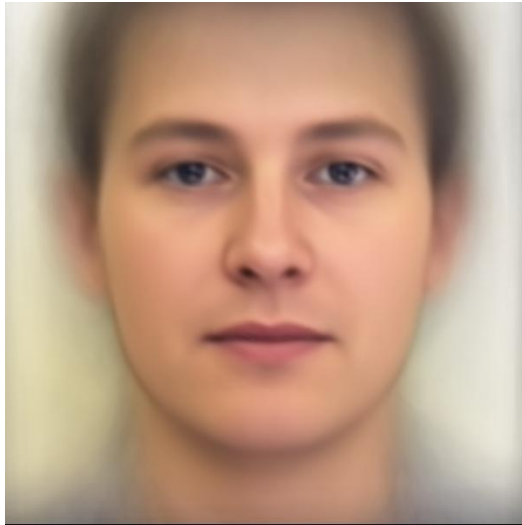


A. PCA of colored faces

1. (.5%) 請畫出所有臉的平均。



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

1st Eigenface

2nd

3rd

4th

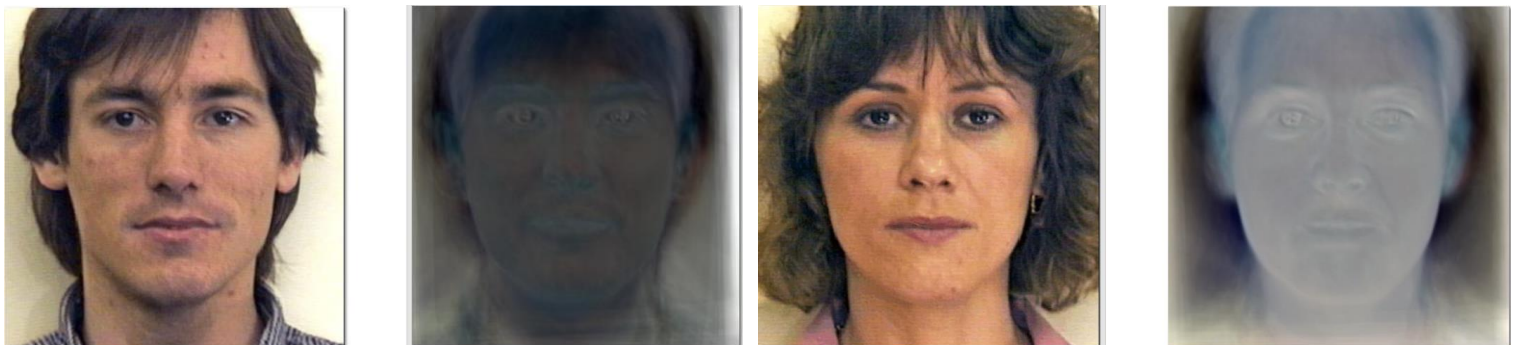


3. (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。

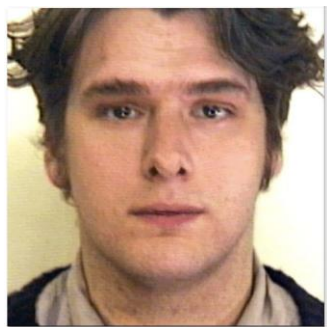
以下為分別使用 2.jpg、4.jpg、6.jpg、15.jpg 進行 reconstruction，左邊為原圖，右邊為 reconstruction 的結果

