

## 深度學習期末報告 – 動漫風格轉換

王俊貴 108321018

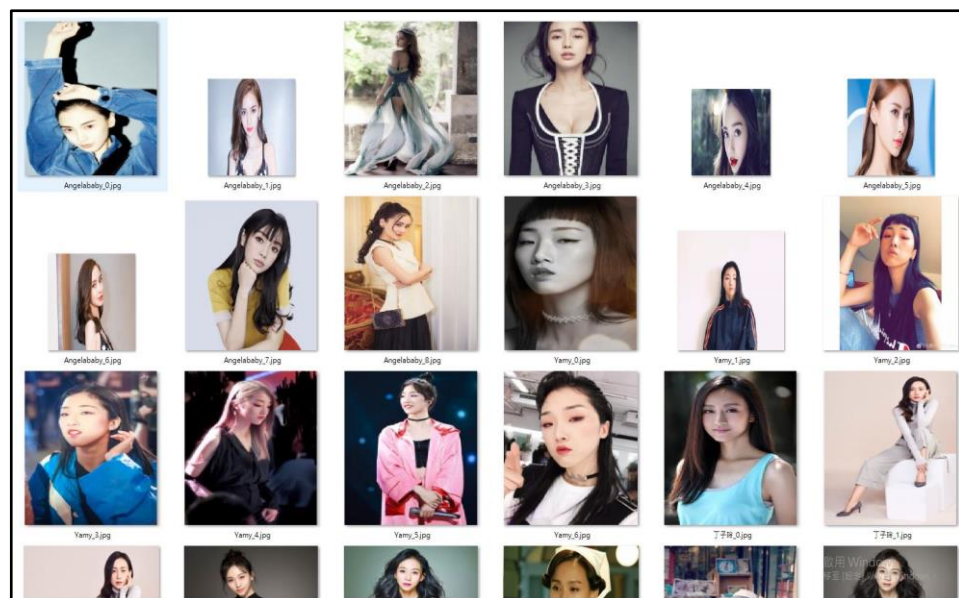
王奇立 108321006

楊曜瑋 107321011

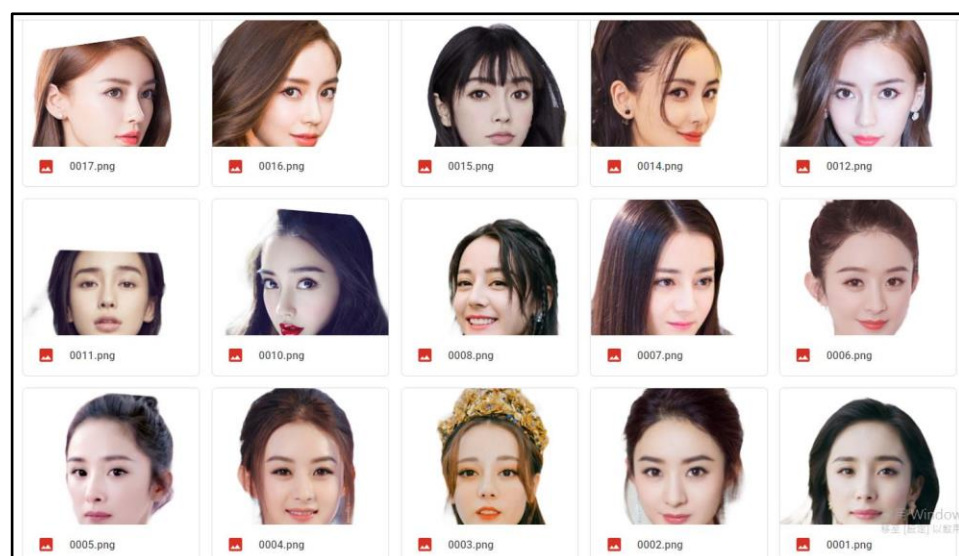
### 一、 使用的資料集

#### a. 自建亞洲人臉資料集

使用網路自建的亞洲人臉資料集，將其中女性的人臉挑選出來，之後經過程式進行前處理，將人臉的部分切割並且擺正之後進行訓練。



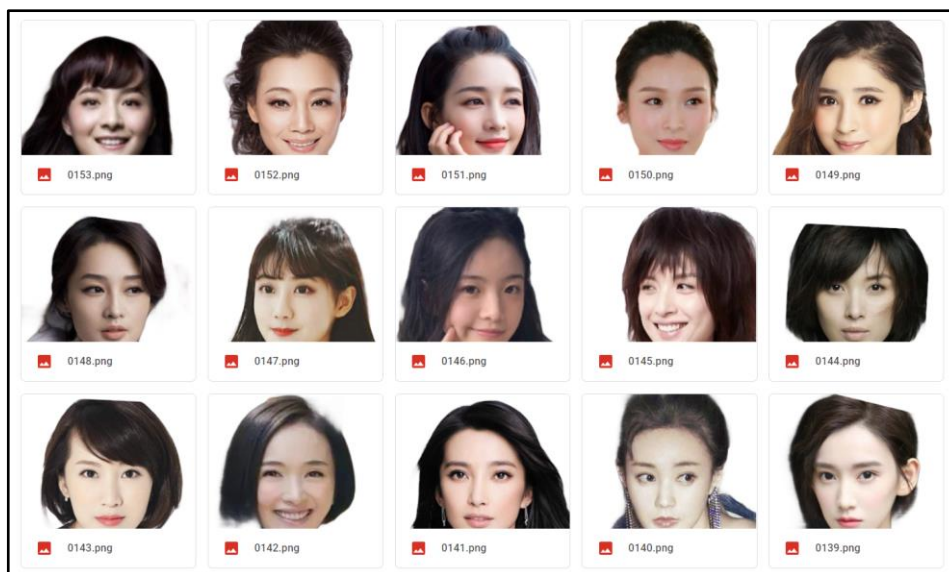
自建亞洲人臉資料集



預處理過後的資料集

## b. 人臉產生器所產出的照片

由網路上的人臉產生器提取各項特徵進行排列組合的結果，但也因此會感覺整體照片會長得很像，好處是產生的照片皆是正臉，就算不進行預處理也能直接使用。



人臉產生器資料集

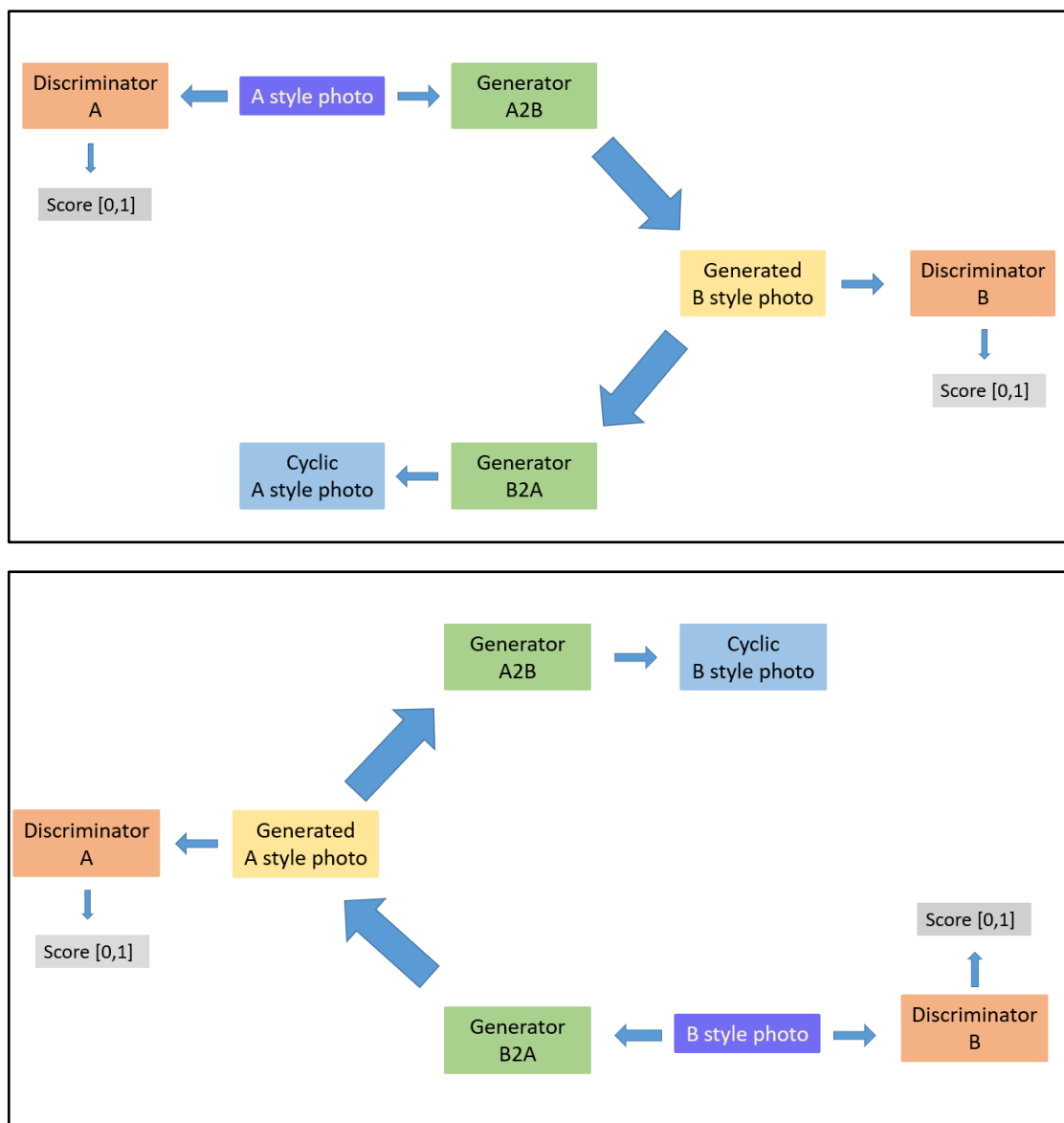
## ➤ 小結

因為上述兩種資料集在訓練後進行比較，發現結果差異不大，基於自建亞洲人臉資料集的人臉擁有較豐富的素材，因此我們後續皆採用此資料集進行訓練。

## 二、 Model 架構

### a. CycleGAN

由兩組 GAN 所構成，總共有兩個 generator (一個負責產生假的 A 資料、一個負責產生假的 B 資料)，兩組 GAN 各帶有一組 discriminator 負責分辨資料的真偽。目的是透過 generator 產生假資料，並由 discriminator 判斷真假，相互進行訓練，以達到較佳的訓練效果。



