

一、 程式結構：

hw6_313657003

├─ code

│ ├── crop.py /Data preprocessing

│ ├── train.py/Train Diffusion Unet

│ └── test.py/ Generate and Evaluate pictures

├─ model/<https://drive.google.com/file/d/1fQZg9iXBWY3sTKwjl70Tp53w0TQorwqo/view?usp=sharing>

├─ results/ Save generated 1068 images

└─ report.pdf

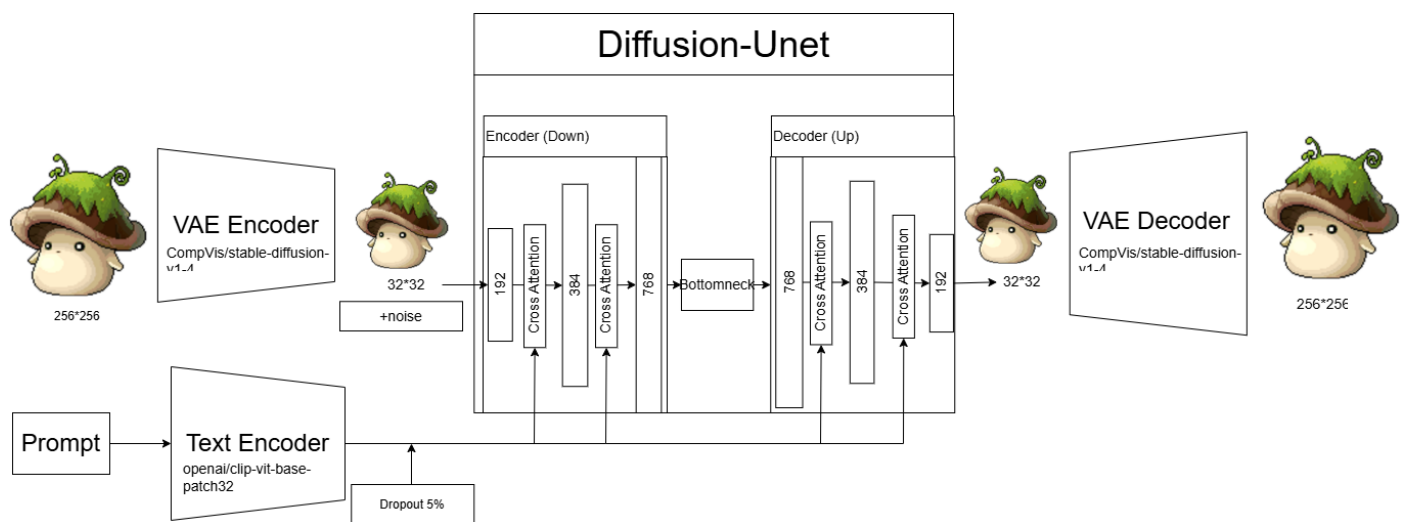
模型雲端連結：

<https://drive.google.com/file/d/1fQZg9iXBWY3sTKwjl70Tp53w0TQorwqo/view?usp=sharing>

