

# 成果報告書

AI臉部表情辨識與風格化形象生成系統

指導教授：林耀鈴 教授

專題學生：

資工四B 411147770 馬詠慶

資工四B 411147754 陳祺紜

資工四B 411147542 洪偉倫

資工四B 411100605 邱博昱

資工四B 411147487 朱聖翔

 **目錄**

章節	標題	頁碼
<b>目錄</b>		2
<b>第一章</b>	<b>序論</b>	3
壹、	前言	3
貳、	動機	3
<b>第二章</b>	<b>報告介紹</b>	3
壹、	實作內容概述	3
貳、	系統架構圖	4
參、	功能概述	5
肆、	使用對象	5
伍、	開發工具	5
<b>第三章</b>	<b>網頁介紹</b>	6
壹、	網頁架構圖	6
貳、	系統畫面流程	7
<b>第四章</b>	<b>模型介紹</b>	7
壹、	模型功能與架構	7
<b>第五章</b>	<b>補充說明</b>	8
壹、	遇到的困難	8
貳、	結論及未來發展	8
參、	經費預算需求表	9
肆、	工作分配	9
伍、	名詞解釋	10

陸、	參考資料	101
----	------	-----

## 第一章 序論

### 壹、前言

本專題目標是整合多個AI模型，建立一個穩定、可操作、可重複生成結果的**自動化風格貼圖生成系統**。該系統旨在將使用者上傳的臉部照片風格與特徵，結合文字提示和姿勢圖像，輸出具有卡通風格化的角色形象。

### 貳、動機

透過整合 SDXL、IP-Adapter、ControlNet、SAM 等多模型，串接成一條**自動化工作流**，解決單一模型在圖像生成品質、姿勢控制和臉部細節上的不足。最終建立一個可重用的 Web 系統架構，實現多任務生成，並透過友善介面服務使用者。

