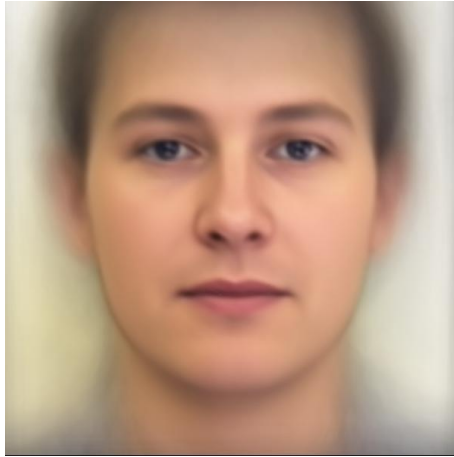


A. PCA of colored faces

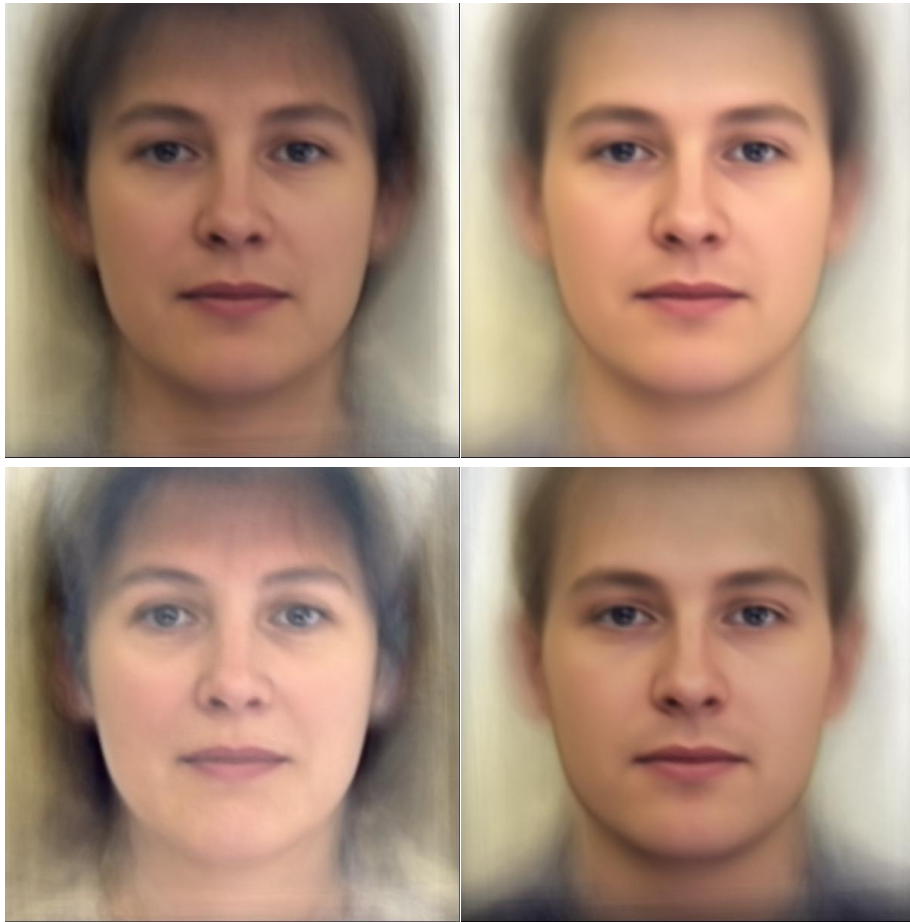
A.1. 請畫出所有臉的平均。



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



- 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。



分別是 0, 3, 4, 6

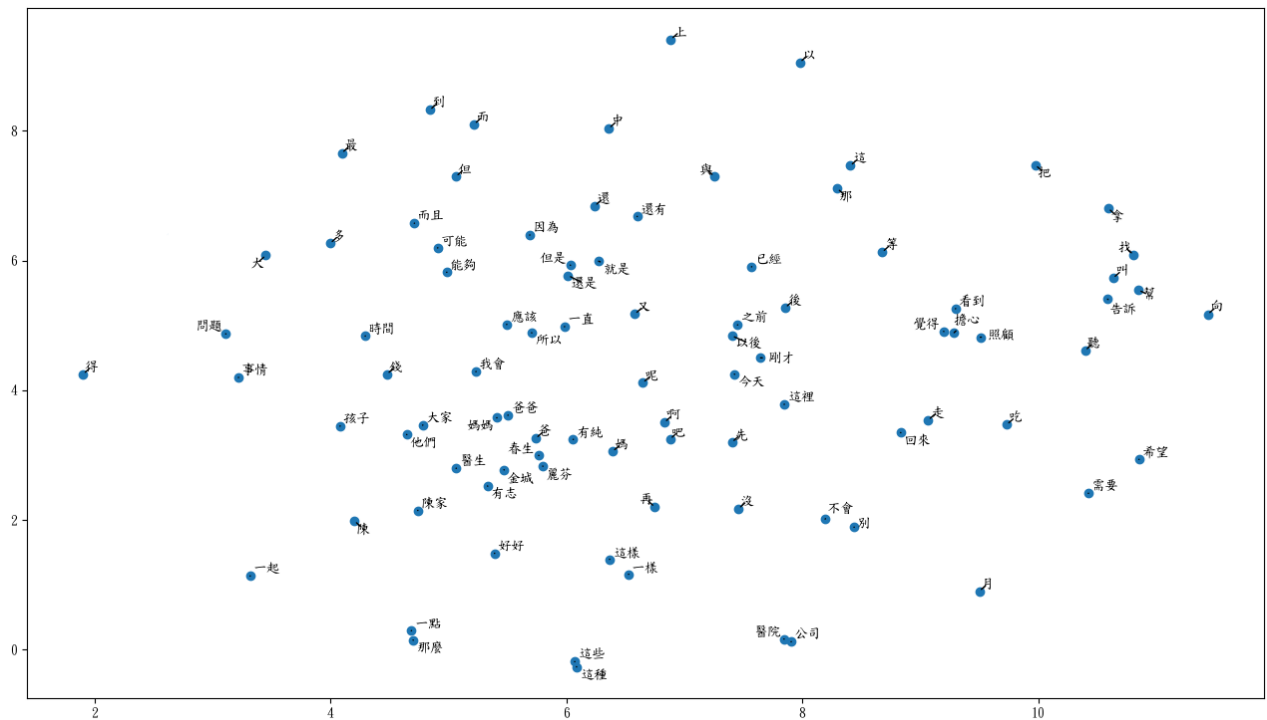
- A.4. 請寫出前四大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入到小數點後一位。
4.1% 2.9% 2.4% 2.2%

B. Visualization of Chinese word embedding

(collaborator:b04901060 黃文璽)

- B.1. 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。
使用 Word2Vec，每個詞用 300 維的向量表示

