

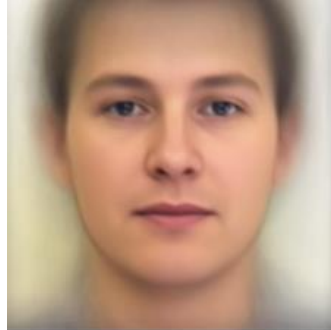
HW4

學號：r06944051 系級：網媒碩一 姓名：郭柏辰

A. PCA of colored faces

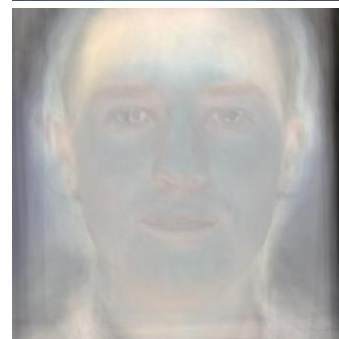
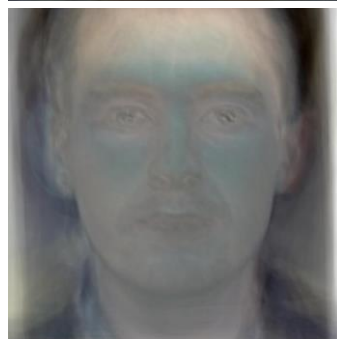
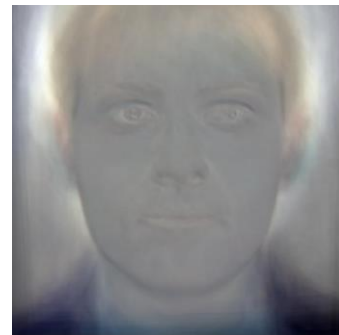
A.1. (.5%) 請畫出所有臉的平均。

Sol:



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

Sol:

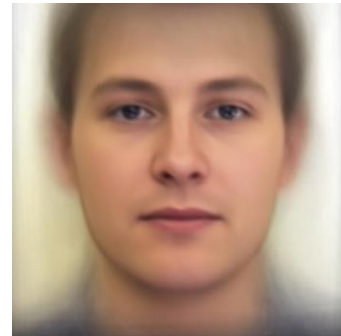


- A.3. (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。

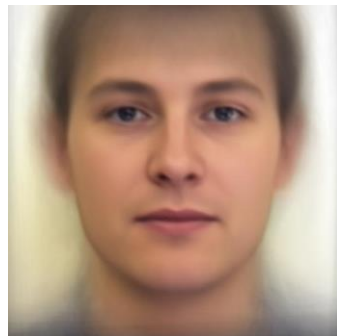
Sol:



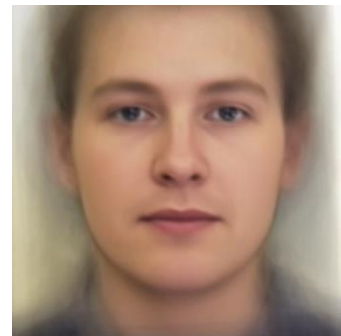
Index = 20



Index = 80



Index = 180



Index = 280

- 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。

Sol:

- (1) 4.1%
- (2) 2.9%
- (3) 2.4%
- (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)

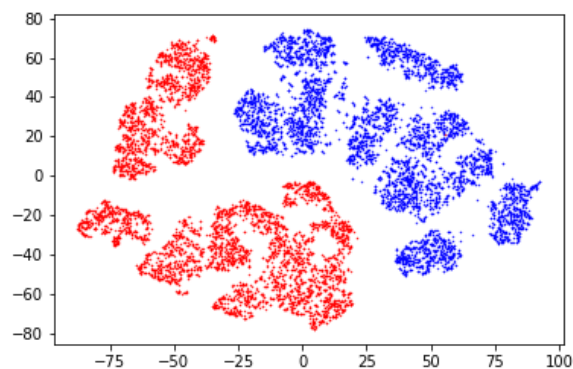
Sol:

	Public Score	Private Score
PCA	0.99995	0.99995
Autoencoder	0.99690	0.99695

我實作 Autoencoder 降到 32 維，在 kaggle 分數為 public score: 0.99690, private score: 0.99695。實作 PCA 降到 200 維，在 kaggle 分數為 public score: 0.99995, private score: 0.99995。雖然 PCA 在分數上的表現較好，但是降的維度沒有 Autoencoder 降的多。

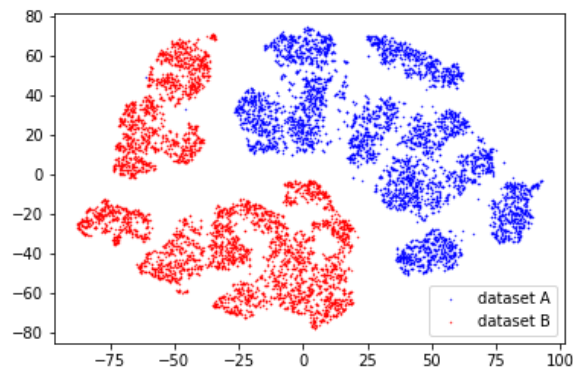
- B.2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。

Sol:



- 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。

Sol:



跟上題相比，可以發現預測的結果跟真實結果是差不多的。因為在 kaggle 上的準確率 **Public Score: 0.99690, Private Score: 0.99695** 幾乎等於 1 了，所以可預期預測結果跟真實結果幾乎會一樣。

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）

Sol:

我在 hw3 實作 ensemble learning，我訓練了 4 個 model，分別取出 4 個不同 model 的 weight，算出不同 model weight 的平均值，以決定最後的輸出。

	Public Score	Private Score
實作前	0.69350	0.67957
實作後	0.69350	0.68626

實作後發現在 **Private Score** 略高一點，可能是我在訓練 4 個不同 model 時，訓練時間不夠長的原因。

Reference: <https://github.com/WaitingLin/ML2017FALL/tree/master/hw6>