img*, *gt*, *mask*, *prefix='stage1_'*, *fake_local=None*)

Calculate multiple types of losses.

Parameters

- **loss_type** (*str*) – Type of the loss.
- **fake_res** (*torch.Tensor*) – Direct results from model.
- **fake_img** (*torch.Tensor*) – Composited results from model.
- **gt** (*torch.Tensor*) – Ground-truth tensor.
- **mask** (*torch.Tensor*) – Mask tensor.
- **prefix** (*str, optional*) – Prefix for loss name. Defaults to 'stage1_'. # noqa
- **fake_local** (*torch.Tensor, optional*) – Local results from model. Defaults to None.

Returns Contain loss value with its name.

Return type dict

train_step(*data: List[dict]*, *optim_wrapper*)

Train step function.

In this function, the inpaintor will finish the train step following the pipeline:

1. get fake res/image
2. optimize discriminator (if have)
3. optimize generator

If *self.train_cfg.disc_step > 1*, the train step will contain multiple iterations for optimizing discriminator with different input data and only one iteration for optimizing generator after *disc_step* iterations for discriminator.

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

class mmagic.models.editors.**DeepFillEncoderDecoder**(*stage1=dict(type='GLEncoderDecoder', encoder=dict(type='DeepFillEncoder'), decoder=dict(type='DeepFillDecoder', in_channels=128), dilation_neck=dict(type='GLDilationNeck', in_channels=128, act_cfg=dict(type='ELU'))), stage2=dict(type='DeepFillRefiner'), return_offset=False*)

Bases: mmengine.model.BaseModule

Two-stage encoder-decoder structure used in DeepFill model.

The details are in: Generative Image Inpainting with Contextual Attention

Parameters

- **stage1** (*dict*) – Config dict for building stage1 model. As DeepFill model uses Global&Local model as baseline in first stage, the stage1 model can be easily built with *GLEncoderDecoder*.
- **stage2** (*dict*) – Config dict for building stage2 model.
- **return_offset** (*bool*) – Whether to return offset feature in contextual attention module. Default: False.

forward(*x*)

Forward function.

Parameters **x** (*torch.Tensor*) – This input tensor has the shape of (n, 5, h, w). In channel dimension, we concatenate [masked_img, ones, mask] as DeepFillv1 models do.

Returns The first two item is the results from first and second stage. If set *return_offset* as True, the offset will be returned as the third item.

Return type tuple[torch.Tensor]

init_weights()

Init weights for models.

class mmagic.models.editors.**DIC**(*generator*, *pixel_loss*, *align_loss*, *discriminator=None*, *gan_loss=None*, *feature_loss=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.editors.srgan.SRGAN

DIC model for Face Super-Resolution.

Paper: Deep Face Super-Resolution with Iterative Collaboration between Attentive Recovery and Landmark Estimation.

Parameters

- **generator** (*dict*) – Config for the generator.
- **pixel_loss** (*dict*) – Config for the pixel loss.
- **align_loss** (*dict*) – Config for the align loss.
- **discriminator** (*dict*) – Config for the discriminator. Default: None.
- **gan_loss** (*dict*) – Config for the gan loss. Default: None.
- **feature_loss** (*dict*) – Config for the feature loss. Default: None.
- **train_cfg** (*dict*) – Config for train. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor. Default: None.

forward_tensor(*inputs*, *data_samples=None*, *training=False*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.

- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.
- **training** (*bool*) – Whether is training. Default: False.

Returns

results of forward inference and forward train.

Return type (Tensor | Tuple[List[Tensor]])

if_run_g()

Calculates whether need to run the generator step.

if_run_d()

Calculates whether need to run the discriminator step.

g_step(*batch_outputs*, *batch_gt_data*)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step of GAN-based method.

Parameters

- **data** (*List[dict]*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, torch.Tensor]

static extract_gt_data(*data_samples*)

extract gt data from data samples.

Parameters **data_samples** (*list*) – List of DataSample.

Returns Extract gt data.

Return type Tensor

class mmagic.models.editors.**DICNet**(*in_channels*, *out_channels*, *mid_channels*, *num_blocks=6*, *hg_mid_channels=256*, *hg_num_keypoints=68*, *num_steps=4*, *upscale_factor=8*, *detach_attention=False*, *prelu_init=0.2*, *num_heatmaps=5*, *num_fusion_blocks=7*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

DIC network structure for face super-resolution.

Paper: Deep Face Super-Resolution with Iterative Collaboration between Attentive Recovery and Landmark Estimation

Parameters

- **in_channels** (*int*) – Number of channels in the input image
- **out_channels** (*int*) – Number of channels in the output image
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64
- **num_blocks** (*tuple[int]*) – Block numbers in the trunk network. Default: 6
- **hg_mid_channels** (*int*) – Channel number of intermediate features of HourGlass. Default: 256
- **hg_num_keypoints** (*int*) – Keypoint number of HourGlass. Default: 68
- **num_steps** (*int*) – Number of iterative steps. Default: 4
- **upscale_factor** (*int*) – Upsampling factor. Default: 8
- **detach_attention** (*bool*) – Detached from the current tensor for heatmap or not.
- **prelu_init** (*float*) – *init* of PReLU. Default: 0.2
- **num_heatmaps** (*int*) – Number of heatmaps. Default: 5
- **num_fusion_blocks** (*int*) – Number of fusion blocks. Default: 7
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor.

Returns Forward results. sr_outputs (list[Tensor]): forward sr results. heatmap_outputs (list[Tensor]): forward heatmap results.

Return type Tensor

class mmagic.models.editors.**FeedbackBlock**(*mid_channels*, *num_blocks*, *upscale_factor*, *padding=2*, *prelu_init=0.2*)

Bases: torch.nn.Module

Feedback Block of DIC.

It has a style of:

```
----- Module ----->
  ^            |
  |____________|
```

Parameters

- **mid_channels** (*int*) – Number of channels in the intermediate features.
- **num_blocks** (*int*) – Number of blocks.
- **upscale_factor** (*int*) – upscale factor.
- **padding** (*int*) – Padding size. Default: 2.
- **prelu_init** (*float*) – *init* of PReLU. Default: 0.2

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**FeedbackBlockCustom**(*in_channels*, *mid_channels*, *num_blocks*, *upscale_factor*)

Bases: *FeedbackBlock*

Custom feedback block, will be used as the first feedback block.

Parameters

- **in_channels** (*int*) – Number of channels in the input features.
- **mid_channels** (*int*) – Number of channels in the intermediate features.
- **num_blocks** (*int*) – Number of blocks.
- **upscale_factor** (*int*) – upscale factor.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**FeedbackBlockHeatmapAttention**(*mid_channels*, *num_blocks*, *upscale_factor*, *num_heatmaps*, *num_fusion_blocks*, *padding=2*, *prelu_init=0.2*)

Bases: *FeedbackBlock*

Feedback block with HeatmapAttention.

Parameters

- **in_channels** (*int*) – Number of channels in the input features.
- **mid_channels** (*int*) – Number of channels in the intermediate features.
- **num_blocks** (*int*) – Number of blocks.
- **upscale_factor** (*int*) – upscale factor.
- **padding** (*int*) – Padding size. Default: 2.
- **prelu_init** (*float*) – *init* of PReLU. Default: 0.2

forward(*x*, *heatmap*)

Forward function.

Parameters

- **x** (*Tensor*) – Input feature tensor.
- **heatmap** (*Tensor*) – Input heatmap tensor.

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**LightCNN**(*in_channels*)

Bases: mmengine.model.BaseModule

LightCNN discriminator with input size 128 x 128.

It is used to train DICGAN.

Parameters **in_channels** (*int*) – Channel number of inputs.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor.

Returns Forward results.

Return type Tensor

init_weights(*pretrained=None*, *strict=True*)

Init weights for models.

Parameters

- **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.
- **strict** (*boo, optional*) – Whether strictly load the pretrained model. Defaults to True.

class mmagic.models.editors.**MaxFeature**(*in_channels*, *out_channels*, *kernel_size=3*, *stride=1*, *padding=1*, *filter_type='conv2d'*)

Bases: torch.nn.Module

Conv2d or Linear layer with max feature selector.

Generate feature maps with double channels, split them and select the max feature.

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **kernel_size** (*int or tuple*) – Size of the convolving kernel.
- **stride** (*int or tuple, optional*) – Stride of the convolution. Default: 1
- **padding** (*int or tuple, optional*) – Zero-padding added to both sides of the input. Default: 1
- **filter_type** (*str*) – Type of filter. Options are 'conv2d' and 'linear'. Default: 'conv2d'.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor.

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**DIM**(*data_preprocessor*, *backbone*, *refiner=None*, *train_cfg=None*, *test_cfg=None*, *loss_alpha=None*, *loss_comp=None*, *loss_refine=None*, *init_cfg: Optional[dict] = None*)

Bases: *mmagic.models.base_models.BaseMattor*

Deep Image Matting model.

https://arxiv.org/abs/1703.03872

Note: For `(self.train_cfg.train_backbone, self.train_cfg.train_refiner)`:

- `(True, False)` corresponds to the encoder-decoder stage in the paper.
- `(False, True)` corresponds to the refinement stage in the paper.
- `(True, True)` corresponds to the fine-tune stage in the paper.

Parameters

- **data_preprocessor** (*dict, optional*) – Config of data pre-processor.
- **backbone** (*dict*) – Config of backbone.
- **refiner** (*dict*) – Config of refiner.
- **loss_alpha** (*dict*) – Config of the alpha prediction loss. Default: None.
- **loss_comp** (*dict*) – Config of the composition loss. Default: None.
- **loss_refine** (*dict*) – Config of the loss of the refiner. Default: None.
- **train_cfg** (*dict*) – Config of training. In `train_cfg`, `train_backbone` should be specified. If the model has a refiner, `train_refiner` should be specified.
- **test_cfg** (*dict*) – Config of testing. In `test_cfg`, If the model has a refiner, `train_refiner` should be specified.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Default: None.

property with_refiner

Whether the matting model has a refiner.

init_weights()

Initialize the model network weights.

train(*mode=True*)

Mode switcher.

Parameters **mode** (*bool*) – whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

freeze_backbone()

Freeze the backbone and only train the refiner.

_forward(*x: torch.Tensor*, *, *refine: bool = True*) → Tuple[torch.Tensor, torch.Tensor]

Raw forward function.

Parameters

- **x** (*torch.Tensor*) – Concatenation of merged image and trimap with shape (N, 4, H, W)

- **refine** (*bool*) – if forward through refiner

Returns pred_alpha, with shape (N, 1, H, W) torch.Tensor: pred_refine, with shape (N, 4, H, W)

Return type torch.Tensor

_forward_test(*inputs*)

Forward to get alpha prediction.

_forward_train(*inputs*, *data_samples*)

Defines the computation performed at every training call.

Parameters

- **inputs** (*torch.Tensor*) – Concatenation of normalized image and trimap shape (N, 4, H, W)
- **data_samples** (*list[DataSample]*) – Data samples containing: - gt_alpha (Tensor): Ground-truth of alpha

  shape (N, 1, H, W), normalized to 0 to 1.

  - **gt_fg (Tensor): Ground-truth of foreground** shape (N, C, H, W), normalized to 0 to 1.
  - **gt_bg (Tensor): Ground-truth of background** shape (N, C, H, W), normalized to 0 to 1.

Returns Contains the loss items and batch information.

Return type dict

class mmagic.models.editors.**ClipWrapper**(*clip_type*, **args*, ***kwargs*)

Bases: torch.nn.Module

Clip Models wrapper.

We provide wrappers for the clip models of `openai` and `mlfoundations`, where the user can specify `clip_type` as `clip` or `open_clip`, and then initialize a clip model using the same arguments as in the original codebase. The following clip models settings are provided in the official repo of disco diffusion: | Setting | Source | Arguments | # noqa |:----------:|————|----------| # noqa | ViTB32 | clip | name='ViT-B/32', jit=False | # noqa | ViTB16 | clip | name='ViT-B/16', jit=False | # noqa | ViTL14 | clip | name='ViT-L/14', jit=False | # noqa | ViTL14_336px | clip | name='ViT-L/14@336px', jit=False | # noqa | RN50 | clip | name='RN50', jit=False | # noqa | RN50x4 | clip | name='RN50x4', jit=False | # noqa | RN50x16 | clip | name='RN50x16', jit=False | # noqa | RN50x64 | clip | name='RN50x64', jit=False | # noqa | RN101 | clip | name='RN101', jit=False | # noqa | ViTB32_laion2b_e16 | open_clip | name='ViT-B-32', pretrained='laion2b_e16' | # noqa | ViTB32_laion400m_e31 | open_clip | model_name='ViT-B-32', pretrained='laion400m_e31' | # noqa | ViTB32_laion400m_32 | open_clip | model_name='ViT-B-32', pretrained='laion400m_e32' | # noqa | ViTB32quickgelu_laion400m_e31 | open_clip | model_name='ViT-B-32-quickgelu', pretrained='laion400m_e31' | # noqa | ViTB32quickgelu_laion400m_e32 | open_clip | model_name='ViT-B-32-quickgelu', pretrained='laion400m_e32' | # noqa | ViTB16_laion400m_e31 | open_clip | model_name='ViT-B-16', pretrained='laion400m_e31' | # noqa | ViTB16_laion400m_e32 | open_clip | model_name='ViT-B-16', pretrained='laion400m_e32' | # noqa | RN50_yffcc15m | open_clip | model_name='RN50', pretrained='yfcc15m' | # noqa | RN50_cc12m | open_clip | model_name='RN50', pretrained='cc12m' | # noqa | RN50_quickgelu_yfcc15m | open_clip | model_name='RN50-quickgelu', pretrained='yfcc15m' | # noqa | RN50_quickgelu_cc12m | open_clip | model_name='RN50-quickgelu', pretrained='cc12m' | # noqa | RN101_yfcc15m | open_clip | model_name='RN101', pretrained='yfcc15m' | # noqa | RN101_quickgelu_yfcc15m | open_clip | model_name='RN101-quickgelu', pretrained='yfcc15m' | # noqa

An example of a `clip_modes_cfg` is as follows:

Examples:

```
>>> # Use OpenAI's CLIP
>>> config = dict(
>>>     type='ClipWrapper',
>>>     clip_type='clip',
>>>     name='ViT-B/32',
>>>     jit=False)
```

```
>>> # Use OpenCLIP
>>> config = dict(
>>>     type='ClipWrapper',
>>>     clip_type='open_clip',
>>>     model_name='RN50',
>>>     pretrained='yfcc15m')
```

```
>>> # Use CLIP from Hugging Face Transformers
>>> config = dict(
>>>     type='ClipWrapper',
>>>     clip_type='huggingface',
>>>     pretrained_model_name_or_path='runwayml/stable-diffusion-v1-5',
>>>     subfolder='text_encoder')
```

Parameters

- **clip_type** (*List[Dict]*) – The original source of the clip model. Whether be `clip`, `open_clip` or `hugging_face`.
- ***args** – Arguments to initialize corresponding clip model.
- ****kwargs** – Arguments to initialize corresponding clip model.

get_embedding_layer()

Function to get embedding layer of the clip model.

Only support for CLIPTextModel currently.

add_embedding(*embeddings: Union[dict, List[dict]]*)

set_only_embedding_trainable()

set_embedding_layer()

unset_embedding_layer()

forward(**args*, ***kwargs*)

Forward function.

class mmagic.models.editors.**DiscoDiffusion**(*unet*, *diffusion_scheduler*, *secondary_model=None*, *clip_models=[]*, *use_fp16=False*, *pretrained_cfgs=None*)

Bases: `torch.nn.Module`

Disco Diffusion (DD) is a Google Colab Notebook which leverages an AI Image generating technique called CLIP-Guided Diffusion to allow you to create compelling and beautiful images from just text inputs. Created by Somnai, augmented by Gandamu, and building on the work of RiversHaveWings, nshepperd, and many others.

Ref: Github Repo: https://github.com/alembics/disco-diffusion Colab: https://colab.research.google.com/github/alembics/disco-diffusion/blob/main/Disco_Diffusion.ipynb # noqa

Parameters

- **unet** (*ModelType*) – Config of denoising Unet.
- **diffusion_scheduler** (*ModelType*) – Config of diffusion_scheduler scheduler.
- **secondary_model** (*ModelType*) – A smaller secondary diffusion model trained by Katherine Crowson to remove noise from intermediate timesteps to prepare them for CLIP. Ref: https://twitter.com/rivershavewings/status/1462859669454536711 # noqa Defaults to None.
- **clip_models** (*list*) – Config of clip models. Defaults to [].
- **use_fp16** (*bool*) – Whether to use fp16 for unet model. Defaults to False.
- **pretrained_cfgs** (*dict*) – Path Config for pretrained weights. Usually this is a dict contains module name and the corresponding ckpt path. Defaults to None.

property device

Get current device of the model.

Returns The current device of the model.

Return type torch.device

load_pretrained_models(*pretrained_cfgs*)

Loading pretrained weights to model. `pretrained_cfgs` is a dict consist of module name as key and checkpoint path as value.

Parameters

- **pretrained_cfgs** (*dict*) – Path Config for pretrained weights.
- **the** (*Usually this is a dict contains module name and*) –
- **None.** (*corresponding ckpt path. Defaults to*) –

infer(*scheduler_kwargs=None*, *height=None*, *width=None*, *init_image=None*, *batch_size=1*, *num_inference_steps=100*, *skip_steps=0*, *show_progress=True*, *text_prompts=[]*, *image_prompts=[]*, *eta=0.8*, *clip_guidance_scale=5000*, *init_scale=1000*, *tv_scale=0.0*, *sat_scale=0.0*, *range_scale=150*, *cut_overview=[12] * 400 + [4] * 600*, *cut_innercut=[4] * 400 + [12] * 600*, *cut_ic_pow=[1] * 1000*, *cut_icgray_p=[0.2] * 400 + [0] * 600*, *cutn_batches=4*, *seed=None*)

Inference API for disco diffusion.

Parameters

- **scheduler_kwargs** (*dict*) – Args for infer time diffusion scheduler. Defaults to None.
- **height** (*int*) – Height of output image. Defaults to None.
- **width** (*int*) – Width of output image. Defaults to None.
- **init_image** (*str*) – Initial image at the start point of denoising. Defaults to None.
- **batch_size** (*int*) – Batch size. Defaults to 1.
- **num_inference_steps** (*int*) – Number of inference steps. Defaults to 1000.
- **skip_steps** (*int*) – Denoising steps to skip, usually set with `init_image`. Defaults to 0.
- **show_progress** (*bool*) – Whether to show progress. Defaults to False.
- **text_prompts** (*list*) – Text prompts. Defaults to [].

- **image_prompts** (*list*) – Image prompts, this is not the same as `init_image`, they works the same way with `text_prompts`. Defaults to [].
- **eta** (*float*) – Eta for ddim sampling. Defaults to 0.8.
- **clip_guidance_scale** (*int*) – The Scale of influence of prompts on output image. Defaults to 1000.
- **seed** (*int*) – Sampling seed. Defaults to None.

class mmagic.models.editors.**DreamBooth**(*vae: ModelType, text_encoder: ModelType, tokenizer: str, unet: ModelType, scheduler: ModelType, test_scheduler: Optional[ModelType] = None, lora_config: Optional[dict] = None, val_prompts: Union[str, List[str]] = None, class_prior_prompt: Optional[str] = None, num_class_images: Optional[int] = 3, prior_loss_weight: float = 0, finetune_text_encoder: bool = False, dtype: str = 'fp16', enable_xformers: bool = True, noise_offset_weight: float = 0, tomesd_cfg: Optional[dict] = None, data_preprocessor: Optional[ModelType] = dict(type='DataPreprocessor'), init_cfg: Optional[dict] = None*)

Bases: `mmagic.models.editors.stable_diffusion.stable_diffusion.StableDiffusion`

Implementation of `DreamBooth with Stable Diffusion.

<https://arxiv.org/abs/2208.12242>`_ (DreamBooth).

Parameters

- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **lora_config** (*dict, optional*) – The config for LoRA finetuning. Defaults to None.
- **val_prompts** (*Union[str, List[str]], optional*) – The prompts for validation. Defaults to None.
- **class_prior_prompt** (*str, optional*) – The prompt for class prior loss.
- **num_class_images** (*int, optional*) – The number of images for class prior. Defaults to 3.
- **prior_loss_weight** (*float, optional*) – The weight for class prior loss. Defaults to 0.
- **finetune_text_encoder** (*bool, optional*) – Whether to fine-tune text encoder. Defaults to False.
- **dtype** (*str, optional*) – The dtype for the model. Defaults to 'fp16'.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise # noqa Defaults to 0.

- **tomesd_cfg** (*dict, optional*) – The config for TOMESD. Please refers to https://github.com/dbolya/tomesd and https://github.com/open-mmlab/mmagic/blob/main/mmagic/models/utils/tome_utils.py for detail. # noqa Defaults to None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Defaults to

  dict(type='DataPreprocessor').
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Defaults to None/

generate_class_prior_images(*num_batches=None*)

Generate images for class prior loss.

Parameters **num_batches** (*int*) – Number of batches to generate images. If not passed, all images will be generated in one forward. Defaults to None.

prepare_model()

Prepare model for training.

Move model to target dtype and disable gradient for some models.

set_lora()

Set LORA for model.

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

train_step(*data, optim_wrapper*)

Implements the default model training process including preprocessing, model forward propagation, loss calculation, optimization, and back-propagation.

During non-distributed training. If subclasses do not override the *train_step()*, `EpochBasedTrainLoop` or `IterBasedTrainLoop` will call this method to update model parameters. The default parameter update process is as follows:

1. Calls `self.data_processor(data, training=False)` to collect batch_inputs and corresponding data_samples(labels).
2. Calls `self(batch_inputs, data_samples, mode='loss')` to get raw loss
3. Calls `self.parse_losses` to get `parsed_losses` tensor used to backward and dict of loss tensor used to log messages.
4. Calls `optim_wrapper.update_params(loss)` to update model.

Parameters

- **data** (*dict or tuple or list*) – Data sampled from dataset.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, torch.Tensor]

abstract forward(*inputs: torch.Tensor*, *data_samples: Optional[list] = None*, *mode: str = 'tensor'*) → Union[Dict[str, torch.Tensor], list]

forward is not implemented now.

class mmagic.models.editors.**EDSRNet**(*in_channels*, *out_channels*, *mid_channels=64*, *num_blocks=16*, *upscale_factor=4*, *res_scale=1*, *rgb_mean=[0.4488, 0.4371, 0.404]*, *rgb_std=[1.0, 1.0, 1.0]*)

Bases: mmengine.model.BaseModule

EDSR network structure.

Paper: Enhanced Deep Residual Networks for Single Image Super-Resolution. Ref repo: https://github.com/thstkdgus35/EDSR-PyTorch

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64.
- **num_blocks** (*int*) – Block number in the trunk network. Default: 16.
- **upscale_factor** (*int*) – Upsampling factor. Support 2^n and 3. Default: 4.
- **res_scale** (*float*) – Used to scale the residual in residual block. Default: 1.
- **rgb_mean** (*list[float]*) – Image mean in RGB orders. Default: [0.4488, 0.4371, 0.4040], calculated from DIV2K dataset.
- **rgb_std** (*list[float]*) – Image std in RGB orders. In EDSR, it uses [1.0, 1.0, 1.0]. Default: [1.0, 1.0, 1.0].

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**EDVR**(*generator*, *pixel_loss*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.BaseEditModel

EDVR model for video super-resolution.

EDVR: Video Restoration with Enhanced Deformable Convolutional Networks.

Parameters

- **generator** (*dict*) – Config for the generator structure.

- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.

forward_train(*inputs*, *data_samples=None*)

Forward training. Returns dict of losses of training.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.

Returns Dict of losses.

Return type dict

class mmagic.models.editors.**EDVRNet**(*in_channels*, *out_channels*, *mid_channels=64*, *num_frames=5*, *deform_groups=8*, *num_blocks_extraction=5*, *num_blocks_reconstruction=10*, *center_frame_idx=2*, *with_tsa=True*, *init_cfg=None*)

Bases: `mmengine.model.BaseModule`

EDVR network structure for video super-resolution.

Now only support X4 upsampling factor. Paper: EDVR: Video Restoration with Enhanced Deformable Convolutional Networks.

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64.
- **num_frames** (*int*) – Number of input frames. Default: 5.
- **deform_groups** (*int*) – Deformable groups. Defaults: 8.
- **num_blocks_extraction** (*int*) – Number of blocks for feature extraction. Default: 5.
- **num_blocks_reconstruction** (*int*) – Number of blocks for reconstruction. Default: 10.
- **center_frame_idx** (*int*) – The index of center frame. Frame counting from 0. Default: 2.
- **with_tsa** (*bool*) – Whether to use TSA module. Default: True.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

forward(*x*)

Forward function for EDVRNet.

Parameters **x** (*Tensor*) – Input tensor with shape (n, t, c, h, w).

Returns SR center frame with shape (n, c, h, w).

Return type Tensor

init_weights()

Init weights for models.

class mmagic.models.editors.**EG3D**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *camera: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseConditionalGAN*

Implementation of *Efficient Geometry-aware 3D Generative Adversarial Networks*

<https://openaccess.thecvf.com/content/CVPR2022/papers/Chan_Efficient_Geometry-Aware_3D_Generative_Adversarial_Networks_CVPR_2022_paper.pdf>_ (EG3D). # noqa

Detailed architecture can be found in `TriplaneGenerator` and `DualDiscriminator`

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **camera** (*Optional[ModelType]*) – The pre-defined camera to sample random camera position. If you want to generate images or videos via high-level API, you must set this argument. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or `DataPreprocessor`.
- **generator_steps** (*int*) – Number of times the generator was completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – Number of times the discriminator was completely updated before the generator is updated. Defaults to 1.
- **noise_size** (*Optional[int]*) – Size of the input noise vector. Default to 128.
- **num_classes** (*Optional[int]*) – The number classes you would like to generate. Defaults to None.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.
- **loss_config** (*Optional[Dict]*) – The config for training losses. Defaults to None.

label_fn(*label: Optional[torch.Tensor] = None*, *num_batches: int = 1*) → torch.Tensor

Label sampling function for EG3D model.

Parameters **label** (*Optional[Tensor]*) – Conditional for EG3D model. If not passed, `self.camera` will be used to sample random camera-to-world and intrinsics matrix. Defaults to None.

Returns Conditional input for EG3D model.

