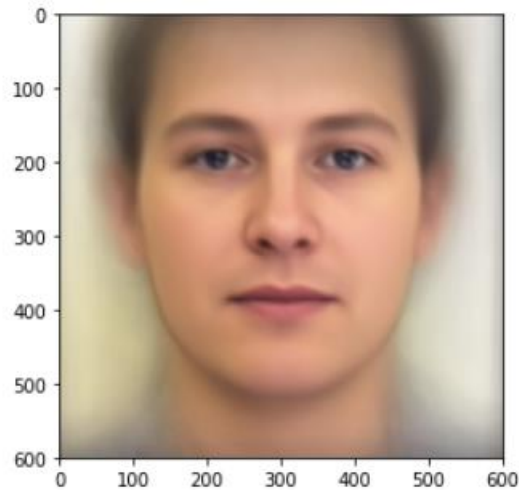


HW4

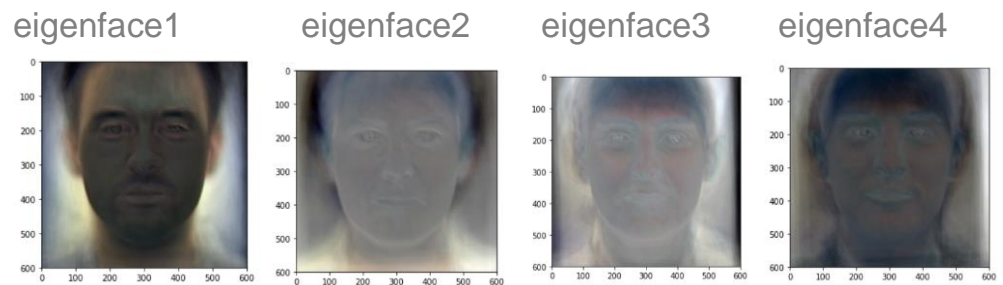
學號：r06922086 系級：資工所 姓名：林凡煒

A. PCA of colored faces

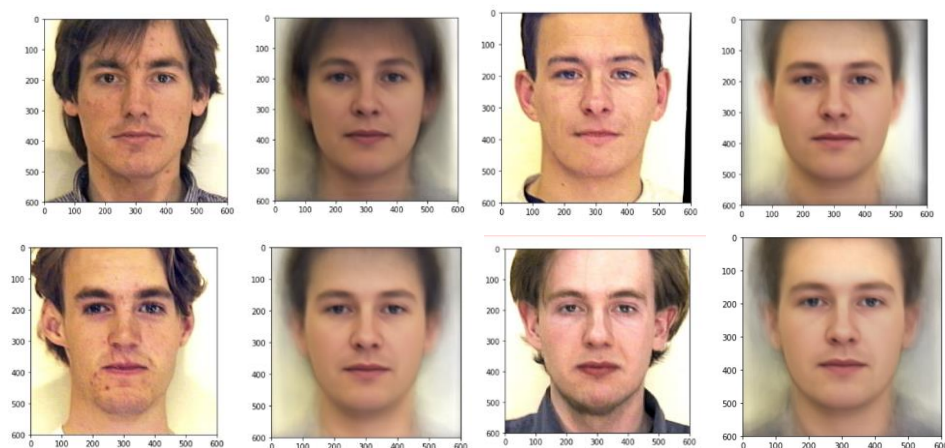
A.1. 請畫出所有臉的平均。



A.2. 請畫出前四個 Eigenfaces。



A.3. 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。

Ratio 1: 4.1%, Ratio 2: 2.9%, Ratio 3: 2.4%, Ratio 4: 2.2%

B. Image clustering

B.1. 請比較至少兩種不同的 **feature extraction** 及其結果。

1. PCA 取 50 Components 作 K-means:

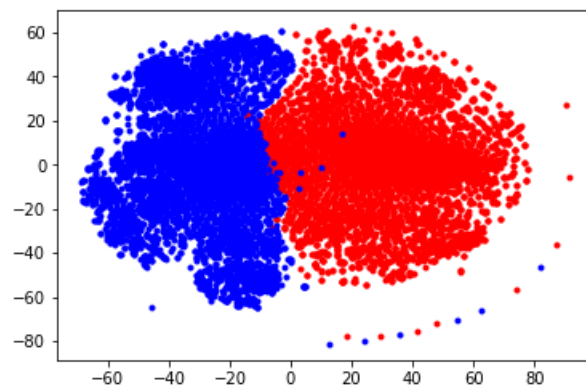
Kaggle' s private/public score: 0.52073 / 0.52011

2. PCA 取 50 Components + t-SNE 降成 2 維 作 K-means

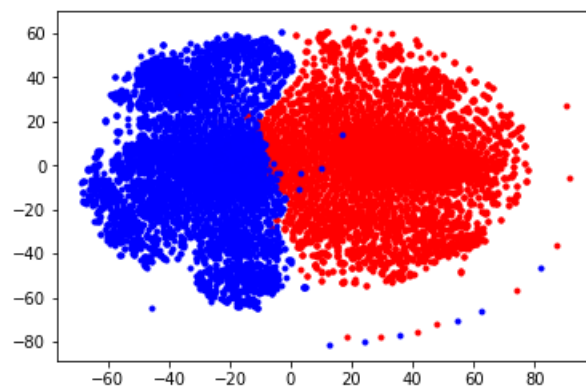
Kaggle' s private/public score: 0.69837 / 0.69773

可以看出有作 t-SNE 後的效果是比較好的。

B.2. 預測 `visualization.npy` 中的 **label**，在二維平面上視覺化 **label** 的分佈。



B.3. `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。



與預測結果一模一樣。

C. Ensemble learning

C.1. 請在 hw3 的 task 上實作 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。

用 uniform voting 的方式去決定最後 predict 的結果，其 performance 表現如下：

Model1 private/public score: 0.67929 / 0.67233

Model2 private/public score: 0.67679 / 0.68152

Model3 private/public score: 0.66341 / 0.67316

Ensemble model private/public score: 0.70214 / 0.69824

由結果可以看出，即使是使用 simple uniform 的方式作 ensemble，確實可以使得整體表現往上再提升。