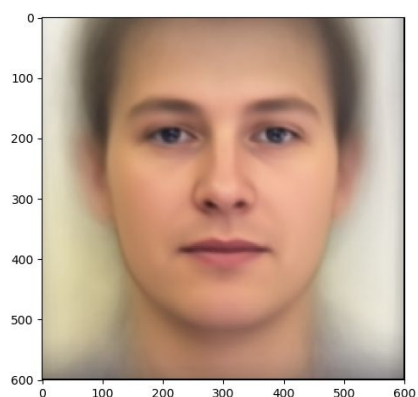


## HW4

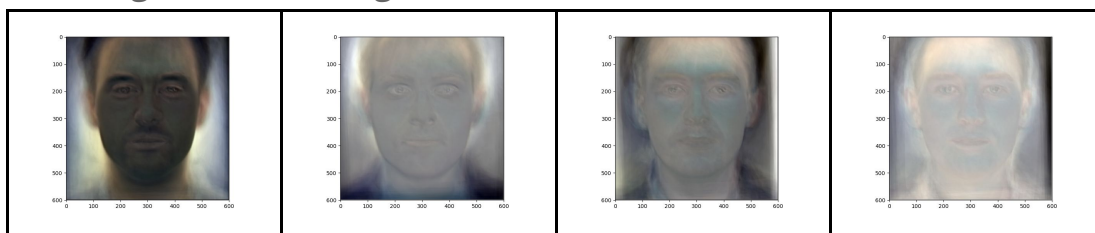
學號：b04705043 系級：資管三 姓名：張凱庭

### A. PCA of colored faces

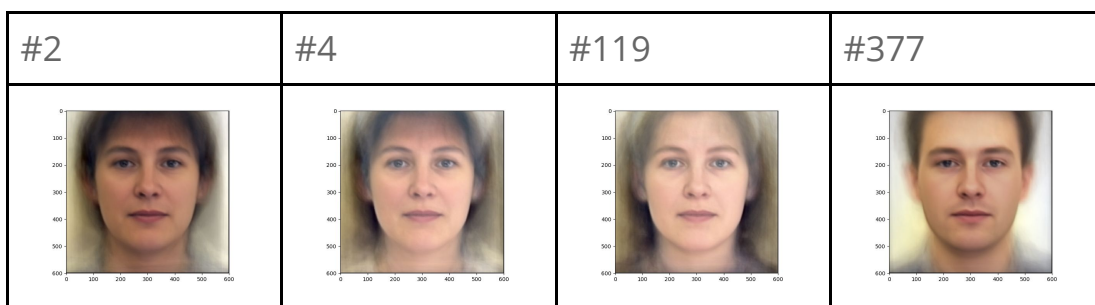
A.1. (.5%) 請畫出所有臉的平均。



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



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。



A.4. (.5%) 請寫出前四大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入到小數點後一位。

4.1%	2.9%	2.4%	2.2%
------	------	------	------

## B. Image clustering

B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

Autoencoder : 架構使用 128-64-64-64-128 , 將資料降到64維

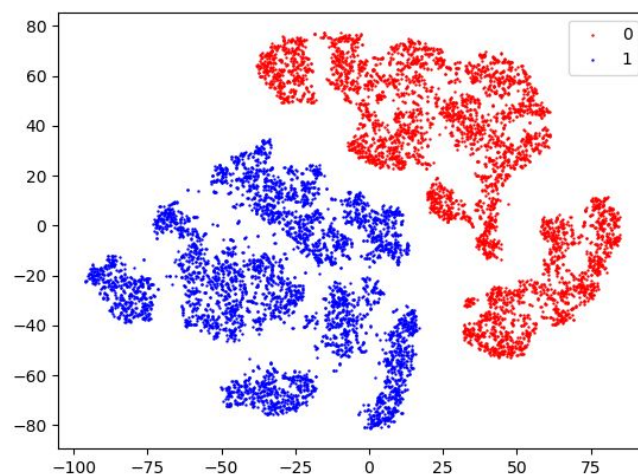
cluster 使用KMeans(n\_clusters=2)

PCA:將資料降到64維 , cluster 使用KMeans(n\_clusters=7)

比較:Autoencoder 可能考慮更複雜的關係 , 即使抽出的特徵較少(64個)分類表現仍遠大於PCA。PCA在圖片上的表現似乎是不太好使用KMeans(n\_clusters=7) , 正確率仍然無法拉高。

	auto encoder	pca
public score	1.00000	0.75065
private score	1.00000	0.75123

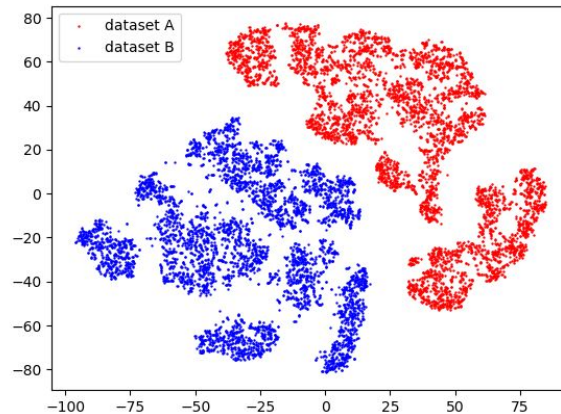
B.2. (.5%) 預測 visualization.npy 中的 label , 在二維平面上視覺化 label 的分佈。



預測結果兩種資料分得很開 , 並沒有混雜在一起

B.3. (.5%) visualization.npy 中前 5000 個 images 跟後 5000 個 images 來自不同 dataset。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。



可以看到和預測結果完全相同，也符合model在kaggle上所得到的分數 (1.00000)，auto encoder 在特徵抽取上有卓越的表現

## C. Ensemble learning

C.1. (1.5%) 請在hw1/hw2/hw3的task上擇一實作ensemble learning，請比較其與未使用ensemble method的模型在 public/private score 的表現並詳細說明你實作的方法。（所有跟ensemble learning有關的方法都可以，不需要像hw3的要求硬塞到同一個model中）

我在hw3實作ensemble learning，實作的方法很簡單，將多個模型所預測出的結果進行平均即生程新的結果。我一共挑選了8組來自四個架構的model來進行ensemble，1~2號model採用一樣架構但是不同的訓練epoch，3~4號model採用一樣架構但是不同的訓練epoch，以此類推5~6，7~8，其中挑選epoch則依據valid\_loss來決定，基本上四個架構model都是參考vgg16，其實彼此差異不會太大，但經過voting後，分數從0.66~0.68上升到0.71左右，直觀上來從機率來看，經過投票會降低錯的發生，強化預測的穩固度。

	epoch	valid_loss
<b>1</b>	146	0.67200
<b>2</b>	253	0.67266
<b>3</b>	166	0.66733
<b>4</b>	168	0.67633

<b>5</b>	248	0.68267
<b>6</b>	252	0.68433
<b>7</b>	157	0.67367
<b>8</b>	158	0.67400

	ensemble	not ensemble(#5)
public score	0.71997	0.68960
private score	0.71858	0.68403
valid_loss	0.71	0.68267