Return type torch.Tensor

data_sample_to_label(*data_sample: mmagic.utils.typing.SampleList*) → Optional[torch.Tensor]

Get labels from input *data_sample* and pack to *torch.Tensor*. If no label is found in the passed *data_sample*, *None* would be returned.

Parameters **data_sample** (*List[DataSample]*) – Input data samples.

Returns Packed label tensor.

Return type Optional[torch.Tensor]

pack_to_data_sample(*output: Dict[str, torch.Tensor]*, *data_sample: Optional[*mmagic.structures.DataSample*] = None*) → *mmagic.structures.DataSample*

Pack output to data sample. If `data_sample` is not passed, a new DataSample will be instantiated. Otherwise, outputs will be added to the passed datasample.

Parameters

- **output** (*Dict[Tensor]*) – Output of the model.
- **index** (*int*) – The index to save.
- **data_sample** (*DataSample, optional*) – Data sample to save outputs. Defaults to None.

Returns Data sample with packed outputs.

Return type DataSample

forward(*inputs: mmagic.utils.typing.ForwardInputs*, *data_samples: Optional[list] = None*, *mode: Optional[str] = None*) → List[*mmagic.structures.DataSample*]

Sample images with the given inputs. If forward mode is 'ema' or 'orig', the image generated by corresponding generator will be returned. If forward mode is 'ema/orig', images generated by original generator and EMA generator will both be returned in a dict.

Parameters

- **inputs** (*ForwardInputs*) – Dict containing the necessary information (e.g. noise, num_batches, mode) to generate image.
- **data_samples** (*Optional[list]*) – Data samples collated by `data_preprocessor`. Defaults to None.
- **mode** (*Optional[str]*) – *mode* is not used in `BaseConditionalGAN`. Defaults to None.

Returns Generated images or image dict.

Return type List[DataSample]

interpolation(*num_images: int*, *num_batches: int = 4*, *mode: str = 'both'*, *sample_model: str = 'orig'*, *show_pbar: bool = True*) → List[dict]

Interpolation input and return a list of output results. We support three kinds of interpolation mode:

- **'camera': First generate style code with random noise and forward** camera. Then synthesis images with interpolated camera position and fixed style code.
- **'conditioning': First generate style code with fixed noise and** interpolated camera. Then synthesis images with style codes and forward camera.
- 'both': Generate images with interpolated camera position.

Parameters

- **num_images** (*int*) – The number of images want to generate.
- **num_batches** (*int, optional*) – The number of batches to generate at one time. Defaults to 4.
- **mode** (*str, optional*) – The interpolation mode. Supported choices are 'both', 'camera', and 'conditioning'. Defaults to 'both'.
- **sample_model** (*str, optional*) – The model used to generate images, support 'orig' and 'ema'. Defaults to 'orig'.
- **show_pbar** (*bool, optional*) – Whether display a progress bar during interpolation. Defaults to True.

Returns The list of output dict of each frame.

Return type List[dict]

class mmagic.models.editors.**ESRGAN**(*generator*, *discriminator=None*, *gan_loss=None*, *pixel_loss=None*, *perceptual_loss=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.editors.srgan.SRGAN

Enhanced SRGAN model for single image super-resolution.

Ref: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. It uses RaGAN for GAN updates: The relativistic discriminator: a key element missing from standard GAN.

Parameters

- **generator** (*dict*) – Config for the generator.
- **discriminator** (*dict*) – Config for the discriminator. Default: None.
- **gan_loss** (*dict*) – Config for the gan loss. Note that the loss weight in gan loss is only for the generator.
- **pixel_loss** (*dict*) – Config for the pixel loss. Default: None.
- **perceptual_loss** (*dict*) – Config for the perceptual loss. Default: None.
- **train_cfg** (*dict*) – Config for training. Default: None. You may change the training of gan by setting: *disc_steps*: how many discriminator updates after one generate update; *disc_init_steps*: how many discriminator updates at the start of the training. These two keys are useful when training with WGAN.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule. Default: None.

g_step(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

d_step_real(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*)

D step of real data.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

d_step_fake(*batch_outputs: torch.Tensor*, *batch_gt_data*)

D step of fake data.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

class mmagic.models.editors.**RRDBNet**(*in_channels*, *out_channels*, *mid_channels=64*, *num_blocks=23*, *growth_channels=32*, *upscale_factor=4*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

Networks consisting of Residual in Residual Dense Block, which is used in ESRGAN and Real-ESRGAN.

ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. # noqa: E501 Currently, it supports [x1/x2/x4] upsampling scale factor.

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64
- **num_blocks** (*int*) – Block number in the trunk network. Defaults: 23
- **growth_channels** (*int*) – Channels for each growth. Default: 32.
- **upscale_factor** (*int*) – Upsampling factor. Support x1, x2 and x4. Default: 4.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

_supported_upscale_factors = [1, 2, 4]

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

init_weights()

Init weights for models.

class mmagic.models.editors.**FastComposer**(*pretrained_cfg: dict*, *vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: str = 'fp32'*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor=dict(type='DataPreprocessor')*, *init_cfg: Optional[dict] = None*)

Bases: mmagic.models.editors.stable_diffusion.StableDiffusion

Class for Stable Diffusion. Refers to https://github.com/Stability- AI/stablediffusion and https://github.com/huggingface/diffusers/blob/main/s rc/diffusers/pipelines/stable_diffusion/pipeline_stable_diffusion_attend_and_excite.py # noqa.

Parameters

- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model This argument will not work when dtype is defined for submodels. Defaults to None.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise Defaults to 0.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.

_tokenize_and_mask_noun_phrases_ends(*caption*)

Augment the text embedding.

_encode_augmented_prompt(*prompt: str*, *reference_images: List[PIL.Image.Image]*, *device: torch.device*, *weight_dtype: torch.dtype*)

Encode reference images.

Parameters

- **prompt** (*str or list(int)*) – prompt to be encoded.
- **reference_images** – (List[Image.Image]): List of reference images.
- **device** (*torch.device*) – torch device.
- **weight_dtype** (*torch.dtype*) – torch.dtype.

Returns

text embeddings generated by clip text encoder.

Return type text_embeddings (torch.Tensor)

infer(*prompt: Union[str, List[str]] = None*, *height: Optional[int] = None*, *width: Optional[int] = None*, *num_inference_steps: int = 50*, *guidance_scale: float = 7.5*, *negative_prompt: Optional[Union[str, List[str]]] = None*, *num_images_per_prompt: Optional[int] = 1*, *eta: float = 0.0*, *generator: Optional[Union[torch.Generator, List[torch.Generator]]] = None*, *latents: Optional[torch.FloatTensor] = None*, *prompt_embeds: Optional[torch.FloatTensor] = None*, *negative_prompt_embeds: Optional[torch.FloatTensor] = None*, *output_type: Optional[str] = 'pil'*, *return_dict: bool = True*, *callback: Optional[Callable[[int, int, torch.FloatTensor], None]] = None*, *callback_steps: int = 1*, *cross_attention_kwargs: Optional[Dict[str, Any]] = None*, *alpha_: float = 0.7*, *reference_subject_images: List[PIL.Image.Image] = None*, *augmented_prompt_embeds: Optional[torch.FloatTensor] = None*, *show_progress: bool = True*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts to guide the image generation. If not defined, one has to pass *prompt_embeds*. instead.
- **height** (*int*, *optional*) – defaults to self.unet.config.sample_size * self.vae_scale_factor The height in pixels of the generated image.
- **width** (*int*, *optional*) – defaults to self.unet.config.sample_size * self.vae_scale_factor The width in pixels of the generated image.
- **num_inference_steps** (*int*, *optional*, defaults to 50) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
- **guidance_scale** (*float*, *optional*, defaults to 7.5) – Guidance scale as defined in [Classifier-Free Diffusion Guidance] (https://arxiv.org/abs/2207.12598). *guidance_scale* is defined as *w* of equation 2. of [Imagen Paper](https://arxiv.org/pdf/2205.11487.pdf). Guidance scale is enabled by setting *guidance_scale > 1*. Higher guidance scale encourages to generate images that are closely linked to the text *prompt*, usually at the expense of lower image quality.
- **negative_prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts not to guide the image generation. If not defined, one has to pass *negative_prompt_embeds* instead. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).
- **num_images_per_prompt** (*int*, *optional*, defaults to 1) – The number of images to generate per prompt.
- **eta** (*float*, *optional*, defaults to 0.0) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to [*schedulers.DDIMScheduler*], will be ignored for others.
- **List[torch.Generator]** (*generator (torch.Generator or*) –

  ***optional*):** One or a list of [torch generator(s)](https://pytorch.org/docs/stable/generated/torch.Generator.html) to make generation deterministic.

:param :

***optional*):** One or a list of [torch generator(s)](https://pytorch.org/docs/stable/generated/torch.Generator.html) to make generation deterministic.

Parameters

- **latents** (*torch.FloatTensor*, *optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak

the same generation with different prompts. If not provided, a latents tensor will ge generated by sampling using the supplied random *generator*.

- **prompt_embeds** (*torch.FloatTensor*, *optional*) – Pre-generated text embeddings. Can be used to easily tweak text inputs, *e.g.* prompt weighting. If not provided, text embeddings will be generated from *prompt* input argument.
- **negative_prompt_embeds** (*torch.FloatTensor*, *optional*) – Pre-generated negative text embeddings. Can be used to easily tweak text inputs, *e.g.* prompt weighting. If not provided, negative_prompt_embeds will be generated from *negative_prompt* input argument.
- **output_type** (*str*, *optional*, defaults to *"pil"*) – The output format of the generate image. Choose between [PIL](https://pillow.readthedocs.io/en/stable/): *PIL.Image.Image* or *np.array*.
- **return_dict** (*bool*, *optional*, defaults to *True*) – Whether or not to return a [*~pipelines.stable_diffusion.StableDiffusionPipelineOutput*] instead of a plain tuple.
- **callback** (*Callable*, *optional*) – A function that will be called every *callback_steps* steps during inference. The function will be called with the following arguments: *callback(step: int, timestep: int, latents: torch.FloatTensor)*.
- **callback_steps** (*int*, *optional*, defaults to 1) – The frequency at which the *callback* function will be called. If not specified, the callback will be called at every step.
- **cross_attention_kwargs** (*dict*, *optional*) – A kwargs dictionary that if specified is passed along to the *AttentionProcessor* as defined under *self.processor* in [diffusers.cross_attention](https://github.com/huggingface/ diffusers/blob/main/src/diffusers/models/cross_attention.py).
- **alpha** (*float*, defaults to 0.7) – The ratio of subject conditioning. If *alpha_* is 0.7, the beginning 30% of denoising steps use text prompts, while the last 70% utilize image-augmented prompts. Increase alpha for identity preservation, decrease it for prompt consistency.
- **reference_subject_images** (*List[PIL.Image.Image]*) – a list of PIL images that are used as reference subjects. The number of images should be equal to the number of augmented tokens in the prompts.
- **augmented_prompt_embeds** – (*torch.FloatTensor*, *optional*): Pre-generated image augmented text embeddings. If not provided, embeddings will be generated from *prompt* and *reference_subject_images*.
- **show_progress** – ('bool'): show progress or not.

Examples:

Returns *OrderedDict* if *return_dict* is True, otherwise a *tuple*. *When returning a tuple, the first element is a list with the generated images, and the second element is a list of `bool`s denoting whether the corresponding generated image likely represents "not-safe-for-work" (nsfw) content, according to the `safety_checker*.

Return type *OrderedDict* or *tuple*

class mmagic.models.editors.**FBADecoder**(*pool_scales*, *in_channels*, *channels*, *conv_cfg=None*, *norm_cfg=dict(type='BN')*, *act_cfg=dict(type='ReLU')*, *align_corners=False*)

Bases: torch.nn.Module

Decoder for FBA matting.

Parameters

- **pool_scales** (*tuple[int]*) – Pooling scales used in
- **Module.** (*Pooling Pyramid*) –
- **in_channels** (*int*) – Input channels.
- **channels** (*int*) – Channels after modules, before conv_seg.
- **conv_cfg** (*dict | None*) – Config of conv layers.
- **norm_cfg** (*dict | None*) – Config of norm layers.
- **act_cfg** (*dict*) – Config of activation layers.
- **align_corners** (*bool*) – align_corners argument of F.interpolate.

init_weights(*pretrained=None*)

Init weights for the model.

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

forward(*inputs*)

Forward function.

Parameters **inputs** (*dict*) – Output dict of FbaEncoder.

Returns Predicted alpha, fg and bg of the current batch.

Return type tuple(Tensor)

class mmagic.models.editors.**FBAResnetDilated**(*depth: int, in_channels: int = 3, stem_channels: int = 64, base_channels: int = 64, num_stages: int = 4, strides: Sequence[int] = (1, 2, 2, 2), dilations: Sequence[int] = (1, 1, 2, 4), deep_stem: bool = False, avg_down: bool = False, frozen_stages: int = - 1, act_cfg: dict = dict(type='ReLU'), conv_cfg: Optional[dict] = None, norm_cfg: dict = dict(type='BN'), with_cp: bool = False, multi_grid: Optional[Sequence[int]] = None, contract_dilation: bool = False, zero_init_residual: bool = True*)

Bases: *mmagic.models.archs.ResNet*

ResNet-based encoder for FBA image matting.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (N, C, H, W).

Returns Output tensor.

Return type Tensor

class mmagic.models.editors.**FLAVR**(*generator: dict, pixel_loss: dict, train_cfg: Optional[dict] = None, test_cfg: Optional[dict] = None, required_frames: int = 2, step_frames: int = 1, init_cfg: Optional[dict] = None, data_preprocessor: Optional[dict] = None*)

Bases: *mmagic.models.base_models.BasicInterpolator*

FLAVR model for video interpolation.

Paper: FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation

Ref repo: https://github.com/tarun005/FLAVR

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **required_frames** (*int*) – Required frames in each process. Default: 2
- **step_frames** (*int*) – Step size of video frame interpolation. Default: 1
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.

init_cfg

Initialization config dict.

Type dict, optional

data_preprocessor

Used for pre-processing data sampled by dataloader to the format accepted by `forward()`.

Type `BaseDataPreprocessor`

static merge_frames(*input_tensors*, *output_tensors*)

merge input frames and output frames.

Interpolate a frame between the given two frames.

Merged from [[in1, in2, in3, in4], [in2, in3, in4, in5], . . .] [[out1], [out2], [out3], . . .]

to [in1, in2, out1, in3, out2, . . . , in(-3), out(-1), in(-2), in(-1)]

Parameters

- **input_tensors** (*Tensor*) – The input frames with shape [n, 4, c, h, w]
- **output_tensors** (*Tensor*) – The output frames with shape [n, 1, c, h, w].

Returns The final frames.

Return type list[np.array]

class mmagic.models.editors.**FLAVRNet**(*num_input_frames*, *num_output_frames*, *mid_channels_list=[512, 256, 128, 64]*, *encoder_layers_list=[2, 2, 2, 2]*, *bias=False*, *norm_cfg=None*, *join_type='concat'*, *up_mode='transpose'*, *init_cfg=None*)

Bases: `mmengine.model.BaseModule`

PyTorch implementation of FLAVR for video frame interpolation.

Paper: FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation

Ref repo: https://github.com/tarun005/FLAVR

Parameters

- **num_input_frames** (*int*) – Number of input frames.
- **num_output_frames** (*int*) – Number of output frames.
- **mid_channels_list** (*list[int]*) – List of number of mid channels. Default: [512, 256, 128, 64]
- **encoder_layers_list** (*list[int]*) – List of number of layers in encoder. Default: [2, 2, 2, 2]
- **bias** (*bool*) – If `True`, adds a learnable bias to the conv layers. Default: `True`
- **norm_cfg** (*dict | None*) – Config dict for normalization layer. Default: None
- **join_type** (*str*) – Join type of tensors from decoder and encoder. Candidates are `concat` and `add`. Default: `concat`
- **up_mode** (*str*) – Up-mode UpConv3d, candidates are `transpose` and `trilinear`. Default: `transpose`
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

forward(*images: torch.Tensor*)

Forward function.

Parameters **images** (*Tensor*) – Input frames tensor with shape (N, T, C, H, W).

Returns Output tensor.

Return type out (Tensor)

class mmagic.models.editors.**GCA**(*data_preprocessor*, *backbone*, *loss_alpha=None*, *init_cfg: Optional[dict] = None*, *train_cfg=None*, *test_cfg=None*)

Bases: *mmagic.models.base_models.BaseMattor*

Guided Contextual Attention image matting model.

https://arxiv.org/abs/2001.04069

Parameters

- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.
- **backbone** (*dict*) – Config of backbone.
- **loss_alpha** (*dict*) – Config of the alpha prediction loss. Default: None.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.
- **train_cfg** (*dict*) – Config of training. In `train_cfg`, `train_backbone` should be specified. If the model has a refiner, `train_refiner` should be specified.
- **test_cfg** (*dict*) – Config of testing. In `test_cfg`, If the model has a refiner, `train_refiner` should be specified.

_forward(*inputs*)

Forward function.

Parameters **inputs** (*torch.Tensor*) – Input tensor.

Returns Output tensor.

Return type Tensor

_forward_test(*inputs*)

Forward function for testing GCA model.

Parameters **inputs** (*torch.Tensor*) – batch input tensor.

Returns Output tensor of model.

Return type Tensor

_forward_train(*inputs*, *data_samples*)

Forward function for training GCA model.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement]*) – data samples collated by `data_preprocessor`.

Returns Contains the loss items and batch information.

Return type dict

class mmagic.models.editors.**GGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseGAN*

Implementation of *Geometric GAN*.

<https://arxiv.org/abs/1705.02894>`_(GGAN).

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple

Get disc loss. GGAN use hinge loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake*)

Get disc loss. GGAN use hinge loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.

- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.editors.**GLEANStyleGANv2**(*in_size*, *out_size*, *img_channels=3*, *rrdb_channels=64*, *num_rrdbs=23*, *style_channels=512*, *num_mlps=8*, *channel_multiplier=2*, *blur_kernel=[1, 3, 3, 1]*, *lr_mlp=0.01*, *default_style_mode='mix'*, *eval_style_mode='single'*, *mix_prob=0.9*, *init_cfg=None*, *fp16_enabled=False*, *bgr2rgb=False*)

Bases: `mmengine.model.BaseModule`

GLEAN (using StyleGANv2) architecture for super-resolution.

Paper: GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution, CVPR, 2021

This method makes use of StyleGAN2 and hence the arguments mostly follow that in 'StyleGAN2v2Generator'.

In StyleGAN2, we use a static architecture composing of a style mapping module and number of convolutional style blocks. More details can be found in: Analyzing and Improving the Image Quality of StyleGAN CVPR2020.

You can load pretrained model through passing information into `pretrained` argument. We have already offered official weights as follows:

- stylegan2-ffhq-config-f: http://download.openmmlab.com/mmediting/stylegan2/official_weights/stylegan2-ffhq-config-f-official_20210327_171224-bce9310c.pth # noqa
- stylegan2-horse-config-f: http://download.openmmlab.com/mmediting/stylegan2/official_weights/stylegan2-horse-config-f-official_20210327_173203-ef3e69ca.pth # noqa
- stylegan2-car-config-f: http://download.openmmlab.com/mmediting/stylegan2/official_weights/stylegan2-car-config-f-official_20210327_172340-8cfe053c.pth # noqa
- stylegan2-cat-config-f: http://download.openmmlab.com/mmediting/stylegan2/official_weights/stylegan2-cat-config-f-official_20210327_172444-15bc485b.pth # noqa
- stylegan2-church-config-f: http://download.openmmlab.com/mmediting/stylegan2/official_weights/stylegan2-church-config-f-official_20210327_172657-1d42b7d1.pth # noqa

If you want to load the ema model, you can just use following codes:

```
# ckpt_http is one of the valid path from http source
generator = StyleGANv2Generator(1024, 512,
                                pretrained=dict(
                                    ckpt_path=ckpt_http,
                                    prefix='generator_ema'))
```

Of course, you can also download the checkpoint in advance and set `ckpt_path` with local path. If you just want to load the original generator (not the ema model), please set the prefix with 'generator'.

Note that our implementation allows to generate BGR image, while the original StyleGAN2 outputs RGB images by default. Thus, we provide `bgr2rgb` argument to convert the image space.

Parameters

- **in_size** (*int*) – The size of the input image.
- **out_size** (*int*) – The output size of the StyleGAN2 generator.
- **img_channels** (*int*) – Number of channels of the input images. 3 for RGB image and 1 for grayscale image. Default: 3.
- **rrdb_channels** (*int*) – Number of channels of the RRDB features. Default: 64.
- **num_rrdbs** (*int*) – Number of RRDB blocks in the encoder. Default: 23.
- **style_channels** (*int*) – The number of channels for style code. Default: 512.
- **num_mlps** (*int, optional*) – The number of MLP layers. Defaults to 8.
- **channel_multiplier** (*int, optional*) – The multiplier factor for the channel number. Defaults to 2.
- **blur_kernel** (*list, optional*) – The blurry kernel. Defaults to [1, 3, 3, 1].
- **lr_mlp** (*float, optional*) – The learning rate for the style mapping layer. Defaults to 0.01.
- **default_style_mode** (*str, optional*) – The default mode of style mixing. In training, we adopt mixing style mode in default. However, in the evaluation, we use 'single' style mode. *['mix', 'single']* are currently supported. Defaults to 'mix'.
- **eval_style_mode** (*str, optional*) – The evaluation mode of style mixing. Defaults to 'single'.
- **mix_prob** (*float, optional*) – Mixing probability. The value should be in range of [0, 1]. Defaults to 0.9.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.
- **fp16_enabled** (*bool, optional*) – Whether to use fp16 training in this module. Defaults to False.
- **bgr2rgb** (*bool, optional*) – Whether to flip the image channel dimension. Defaults to False.

forward(*lq*)

Forward function.

Parameters **lq** (*Tensor*) – Input LR image with shape (n, c, h, w).

Returns Output HR image.

Return type Tensor

class mmagic.models.editors.**GLDecoder**(*in_channels=256*, *norm_cfg=None*, *act_cfg=dict(type='ReLU')*, *out_act='clip'*)

Bases: mmengine.model.BaseModule

Decoder used in Global&Local model.

This implementation follows: Globally and locally Consistent Image Completion

Parameters

- **in_channels** (*int*) – Channel number of input feature.
- **norm_cfg** (*dict*) – Config dict to build norm layer.
- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.
- **out_act** (*str*) – Output activation type, "clip" by default. Noted that in our implementation, we clip the output with range [-1, 1].

forward(*x*)

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**GLDilationNeck**(*in_channels=256*, *conv_type='conv'*, *norm_cfg=None*, *act_cfg=dict(type='ReLU')*, ***kwargs*)

Bases: mmengine.model.BaseModule

Dilation Backbone used in Global&Local model.

This implementation follows: Globally and locally Consistent Image Completion

Parameters

- **in_channels** (*int*) – Channel number of input feature.
- **conv_type** (*str*) – The type of conv module. In DeepFillv1 model, the *conv_type* should be 'conv'. In DeepFillv2 model, the *conv_type* should be 'gated_conv'.
- **norm_cfg** (*dict*) – Config dict to build norm layer.
- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.
- **kwargs** (*keyword arguments*) –

_conv_type

forward(*x*)

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**GLEncoder**(*norm_cfg=None*, *act_cfg=dict(type='ReLU')*)

Bases: mmengine.model.BaseModule

Encoder used in Global&Local model.

This implementation follows: Globally and locally Consistent Image Completion

Parameters

- **norm_cfg** (*dict*) – Config dict to build norm layer.
- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.

forward(*x*)

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**GLEncoderDecoder**(*encoder=dict(type='GLEncoder')*, *decoder=dict(type='GLDecoder')*, *dilation_neck=dict(type='GLDilationNeck')*)

Bases: mmengine.model.BaseModule

Encoder-Decoder used in Global&Local model.

This implementation follows: Globally and locally Consistent Image Completion

The architecture of the encoder-decoder is: (conv2d x 6) –> (dilated conv2d x 4) –> (conv2d or deconv2d x 7)

Parameters

- **encoder** (*dict*) – Config dict to encoder.
- **decoder** (*dict*) – Config dict to build decoder.
- **dilation_neck** (*dict*) – Config dict to build dilation neck.

forward(*x*)

Forward Function.

Parameters **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**AblatedDiffusionModel**(*data_preprocessor*, *unet*, *diffusion_scheduler*, *use_fp16=False*, *classifier=None*, *classifier_scale=1.0*, *rgb2bgr=False*, *pretrained_cfgs=None*)

Bases: mmengine.model.BaseModel

Guided diffusion Model.

Parameters

- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **unet** (*ModelType*) – Config of denoising Unet.
- **diffusion_scheduler** (*ModelType*) – Config of diffusion_scheduler scheduler.
- **use_fp16** (*bool*) – Whether to use fp16 for unet model. Defaults to False.
- **classifier** (*ModelType*) – Config of classifier. Defaults to None.
- **pretrained_cfgs** (*dict*) – Path Config for pretrained weights. Usually this is a dict contains module name and the corresponding ckpt path.Defaults to None.

property device

Get current device of the model.

Returns The current device of the model.

Return type torch.device

load_pretrained_models(*pretrained_cfgs*)

summary

Parameters **pretrained_cfgs** (*_type_*) – _description_

infer(*scheduler_kwargs=None*, *init_image=None*, *batch_size=1*, *num_inference_steps=1000*, *labels=None*, *classifier_scale=0.0*, *show_progress=False*)

summary

Parameters

- **init_image** (*_type_, optional*) – _description_. Defaults to None.
- **batch_size** (*int, optional*) – _description_. Defaults to 1.
- **num_inference_steps** (*int, optional*) – _description_. Defaults to 1000.
- **labels** (*_type_, optional*) – _description_. Defaults to None.
- **show_progress** (*bool, optional*) – _description_. Defaults to False.

Returns _description_

Return type _type_

forward(*inputs: mmagic.utils.typing.ForwardInputs*, *data_samples: Optional[list] = None*, *mode: Optional[str] = None*) → List[*mmagic.structures.DataSample*]

summary

Parameters

- **inputs** (*ForwardInputs*) – _description_
- **data_samples** (*Optional[list], optional*) – _description_. Defaults to None.
- **mode** (*Optional[str], optional*) – _description_. Defaults to None.

Returns _description_

Return type List[DataSample]

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data.