CycleGAN 流程示意圖

### b. Model 各層架構

| Layer type      | #   | Layer type  | #  |
|-----------------|-----|-------------|----|
| ReflectionPad2d | 190 | ResnetBlock | 4  |
| Conv2d          | 205 | Linear      | 52 |
| InstanceNorm2d  | 183 | adaLIN      | 8  |

|                |     |                       |            |
|----------------|-----|-----------------------|------------|
| ReLU           | 204 | SoftAdaLIN            | 8          |
| ConvBlock      | 56  | ResnetSoftAdaLINBlock | 4          |
| HourGlass      | 4   | Tanh                  | 1          |
| HourGlassBlock | 4   | Upsample              | 2          |
| LIN            | 2   | 總計                    | 927 layers |

c. 參數量

|  |
|--|
| Total params 3,770,375                 |
| Trainable params 3,770,375             |
| Non-trainable params 0                 |
| -----                                  |
| Input size (MB) 0.75                   |
| Forwardbackward pass size (MB) 2389.99 |
| Params size (MB) 14.38                 |
| Estimated Total Size (MB) 2405.12      |
| -----                                  |

d. Loss function

I. MSE\_Loss : Mean Square Error

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (y - \hat{y}_i)^2$$

II. L1\_Loss : 輸出和模板之間相對應的元素相減後的總和

$$L1LossFunction = \sum_{i=1}^n |y_{true} - y_{predicted}|$$

III. BCEWithLogitsLoss : Binary Cross Entropy + sigmoid

$$CE = - \sum_{i=1}^{C'=2} t_i \log(f(s_i)) = -t_1 \log(f(s_1)) - (1 - t_1) \log(1 - f(s_1))$$

