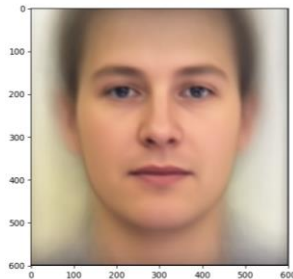


Part A: PCA of colored faces

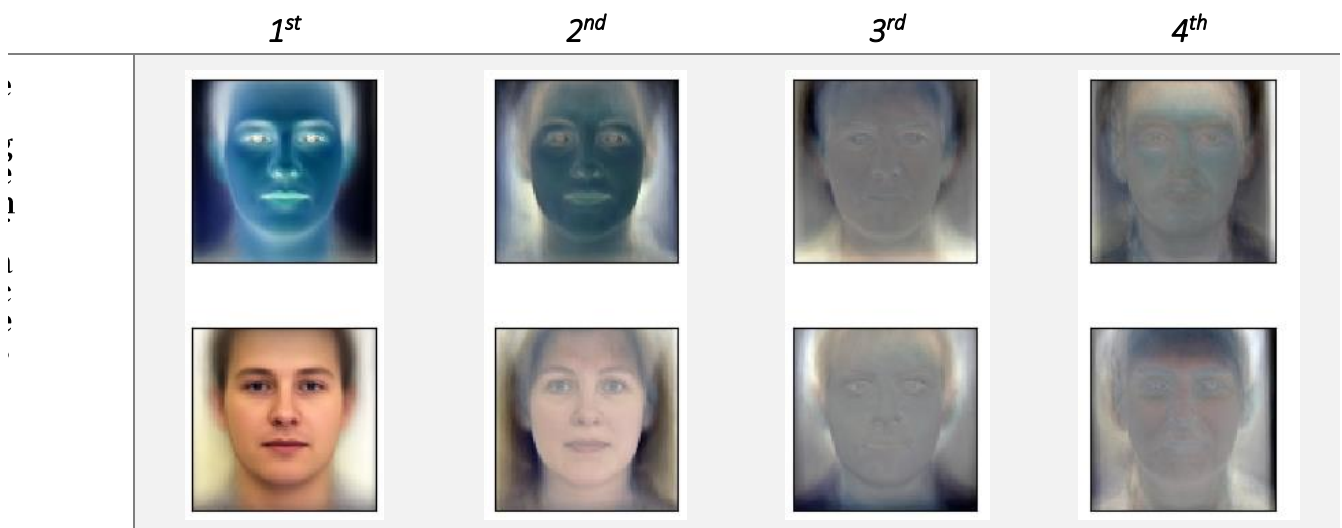
Collaborators : 黃禹程 R06944034、鄭克宣 R06921083、蔡孟庭 R05922078