Calls `self.data_preprocessor(data)` and `self(inputs, data_sample, mode=None)` in order. Return the generated results which will be passed to evaluator.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as *val_step()*.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type List[DataSample]

`train_step`(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

summary

Parameters

- **data** (*dict*) – _description_
- **optim_wrapper** (*OptimWrapperDict*) – _description_

Returns _description_

Return type _type_

`get_module`(*model: torch.nn.Module*, *module_name: str*) → torch.nn.Module

Get an inner module from model.

Since we will wrapper DDP for some model, we have to judge whether the module can be indexed directly.

Parameters

- **model** (*nn.Module*) – This model may wrapped with DDP or not.
- **module_name** (*str*) – The name of specific module.

Returns Returned sub module.

Return type nn.Module

class `mmagic.models.editors.`**`IconVSRNet`**(*mid_channels=64*, *num_blocks=30*, *keyframe_stride=5*, *padding=2*, *spynet_pretrained=None*, *edvr_pretrained=None*)

Bases: `mmengine.model.BaseModule`

IconVSR network structure for video super-resolution.

Support only x4 upsampling.

Paper: BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond, CVPR, 2021

Parameters

- **mid_channels** (*int*) – Channel number of the intermediate features. Default: 64.
- **num_blocks** (*int*) – Number of residual blocks in each propagation branch. Default: 30.
- **keyframe_stride** (*int*) – Number determining the keyframes. If stride=5, then the (0, 5, 10, 15, ...)-th frame will be the keyframes. Default: 5.
- **padding** (*int*) – Number of frames to be padded at two ends of the sequence. 2 for REDS and 3 for Vimeo-90K. Default: 2.
- **spynet_pretrained** (*str*) – Pre-trained model path of SPyNet. Default: None.
- **edvr_pretrained** (*str*) – Pre-trained model path of EDVR (for refill). Default: None.

`spatial_padding`(*lrs*)

Apply padding spatially.

Since the PCD module in EDVR requires that the resolution is a multiple of 4, we apply padding to the input LR images if their resolution is not divisible by 4.

Parameters **lrs** (*Tensor*) – Input LR sequence with shape (n, t, c, h, w).

Returns Padded LR sequence with shape (n, t, c, h_pad, w_pad).

Return type Tensor

`check_if_mirror_extended`(*lrs*)

Check whether the input is a mirror-extended sequence.

If mirror-extended, the i-th (i=0, ..., t-1) frame is equal to the (t-1-i)-th frame.

Parameters **lrs** (*tensor*) – Input LR images with shape (n, t, c, h, w)

`compute_refill_features`(*lrs*, *keyframe_idx*)

Compute keyframe features for information-refill.

Since EDVR-M is used, padding is performed before feature computation. :param lrs: Input LR images with shape (n, t, c, h, w) :type lrs: Tensor :param keyframe_idx: The indices specifying the keyframes. :type keyframe_idx: list(int)

Returns

The keyframe features. Each key corresponds to the indices in keyframe_idx.

Return type dict(Tensor)

`compute_flow`(*lrs*)

Compute optical flow using SPyNet for feature warping.

Note that if the input is an mirror-extended sequence, 'flows_forward' is not needed, since it is equal to 'flows_backward.flip(1)'.

Parameters **lrs** (*tensor*) – Input LR images with shape (n, t, c, h, w)

Returns

Optical flow. 'flows_forward' corresponds to the flows used for forward-time propagation (current to previous). 'flows_backward' corresponds to the flows used for backward-time propagation (current to next).

Return type tuple(Tensor)

`forward`(*lrs*)

Forward function for IconVSR.

Parameters **lrs** (*Tensor*) – Input LR tensor with shape (n, t, c, h, w).

Returns Output HR tensor with shape (n, t, c, 4h, 4w).

Return type Tensor

class `mmagic.models.editors.`**`DepthwiseIndexBlock`**(*in_channels*, *norm_cfg=dict(type='BN')*, *use_context=False*, *use_nonlinear=False*, *mode='o2o'*)

Bases: `mmengine.model.BaseModule`

Depthwise index block.

From https://arxiv.org/abs/1908.00672.

Parameters

- **in_channels** (*int*) – Input channels of the holistic index block.
- **norm_cfg** (*dict*) – Config dict for normalization layer. Default: dict(type='BN').
- **use_context** (*bool, optional*) – Whether use larger kernel size in index block. Refer to the paper for more information. Defaults to False.

- **use_nonlinear** (*bool*) – Whether add a non-linear conv layer in the index blocks. Default: False.
- **mode** (*str*) – Mode of index block. Should be 'o2o' or 'm2o'. In 'o2o' mode, the group of the conv layers is 1; In 'm2o' mode, the group of the conv layer is *in_channels*.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input feature map with shape (N, C, H, W).

Returns Encoder index feature and decoder index feature.

Return type tuple(Tensor)

class mmagic.models.editors.**HolisticIndexBlock**(*in_channels*, *norm_cfg=dict(type='BN')*, *use_context=False*, *use_nonlinear=False*)

Bases: mmengine.model.BaseModule

Holistic Index Block.

From https://arxiv.org/abs/1908.00672.

Parameters

- **in_channels** (*int*) – Input channels of the holistic index block.
- **norm_cfg** (*dict*) – Config dict for normalization layer. Default: dict(type='BN').
- **use_context** (*bool, optional*) – Whether use larger kernel size in index block. Refer to the paper for more information. Defaults to False.
- **use_nonlinear** (*bool*) – Whether add a non-linear conv layer in the index block. Default: False.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input feature map with shape (N, C, H, W).

Returns Encoder index feature and decoder index feature.

Return type tuple(Tensor)

class mmagic.models.editors.**IndexedUpsample**(*in_channels*, *out_channels*, *kernel_size=5*, *norm_cfg=dict(type='BN')*, *conv_module=ConvModule*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModule

Indexed upsample module.

Parameters

- **in_channels** (*int*) – Input channels.
- **out_channels** (*int*) – Output channels.
- **kernel_size** (*int, optional*) – Kernel size of the convolution layer. Defaults to 5.
- **norm_cfg** (*dict, optional*) – Config dict for normalization layer. Defaults to dict(type='BN').
- **conv_module** (*ConvModule | DepthwiseSeparableConvModule, optional*) – Conv module. Defaults to ConvModule.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

init_weights()

Init weights for the module.

forward(*x*, *shortcut*, *dec_idx_feat=None*)

Forward function.

Parameters

- **x** (*Tensor*) – Input feature map with shape (N, C, H, W).
- **shortcut** (*Tensor*) – The shortcut connection with shape (N, C, H', W').
- **dec_idx_feat** (*Tensor, optional*) – The decode index feature map with shape (N, C, H', W'). Defaults to None.

Returns Output tensor with shape (N, C, H', W').

Return type Tensor

class mmagic.models.editors.**IndexNet**(*data_preprocessor*, *backbone*, *loss_alpha=None*, *loss_comp=None*, *init_cfg=None*, *train_cfg=None*, *test_cfg=None*)

Bases: *mmagic.models.base_models.BaseMattor*

IndexNet matting model.

This implementation follows: Indices Matter: Learning to Index for Deep Image Matting

Parameters

- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **backbone** (*dict*) – Config of backbone.
- **train_cfg** (*dict*) – Config of training. In 'train_cfg', 'train_backbone' should be specified.
- **test_cfg** (*dict*) – Config of testing.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.
- **loss_alpha** (*dict*) – Config of the alpha prediction loss. Default: None.
- **loss_comp** (*dict*) – Config of the composition loss. Default: None.

_forward(*inputs*)

Forward function.

Parameters **inputs** (*torch.Tensor*) – Input tensor.

Returns Output tensor.

Return type Tensor

_forward_test(*inputs*)

Forward function for testing IndexNet model.

Parameters **inputs** (*torch.Tensor*) – batch input tensor.

Returns Output tensor of model.

Return type Tensor

_forward_train(*inputs*, *data_samples*)

Forward function for training IndexNet model.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*List[BaseDataElement]*) – data samples collated by data_preprocessor.

Returns Contains the loss items and batch information.

Return type dict

class mmagic.models.editors.**IndexNetDecoder**(*in_channels*, *kernel_size=5*, *norm_cfg=dict(type='BN')*, *separable_conv=False*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModule

Decoder for IndexNet.

Please refer to https://arxiv.org/abs/1908.00672.

Parameters

- **in_channels** (*int*) – Input channels of the decoder.
- **kernel_size** (*int, optional*) – Kernel size of the convolution layer. Defaults to 5.
- **norm_cfg** (*None | dict, optional*) – Config dict for normalization layer. Defaults to dict(type='BN').
- **separable_conv** (*bool*) – Whether to use separable conv. Default: False.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

init_weights()

Init weights for the module.

forward(*inputs*)

Forward function.

Parameters **inputs** (*dict*) – Output dict of IndexNetEncoder.

Returns Predicted alpha matte of the current batch.

Return type Tensor

class mmagic.models.editors.**IndexNetEncoder**(*in_channels*, *out_stride=32*, *width_mult=1*, *index_mode='m2o'*, *aspp=True*, *norm_cfg=dict(type='BN')*, *freeze_bn=False*, *use_nonlinear=True*, *use_context=True*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModule

Encoder for IndexNet.

Please refer to https://arxiv.org/abs/1908.00672.

Parameters

- **in_channels** (*int, optional*) – Input channels of the encoder.
- **out_stride** (*int, optional*) – Output stride of the encoder. For example, if *out_stride* is 32, the input feature map or image will be downsample to the 1/32 of original size. Defaults to 32.

- **width_mult** (*int, optional*) – Width multiplication factor of channel dimension in MobileNetV2. Defaults to 1.
- **index_mode** (*str, optional*) – Index mode of the index network. It must be one of {'holistic', 'o2o', 'm2o'}. If it is set to 'holistic', then Holistic index network will be used as the index network. If it is set to 'o2o' (or 'm2o'), when O2O (or M2O) Depthwise index network will be used as the index network. Defaults to 'm2o'.
- **aspp** (*bool, optional*) – Whether use ASPP module to augment output feature. Defaults to True.
- **norm_cfg** (*None | dict, optional*) – Config dict for normalization layer. Defaults to dict(type='BN').
- **freeze_bn** (*bool, optional*) – Whether freeze batch norm layer. Defaults to False.
- **use_nonlinear** (*bool, optional*) – Whether use nonlinearity in index network. Refer to the paper for more information. Defaults to True.
- **use_context** (*bool, optional*) – Whether use larger kernel size in index network. Refer to the paper for more information. Defaults to True.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

Raises

- **ValueError** – out_stride must 16 or 32.
- **NameError** – Supported index_mode are {'holistic', 'o2o', 'm2o'}.

_make_layer(*layer_setting*, *norm_cfg*)

train(*mode=True*)

Set BatchNorm modules in the model to evaluation mode.

init_weights()

Init weights for the model.

Initialization is based on self._init_cfg

Parameters **pretrained** (*str, optional*) – Path for pretrained weights. If given None, pretrained weights will not be loaded. Defaults to None.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input feature map with shape (N, C, H, W).

Returns Output tensor, shortcut feature and decoder index feature.

Return type dict

class mmagic.models.editors.**InstColorization**(*data_preprocessor: Union[dict, mmengine.config.Config]*, *image_model*, *instance_model*, *fusion_model*, *color_data_opt*, *which_direction='AtoB'*, *loss=None*, *init_cfg=None*, *train_cfg=None*, *test_cfg=None*)

Bases: mmengine.model.BaseModel

Colorization InstColorization method.

This Colorization is implemented according to the paper: Instance-aware Image Colorization, CVPR 2020

Adapted from 'https://github.com/ericsujw/InstColorization.git' 'InstColorization/models/train_model' Copyright (c) 2020, Su, under MIT License.

Parameters

- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **image_model** (*dict*) – Config for single image model
- **instance_model** (*dict*) – Config for instance model
- **fusion_model** (*dict*) – Config for fusion model
- **color_data_opt** (*dict*) – Option for colorspace conversion
- **which_direction** (*str*) – AtoB or BtoA
- **loss** (*dict*) – Config for loss.
- **init_cfg** (*str*) – Initialization config dict. Default: None.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.

forward(*inputs: torch.Tensor, data_samples: Optional[List[*mmagic.structures.DataSample*]] = None, mode: str = 'tensor', **kwargs*)

Returns losses or predictions of training, validation, testing, and simple inference process.

forward method of BaseModel is an abstract method, its subclasses must implement this method.

Accepts inputs and data_samples processed by data_preprocessor, and returns results according to mode arguments.

During non-distributed training, validation, and testing process, forward will be called by BaseModel.train_step, BaseModel.val_step and BaseModel.val_step directly.

During distributed data parallel training process, MMSeparateDistributedDataParallel.train_step will first call DistributedDataParallel.forward to enable automatic gradient synchronization, and then call forward to get training loss.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.
- **mode** (*str*) – mode should be one of loss, predict and tensor. Default: 'tensor'.
  - loss: Called by train_step and return loss dict used for logging
  - predict: Called by val_step and test_step and return list of BaseDataElement results used for computing metric.
  - tensor: Called by custom use to get Tensor type results.

Returns

- If mode == loss, return a dict of loss tensor used for backward and logging.
- If mode == predict, return a list of BaseDataElement for computing metric and getting inference result.
- If mode == tensor, return a tensor or tuple of tensor or dict or tensor for custom use.

Return type ForwardResults

convert_to_datasample(*inputs*, *data_samples*)

Add predictions and destructed inputs (if passed) to data samples.

Parameters

- **inputs** (*Optional[torch.Tensor]*) – The input of model. Defaults to None.
- **data_samples** (*List[DataSample]*) – The data samples loaded from dataloader.

Returns Modified data samples.

Return type List[DataSample]

abstract forward_train(*inputs*, *data_samples=None*, ***kwargs*)

Forward function for training.

abstract train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step function.

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns

Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

forward_inference(*inputs*, *data_samples=None*, ***kwargs*)

Forward inference. Returns predictions of validation, testing.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.

Returns predictions.

Return type List[DataSample]

forward_tensor(*inputs*, *data_samples*)

Forward function in tensor mode.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_sample** (*dict*) – Dict contains data sample.

Returns Dict contains output results.

Return type dict

class mmagic.models.editors.**LIIF**(*generator: dict*, *pixel_loss: dict*, *train_cfg: Optional[dict] = None*, *test_cfg: Optional[dict] = None*, *init_cfg: Optional[dict] = None*, *data_preprocessor: Optional[dict] = None*)

Bases: *mmagic.models.base_models.BaseEditModel*

LIIF model for single image super-resolution.

Paper: Learning Continuous Image Representation with Local Implicit Image Function

Parameters

- **generator** (*dict*) – Config for the generator.
- **pixel_loss** (*dict*) – Config for the pixel loss.
- **pretrained** (*str*) – Path for pretrained model. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.

forward_tensor(*inputs*, *data_samples=None*, ***kwargs*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.

Returns result of simple forward.

Return type Tensor

forward_inference(*inputs*, *data_samples=None*, ***kwargs*)

Forward inference. Returns predictions of validation, testing, and simple inference.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.
- **data_samples** (*BaseDataElement, optional*) – data samples collated by data_preprocessor.

Returns predictions.

Return type List[DataSample]

class mmagic.models.editors.**MLPRefiner**(*in_dim*, *out_dim*, *hidden_list*)

Bases: mmengine.model.BaseModule

Multilayer perceptrons (MLPs), refiner used in LIIF.

Parameters

- **in_dim** (*int*) – Input dimension.
- **out_dim** (*int*) – Output dimension.
- **hidden_list** (*list[int]*) – List of hidden dimensions.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – The input of MLP.

Returns The output of MLP.

Return type Tensor

class `mmagic.models.editors.LSGAN`(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = None*, *ema_config: Optional[Dict] = None*, *loss_config: Optional[Dict] = None*)

Bases: `mmagic.models.base_models.BaseGAN`

Implementation of *Least Squares Generative Adversarial Networks*.

Paper link: https://arxiv.org/pdf/1611.04076.pdf

Detailed architecture can be found in `LSGANGenerator` and `LSGANDiscriminator`

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple

Get disc loss. LSGAN use the least squares loss to train the discriminator.

$$L_D = (D\left(X_{\text{data}}\right) - 1)^2 + (D(G(z)))^2$$

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple

Get gen loss. LSGAN use the least squares loss to train the generator.

$$L_G = (D(G(z)) - 1)^2$$

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

class `mmagic.models.editors.MSPIEStyleGAN2`(**args*, *train_settings=dict()*, ***kwargs*)

Bases: `mmagic.models.editors.stylegan2.StyleGAN2`

MS-PIE StyleGAN2.

In this GAN, we adopt the MS-PIE training schedule so that multi-scale images can be generated with a single generator. Details can be found in: Positional Encoding as Spatial Inductive Bias in GANs, CVPR2021.

Parameters **train_settings** (*dict*) – Config for training settings. Defaults to *dict()*.

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train GAN model. In the training of GAN models, generator and discriminator are updated alternatively. In MMagic's design, *self.train_step* is called with data input. Therefore we always update discriminator, whose updating is relay on real data, and then determine if the generator needs to be updated based on the current number of iterations. More details about whether to update generator can be found in `should_gen_update()`.

Parameters

- **data** (*dict*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapperDict*) – OptimWrapperDict instance contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*TrainInput*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*TrainInput*) – Inputs from dataloader.

- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.editors.**PESinGAN**(*generator: ModelType*, *discriminator: Optional[ModelType]*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *num_scales: Optional[int] = None*, *fixed_noise_with_pad: bool = False*, *first_fixed_noises_ch: int = 1*, *iters_per_scale: int = 200*, *noise_weight_init: int = 0.1*, *lr_scheduler_args: Optional[dict] = None*, *test_pkl_data: Optional[str] = None*, *ema_confg: Optional[dict] = None*)

Bases: mmagic.models.editors.singan.SinGAN

Positional Encoding in SinGAN.

This modified SinGAN is used to reimplement the experiments in: Positional Encoding as Spatial Inductive Bias in GANs, CVPR2021.

construct_fixed_noises()

Construct the fixed noises list used in SinGAN.

class mmagic.models.editors.**NAFBaseline**(*img_channel=3*, *mid_channels=16*, *middle_blk_num=1*, *enc_blk_nums=[1, 1, 1, 28]*, *dec_blk_nums=[1, 1, 1, 1]*, *dw_expand=1*, *ffn_expand=2*)

Bases: mmengine.model.BaseModule

The original version of Baseline model in "Simple Baseline for Image Restoration".

Parameters

- **img_channels** (*int*) – Channel number of inputs.
- **mid_channels** (*int*) – Channel number of intermediate features.
- **middle_blk_num** (*int*) – Number of middle blocks.
- **enc_blk_nums** (*List of int*) – Number of blocks for each encoder.
- **dec_blk_nums** (*List of int*) – Number of blocks for each decoder.

forward(*inp*)

Forward function.

Parameters **inp** – input tensor image with (B, C, H, W) shape

check_image_size(*x*)

Check image size and pad images so that it has enough dimension do downsample.

Parameters **x** – input tensor image with (B, C, H, W) shape.

class mmagic.models.editors.**NAFBaselineLocal**(**args*, *train_size=(1, 3, 256, 256)*, *fast_imp=False*, ***kwargs*)

Bases: mmagic.models.editors.nafnet.naf_avgpool2d.Local_Base, *NAFBaseline*

The original version of Baseline model in "Simple Baseline for Image Restoration".

Parameters

- **img_channels** (*int*) – Channel number of inputs.
- **mid_channels** (*int*) – Channel number of intermediate features.
- **middle_blk_num** (*int*) – Number of middle blocks.
- **enc_blk_nums** (*List of int*) – Number of blocks for each encoder.
- **dec_blk_nums** (*L`ist of int*) – Number of blocks for each decoder.

class mmagic.models.editors.**NAFNet**(*img_channels=3*, *mid_channels=16*, *middle_blk_num=1*, *enc_blk_nums=[]*, *dec_blk_nums=[]*)

Bases: mmengine.model.BaseModule

NAFNet.

The original version of NAFNet in "Simple Baseline for Image Restoration".

Parameters

- **img_channels** (*int*) – Channel number of inputs.
- **mid_channels** (*int*) – Channel number of intermediate features.
- **middle_blk_num** (*int*) – Number of middle blocks.
- **enc_blk_nums** (*List of int*) – Number of blocks for each encoder.
- **dec_blk_nums** (*List of int*) – Number of blocks for each decoder.

forward(*inp*)

Forward function.

Parameters **inp** – input tensor image with (B, C, H, W) shape

check_image_size(*x*)

Check image size and pad images so that it has enough dimension do downsample.

Parameters **x** – input tensor image with (B, C, H, W) shape.

class mmagic.models.editors.**NAFNetLocal**(**args*, *train_size=(1, 3, 256, 256)*, *fast_imp=False*, ***kwargs*)

Bases: mmagic.models.editors.nafnet.naf_avgpool2d.Local_Base, *NAFNet*

The original version of NAFNetLocal in "Simple Baseline for Image Restoration".

NAFNetLocal uses local average pooling modules than NAFNet.

Parameters

- **img_channels** (*int*) – Channel number of inputs.
- **mid_channels** (*int*) – Channel number of intermediate features.
- **middle_blk_num** (*int*) – Number of middle blocks.
- **enc_blk_nums** (*List of int*) – Number of blocks for each encoder.
- **dec_blk_nums** (*List of int*) – Number of blocks for each decoder.

class mmagic.models.editors.**MaskConvModule**(**args*, ***kwargs*)

Bases: mmcv.cnn.ConvModule

Mask convolution module.

This is a simple wrapper for mask convolution like: 'partial conv'. Convolutions in this module always need a mask as extra input.

Parameters

- **in_channels** (*int*) – Same as nn.Conv2d.
- **out_channels** (*int*) – Same as nn.Conv2d.
- **kernel_size** (*int or tuple[int]*) – Same as nn.Conv2d.
- **stride** (*int or tuple[int]*) – Same as nn.Conv2d.
- **padding** (*int or tuple[int]*) – Same as nn.Conv2d.
- **dilation** (*int or tuple[int]*) – Same as nn.Conv2d.
- **groups** (*int*) – Same as nn.Conv2d.
- **bias** (*bool or str*) – If specified as *auto*, it will be decided by the norm_cfg. Bias will be set as True if norm_cfg is None, otherwise False.
- **conv_cfg** (*dict*) – Config dict for convolution layer.
- **norm_cfg** (*dict*) – Config dict for normalization layer.
- **act_cfg** (*dict*) – Config dict for activation layer, "relu" by default.
- **inplace** (*bool*) – Whether to use inplace mode for activation.
- **with_spectral_norm** (*bool*) – Whether use spectral norm in conv module.
- **padding_mode** (*str*) – If the *padding_mode* has not been supported by current *Conv2d* in Pytorch, we will use our own padding layer instead. Currently, we support ['zeros', 'circular'] with official implementation and ['reflect'] with our own implementation. Default: 'zeros'.
- **order** (*tuple[str]*) – The order of conv/norm/activation layers. It is a sequence of "conv", "norm" and "act". Examples are ("conv", "norm", "act") and ("act", "conv", "norm").

supported_conv_list = ['PConv']

forward(*x*, *mask=None*, *activate=True*, *norm=True*, *return_mask=True*)

Forward function for partial conv2d.

Parameters

- **x** (*torch.Tensor*) – Tensor with shape of (n, c, h, w).
- **mask** (*torch.Tensor*) – Tensor with shape of (n, c, h, w) or (n, 1, h, w). If mask is not given, the function will work as standard conv2d. Default: None.
- **activate** (*bool*) – Whether use activation layer.
- **norm** (*bool*) – Whether use norm layer.
- **return_mask** (*bool*) – If True and mask is not None, the updated mask will be returned. Default: True.

Returns

Result Tensor or 2-tuple of

`Tensor`: Results after partial conv.

`Tensor`: Updated mask will be returned if mask is given and *return_mask* is True.

Return type Tensor or tuple

class mmagic.models.editors.**PartialConv2d**(**args*, *multi_channel=False*, *eps=1e-08*, ***kwargs*)

Bases: torch.nn.Conv2d

Implementation for partial convolution.

Image Inpainting for Irregular Holes Using Partial Convolutions [https://arxiv.org/abs/1804.07723]

Parameters

- **multi_channel** (*bool*) – If True, the mask is multi-channel. Otherwise, the mask is single-channel.
- **eps** (*float*) – Need to be changed for mixed precision training. For mixed precision training, you need change 1e-8 to 1e-6.

forward(*input*, *mask=None*, *return_mask=True*)

Forward function for partial conv2d.

Parameters

- **input** (*torch.Tensor*) – Tensor with shape of (n, c, h, w).
- **mask** (*torch.Tensor*) – Tensor with shape of (n, c, h, w) or (n, 1, h, w). If mask is not given, the function will work as standard conv2d. Default: None.
- **return_mask** (*bool*) – If True and mask is not None, the updated mask will be returned. Default: True.

Returns Results after partial conv. torch.Tensor : Updated mask will be returned if mask is given and return_mask is True.

Return type torch.Tensor

class mmagic.models.editors.**PConvDecoder**(*num_layers=7*, *interpolation='nearest'*, *conv_cfg=dict(type='PConv', multi_channel=True)*, *norm_cfg=dict(type='BN')*)

Bases: mmengine.model.BaseModule

Decoder with partial conv.

About the details for this architecture, pls see: Image Inpainting for Irregular Holes Using Partial Convolutions

Parameters

- **num_layers** (*int*) – The number of convolutional layers. Default: 7.
- **interpolation** (*str*) – The upsample mode. Default: 'nearest'.
- **conv_cfg** (*dict*) – Config for convolution module. Default: {'type': 'PConv', 'multi_channel': True}.
- **norm_cfg** (*dict*) – Config for norm layer. Default: {'type': 'BN'}.

forward(*input_dict*)

Forward Function.

Parameters **input_dict** (*dict | torch.Tensor*) – Input dict with middle features or torch.Tensor.

Returns Output tensor with shape of (n, c, h, w).

Return type torch.Tensor

class mmagic.models.editors.**PConvEncoder**(*in_channels=3*, *num_layers=7*, *conv_cfg=dict(type='PConv', multi_channel=True)*, *norm_cfg=dict(type='BN', requires_grad=True)*, *norm_eval=False*)

Bases: mmengine.model.BaseModule

Encoder with partial conv.

