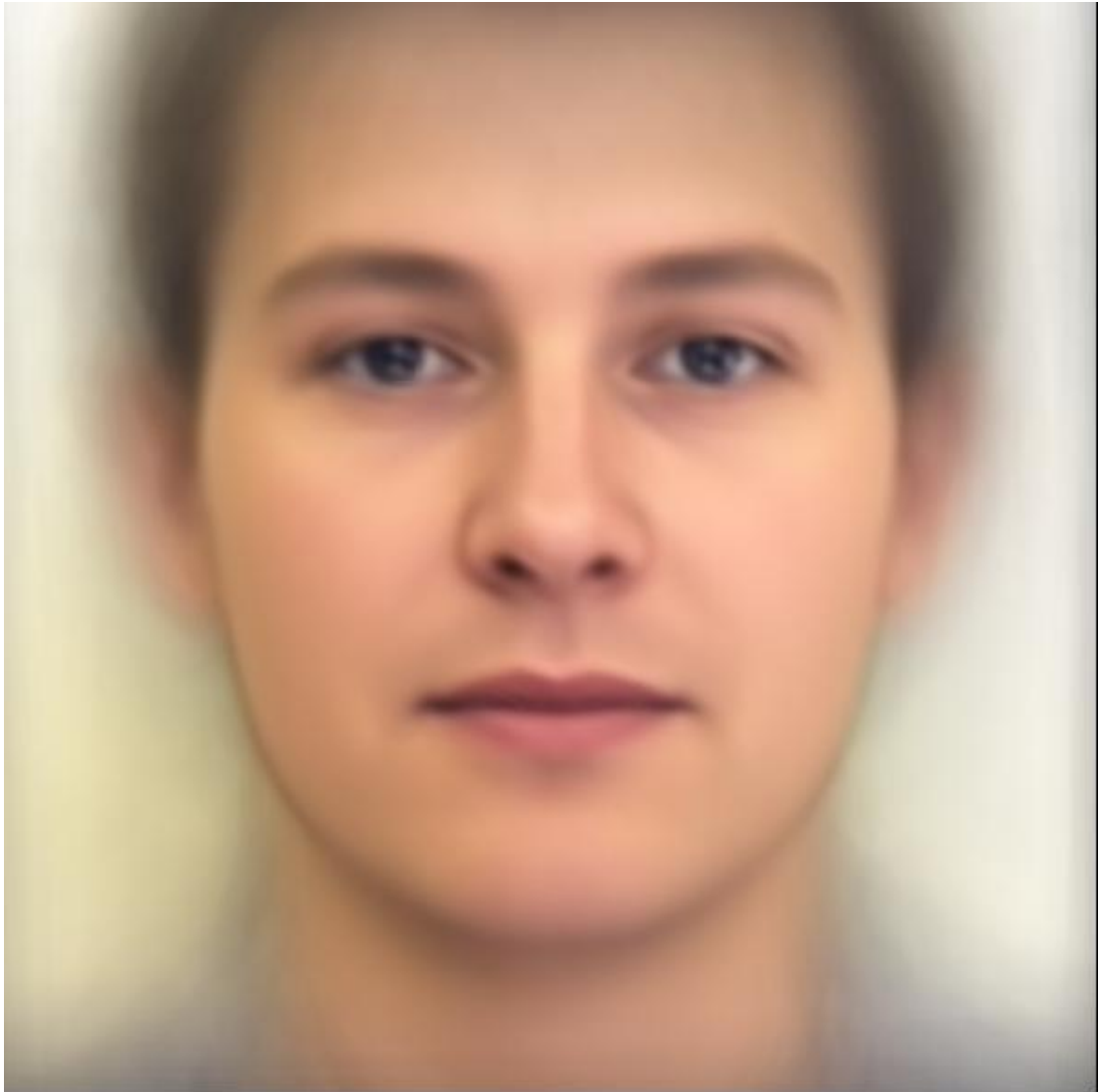


學號：R05943136 系級： 電子碩二 姓名：盧真玄

A. PCA of colored faces

1. (.5%) 請畫出所有臉的平均。



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



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。

Original face:



Reconstruct:



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。

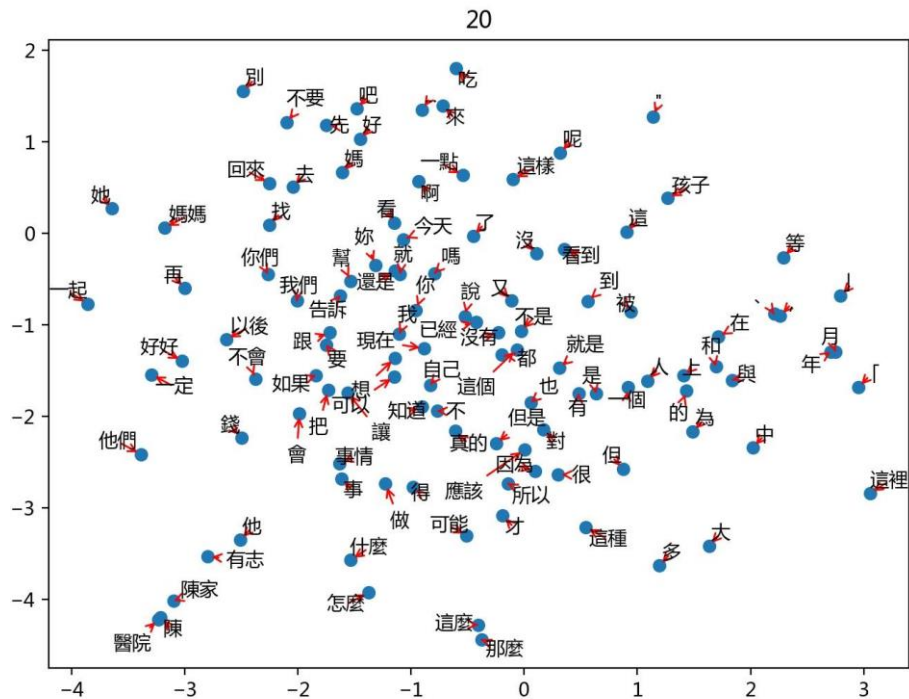
Percentage	4.1%	2.9%	2.4%	2.2%
------------	------	------	------	------

B. 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。

Jieba 和 gensim

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。

從圖中可以看出一些詞的分類，例如左下角把人名放在一起，或者在中間下方把相近的詞意放在一起還有右方把年跟月放在一起。

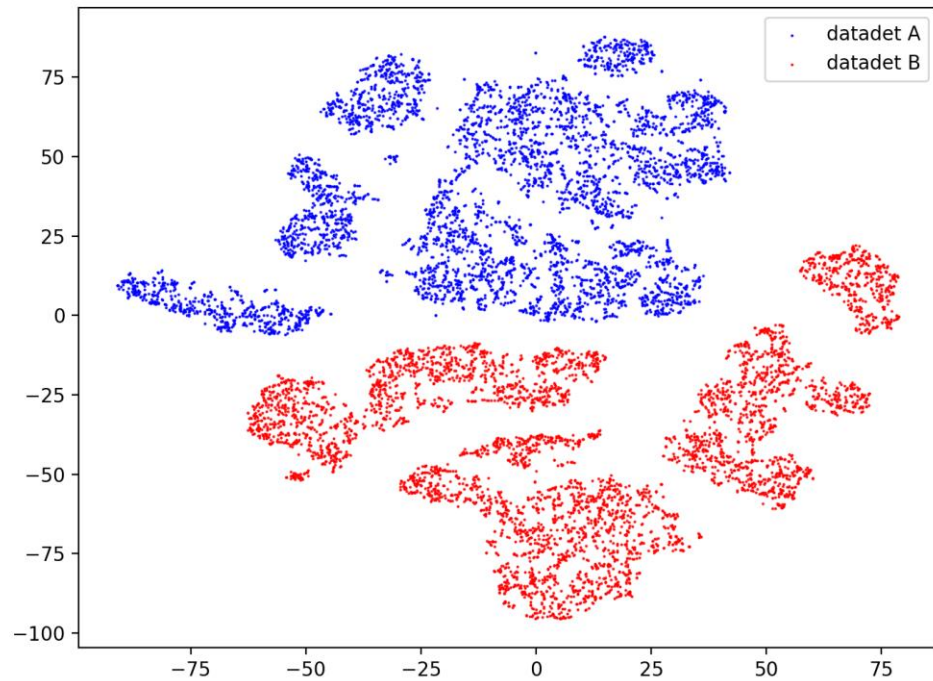
C. Image clustering

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)

Autoencoder 直接降至二維的結果約為 0.03

Autoencoder 降至 16 維再加 PCA 則上升到 0.98~1

2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。



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。從結果看出視覺化的 label 分布和自己預測的 label 幾乎一模一樣。