**委外專題 : 影片即時 AI 換臉**

* 目錄
* 專題題目說明 (未含個人特質之音樂)
* 專題成果標準
* 可能運用的技術
* 訓練環境
* 範例與建議
* 功能要求
* 專題題目說明 :

使用者可以透過 AI 技術將使用者的臉換到指定影片中的特定角色上，且不影響影片動態效果。

* 專題成果標準 :

換臉應不影響角色在影片中做的各種動作、角度 (包含轉圈、甩頭、遮臉等複雜的動作)。臉不能有邊緣感，不自然的抖動，或者在任何角度有任何違和、不自然之處。需做到不會邊緣模糊、陰影不均勻、影像閃動和頭髮跟背景交界穿幫等情況並可眨眼且眨眼自然

須以”寫實風格”、”擬真” 為優先，但亦能兼顧油畫、漫畫、水彩等其他風格。

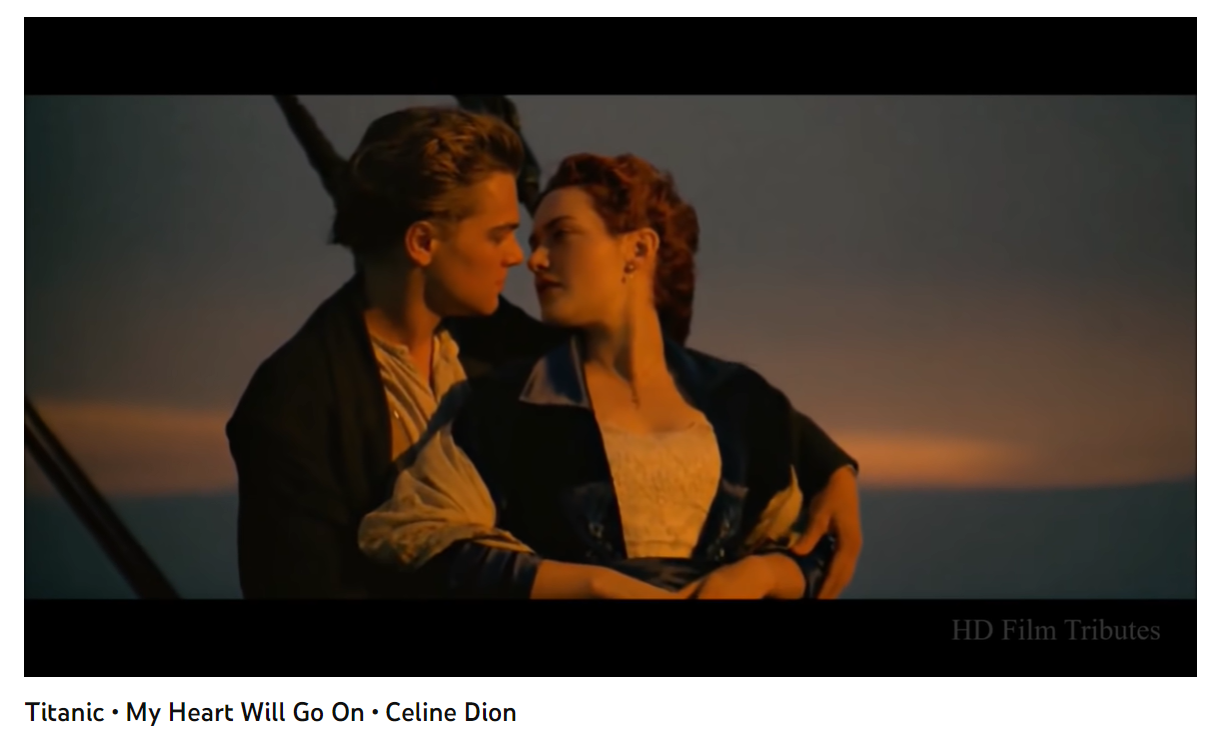
* 可能運用的技術 :

|  |  |  |
| --- | --- | --- |
| **No.** | **Technology** | **Description** |
| 1 | Stable Diffusion | 文生圖擴散模型  官網 : <https://stablediffusionweb.com/>  架設教學 : <https://blog.jimmycstudio.com/stablediffusion-local-install/>  維基百科 : <https://zh.wikipedia.org/zh-tw/Stable_Diffusion>  硬體需求 : 能在 NV 30xx/40xx系列之顯卡上執行 |
| 2 | ffmpeg | 開放原始碼的自由軟體, 可用於錄影, 轉檔, 串流功能  官網 : <https://ffmpeg.org/>  ffmpeg 常用指令 : <https://hackmd.io/@kd01/B137F9ghO?type=view> |
| 3 | Lora | 以Lora fine-tune dreambooth 達到更好的生成效果  Lora github : <https://github.com/cloneofsimo/lora>  Lora 介紹 : <https://replicate.com/blog/lora-faster-fine-tuning-of-stable-diffusion> |
| 4 | Batch-Face-Swap | Stable Diffusion Extension批次臉部遮罩  自動偵測圖片中的人臉並形成遮罩，是一款專門用於 AI 換臉的插件  Batch-Face-Swap github : <https://github.com/kex0/batch-face-swap> |

