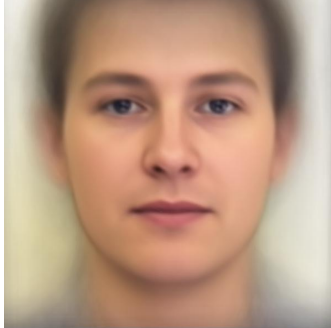


Homework 4

學號： b05902008 系級： 資工二 姓名： 王行健

A. PCA of coloured 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。

Eigface1

Eigface2

Eigface3

Eigface4



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。

0.jpg

1.jpg

2.jpg

3.jpg

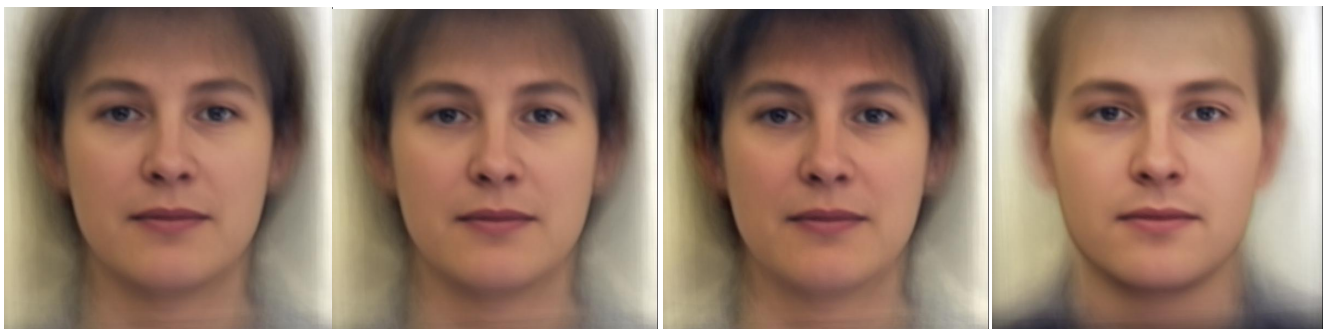


reconstruct0

reconstruct1

reconstruct2

reconstruct3



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。

eigface0 : 4.1%
eigface1 : 2.9%
eigface2 : 2.4%
eigface3 : 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)

(1)CNNAutoEncoder(降到10為)+KMeans(2 clusters) : Kaggle Score 1

(2)Naive Feature Extraction(取17個feature)+If-Else Dividing : Kaggle Score 1

features :

最長縱/橫向非零序列長度

column0 非零byte數量

column1 非零byte數量

column2 非零byte數量

column3 非零byte數量

column24 非零byte數量

column25 非零byte數量

column26 非零byte數量

column27 非零byte數量

row0 非零byte數量

row1 非零byte數量

row2 非零byte數量

row3 非零byte數量

row24 非零byte數量

row25 非零byte數量

row26 非零byte數量

row27 非零byte數量

Dividing :

Label1 :

