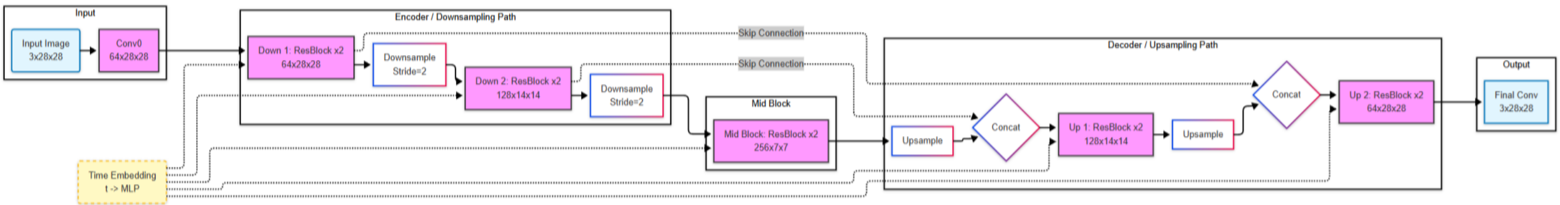
HW3- Image Generation

1. Model Description

➢ Architecture

- Using U-Net-based Denoising Diffusion Probabilistic Models.



2. Implementation Details

➢ Preprocessing

- Convert to RGB.
- Adjust the input image size to 28 x 28 pixels.
- Convert the image data type from PIL Image (integer 0-255) to PyTorch Tensor (floating-point 0.0-1.0).
- Normalization, Convert the numerical range from [0, 1] to [-1, 1].

➢ Hyperparameters

- Training

Parameter	Value
lr	2e-4
batch size	64
epochs	50
optimizer	Adam
loss	MSE Loss

- Diffusion Process

Parameter	Value
time steps	1000
beta schedule	linear
- start	0.0001
- end	0.02

- Model Architecture

Parameter	Value
input image size	28x28
input channels	3
down sampling channels	[64, 128, 256]
up sampling channels	[256, 128, 64]
time embedding dimension	256
GroupNorm groups	8

- Inference

Parameter	Value
total images	10000
steps	1000
visual grid size	8x8
image size	28x28
channels	3

➢ Loss functions

- Mean Squared Error

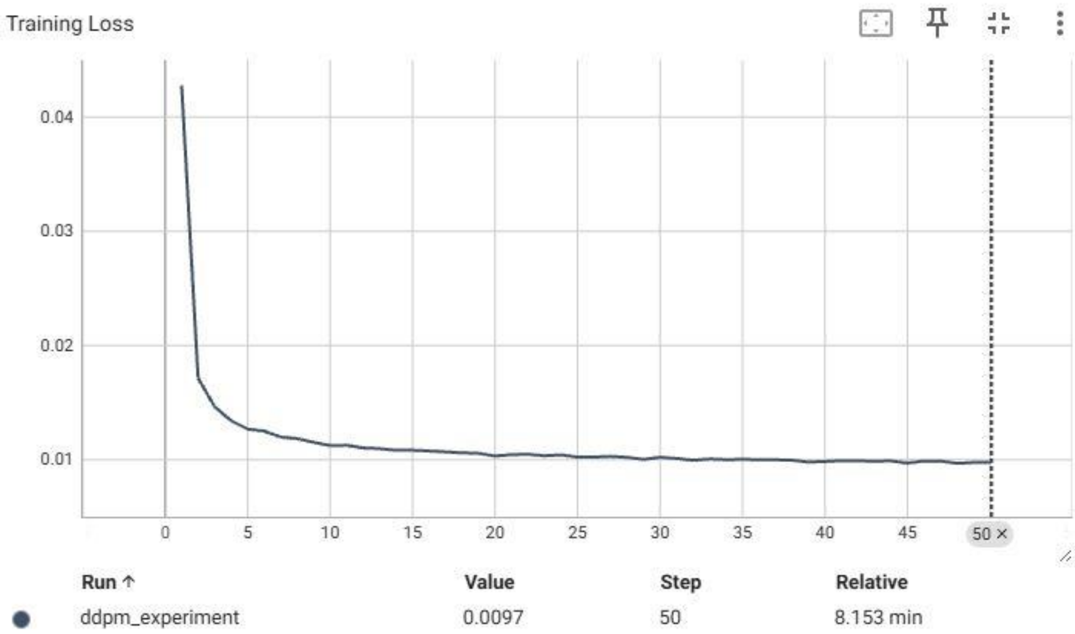
➢ Training strategies

- The model constructed in this work is a Denoising Diffusion Probabilistic Model (DDPM) based on the U-Net architecture. It applies standardized preprocessing to 28x28 RGB MNIST images and is configured with a 1000-step linear noise schedule and the Mean Squared Error (MSE) loss function. The overall implementation adheres to the description in the DDPM paper, integrating modules such as Sinusoidal Position Embedding, Residual Blocks, and Down/Upsample layers to fully reconstruct the generative architecture. Furthermore, a Best Model Checkpoint mechanism is employed to ensure that the numerically most converged model is utilized during inference.