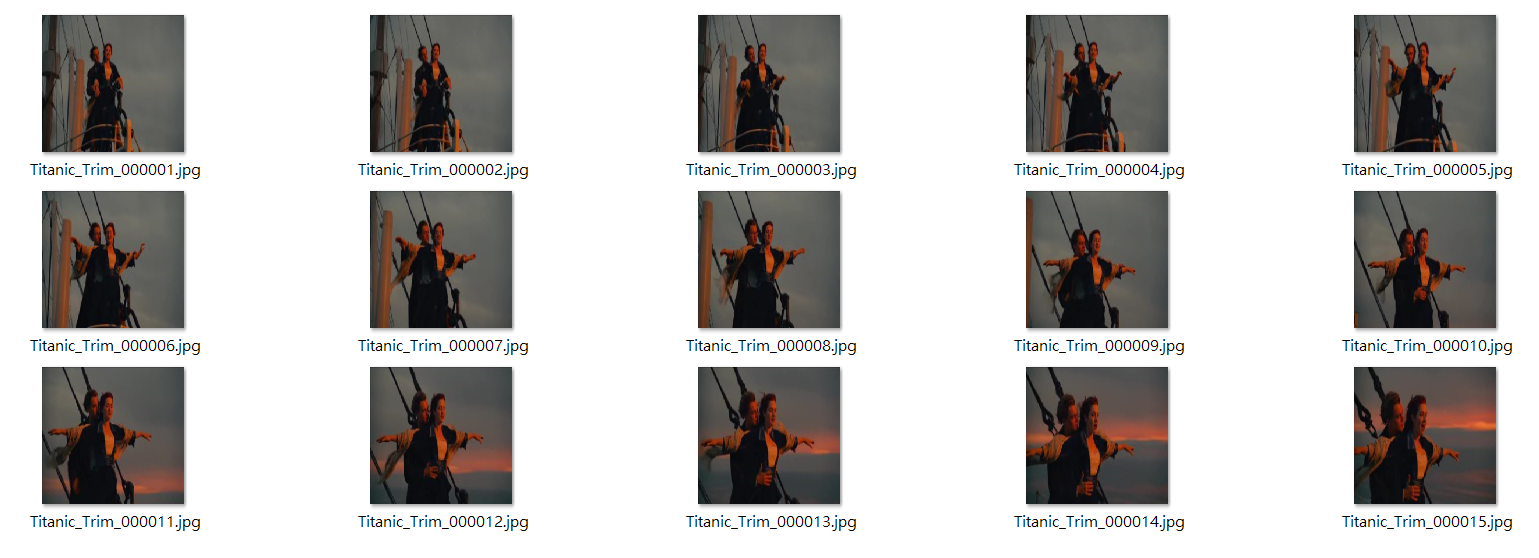
* 訓練環境
  + 北極雲海 (Linux虛擬機) (ucloud.moremote.com) RegionTwo
  + Ubuntu20.04
  + 3080 / 16G VRAM
  + M2模板
* 範例與建議 :

1. 此處以 : 鐵達尼號經典片段(近)、周子瑜熱舞(中)、韓國不知名舞者熱舞(遠) 作為實驗題材 :

|  |  |  |
| --- | --- | --- |
| **目** | **項目** | **標準** |
| 1 | 影片畫質 | 畫質過低 (480p 或以下) 的影片會影響 batch face swap 插件的人臉偵測，以**720P**為最低標準 |
| 2 | 影片長度 | 30s ~1min |
| 3 | 著作權問題 | 暫不考慮 |
| 4 | 實驗影片網址 | * 鐵達尼 : <https://www.youtube.com/watch?v=F2RnxZnubCM> * 周子瑜 : <https://www.youtube.com/watch?v=MTjud9r6g8Y> * 韓國 Youtuber imlisarhee 舞蹈(1:35~1:50) : <https://www.youtube.com/watch?v=wMypHzyVvQs&ab_channel=imlisarhee> |