最長縱/橫向序列長度 ≥ 22

column0 非零byte數量 ≥ 2 and max(column1,2,3非零byte數量) ≥ 10

column27 非零byte數量 ≥ 2 and max(column26,25,24非零byte數量) ≥ 10

row0 非零byte數量 ≥ 2 and max(row1,2,3非零byte數量) ≥ 10

row27 非零byte數量 ≥ 2 and max(row26,row25,row24非零byte數量) ≥ 10

column0,1,26,27 非零byte數量 > 0

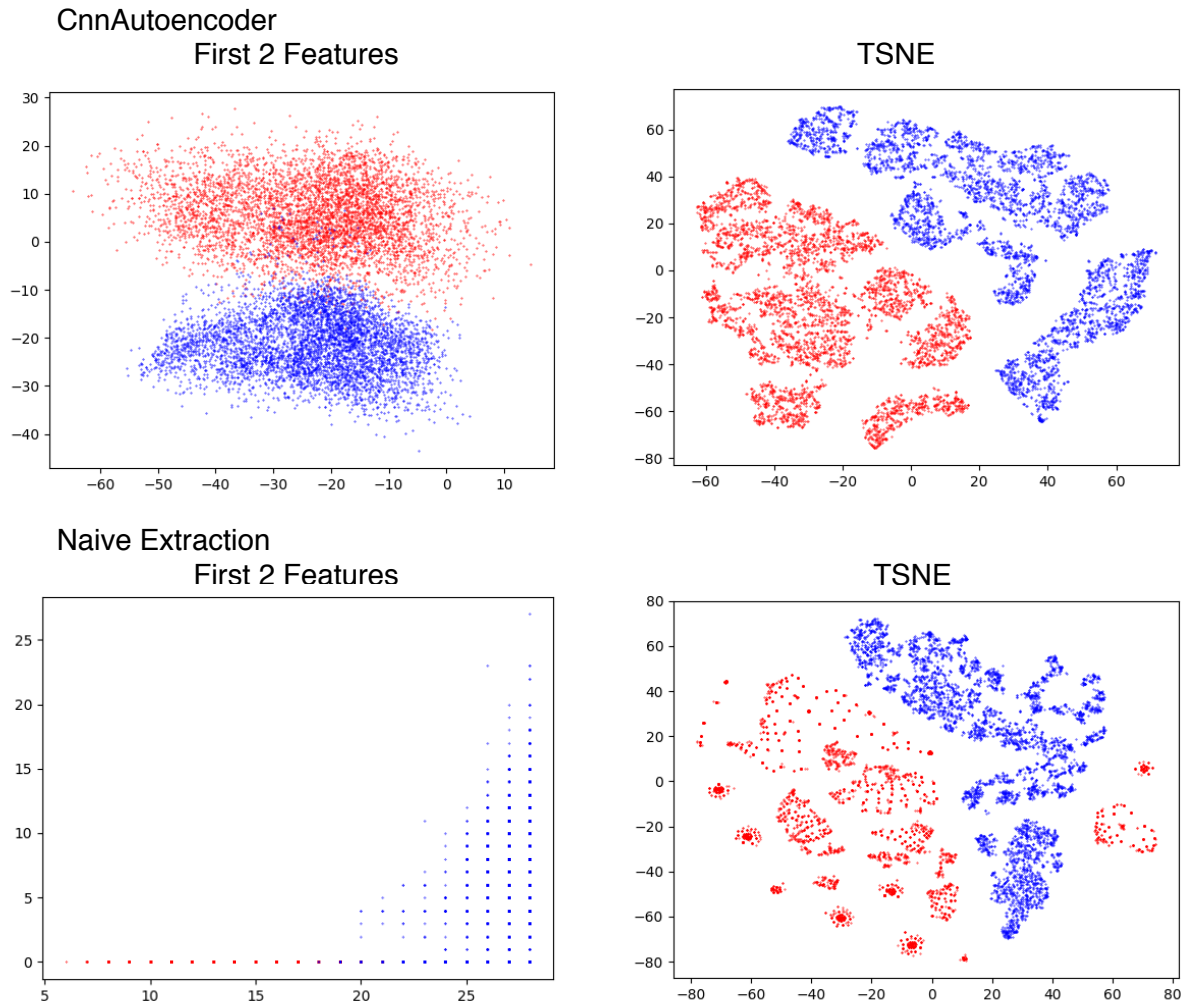
row0,27 非零byte數量 > 0 and row1,26 非零byte數量 ≥ 5

Label0 :

其他

****兩者標出的label完全相同**

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。

預測label和實際label完全一樣，visualize的圖也相同(礙於頁數不再重複放一次)

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）

task : hw2

approach : 建99個NN classifier，然後投票決定最後結果。投票時每個模型的權重相等，換句話說，就是以多數決來選定結果。

accuracy : before ensemble [public : 0.84373, private : 0.84191]
after ensemble [public : 0.87346, private : 0.86807]

可以看到，在ensemble後，不論public或private都有約3%的提升。這顯示train多個模型並投票決定結果的確可以使其互相修正錯誤。