

HW4

學號：B05902019 系級：資工二 姓名：蔡青邑

### 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。

由左而右1, 2, 3, 4



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。



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。 4.1% 3.0% 2.4% 2.2%

## B. Image clustering

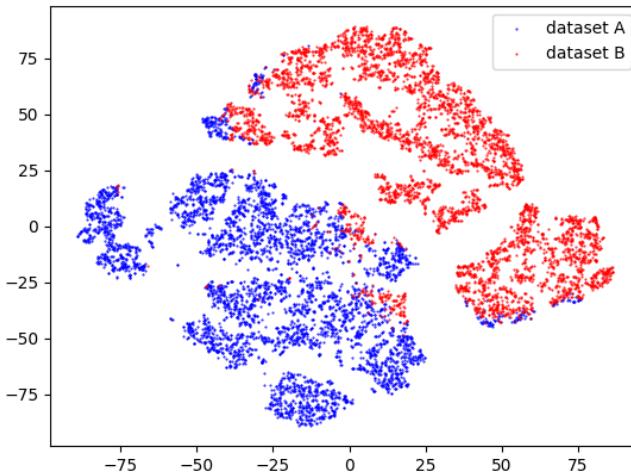
B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

|   |         |         |                                     |
|---|---------|---------|-------------------------------------|
| cnn_prediction.csv<br>a day ago by B05902019蔡青邑<br>add submission details | 0.66244 | 0.66223 | <input type="checkbox"/>            |
| prediction.csv<br>2 days ago by B05902019蔡青邑<br>add submission details    | 0.96338 | 0.96337 | <input checked="" type="checkbox"/> |

(下面的)DNN: 96%, (上面的)CNN: 66%

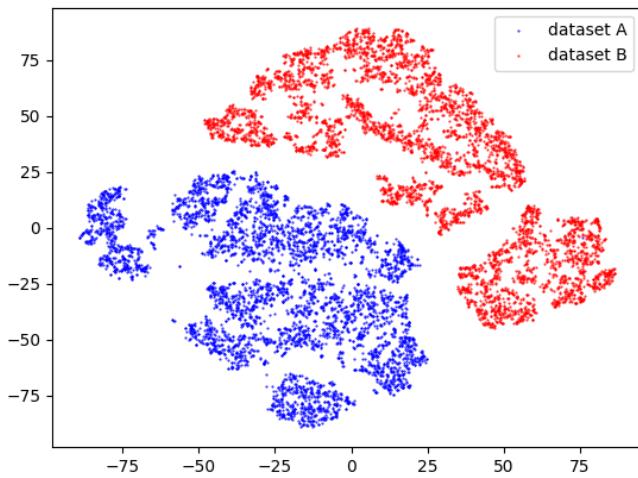
我認為我的CNN model產生了過多不必要的feature，因此表現較差（因為我有減少過CNN的parameters，結果真的能提升accuracy 33%->66%）

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



利用TSNE 投影

B.2. (.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。



由於TSNE降維會保留原本維度的相對性，所以降維後還可以出現明顯的小分群。而我的預測結果雖然有點準卻切過了許多小分群，代表我DNN的降維做法可能沒有保留特定的關係(比如說高維與低維)，而是直接切開。

### 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）

|   |         |         |                                     |
|---|---------|---------|-------------------------------------|
| <a href="#">3in1.csv</a><br>14 days ago by B05902019蔡青邑<br><a href="#">add submission details</a>               | 0.65087 | 0.65867 | <input type="checkbox"/>            |
| <a href="#">medium.csv</a><br>15 days ago by B05902019蔡青邑<br><a href="#">add submission details</a>             | 0.64558 | 0.64502 | <input checked="" type="checkbox"/> |
| <a href="#">10addmoreconvandNN.csv</a><br>21 days ago by B05902019蔡青邑<br><a href="#">add submission details</a> | 0.64641 | 0.64809 | <input type="checkbox"/>            |
| <a href="#">11addNN.csv</a><br>20 days ago by B05902019蔡青邑<br><a href="#">add submission details</a>            | 0.61799 | 0.63081 | <input type="checkbox"/>            |

我在hw3時就做過ensemble，最上面的”3in1.csv”是下面三個model合起來的結果。主要的實作方式就是將下面三個model對各個testcase的6個類別的預估機率相加，然後取argmax，以三個model預測機率加總最大的那個類別作為最後的預測結果。