1. 可利用 ffmpeg ，以30 fps自動切割影片為影格圖片 ( 如圖 )



1. 利用 **SD + Lora** 來訓練使用者的人臉模組，標準參考以下 :

|  |  |  |  |
| --- | --- | --- | --- |
| **目** | **項目** | **標準** | **說明** |
| 1 | 輸入照片數量 | 5~15 | 以5張為要求，最多不超過15張 |
| 2 | 照片大小 | 512\*512 分辨率不低於 96dpi | 若使用者的圖片小於 512\*512或分辨率低於96dpi，則自動提高圖片畫質、清晰度等 |
| 3 | 照片標準 | 不同背景、衣服、場景的臉部清晰、高解析度自拍照 | 需使用者上傳時，需有明顯字眼提示使用者 |
| 4 | 照片格式 | JPEG, PNG 檔 |  |
| 5 | 訓練用 SD model | V1.5 | 用官方標準1.5 或以上之版本 |
| 6 | 產圖用 SD model | V1.5 |  |
| 7 | 建議訓練步數 | 100(待定) | 在 3080上應該是10-15 分鐘，目標3分鐘完成 |
| 8 | 建議產圖步數 | 10(待定) | 目標3-5秒內完成 |
| 9 | 人像相似度 | 需接近本人而非與自己 “神韻相似”的人像 |  |
|  | | | | |

1. 最終, 使用經訓練出來個人的模型逐一置換影格，再利用 ffmpeg 串成影片成果如下

|  |  |
| --- | --- |
|  | **實驗成果 1-鐵達尼** |
|  |  |
| **成果檔 2** | **實驗成果 2-周子瑜** |
|  |  |
| **成果檔 3** | **實驗成果 3-韓國 Youtuber imlisarhee 舞蹈** |
|  |  |

* 功能要求 :

|  |  |  |
| --- | --- | --- |
| **目** | **功能** | **描述** |
| 1 | 模型訓練 | 參考 “訓練環境 “ 部分 |
| 2 | 一鍵換臉到影片 | 1. 一鍵換臉成為影片的歩驟   * 前台 :   1. 使用者上傳5-15張大頭照   2. 點按訓練等待 3 分鐘即可得到客製化人像模組   3. 到 template 中可以選擇內建影片, 或自行上傳畫質不低於 720p 的影片   4. 選擇訓練好的模組   5. 點按 “switch”等待2-5 分鐘即可得到換臉影片   2. 後台要提供API給前台 :  (1) 訓練照片的所有參數及參數定義  (2) 影片-影格切割數據  (3) 人臉融合參數 (module, prompt, step, seed, CFG etc.)  (4) 影格-影片串聯數據 |
| 3 | Text to video 影格間的不連貫 | 需解決利用 Lora 所訓練之模型，透過擴散方式生成的人像不連貫、不可控等問題 |
| 4 | 訓練耗時及處理影片時間 | 1. 目前在(RTX 3080)訓練1500步耗時約15~20分鐘，  這個時長以產品角度看，遠遠超過一般使用者的等待體驗時間，若產品化，人像尚未訓練完成使用者就已離開網頁，做到目標3分鐘完成  2. 影片產生時間需在 2-5 分內 |
| 5 | 生成之臉部表情不自然 | 需解決由AI生成的人臉看出明顯的生硬、不自然感 |
| 6 | 支持多臉融合 | 支援單臉、多臉、選臉融合，最多支援指定融合6張人臉  可應用在全家福、與明星合照等多人互動場景。適用於互動娛樂、相機、美妝、換臉、換髮 |
| 7 | 支持風格替換 | 融合效果自然，相容真人、油畫、水彩等風格。 |
| 8 | 支持人像精修 | 參數精細調整修改融合相似度，調整五官和臉型參數，可精細調整用戶人臉和特定形象的融合比例。  須提供支援圖片、影片修圖相關功能的 API  (可參考 IPhone 內建圖片、影片編輯功能 ) :   * + - * + 簡易剪輯(限定影片)         + 曝光參數         + 亮度參數         + 陰影參數         + 對比參數         + 飽和度參數         + 色溫參數         + 色調參數         + 清晰度參數         + 畫質參數         + 雜點消除參數 |
| 9 | 有五套劇本供換臉使用 | 有內建受歡迎的換臉影片腳本，供使用者選擇 |

請有意願了解執行的老師，聯繫本會

Line@: @749bdvym

sec@tsta-tw.org.tw