About the details for this architecture, pls see: Image Inpainting for Irregular Holes Using Partial Convolutions

Parameters

- **in_channels** (*int*) – The number of input channels. Default: 3.
- **num_layers** (*int*) – The number of convolutional layers. Default: 7.
- **conv_cfg** (*dict*) – Config for convolution module. Default: {'type': 'PConv', 'multi_channel': True}.
- **norm_cfg** (*dict*) – Config for norm layer. Default: {'type': 'BN'}.
- **norm_eval** (*bool*) – Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effective on Batch Norm and its variants only. Default: False.

train(*mode=True*)

Set BatchNorm modules in the model to evaluation mode.

forward(*x*, *mask*)

Forward function for partial conv encoder.

Parameters

- **x** (*torch.Tensor*) – Masked image with shape (n, c, h, w).
- **mask** (*torch.Tensor*) – Mask tensor with shape (n, c, h, w).

Returns Contains the results and middle level features in this module. *hidden_feats* contain the middle feature maps and *hidden_masks* store updated masks.

Return type dict

class mmagic.models.editors.**PConvEncoderDecoder**(*encoder*, *decoder*)

Bases: mmengine.model.BaseModule

Encoder-Decoder with partial conv module.

Parameters

- **encoder** (*dict*) – Config of the encoder.
- **decoder** (*dict*) – Config of the decoder.

forward(*x*, *mask_in*)

Forward Function.

Parameters

- **x** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).
- **mask_in** (*torch.Tensor*) – Input tensor with shape of (n, c, h, w).

Returns Output tensor with shape of (n, c, h', w').

Return type torch.Tensor

class mmagic.models.editors.**PConvInpaintor**(*data_preprocessor: Union[dict, mmengine.config.Config], encdec: dict, disc: Optional[dict] = None, loss_gan: Optional[dict] = None, loss_gp: Optional[dict] = None, loss_disc_shift: Optional[dict] = None, loss_composed_percep: Optional[dict] = None, loss_out_percep: bool = False, loss_l1_hole: Optional[dict] = None, loss_l1_valid: Optional[dict] = None, loss_tv: Optional[dict] = None, train_cfg: Optional[dict] = None, test_cfg: Optional[dict] = None, init_cfg: Optional[dict] = None*)

Bases: *mmagic.models.base_models.OneStageInpaintor*

Inpaintor for Partial Convolution method.

This inpaintor is implemented according to the paper: Image inpainting for irregular holes using partial convolutions

forward_tensor(*inputs, data_samples*)

Forward function in tensor mode.

Parameters

- **inputs** (*torch.Tensor*) – Input tensor.
- **data_sample** (*dict*) – Dict contains data sample.

Returns Dict contains output results.

Return type dict

train_step(*data: List[dict], optim_wrapper*)

Train step function.

In this function, the inpaintor will finish the train step following the pipeline:

1. get fake res/image
2. optimize discriminator (if have)
3. optimize generator

If *self.train_cfg.disc_step > 1*, the train step will contain multiple iterations for optimizing discriminator with different input data and only one iteration for optimizing generator after *disc_step* iterations for discriminator.

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*dict[torch.optim.Optimizer]*) – Dict with optimizers for generator and discriminator (if have).

Returns Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

class mmagic.models.editors.**ProgressiveGrowingGAN**(*generator, discriminator, data_preprocessor, nkimgs_per_scale, noise_size=None, interp_real=None, transition_kimgs: int = 600, prev_stage: int = 0, ema_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseGAN*

Progressive Growing Unconditional GAN.

In this GAN model, we implement progressive growing training schedule, which is proposed in Progressive Growing of GANs for improved Quality, Stability and Variation, ICLR 2018.

We highly recommend to use `GrowScaleImgDataset` for saving computational load in data pre-processing.

Notes for **using PGGAN**:

1. In official implementation, Tero uses gradient penalty with `norm_mode="HWC"`
2. We do not implement `minibatch_repeats` where has been used in official Tensorflow implementation.

Notes for resuming progressive growing GANs: Users should specify the `prev_stage` in `train_cfg`. Otherwise, the model is possible to reset the optimizer status, which will bring inferior performance. For example, if your model is resumed from the *256* stage, you should set `train_cfg=dict(prev_stage=256)`.

Parameters

- **generator** (*dict*) – Config for generator.
- **discriminator** (*dict*) – Config for discriminator.

forward(*inputs: mmagic.utils.typing.ForwardInputs*, *data_samples: Optional[list] = None*, *mode: Optional[str] = None*) → mmagic.utils.typing.SampleList

Sample images from noises by using the generator.

Parameters

- **batch_inputs** (*ForwardInputs*) – Dict containing the necessary information (e.g. noise, num_batches, mode) to generate image.
- **data_samples** (*Optional[list]*) – Data samples collated by `data_preprocessor`. Defaults to None.
- **mode** (*Optional[str]*) – *mode* is not used in *ProgressiveGrowingGAN*. Defaults to None.

Returns A list of `DataSample` contain generated results.

Return type SampleList

train_discriminator(*inputs: torch.Tensor*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*Tensor*) – Inputs from current resolution training.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*, *fake_data: torch.Tensor*, *real_data: torch.Tensor*) → Tuple[torch.Tensor, dict]

Get disc loss. PGGAN use WGAN-GP's loss and discriminator shift loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.
- **fake_data** (*Tensor*) – Generated images, used to calculate gradient penalty.
- **real_data** (*Tensor*) – Real images, used to calculate gradient penalty.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

train_generator(*inputs: torch.Tensor*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*Tensor*) – Inputs from current resolution training.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple[torch.Tensor, dict]

Generator loss for PGGAN. PGGAN use WGAN's loss to train the generator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **recon_imgs** (*Tensor*) – Reconstructive images.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

Train step function.

This function implements the standard training iteration for asynchronous adversarial training. Namely, in each iteration, we first update discriminator and then compute loss for generator with the newly updated discriminator.

As for distributed training, we use the `reducer` from ddp to synchronize the necessary params in current computational graph.

Parameters

- **data_batch** (*dict*) – Input data from dataloader.
- **optimizer** (*dict*) – Dict contains optimizer for generator and discriminator.

- **ddp_reducer** (Reducer | None, optional) – Reducer from ddp. It is used to prepare for `backward()` in ddp. Defaults to None.
- **running_status** (*dict | None, optional*) – Contains necessary basic information for training, e.g., iteration number. Defaults to None.

Returns Contains 'log_vars', 'num_samples', and 'results'.

Return type dict

class mmagic.models.editors.**Pix2Pix**(**args*, ***kwargs*)

Bases: *mmagic.models.base_models.BaseTranslationModel*

Pix2Pix model for paired image-to-image translation.

Ref: Image-to-Image Translation with Conditional Adversarial Networks

forward_test(*img*, *target_domain*, ***kwargs*)

Forward function for testing.

Parameters

- **img** (*tensor*) – Input image tensor.
- **target_domain** (*str*) – Target domain of output image.
- **kwargs** (*dict*) – Other arguments.

Returns Forward results.

Return type dict

_get_disc_loss(*outputs*)

Get the loss of discriminator.

Parameters **outputs** (*dict*) – A dict of output.

Returns Loss and a dict of log of loss terms.

Return type Tuple

_get_gen_loss(*outputs*)

Get the loss of generator.

Parameters **outputs** (*dict*) – A dict of output.

Returns Loss and a dict of log of loss terms.

Return type Tuple

train_step(*data*, *optim_wrapper=None*)

Training step function.

Parameters

- **data_batch** (*dict*) – Dict of the input data batch.
- **optimizer** (*dict[torch.optim.Optimizer]*) – Dict of optimizers for the generator and discriminator.
- **ddp_reducer** (Reducer | None, optional) – Reducer from ddp. It is used to prepare for `backward()` in ddp. Defaults to None.
- **running_status** (*dict | None, optional*) – Contains necessary basic information for training, e.g., iteration number. Defaults to None.

Returns Dict of loss, information for logger, the number of samples and results for visualization.

Return type dict

`test_step`(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as `val_step()`.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type List[DataSample]

`val_step`(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as `val_step()`.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type List[DataSample]

class `mmagic.models.editors.PlainDecoder`(*in_channels*, *init_cfg: Optional[dict] = None*)

Bases: `mmengine.model.BaseModule`

Simple decoder from Deep Image Matting.

Parameters

- **in_channels** (*int*) – Channel num of input features.
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

`init_weights`()

Init weights for the module.

`forward`(*inputs*)

Forward function of PlainDecoder.

Parameters **inputs** (*dict*) – Output dictionary of the VGG encoder containing:

- out (Tensor): Output of the VGG encoder.
- max_idx_1 (Tensor): Index of the first maxpooling layer in the VGG encoder.
- max_idx_2 (Tensor): Index of the second maxpooling layer in the VGG encoder.
- max_idx_3 (Tensor): Index of the third maxpooling layer in the VGG encoder.
- max_idx_4 (Tensor): Index of the fourth maxpooling layer in the VGG encoder.
- max_idx_5 (Tensor): Index of the fifth maxpooling layer in the VGG encoder.

Returns Output tensor.

Return type Tensor

class `mmagic.models.editors.PlainRefiner`(*conv_channels=64*, *init_cfg=None*)

Bases: `mmengine.model.BaseModule`

Simple refiner from Deep Image Matting.

Parameters

- **conv_channels** (*int*) – Number of channels produced by the three main convolutional layer. Default: 64.
- **pretrained** (*str*) – Name of pretrained model. Default: None.

init_weights()

Init weights for the module.

forward(*x*, *raw_alpha*)

Forward function.

Parameters

- **x** (*Tensor*) – The input feature map of refiner.
- **raw_alpha** (*Tensor*) – The raw predicted alpha matte.

Returns The refined alpha matte.

Return type Tensor

class mmagic.models.editors.**RDNNet**(*in_channels*, *out_channels*, *mid_channels=64*, *num_blocks=16*, *upscale_factor=4*, *num_layers=8*, *channel_growth=64*)

Bases: mmengine.model.BaseModule

RDN model for single image super-resolution.

Paper: Residual Dense Network for Image Super-Resolution

Adapted from 'https://github.com/yjn870/RDN-pytorch.git' 'RDN-pytorch/blob/master/models.py' Copyright (c) 2021, JaeYun Yeo, under MIT License.

Most of the implementation follows the implementation in: 'https://github.com/sanghyun-son/EDSR-PyTorch.git' 'EDSR-PyTorch/blob/master/src/model/rdn.py' Copyright (c) 2017, sanghyun-son, under MIT license.

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64.
- **num_blocks** (*int*) – Block number in the trunk network. Default: 16.
- **upscale_factor** (*int*) – Upsampling factor. Support 2^n and 3. Default: 4.
- **num_layer** (*int*) – Layer number in the Residual Dense Block. Default: 8.
- **channel_growth** (*int*) – Channels growth in each layer of RDB. Default: 64.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**RealBasicVSR**(*generator*, *discriminator=None*, *gan_loss=None*, *pixel_loss=None*, *cleaning_loss=None*, *perceptual_loss=None*, *is_use_sharpened_gt_in_pixel=False*, *is_use_sharpened_gt_in_percep=False*, *is_use_sharpened_gt_in_gan=False*, *is_use_ema=False*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: `mmagic.models.editors.real_esrgan.RealESRGAN`

RealBasicVSR model for real-world video super-resolution.

Ref: Investigating Tradeoffs in Real-World Video Super-Resolution, arXiv

Parameters

- **generator** (*dict*) – Config for the generator.
- **discriminator** (*dict, optional*) – Config for the discriminator. Default: None.
- **gan_loss** (*dict, optional*) – Config for the gan loss. Note that the loss weight in gan loss is only for the generator.
- **pixel_loss** (*dict, optional*) – Config for the pixel loss. Default: None.
- **cleaning_loss** (*dict, optional*) – Config for the image cleaning loss. Default: None.
- **perceptual_loss** (*dict, optional*) – Config for the perceptual loss. Default: None.
- **is_use_sharpened_gt_in_pixel** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for pixel loss. Default: False.
- **is_use_sharpened_gt_in_percep** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for perceptual loss. Default: False.
- **is_use_sharpened_gt_in_gan** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for adversarial loss. Default: False.
- **train_cfg** (*dict*) – Config for training. Default: None. You may change the training of gan by setting: *disc_steps*: how many discriminator updates after one generate update; *disc_init_steps*: how many discriminator updates at the start of the training. These two keys are useful when training with WGAN.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Default: None.

extract_gt_data(*data_samples*)

extract gt data from data samples.

Parameters **data_samples** (*list*) – List of DataSample.

Returns Extract gt data.

Return type Tensor

g_step(*batch_outputs*, *batch_gt_data*)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tuple[Tensor]*) – Batch GT data.

Returns Dict of losses.

Return type dict

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step of GAN-based method.

Parameters

- **data** (*List[dict]*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

forward_train(*batch_inputs*, *data_samples=None*)

Forward Train.

Run forward of generator with `return_lqs=True`

Parameters

- **batch_inputs** (*Tensor*) – Batch inputs.
- **data_samples** (*List[DataSample]*) – Data samples of Editing. Default:None

Returns

Result of generator. (outputs, lqs)

Return type Tuple[Tensor]

class mmagic.models.editors.**RealBasicVSRNet**(*mid_channels=64*, *num_propagation_blocks=20*, *num_cleaning_blocks=20*, *dynamic_refine_thres=255*, *spynet_pretrained=None*, *is_fix_cleaning=False*, *is_sequential_cleaning=False*)

Bases: `mmengine.model.BaseModule`

RealBasicVSR network structure for real-world video super-resolution.

Support only x4 upsampling.

Paper: Investigating Tradeoffs in Real-World Video Super-Resolution, arXiv

Parameters

- **mid_channels** (*int, optional*) – Channel number of the intermediate features. Default: 64.
- **num_propagation_blocks** (*int, optional*) – Number of residual blocks in each propagation branch. Default: 20.
- **num_cleaning_blocks** (*int, optional*) – Number of residual blocks in the image cleaning module. Default: 20.
- **dynamic_refine_thres** (*int, optional*) – Stop cleaning the images when the residue is smaller than this value. Default: 255.
- **spynet_pretrained** (*str, optional*) – Pre-trained model path of SPyNet. Default: None.
- **is_fix_cleaning** (*bool, optional*) – Whether to fix the weights of the image cleaning module during training. Default: False.

- **is_sequential_cleaning** (*bool, optional*) – Whether to clean the images sequentially. This is used to save GPU memory, but the speed is slightly slower. Default: False.

forward(*lqs*, *return_lqs=False*)

Forward function for BasicVSR++.

Parameters

- **lqs** (*tensor*) – Input low quality (LQ) sequence with shape (n, t, c, h, w).
- **return_lqs** (*bool*) – Whether to return LQ sequence. Default: False.

Returns Output HR sequence.

Return type Tensor

class mmagic.models.editors.**RealESRGAN**(*generator*, *discriminator=None*, *gan_loss=None*, *pixel_loss=None*, *perceptual_loss=None*, *is_use_sharpened_gt_in_pixel=False*, *is_use_sharpened_gt_in_percep=False*, *is_use_sharpened_gt_in_gan=False*, *is_use_ema=True*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.editors.srgan.SRGAN

Real-ESRGAN model for single image super-resolution.

Ref: Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data, 2021.

Note: generator_ema is realized in EMA_HOOK

Parameters

- **generator** (*dict*) – Config for the generator.
- **discriminator** (*dict, optional*) – Config for the discriminator. Default: None.
- **gan_loss** (*dict, optional*) – Config for the gan loss. Note that the loss weight in gan loss is only for the generator.
- **pixel_loss** (*dict, optional*) – Config for the pixel loss. Default: None.
- **perceptual_loss** (*dict, optional*) – Config for the perceptual loss. Default: None.
- **is_use_sharpened_gt_in_pixel** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for pixel loss. Default: False.
- **is_use_sharpened_gt_in_percep** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for perceptual loss. Default: False.
- **is_use_sharpened_gt_in_gan** (*bool, optional*) – Whether to use the image sharpened by unsharp masking as the GT for adversarial loss. Default: False.
- **is_use_ema** (*bool, optional*) – When to apply exponential moving average on the network weights. Default: True.
- **train_cfg** (*dict*) – Config for training. Default: None. You may change the training of gan by setting: *disc_steps*: how many discriminator updates after one generate update; *disc_init_steps*: how many discriminator updates at the start of the training. These two keys are useful when training with WGAN.
- **test_cfg** (*dict*) – Config for testing. Default: None.

- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Default: None.

forward_tensor(*inputs*, *data_samples=None*, *training=False*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.
- **training** (*bool*) – Whether is training. Default: False.

Returns result of simple forward.

Return type Tensor

g_step(*batch_outputs*, *batch_gt_data*)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tuple[Tensor]*) – Batch GT data.

Returns Dict of losses.

Return type dict

d_step_real(*batch_outputs*, *batch_gt_data: torch.Tensor*)

Real part of D step.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tuple[Tensor]*) – Batch GT data.

Returns Real part of gan_loss for discriminator.

Return type Tensor

d_step_fake(*batch_outputs*, *batch_gt_data*)

Fake part of D step.

Parameters

- **batch_outputs** (*Tensor*) – Output of generator.
- **batch_gt_data** (*Tuple[Tensor]*) – Batch GT data.

Returns Fake part of gan_loss for discriminator.

Return type Tensor

extract_gt_data(*data_samples*)

extract gt data from data samples.

Parameters **data_samples** (*list*) – List of DataSample.

Returns Extract gt data.

Return type Tensor

class mmagic.models.editors.**UNetDiscriminatorWithSpectralNorm**(*in_channels*, *mid_channels=64*, *skip_connection=True*)

Bases: mmengine.model.BaseModule

A U-Net discriminator with spectral normalization.

Parameters

- **in_channels** (*int*) – Channel number of the input.
- **mid_channels** (*int, optional*) – Channel number of the intermediate features. Default: 64.
- **skip_connection** (*bool, optional*) – Whether to use skip connection. Default: True.

forward(*img*)

Forward function.

Parameters **img** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**Restormer**(*inp_channels=3*, *out_channels=3*, *dim=48*, *num_blocks=[4, 6, 6, 8]*, *num_refinement_blocks=4*, *heads=[1, 2, 4, 8]*, *ffn_expansion_factor=2.66*, *bias=False*, *LayerNorm_type='WithBias'*, *dual_pixel_task=False*, *dual_keys=['imgL', 'imgR']*)

Bases: mmengine.model.BaseModule

Restormer A PyTorch impl of: *Restormer: Efficient Transformer for High- Resolution Image Restoration*. Ref repo: https://github.com/swz30/Restormer.

Parameters

- **inp_channels** (*int*) – Number of input image channels. Default: 3.
- **out_channels** (*int*) – Number of output image channels: 3.
- **dim** (*int*) – Number of feature dimension. Default: 48.
- **num_blocks** (*List(int)*) – Depth of each Transformer layer. Default: [4, 6, 6, 8].
- **num_refinement_blocks** (*int*) – Number of refinement blocks. Default: 4.
- **heads** (*List(int)*) – Number of attention heads in different layers. Default: 7.
- **ffn_expansion_factor** (*float*) – Ratio of feed forward network expansion. Default: 2.66.
- **bias** (*bool*) – The bias of convolution. Default: False
- **LayerNorm_type** (*str|optional*) – Select layer Normalization type. Optional: 'WithBias','BiasFree' Default: 'WithBias'.
- **dual_pixel_task** (*bool*) – True for dual-pixel defocus deblurring only. Also set inp_channels=6. Default: False.
- **dual_keys** (*List*) – Keys of dual images in inputs. Default: ['imgL', 'imgR'].

forward(*inp_img*)

Forward function.

Parameters **inp_img** (*Tensor*) – Input tensor with shape (B, C, H, W).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**SAGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *noise_size: Optional[int] = 128*, *num_classes: Optional[int] = None*, *ema_config: Optional[Dict] = None*)

Bases: *mmagic.models.base_models.BaseConditionalGAN*

Implementation of *Self-Attention Generative Adversarial Networks*.

<https://arxiv.org/abs/1805.08318>`_ (SAGAN), Spectral Normalization for Generative Adversarial Networks (SNGAN), and cGANs with Projection Discriminator (Proj-GAN).

Detailed architecture can be found in `SNGANGenerator` and `ProjDiscriminator`

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or `DataPreprocessor`.
- **generator_steps** (*int*) – Number of times the generator was completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – Number of times the discriminator was completely updated before the generator is updated. Defaults to 1.
- **noise_size** (*Optional[int]*) – Size of the input noise vector. Default to 128.
- **num_classes** (*Optional[int]*) – The number classes you would like to generate. Defaults to None.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple[torch.Tensor, dict]

Get disc loss. SAGAN, SNGAN and Proj-GAN use hinge loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple[torch.Tensor, dict]

Get disc loss. SAGAN, SNGAN and Proj-GAN use hinge loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

class mmagic.models.editors.**SinGAN**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *num_scales: Optional[int] = None*, *iters_per_scale: int = 2000*, *noise_weight_init: int = 0.1*, *lr_scheduler_args: Optional[dict] = None*, *test_pkl_data: Optional[str] = None*, *ema_confg: Optional[dict] = None*)

Bases: *mmagic.models.base_models.BaseGAN*

SinGAN.

This model implement the single image generative adversarial model proposed in: Singan: Learning a Generative Model from a Single Natural Image, ICCV'19.

Notes for training:

- This model should be trained with our dataset `SinGANDataset`.
- In training, the `total_iters` arguments is related to the number of scales in the image pyramid and `iters_per_scale` in the `train_cfg`. You should set it carefully in the training config file.

Notes for model architectures:

- The generator and discriminator need `num_scales` in initialization. However, this arguments is generated by `create_real_pyramid` function from the `singan_dataset.py`. The last element in the returned list (`stop_scale`) is the value for `num_scales`. Pay attention that this scale is counted from zero. Please see our tutorial for SinGAN to obtain more details or our standard config for reference.

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or `DataPreprocessor`.
- **generator_steps** (*int*) – The number of times the generator is completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – The number of times the discriminator is completely updated before the generator is updated. Defaults to 1.
- **num_scales** (*int*) – The number of scales/stages in generator/ discriminator. Note that this number is counted from zero, which is the same as the original paper. Defaults to None.
- **iters_per_scale** (*int*) – The training iteration for each resolution scale. Defaults to 2000.
- **noise_weight_init** (*float*) – The initialize weight of fixed noise. Defaults to 0.1
- **lr_scheduler_args** (*Optional[dict]*) – Arguments for learning schedulers. Note that in SinGAN, we use MultiStepLR, which is the same as the original paper. If not passed, no learning schedule will be used. Defaults to None.
- **test_pkl_data** (*Optional[str]*) – The path of pickle file which contains fixed noise and noise weight. This is must for test. Defaults to None.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

load_test_pkl()

Load pickle for test.

_from_numpy(*data: Tuple[list, numpy.ndarray]*) → Tuple[torch.Tensor, List[torch.Tensor]]

Convert input numpy array or list of numpy array to Tensor or list of Tensor.

Parameters **data** (*Tuple[list, np.ndarray]*) – Input data to convert.

Returns Converted Tensor or list of tensor.

Return type Tuple[Tensor, List[Tensor]]

get_module(*model: torch.nn.Module*, *module_name: str*) → torch.nn.Module

Get an inner module from model.

Since we will wrapper DDP for some model, we have to judge whether the module can be indexed directly.

Parameters

- **model** (*nn.Module*) – This model may wrapped with DDP or not.
- **module_name** (*str*) – The name of specific module.

Returns Returned sub module.

Return type nn.Module

construct_fixed_noises()

Construct the fixed noises list used in SinGAN.

forward(*inputs: mmagic.utils.ForwardInputs*, *data_samples: Optional[list] = None*, *mode=None*) → List[*mmagic.structures.DataSample*]

Forward function for SinGAN. For SinGAN, *inputs* should be a dict contains 'num_batches', 'mode' and other input arguments for the generator.

Parameters

- **inputs** (*dict*) – Dict containing the necessary information (e.g., noise, num_batches, mode) to generate image.
- **data_samples** (*Optional[list]*) – Data samples collated by `data_preprocessor`. Defaults to None.
- **mode** (*Optional[str]*) – *mode* is not used in `BaseConditionalGAN`. Defaults to None.

gen_loss(*disc_pred_fake: torch.Tensor*, *recon_imgs: torch.Tensor*) → Tuple[torch.Tensor, dict]

Generator loss for SinGAN. SinGAN use WGAN's loss and MSE loss to train the generator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **recon_imgs** (*Tensor*) – Reconstructive images.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*, *fake_data: torch.Tensor*, *real_data: torch.Tensor*) → Tuple[torch.Tensor, dict]

Get disc loss. SAGAN, SNGAN and Proj-GAN use hinge loss to train the generator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.
- **fake_data** (*Tensor*) – Generated images, used to calculate gradient penalty.
- **real_data** (*Tensor*) – Real images, used to calculate gradient penalty.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

train_generator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_discriminator(*inputs: dict*, *data_samples: List[*mmagic.structures.DataSample*]*, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*List[DataSample]*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_gan(*inputs_dict: dict*, *data_sample: List[*mmagic.structures.DataSample*]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train GAN model. In the training of GAN models, generator and discriminator are updated alternatively. In MMagic's design, *self.train_step* is called with data input. Therefore we always update discriminator, whose updating is relay on real data, and then determine if the generator needs to be updated based on the current number of iterations. More details about whether to update generator can be found in `should_gen_update()`.

Parameters

- **data** (*dict*) – Data sampled from dataloader.
- **data_sample** (*List[DataSample]*) – List of data sample contains GT and meta information.
- **optim_wrapper** (*OptimWrapperDict*) – OptimWrapperDict instance contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step for SinGAN model. SinGAN is trained with multi-resolution images, and each resolution is trained for *:attr:self.iters_per_scale* times.

We initialize the weight and learning rate scheduler of the corresponding module at the start of each resolution's training. At the end of each resolution's training, we update the weight of the noise of current resolution by mse loss between reconstructed image and real image.

Parameters

- **data** (*dict*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapperDict*) – OptimWrapperDict instance contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

test_step(*data: dict*) → mmagic.utils.SampleList

Gets the generated image of given data in test progress. Before generate images, we call *:meth:self.load_test_pkl* to load the fixed noise and current stage of the model from the pickle file.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

class `mmagic.models.editors.SRCNNNet`(*channels=(3, 64, 32, 3)*, *kernel_sizes=(9, 1, 5)*, *upscale_factor=4*)

Bases: `mmengine.model.BaseModule`

SRCNN network structure for image super resolution.

SRCNN has three conv layers. For each layer, we can define the *in_channels*, *out_channels* and *kernel_size*. The input image will first be upsampled with a bicubic upsampler, and then super-resolved in the HR spatial size.