## 第二章 報告介紹

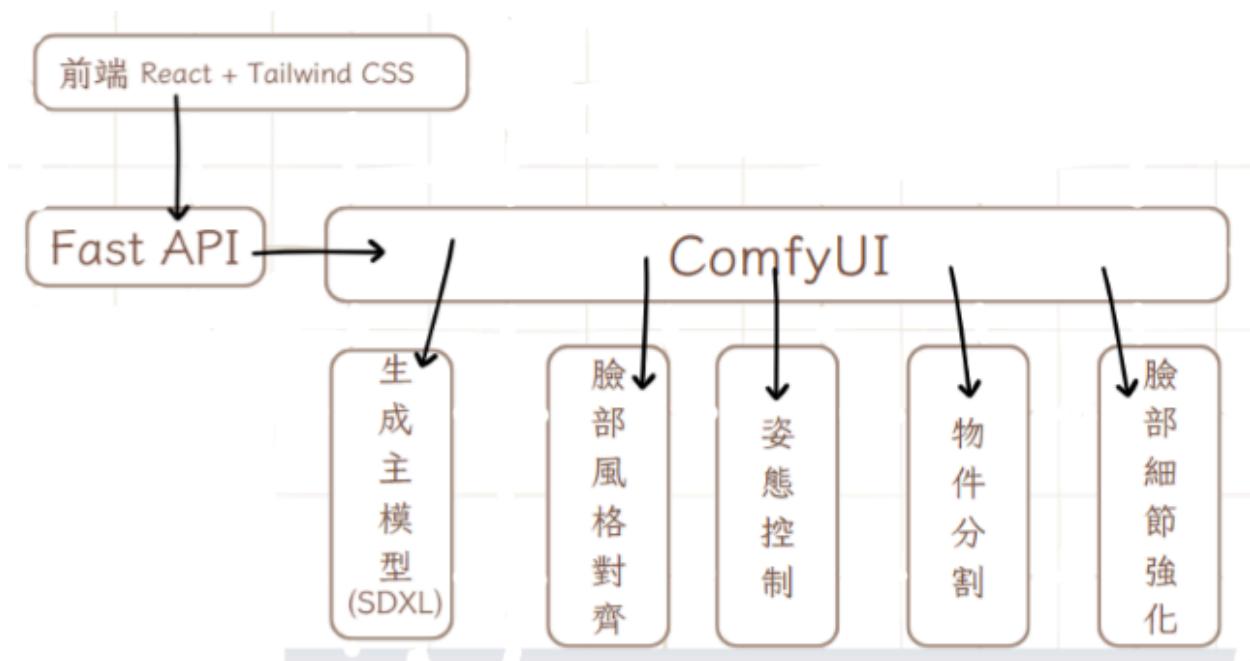
### 壹、實作內容概述

本專題的實作核心是整合多個 AI 模組，建立一個**穩定、可操作、可重複生成結果的自動化風格貼圖生成系統**。

面向	實作內容	技術重點
系統整合與任務自動化	SDXL, IP-Adapter, ControlNet, SAM 等多模型串接成一條自動化工作流，讓使用者上傳人臉與姿勢圖後，系統自動完成生成流程。	使用 ComfyUI Workflow 架構，確保多模型能正確協同運作。使用 FastAPI + WebSocket，實現即時任務派送與進度回傳。
影像處理流程	多階段生成：人臉融合 → 姿勢對齊 → 風格化輸出。	使用 ControlNet + IP-Adapter 實現條件控制式生成。
前後端整合	React + Tailwind CSS 製作互動介面，後端以 FastAPI 控制 ComfyUI。	建立可重用的 Web 系統架構，支援多任務生成。

## 貳、系統架構圖

本系統採用前後端分離，以 FastAPI 作為 API 伺服器，負責與 ComfyUI(模型引擎)進行通訊和任務管理。



角色/元件	服務內容	資料來源/去向
使用者	網頁瀏覽器操作。	上傳人臉圖片、姿勢圖片，接收風格化貼圖結果。
應用程式伺服器 (FastAPI)	任務管理、WebSocket 連線、派送生成任務至 ComfyUI。	接收前端請求，發送指令到 ComfyUI 引擎。
生成引擎 (ComfyUI)	生成主模型 (SDXL)、臉部風格對齊 (IP-Adapter Plus)、姿態控制 (ControlNet)、物件分割 (SAM)、臉部細節強化 (FaceDetailer)。	處理輸入圖片與文字提示，輸出最終圖像。

## 參、功能概述

- 風格化貼圖生成**: 將使用者人臉照片，轉換為指定風格(例如:3D 卡通)的圖像。
- 人臉特徵融合**: 確保生成圖像保留使用者臉部的個人特徵和風格(透過 IP-Adapter 實現)。
- 姿勢精確控制**: 允許使用者上傳姿勢圖，以精確控制生成角色的身體姿態(透過 ControlNet 實現)。
- 臉部細節強化**: 自動對生成圖像的臉部進行精修，提高五官的清晰度和自然度(透過 FaceDetailer 實現)。