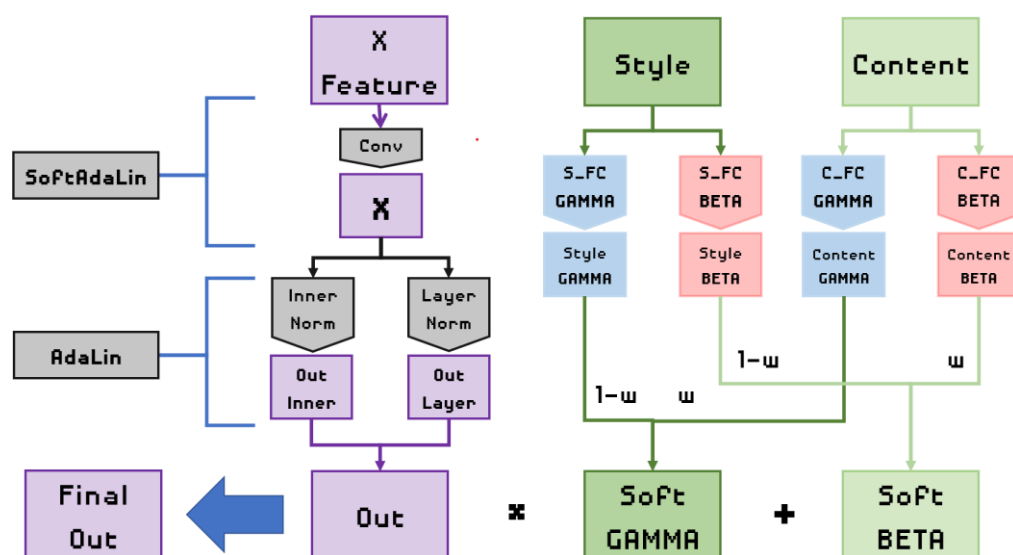
e. AdaLIN

將 Attention Feature Maps 放到 fully connected layer 學習兩個參數：scale 的  $\gamma$  以及 shift 的  $\beta$  映射到位於目標風格的 space，讓神經網路自己選擇應該要用 Instance Norm 或是 Layer Norm。

$$AdaLIN(a, \gamma, \beta) = \gamma \cdot (\rho \cdot \hat{a}_I + (1 - \rho) \cdot \hat{a}_L) + \beta,$$

$$\hat{a}_I = \frac{a - \mu_I}{\sqrt{\sigma_I^2 + \epsilon}}, \hat{a}_L = \frac{a - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}},$$