Paper: Learning a Deep Convolutional Network for Image Super-Resolution.

Parameters

- **channels** (*tuple[int]*) – A tuple of channel numbers for each layer including channels of input and output . Default: (3, 64, 32, 3).
- **kernel_sizes** (*tuple[int]*) – A tuple of kernel sizes for each conv layer. Default: (9, 1, 5).
- **upscale_factor** (*int*) – Upsampling factor. Default: 4.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class `mmagic.models.editors.SRGAN`(*generator*, *discriminator=None*, *gan_loss=None*, *pixel_loss=None*, *perceptual_loss=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: *mmagic.models.base_models.BaseEditModel*

SRGAN model for single image super-resolution.

Ref: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.

Parameters

- **generator** (*dict*) – Config for the generator.
- **discriminator** (*dict*) – Config for the discriminator. Default: None.
- **gan_loss** (*dict*) – Config for the gan loss. Note that the loss weight in gan loss is only for the generator.
- **pixel_loss** (*dict*) – Config for the pixel loss. Default: None.
- **perceptual_loss** (*dict*) – Config for the perceptual loss. Default: None.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Default: None.

forward_train(*inputs*, *data_samples=None*, ***kwargs*)

Forward training. Losses of training is calculated in train_step.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.

Returns Result of `forward_tensor` with `training=True`.

Return type Tensor

forward_tensor(*inputs*, *data_samples=None*, *training=False*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.
- **training** (*bool*) – Whether is training. Default: False.

Returns result of simple forward.

Return type Tensor

if_run_g()

Calculates whether need to run the generator step.

if_run_d()

Calculates whether need to run the discriminator step.

g_step(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

d_step_real(*batch_outputs*, *batch_gt_data: torch.Tensor*)

Real part of D step.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Real part of gan_loss for discriminator.

Return type Tensor

d_step_fake(*batch_outputs: torch.Tensor*, *batch_gt_data*)

Fake part of D step.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Fake part of gan_loss for discriminator.

Return type Tensor

g_step_with_optim(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

G step with optim of GAN: Calculate losses of generator and run optim.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.
- **optim_wrapper** (*OptimWrapperDict*) – Optim wrapper dict.

Returns Dict of parsed losses.

Return type dict

d_step_with_optim(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

D step with optim of GAN: Calculate losses of discriminator and run optim.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.
- **optim_wrapper** (*OptimWrapperDict*) – Optim wrapper dict.

Returns Dict of parsed losses.

Return type dict

extract_gt_data(*data_samples*)

extract gt data from data samples.

Parameters **data_samples** (*list*) – List of DataSample.

Returns Extract gt data.

Return type Tensor

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step of GAN-based method.

Parameters

- **data** (*List[dict]*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

class mmagic.models.editors.**ModifiedVGG**(*in_channels*, *mid_channels*)

Bases: mmengine.model.BaseModule

A modified VGG discriminator with input size 128 x 128.

It is used to train SRGAN and ESRGAN.

Parameters

- **in_channels** (*int*) – Channel number of inputs. Default: 3.
- **mid_channels** (*int*) – Channel number of base intermediate features. Default: 64.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**MSRResNet**(*in_channels*, *out_channels*, *mid_channels=64*, *num_blocks=16*, *upscale_factor=4*)

Bases: mmengine.model.BaseModule

Modified SRResNet.

A compacted version modified from SRResNet in "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network".

It uses residual blocks without BN, similar to EDSR. Currently, it supports x2, x3 and x4 upsampling scale factor.

Parameters

- **in_channels** (*int*) – Channel number of inputs.
- **out_channels** (*int*) – Channel number of outputs.
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64.
- **num_blocks** (*int*) – Block number in the trunk network. Default: 16.
- **upscale_factor** (*int*) – Upsampling factor. Support x2, x3 and x4. Default: 4.

_supported_upscale_factors = [2, 3, 4]

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

init_weights()

Init weights for models.

class mmagic.models.editors.**StableDiffusion**(*vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: Optional[str] = None*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor: Optional[ModelType] = dict(type='DataPreprocessor')*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Class for Stable Diffusion. Refers to https://github.com/Stability- AI/stablediffusion and https://github.com/huggingface/diffusers/blob/main/s rc/diffusers/pipelines/stable_diffusion/pipeline_stable_diffusion_attend_and_excite.py # noqa.

Parameters

- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model This argument will not work when dtype is defined for submodels. Defaults to None.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise Defaults to 0.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.

property device

set_xformers(*module: Optional[torch.nn.Module] = None*) → torch.nn.Module

Set xformers for the model.

Returns The model with xformers.

Return type nn.Module

set_tomesd() → torch.nn.Module

Set ToMe for the stable diffusion model.

Returns The model with ToMe.

Return type nn.Module

train(*mode: bool = True*)

Set train/eval mode.

Parameters **mode** (*bool, optional*) – Whether set train mode. Defaults to True.

infer(*prompt: Union[str, List[str]]*, *height: Optional[int] = None*, *width: Optional[int] = None*, *num_inference_steps: int = 50*, *guidance_scale: float = 7.5*, *negative_prompt: Optional[Union[str, List[str]]] = None*, *num_images_per_prompt: Optional[int] = 1*, *eta: float = 0.0*, *generator: Optional[torch.Generator] = None*, *latents: Optional[torch.FloatTensor] = None*, *show_progress=True*, *seed=1*, *return_type='image'*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str* or *List[str]*) – The prompt or prompts to guide the image generation.
- **(int** (*height*) – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The height in pixels of the generated image.

Parameters

- **(int** (*width*) – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The width in pixels of the generated image.

Parameters

- **num_inference_steps** (*int*, *optional*, defaults to 50) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
- **guidance_scale** (*float*, *optional*, defaults to 7.5) – Guidance scale as defined in [Classifier-Free Diffusion Guidance] (https://arxiv.org/abs/2207.12598).
- **negative_prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).
- **num_images_per_prompt** (*int*, *optional*, defaults to 1) – The number of images to generate per prompt.
- **eta** (*float*, *optional*, defaults to 0.0) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to [*schedulers.DDIMScheduler*], will be ignored for others.
- **generator** (*torch.Generator*, *optional*) – A [torch generator] to make generation deterministic.
- **latents** (*torch.FloatTensor*, *optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*.

- **return_type** (*str*) – The return type of the inference results. Supported types are 'image', 'numpy', 'tensor'. If 'image' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

Returns A dict containing the generated images.

Return type dict

output_to_pil(*image*) → List[PIL.Image.Image]

Convert output tensor to PIL image. Output tensor will be de-normed to [0, 255] by *DataPreprocessor.destruct*. Due to no *data_samples* is passed, color order conversion will not be performed.

Parameters **image** (*torch.Tensor*) – The output tensor of the decoder.

Returns The list of processed PIL images.

Return type List[Image.Image]

_encode_prompt(*prompt*, *device*, *num_images_per_prompt*, *do_classifier_free_guidance*, *negative_prompt*)

Encodes the prompt into text encoder hidden states.

Parameters

- **prompt** (*str or list(int)*) – prompt to be encoded.
- **device** – (torch.device): torch device.
- **num_images_per_prompt** (*int*) – number of images that should be generated per prompt.
- **do_classifier_free_guidance** (*bool*) – whether to use classifier free guidance or not.
- **negative_prompt** (*str or List[str]*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).

Returns

text embeddings generated by clip text encoder.

Return type text_embeddings (torch.Tensor)

decode_latents(*latents*)

use vae to decode latents.

Parameters **latents** (*torch.Tensor*) – latents to decode.

Returns image result.

Return type image (torch.Tensor)

prepare_extra_step_kwargs(*generator*, *eta*)

prepare extra kwargs for the scheduler step.

Parameters

- **generator** (*torch.Generator*) – generator for random functions.
- **eta** (*float*) – eta () is only used with the DDIMScheduler, it will be ignored for other schedulers. eta corresponds to in DDIM paper: https://arxiv.org/abs/2010.02502 and should be between [0, 1]

Returns dict contains 'generator' and 'eta'

Return type extra_step_kwargs (dict)

prepare_test_scheduler_extra_step_kwargs(*generator*, *eta*)

prepare extra kwargs for the scheduler step.

Parameters

- **generator** (*torch.Generator*) – generator for random functions.
- **eta** (*float*) – eta () is only used with the DDIMScheduler, it will be ignored for other schedulers. eta corresponds to in DDIM paper: https://arxiv.org/abs/2010.02502 and should be between [0, 1]

Returns dict contains 'generator' and 'eta'

Return type extra_step_kwargs (dict)

check_inputs(*prompt*, *height*, *width*)

check whether inputs are in suitable format or not.

prepare_latents(*batch_size*, *num_channels_latents*, *height*, *width*, *dtype*, *device*, *generator*, *latents=None*)

prepare latents for diffusion to run in latent space.

Parameters

- **batch_size** (*int*) – batch size.
- **num_channels_latents** (*int*) – latent channel nums.
- **height** (*int*) – image height.
- **width** (*int*) – image width.
- **dtype** (*torch.dtype*) – float type.
- **device** (*torch.device*) – torch device.
- **generator** (*torch.Generator*) – generator for random functions, defaults to None.
- **latents** (*torch.Tensor*) – Pre-generated noisy latents, defaults to None.

Returns prepared latents.

Return type latents (torch.Tensor)

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the predictions of given data.

Calls `self.data_preprocessor(data, False)` and `self(inputs, data_sample, mode='predict')` in order. Return the predictions which will be passed to evaluator.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns The predictions of given data.

Return type list

test_step(*data: dict*) → mmagic.utils.typing.SampleList

`BaseModel` implements `test_step` the same as `val_step`.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns The predictions of given data.

Return type list

train_step(*data*, *optim_wrapper_dict*)

Implements the default model training process including preprocessing, model forward propagation, loss calculation, optimization, and back-propagation.

During non-distributed training. If subclasses do not override the *train_step()*, EpochBasedTrainLoop or IterBasedTrainLoop will call this method to update model parameters. The default parameter update process is as follows:

1. Calls `self.data_processor(data, training=False)` to collect batch_inputs and corresponding data_samples(labels).
2. Calls `self(batch_inputs, data_samples, mode='loss')` to get raw loss
3. Calls `self.parse_losses` to get `parsed_losses` tensor used to backward and dict of loss tensor used to log messages.
4. Calls `optim_wrapper.update_params(loss)` to update model.

Parameters

- **data** (*dict or tuple or list*) – Data sampled from dataset.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

abstract forward(*inputs: torch.Tensor*, *data_samples: Optional[list] = None*, *mode: str = 'tensor'*) → Union[Dict[str, torch.Tensor], list]

forward is not implemented now.

class mmagic.models.editors.**StableDiffusionInpaint**(**args*, ***kwargs*)

Bases: `mmagic.models.editors.stable_diffusion.stable_diffusion.StableDiffusion`

Class for Stable Diffusion. Refers to https://github.com/Stability- AI/stablediffusion and https://github.com/huggingface/diffusers/blob/main/s rc/diffusers/pipelines/stable_diffusion/pipeline_stable_diffusion_attend_and_excite.py # noqa.

Parameters

- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model This argument will not work when dtype is defined for submodels. Defaults to None.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise Defaults to 0.

- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule.

infer(*prompt: Union[str, List[str]], image: Union[torch.FloatTensor, PIL.Image.Image] = None, mask_image: Union[torch.FloatTensor, PIL.Image.Image] = None, height: Optional[int] = None, width: Optional[int] = None, num_inference_steps: int = 50, guidance_scale: float = 7.5, negative_prompt: Optional[Union[str, List[str]]] = None, num_images_per_prompt: Optional[int] = 1, eta: float = 0.0, generator: Optional[torch.Generator] = None, latents: Optional[torch.FloatTensor] = None, show_progress=True, seed=1, return_type='image'*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str* or *List[str]*) – The prompt or prompts to guide the image generation.
- **image** (*Union[torch.FloatTensor, Image.Image]*) – The image to inpaint.
- **mask_image** (*Union[torch.FloatTensor, Image.Image]*) – The mask to apply to the image, i.e. regions to inpaint.
- **(int** (*height*) – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The height in pixels of the generated image.

Parameters

- **(int** (*width*) – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The width in pixels of the generated image.

Parameters

- **num_inference_steps** (*int*, *optional*, defaults to 50) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
- **guidance_scale** (*float*, *optional*, defaults to 7.5) – Guidance scale as defined in [Classifier-Free Diffusion Guidance] (https://arxiv.org/abs/2207.12598).
- **negative_prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).
- **num_images_per_prompt** (*int*, *optional*, defaults to 1) – The number of images to generate per prompt.
- **eta** (*float*, *optional*, defaults to 0.0) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to [*schedulers.DDIMScheduler*], will be ignored for others.

- **generator** (*torch.Generator*, *optional*) – A [torch generator] to make generation deterministic.
- **latents** (*torch.FloatTensor*, *optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*.
- **return_type** (*str*) – The return type of the inference results. Supported types are 'image', 'numpy', 'tensor'. If 'image' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

Returns A dict containing the generated images.

Return type dict

prepare_mask_latents(*mask*, *masked_image*, *batch_size*, *num_channels_latents*, *height*, *width*, *dtype*, *device*, *generator*, *do_classifier_free_guidance*)

prepare latents for diffusion to run in latent space.

Parameters

- **mask** (*torch.Tensor*) – The mask to apply to the image, i.e. regions to inpaint.
- **image** (*torch.Tensor*) – The image to be masked.
- **batch_size** (*int*) – batch size.
- **num_channels_latents** (*int*) – latent channel nums.
- **height** (*int*) – image height.
- **width** (*int*) – image width.
- **dtype** (*torch.dtype*) – float type.
- **device** (*torch.device*) – torch device.
- **generator** (*torch.Generator*) – generator for random functions, defaults to None.
- **latents** (*torch.Tensor*) – Pre-generated noisy latents, defaults to None.
- **do_classifier_free_guidance** (*bool*) – Whether to apply classifier-free guidance.

Returns prepared latents.

Return type latents (torch.Tensor)

abstract val_step(*data: dict*) → mmagic.utils.typing.SampleList

Performs a validation step on the provided data.

This method is decorated with *torch.no_grad()* which indicates no gradients will be computed during the operations. This ensures efficient memory usage during testing.

Parameters **data** (*dict*) – Dictionary containing input data for testing.

Returns List of samples processed during the testing step.

Return type SampleList

Raises **NotImplementedError** – This method has not been implemented.

abstract test_step(*data: dict*) → mmagic.utils.typing.SampleList

Performs a testing step on the provided data.

This method is decorated with *torch.no_grad()* which indicates no gradients will be computed during the operations. This ensures efficient memory usage during testing.

Parameters **data** (*dict*) – Dictionary containing input data for testing.

Returns List of samples processed during the testing step.

Return type SampleList

Raises **NotImplementedError** – This method has not been implemented.

abstract train_step(*data*, *optim_wrapper_dict*)

Performs a training step on the provided data.

Parameters

- **data** – Input data for training.
- **optim_wrapper_dict** – Dictionary containing optimizer wrappers which may contain optimizers, schedulers, etc. required for the training step.

Raises **NotImplementedError** – This method has not been implemented.

class mmagic.models.editors.**StableDiffusionXL**(*vae: ModelType*, *text_encoder_one: ModelType*, *tokenizer_one: str*, *text_encoder_two: ModelType*, *tokenizer_two: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: Optional[str] = None*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor: Optional[ModelType] = dict(type='DataPreprocessor')*, *lora_config: Optional[dict] = None*, *val_prompts: Union[str, List[str]] = None*, *finetune_text_encoder: bool = False*, *force_zeros_for_empty_prompt: bool = True*, *init_cfg: Optional[dict] = None*)

Bases: mmengine.model.BaseModel

Class for Stable Diffusion XL. Refers to https://github.com/Stability- AI.

/generative-models and https://github.com/huggingface/diffusers/blob/main/src/diffusers/pipelines/stable_diffusion_xl/pipeline_stable_diffusion_xl.py

Parameters

- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **text_encoder_one** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer_one** (*str*) – The **name** for CLIP tokenizer.
- **text_encoder_two** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer_two** (*str*) – The **name** for CLIP tokenizer.
- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.

- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model This argument will not work when dtype is defined for submodels. Defaults to None.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise Defaults to 0.
- **tomesd_cfg** (*dict, optional*) – The config for TOMESD. Please refers to https://github.com/dbolya/tomesd and https://github.com/open-mmlab/mmagic/blob/main/mmagic/models/utils/tome_utils.py for detail. # noqa Defaults to None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.
- **lora_config** (*dict, optional*) – The config for LoRA finetuning. Defaults to None.
- **val_prompts** (*Union[str, List[str]], optional*) – The prompts for validation. Defaults to None.
- **finetune_text_encoder** (*bool, optional*) – Whether to fine-tune text encoder. Defaults to False.
- **force_zeros_for_empty_prompt** (*bool*) – Whether the negative prompt embeddings shall be forced to always be set to 0. Defaults to True.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`.

property device

prepare_model()

Prepare model for training.

Move model to target dtype and disable gradient for some models.

set_lora()

Set LORA for model.

set_xformers(*module: Optional[torch.nn.Module] = None*) → torch.nn.Module

Set xformers for the model.

Returns The model with xformers.

Return type nn.Module

set_tomesd() → torch.nn.Module

Set ToMe for the stable diffusion model.

Returns The model with ToMe.

Return type nn.Module

train(*mode: bool = True*)

Set train/eval mode.

Parameters **mode** (*bool, optional*) – Whether set train mode. Defaults to True.

infer(*prompt: Union[str, List[str]]*, *prompt_2: Optional[Union[str, List[str]]] = None*, *height: Optional[int] = None*, *width: Optional[int] = None*, *num_inference_steps: int = 50*, *denoising_end: Optional[float] = None*, *guidance_scale: float = 7.5*, *negative_prompt: Optional[Union[str, List[str]]] = None*, *negative_prompt_2: Optional[Union[str, List[str]]] = None*, *num_images_per_prompt: Optional[int] = 1*, *eta: float = 0.0*, *generator: Optional[torch.Generator] = None*, *latents: Optional[torch.FloatTensor] = None*, *show_progress: bool = True*, *seed: int = 1*, *original_size: Optional[Tuple[int, int]] = None*, *crops_coords_top_left: Tuple[int, int] = (0, 0)*, *target_size: Optional[Tuple[int, int]] = None*, *negative_original_size: Optional[Tuple[int, int]] = None*, *negative_crops_coords_top_left: Tuple[int, int] = (0, 0)*, *negative_target_size: Optional[Tuple[int, int]] = None*, *return_type='image'*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str* or *List[str]*) – The prompt or prompts to guide the image generation.
- **prompt2** (*str* or *List[str]*, *optional*) – The prompt or prompts to be sent to the *tokenizer_two* and *text_encoder_two*. If not defined, *prompt* is used in both text-encoders. Defaults to None.
- **(int** (*height*) – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The height in pixels of the generated image.

Parameters

- **(int** (*width*) – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The width in pixels of the generated image.

Parameters

- **num_inference_steps** (*int*, *optional*, defaults to 50) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
- **denoising_end** (*float*, *optional*) – When specified, determines the fraction (between 0.0 and 1.0) of the total denoising process to be completed before it is intentionally prematurely terminated. As a result, the returned sample will still retain a substantial amount of noise as determined by the discrete timesteps selected by the scheduler. The denoising_end parameter should ideally be utilized when this pipeline forms a part of a "Mixture of Denoisers" multi-pipeline setup, as elaborated in [**Refining the Image Output**](https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/stable_diffusion_xl#refining-the-image-output)
- **guidance_scale** (*float*, *optional*, defaults to 7.5) – Guidance scale as defined in [Classifier-Free Diffusion Guidance] (https://arxiv.org/abs/2207.12598).

- **negative_prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).
- **negative_prompt_2** (*str* or *List[str]*, *optional*)) – The *negative_prompt* to be sent to the *tokenizer_two* and *text_encoder_two*. If not defined, *negative_prompt* is used in both text-encoders. Defaults to None.
- **num_images_per_prompt** (*int*, *optional*, defaults to 1) – The number of images to generate per prompt.
- **eta** (*float*, *optional*, defaults to 0.0) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to [*schedulers.DDIMScheduler*], will be ignored for others.
- **generator** (*torch.Generator*, *optional*) – A [torch generator] to make generation deterministic.
- **latents** (*torch.FloatTensor*, *optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*.
- **show_progress** (*bool*) – Whether to show progress. Defaults to False.
- **seed** (*int*) – Seed to be used. Defaults to 1.
- **original_size** (*Tuple[int]*, *optional*) – If *original_size* is not the same as *target_size* the image will appear to be down- or upsampled. If *original_size* is *(width, height)* if not specified. Defaults to None.
- **crops_coords_top_left** (*Tuple[int]*, *optional*) – *crops_coords_top_left* can be used to generate an image that appears to be "cropped" from the position. Favorable, well-centered images are usually achieved by setting *crops_coords_top_left* to (0, 0). Defaults to (0, 0).
- **target_size** (*Tuple[int]*, *optional*) – For most cases, *target_size* should be set to the desired height and width of the generated image. If not specified it will be *(width, height)*. Defaults to None.
- **negative_original_size** (*Tuple[int]*, *optional*) – To negatively condition the generation process based on a specific image resolution. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208. Defaults to None.
- **negative_crops_coords_top_left** (*Tuple[int]*, *optional*) – To negatively condition the generation process based on a specific crop coordinates. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208. Defaults to (0, 0).
- **negative_target_size** (*Tuple[int]*, *optional*) – To negatively condition the generation process based on a target image resolution. It should be as same as the *target_size* for most cases. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208. Defaults to None.
- **return_type** (*str*) – The return type of the inference results. Supported types are 'image', 'numpy', 'tensor'. If 'image' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

Returns A dict containing the generated images.

Return type dict

_get_add_time_ids(*original_size: Optional[Tuple[int, int]]*, *crops_coords_top_left: Tuple[int, int]*, *target_size: Optional[Tuple[int, int]]*, *dtype*)

Get *add_time_ids*.

Parameters

- **original_size** (*Tuple[int]*, *optional*) – If *original_size* is not the same as *target_size* the image will appear to be down- or upsampled. If *original_size* is *(width, height)* if not specified. Defaults to None.
- **crops_coords_top_left** (*Tuple[int]*, *optional*) – *crops_coords_top_left* can be used to generate an image that appears to be "cropped" from the position. Favorable, well-centered images are usually achieved by setting *crops_coords_top_left* to (0, 0). Defaults to (0, 0).
- **target_size** (*Tuple[int]*, *optional*) – For most cases, *target_size* should be set to the desired height and width of the generated image. If not specified it will be *(width, height)*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the embeddings.

Returns time ids for time embeddings layer.

Return type add_time_ids (torch.Tensor)

output_to_pil(*image*) → List[PIL.Image.Image]

Convert output tensor to PIL image. Output tensor will be de-normed to [0, 255] by *DataPreprocessor.destruct*. Due to no *data_samples* is passed, color order conversion will not be performed.

Parameters **image** (*torch.Tensor*) – The output tensor of the decoder.

Returns The list of processed PIL images.

Return type List[Image.Image]

_encode_prompt(*prompt*, *prompt_2*, *device*, *num_images_per_prompt*, *do_classifier_free_guidance*, *negative_prompt*, *negative_prompt_2*)

Encodes the prompt into text encoder hidden states.

Parameters

- **prompt** (*str or list(int)*) – prompt to be encoded.
- **prompt_2** (*str or list(int)*) – prompt to be encoded. Send to the *tokenizer_two* and *text_encoder_two*. If not defined, *prompt* is used in both text-encoders.
- **device** – (torch.device): torch device.
- **num_images_per_prompt** (*int*) – number of images that should be generated per prompt.
- **do_classifier_free_guidance** (*bool*) – whether to use classifier free guidance or not.
- **negative_prompt** (*str or List[str]*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).

- **negative_prompt_2** (*str or List[str]*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*). Send to *tokenizer_two* and *text_encoder_two*. If not defined, *negative_prompt* is used in both text-encoders

Returns

text embeddings generated by clip text encoder.

Return type text_embeddings (torch.Tensor)

decode_latents(*latents*)

use vae to decode latents.

Parameters **latents** (*torch.Tensor*) – latents to decode.

Returns image result.

Return type image (torch.Tensor)

prepare_extra_step_kwargs(*generator*, *eta*)

prepare extra kwargs for the scheduler step.

Parameters

- **generator** (*torch.Generator*) – generator for random functions.
- **eta** (*float*) – eta () is only used with the DDIMScheduler, it will be ignored for other schedulers. eta corresponds to in DDIM paper: https://arxiv.org/abs/2010.02502 and should be between [0, 1]

Returns dict contains 'generator' and 'eta'

Return type extra_step_kwargs (dict)

prepare_test_scheduler_extra_step_kwargs(*generator*, *eta*)

prepare extra kwargs for the scheduler step.

Parameters

- **generator** (*torch.Generator*) – generator for random functions.
- **eta** (*float*) – eta () is only used with the DDIMScheduler, it will be ignored for other schedulers. eta corresponds to in DDIM paper: https://arxiv.org/abs/2010.02502 and should be between [0, 1]

Returns dict contains 'generator' and 'eta'

Return type extra_step_kwargs (dict)

check_inputs(*prompt*, *height*, *width*)

check whether inputs are in suitable format or not.

prepare_latents(*batch_size*, *num_channels_latents*, *height*, *width*, *dtype*, *device*, *generator*, *latents=None*)

prepare latents for diffusion to run in latent space.

Parameters

- **batch_size** (*int*) – batch size.
- **num_channels_latents** (*int*) – latent channel nums.
- **height** (*int*) – image height.
- **width** (*int*) – image width.

- **dtype** (*torch.dtype*) – float type.
- **device** (*torch.device*) – torch device.
- **generator** (*torch.Generator*) – generator for random functions, defaults to None.
- **latents** (*torch.Tensor*) – Pre-generated noisy latents, defaults to None.

Returns prepared latents.

Return type latents (torch.Tensor)

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as *val_step()*.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

encode_prompt_train(*text_one*, *text_two*)

Encode prompt for training.

Parameters

- **text_one** (*torch.tensor*) – Input ids from tokenizer_one.
- **text_two** (*torch.tensor*) – Input ids from tokenizer_two.

Returns Prompt embedings. pooled_prompt_embeds (torch.tensor): Pooled prompt embeddings.

Return type prompt_embeds (torch.tensor)

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

Train step function.