## 肆、使用對象

- AI 圖像設計與內容創作者。
- 虛擬角色 (VTuber) 開發者。
- 對 AI 生成技術感興趣的一般使用者。

## 伍、開發工具

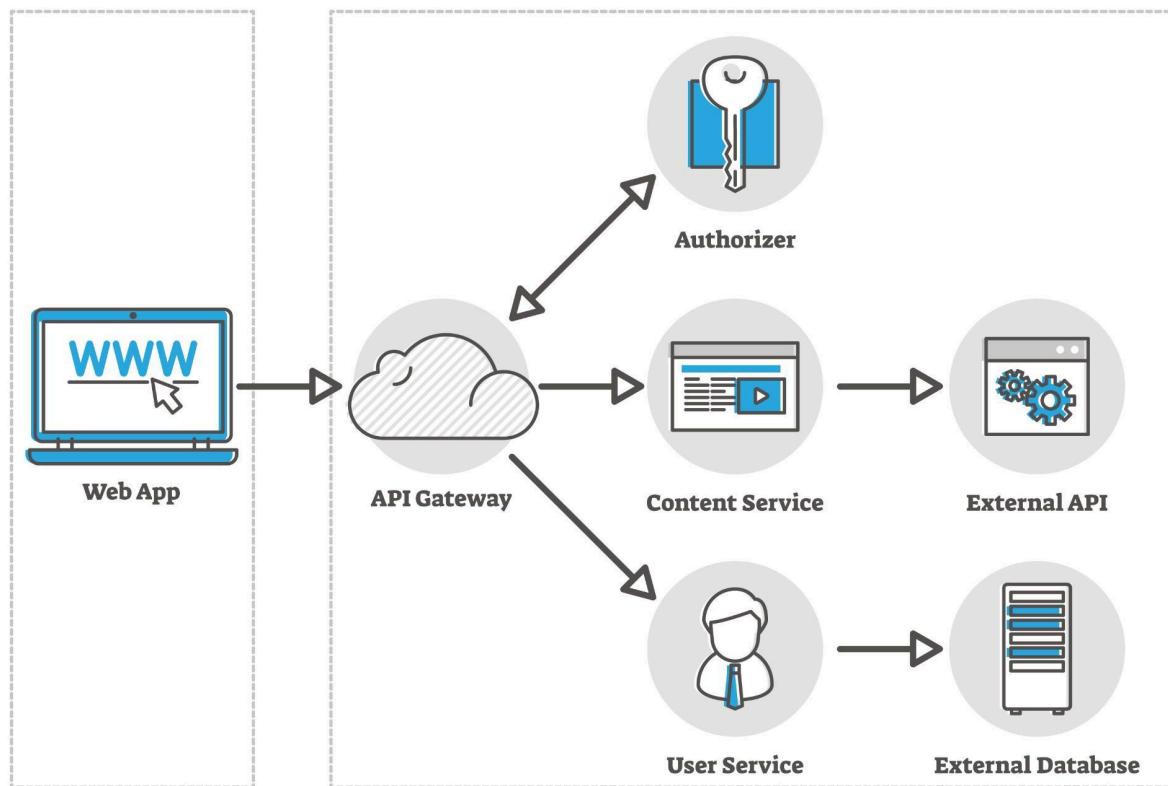
類別	工具名稱/套件	備註
模型引擎	ComfyUI	視覺化 Stable Diffusion 工作流管理介面。
主模型	SDXL (Stable Diffusion XL)	文本到影像生成模型。
前端	React + Tailwind CSS	製作互動介面。
後端	FastAPI	負責控制 ComfyUI 及 WebSocket 連線。
核心 AI 模組	IP-Adapter Plus, ControlNet (DWPose), SAM, FaceDetailer	實現風格對齊、姿勢控制、物件分割與臉部精修。

## 第三章 網頁介紹

### 壹、網頁架構圖

網頁的核心功能為圖像生成介面與任務排隊/進度顯示。

# SERVERLESS



模組	內容
模組	內容

<b>主視覺/首頁</b>	展現系統的風格與功能介紹。
<b>生成介面</b>	<b>核心功能:</b> 包含人臉圖上傳、姿勢圖上傳、風格提示輸入、生成按鈕。
<b>任務佇列/紀錄</b>	顯示當前任務的佇列狀態、進度百分比、執行時間。
<b>結果展示</b>	顯示最終生成的風格化圖像。

## 貳、系統畫面流程

### 系統畫面流程(圖像生成)

1. **開始:**使用者進入生成介面。
2. **使用者上傳圖片:**上傳 Face Image (人臉) 和 Pose Image (姿勢)。
3. **使用者輸入提示:**輸入 Text Prompt (風格描述)。
4. **FastAPI 派送任務:**將所有輸入資料打包，派送至 ComfyUI 佇列。
5. **ComfyUI 執行工作流:**
  - CLIP/IP-Adapter Plus 提取人臉風格。
  - ControlNet (DWPose) 提取姿勢骨架。
  - SDXL 進行風格化圖像生成。
  - SAM/FaceDetailer 進行臉部細節強化。
6. **WebSocket 即時回傳進度:**顯示百分比 (例如: 164.07%)。
7. **顯示生成結果:**在網頁上展示最終的風格化圖像。
8. **結束。**