$$\rho \leftarrow clip_{[0,1]}(\rho - \tau \Delta \rho)$$



AdaLIN 示意圖

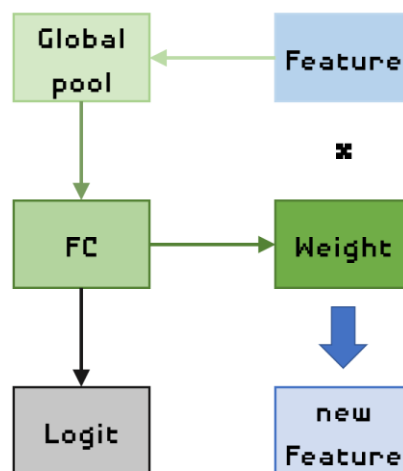
#### f. Optimizer

使用 Adam Optimizer

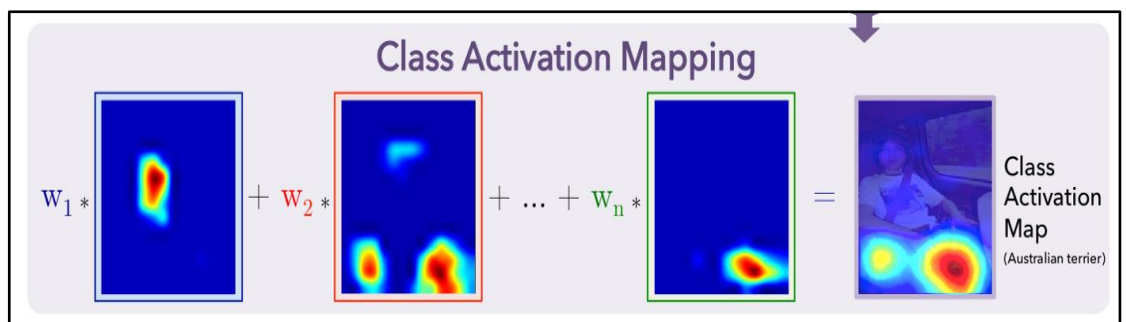
|                         | Learning rate | Betas         | Weight decay |
|-------------------------|---------------|---------------|--------------|
| Generator optimizer     | 0.0001        | (0.5 , 0.999) | 0.0001       |
| Discriminator optimizer | 0.0001        | (0.5 , 0.999) | 0.0001       |

#### g. CAM

在結束 convolution 後，套入一層 GAP 層(Global Average Pooling Layer)，每一張特徵圖經過 GAP 的轉換後將特徵圖的訊息壓縮成一個一個的神經元，因此經過 GAP 轉換後的每一個神經元分別對應到了最後一層的某一張特徵圖，而 GAP 層所連接的權重即可視為每一張特徵圖對於模型預測類別的重要性，最後將每一張特徵圖依照其對應的權重進行加權即得到 CAM (Class Activation Map)。



CAM 流程圖



CAM 權重疊加方式

### 三、 CAM 機制的影響

因為我們推估，加入 CAM 機制，可以有效提升特徵提取的能力，因此將 CAM 的權重分別設為 1000 及 0 進行比較。然而卻會發現，加入 CAM，反而使得整體效果變差了，與我們的推測有所不同。



CAM 權重 = 0, epochs = 8000



CAM 權重 = 1000, epochs = 8000











CAM 權重 = 0, epochs = 13490









CAM 權重 = 1000, epochs = 13950







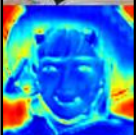
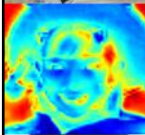
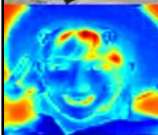
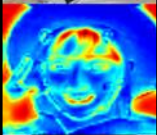
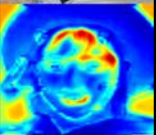
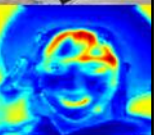






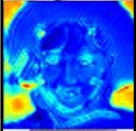

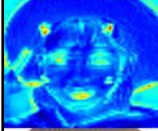

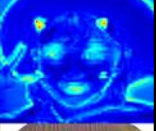







#### 四、 成果展示&發現

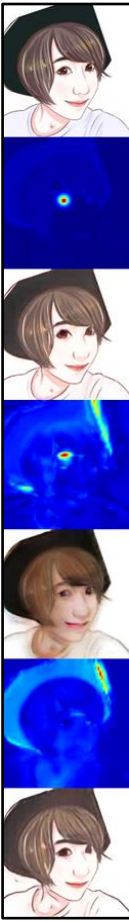
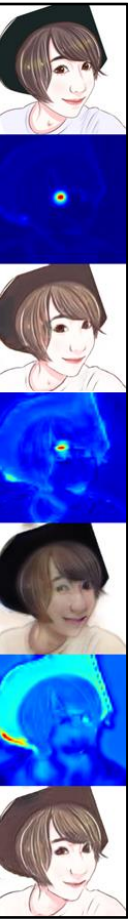
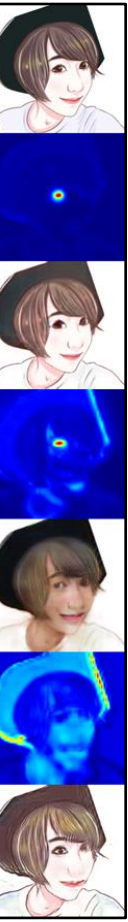
##### a. 測試結果