- A.1 (.5%) 請畫出所有臉的平均。



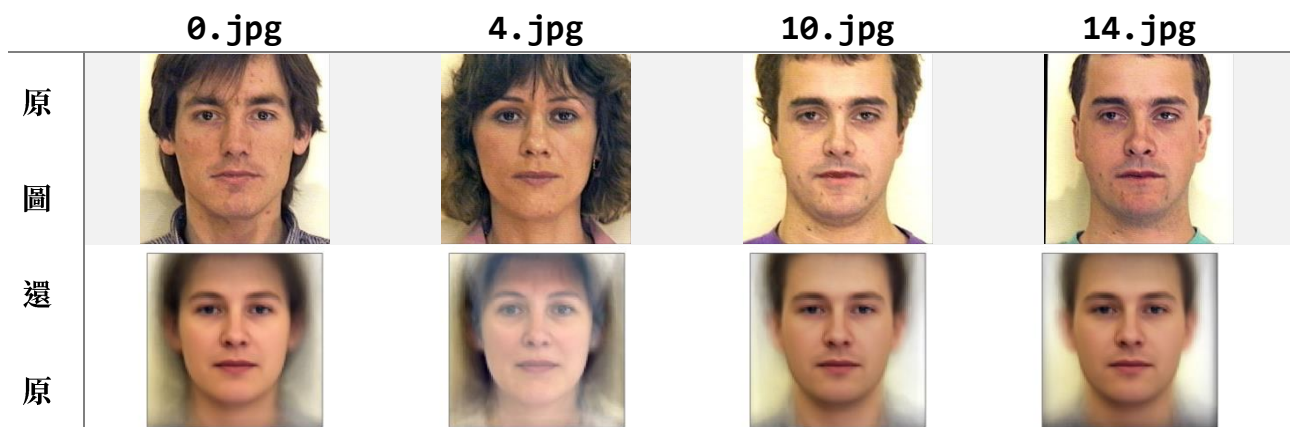
- 將415張600*600*3之影像加總並平均之

- A.2 (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。



- 由左到右依序為最大、次大、第三、第四大之 eigenface
- 礙於計算資源有限，影像已先resize成100*100*3
- 為方便助教批改，故列舉正反之兩種可能性

- A.3 (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。



Part C: Image clustering

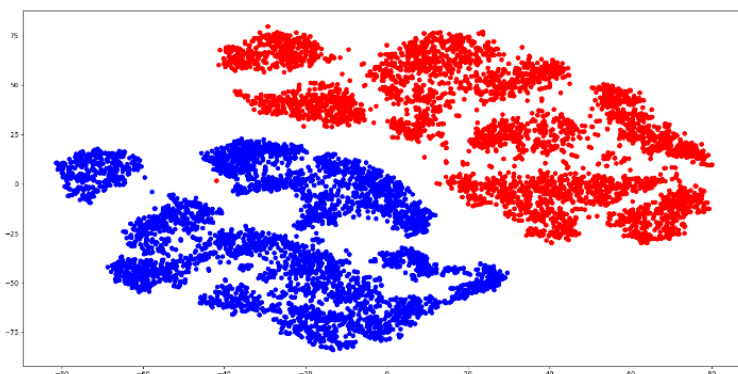
Collaborators : 黃禹程 R06944034、丁繼楷 R06922129、葉孟元 R04921094

- C.1 (.5%) 請比較至少兩種不同的 *feature extraction* 及其結果。

方法	PRIVATE 分數	PUBLIC 分數	平均
AUTOENCODERS+KMEANS	1	1	1
PCA+KMEANS	0.03040	0.03020	0.03030

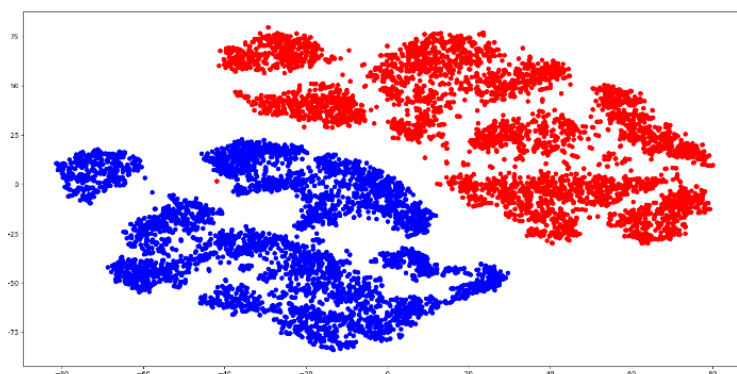
- 採 TSNE 降維，並用 matplotlib 之 scatter 繪圖
- 紅色來自 dataset A，藍色來自 dataset B，於圖上呈兩大區塊分布
- Autoencoder 參考助教之設計，PCA 之參數則是 `n_components=3`
- Autoencoder 約在 `train epoch=500` 約可達到 `0.93`，而附上之版本則為 `epoch=14900` 次
- 與同學討論時，發現 PCA 亦有潛力提高準確率，礙於時間有限並無成功實作

- C.2 (.5%) 預測 `visualization.npy` 中的 `label`，在二維平面上視覺化 `label` 的分佈。



- 採 TSNE 降維，並用 matplotlib 之 scatter 繪圖
- 紅色來自 dataset A，藍色來自 dataset B，於圖上呈兩大區塊

- C.3 (.5%) `visualization.npy` 中前 5000 個 `images` 來自 dataset A，後 5000 個 `images` 來自 dataset B。請根據這個資訊，在二維平面上視覺化 `label` 的分佈，接著比較和自己預測的 `label` 之間有何不同。



- 紅色來自 dataset A，藍色來自 dataset B，於圖上亦呈兩大區塊
- 與自行預測的 `label`（上上圖）相似度極高，應該是 `model` 於 `kaggle` 上高準確率之表現有關