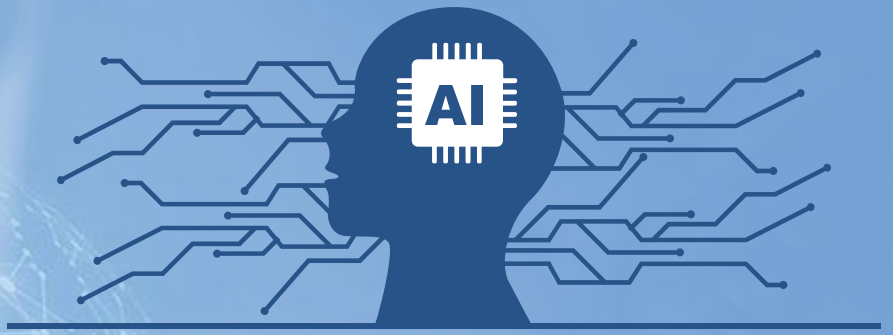
CVPDL: Computer Vision Practice with Deep Learning



Homework #3
Image Generation

Wen-Huang Cheng (鄭文皇)

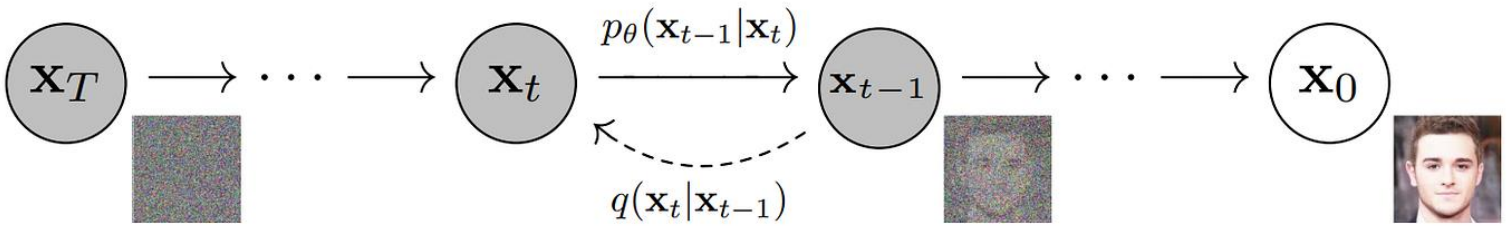
National Taiwan University

wenhuang@csie.ntu.edu.tw

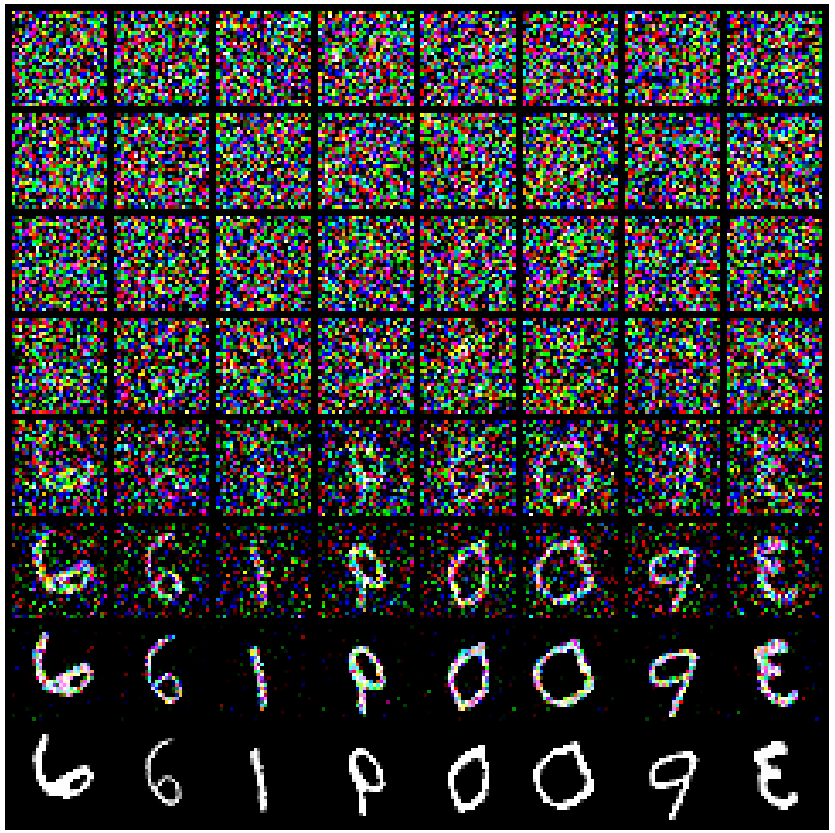
HW3 – Image Generation

TOPIC: Image Generation for Handwritten Digits

- Implementing diffusion generative models.
- Evaluating generative models in terms of FID.



