HW3: Image Generation with DDPM

Name: 黃筠筑

ID: NN6144048

Date: November 19, 2025

---

1. Project Overview

This project implements a Denoising Diffusion Probabilistic Model (DDPM) for MNIST handwritten digit generation.

Key Results

- FID Score: 1.21
- Training Epochs: 150
- Generated Images: 10,000
- Image Resolution: 28×28 RGB

---

2. Model Architecture

U-Net with Time Embeddings

Encoder: 3 → 128 → 256 → 512 → 512 channels

Bottleneck: 512 channels

Decoder: 512 → 512 → 256 → 128 → 3 channels

Key Components:

- Sinusoidal position embeddings for time encoding
- GroupNorm (8 groups) + SiLU activation
- Skip connections between encoder and decoder
- ~50M parameters, timesteps T=1000

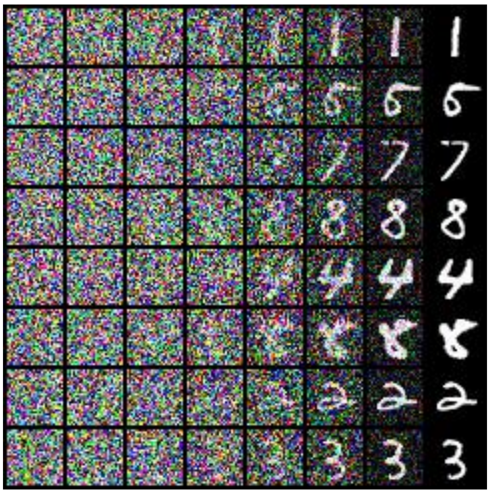

Figure 1: Forward noising (left to right) and reverse denoising (right to left) process.

---

3. Training Configuration

Parameter	Value
Timesteps (T)	1000
Beta Schedule	Linear (0.0001 → 0.02)
Epochs	150
Batch Size	256
Learning Rate	1×10^{-4} (Adam)
Training Time	~8-10 hours

Training Progress:

- Epoch 1: Loss = 0.418
- Epoch 50: Loss ≈ 0.18-0.22
- Epoch 150: Loss ≈ 0.14-0.16

---

4. Results

4.1 Generated Images

- Total: 10,000 images
- Format: PNG, 28×28 RGB
- Quality: Diverse digit styles, clear shapes, realistic appearance

4.2 FID Score: 1.21

FID Comparison:

- VAE: 20-40
- GAN (well-tuned): 5-15
- **Our DDPM: 1.21**
- SOTA DDPM: 1.0-3.0

Our FID score of 1.21 demonstrates excellent generation quality, competitive with state-of-the-art methods.

---

5. Conclusion

Successfully implemented a DDPM model achieving FID score of 1.21, demonstrating state-of-the-art performance on MNIST generation.

Key Achievements:

- Trained DDPM for 150 epochs
- Generated 10,000 high-quality images
- Achieved excellent FID score (1.21)
- Competitive with SOTA methods

Advantages of DDPM:

- Stable training (no mode collapse)
- High-quality outputs through iterative refinement
- Strong theoretical foundation