B.2. 請在 Report 上放上你 visualization 的結果。



B.3. 請討論你從 visualization 的結果觀察到什麼。

請類似的字很近，如:媽 媽媽 爸 爸爸

語尾助詞很近，如:呢 啊 吧

C. Image clustering

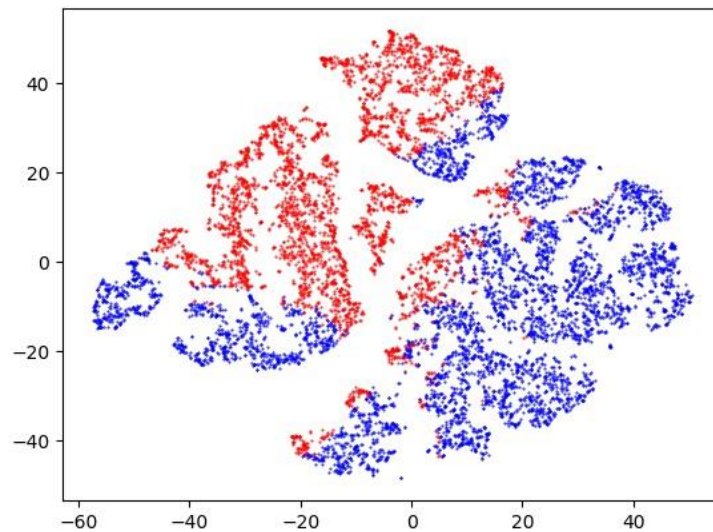
C.1. 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

用 autoencoder 降維成 32 維後使用 K mean

Cluster = 2 : 0.37485

Cluster = 16: 0.94902 (分成 16 類後隨機看圖分成兩類)

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



C.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。

在正確的 `label` 中(下圖)可以看到除了零星的點以外其他基本上都可以藉由 `tsne` 分離兩類

如果是自己預測的(C.2. 圖) 在交界處則有大範圍誤判的情形

