專題流程

WZX
7 min read · Just now

技術整理摘要

1. 前端 (Frontend)

框架/函式庫 (Framework/Library): React

語言 (Language): TypeScript

樣式 (Styling): Tailwind CSS

主要功能:

- 提供使用者介面上傳兩張圖片。
- 呼叫後端 API 並傳送圖片數據。
- 接收後端回傳的影片數據。
- 在介面上播放/顯示結果影片。

2. 後端 (Backend)

框架 (Framework): FastAPI (Python)

聯網工具 (Networking): ngrok (用於將本地開發的 FastAPI 服務放到公網，方便前端訪問或展示)

主要功能:

- 提供 API 端點 (Endpoint) 接收前端的圖片上傳請求。
- 協調執行整個影像處理與 AI 生成的流程。
- 管理和調用不同的 AI 模型與算法庫。
- 將最終生成的影片回傳給前端。

3. 核心演算法 & AI 模型 (Core Algorithm & AI Models)

影像對齊 (Image Alignment):

- **SIFT (Scale-Invariant Feature Transform):** 用於偵測影像中的關鍵特徵點。
- **FLANN (Fast Library for Approximate Nearest Neighbors):** 用於快速匹配兩張影像中的 SIFT 特徵點。
- **MAGSAC++:** 先進的 RANSAC 變體，用於從匹配的特徵點中穩健地估計轉換模型（例如：單應性矩陣 Homography），以對齊影像，能有效處理大量的離群點 (outliers)。

圖像轉文字 (Image-to-Text / Image Captioning):

- **Florence-2-large:** Microsoft 的視覺基礎模型，能夠執行多種視覺任務，在此用於生成輸入圖片的詳細文字描述。

文字轉 Prompt (Text-to-Prompt Generation):

- **Phi-4 :** 大型語言模型 (LLM)，在此用於理解兩段圖像描述，並生成一系列引導 Stable Diffusion 進行語義過渡的 Prompts。

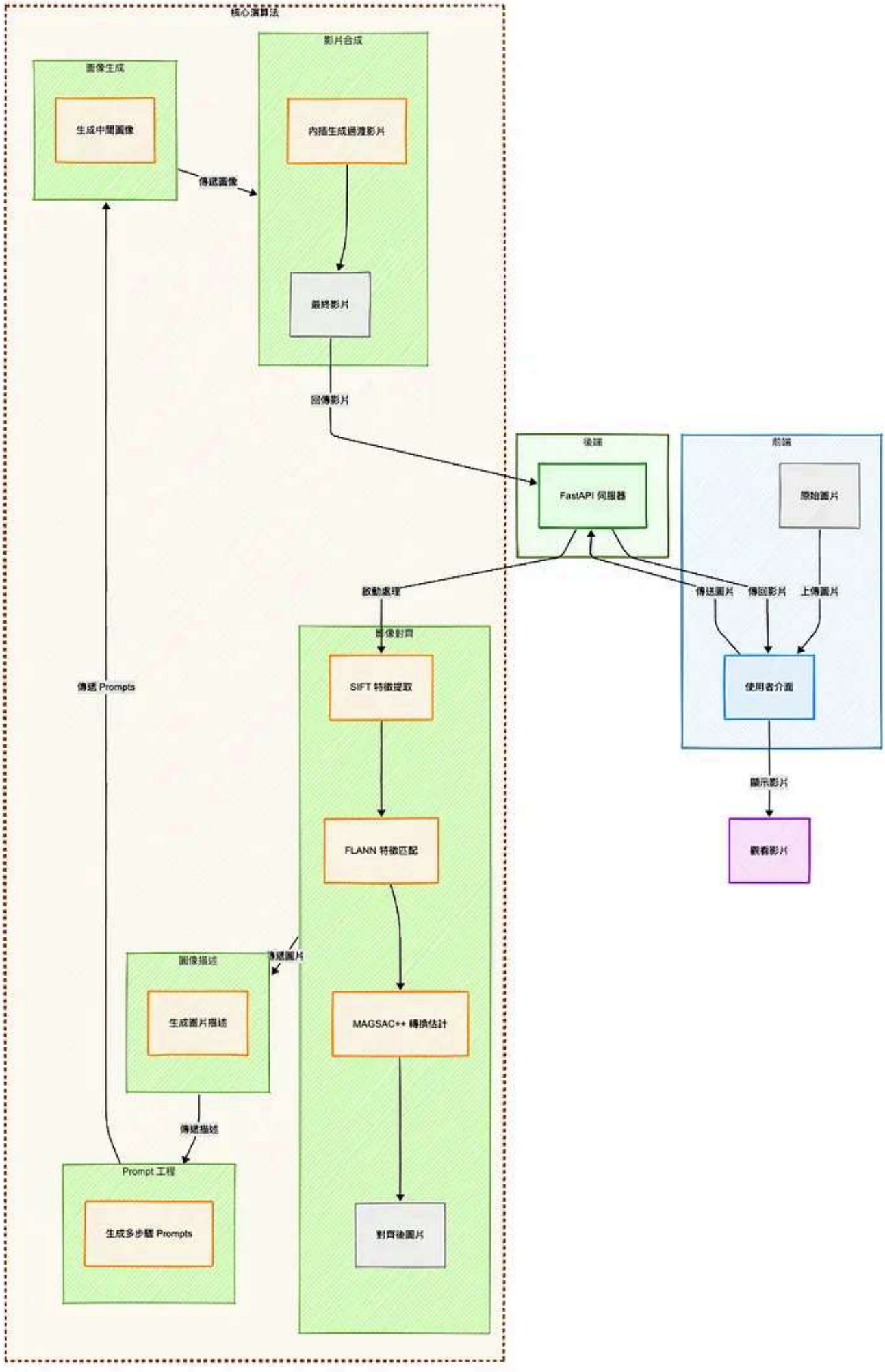
圖像生成 (Image Generation):

