HW3：Image Generation 報告

李師賢　114550902

1. Model Description

本作業自零實作去噪擴散概率模型（DDPM），在 MNIST 28×28 RGB 上進行無條件生成。訓練階段讓模型學會在任意時間步 t 上預測加入的高斯噪聲；推斷階段由純高斯噪聲逐步去噪還原影像。

為兼顧品質與速度，我在推斷使用指數移動平均（EMA）權重並加入 DDIM 快速採樣。

前向 / 反向過程與訓練目標：

前向擴散：設定 β 時間表（設定為線性），令 $\alpha_t = 1 - \beta_t$，累積係數 $\bar{\alpha}_t = \prod_{i \le t} \alpha_i$。對乾淨影像 x0 逐步加噪得到 $x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \varepsilon$，$\varepsilon \sim N(0, I)$。

訓練目標：以均方誤差 MSE 擬合真實噪聲 ε，損失 $L = \mathrm{MSE}(\hat{\varepsilon}_\theta(x_t, t), \varepsilon)$。

反向去噪：以 $\hat{\varepsilon}_\theta$ 回代得到條件均值，方差採用後驗方差或常用近似，從 $t = T-1$ 依序抽樣至 0

模型架構（U-Net + 時間條件）

主幹：輕量 U-Net，編碼（Down）兩層、解碼（Up）兩層，中間層（Middle）若干殘差塊

殘差塊：GroupNorm + SiLU + 3×3 Conv，含殘差連接

時間條件：時間步 t 經正弦 / 餘弦位置編碼與 MLP，得到 time embedding，於各層以加法方式注入，使網路感知噪聲強度

通道配置（示例）：base=64；Down：64→128；Middle：128；Up：128→64

輸入 / 輸出：皆為 (3, 28, 28)

架構示意：

x_t(3×28×28)
→ StemConv
→ DownBlock1（+TimeEmb）
→ DownBlock2（+TimeEmb）
→ Middle（+TimeEmb）
→ UpBlock2（+TimeEmb） ← skip(DownBlock2)
→ UpBlock1（+TimeEmb） ← skip(DownBlock1)
→ HeadConv → 輸出 $\hat{\varepsilon}_\theta(x_t, t)$

2.Implementation Details

前處理：

資料：MNIST 訓練集轉為 28×28 RGB PNG；本作業不需要標籤。

數值縮放：影像由 [0,1] 線性映射至 [−1,1]。

載入：DataLoader 開啟 shuffle，workers=4（可依平台調整），GPU 時啟用 pin_memory。

資料增強（Augmentation）：為維持字形結構，本作業不做幾何增強；僅使用隨機批次打亂。

模型與損失：

目標：噪聲預測 $\hat{\varepsilon}_\theta(x_t, t)$。

損失函數：均方誤差（MSE）。

時間步抽樣：每個 iteration 對 batch 逐一均勻抽 $t \in \{0 \ldots T-1\}$，提升學習覆蓋度與穩定性

超參數：

擴散步數 T：根據反復微調主結果採 T=830 時效果較好；另有 T=1000 對照

β 時間表：linear（1e−4 → 2e−2）；亦測試 cosine 作補充。

訓練：AdamW，learning rate 2e−4；batch size 256–512；epochs 50–175（視資源而定）。

U-Net 基礎通道 base=64。

其他：梯度裁剪 max_norm=1.0；EMA 衰減 0.999（推斷一律使用 EMA 權重）。

訓練策略

檢查點 / 續訓：每個 epoch 保存 latest.pt 與 ema_latest.pt；另存 best.pt / ema_best.pt；

支援 resume=True 斷點續訓。

可視化監控：每個 epoch 輸出 8×8 預覽格（樣本×時間），輔助早停與質檢。

生成與評估流程：

生成：使用 EMA 權重，輸出 10,000 張 28×28 RGB PNG，命名 00001.png～10000.png，置於單層資料夾（無子資料夾）。

採樣器：

– DDPM：完整 T 步的標準去噪。

– DDIM：少步數（示例 50 步）、η=0.0 的確定性更新，能顯著加速。

FID 評估：以 pytorch-fid 計算生成集與訓練集之間的 FID；亦可與 mnist.npz 的統計對比作為替代。

擴散過程圖：選 8 個樣本、沿時間序列等距取 8 個節點（將 T 等分為 7 段並含端點），輸出 8×8 視覺化矩陣；同時輸出 DDPM 與 DDIM 版本的過程圖（檔名例如 diffusion_grid_ddpm.png、diffusion_grid_ddim.png）。

