

比較 VAE、GAN、cGAN 與 Diffusion Model 在 MNIST 手寫數字生成的應用

作業目標

- 理解四種生成模型 (VAE、GAN、cGAN、Diffusion) 的基本設計理念
 - 在相同資料集上實作並訓練四種模型
 - 比較它們在生成影像上的 清晰度、穩定性、可控性、效率
-

1) 資料

- MNIST (28×28, 灰階)
 - 使用 `torchvision.datasets.MNIST` 下載
-

2) 模型設計

VAE

- Encoder: 將輸入影像展平, 壓縮成潛在空間 (z), 輸出均值 μ 與對數方差 $\log\sigma^2$
- Decoder: 從 z 還原影像 (28×28)

GAN

- Generator: 輸入隨機噪聲 z (維度 100), 輸出 28×28 假影像
- Discriminator: 輸入影像, 判斷真 / 假

cGAN

- Generator: 輸入隨機噪聲 z + 類別標籤 (one-hot), 輸出指定類別影像

- Discriminator: 輸入影像 + 類別標籤, 輸出真 / 假

Diffusion Model

- Forward: 逐步將圖片加上高斯噪聲
 - Reverse: 訓練一個模型逐步「去噪」生成圖片
 - 使用簡化版本 (如 DDPM / 小型 U-Net) 即可
-

3) 訓練設定

固定

- Batch size: 128
- Optimizer: Adam (建議 $lr=1e-3$ for VAE, $2e-4$ for GAN/cGAN)
- Loss:
 - VAE: 重建 BCE + KLD
 - GAN/cGAN: 對抗 BCE (cGAN 判別器對真樣本採 label smoothing)
 - Diffusion: MSE 去噪損失
- 隨機種子: 固定 seed=42

可彈性調整

- Epoch (建議 30, 可增加至 50+)
 - 隱藏層大小
 - 激活函數
-

4) 輸出結果

1. **VAE**: 隨機生成 10 張影像
2. **GAN**: 隨機生成 10 張影像
3. **cGAN**: 生成數字 0-9 各 10 張 (排成 10×10 圖格)
4. **Diffusion**: 隨機生成 10 張影像

5. 對照圖:將四種方法的結果放在一起, 方便比較
6. 分析四個模型項目
 - 清晰度比較
 - 可控性(是否能指定數字)
 - 訓練/推理效率
 - 穩定性(是否出現模糊或 mode collapse)

作業繳交格式

- 作業內容及輸出範例請參考附檔
- 程式檔 – 包含4個模型, 以**colab**方式繳交
- 文字分析檔, 直接以**word**繳交
- 派一個當代表繳交即可

注意:

- 1. 以 **Colab** 環境撰寫
- 2. 上傳到 **GitHub**
- 3. 繳交 **GitHub** 連結