3. Result Analysis

➢ Quantitative improvements

- Training Loss Curve: Use the TensorBoard package to plot the loss curve.

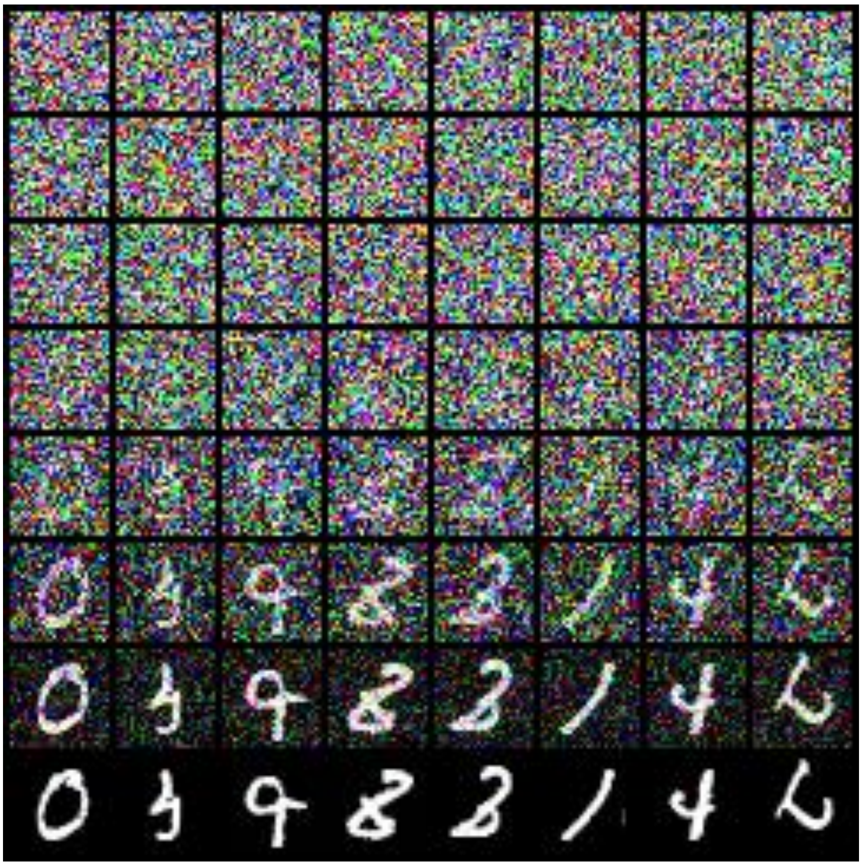


- FID Score

Comparison Target	Dataset Source	FID Score
Training Consistency	Generated vs. Training Set	3.9573
Generalization	Generated vs. Test Set (mnist.npz)	3.9574

```
(CVPDL3) m11315025@ywu:~/Desktop/CVPDL/hw3_M11315025/code_M11315025/src$ python -m pytorch_fid ./img_M11315025 ./datasets
100%|                                                        | 200/200 [00:04<00:00, 42.50it/s]
100%|                                                        | 1200/1200 [00:26<00:00, 44.52it/s]
FID:  3.9573685862829393
(CVPDL3) m11315025@ywu:~/Desktop/CVPDL/hw3_M11315025/code_M11315025/src$ python -m pytorch_fid ./img_M11315025 ./mnist.npz
100%|                                                        | 200/200 [00:04<00:00, 42.48it/s]
FID:  3.9574497541908045
```

- **Diffusion Process Visualizations**



4. Short conclusion

- In this experiment, we implemented a U-Net-based DDPM model applied to the MNIST handwritten dataset. After 50 epochs of training, the model demonstrated excellent convergence and produced high-quality images. In terms of quantitative evaluation, the model achieved an outstanding FID score of 3.9573, indicating that the generated images closely resemble real data in both perceptual quality and diversity. Notably, the FID scores calculated against the training set and the standard test set were nearly identical. Through visual analysis, we observed that the model successfully reconstructed clear, sharp-edged, and diverse RGB digit images from pure Gaussian noise. This experiment verified the stability and high efficiency of DDPM in low-resolution image generation tasks, with the 'Best Model Checkpoint' mechanism ensuring optimal performance during the inference phase.