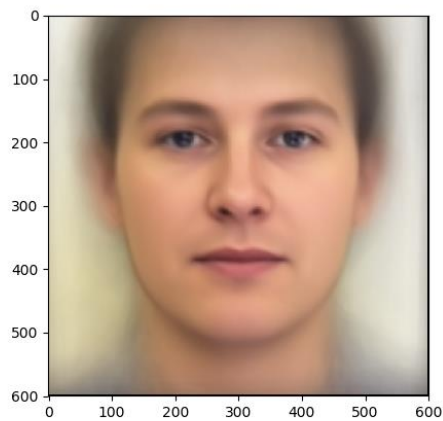


## HW4

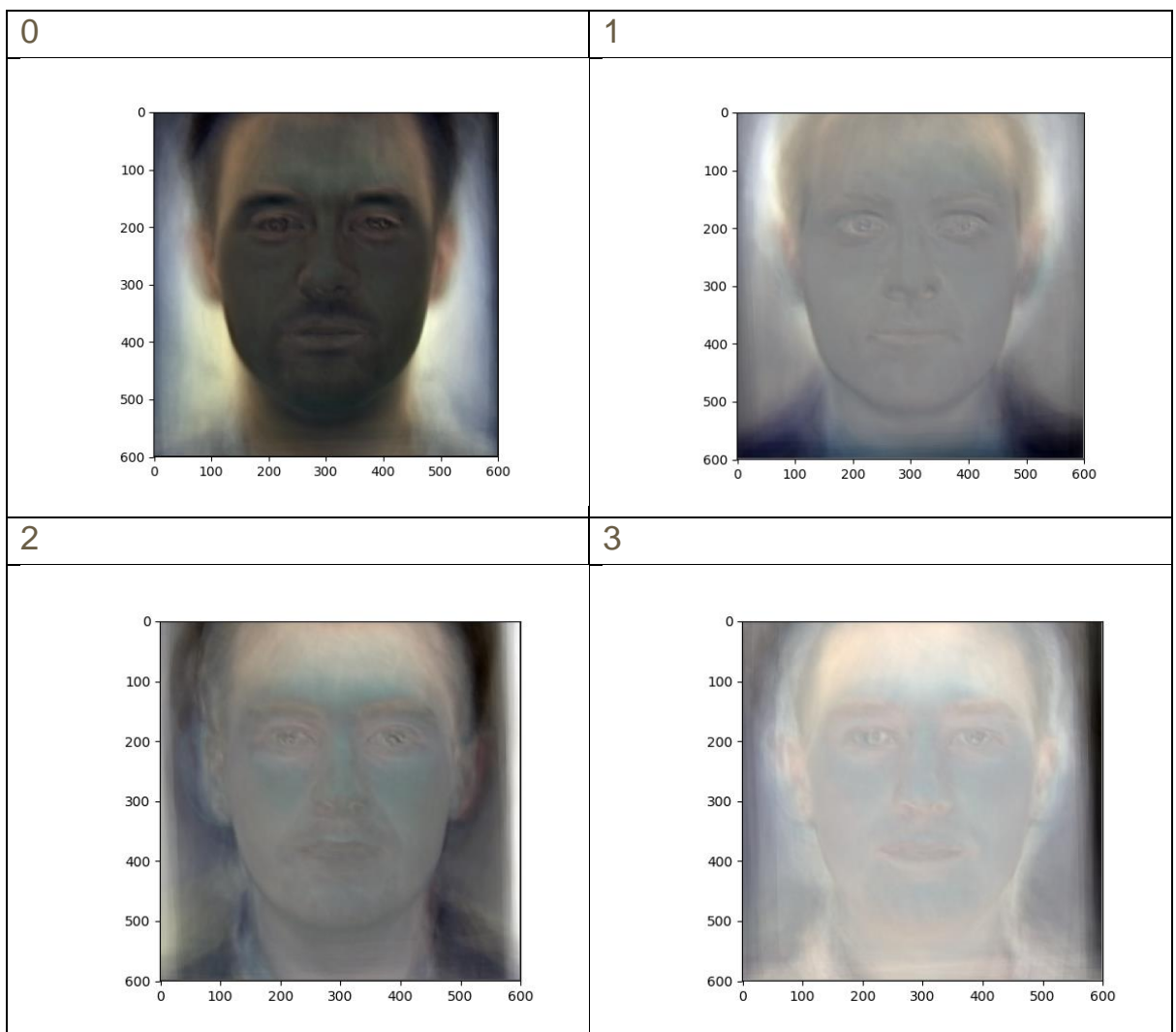
學號：r06921058 系級：電機碩一 姓名：方浩宇

### A. PCA of colored faces

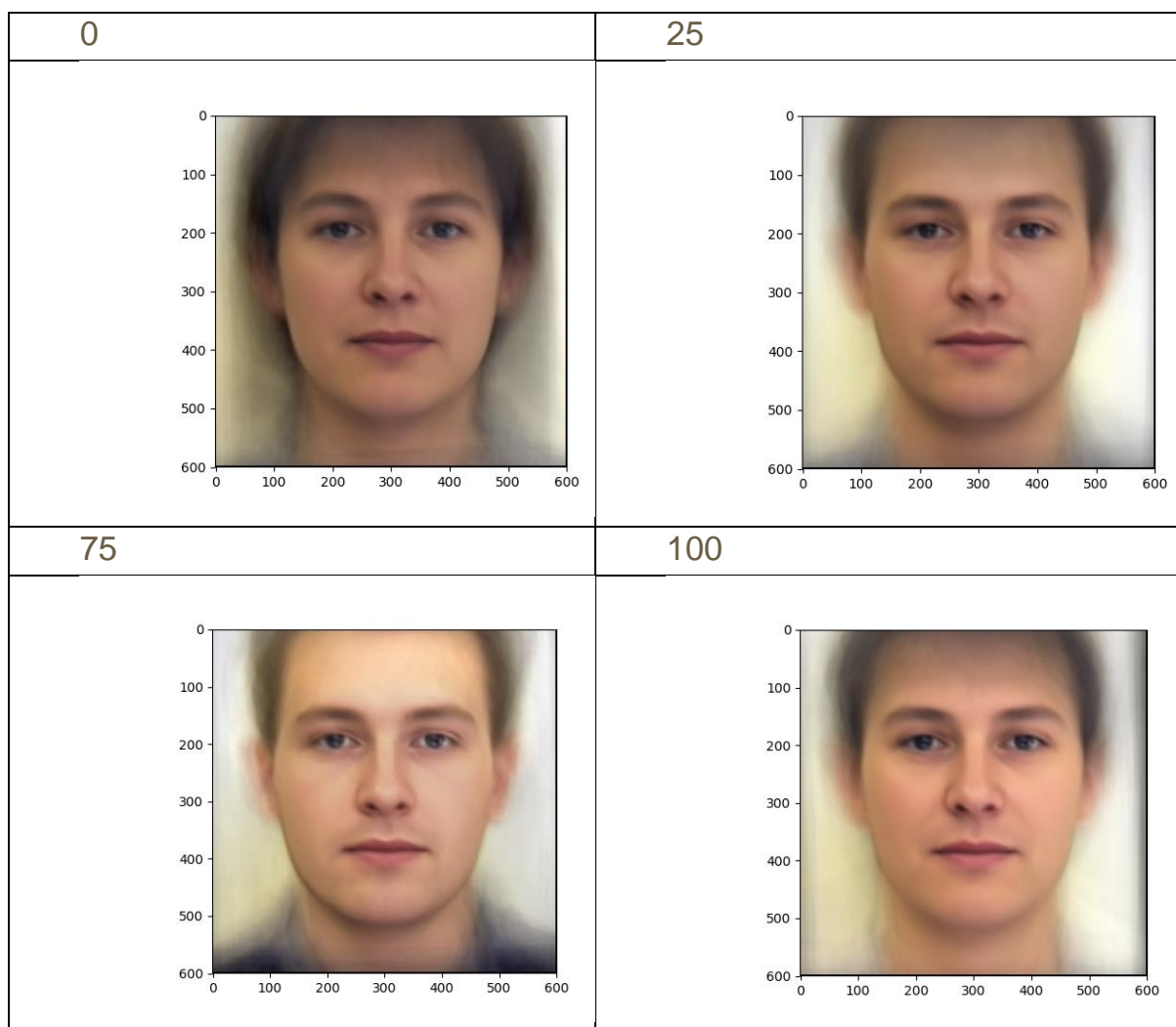
A.1. (.5%) 請畫出所有臉的平均。



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



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。

```
0.04142863127519326 S0 %
0.029472577645606967 S1 %
0.023867038488061692 S2 %
0.022068041899087336 S3 %
