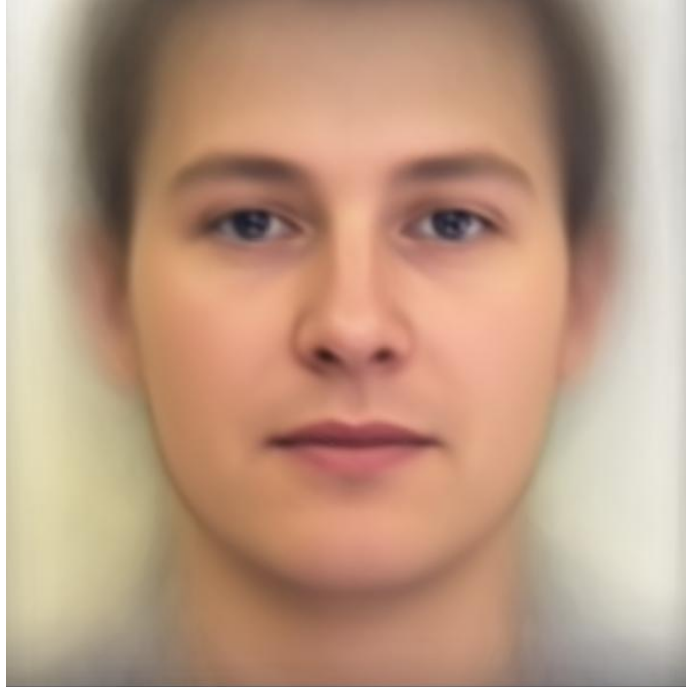


學號：B04901025 系級：電機三 姓名：陳鴻智

A. PCA of colored faces

A.1. (.5%) 請畫出所有臉的平均。

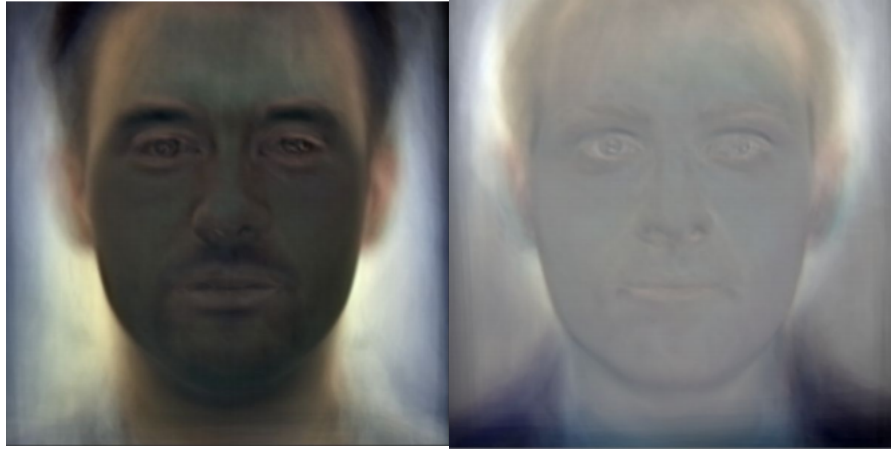


A.2.

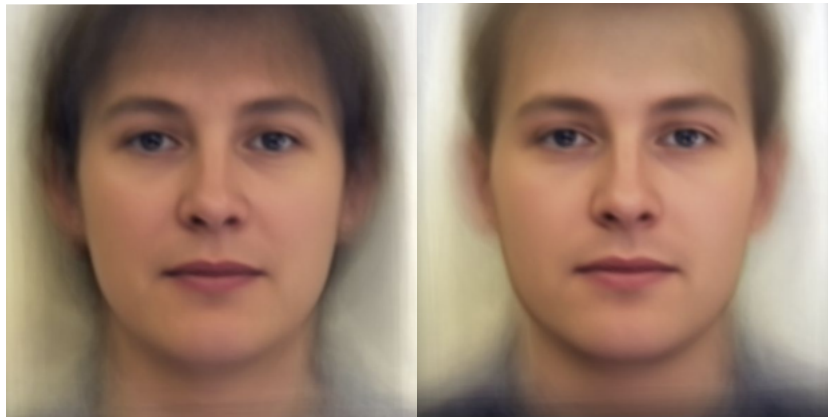
A.3. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。



