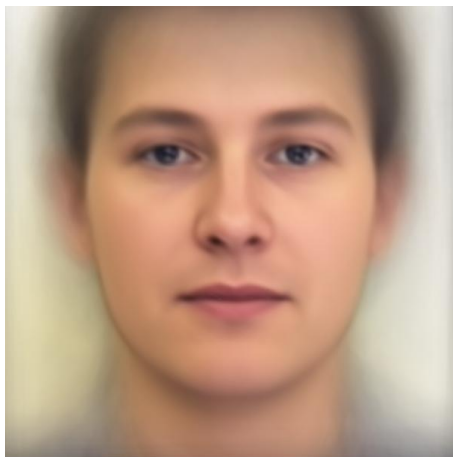


HW4

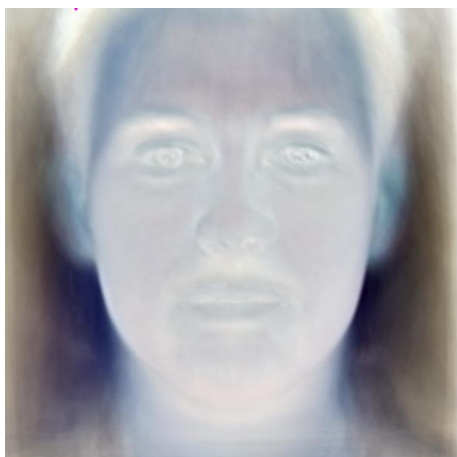
學號：B05902002 系級：資工二 姓名：李栢淵

PCA of colored faces

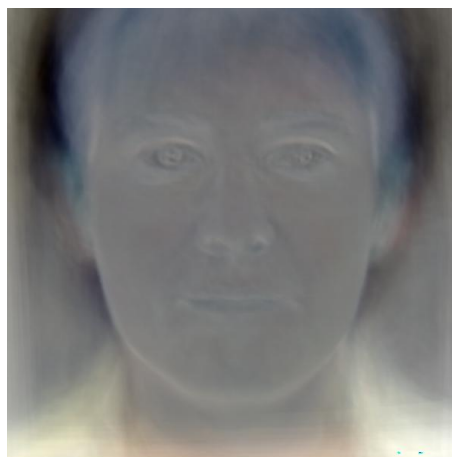
(.5%) 請畫出所有臉的平均。



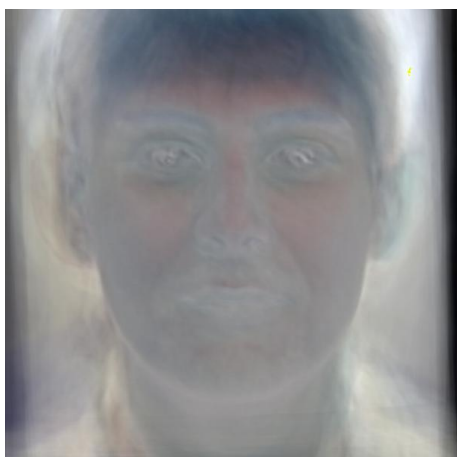
(.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。



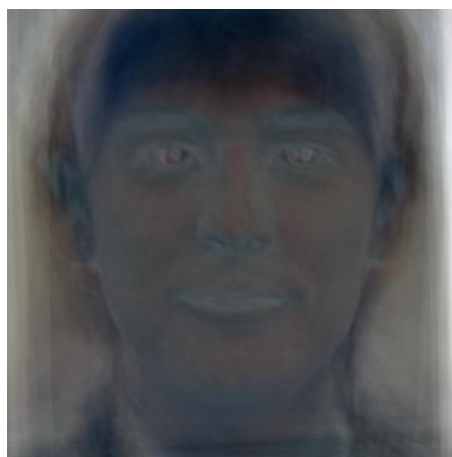
1



2



3



4

(.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

(.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。

No.1 → 4.1% (0.041475)

No.2 → 3.0% (0.029508)

No.3 → 2.4% (0.023894)

No.4 → 2.2% (0.02209)

Image clustering

(.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)