Parameters

- **data** (*List[dict]*) – Batch of data as input.
- **optim_wrapper** (*OptimWrapperDict*) – Dict with optimizers for generator and discriminator (if have).

Returns Dict with loss, information for logger, the number of samples and results for visualization.

Return type dict

abstract forward(*inputs: torch.Tensor*, *data_samples: Optional[list] = None*, *mode: str = 'tensor'*) → Union[Dict[str, torch.Tensor], list]

forward is not implemented now.

class mmagic.models.editors.**StyleGAN1**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *style_channels: int = 512*, *nkimgs_per_scale: dict = {}*, *interp_real: Optional[dict] = None*, *transition_kimgs: int = 600*, *prev_stage: int = 0*, *ema_config: Optional[Dict] = None*)

Bases: mmagic.models.editors.pggan.ProgressiveGrowingGAN

Implementation of *A Style-Based Generator Architecture for Generative Adversarial Networks*.

<https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html>`_ # noqa (StyleGANv1). This class is inherited from *ProgressiveGrowingGAN* to support progressive training.

Detailed architecture can be found in StyleGAN1Generator and StyleGAN1Discriminator

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or DataPreprocessor.
- **style_channels** (*int*) – The number of channels for style code. Defaults to 128.
- **nkimgs_per_scale** (*dict*) – The number of images need for each resolution's training. Defaults to *{}*.
- **intep_real** (*dict, optional*) – The config of interpolation method for real images. If not passed, bilinear interpolation with align_corners will be used. Defaults to None.
- **transition_kimgs** (*int, optional*) – The number of images during used to transit from the previous torgb layer to newer torgb layer. Defaults to 600.
- **prev_stage** (*int, optional*) – The resolution of previous stage. Used for resume training. Defaults to 0.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*, *fake_data: torch.Tensor*, *real_data: torch.Tensor*) → Tuple[torch.Tensor, dict]

Get disc loss. StyleGANv1 use non-saturating gan loss and R1 gradient penalty. loss to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.
- **fake_data** (*Tensor*) – Generated images, used to calculate gradient penalty.
- **real_data** (*Tensor*) – Real images, used to calculate gradient penalty.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple[torch.Tensor, dict]

Generator loss for PGGAN. PGGAN use WGAN's loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type Tuple[Tensor, dict]

class mmagic.models.editors.**StyleGAN2**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *ema_config: Optional[Dict] = None*, *loss_config=dict()*)

Bases: *mmagic.models.base_models.BaseGAN*

Implementation of *Analyzing and Improving the Image Quality of Stylegan*. # noqa.

Paper link: https://openaccess.thecvf.com/content_CVPR_2020/html/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.html. # noqa

`StyleGAN2Generator` and `StyleGAN2Discriminator`

Parameters

- **generator** (*ModelType*) – The config or model of the generator.
- **discriminator** (*Optional[ModelType]*) – The config or model of the discriminator. Defaults to None.
- **data_preprocessor** (*Optional[Union[dict, Config]]*) – The pre-process config or `DataPreprocessor`.
- **generator_steps** (*int*) – The number of times the generator is completely updated before the discriminator is updated. Defaults to 1.
- **discriminator_steps** (*int*) – The number of times the discriminator is completely updated before the generator is updated. Defaults to 1.
- **ema_config** (*Optional[Dict]*) – The config for generator's exponential moving average setting. Defaults to None.

disc_loss(*disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*, *real_imgs: torch.Tensor*) → Tuple

Get disc loss. StyleGANv2 use the non-saturating loss and R1 gradient penalty to train the discriminator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.
- **real_imgs** (*Tensor*) – Input real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*, *batch_size: int*) → Tuple

Get gen loss. StyleGANv2 use the non-saturating loss and generator path regularization to train the generator.

Parameters

- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.
- **batch_size** (*int*) – Batch size for generating fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_step(*data: dict*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train GAN model. In the training of GAN models, generator and discriminator are updated alternatively. In MMagic's design, *self.train_step* is called with data input. Therefore we always update discriminator, whose updating is relay on real data, and then determine if the generator needs to be updated based on the current number of iterations. More details about whether to update generator can be found in `should_gen_update()`.

Parameters

- **data** (*dict*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapperDict*) – OptimWrapperDict instance contains OptimWrapper of generator and discriminator.

Returns A `dict` of tensor for logging.

Return type Dict[str, torch.Tensor]

class mmagic.models.editors.**StyleGAN3**(*generator: ModelType*, *discriminator: Optional[ModelType] = None*, *data_preprocessor: Optional[Union[dict, mmengine.Config]] = None*, *generator_steps: int = 1*, *discriminator_steps: int = 1*, *forward_kwargs: Optional[Dict] = None*, *ema_config: Optional[Dict] = None*, *loss_config=dict()*)

Bases: `mmagic.models.editors.stylegan2.StyleGAN2`

Implementation of *Alias-Free Generative Adversarial Networks*. # noqa.

Paper link: https://nvlabs-fi-cdn.nvidia.com/stylegan3/stylegan3-paper.pdf # noqa

Detailed architecture can be found in

`StyleGAN3Generator` and `StyleGAN2Discriminator`

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as `val_step()`.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Same as `val_step()`.

Parameters **data** (*dict*) – Data sampled from metric specific sampler. More details in *Metrics* and *Evaluator*.

Returns A list of `DataSample` contain generated results.

Return type SampleList

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

sample_equivarience_pairs(*batch_size*, *sample_mode='ema'*, *eq_cfg=dict(compute_eqt_int=False, compute_eqt_frac=False, compute_eqr=False, translate_max=0.125, rotate_max=1)*, *sample_kwargs=dict()*)

class mmagic.models.editors.**StyleGAN3Generator**(*out_size*, *style_channels*, *img_channels*, *noise_size=512*, *rgb2bgr=False*, *pretrained=None*, *synthesis_cfg=dict(type='SynthesisNetwork')*, *mapping_cfg=dict(type='MappingNetwork')*)

Bases: `mmengine.model.BaseModule`

StyleGAN3 Generator.

In StyleGAN3, we make several changes to StyleGANv2's generator which include transformed fourier features, filtered nonlinearity and non-critical sampling, etc. More details can be found in: Alias-Free Generative Adversarial Networks NeurIPS'2021.

Ref: https://github.com/NVlabs/stylegan3

Parameters

- **out_size** (*int*) – The output size of the StyleGAN3 generator.
- **style_channels** (*int*) – The number of channels for style code.
- **img_channels** (*int*) – The number of output's channels.
- **noise_size** (*int, optional*) – Size of the input noise vector. Defaults to 512.
- **rgb2bgr** (*bool, optional*) – Whether to reformat the output channels with order *bgr*. We provide several pre-trained StyleGAN3 weights whose output channels order is *rgb*. You can set this argument to True to use the weights.
- **pretrained** (*str | dict, optional*) – Path for the pretrained model or dict containing information for pretrained models whose necessary key is 'ckpt_path'. Besides, you can also provide 'prefix' to load the generator part from the whole state dict. Defaults to None.
- **synthesis_cfg** (*dict, optional*) – Config for synthesis network. Defaults to dict(type='SynthesisNetwork').
- **mapping_cfg** (*dict, optional*) – Config for mapping network. Defaults to dict(type='MappingNetwork').

_load_pretrained_model(*ckpt_path*, *prefix=''*, *map_location='cpu'*, *strict=True*)

forward(*noise*, *num_batches=0*, *input_is_latent=False*, *truncation=1*, *num_truncation_layer=None*, *update_emas=False*, *force_fp32=True*, *return_noise=False*, *return_latents=False*)

Forward Function for stylegan3.

Parameters

- **noise** (*torch.Tensor | callable | None*) – You can directly give a batch of noise through a `torch.Tensor` or offer a callable function to sample a batch of noise data. Otherwise, the `None` indicates to use the default noise sampler.
- **num_batches** (*int, optional*) – The number of batch size. Defaults to 0.
- **input_is_latent** (*bool, optional*) – If *True*, the input tensor is the latent tensor. Defaults to False.
- **truncation** (*float, optional*) – Truncation factor. Give value less than 1., the truncation trick will be adopted. Defaults to 1.

- **num_truncation_layer** (*int, optional*) – Number of layers use truncated latent. Defaults to None.
- **update_emas** (*bool, optional*) – Whether update moving average of mean latent. Defaults to False.
- **force_fp32** (*bool, optional*) – Force fp32 ignore the weights. Defaults to True.
- **return_noise** (*bool, optional*) – If True, `noise_batch` will be returned in a dict with `fake_img`. Defaults to False.
- **return_latents** (*bool, optional*) – If True, `latent` will be returned in a dict with `fake_img`. Defaults to False.

Returns Generated image tensor or dictionary containing more data.

Return type torch.Tensor | dict

get_mean_latent(*num_samples=4096*, ***kwargs*)

Get mean latent of W space in this generator.

Parameters **num_samples** (*int, optional*) – Number of sample times. Defaults to 4096.

Returns Mean latent of this generator.

Return type Tensor

get_training_kwargs(*phase*)

Get training kwargs. In StyleGANv3, we enable fp16, and update magnitude ema during training of discriminator. This function is used to pass related arguments.

Parameters **phase** (*str*) – Current training phase.

Returns Training kwargs.

Return type dict

class mmagic.models.editors.**SwinIRNet**(*img_size=64*, *patch_size=1*, *in_chans=3*, *embed_dim=96*, *depths=[6, 6, 6, 6]*, *num_heads=[6, 6, 6, 6]*, *window_size=7*, *mlp_ratio=4.0*, *qkv_bias=True*, *qk_scale=None*, *drop_rate=0.0*, *attn_drop_rate=0.0*, *drop_path_rate=0.1*, *norm_layer=nn.LayerNorm*, *ape=False*, *patch_norm=True*, *use_checkpoint=False*, *upscale=2*, *img_range=1.0*, *upsampler=''*, *resi_connection='1conv'*, ***kwargs*)

Bases: `mmengine.model.BaseModule`

SwinIR A PyTorch impl of: *SwinIR: Image Restoration Using Swin Transformer*, based on Swin Transformer. Ref repo: https://github.com/JingyunLiang/SwinIR

Parameters

- **img_size** (*int | tuple(int)*) – Input image size. Default 64
- **patch_size** (*int | tuple(int)*) – Patch size. Default: 1
- **in_chans** (*int*) – Number of input image channels. Default: 3
- **embed_dim** (*int*) – Patch embedding dimension. Default: 96
- **depths** (*tuple(int)*) – Depth of each Swin Transformer layer. Default: [6, 6, 6, 6]
- **num_heads** (*tuple(int)*) – Number of attention heads in different layers. Default: [6, 6, 6, 6]
- **window_size** (*int*) – Window size. Default: 7

- **mlp_ratio** (*float*) – Ratio of mlp hidden dim to embedding dim. Default: 4
- **qkv_bias** (*bool*) – If True, add a learnable bias to query, key, value. Default: True
- **qk_scale** (*float*) – Override default qk scale of head_dim ** -0.5 if set. Default: None
- **drop_rate** (*float*) – Dropout rate. Default: 0
- **attn_drop_rate** (*float*) – Attention dropout rate. Default: 0
- **drop_path_rate** (*float*) – Stochastic depth rate. Default: 0.1
- **norm_layer** (*nn.Module*) – Normalization layer. Default: nn.LayerNorm.
- **ape** (*bool*) – If True, add absolute position embedding to the patch embedding. Default: False
- **patch_norm** (*bool*) – If True, add normalization after patch embedding. Default: True
- **use_checkpoint** (*bool*) – Whether to use checkpointing to save memory. Default: False
- **upscale** (*int*) – Upscale factor. 2/3/4/8 for image SR, 1 for denoising and compress artifact reduction. Default: 2
- **img_range** (*float*) – Image range. 1. or 255. Default: 1.0
- **upsampler** (*string, optional*) – The reconstruction module. 'pixelshuffle' / 'pixelshuffledirect' /'nearest+conv'/None. Default: ''
- **resi_connection** (*string*) – The convolutional block before residual connection. '1conv'/'3conv'. Default: '1conv'

_init_weights(*m*)

no_weight_decay()

no_weight_decay_keywords()

check_image_size(*x*)

Check image size and pad images so that it has enough dimension do window size.

Parameters **x** – input tensor image with (B, C, H, W) shape.

forward_features(*x*)

Forward function of Deep Feature Extraction.

Parameters **x** (*Tensor*) – Input tensor with shape (B, C, H, W).

Returns Forward results.

Return type Tensor

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (B, C, H, W).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**TDAN**(*generator*, *pixel_loss*, *lq_pixel_loss*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.BaseEditModel

TDAN model for video super-resolution.

Paper: TDAN: Temporally-Deformable Alignment Network for Video Super- Resolution, CVPR, 2020

Parameters

- **generator** (*dict*) – Config for the generator structure.
- **pixel_loss** (*dict*) – Config for pixel-wise loss.
- **lq_pixel_loss** (*dict*) – Config for pixel-wise loss for the LQ images.
- **train_cfg** (*dict*) – Config for training. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`.

forward_train(*inputs*, *data_samples=None*, ***kwargs*)

Forward training. Returns dict of losses of training.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.

Returns Dict of losses.

Return type dict

forward_tensor(*inputs*, *data_samples=None*, *training=False*, ***kwargs*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by `data_preprocessor`.
- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by `data_preprocessor`.
- **training** (*bool*) – Whether is training. Default: False.

Returns

results of forward inference and forward train.

Return type (Tensor | List[Tensor])

class mmagic.models.editors.**TDANNet**(*in_channels=3*, *mid_channels=64*, *out_channels=3*, *num_blocks_before_align=5*, *num_blocks_after_align=10*)

Bases: `mmengine.model.BaseModule`

TDAN network structure for video super-resolution.

Support only x4 upsampling.

Paper: TDAN: Temporally-Deformable Alignment Network for Video Super- Resolution, CVPR, 2020

Parameters

- **in_channels** (*int*) – Number of channels of the input image. Default: 3.
- **mid_channels** (*int*) – Number of channels of the intermediate features. Default: 64.

- **out_channels** (*int*) – Number of channels of the output image. Default: 3.
- **num_blocks_before_align** (*int*) – Number of residual blocks before temporal alignment. Default: 5.
- **num_blocks_after_align** (*int*) – Number of residual blocks after temporal alignment. Default: 10.

forward(*lrs*)

Forward function for TDANNet.

Parameters **lrs** (*Tensor*) – Input LR sequence with shape (n, t, c, h, w).

Returns Output HR image with shape (n, c, 4h, 4w) and aligned LR images with shape (n, t, c, h, w).

Return type tuple[Tensor]

class mmagic.models.editors.**TextualInversion**(*placeholder_token: str*, *vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *dtype: Optional[str] = None*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *initialize_token: Optional[str] = None*, *num_vectors_per_token: int = 1*, *val_prompts=None*, *data_preprocessor: Optional[ModelType] = dict(type='DataPreprocessor')*, *init_cfg: Optional[dict] = None*)

Bases: mmagic.models.editors.stable_diffusion.stable_diffusion.StableDiffusion

Implementation of `An Image is Worth One Word: Personalizing Text-to- Image Generation using Textual Inversion.

<https://arxiv.org/abs/2208.01618>`_ (Textual Inversion).

Parameters

- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **dtype** (*str, optional*) – The dtype for the model. Defaults to 'fp16'.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise # noqa Defaults to 0.
- **tomesd_cfg** (*dict, optional*) – The config for TOMESD. Please refers to https://github.com/dbolya/tomesd and https://github.com/open-mmlab/mmagic/blob/main/mmagic/models/utils/tome_utils.py for detail. # noqa Defaults to None.

- **initialize_token** (*str, optional*) – The initialization token for textual embedding to train. Defaults to None.
- **num_vefctor_per_token** (*int*) – The length of the learnable embedding. Defaults to 1.
- **val_prompts** (*Union[str, List[str]], optional*) – The prompts for validation. Defaults to None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Defaults to

  dict(type='DataPreprocessor').
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Defaults to None/

prepare_models()

Disable gradient for untrainable modules to save memory.

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

add_tokens(*placeholder_token: str*, *initialize_token: str = None*, *num_vectors_per_token: int = 1*)

Add token for training.

TODO: support add tokens as dict, then we can load pretrained tokens.

train_step(*data*, *optim_wrapper*)

Training step.

class mmagic.models.editors.**TOFlowVFINet**(*rgb_mean=[0.485, 0.456, 0.406]*, *rgb_std=[0.229, 0.224, 0.225]*, *flow_cfg=dict(norm_cfg=None, pretrained=None)*, *init_cfg=None*)

Bases: `mmengine.model.BaseModule`

PyTorch implementation of TOFlow for video frame interpolation.

Paper: Xue et al., Video Enhancement with Task-Oriented Flow, IJCV 2018 Code reference:

1. https://github.com/anchen1011/toflow
2. https://github.com/Coldog2333/pytoflow

Parameters

- **rgb_mean** (*list[float]*) – Image mean in RGB orders. Default: [0.485, 0.456, 0.406]

- **rgb_std** (*list[float]*) – Image std in RGB orders. Default: [0.229, 0.224, 0.225]
- **flow_cfg** (*dict*) – Config of SPyNet. Default: dict(norm_cfg=None, pretrained=None)
- **init_cfg** (*dict, optional*) – Initialization config dict. Default: None.

forward(*imgs*)

Parameters **imgs** – Input frames with shape of (b, 2, 3, h, w).

Returns Interpolated frame with shape of (b, 3, h, w).

Return type Tensor

class mmagic.models.editors.**TOFlowVSRNet**(*adapt_official_weights=False*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

PyTorch implementation of TOFlow.

In TOFlow, the LR frames are pre-upsampled and have the same size with the GT frames.

Paper: Xue et al., Video Enhancement with Task-Oriented Flow, IJCV 2018 Code reference:

1. https://github.com/anchen1011/toflow
2. https://github.com/Coldog2333/pytoflow

Parameters **adapt_official_weights** (*bool*) – Whether to adapt the weights translated from the official implementation. Set to false if you want to train from scratch. Default: False

forward(*lrs*)

Parameters **lrs** – Input lr frames: (b, 7, 3, h, w).

Returns SR frame: (b, 3, h, w).

Return type Tensor

class mmagic.models.editors.**ToFResBlock**

Bases: torch.nn.Module

ResNet architecture.

Three-layers ResNet/ResBlock

forward(*frames*)

Parameters **frames** (*Tensor*) – Tensor with shape of (b, 2, 3, h, w).

Returns Interpolated frame with shape of (b, 3, h, w).

Return type Tensor

class mmagic.models.editors.**LTE**(*requires_grad=True*, *pixel_range=1.0*, *load_pretrained_vgg=True*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

Learnable Texture Extractor.

Based on pretrained VGG19. Generate features in 3 levels.

Parameters

- **requires_grad** (*bool*) – Require grad or not. Default: True.
- **pixel_range** (*float*) – Pixel range of feature. Default: 1.

- **load_pretrained_vgg** (*bool*) – Load pretrained VGG from torchvision. Default: True. Train: must load pretrained VGG. Eval: needn't load pretrained VGG, because we will load pretrained LTE.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, 3, h, w).

Returns

Forward results in 3 levels. x_level3: Forward results in level 3 (n, 256, h/4, w/4). x_level2: Forward results in level 2 (n, 128, h/2, w/2). x_level1: Forward results in level 1 (n, 64, h, w).

Return type Tuple[Tensor]

class mmagic.models.editors.**TTSR**(*generator*, *extractor*, *transformer*, *pixel_loss*, *discriminator=None*, *perceptual_loss=None*, *transferal_perceptual_loss=None*, *gan_loss=None*, *train_cfg=None*, *test_cfg=None*, *init_cfg=None*, *data_preprocessor=None*)

Bases: mmagic.models.editors.srgan.SRGAN

TTSR model for Reference-based Image Super-Resolution.

Paper: Learning Texture Transformer Network for Image Super-Resolution.

Parameters

- **generator** (*dict*) – Config for the generator.
- **extractor** (*dict*) – Config for the extractor.
- **transformer** (*dict*) – Config for the transformer.
- **pixel_loss** (*dict*) – Config for the pixel loss.
- **discriminator** (*dict*) – Config for the discriminator. Default: None.
- **perceptual_loss** (*dict*) – Config for the perceptual loss. Default: None.
- **transferal_perceptual_loss** (*dict*) – Config for the transferal perceptual loss. Default: None.
- **gan_loss** (*dict*) – Config for the GAN loss. Default: None
- **train_cfg** (*dict*) – Config for train. Default: None.
- **test_cfg** (*dict*) – Config for testing. Default: None.
- **init_cfg** (*dict, optional*) – The weight initialized config for BaseModule. Default: None.
- **data_preprocessor** (*dict, optional*) – The pre-process config of BaseDataPreprocessor. Default: None.

forward_tensor(*inputs*, *data_samples=None*, *training=False*)

Forward tensor. Returns result of simple forward.

Parameters

- **inputs** (*torch.Tensor*) – batch input tensor collated by data_preprocessor.

- **data_samples** (*List[BaseDataElement], optional*) – data samples collated by data_preprocessor.
- **training** (*bool*) – Whether is training. Default: False.

Returns

results of forward inference and forward train.

Return type (Tensor | Tuple[List[Tensor]])

if_run_g()

Calculates whether need to run the generator step.

if_run_d()

Calculates whether need to run the discriminator step.

g_step(*batch_outputs*, *batch_gt_data:* mmagic.structures.DataSample)

G step of GAN: Calculate losses of generator.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.

Returns Dict of losses.

Return type dict

g_step_with_optim(*batch_outputs: torch.Tensor*, *batch_gt_data: torch.Tensor*, *optim_wrapper: mmengine.optim.OptimWrapperDict*)

G step with optim of GAN: Calculate losses of generator and run optim.

Parameters

- **batch_outputs** (*Tensor*) – Batch output of generator.
- **batch_gt_data** (*Tensor*) – Batch GT data.
- **optim_wrapper** (*OptimWrapperDict*) – Optim wrapper dict.

Returns Dict of parsed losses.

Return type dict

train_step(*data: List[dict]*, *optim_wrapper: mmengine.optim.OptimWrapperDict*) → Dict[str, torch.Tensor]

Train step of GAN-based method.

Parameters

- **data** (*List[dict]*) – Data sampled from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A dict of tensor for logging.

Return type Dict[str, torch.Tensor]

class mmagic.models.editors.**SearchTransformer**(*init_cfg: Union[dict, List[dict], None] = None*)

Bases: mmengine.model.BaseModule

Search texture reference by transformer.

Include relevance embedding, hard-attention and soft-attention.

gather(*inputs*, *dim*, *index*)

Hard Attention. Gathers values along an axis specified by dim.

Parameters

- **inputs** (*Tensor*) – The source tensor. (N, C*k*k, H*W)
- **dim** (*int*) – The axis along which to index.
- **index** (*Tensor*) – The indices of elements to gather. (N, H*W)

results: outputs (Tensor): The result tensor. (N, C*k*k, H*W)

forward(*img_lq*, *ref_lq*, *refs*)

Texture transformer.

Q = LTE(img_lq) K = LTE(ref_lq) V = LTE(ref), from V_level_n to V_level_1

Relevance embedding aims to embed the relevance between the LQ and Ref image by estimating the similarity between Q and K.

Hard-Attention: Only transfer features from the most relevant position in V for each query.

Soft-Attention: synthesize features from the transferred GT texture features T and the LQ features F from the backbone.

Parameters

- **extractor** (*All args are features come from*) – These features contain 3 levels. When upscale_factor=4, the size ratio of these features is level3:level2:level1 = 1:2:4.
- **img_lq** (*Tensor*) – Tensor of 4x bicubic-upsampled lq image. (N, C, H, W)
- **ref_lq** (*Tensor*) – Tensor of ref_lq. ref_lq is obtained by applying bicubic down-sampling and up-sampling with factor 4x on ref. (N, C, H, W)
- **refs** (*Tuple[Tensor]*) – Tuple of ref tensors. [(N, C, H, W), (N, C/2, 2H, 2W), . . .]

Returns

tuple contains: soft_attention (Tensor): Soft-Attention tensor. (N, 1, H, W)

textures (Tuple[Tensor]): Transferred GT textures. [(N, C, H, W), (N, C/2, 2H, 2W), . . .]

Return type tuple

class mmagic.models.editors.**TTSRDiscriminator**(*in_channels=3*, *in_size=160*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

A discriminator for TTSR.

Parameters

- **in_channels** (*int*) – Channel number of inputs. Default: 3.
- **in_size** (*int*) – Size of input image. Default: 160.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward(*x*)

Forward function.

Parameters **x** (*Tensor*) – Input tensor with shape (n, c, h, w).

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**TTSRNet**(*in_channels*, *out_channels*, *mid_channels=64*, *texture_channels=64*, *num_blocks=(16, 16, 8, 4)*, *res_scale=1.0*, *init_cfg=None*)

Bases: mmengine.model.BaseModule

TTSR network structure (main-net) for reference-based super-resolution.

Paper: Learning Texture Transformer Network for Image Super-Resolution

Adapted from 'https://github.com/researchmm/TTSR.git' 'https://github.com/researchmm/TTSR' Copyright permission at 'https://github.com/researchmm/TTSR/issues/38'.

Parameters

- **in_channels** (*int*) – Number of channels in the input image
- **out_channels** (*int*) – Number of channels in the output image
- **mid_channels** (*int*) – Channel number of intermediate features. Default: 64
- **texture_channels** (*int*) – Number of texture channels. Default: 64.
- **num_blocks** (*tuple[int]*) – Block numbers in the trunk network. Default: (16, 16, 8, 4)
- **res_scale** (*float*) – Used to scale the residual in residual block. Default: 1.
- **init_cfg** (*dict, optional*) – Initialization config dict.

forward(*x*, *soft_attention*, *textures*)

Forward function.

Parameters

- **x** (*Tensor*) – Input tensor with shape (n, c, h, w).
- **soft_attention** (*Tensor*) – Soft-Attention tensor with shape (n, 1, h, w).
- **textures** (*Tuple[Tensor]*) – Transferred HR texture tensors. [(N, C, H, W), (N, C/2, 2H, 2W), . . .]

Returns Forward results.

