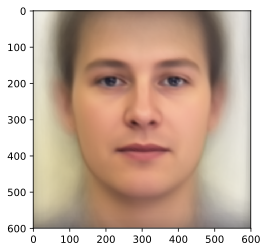
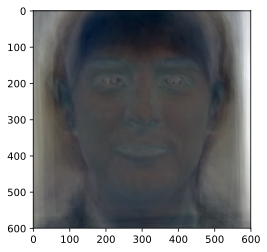
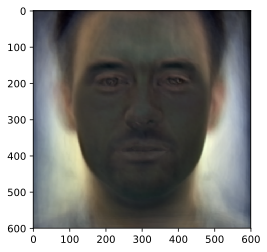
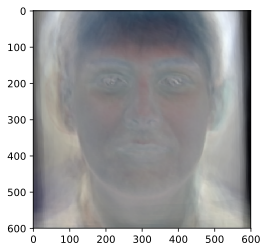
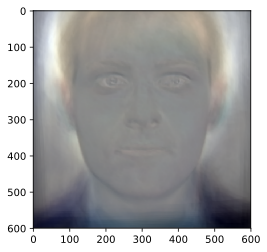
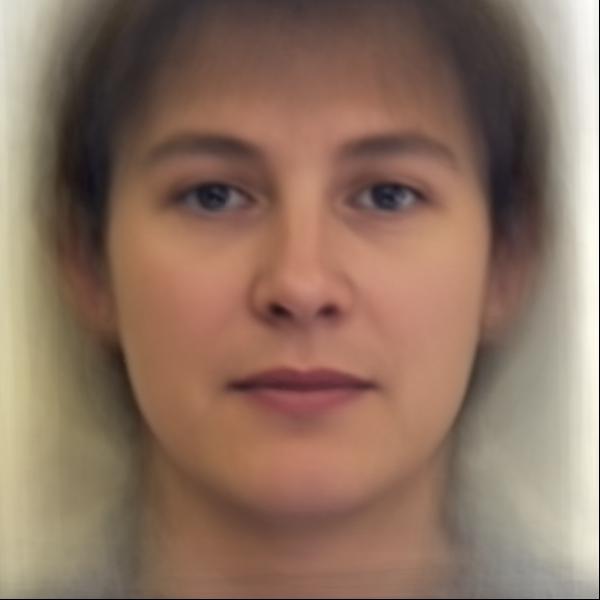
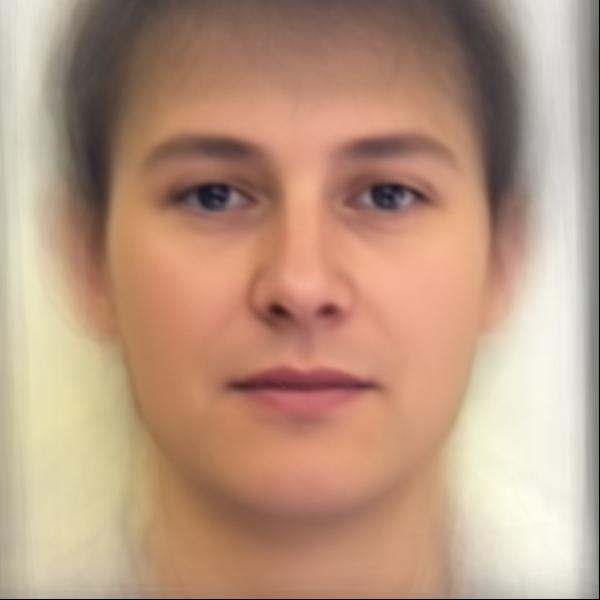
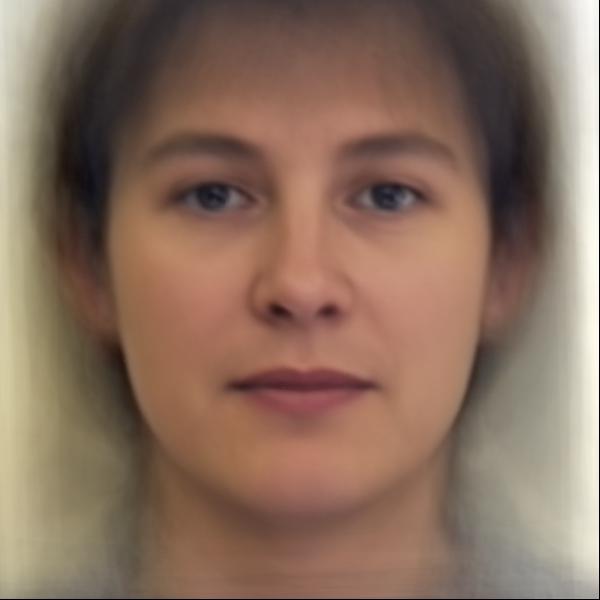
學號：R05546022 系級： 工工所碩二 姓名：謝立成