## 第四章 模型介紹

### 壹、模型功能與架構

模型/模組	功能概述	技術細節
<b>生成主模型 (SDXL)</b>	負責影像風格的文本到影像生成。	Base + Refiner 雙階段架構。本系統採用 Samaritan 3d Cartoon 模型。
<b>CLIP / IP-Adapter Plus</b>	<b>臉部風格對齊:</b> 將人臉特徵嵌入生成圖像。	使用 CLIP Vision 提取人臉特徵向量。架構中包含 Linear、LN、Decoupled Cross-Attention 層。

<b>ControlNet (DWPose)</b>	<b>姿態控制</b> : 精確控制生成人物的姿勢。	DWPose 比 OpenPose 更精準，特別適用於 SDXL。
<b>物件分割 (SAM)</b>	<b>高精度遮罩</b> : 自動偵測並分割影像中的人物、臉部或配件。	用於輔助 FaceDetailer 進行精準局部修飾。
<b>臉部細節強化 (FaceDetailer)</b>	<b>臉部 Inpainting</b> : 自動化重繪臉部五官，提升臉部解析度。	解決 SDXL 生成臉部模糊問題，支援多輪 Refinement。

## 第五章 補充說明

### 壹、遇到的困難

困難描述	改善方式
<b>多模型串接複雜度高</b>	採用 ComfyUI Workflow 進行視覺化管理，確保多模型能正確協同運作。
<b>臉部細節生成不自然</b>	導入 FaceDetailer 模組，專門針對臉部進行重繪強化。
<b>前後端即時通訊</b>	使用 FastAPI + WebSocket 實現即時任務派送與進度回傳。

### 貳、結論及未來發展

項目	內容
<b>結論</b>	成功整合多個最先進的 AI 模組，形成一個完整的可用系統，實現了穩定、可控、具備高品質輸出能力的自動化風格貼圖生成系統。
<b>短期目標</b>	1. 減少圖片生成所耗費的時間 2. 優化模型結構
<b>長期目標</b>	更多生圖功能開發

### 參、經費預算需求表 (執行中所需之經費項目單價明細)

項目名稱	說明	單位	數量	單價 臺幣 (元)	小計 臺幣 (元)	備註
個人電腦	專案之進行	部	1	26000	26000	系上實驗室提供
雷射印表機	列印所需文件	部	1	100	100	系上實驗室提供
消耗性器材	印表機消耗材料、紙張等	批	1	200	200	系上實驗室提供
消耗性器材	光碟片、隨身碟、外接硬碟等	批	1	300	300	自行負擔
共計					26600	

#### 肆、工作分配 (詳述參與人員分工)

學生姓名	分工內容
馬詠慶	ComfyUI工作流設計
陳祺紘	網頁開發、設計
朱聖翔	製作流程圖
洪偉倫	製作流程圖
邱博昱	網頁開發、設計

#### 伍、名詞解釋

詞彙	定義

<b>SDXL</b>	Stable Diffusion XL, 高階文本到影像生成模型，採用 Base + Refiner 雙階段架構。
<b>ComfyUI</b>	開源視覺化的 Stable Diffusion 工作流管理介面，允許透過拖曳節點組合 AI 模組。
<b>IP-Adapter</b>	Image Prompt Adapter, 將輸入圖像的風格與特徵嵌入擴散模型，用於人臉融合。
<b>ControlNet</b>	一種神經網路架構，為擴散模型提供額外的條件控制(例如：姿勢、邊緣)。
<b>DWPose</b>	ControlNet 體系中用於人物姿勢控制的預處理器和模型，比 OpenPose 更精確。
<b>SAM</b>	Segment Anything Model, 用於影像中的自動物件分割，提供高精度遮罩。

## 陸、參考資料

- <https://www.patreon.com/posts/comfyuijiao-xue-97348864>
- <https://chrislee0728.medium.com/stable-diffusion-xl-%E8%AA%AA%E6%98%8E%E5%8F%8A-sdxl1-0-%E5%AE%89%E8%A3%9D%E6%95%99%E5%AD%B8-b2f7648d49ff>
- <https://chrislee0728.medium.com/stable-diffusion-xl-%E8%AA%AA%E6%98%8E%E5%8F%8A-sdxl1-0-%E5%AE%89%E8%A3%9D%E6%95%99%E5%AD%B8-b2f7648d49ff>
- <https://github.com/comfyanonymous/ComfyUI>