Return type Tensor

class mmagic.models.editors.**ViCo**(*vae: ModelType*, *text_encoder: ModelType*, *tokenizer: str*, *unet: ModelType*, *scheduler: ModelType*, *test_scheduler: Optional[ModelType] = None*, *val_prompts: Union[str, List[str]] = None*, *dtype: str = 'fp16'*, *enable_xformers: bool = True*, *noise_offset_weight: float = 0*, *tomesd_cfg: Optional[dict] = None*, *data_preprocessor: Optional[ModelType] = dict(type='DataPreprocessor')*, *init_cfg: Optional[dict] = None*, *image_cross_layers: List[int] = None*, *reg_loss_weight: float = 0*, *placeholder: str = None*, *initialize_token: str = None*, *num_vectors_per_token: int = 1*)

Bases: mmagic.models.editors.stable_diffusion.stable_diffusion.StableDiffusion

Implementation of `ViCo with Stable Diffusion.

<https://arxiv.org/abs/2306.00971>`_ (ViCo).

Parameters

- **vae** (*Union[dict, nn.Module]*) – The config or module for VAE model.
- **text_encoder** (*Union[dict, nn.Module]*) – The config or module for text encoder.
- **tokenizer** (*str*) – The **name** for CLIP tokenizer.
- **unet** (*Union[dict, nn.Module]*) – The config or module for Unet model.
- **schedule** (*Union[dict, nn.Module]*) – The config or module for diffusion scheduler.
- **test_scheduler** (*Union[dict, nn.Module], optional*) – The config or module for diffusion scheduler in test stage (*self.infer*). If not passed, will use the same scheduler as *schedule*. Defaults to None.
- **val_prompts** (*Union[str, List[str]], optional*) – The prompts for validation. Defaults to None.
- **num_class_images** (*int, optional*) – The number of images for class prior. Defaults to 3.
- **dtype** (*str, optional*) – The dtype for the model. Defaults to 'fp16'.
- **enable_xformers** (*bool, optional*) – Whether to use xformers. Defaults to True.
- **noise_offset_weight** (*bool, optional*) – The weight of noise offset introduced in https://www.crosslabs.org/blog/diffusion-with-offset-noise # noqa Defaults to 0.
- **data_preprocessor** (*dict, optional*) – The pre-process config of `BaseDataPreprocessor`. Defaults to

  dict(type='DataPreprocessor').
- **init_cfg** (*dict, optional*) – The weight initialized config for `BaseModule`. Defaults to None/
- **image_cross_layers** (*List[int], optional*) – The layers to use image cross attention. Defaults to None.
- **reg_loss_weight** (*float, optional*) – The weight of regularization loss. Defaults to 0.
- **placeholder** (*str, optional*) – The placeholder token. Defaults to None.
- **initialize_token** (*str, optional*) – The token to initialize the placeholder. Defaults to None.
- **num_vectors_per_token** (*int, optional*) – The number of vectors per token.

prepare_models()

Prepare model for training.

Move model to target dtype and disable gradient for some models.

set_vico(*have_image_cross_attention: List[int]*)

Set ViCo for model.

set_only_imca_trainable()

Set only image cross attention trainable.

add_tokens(*placeholder_token: str*, *initialize_token: str = None*, *num_vectors_per_token: int = 1*)

Add token for training.

TODO: support add tokens as dict, then we can load pretrained tokens.

val_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

test_step(*data: dict*) → mmagic.utils.typing.SampleList

Gets the generated image of given data. Calls `self.data_preprocessor` and `self.infer` in order. Return the generated results which will be passed to evaluator or visualizer.

Parameters **data** (*dict or tuple or list*) – Data sampled from dataset.

Returns Generated image or image dict.

Return type SampleList

prepare_reference(*image_ref: Union[PIL.Image.Image, torch.Tensor]*, *height: Optional[int] = 512*, *width: Optional[int] = 512*)

train_step(*data*, *optim_wrapper*)

Training step.

infer(*prompt: Union[str, List[str]]*, *image_reference: PIL.Image.Image = None*, *height: Optional[int] = None*, *width: Optional[int] = None*, *num_inference_steps: int = 50*, *guidance_scale: float = 7.5*, *negative_prompt: Optional[Union[str, List[str]]] = None*, *num_images_per_prompt: Optional[int] = 1*, *eta: float = 0.0*, *generator: Optional[torch.Generator] = None*, *latents: Optional[torch.FloatTensor] = None*, *show_progress=True*, *seed=1*, *return_type='image'*)

Function invoked when calling the pipeline for generation.

Parameters

- **prompt** (*str* or *List[str]*) – The prompt or prompts to guide the image generation.
- **(int** (*height*) – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The height in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The height in pixels of the generated image.

Parameters

- **(int** (*width*) – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.
- **optional** – defaults to self.unet_sample_size * self.vae_scale_factor): The width in pixels of the generated image.

:param [defaults to self.unet_sample_size * self.vae_scale_factor):] The width in pixels of the generated image.

Parameters

- **num_inference_steps** (*int*, *optional*, defaults to 50) – The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
- **guidance_scale** (*float*, *optional*, defaults to 7.5) – Guidance scale as defined in [Classifier-Free Diffusion Guidance] (https://arxiv.org/abs/2207.12598).
- **negative_prompt** (*str* or *List[str]*, *optional*) – The prompt or prompts not to guide the image generation. Ignored when not using guidance (i.e., ignored if *guidance_scale* is less than *1*).
- **num_images_per_prompt** (*int*, *optional*, defaults to 1) – The number of images to generate per prompt.
- **eta** (*float*, *optional*, defaults to 0.0) – Corresponds to parameter eta () in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to [*schedulers.DDIMScheduler*], will be ignored for others.
- **generator** (*torch.Generator*, *optional*) – A [torch generator] to make generation deterministic.
- **latents** (*torch.FloatTensor*, *optional*) – Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will be generated by sampling using the supplied random *generator*.
- **return_type** (*str*) – The return type of the inference results. Supported types are 'image', 'numpy', 'tensor'. If 'image' is passed, a list of PIL images will be returned. If 'numpy' is passed, a numpy array with shape [N, C, H, W] will be returned, and the value range will be same as decoder's output range. If 'tensor' is passed, the decoder's output will be returned. Defaults to 'image'.

Returns A dict containing the generated images.

Return type dict

abstract forward(*inputs: torch.Tensor*, *data_samples: Optional[list] = None*, *mode: str = 'tensor'*) → Union[Dict[str, torch.Tensor], list]

forward is not implemented now.

class mmagic.models.editors.**WGANGP**(**args*, ***kwargs*)

Bases: *mmagic.models.base_models.BaseGAN*

Implementation of *Improved Training of Wasserstein GANs*.

Paper link: https://arxiv.org/pdf/1704.00028

Detailed architecture can be found in `WGANGPGenerator` and `WGANGPDiscriminator`

disc_loss(*real_data: torch.Tensor*, *fake_data: torch.Tensor*, *disc_pred_fake: torch.Tensor*, *disc_pred_real: torch.Tensor*) → Tuple

Get disc loss. WGAN-GP use the wgan loss and gradient penalty to train the discriminator.

Parameters

- **real_data** (*Tensor*) – Real input data.
- **fake_data** (*Tensor*) – Fake input data.
- **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

- **disc_pred_real** (*Tensor*) – Discriminator's prediction of the real images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

gen_loss(*disc_pred_fake: torch.Tensor*) → Tuple

Get gen loss. DCGAN use the wgan loss to train the generator.

Parameters **disc_pred_fake** (*Tensor*) – Discriminator's prediction of the fake images.

Returns Loss value and a dict of log variables.

Return type tuple[Tensor, dict]

train_discriminator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train discriminator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

train_generator(*inputs: dict*, *data_samples:* mmagic.structures.DataSample, *optimizer_wrapper: mmengine.optim.OptimWrapper*) → Dict[str, torch.Tensor]

Train generator.

Parameters

- **inputs** (*dict*) – Inputs from dataloader.
- **data_samples** (*DataSample*) – Data samples from dataloader. Do not used in generator's training.
- **optim_wrapper** (*OptimWrapper*) – OptimWrapper instance used to update model parameters.

Returns A `dict` of tensor for logging.

Return type Dict[str, Tensor]

1.92 mmagic.utils

1.92.1 Package Contents

Functions

modify_args()	Modify args of argparse.ArgumentParser.
all_to_tensor(value)	Trans image and sequence of frames to tensor.
can_convert_to_image(value)	Judge whether the input value can be converted to image tensor via
get_box_info(pred_bbox, original_shape, final_size)	**param pred_bbox** The bounding box for the instance
reorder_image(img[, input_order])	Reorder images to 'HWC' order.
tensor2img(tensor[, out_type, min_max])	Convert torch Tensors into image numpy arrays.
to_numpy(img[, dtype])	Convert data into numpy arrays of dtype.
download_from_url(url[, dest_path, dest_dir, hash_prefix])	Download object at the given URL to a local path.
print_colored_log(msg[, level, color])	Print colored log with default logger.
get_sampler(sample_kwargs, runner)	Get a sampler to loop input data.
register_all_modules(→ None)	Register all modules in mmagic into the registries.
try_import(→ Optional[types.ModuleType])	Try to import a module.
add_gaussian_noise(img, mu, sigma)	Add Gaussian Noise on the input image.
adjust_gamma(image[, gamma, gain])	Performs Gamma Correction on the input image.
bbox2mask(img_shape, bbox[, dtype])	Generate mask in np.ndarray from bbox.
brush_stroke_mask(img_shape[, num_vertices, ...])	Generate free-form mask.
get_irregular_mask(img_shape[, area_ratio_range])	Get irregular mask with the constraints in mask ratio.
make_coord(shape[, ranges, flatten])	Make coordinates at grid centers.
random_bbox(img_shape, max_bbox_shape[, ...])	Generate a random bbox for the mask on a given image.
random_choose_unknown(unknown, crop_size)	Randomly choose an unknown start (top-left) point for a given crop_size.

Attributes

MMAGIC_CACHE_DIR
ConfigType
ForwardInputs
LabelVar
NoiseVar
SampleList

`mmagic.utils.`**`modify_args`**`()`

Modify args of argparse.ArgumentParser.

`mmagic.utils.`**`all_to_tensor`**`(`*value*`)`

Trans image and sequence of frames to tensor.

Parameters **value** (*np.ndarray | list[np.ndarray] | Tuple[np.ndarray]*) – The original image or list of frames.

Returns The output tensor.

Return type Tensor

mmagic.utils.**can_convert_to_image**(*value*)

Judge whether the input value can be converted to image tensor via `images_to_tensor()` function.

Parameters **value** (*any*) – The input value.

Returns

If true, the input value can convert to image with `images_to_tensor()`, and vice versa.

Return type bool

mmagic.utils.**get_box_info**(*pred_bbox*, *original_shape*, *final_size*)

Parameters

- **pred_bbox** – The bounding box for the instance
- **original_shape** – Original image shape
- **final_size** – Size of the final output

Returns [L_pad, R_pad, T_pad, B_pad, rh, rw]

Return type List

mmagic.utils.**reorder_image**(*img*, *input_order='HWC'*)

Reorder images to 'HWC' order.

If the input_order is (h, w), return (h, w, 1); If the input_order is (c, h, w), return (h, w, c); If the input_order is (h, w, c), return as it is.

Parameters

- **img** (*np.ndarray*) – Input image.
- **input_order** (*str*) – Whether the input order is 'HWC' or 'CHW'. If the input image shape is (h, w), input_order will not have effects. Default: 'HWC'.

Returns Reordered image.

Return type np.ndarray

mmagic.utils.**tensor2img**(*tensor*, *out_type=np.uint8*, *min_max=(0, 1)*)

Convert torch Tensors into image numpy arrays.

After clamping to (min, max), image values will be normalized to [0, 1].

For different tensor shapes, this function will have different behaviors:

1. **4D mini-batch Tensor of shape (N x 3/1 x H x W):** Use *make_grid* to stitch images in the batch dimension, and then convert it to numpy array.
2. **3D Tensor of shape (3/1 x H x W) and 2D Tensor of shape (H x W):** Directly change to numpy array.

Note that the image channel in input tensors should be RGB order. This function will convert it to cv2 convention, i.e., (H x W x C) with BGR order.

Parameters

- **tensor** (*Tensor | list[Tensor]*) – Input tensors.

- **out_type** (*numpy type*) – Output types. If `np.uint8`, transform outputs to uint8 type with range [0, 255]; otherwise, float type with range [0, 1]. Default: `np.uint8`.
- **min_max** (*tuple*) – min and max values for clamp.

Returns 3D ndarray of shape (H x W x C) or 2D ndarray of shape (H x W).

Return type (Tensor | list[Tensor])

`mmagic.utils.`**`to_numpy`**(*img*, *dtype=np.float64*)

Convert data into numpy arrays of dtype.

Parameters

- **img** (*Tensor | np.ndarray*) – Input data.
- **dtype** (*np.dtype*) – Set the data type of the output. Default: np.float64

Returns Converted numpy arrays data.

Return type img (np.ndarray)

`mmagic.utils.`**`MMAGIC_CACHE_DIR`**

`mmagic.utils.`**`download_from_url`**(*url*, *dest_path=None*, *dest_dir=MMAGIC_CACHE_DIR*, *hash_prefix=None*)

Download object at the given URL to a local path.

Parameters

- **url** (*str*) – URL of the object to download.
- **dest_path** (*str*) – Path where object will be saved.
- **dest_dir** (*str*) – The directory of the destination. Defaults to `'~/.cache/openmmlab/mmagic/'`.
- **hash_prefix** (*string, optional*) – If not None, the SHA256 downloaded file should start with *hash_prefix*. Default: None.

Returns path for the downloaded file.

Return type str

`mmagic.utils.`**`print_colored_log`**(*msg*, *level=logging.INFO*, *color='magenta'*)

Print colored log with default logger.

Parameters

- **msg** (*str*) – Message to log.
- **level** (*int*) – The root logger level. Note that only the process of rank 0 is affected, while other processes will set the level to "Error" and be silent most of the time.Log level, default to 'info'.
- **color** (*str, optional*) – Color 'magenta'.

`mmagic.utils.`**`get_sampler`**(*sample_kwargs: dict*, *runner: Optional[mmengine.runner.Runner]*)

Get a sampler to loop input data.

Parameters

- **sample_kwargs** (*dict*) – _description_
- **runner** (*Optional[Runner]*) – _description_

Returns _description_

Return type _type_

`mmagic.utils.register_all_modules(init_default_scope: bool = True) → None`

Register all modules in mmagic into the registries.

Parameters **init_default_scope** (*bool*) – Whether initialize the mmagic default scope. When *init_default_scope=True*, the global default scope will be set to *mmagic*, and all registries will build modules from mmagic's registry node. To understand more about the registry, please refer to https://mmengine.readthedocs.io/en/latest/advanced_tutorials/registry.html Defaults to True.

`mmagic.utils.try_import(name: str) → Optional[types.ModuleType]`

Try to import a module.

Parameters **name** (*str*) – Specifies what module to import in absolute or relative terms (e.g. either pkg.mod or ..mod).

Returns If importing successfully, returns the imported module, otherwise returns None.

Return type ModuleType or None

`mmagic.utils.add_gaussian_noise(img: numpy.ndarray, mu, sigma)`

Add Gaussian Noise on the input image.

Parameters

- **img** (*np.ndarray*) – Input image.
- **mu** (*float*) – The mu value of the Gaussian function.
- **sigma** (*float*) – The sigma value of the Gaussian function.

Returns Gaussian noisy output image.

Return type noisy_img (np.ndarray)

`mmagic.utils.adjust_gamma(image, gamma=1, gain=1)`

Performs Gamma Correction on the input image.

This function is adopted from skimage: https://github.com/scikit-image/scikit-image/blob/7e4840bd9439d1dfb6beaf549998452c99f97fdd/skimage/exposure/ exposure.py#L439-L494

Also known as Power Law Transform. This function transforms the input image pixelwise according to the equation `O = I**gamma` after scaling each pixel to the range 0 to 1.

Parameters

- **image** (*np.ndarray*) – Input image.
- **gamma** (*float, optional*) – Non negative real number. Defaults to 1.
- **gain** (*float, optional*) – The constant multiplier. Defaults to 1.

Returns Gamma corrected output image.

Return type np.ndarray

`mmagic.utils.bbox2mask(img_shape, bbox, dtype='uint8')`

Generate mask in np.ndarray from bbox.

The returned mask has the shape of (h, w, 1). '1' indicates the hole and '0' indicates the valid regions.

We prefer to use *uint8* as the data type of masks, which may be different from other codes in the community.

Parameters

- **img_shape** (*tuple[int]*) – The size of the image.
- **bbox** (*tuple[int]*) – Configuration tuple, (top, left, height, width)
- **np.dtype** (*str*) – Indicate the data type of returned masks. Default: 'uint8'

Returns Mask in the shape of (h, w, 1).

Return type mask (np.ndarray)

`mmagic.utils.brush_stroke_mask(img_shape, num_vertices=(4, 12), mean_angle=2 * math.pi / 5, angle_range=2 * math.pi / 15, brush_width=(12, 40), max_loops=4, dtype='uint8')`

Generate free-form mask.

The method of generating free-form mask is in the following paper: Free-Form Image Inpainting with Gated Convolution.

When you set the config of this type of mask. You may note the usage of *np.random.randint* and the range of *np.random.randint* is [left, right).

We prefer to use *uint8* as the data type of masks, which may be different from other codes in the community.

TODO: Rewrite the implementation of this function.

Parameters

- **img_shape** (*tuple[int]*) – Size of the image.
- **num_vertices** (*int | tuple[int]*) – Min and max number of vertices. If only give an integer, we will fix the number of vertices. Default: (4, 12).
- **mean_angle** (*float*) – Mean value of the angle in each vertex. The angle is measured in radians. Default: 2 * math.pi / 5.
- **angle_range** (*float*) – Range of the random angle. Default: 2 * math.pi / 15.
- **brush_width** (*int | tuple[int]*) – (min_width, max_width). If only give an integer, we will fix the width of brush. Default: (12, 40).
- **max_loops** (*int*) – The max number of for loops of drawing strokes. Default: 4.
- **np.dtype** (*str*) – Indicate the data type of returned masks. Default: 'uint8'.

Returns Mask in the shape of (h, w, 1).

Return type mask (np.ndarray)

`mmagic.utils.get_irregular_mask(img_shape, area_ratio_range=(0.15, 0.5), **kwargs)`

Get irregular mask with the constraints in mask ratio.

Parameters

- **img_shape** (*tuple[int]*) – Size of the image.
- **area_ratio_range** (*tuple(float)*) – Contain the minimum and maximum area
- **Default** (*ratio.*) – (0.15, 0.5).

Returns Mask in the shape of (h, w, 1).

Return type mask (np.ndarray)

`mmagic.utils.make_coord(shape, ranges=None, flatten=True)`

Make coordinates at grid centers.

Parameters

- **shape** (*tuple*) – shape of image.
- **ranges** (*tuple*) – range of coordinate value. Default: None.
- **flatten** (*bool*) – flatten to (n, 2) or Not. Default: True.

Returns coordinates.

Return type coord (Tensor)

mmagic.utils.**random_bbox**(*img_shape*, *max_bbox_shape*, *max_bbox_delta=40*, *min_margin=20*)

Generate a random bbox for the mask on a given image.

In our implementation, the max value cannot be obtained since we use *np.random.randint*. And this may be different with other standard scripts in the community.

Parameters

- **img_shape** (*tuple[int]*) – The size of a image, in the form of (h, w).
- **max_bbox_shape** (*int | tuple[int]*) – Maximum shape of the mask box, in the form of (h, w). If it is an integer, the mask box will be square.
- **max_bbox_delta** (*int | tuple[int]*) – Maximum delta of the mask box, in the form of (delta_h, delta_w). If it is an integer, delta_h and delta_w will be the same. Mask shape will be randomly sampled from the range of *max_bbox_shape - max_bbox_delta* and *max_bbox_shape*. Default: (40, 40).
- **min_margin** (*int | tuple[int]*) – The minimum margin size from the edges of mask box to the image boarder, in the form of (margin_h, margin_w). If it is an integer, margin_h and margin_w will be the same. Default: (20, 20).

Returns The generated box, (top, left, h, w).

Return type tuple[int]

mmagic.utils.**random_choose_unknown**(*unknown*, *crop_size*)

Randomly choose an unknown start (top-left) point for a given crop_size.

Parameters

- **unknown** (*np.ndarray*) – The binary unknown mask.
- **crop_size** (*tuple[int]*) – The given crop size.

Returns The top-left point of the chosen bbox.

Return type tuple[int]

mmagic.utils.**ConfigType**

mmagic.utils.**ForwardInputs**

mmagic.utils.**LabelVar**

mmagic.utils.**NoiseVar**

mmagic.utils.**SampleList**

1.93 Overview

This section introduces the following contents in terms of migration from MMEditing 0.x

- *Overview*
  - *New dependencies*
  - *Overall structures*
  - *Other config settings*

1.93.1 New dependencies

MMagic 1.x depends on some new packages, you can prepare a new clean environment and install it again according to the *install tutorial*.

1.93.2 Overall structures

We refactor overall structures in MMagic 1.x as follows.

- The `core` in the old versions of MMEdit is split into `engine`, `evaluation`, `structures`, and `visualization`
- The `pipelines` of `datasets` in the old versions of MMEdit is refactored to `transforms`
- The `models` in MMagic 1.x is refactored to six parts: `archs`, `base_models`, `data_preprocessors`, `editors`, `diffusion_schedulers`, and `losses`.

1.93.3 Other config settings

We rename the config file to the new template: `{model_settings}_{module_setting}_{training_setting}_{datasets_info}`.

More details of config are shown in *config guides*.

1.94 Migration of Runtime Settings

We update runtime settings in MMagic 1.x. Important modifications are as following.

- The `checkpoint_config` is moved to `default_hooks.checkpoint` and the `log_config` is moved to `default_hooks.logger`. And we move many hooks settings from the script code to the `default_hooks` field in the runtime configuration.
- The `resume_from` is removed. And we use `resume` to replace it.
  - If resume=True and load_from is not None, resume training from the checkpoint in load_from.
  - If resume=True and load_from is None, try to resume from the latest checkpoint in the work directory.
  - If resume=False and load_from is not None, only load the checkpoint, not resume training.
  - If resume=False and load_from is None, do not load nor resume.
- The `dist_params` field is a sub field of `env_cfg` now. And there are some new configurations in the `env_cfg`.
- The `workflow` related functionalities are removed.
- New field `visualizer`: The visualizer is a new design. We use a visualizer instance in the runner to handle results & log visualization and save to different backends, like Local, TensorBoard and Wandb.

- New field `default_scope`: The start point to search module for all registries.

```
checkpoint_config = dict(  # Config to set the checkpoint hook, Refer to https://github.
↪com/open-mmlab/mmcv/blob/master/mmcv/runner/hooks/checkpoint.py for implementation.
    interval=5000,  # The save interval is 5000 iterations
    save_optimizer=True,  # Also save optimizers
    by_epoch=False)  # Count by iterations
log_config = dict(  # Config to register logger hook
    interval=100,  # Interval to print the log
    hooks=[
        dict(type='TextLoggerHook', by_epoch=False),  # The logger used to record the␣
↪training process
        dict(type='TensorboardLoggerHook'),  # The Tensorboard logger is also supported
    ])
visual_config = None  # Visual config, we do not use it.
# runtime settings
dist_params = dict(backend='nccl')  # Parameters to setup distributed training, the port␣
↪can also be set
log_level = 'INFO'  # The level of logging
load_from = None # load models as a pre-trained model from a given path. This will not␣
↪resume training
resume_from = None # Resume checkpoints from a given path, the training will be resumed␣
↪from the iteration when the checkpoint's is saved
workflow = [('train', 1)]  # Workflow for runner. [('train', 1)] means there is only one␣
↪workflow and the workflow named 'train' is executed once. Keep this unchanged when␣
↪training current matting models
```

```
default_hooks = dict(  # Used to build default hooks
    checkpoint=dict(  # Config to set the checkpoint hook
        type='CheckpointHook',
        interval=5000,  # The save interval is 5000 iterations
        save_optimizer=True,
        by_epoch=False,  # Count by iterations
        out_dir=save_dir,
    ),
    timer=dict(type='IterTimerHook'),
    logger=dict(type='LoggerHook', interval=100),  # Config to register logger hook
    param_scheduler=dict(type='ParamSchedulerHook'),
    sampler_seed=dict(type='DistSamplerSeedHook'),
)
default_scope = 'mmedit' # Used to set registries location
env_cfg = dict(  # Parameters to setup distributed training, the port can also be set
    cudnn_benchmark=False,
    mp_cfg=dict(mp_start_method='fork', opencv_num_threads=4),
    dist_cfg=dict(backend='nccl'),
)
log_level = 'INFO'  # The level of logging
log_processor = dict(type='LogProcessor', window_size=100, by_epoch=False)  # Used to␣
↪build log processor
load_from = None  # load models as a pre-trained model from a given path. This will not␣
↪resume training.
resume = False  # Resume checkpoints from a given path, the training will be resumed␣
↪from the epoch when the checkpoint's is saved.
```

1.95 Migration of Model Settings

We update model settings in MMagic 1.x. Important modifications are as following.

- Remove `pretrained` fields.
- Add `train_cfg` and `test_cfg` fields in model settings.
- Add `data_preprocessor` fields. Normalization and color space transforms operations are moved from datasets transforms pipelines to data_preprocessor. We will introduce data_preprocessor later.

```
model = dict(
    type='BasicRestorer',  # Name of the model
    generator=dict(  # Config of the generator
        type='EDSR',  # Type of the generator
        in_channels=3,  # Channel number of inputs
        out_channels=3,  # Channel number of outputs
        mid_channels=64,  # Channel number of intermediate features
        num_blocks=16,  # Block number in the trunk network
        upscale_factor=scale, # Upsampling factor
        res_scale=1,  # Used to scale the residual in residual block
        rgb_mean=(0.4488, 0.4371, 0.4040),  # Image mean in RGB orders
        rgb_std=(1.0, 1.0, 1.0)),  # Image std in RGB orders
    pretrained=None,
    pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean'))  # Config for␣
↪pixel loss model training and testing settings
```

```
model = dict(
    type='BaseEditModel',  # Name of the model
    generator=dict(  # Config of the generator
        type='EDSRNet',  # Type of the generator
        in_channels=3,  # Channel number of inputs
        out_channels=3,  # Channel number of outputs
        mid_channels=64,  # Channel number of intermediate features
        num_blocks=16,  # Block number in the trunk network
        upscale_factor=scale, # Upsampling factor
        res_scale=1,  # Used to scale the residual in residual block
        rgb_mean=(0.4488, 0.4371, 0.4040),  # Image mean in RGB orders
        rgb_std=(1.0, 1.0, 1.0)),  # Image std in RGB orders
    pixel_loss=dict(type='L1Loss', loss_weight=1.0, reduction='mean')  # Config for␣
↪pixel loss
    train_cfg=dict(),  # Config of training model.
    test_cfg=dict(),  # Config of testing model.
    data_preprocessor=dict(  # The Config to build data preprocessor
        type='DataPreprocessor', mean=[0., 0., 0.], std=[255., 255.,
                                                         255.]))
```

We refactor models in MMagic 1.x. Important modifications are as following.

