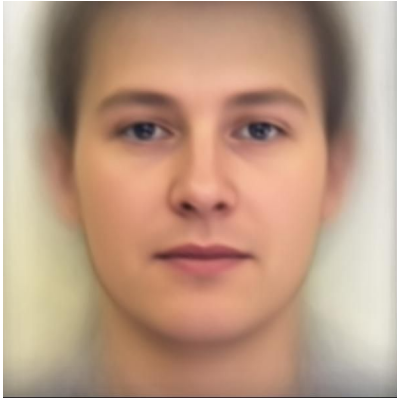


HW4

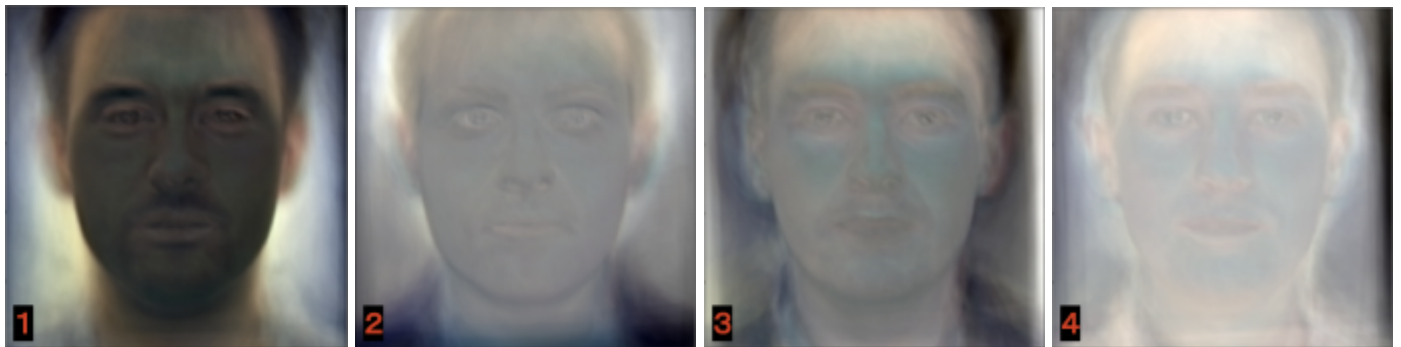
學號：r06921081 系級：電機碩一 姓名：張邵瑀

A. PCA of colored faces

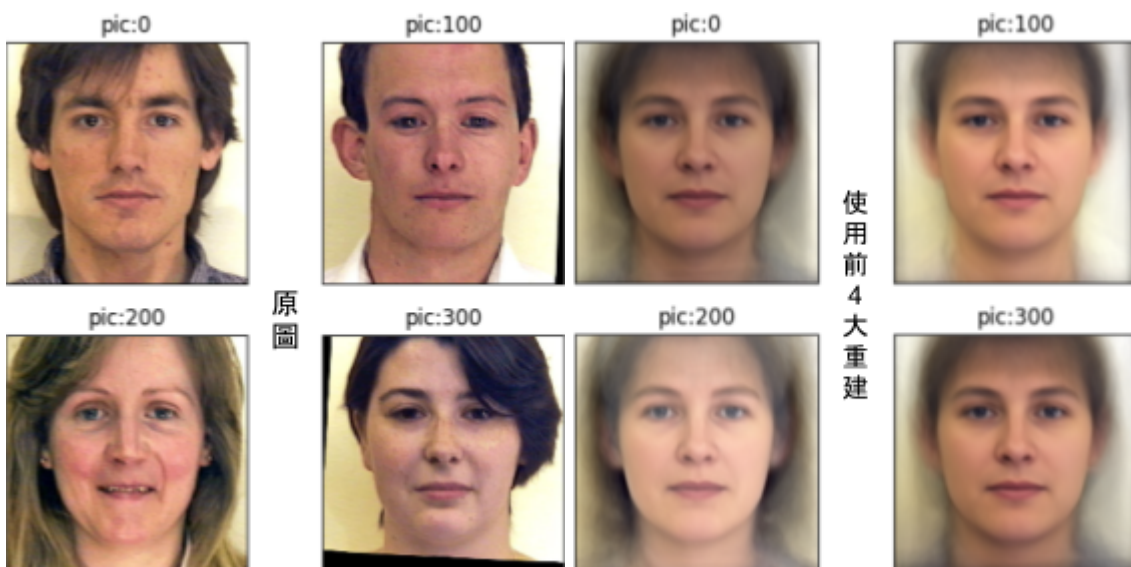
A.1. (.5%) 請畫出所有臉的平均。



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。

第1大 eigenvalue	第2大 eigenvalue	第3大 eigenvalue	第4大 eigenvalue
4.1%	2.9%	2.4%	2.2%

B. Image clustering

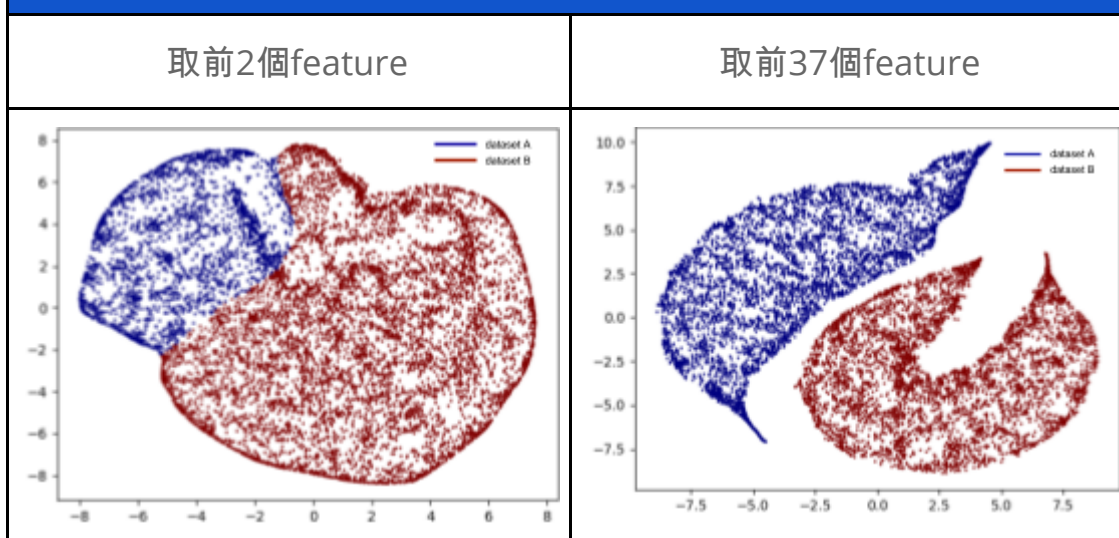
- B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。

	PCA	CNN autoencoder
降維至/Score	410/1.00000	600/0.99836

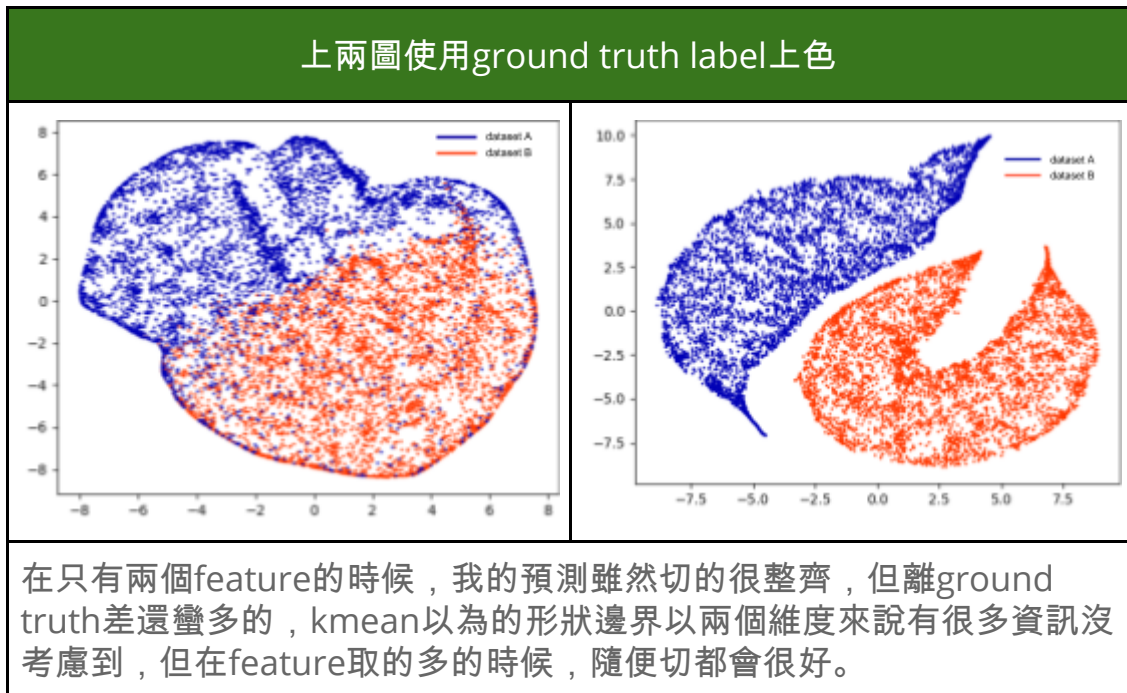
兩者都在使用上述方法降維後，再用K-means分成兩堆，想是我的CNN model疊的不夠好，不然應該可以降更多維，並且如果CNN做完後再用PCA降維一次就能達到1，我在想應該是whiten這個參數影響重大。

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

PCA降維，再用tsne投到2維平面上，使用kmeans label預測結果上色，tsne參數：random_state=30, n_iter=280, perplexity=100



- 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。



C. Ensemble learning

- C.1. (1.5%) 請在hw3的task上實作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。

分數：private/public。

方法：data切10份輪流當 validation set用同一model訓練，用其預測結果做weighted voting，將所有model預測出來的softmax值相加取最大。

1	2	3	4	5
0.69295 0.70409	0.69072 0.69434	0.68821 0.68904	0.69127 0.69768	0.68682 0.69490
6	7	8	9	10
0.69044 0.69740	0.69963 0.70242	0.69183 0.70242	0.68821 0.69434	0.69044 0.69768
ensemble	0.70298 0.71022	大約提升1%左右		

效果不是太好，所以我把之前最高分的model(0.70799/0.71022)加進去做weighted，並把分數較低的幾個剔除掉取(1,2,4,7,8,10)weight為1 : 1 : 1 : 1 : 1 : 5.8 結果為(0.71440/0.71412)。