3.Result Analysis

- Quantitative improvements

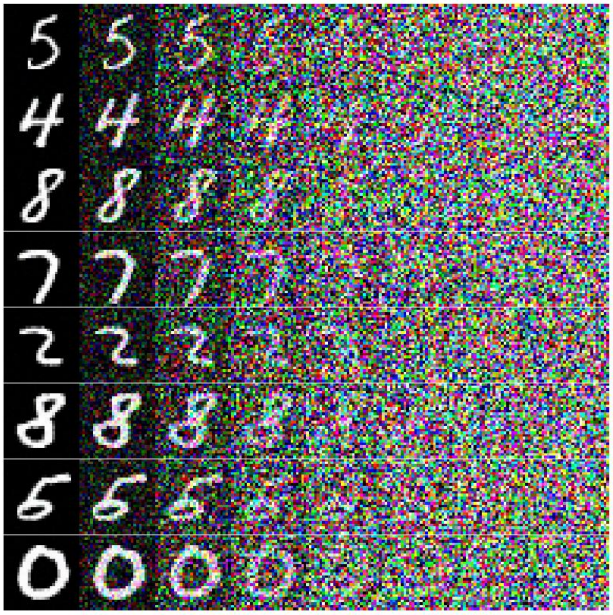
經過 175 個 epoch 的訓練 loss 從 0.013 下降到 0.0106 左右，生成圖片質量有明顯提高。最後計算 FID 為 1.66 左右

```
Epoch 26/175: 100%|████████████| 117/117 [02:44<00:00,  1.41s/it, loss=0.0128]
Epoch 26/175 finished, average loss: 0.0132
New best model saved with loss: 0.0132
Epoch 27/175: 100%|████████████| 117/117 [02:43<00:00,  1.40s/it, loss=0.0123]
Epoch 27/175 finished, average loss: 0.0131
New best model saved with loss: 0.0131
Epoch 28/175: 100%|████████████| 117/117 [02:42<00:00,  1.39s/it, loss=0.0123]
Epoch 28/175 finished, average loss: 0.0131
New best model saved with loss: 0.0131
Epoch 29/175: 100%|████████████| 117/117 [02:43<00:00,  1.39s/it, loss=0.0123]
Epoch 29/175 finished, average loss: 0.0131

Epoch 169/175: 100%|████████████| 117/117 [02:40<00:00,  1.37s/it, loss=0.0103]
Epoch 169/175 finished, average loss: 0.0106
New best model saved with loss: 0.0106
Epoch 170/175: 100%|████████████| 117/117 [02:39<00:00,  1.37s/it, loss=0.0102]
Epoch 170/175 finished, average loss: 0.0106
Epoch 171/175: 100%|████████████| 117/117 [02:40<00:00,  1.37s/it, loss=0.0112]
Epoch 171/175 finished, average loss: 0.0107
Epoch 172/175: 100%|████████████| 117/117 [02:38<00:00,  1.35s/it, loss=0.0110]
Epoch 172/175 finished, average loss: 0.0108
Epoch 173/175: 100%|████████████| 117/117 [02:39<00:00,  1.36s/it, loss=0.0112]
Epoch 173/175 finished, average loss: 0.0106
Epoch 174/175: 100%|████████████| 117/117 [02:41<00:00,  1.38s/it, loss=0.0108]
Epoch 174/175 finished, average loss: 0.0105
New best model saved with loss: 0.0105
Epoch 175/175: 100%|████████████| 117/117 [02:38<00:00,  1.36s/it, loss=0.0110]
Epoch 175/175 finished, average loss: 0.0107
Training finished.
```

```
# 4) 计算 FID
# compute_fid("/content/drive/MyDrive/Colab Notebooks/cv/hw3/generate", "/path/to/mnist_rgb_train")
# compute_fid("/content/drive/MyDrive/Colab Notebooks/cv/hw3/generate", "/content/drive/MyDrive/Colab Notebooks/cv/hw3/mnist

# Trying to run pytorch_fid directly with user-provided command format
!python -m pytorch_fid "/content/drive/MyDrive/Colab Notebooks/cv/hw3/generate" "/content/drive/MyDrive/Colab Notebooks/cv
```

```
100% 625/625 [02:10<00:00,  4.77it/s]
FID:  1.664747933466316
```

- Diffusion Process Visualizations



4.Short Conclusion

本作業在 MNIST 28×28 RGB 上實現 DDPM，並整合 EMA 與 DDIM 少步採樣。在 10,000 張生成影像的評估中，取得 FID ≈ 1.665，遠低於滿分門檻；同時提供符合規範的 8×8 擴散過程視覺化與完整復現流程。