- **Stable Diffusion:** 基於擴散模型的文本到圖像生成模型，根據 Phi-4 生成的 Prompts 創造中間過渡圖像。

影像處理 (Image Processing):

- **內插法 (Interpolation):** 技術如幀插值 (Frame Interpolation) ，用於在 Stable Diffusion 生成的關鍵幀圖像之間創建平滑的過渡幀，最終組合成影片。

流程圖


核心演算法
影片合成
圖像生成
生成中間圖像
內插生成過渡影片
傳遞圖像
最終影片
回傳影片
後端
FastAPI 伺服器
前端
原始圖片
啟動處理
傳送圖片
傳回影片
上傳圖片
影像對齊
SIFT 特徵提取
使用者介面
傳送 Prompts
顯示影片
FLANN 特徵匹配
觀看影片
傳送圖片
圖像描述
生成圖片描述
MAGSAC++ 轉換估計
傳送描述
Prompt 工程
生成多步驟 Prompts
對齊後圖片

```
---
config:
  theme: forest
  look: handDrawn
  layout: dagre
  fontSize: 16
  flowchart:
    nodeSpacing: 40
    rankSpacing: 80
    padding: 15
---
flowchart TD
 subgraph UserInterface["前端"]
    direction LR
        UI["使用者介面"]
        UserInput["原始圖片"]
  end
 subgraph BackendInfra["後端"]
    direction LR
        API["FastAPI 伺服器"]
  end
 subgraph Align["影像對齊"]
    direction TB
        Flann["FLANN 特徵匹配"]
        Sift["SIFT 特徵提取"]
        Magsac["MAGSAC++ 轉換估計"]
        AlignedImages["對齊後圖片"]
  end
 subgraph Describe["圖像描述"]
    direction TB
        Florence["生成圖片描述"]
  end
 subgraph Prompting["Prompt 工程"]
    direction TB
        Phi3["生成多步驟 Prompts"]
  end
 subgraph Generate["圖像生成"]
    direction TB
        StableDiffusion["生成中間圖像"]
  end
 subgraph Interpolate["影片合成"]
    direction TB
        FinalVideo["最終影片"]
        Interpolation["內插生成過渡影片"]
  end
 subgraph ProcessingPipeline["核心演算法"]
    direction TB
        Align
        Describe
        Prompting
        Generate
```

```
        Interpolate
    end
      UserInput -- 上傳圖片 --> UI
      Sift --> Flann
      Flann --> Magsac
      Magsac --> AlignedImages
      Interpolation --> FinalVideo
      Align -- 傳遞圖片 --> Describe
      Describe -- 傳遞描述 --> Prompting
      Prompting -- 傳遞 Prompts --> Generate
      Generate -- 傳遞圖像 --> Interpolate
      UI -- 傳送圖片 --> API
      API -- 啟動處理 --> Sift
      FinalVideo -- 回傳影片 --> API
      API -- 傳回影片 --> UI
      UI -- 顯示影片 --> UserView["觀看影片"]
       UI:::frontendStyle
       UserInput:::dataStyle
       API:::backendStyle
       Flann:::stepStyle
       Sift:::stepStyle
       Magsac:::stepStyle
       AlignedImages:::dataStyle
       Florence:::stepStyle
       Phi3:::stepStyle
       StableDiffusion:::stepStyle
       FinalVideo:::dataStyle
       Interpolation:::stepStyle
       UserView:::finalStyle
      classDef frontendStyle fill:#e6f3ff,stroke:#1e90ff,stroke-width:2px,border-
      classDef backendStyle fill:#e6ffe6,stroke:#228b22,stroke-width:2px,border-r
      classDef stepStyle fill:#fff5e6,stroke:#ff8c00,stroke-width:2px,border-radi
      classDef dataStyle fill:#f0f0f0,stroke:#666,stroke-width:1px,border-radius:
      classDef finalStyle fill:#fde0ff,stroke:#9400d3,stroke-width:2px,border-rad
      style UserInterface fill:#f0f8ff,stroke:#4682b4,stroke-width:2px,border-rad
      style BackendInfra fill:#f0fff0,stroke:#556b2f,stroke-width:2px,border-radi
      style ProcessingPipeline fill:#fffaf0,stroke:#8b4513,stroke-width:2px,borde
```

Workflow

Edit profile