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。

(.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，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**%

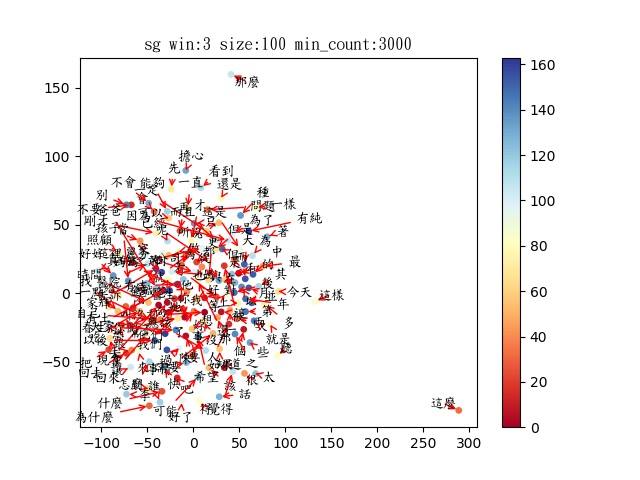
1. Visualization of Chinese word embedding
   1. (.5%) 請說明你用哪一個 word2vec 套件，並針對你有調整的參數說明那個參數的意義。

word2vec使用gensim中的skip-gram model，

number of variable = 100

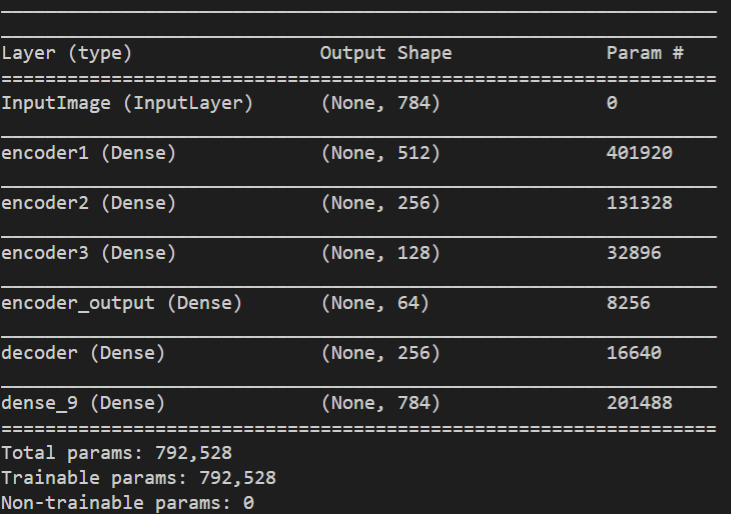
minimum count = 3000 頻率出現3000以上才畫出

window size = 3 移動窗格大小

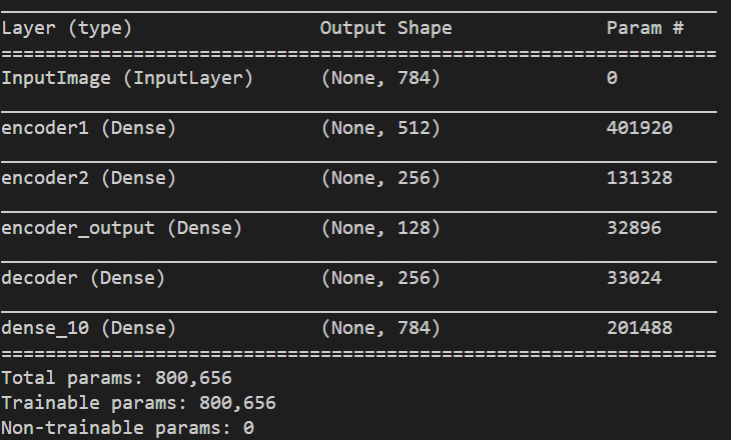
* 1. (.5%) 請在 Report 上放上你 visualization 的結果。
  2. (.5%) 請討論你從 visualization 的結果觀察到什麼。

