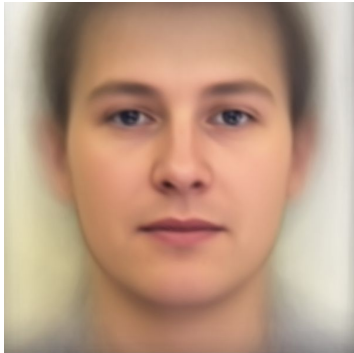


## Homework 4 Report - Unsupervised learning & dimension reduction

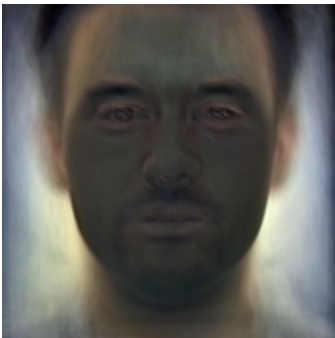
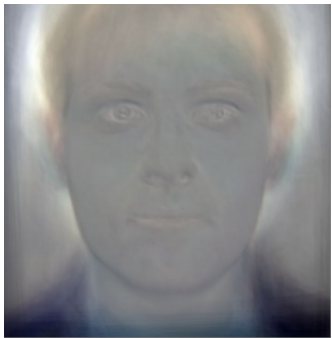
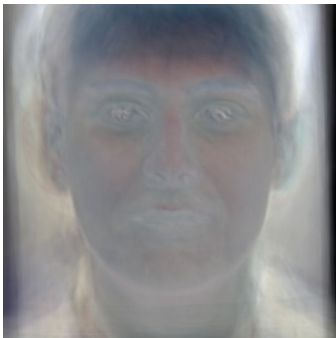