2.jpg

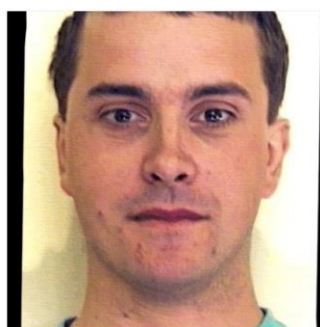
4.jpg



6.jpg



15.jpg



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。

1st: 4.1% 2nd: 2.9% 3rd: 2.4% 4th: 2.2%

B. Image clustering

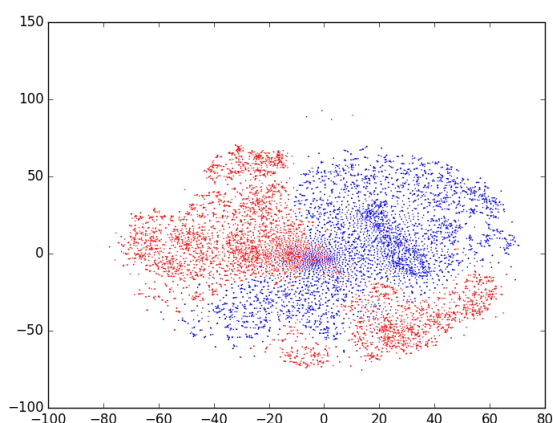
1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

本次作業我使用 PCA(降到 274 維)與 Kmeans 的方法，在 public 跟 private score 的部分都得到了 0.99998 的分數；而另外一個方法則是使用助教提供三層的 autoencoder，加上 Kmeans 的方法來進行比較，結果如下。

可以看出 PCA 有較好的表現，但 PCA 的維度選擇會對準確度造成很大的影響，可能使用 273 維，其準確率會低於 0.5，但使用 274 維，其準確率就會有 0.99，可說是落差非常大，因此使用 PCA 感覺有種僥倖的心態阿。

	Public score	Private score
Autoencoder + Kmeans	0.97796	0.97803
PCA(n=274) + Kmeans	0.99998	0.99998

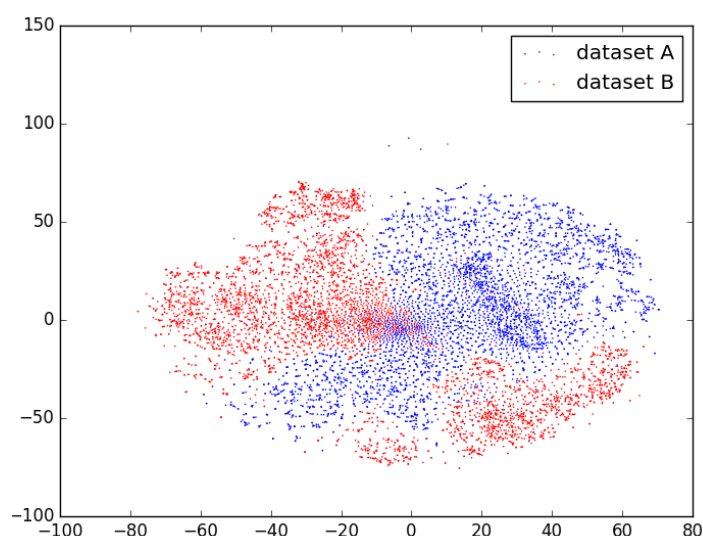
2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。



預測結果如左圖，我先用 PCA 將 data 降到 274 維，再使用 TSNE 降到兩維，並代表每個點的 X 跟 Y。

而圖中的 label 則是使用 PCA 降為到 274 維後，使用 Kmeans 方式所預測出來的結果。

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。



右圖為正確的 label 的分布，與上圖比較起來幾乎沒有差異，這次的預測十分成功。

C. Ensemble learning

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）

在此我選擇以 hw3 作為 ensemble 的目標，事先 train 好 20 個 model 後，將 data 依序輸入這些 model，每筆 data 都在不同的 model 都有不同的結果，這時候再使用 bagging 的 voting 方法，得到一個較好的預測結果，結果如下

	Public score	Private score
沒有使用 ensemble(註)	0.69629	0.69713
使用 ensemble	0.70270	0.69127

註：該筆 model 是從 ensemble 所使用的 20 個 model 中，預測結果最好的

雖然使用 ensemble 後，public score 變的較好，但 private score 不升反降，可能是被其他不夠準確的 model 所拖累，因此在進行 ensemble 時，還是要取用表現最好的那幾個 model 就好，這樣結果應該會變得更好。