```

0	1	2	3
4.1%	2.9%	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)

[2\\_Kmeans\\_PCA.csv](#)  
7 days ago by [r06921058\\_>.O](#)  
[add submission details](#)

0.99998

0.99998



[2\\_Kmeans\\_PCA.csv](#)  
7 days ago by [r06921058\\_>.O](#)  
[add submission details](#)

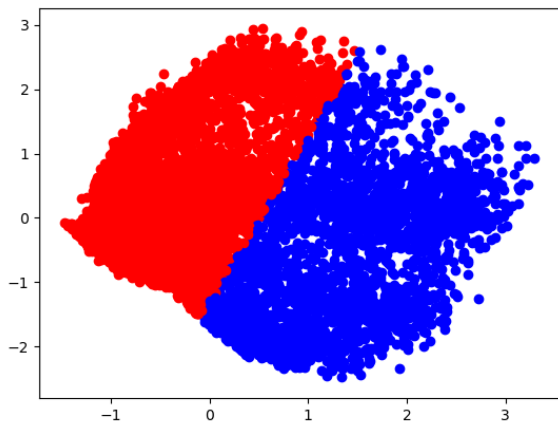
0.51767

0.51691

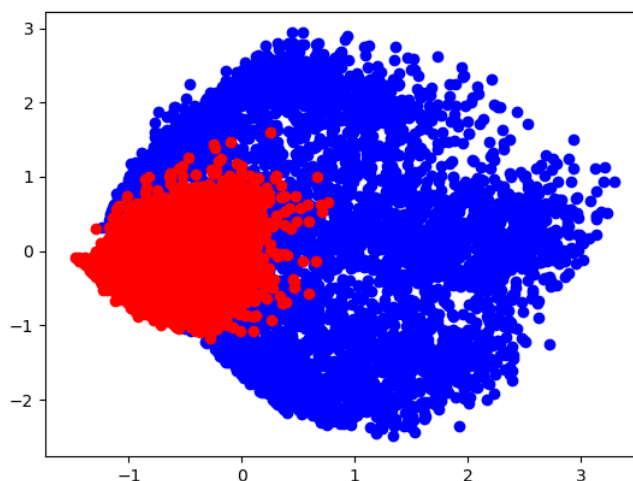


兩個分別是 PCA 有使用 WHITEN 和沒使用的結果，可以發現 Whiten 對結果有顯著的影響。

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。



這是正確的分類的圖，可以發現，如果直接用 PCA 降到 2 維，會有很多資料被搞混(失去太多資訊)。

## 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）

[ans\\_ensemble.csv](#)  
11 minutes ago by r06921058\_>O  
[add submission details](#)

0.85726

0.85896



這是 ensemble 的 MODEL 的結果，是使用簡單的 Voting 的方式，使用的三個

MODEL Public 分別為：8.59,8.58,8.57，可以發現結果介於這三者之間。我實作方式是直接將三個 Predict 的結果做投票，使用基本的多數決。