DDPM (NeurIPS 2020, link)



Submission Deadline

2025/11/28 23:59

NTU COOL: Late Submission Policy is Applicable

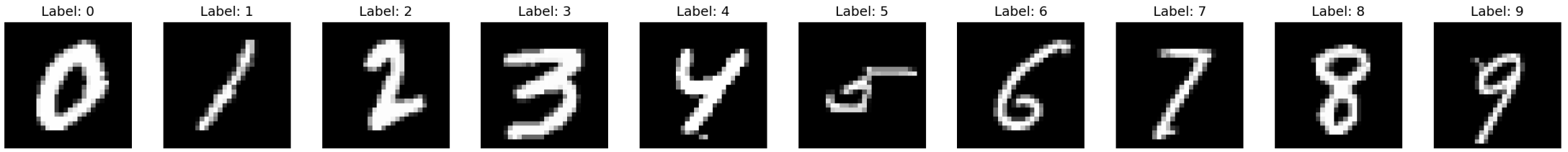
(See TA01-Assignments_Overview.pdf)

NTU & TAICA

- **The grading policy for TAICA students should be determined by your institution‘s teaching assistants or professors. Any questions, please contact them.**

Dataset – MNIST

- Download link: link
- Training set: 60,000 handwritten digits
- Image size: 28x28
- The images have been converted to RGB for simpler implementation



Evaluation Metric – FID

- The Fréchet Inception Distance (FID) is a widely used metric to measure the divergence between two distributions of images, specifically to evaluate the quality of images generated by models.

- A lower FID score indicates that the two distributions are closer, meaning the generated images are more similar to the real ones in terms of quality and diversity.

- For more information, you can refer to link.

FID Implementation

➢ We use `pytorch-fid` (GitHub Repositories) to calculate FID:

```
pip install pytorch-fid
```

➢ Calculate FID (Two ways):

- With the training dataset

```
python -m pytorch_fid path/to/images path/to/mnist
```

- With precalculated mean and covariance

```
python -m pytorch_fid path/to/images path/to/mnist.npz
```

- `path/to/images`: the folder of the generated images.
- `path/to/mnist`: the folder of the training data.
- `path/to/mnist.npz`: the precalculated mean and covariance of training data, which can be downloaded from link.

Grading Policies

- FID (90%) (evaluated between 10000 generated images and training dataset)

Your FID	Points
< 30	90
$30 \sim 70$	Linear between 60 ~ 90
$70 \sim 100$	Linear between 0 ~ 60
> 100	0

- Report (10%)

Programming Spec (1)

- Implement a diffusion-based generative model on your own
- Use Python >= 3.10 (for consistency and reproducibility).
- Pretrained weights and external datasets are forbidden.
- Train your model from scratch.
- You may consult online resources (e.g., link) but you must understand it and write the code yourself.
- No plagiarism.

- Violating the above rules on this page will result in a score of zero.
- If you are uncertain about the legitimacy of the usage, email the TAs for clarification

Programming Spec (2)

➢ Python Package White List:

- Deep learning packages: Tensorflow (including Keras), PyTorch.
- Deep learning related packages: torchvision.
- Machine learning packages: sklearn, xgboost.
- Others: pandas, tqdm, tensorboardX, tensorboard, opencv-python, pillow, scipy, numpy.

- Violating the above rules on this page will result in a score of zero.
- If you are uncertain about the legitimacy of the usage, email the TAs for clarification

Submission File – Source Code

➢ Zip your source code and submit to NTU COOL:

|-------- code_< student-id>.zip

|-------- src/ (Your source code)

|-------- readme.md (Show how to install the environment, dataset, run training & generation)

|-------- requirements.txt (The list of necessary packages)

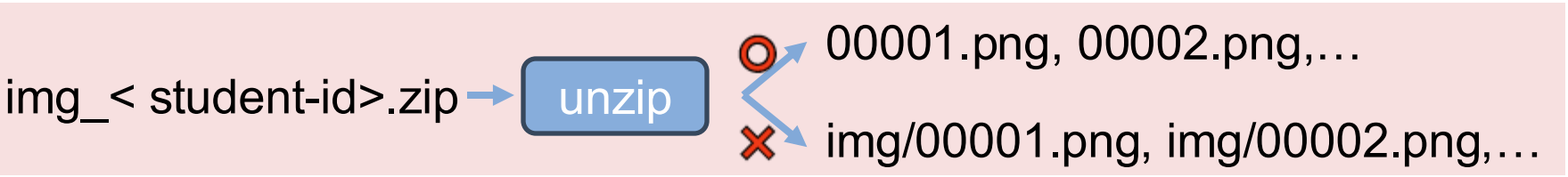
➢ DO NOT include dataset, add the description about dataset path in readme.

➢ DO NOT include model weights.

- An incorrect format will result in a deduction of a -5 score.

