

# Generative AI Homework Week3

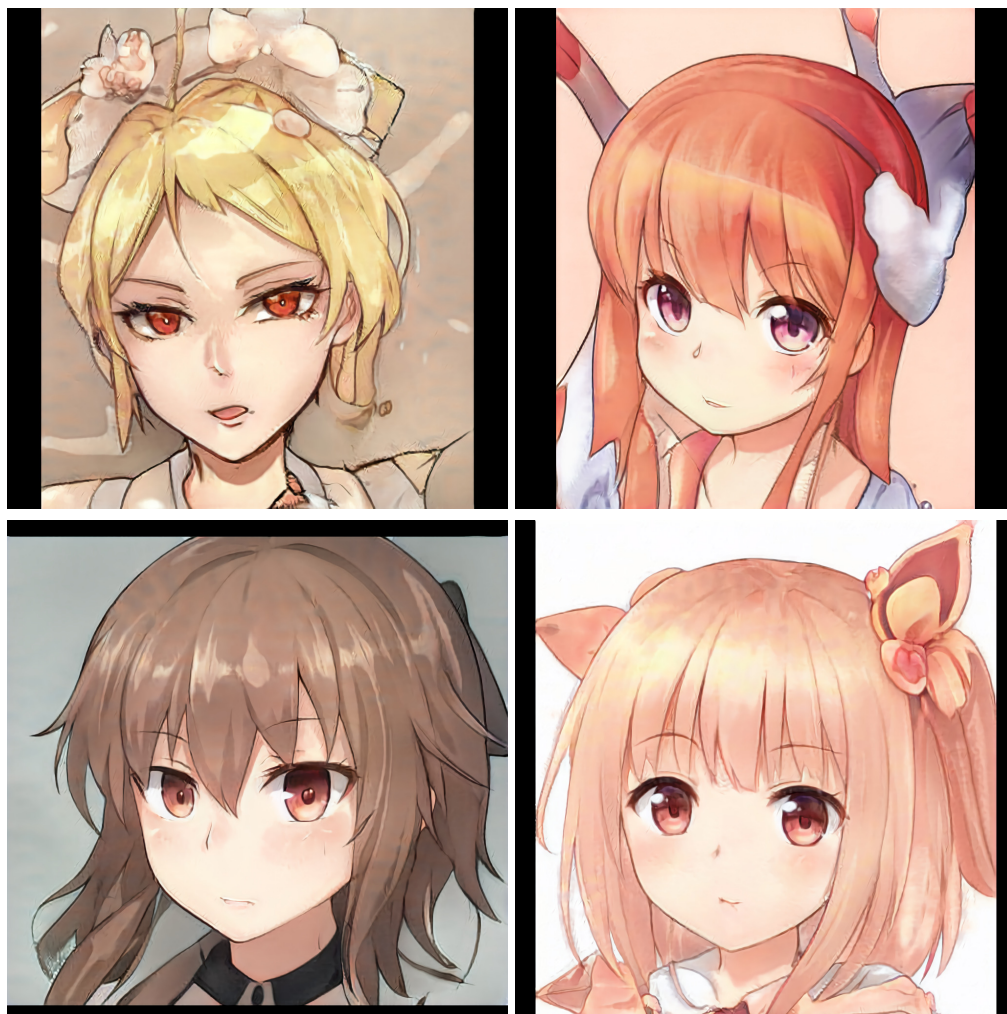
---

- 選擇主題：主題一

## 使用 GAN, Diffusion model 生圖

### GAN

- 使用網站：[this waifu does not exist](https://www.thiswaifudoesnotexist.net/)
  - <https://www.thiswaifudoesnotexist.net/>
- 介紹
  - 這個網站主要基於 StyleGAN (Generative Adversarial Network, 生成對抗網路) 來生成高品質的動漫人臉
  - StyleGAN 透過 風格向量 (Style Vectors) 控制圖像的不同層級特徵，如臉型、表情、頭髮顏色等，使生成結果更加靈活，且無需訓練即可變換特徵透過 潛在空間 (Latent Space) 操作，可以輕鬆調整眼睛、年齡、性別等細節，而不需重新訓練模型
- 生成出來的圖片



## Diffusion model

- 使用網站：[stable diffusion online](https://stablediffusionweb.com)
  - <https://stablediffusionweb.com/zh-tw>
- 介紹：
  - 免費的基於 Stable Diffusion Model 的生成圖片的網頁，使用者可以透過簡單的Prompt 快速生成高品質的圖像
  - Stable Diffusion 是一種潛在擴散模型（Latent Diffusion Model, LDM），能根據文字描述（Prompt）生成高品質圖像。
- 生成出來的圖片(prompt:"Wife of old anime")



## 比較生成結果

- GAN: 明顯在臉部細節上跟 diffusion model 相比有更多不自然的角度，且圖片上有一些噪點
- Diffusion model: 面部細節、線條流暢，雖然仍有不自然的部分，但明顯影像生成品質較前者好

思考: Diffusion model 跟 GAN 的差別，為什麼現在都用 diffusion model

GAN VS. Diffusion model

- GAN (Generative Adversarial Network)
  - 由 Generator 和 Discriminator 兩個神經網路組成，透過對抗訓練的方式學習資料分佈
  - 核心概念: 生成器試圖欺騙鑑別器，使其無法區分「假」的樣本與「真」的樣本
- Diffusion Model
  - 透過 Forward Process 讓影像變得模糊，然後再藉由 Reverse Process 來學習如何從純噪聲中還原影像
  - 核心改念: 透過反向去噪來生成影像

## 為什麼現在都用 Diffusion Model

- 影像品質更高
  - Diffusion Model 生成的影像比 GAN 更清晰、更自然，並且沒有 Mode Collapse
- 更好的控制能力
  - 透過 Stable Diffusion，使用者可以輸入描述性文字來精準控制生成結果，這點是 GAN 很難做到的
- 訓練更穩定
  - GAN 容易因為對抗訓練不穩定，出現 模式崩潰 (Mode Collapse) 或 梯度消失 問題，使得訓練困難。Diffusion Model 訓練雖然慢，但因為採用「去雜訊」方式，學習過程更穩定，可確保模型生成高品質圖像
- 更容易擴展到多模態
  - 圖像 + 文字 + 影片 + 音訊：擴散模型可以用於 Stable Video Diffusion、音訊生成 (AudioLDM)、3D 建模 (DreamFusion)，而 GAN 在這些應用上的效果較差
- 可適應更大規模的計算
  - 透過 Latent Diffusion Model (LDM) (如 Stable Diffusion)，擴散模型可以在較小的潛在空間中運行，提高運算效率，讓消費級 GPU 也能使用