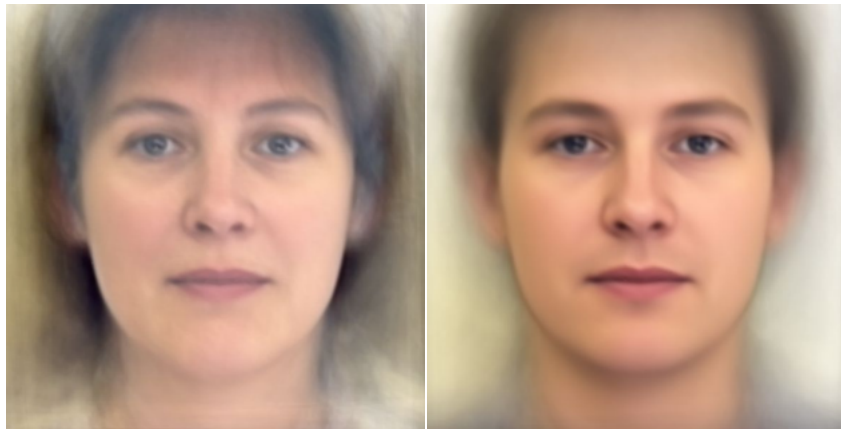
A.4.



A.5. (.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.6.



A.7. (.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。

A.8. ratio of eigenface1 : 4.2%

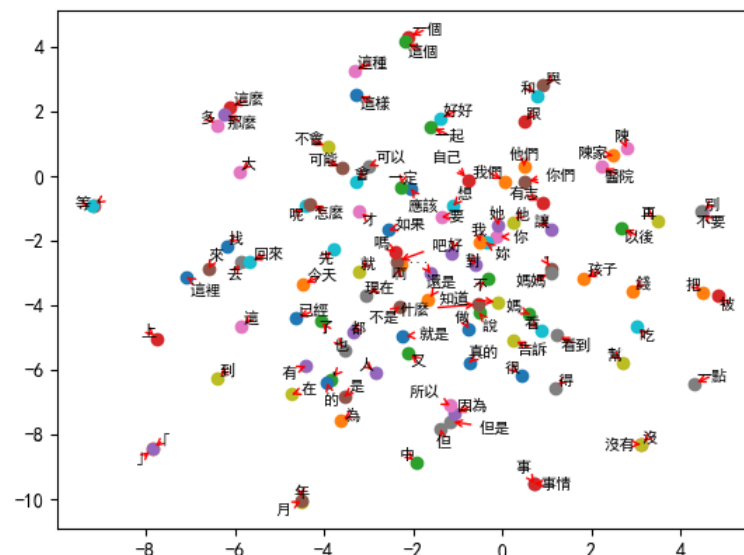
A.9. ratio of eigenface2 : 2.9%

A.10. ratio of eigenface3 : 2.5%

A.1. ratio of eigenface4 : 2.2%

B. Visualization of Chinese word embedding

- B.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。
- B.2. gensim 的 word2vec, min_count=2 代表字出現的最小次數是 2 次，出現一次的就丟棄。
- B.3. (.5%) 請在 Report 上放上你 visualization 的結果。



- B.4.
- B.5.
- B.6. (.5%) 請討論你從 visualization 的結果觀察到什麼。
- B.7. 左上方好像都是連接詞，中間有一部分都是人稱。

C. Image clustering

- C.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)
- C.2. 1.用 pca 降維到 400，再來用 kmeans 作 cluster，F1=100%
- C.3. 2.用 autoencoder，再用 kmeans 作 cluster，F1=100%
- C.4. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。
- C.5.
- C.6. (.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。
- C.7.

