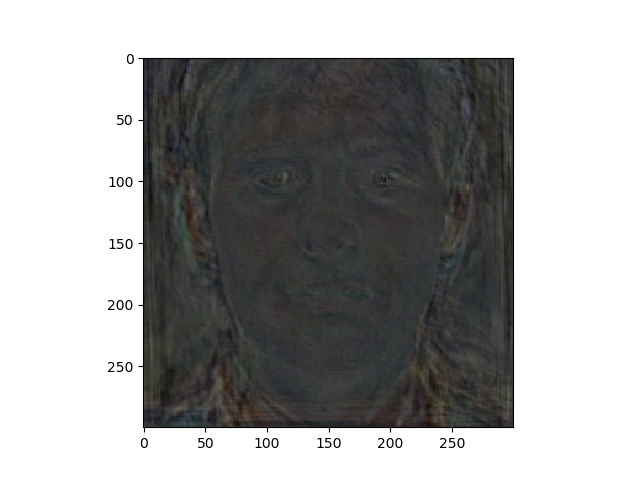
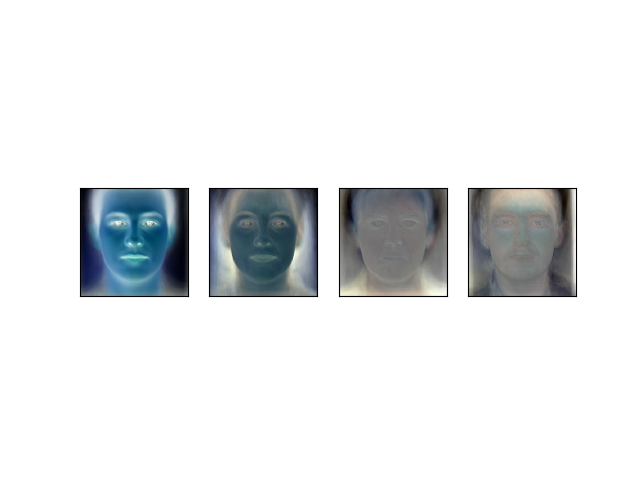
學號：R06922152 系級： 資工碩一 姓名：袁晟峻

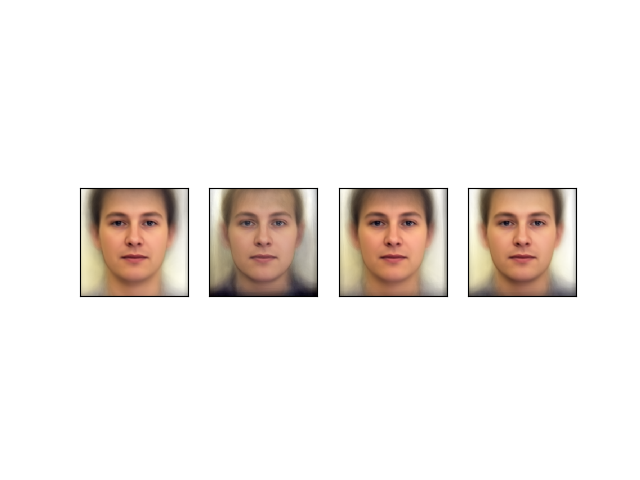
1. PCA of colored faces
   1. (.5%) 請畫出所有臉的平均。



* 1. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。



* 1. (.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。



* 1. (.5%) 請寫出前四大 Eigenfaces 各自所佔的比重 (explained variance ratio)，請四捨五入到小數點後一位。

7.4%、3.7%、2.8%、2.2%

1. Visualization of Chinese word embedding
   1. (.5%) 請說明你用哪一個 word2vec 套件，並針對你有調整的參數說明那個參數的意義。
   2. (.5%) 請在 Report 上放上你 visualization 的結果。
   3. (.5%) 請討論你從 visualization 的結果觀察到什麼。
2. 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)
3. 使用auto-encoder + Kmeans：降維到128維後用Kmeans分群，在kaggle public上得到1.0的分數。
4. 使用TSNE + Kmeans：直接降到二維後用Kmeans分群，但是效果很差，在kaggle public上只得到0.025的分數，而且花的時間還很久。
   1. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。
   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。