- The `models` in MMagic 1.x is refactored to six parts: `archs`, `base_models`, `data_preprocessors`, `editors`, `diffusion_schedulers` and `losses`.
- Add `data_preprocessor` module in `models`. Normalization and color space transforms operations are moved from datasets transforms pipelines to data_preprocessor. The data out from the data pipeline is transformed by this module and then fed into the model.

More details of models are shown in *model guides*.

1.96 Migration of Evaluation and Testing Settings

We update evaluation settings in MMagic 1.x. Important modifications are as following.

- The evaluation field is split to `val_evaluator` and `test_evaluator`. The `interval` is moved to `train_cfg.val_interval`.
- The metrics to evaluation are moved from `test_cfg` to `val_evaluator` and `test_evaluator`.

```
train_cfg = None  # Training config
test_cfg = dict(  # Test config
    metrics=['PSNR'],  # Metrics used during testing
    crop_border=scale)  # Crop border during evaluation

evaluation = dict(  # The config to build the evaluation hook
    interval=5000,  # Evaluation interval
    save_image=True,  # Save images during evaluation
    gpu_collect=True)  # Use gpu collect
```

```
val_evaluator = [
    dict(type='PSNR', crop_border=scale),  # The name of metrics to evaluate
]
test_evaluator = val_evaluator

train_cfg = dict(
    type='IterBasedTrainLoop', max_iters=300000, val_interval=5000)  # Config of train␣
↪loop type
val_cfg = dict(type='ValLoop')  # The name of validation loop type
test_cfg = dict(type='TestLoop')  # The name of test loop type
```

We have merged MMGeneration 1.x into MMagic. Here is migration of Evaluation and Testing Settings about MMGeneration.

The evaluation field is splited to `val_evaluator` and `test_evaluator`. And it won't support `interval` and `save_best` arguments. The `interval` is moved to `train_cfg.val_interval`, see *the schedule settings* and the `save_best` is moved to `default_hooks.checkpoint.save_best`.

```
evaluation = dict(
    type='GenerativeEvalHook',
    interval=10000,
    metrics=[
        dict(
            type='FID',
            num_images=50000,
            bgr2rgb=True,
            inception_args=dict(type='StyleGAN')),
        dict(type='IS', num_images=50000)
    ],
    best_metric=['fid', 'is'],
    sample_kwargs=dict(sample_model='ema'))
```

```
val_evaluator = dict(
    type='Evaluator',
    metrics=[
        dict(
            type='FID',
            prefix='FID-Full-50k',
            fake_nums=50000,
            inception_style='StyleGAN',
            sample_model='orig')
        dict(
            type='IS',
            prefix='IS-50k',
            fake_nums=50000)])
# set best config
default_hooks = dict(
    checkpoint=dict(
        type='CheckpointHook',
        interval=10000,
        by_epoch=False,
        less_keys=['FID-Full-50k/fid'],
        greater_keys=['IS-50k/is'],
        save_optimizer=True,
        save_best=['FID-Full-50k/fid', 'IS-50k/is'],
        rule=['less', 'greater']))
test_evaluator = val_evaluator
```

To evaluate and test the model correctly, we need to set specific loop in `val_cfg` and `test_cfg`.

```
total_iters = 1000000

runner = dict(
    type='DynamicIterBasedRunner',
    is_dynamic_ddp=False,
    pass_training_status=True)
```

```
train_cfg = dict(
    by_epoch=False,  # use iteration based training
    max_iters=1000000,  # max training iteration
    val_begin=1,
    val_interval=10000)  # evaluation interval
val_cfg = dict(type='MultiValLoop')  # specific loop in validation
test_cfg = dict(type='MultiTestLoop')  # specific loop in testing
```

1.97 Migration of Schedule Settings

We update schedule settings in MMagic 1.x. Important modifications are as following.

- Now we use `optim_wrapper` field to specify all configuration about the optimization process. And the `optimizer` is a sub field of `optim_wrapper` now.
- The `lr_config` field is removed and we use new `param_scheduler` to replace it.
- The `total_iters` field is moved to `train_cfg` as `max_iters`, `val_cfg` and `test_cfg`, which configure the loop in training, validation and test.

```
optimizers = dict(generator=dict(type='Adam', lr=1e-4, betas=(0.9, 0.999)))  # Config␣
↪used to build optimizer, support all the optimizers in PyTorch whose arguments are␣
↪also the same as those in PyTorch
total_iters = 300000 # Total training iters
lr_config = dict( # Learning rate scheduler config used to register LrUpdater hook
    policy='Step', by_epoch=False, step=[200000], gamma=0.5)  # The policy of scheduler
```

```
optim_wrapper = dict(
    dict(
        type='OptimWrapper',
        optimizer=dict(type='Adam', lr=1e-4),
    )
)  # Config used to build optimizer, support all the optimizers in PyTorch whose␣
↪arguments are also the same as those in PyTorch.
param_scheduler = dict(  # Config of learning policy
    type='MultiStepLR', by_epoch=False, milestones=[200000], gamma=0.5)  # The policy of␣
↪scheduler
train_cfg = dict(
    type='IterBasedTrainLoop', max_iters=300000, val_interval=5000)  # Config of train␣
↪loop type
val_cfg = dict(type='ValLoop')  # The name of validation loop type
test_cfg = dict(type='TestLoop')  # The name of test loop type
```

More details of schedule settings are shown in MMEngine Documents.

1.98 Migration of Data Settings

This section introduces the migration of data settings:

- *Migration of Data Settings*
  - *Data pipelines*
  - *Dataloader*

1.98.1 Data pipelines

We update data pipelines settings in MMagic 1.x. Important modifications are as following.

- Remove normalization and color space transforms operations. They are moved from datasets transforms pipelines to data_preprocessor.
- The original formatting transforms pipelines `Collect` and `ToTensor` are combined as `PackInputs`. More details of data pipelines are shown in *transform guides*.

```
train_pipeline = [  # Training data processing pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        io_backend='disk',  # io backend
        key='lq',  # Keys in results to find corresponding path
        flag='unchanged'),  # flag for reading images
    dict(type='LoadImageFromFile',  # Load images from files
        io_backend='disk',  # io backend
        key='gt',  # Keys in results to find corresponding path
        flag='unchanged'),  # flag for reading images
    dict(type='RescaleToZeroOne', keys=['lq', 'gt']),  # Rescale images from [0, 255] to␣
↪[0, 1]
    dict(type='Normalize',  # Augmentation pipeline that normalize the input images
        keys=['lq', 'gt'],  # Images to be normalized
        mean=[0, 0, 0],  # Mean values
        std=[1, 1, 1],  # Standard variance
        to_rgb=True),  # Change to RGB channel
    dict(type='PairedRandomCrop', gt_patch_size=96),  # Paired random crop
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='horizontal'),  # Flip direction
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='vertical'),  # Flip direction
    dict(type='RandomTransposeHW',  # Random transpose h and w for images
        keys=['lq', 'gt'],  # Images to be transposed
        transpose_ratio=0.5  # Transpose ratio
        ),
    dict(type='Collect',  # Pipeline that decides which keys in the data should be␣
↪passed to the model
        keys=['lq', 'gt'],  # Keys to pass to the model
        meta_keys=['lq_path', 'gt_path']), # Meta information keys. In training, meta␣
↪information is not needed
    dict(type='ToTensor',  # Convert images to tensor
        keys=['lq', 'gt'])  # Images to be converted to Tensor
]
test_pipeline = [  # Test pipeline
    dict(
        type='LoadImageFromFile',  # Load images from files
        io_backend='disk',  # io backend
        key='lq',  # Keys in results to find corresponding path
        flag='unchanged'),  # flag for reading images
    dict(
```

(continues on next page)

(continued from previous page)

```
        type='LoadImageFromFile',  # Load images from files
        io_backend='disk',  # io backend
        key='gt',  # Keys in results to find corresponding path
        flag='unchanged'),  # flag for reading images
    dict(type='RescaleToZeroOne', keys=['lq', 'gt']),  # Rescale images from [0, 255] to␣
↪[0, 1]
    dict(
        type='Normalize',  # Augmentation pipeline that normalize the input images
        keys=['lq', 'gt'],  # Images to be normalized
        mean=[0, 0, 0],  # Mean values
        std=[1, 1, 1],  # Standard variance
        to_rgb=True),  # Change to RGB channel
    dict(type='Collect',  # Pipeline that decides which keys in the data should be␣
↪passed to the model
        keys=['lq', 'gt'],  # Keys to pass to the model
        meta_keys=['lq_path', 'gt_path']),  # Meta information keys
    dict(type='ToTensor',  # Convert images to tensor
        keys=['lq', 'gt'])  # Images to be converted to Tensor
]
```

```
train_pipeline = [  # Training data processing pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='SetValues', dictionary=dict(scale=scale)),  # Set value to destination␣
↪keys
    dict(type='PairedRandomCrop', gt_patch_size=96),  # Paired random crop
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='horizontal'),  # Flip direction
    dict(type='Flip',  # Flip images
        keys=['lq', 'gt'],  # Images to be flipped
        flip_ratio=0.5,  # Flip ratio
        direction='vertical'),  # Flip direction
    dict(type='RandomTransposeHW',  # Random transpose h and w for images
        keys=['lq', 'gt'],  # Images to be transposed
        transpose_ratio=0.5  # Transpose ratio
        ),
    dict(type='PackInputs')  # The config of collecting data from current pipeline
]
test_pipeline = [  # Test pipeline
    dict(type='LoadImageFromFile',  # Load images from files
        key='img',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
```

(continues on next page)

(continued from previous page)

```
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='LoadImageFromFile',  # Load images from files
        key='gt',  # Keys in results to find corresponding path
        color_type='color',  # Color type of image
        channel_order='rgb',  # Channel order of image
        imdecode_backend='cv2'),  # decode backend
    dict(type='PackInputs')  # The config of collecting data from current pipeline
]
```

1.98.2 Dataloader

We update dataloader settings in MMagic 1.x. Important modifications are as following.

- The original `data` field is split to `train_dataloader`, `val_dataloader` and `test_dataloader`. This allows us to configure them in fine-grained. For example, you can specify different sampler and batch size during training and test.
- The `samples_per_gpu` is renamed to `batch_size`.
- The `workers_per_gpu` is renamed to `num_workers`.

```
data = dict(
    # train
    samples_per_gpu=16,  # Batch size of a single GPU
    workers_per_gpu=4,  # Worker to pre-fetch data for each single GPU
    drop_last=True,  # Use drop_last in data_loader
    train=dict(  # Train dataset config
        type='RepeatDataset',  # Repeated dataset for iter-based model
        times=1000,  # Repeated times for RepeatDataset
        dataset=dict(
            type=train_dataset_type,  # Type of dataset
            lq_folder='data/DIV2K/DIV2K_train_LR_bicubic/X2_sub',  # Path for lq folder
            gt_folder='data/DIV2K/DIV2K_train_HR_sub',  # Path for gt folder
            ann_file='data/DIV2K/meta_info_DIV2K800sub_GT.txt',  # Path for annotation␣
↪file
            pipeline=train_pipeline,  # See above for train_pipeline
            scale=scale)),  # Scale factor for upsampling
    # val
    val_samples_per_gpu=1,  # Batch size of a single GPU for validation
    val_workers_per_gpu=4,  # Worker to pre-fetch data for each single GPU for validation
    val=dict(
        type=val_dataset_type,  # Type of dataset
        lq_folder='data/val_set5/Set5_bicLRx2',  # Path for lq folder
        gt_folder='data/val_set5/Set5_mod12',  # Path for gt folder
        pipeline=test_pipeline,  # See above for test_pipeline
        scale=scale,  # Scale factor for upsampling
        filename_tmpl='{}'),  # filename template
    # test
    test=dict(
        type=val_dataset_type,  # Type of dataset
        lq_folder='data/val_set5/Set5_bicLRx2',  # Path for lq folder
```

(continues on next page)

(continued from previous page)

```
        gt_folder='data/val_set5/Set5_mod12',  # Path for gt folder
        pipeline=test_pipeline,  # See above for test_pipeline
        scale=scale,  # Scale factor for upsampling
        filename_tmpl='{}'))  # filename template
```

```
dataset_type = 'BasicImageDataset'  # The type of dataset
data_root = 'data'  # Root path of data
train_dataloader = dict(
    batch_size=16,
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    sampler=dict(type='InfiniteSampler', shuffle=True),  # The type of data sampler
    dataset=dict(  # Train dataset config
        type=dataset_type,  # Type of dataset
        ann_file='meta_info_DIV2K800sub_GT.txt',  # Path of annotation file
        metainfo=dict(dataset_type='div2k', task_name='sisr'),
        data_root=data_root + '/DIV2K',  # Root path of data
        data_prefix=dict(  # Prefix of image path
            img='DIV2K_train_LR_bicubic/X2_sub', gt='DIV2K_train_HR_sub'),
        filename_tmpl=dict(img='{}', gt='{}'),  # Filename template
        pipeline=train_pipeline))
val_dataloader = dict(
    batch_size=1,
    num_workers=4,  # The number of workers to pre-fetch data for each single GPU
    persistent_workers=False,  # Whether maintain the workers Dataset instances alive
    drop_last=False,  # Whether drop the last incomplete batch
    sampler=dict(type='DefaultSampler', shuffle=False),  # The type of data sampler
    dataset=dict(  # Validation dataset config
        type=dataset_type,  # Type of dataset
        metainfo=dict(dataset_type='set5', task_name='sisr'),
        data_root=data_root + '/Set5',  # Root path of data
        data_prefix=dict(img='LRbicx2', gt='GTmod12'),  # Prefix of image path
        pipeline=test_pipeline))
test_dataloader = val_dataloader
```

1.99 Migration of Distributed Training Settings

We have merged MMGeneration 1.x into MMagic. Here is migration of Distributed Training Settings about MMGeneration.

In 0.x version, MMGeneration uses `DDPWrapper` and `DynamicRunner` to train static and dynamic model (e.g., PGGAN and StyleGANv2) respectively. In 1.x version, we use `MMSeparateDistributedDataParallel` provided by MMEngine to implement distributed training.

The configuration differences are shown below:

```
# Use DDPWrapper
use_ddp_wrapper = True
find_unused_parameters = False

runner = dict(
```

(continues on next page)

(continued from previous page)

```
    type='DynamicIterBasedRunner',
    is_dynamic_ddp=False)
```

```
model_wrapper_cfg = dict(
    type='MMSeparateDistributedDataParallel',
    broadcast_buffers=False,
    find_unused_parameters=False)
```

```
use_ddp_wrapper = False
find_unused_parameters = False

# Use DynamicRunner
runner = dict(
    type='DynamicIterBasedRunner',
    is_dynamic_ddp=True)
```

```
model_wrapper_cfg = dict(
    type='MMSeparateDistributedDataParallel',
    broadcast_buffers=False,
    find_unused_parameters=True) # set `find_unused_parameters` for dynamic models
```

1.100 Migration of Optimizers

We have merged MMGeneration 1.x into MMagic. Here is migration of Optimizers about MMGeneration.

In version 0.x, MMGeneration uses PyTorch's native Optimizer, which only provides general parameter optimization. In version 1.x, we use `OptimizerWrapper` provided by MMEngine.

Compared to PyTorch's `Optimizer`, `OptimizerWrapper` supports the following features:

- `OptimizerWrapper.update_params` implement `zero_grad`, `backward` and `step` in a single function.
- Support gradient accumulation automatically.
- Provide a context manager named `OptimizerWrapper.optim_context` to warp the forward process. `optim_context` can automatically call `torch.no_sync` according to current number of updating iteration. In AMP (auto mixed precision) training, `autocast` is called in `optim_context` as well.

For GAN models, generator and discriminator use different optimizer and training schedule. To ensure that the GAN model's function signature of `train_step` is consistent with other models, we use `OptimWrapperDict`, inherited from `OptimizerWrapper`, to wrap the optimizer of the generator and discriminator. To automate this process MMagic implement `MultiOptimWrapperContructor`. And you should specify this constructor in your config is you want to train GAN model.

The config for the 0.x and 1.x versions are shown below:

```
optimizer = dict(
    generator=dict(type='Adam', lr=0.0001, betas=(0.0, 0.999), eps=1e-6),
    discriminator=dict(type='Adam', lr=0.0004, betas=(0.0, 0.999), eps=1e-6))
```

```
optim_wrapper = dict(
    constructor='MultiOptimWrapperConstructor',
```

(continues on next page)

(continued from previous page)

```
    generator=dict(optimizer=dict(type='Adam', lr=0.0002, betas=(0.0, 0.999), eps=1e-6)),
    discriminator=dict(
        optimizer=dict(type='Adam', lr=0.0004, betas=(0.0, 0.999), eps=1e-6)))
```

Note that, in the 1.x, MMGeneration uses `OptimWrapper` to realize gradient accumulation. This make the config of `discriminator_steps` (training trick for updating the generator once after multiple updates of the discriminator) and gradient accumulation different between 0.x and 1.x version.

- In 0.x version, we use `disc_steps`, `gen_steps` and `batch_accumulation_steps` in configs. `disc_steps` and `batch_accumulation_steps` are counted by the number of calls of `train_step` (is also the number of data reads from the dataloader). Therefore the number of consecutive updates of the discriminator is `disc_steps // batch_accumulation_steps`. And for generators, `gen_steps` is the number of times the generator actually updates continuously.
- In 1.x version, we use `discriminator_steps`, `generator_steps` and `accumulative_counts` in configs. `discriminator_steps` and `generator_steps` are the number of consecutive updates to itself before updating other modules.

Take config of BigGAN-128 as example.

```
model = dict(
    type='BasiccGAN',
    generator=dict(
        type='BigGANGenerator',
        output_scale=128,
        noise_size=120,
        num_classes=1000,
        base_channels=96,
        shared_dim=128,
        with_shared_embedding=True,
        sn_eps=1e-6,
        init_type='ortho',
        act_cfg=dict(type='ReLU', inplace=True),
        split_noise=True,
        auto_sync_bn=False),
    discriminator=dict(
        type='BigGANDiscriminator',
        input_scale=128,
        num_classes=1000,
        base_channels=96,
        sn_eps=1e-6,
        init_type='ortho',
        act_cfg=dict(type='ReLU', inplace=True),
        with_spectral_norm=True),
    gan_loss=dict(type='GANLoss', gan_type='hinge'))

# continuous update discriminator for `disc_steps // batch_accumulation_steps = 8 // 8 =␣
↪1` times
# continuous update generator for `gen_steps = 1` times
# generators and discriminators perform `batch_accumulation_steps = 8` times gradient␣
↪accumulations before each update
train_cfg = dict(
    disc_steps=8, gen_steps=1, batch_accumulation_steps=8, use_ema=True)
```

```
model = dict(
    type='BigGAN',
    num_classes=1000,
    data_preprocessor=dict(type='DataPreprocessor'),
    generator=dict(
        type='BigGANGenerator',
        output_scale=128,
        noise_size=120,
        num_classes=1000,
        base_channels=96,
        shared_dim=128,
        with_shared_embedding=True,
        sn_eps=1e-6,
        init_type='ortho',
        act_cfg=dict(type='ReLU', inplace=True),
        split_noise=True,
        auto_sync_bn=False),
    discriminator=dict(
        type='BigGANDiscriminator',
        input_scale=128,
        num_classes=1000,
        base_channels=96,
        sn_eps=1e-6,
        init_type='ortho',
        act_cfg=dict(type='ReLU', inplace=True),
        with_spectral_norm=True),
    # continuous update discriminator for `discriminator_steps = 1` times
    # continuous update generator for `generator_steps = 1` times
    generator_steps=1,
    discriminator_steps=1)

optim_wrapper = dict(
    constructor='MultiOptimWrapperConstructor',
    generator=dict(
        # generator perform `accumulative_counts = 8` times gradient accumulations before␣
↪each update
        accumulative_counts=8,
        optimizer=dict(type='Adam', lr=0.0001, betas=(0.0, 0.999), eps=1e-6)),
    discriminator=dict(
        # discriminator perform `accumulative_counts = 8` times gradient accumulations␣
↪before each update
        accumulative_counts=8,
        optimizer=dict(type='Adam', lr=0.0004, betas=(0.0, 0.999), eps=1e-6)))
```

1.101 Migration of Visualization

In 0.x, MMEditing use `VisualizationHook` to visualize results in training process. In 1.x version, we unify the function of those hooks into `BasicVisualizationHook` / `VisualizationHook`. Additionally, follow the design of MMEngine, we implement `ConcatImageVisualizer` / `Visualizer` and a group of `VisBackend` to draw and save the visualization results.

```
visual_config = dict(
    type='VisualizationHook',
    output_dir='visual',
    interval=1000,
    res_name_list=['gt_img', 'masked_img', 'fake_res', 'fake_img'],
)
```

```
vis_backends = [dict(type='LocalVisBackend')]
visualizer = dict(
    type='ConcatImageVisualizer',
    vis_backends=vis_backends,
    fn_key='gt_path',
    img_keys=['gt_img', 'input', 'pred_img'],
    bgr2rgb=True)
custom_hooks = [dict(type='BasicVisualizationHook', interval=1)]
```

To learn more about the visualization function, please refers to *this tutorial*.

1.102 Migration of AMP Training

In 0.x, MMEditing do not support AMP training for the entire forward process. Instead, users must use `auto_fp16` decorator to warp the specific submodule and convert the parameter of submodule to fp16. This allows for fine-grained control of the model parameters, but is more cumbersome to use. In addition, users need to handle operations such as scaling of the loss function during the training process by themselves.

MMagic 1.x use `AmpOptimWrapper` provided by MMEngine. In `AmpOptimWrapper.update_params`, gradient scaling and `GradScaler` updating is automatically performed. And in `optim_context` context manager, `auto_cast` is applied to the entire forward process.

Specifically, the difference between the 0.x and 1.x is as follows:

```
# config
runner = dict(fp16_loss_scaler=dict(init_scale=512))
```

```
# code
import torch.nn as nn
from mmedit.models.builder import build_model
from mmedit.core.runners.fp16_utils import auto_fp16


class DemoModule(nn.Module):
    def __init__(self, cfg):
        self.net = build_model(cfg)

    @auto_fp16
```

(continues on next page)

(continued from previous page)

```
    def forward(self, x):
        return self.net(x)

class DemoModel(nn.Module):

    def __init__(self, cfg):
        super().__init__(self)
        self.demo_network = DemoModule(cfg)

    def train_step(self,
                   data_batch,
                   optimizer,
                   ddp_reducer=None,
                   loss_scaler=None,
                   use_apex_amp=False,
                   running_status=None):
        # get data from data_batch
        inputs = data_batch['img']
        output = self.demo_network(inputs)

        optimizer.zero_grad()
        loss, log_vars = self.get_loss(data_dict_)

        if ddp_reducer is not None:
            ddp_reducer.prepare_for_backward(_find_tensors(loss_disc))

        if loss_scaler:
            # add support for fp16
            loss_scaler.scale(loss_disc).backward()
        elif use_apex_amp:
            from apex import amp
            with amp.scale_loss(loss_disc, optimizer,
                    loss_id=0) as scaled_loss_disc:
                scaled_loss_disc.backward()
        else:
            loss_disc.backward()

        if loss_scaler:
            loss_scaler.unscale_(optimizer)
            loss_scaler.step(optimizer)
        else:
            optimizer.step()
```

```
# config
optim_wrapper = dict(
    constructor='OptimWrapperConstructor',
    generator=dict(
        accumulative_counts=8,
        optimizer=dict(type='Adam', lr=0.0001, betas=(0.0, 0.999), eps=1e-06),
        type='AmpOptimWrapper',  # use amp wrapper
        loss_scale='dynamic'),
    discriminator=dict(
```

(continues on next page)

(continued from previous page)

```
        accumulative_counts=8,
        optimizer=dict(type='Adam', lr=0.0004, betas=(0.0, 0.999), eps=1e-06),
        type='AmpOptimWrapper',  # use amp wrapper
        loss_scale='dynamic'))
```

```
# code
import torch.nn as nn
from mmagic.registry import MODULES
from mmengine.model import BaseModel


class DemoModule(nn.Module):
    def __init__(self, cfg):
        self.net = MODULES.build(cfg)

    def forward(self, x):
        return self.net(x)

class DemoModel(BaseModel):
    def __init__(self, cfg):
        super().__init__(self)
        self.demo_network = DemoModule(cfg)

    def train_step(self, data, optim_wrapper):
        # get data from data_batch
        data = self.data_preprocessor(data, True)
        inputs = data['inputs']

        with optim_wrapper.optim_context(self.discriminator):
            output = self.demo_network(inputs)
        loss_dict = self.get_loss(output)
        # use parse_loss provide by `BaseModel`
        loss, log_vars = self.parse_loss(loss_dict)
        optimizer_wrapper.update_params(loss)

        return log_vars
```

To avoid user modifications to the configuration file, MMagic provides the `--amp` option in `train.py`, which allows the user to start AMP training without modifying the configuration file. Users can start AMP training by following command:

```
bash tools/dist_train.sh CONFIG GPUS --amp

# for slurm users
bash tools/slurm_train.sh PARTITION JOB_NAME CONFIG WORK_DIR --amp
```

1.103 NPU (HUAWEI Ascend)

1.103.1 Usage

Please refer to the building documentation of MMCV to install MMCV and mmengine on NPU devices.

Here we use 8 NPUs on your computer to train the model with the following command:

```
bash tools/dist_train.sh configs/edsr/edsr_x2c64b16_1xb16-300k_div2k.py 8
```

Also, you can use only one NPU to train the model with the following command:

```
python tools/train.py configs/edsr/edsr_x2c64b16_1xb16-300k_div2k.py
```

1.103.2 Models Results

Notes:

- If not specially marked, the results on NPU with amp are the basically same as those on the GPU with FP32.

All above models are provided by Huawei Ascend group.

1.104 English

1.105

CHAPTER

TWO

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

m

INDEX

Symbols

A

B

C

D

E

F

G

H

I

L

M

N

O

P

R

S

T

U

V

W