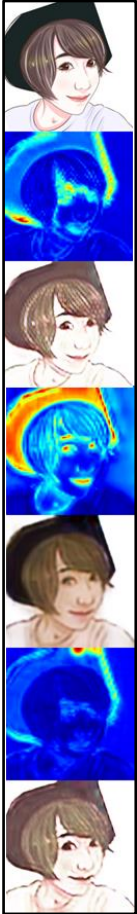
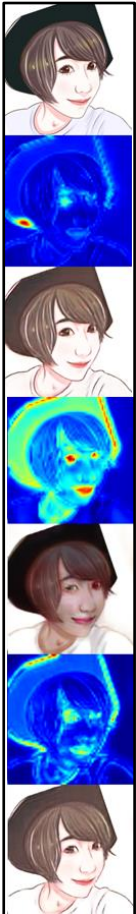
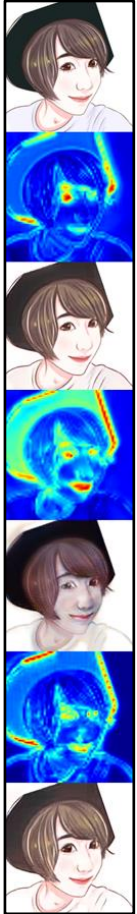
|                   |   |  |   |
|-------------------|---|--|---|
| 測試<br>結果          |   |   |   |
| epoch<br>訓練<br>次數 | 3000  | 7000   | 11950   |
| 測試<br>結果          |  |  |  |
| epoch<br>訓練<br>次數 | 15950   | 23000  | 25000   |
| 測試結果一             |   |  |   |

|                   |   |  |   |
|-------------------|---|--|---|
| 測試<br>結果          |  |  |  |
| epoch<br>訓練<br>次數 | 3000  | 7000   | 11950   |
| 測試<br>結果          |  |  |  |
| epoch<br>訓練<br>次數 | 15950   | 23000  | 25000   |
| 測試結果二             |   |  |   |



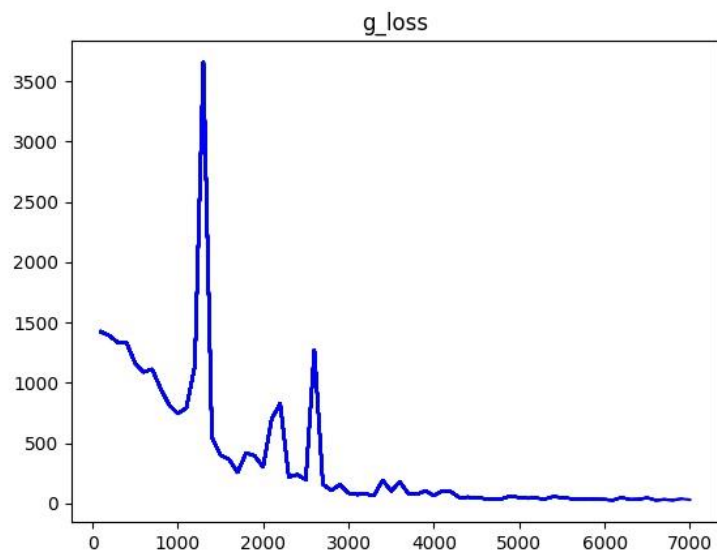
|            |   |   |   |   |   |   |
|------------|---|---|---|---|---|---|
| 結果         |  |  |  |  |  |  |
|            |  |  |  |  |  |  |
|            |  |  |  |  |  |  |
|            |  |  |  |  |  |  |
|            |  |  |  |  |  |  |
| epoch 訓練次數 | 1000  | 3000  | 6000  | 9000  | 11000   | 12000   |
| A2B 變化過程   |   |   |   |   |   |   |

|                       |  |  |  |
|-----------------------|--|--|--|
| 結果                    |  |  |  |
| epoch 訓練次數            | 19000  | 21000  | 25000  |
| B2A 變化過程(CAM 權重=1000) |  |  |  |