學號：b04501095 系級：土木三 姓名：黃平瑋

### A. PCA of colored faces

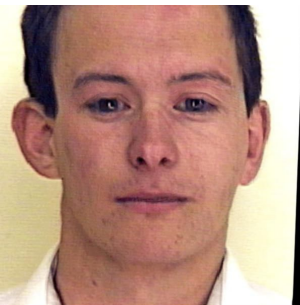

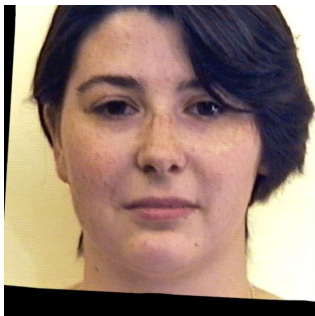

A.1. (.5%) 請畫出所有臉的平均。

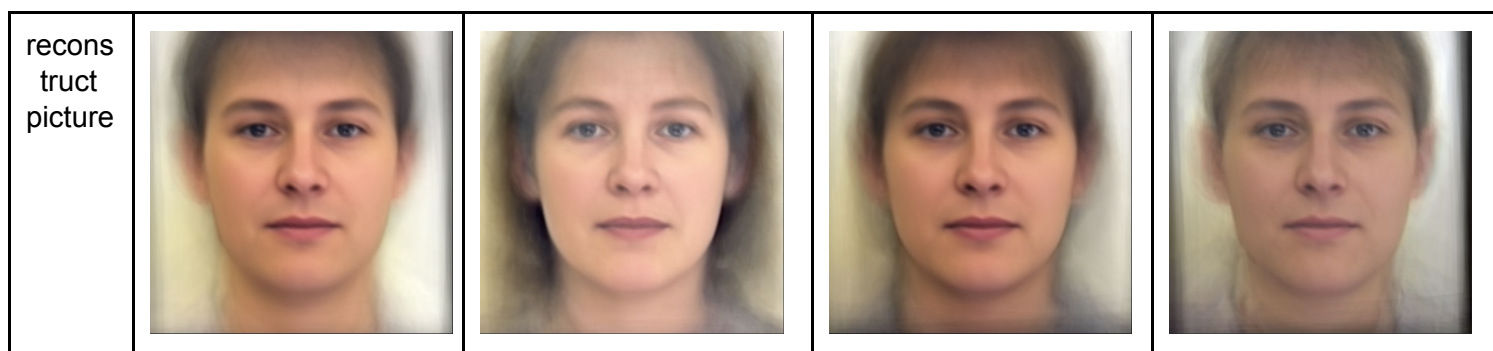


A.2. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。

eigenface1	eigenface2	eigenface3	eigenface4
			

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。

picture name	100.JPG	200.JPG	300.JPG	400.JPG
original picture				



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。

	eigenface1	eigenface2	eigenface3	eigenface4
比重( % )	4.1	2.9	2.4	2.2

## B. Image clustering

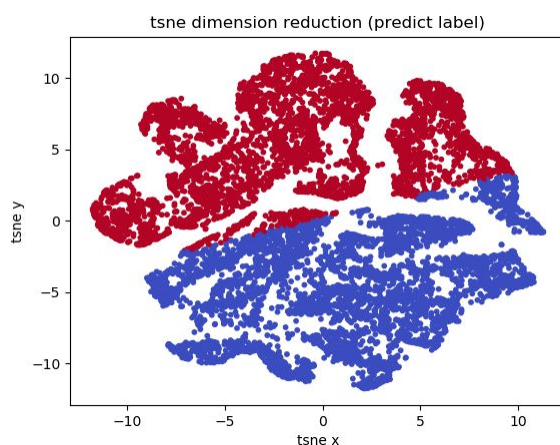
B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

方法	public score/ private score
先用PCA 降到400維, 再用k means分成兩群	100 %
先用PCA 降到50維, 再用tSNE到2維, 再用k means	99.98 %

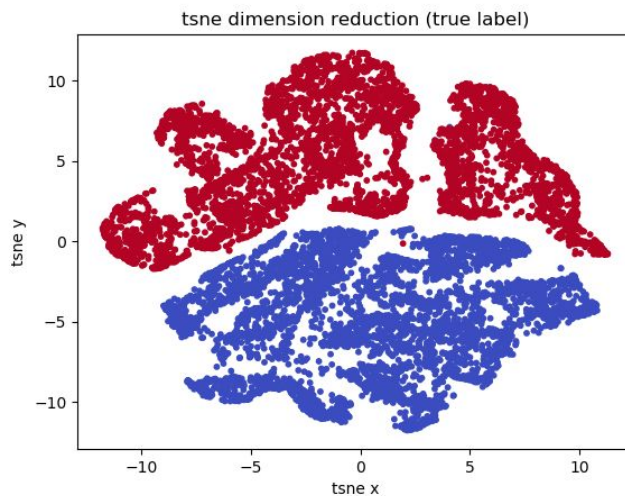
這次作業我實做的方法是第一種, 先用PCA降維再用Kmeans分成兩群, 再用不同的 random seed去調整不同的初始點, 最後恰巧正確的分出100%的正確率

第二個方法先用pca降到一定的維度, 再使用tsne, training的時間明顯比第一種還要久, 但卻沒有達到理想的結果

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。



可以看到我預測的結果和真實的Label有些不同,主要是我切的邊界和正確的label邊界不太一樣,下表示預測和真實結果的差距

預測結果	class 0	class 1
class 0	4750	250
class 1	441	4559

## 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。

在作業三的CNN中我使用了ensemble learning中的bagging來解決分類的問題, 我總共 trained 5個架構不盡相同的model, 將所有model通過sigmoid function的結果加總起來, 最後再用 voting決定是屬於哪個class

	public score	private score
model 1	0.66202	0.65979
model 2	0.66453	0.66369
model 3	0.68403	0.66843
model 4	0.68932	0.68124
model 5	0.67539	0.66258
<b>ensembl model</b>	<b>0.71273</b>	<b>0.70437</b>

可以看到經過ensemble後的結果, 比任何一個model的表現都還要好, 有效降低了variance, 大約增加了3%的準確度