二、 資料前處理：

找出圖片中角色（非白色區域），裁切為正方形並補齊邊緣，最後 **resize** 成 256x256 圖片，讓圖片的角色更清楚。

三、 訓練架構：



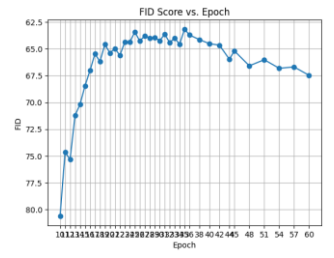
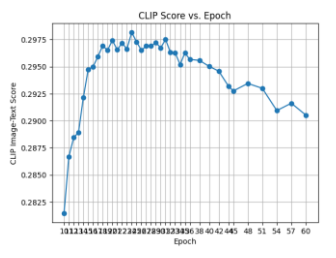
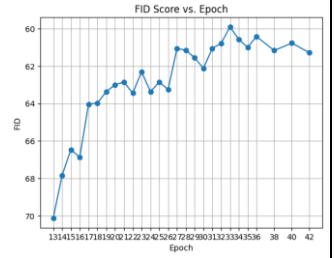
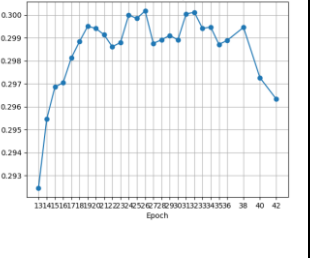
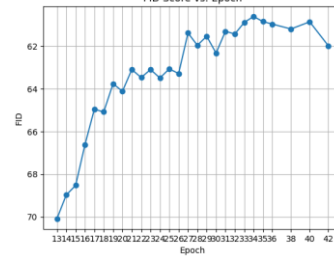
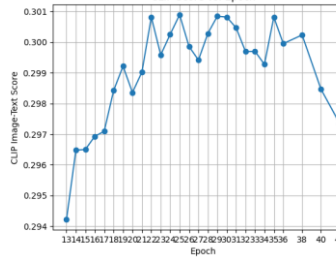
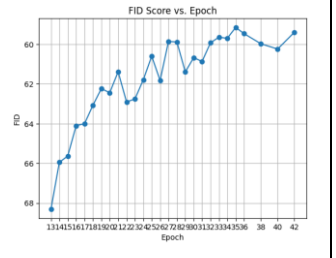
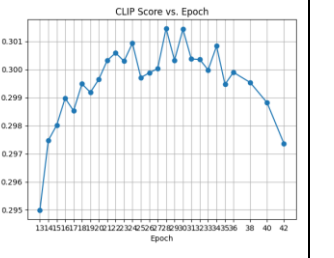
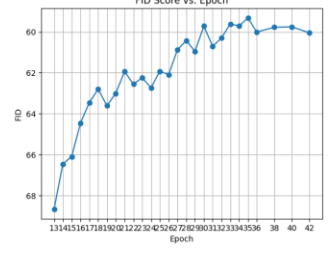
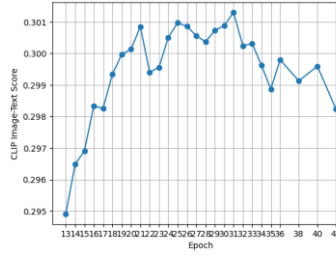
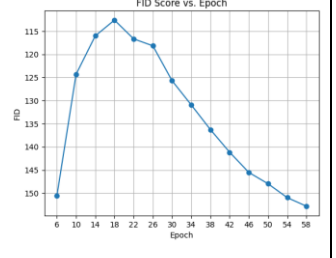
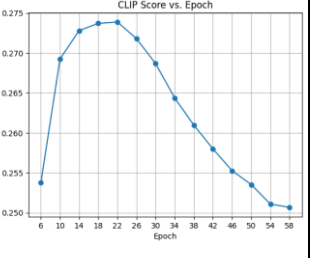
- 訓練參數：Epoch：60, Learning rate: 2×10^{-4} , Batch size：64
- Classifier-Free Guidance：訓練時將 5% 的 prompt 設為空字串，提升模型在推論時引導能力。
- UNet 架構：作為 denoising network，包含 3 層 Down / Up blocks，並在每一層加入 Cross-Attention 以融合由 CLIP Encoder 提供的 Text Embedding。
- Optimizer: AdamW
- LR Scheduler: Cosine with first 5 % steps warmup
- EMA：推論階段使用 EMA 權重（decay = 0.998）以提升穩定性與生成品質

四、 實驗分析：

圖片生成階段嘗試了下列六種組合。其中 DDPM 有使用了 DPMSolverMultistep 作為生成階段的 scheduler。根據表一的結果可發現，使用 DDPM 進行圖片生成的效果整體優於 DDIM。進一步比較不同參數組合的表現後發現：

1. 在相同的 steps 條件下，Classifier-Free Guidance 設為 3 的組合通常優於設為 2 或 4。
2. 當 guidance = 3 時，步數設為 15 的生成品質最佳。雖然 steps = 15 和 steps = 20 的評分結果相近（FID 分別為 59.6989 與 59.7125，CLIP-Text Score 為 0.3008 與 0.3009），但考量到生成時間，最終選擇 steps = 15、guidance = 3 作為最終推論設定。

表一、各種組合 FID 和 CLIP-T 的圖、綜合表現最好的 FID/CLIP-T/CLIP-I

DDPM Step=10, Guidance=2		DDPM Step=10, Guidance=3	
			
Best:63.436/0.297		Best:59.8987/0.2994/0.8246	
DDPM Step=10, Guidance=4		DDPM Step=15, Guidance=3	
			
Best:60.608/0.2993		Best:59.6989/0.3008/0.8276	
DDPM Step=20, Guidance=3		DDIM Step=10, Guidance=2	
			
Best:59.7125/0.3009		Best:110.853/0.275/0.696	