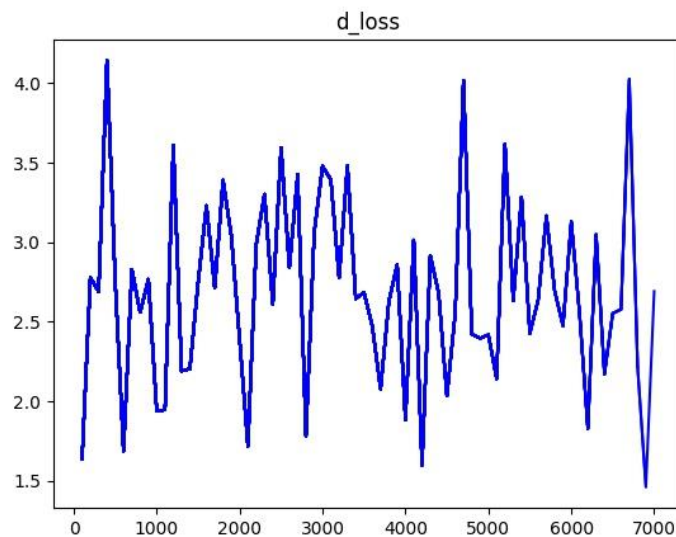
|                    |  |  |  |
|--------------------|--|--|--|
| 結果                 |  |  |  |
| epoch 訓練次數         | 1000   | 7000   | 13000  |
| B2A 變化過程(CAM 權重=0) |  |  |  |

#### b. 數據分析

Discriminator Loss 經常不穩定的跳來跳去，Generator Loss 則是有趨於 0 的趨勢，實際看過也確實 Generator 產生的效果越來越好。



Generator Loss






Discriminator Loss

### c. 問題與發現

- I. 原先預期將 CAM 的權重調高，會有較好的特徵提取效果，然而成果卻反而不如預期。此外，在 B2A 的成果上，也能發現使用 CAM 會使提取特徵集中單一區域，例如眼睛、額頭等...，沒有使用 CAM 反而提取範圍較全面，集中於整個面部區域
- II. 當 epoch 數提高，特徵提取會從原先發散至整張圖片背景逐漸集中於人的臉部，能有效提升成效。
- III. 根據參考的資料，原先預設是要跑 100 萬次 epoch，然而我們跑了 25000 次 epoch 後，其實就有不錯的效果。

d. 組員大頭照結果 ( 25000 epochs , CAM = 1000)

部分特徵(眉毛、眼睛)還是有些瑕疵，但是整體輪廓已經很清楚，並且眼鏡也有很好的畫出來。

|     |  |
|-----|--|
| 王俊貴 |  An AI-generated portrait of a man with short, dark hair, looking directly at the camera. The style is somewhat painterly with visible brushstrokes. The face is well-defined, but there are some artifacts around the eyes and mouth. |
| 王奇立 |  An AI-generated portrait of a man with short, dark hair, looking directly at the camera. The style is painterly. The face is well-defined, but there are some artifacts around the eyes and mouth.                                   |
| 楊曜璋 |  An AI-generated portrait of a man with short, dark hair, wearing glasses. The style is painterly. The face is well-defined, but there are some artifacts around the eyes and mouth.   |

## 五、 結論

- Epoch 數越高，能有效強化特徵提取效果，在風格轉換上能做到更加地精細，使成果越趨完美。
- CAM 好像在風格轉換上的幫助有限，甚至有反效果。
- 有機會可嘗試加入不同性別、種族，抑或是不同風格的訓練資料，看看會不會有更好的效果。
- 參考資料上沒有一個明確判斷風格轉換是否完全的指標，但在網路上查詢到可以判斷 loss 是否有明確改變，如果沒有的話，epoch 就提早停止，未來有機會可以嘗試看看加入這個判斷機制去找到適當的訓練次數。