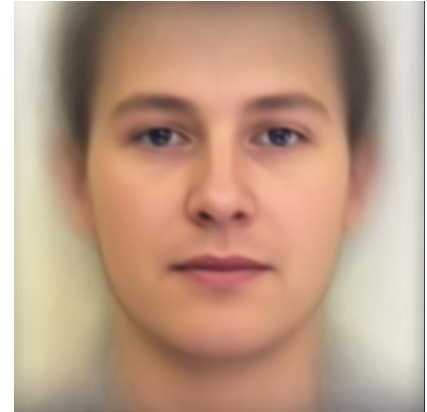


## HW4

學號：B04901067 系級：電機三 姓名：陳博彥

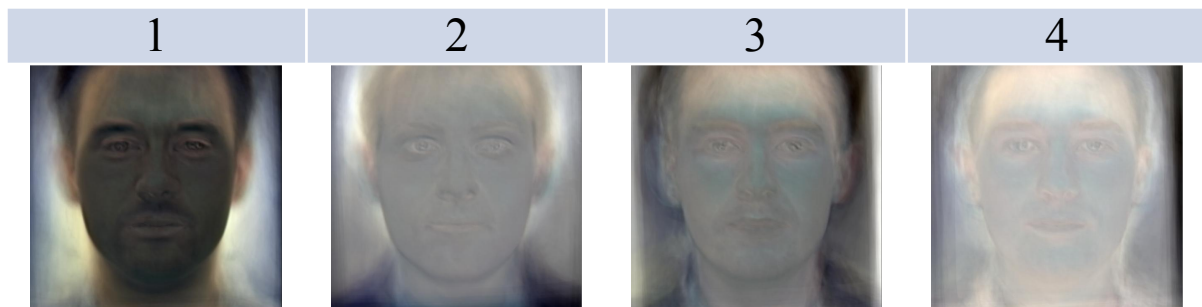
### A. PCA of colored faces

A.1.(.5%) 請畫出所有臉的平均。



平均臉（其實還滿帥的...）

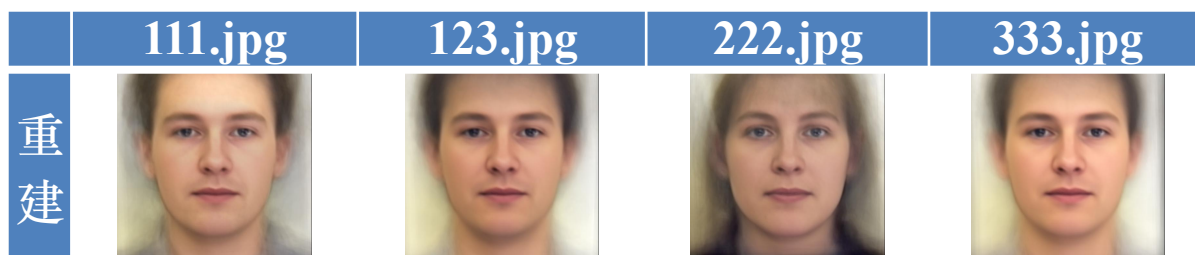
A.2.(.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。



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。

前三張可看見頭髮部分保留了一些特徵，第四張由於頭髮位置跟 mean\_face 差不多，看起來幾乎跟 mean\_face 一樣。





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。

前四大比重（依次由大到小）：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)

方法一：

(1)用autoencoder把影像降到32維

(2)KMeans(n\_clusters=2)

