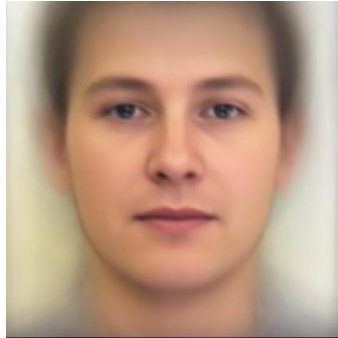
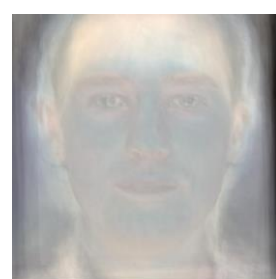
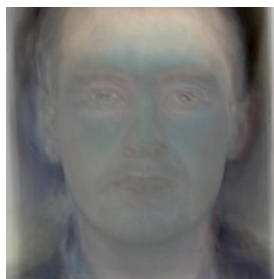
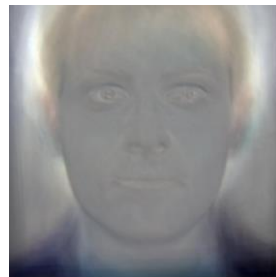
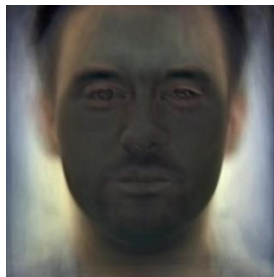


A. PCA of colored faces

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。

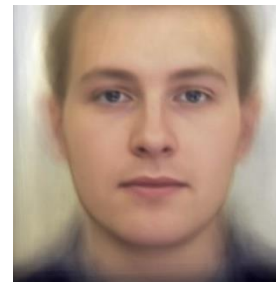


3. (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。

我挑選 index = 66, 67, 77, 87 重建



Index=66



index=67

3. (.5%) 請討論你從 `visualization` 的結果觀察到什麼。
- 我發現我用藍色圈起來的部分，裡面有你、你們、他們、我、我們等等的詞，在二維空間上的距離都蠻近的！和、與相似詞也在隔壁而已，

C. 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)
- 我一開始使用 `pca` 降到 32 維，然後使用 `Kmeans` 分群，丟到 `kaggle` 後結果很差。

prediction.csv
8 days ago by Wei Ting Lin
PCA

0.07963

0.07942



後來改用多層 `autoencoder`，train 兩百個 `epoch` 後得到我最好的 `model`

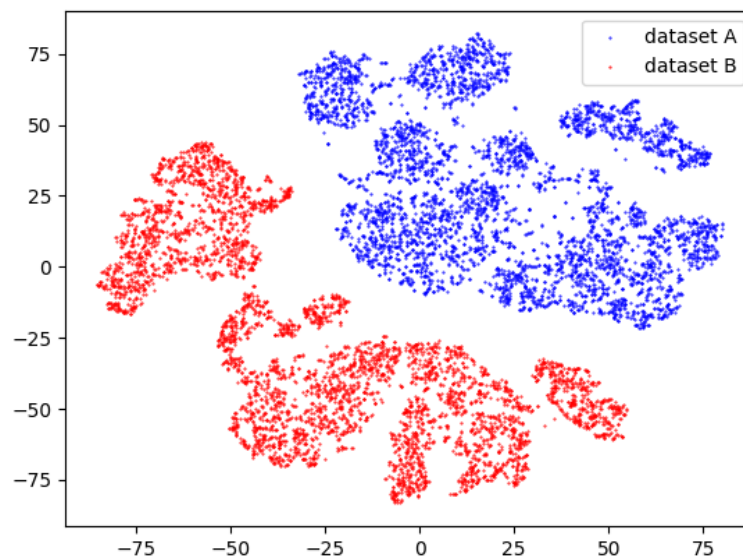
prediction.csv
4 days ago by Wei Ting Lin
Best model

1.00000

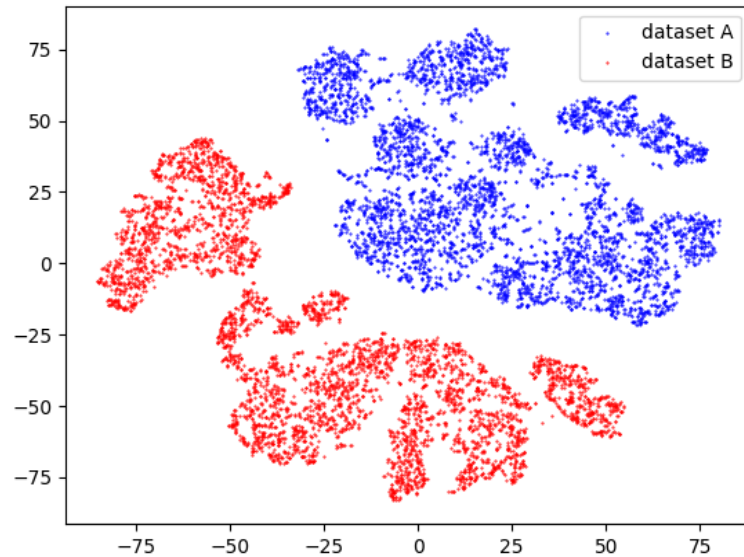
1.00000



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



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。



比較結果發現一樣，於是印出全部的 label 看看。發現前五千個都被分類到 **1**，後五千個都分類到 **0**。因此完全預測正確!