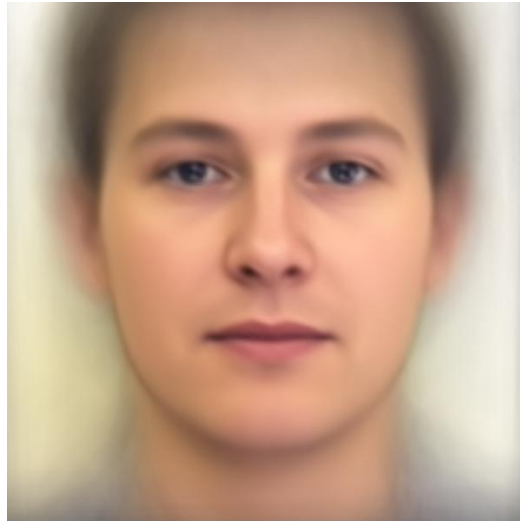


HW4

學號：B04901020 系級：電機三 姓名：解正平

A. PCA of colored faces

A.1. (.5%) 請畫出所有臉的平均。

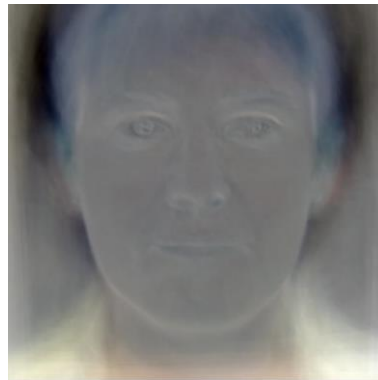


A.2. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。

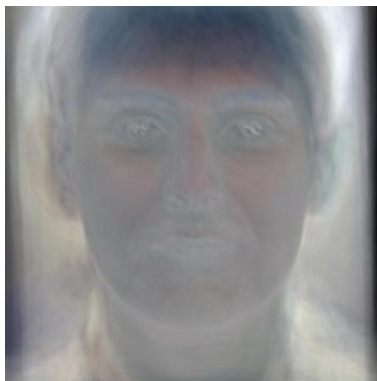
由 Eigenvalues 大到小排列順序的 Eigenfaces 編號 1 到 4



1



2






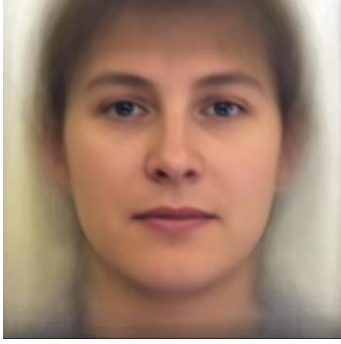

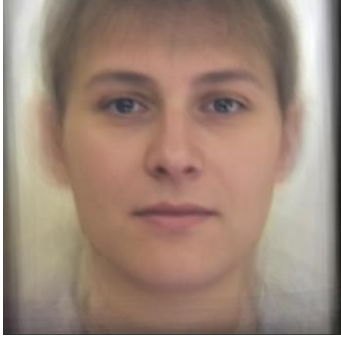


3



4

- 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。

	Original Picture	Reconstruction Picture
100.jpg		
200.jpg		
300.jpg		
400.jpg		

- 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。

左至右大到小	1	2	3	4
比例	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)

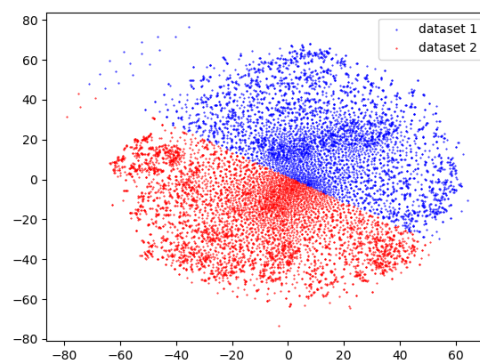
PCA 使用 `n_components=400,whiten=True` 。

Autoencoder 使用 `layer : activation = tanh` ,

`Compile : optimizer=Adam(lr=5e-4)`

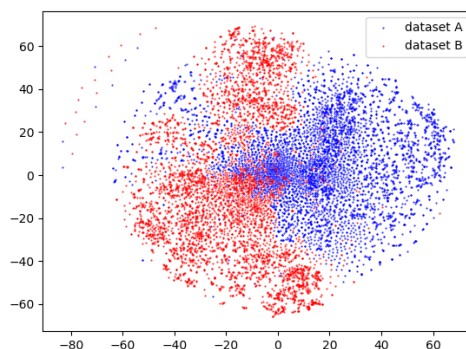
	Public score	Private score
PCA	1.00000	1.00000
Autoencoder	1.00000	1.00000

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



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。

從上圖與下圖比較, 可以發現我們降維之後未能將 **data** 很明顯的分開, 還是仍有部分藍點與紅點混雜在一起。



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 model 實作 ensemble，把 train 出來的五種 model 利用 stacking 的方法產生不同結果，再將每個結果加起來平均，實作 voting 的效果，使得 classification 有很好的表現。

	Public score	Private score
Model1	0.69295	0.68347
Model2	0.67428	0.68041
Model3	0.69350	0.69768
Model4	0.68988	0.69601
Model5	0.69267	0.69155
Ensemble	0.71969	0.72137