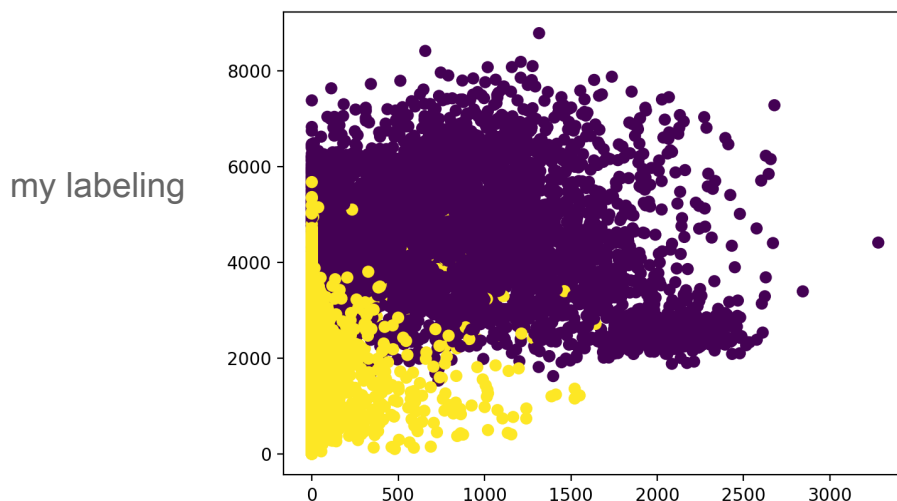
方法二：

(1)PCA(n\_components=400,random\_state=0,whiten=True)

(2)KMeans(n\_clusters=2)

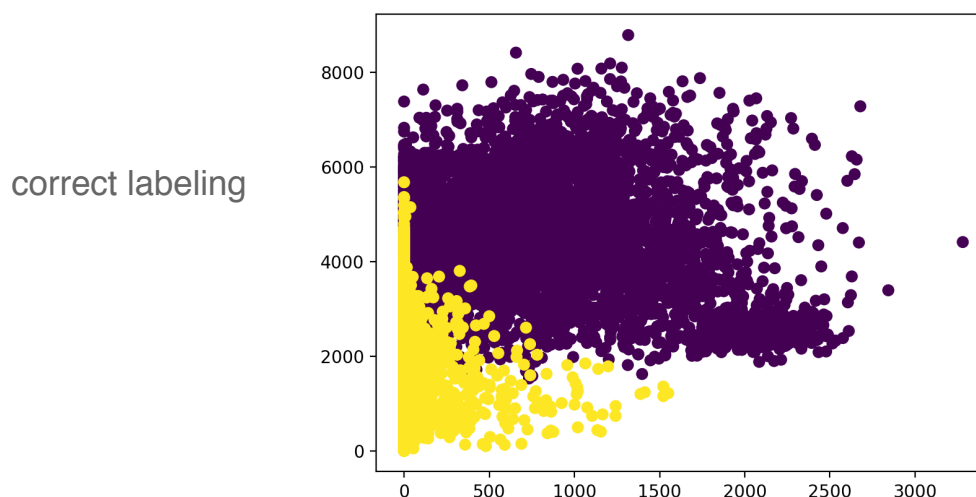
	public score	private score
方法一	0.98596	0.98577
方法二	1.00000	1.00000

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



上圖是我用B.1中方法一來cluster的結果，由於我是降到32微做Kmeans，卻只用前2維來visualize，看不出什麼所以然，資料點都連在一起，但至少兩群大致上有個清楚的分界線。

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。



上圖是根據本題資訊畫出來的，和my labeling比較可看見，兩者分群大致相同，只是my labeling把一些應該是紫色的點label成黃色的（可看見紫色區塊散落著一些黃色的點），這些點應該是在別的維度上跟黃色的點很接近，所以才會被Kmeans誤判。

## 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。

我選擇hw3進行測試，model 1~3 是我以前的model，其差別主要是在activation的方法，還有epoch數。

針對這次的作業，ensemble的方法如下：

(1)用model 1~3分別對test.csv做出predict

(2)取model 1~3的預測中，「信心度的最大值」所對應的類別作為ensemble輸

出的答案。舉例來說，在某一項預測中：

model 1對於七個類別的prediction是[ 0.1 0.4 0.1 0.1 0.1 0.1 0.1 ]

model 2對於七個類別的prediction是[ 0.5 0.0 0.1 0.1 0.1 0.1 0.1 ]

model 3對於七個類別的prediction是[ 0.1 0.3 0.2 0.1 0.1 0.1 0.1 ]

這樣我就選擇第一個類別作為ensemble的輸出答案，因為0.5在這些信心度中是最大的。注意，我不是用投票的方式，也不是把信心度加總，而單純只看最大的信心度出現在哪個類別。「當意見分歧時，就相信最肯定的那個人吧」這是這套做法背後的理由。

	model 1	model 2	model 3	ensemble
public score	0.70632	0.70604	0.69908	0.71551
private score	0.70270	0.70743	0.70047	0.71551

成效確實提升。