Submission File – Generated Images

- The number of generated images: **10000**.
- The dimension of generated images: **28x28 RGB**.
- The format/filename of generated images: **png** (**00001.png ~ 10000.png).**
- Zip these images into a .zip file **without any subdirectory**



|-------- img_< student-id>.zip
|------- 00001.png
|------- 00002.png
…

- **DO NOT** include any files other than images (e.g., for MacOS users, **DO NOT** include .DS_Store and __MACOSX in your zip file).

- Violating the above rules on this page will result in 0 score of FID evaluation.

Submission File – Report (1)

1. **Model Description**
   - Introduce your model (must include an **architecture illustration)**
2. **Implementation Details**
   - Preprocessing, augmentation, hyperparameters, loss functions, training strategies, etc.
3. **Result Analysis**
   - Quantitative improvements (tables, metrics, discussion)
   - Diffusion Process Visualizations (see next page) (5%)
4. **Short conclusion**

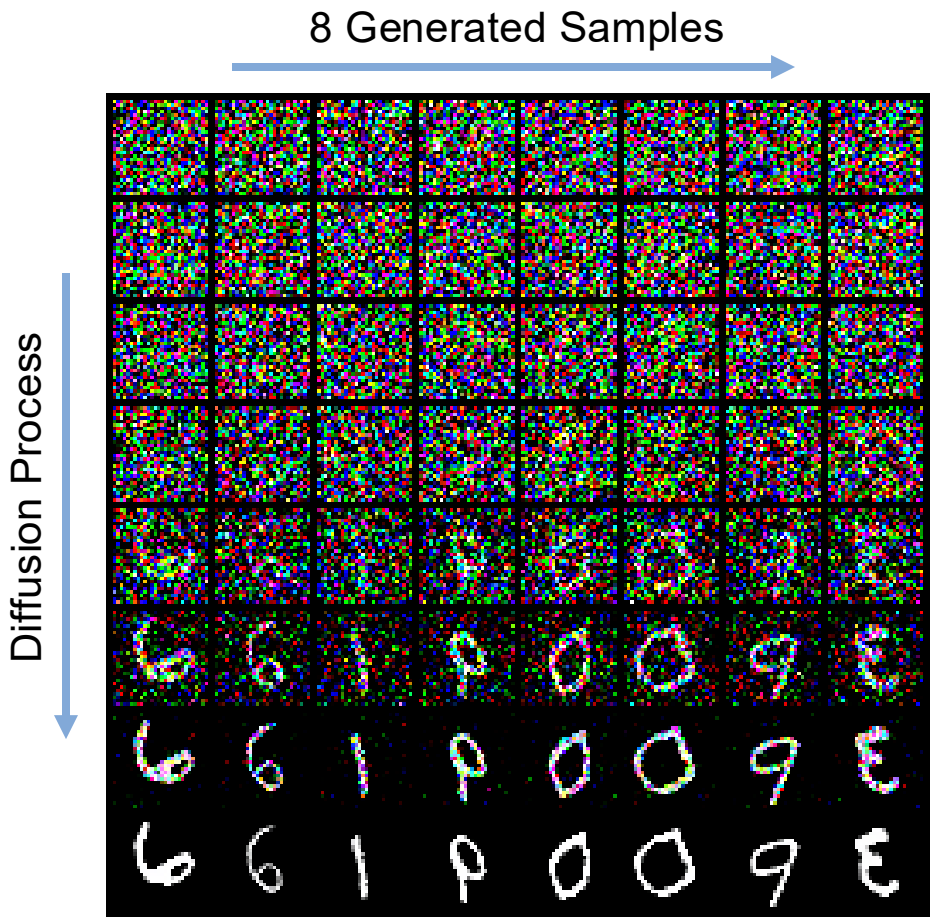
- 3-5 pages (excluding references), exceeding the limit will result in a -5 score

Submission File – Report (2)

➢ Diffusion Process

- Start the diffusion process from noise
- Generate 8 different results.
- Divide the time steps into 7 equal parts.

  For example, if the total number of steps is 1000, record the results every 142 steps.
- Arrange these 64 images into a form similar to the image on the right and paste it on the report.



NTU COOL Submission Rules

➢ Your submission should be a zipped file with the following structure:

- hw3_<student-id>.zip (e.g., hw3_R12345678.zip)

  |-- hw3_<student-id> (Should contain this folder, not separate files)

  |-------- report_<student-id>.pdf (Your report)

  |-------- code_< student-id>.zip (Your source code)

  |-------- img_< student-id>.zip (Your generated images)

- An incorrect format will result in a deduction of a -5 score.
- Failure to re-implement similar performance will result in a 60% discount of the total score.
- Plagiarism in the report or code will result in 0%.

Any Question

Ask peers first

(Join with name: <school_student-id>)

Then ask TAs

(only for NTU students)

cvpdl.ta.fall.2025@gmail.com