Autoencoder(DNN) 降維：

[try.csv](#)
2 days ago by Bai-Yuan Lee
[add submission details](#)

0.99781

☒

PCA 降維：

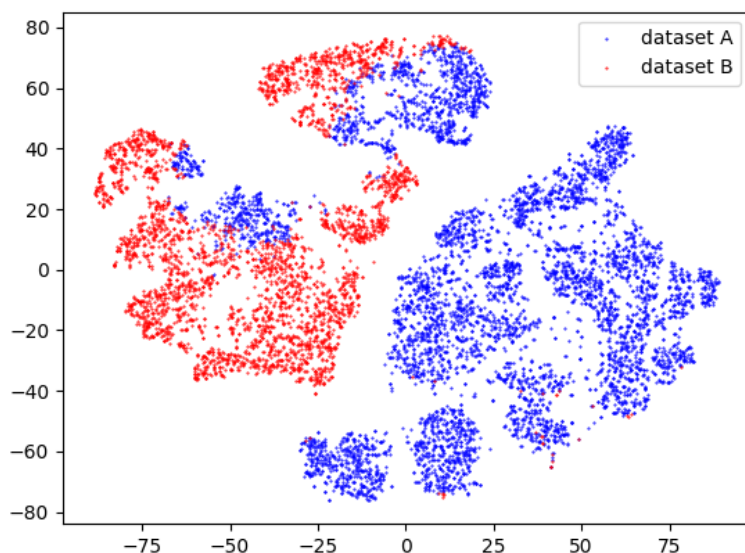
[try_PCA.csv](#)
an hour ago by Bai-Yuan Lee
[add submission details](#)

0.99998

☒

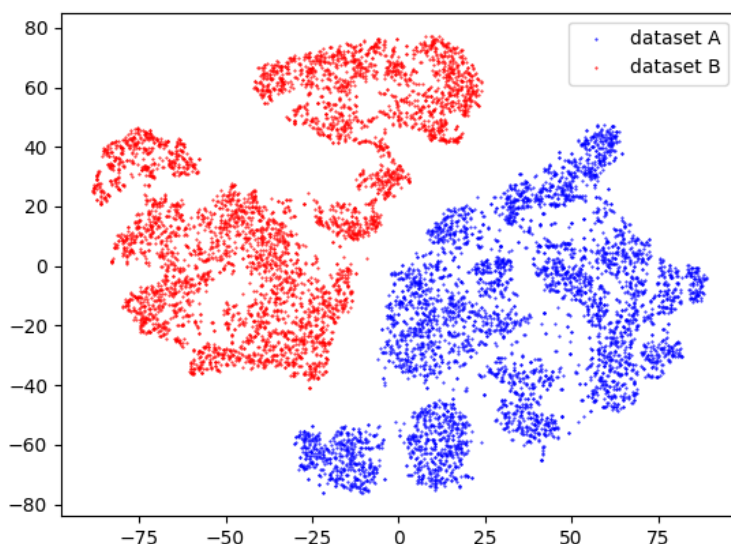
Autoencoder降維的結果比PCA降維顯得差一點，Autoencoder是用DNN降到100維，而PCA降到四百多維。可能是這筆資料比較好分，用PCA就可以很有效。我DNN的epoch數目是50，經觀察發現，如果epoch數目太多，最後的結果反而變差。

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



(.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。

這是我用 Autoencoder 降維的結果，這樣的降維方法最後 accuracy 只有0.99781。
明顯地除了一些邊界，我有一大塊藍色部分判成紅色部分。但整體來說大部分的情形都有判對，還是做得不錯。



Ensemble learning

(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）

final_2.csv 14 days ago by Bai-Yuan Lee add submission details	0.68069	0.69155	<input checked="" type="checkbox"/>
good8.csv 14 days ago by Bai-Yuan Lee add submission details	0.65282	0.65477	<input type="checkbox"/>

圖片的下方（good8.csv）是我做出來的一個cnn的結果，有過Simple但沒過Strong，將其重複train四次存了四個model。因為model的output是一個array，表達各類別的機率，所以我將他們的結果加起來平均，再取機率最高的輸出，就達到上面的結果（final_2.csv），明顯變好很多。