說爸爸、照顧、孩子會聚在一起，但是圖畫起來有點亂。

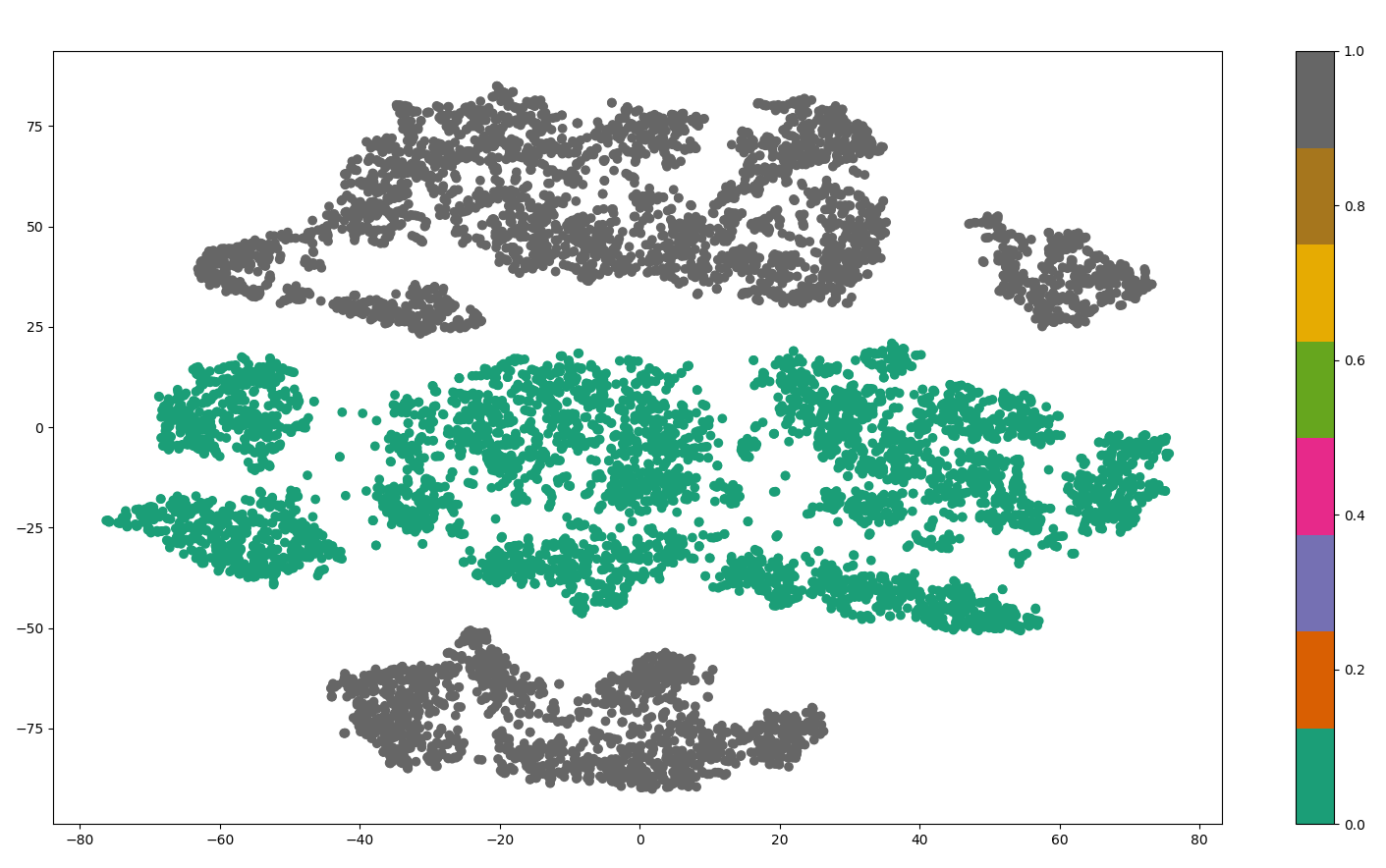
1. Image clustering
   1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

Autoencoder with encoding layer = 64 : f-score = 0.99879

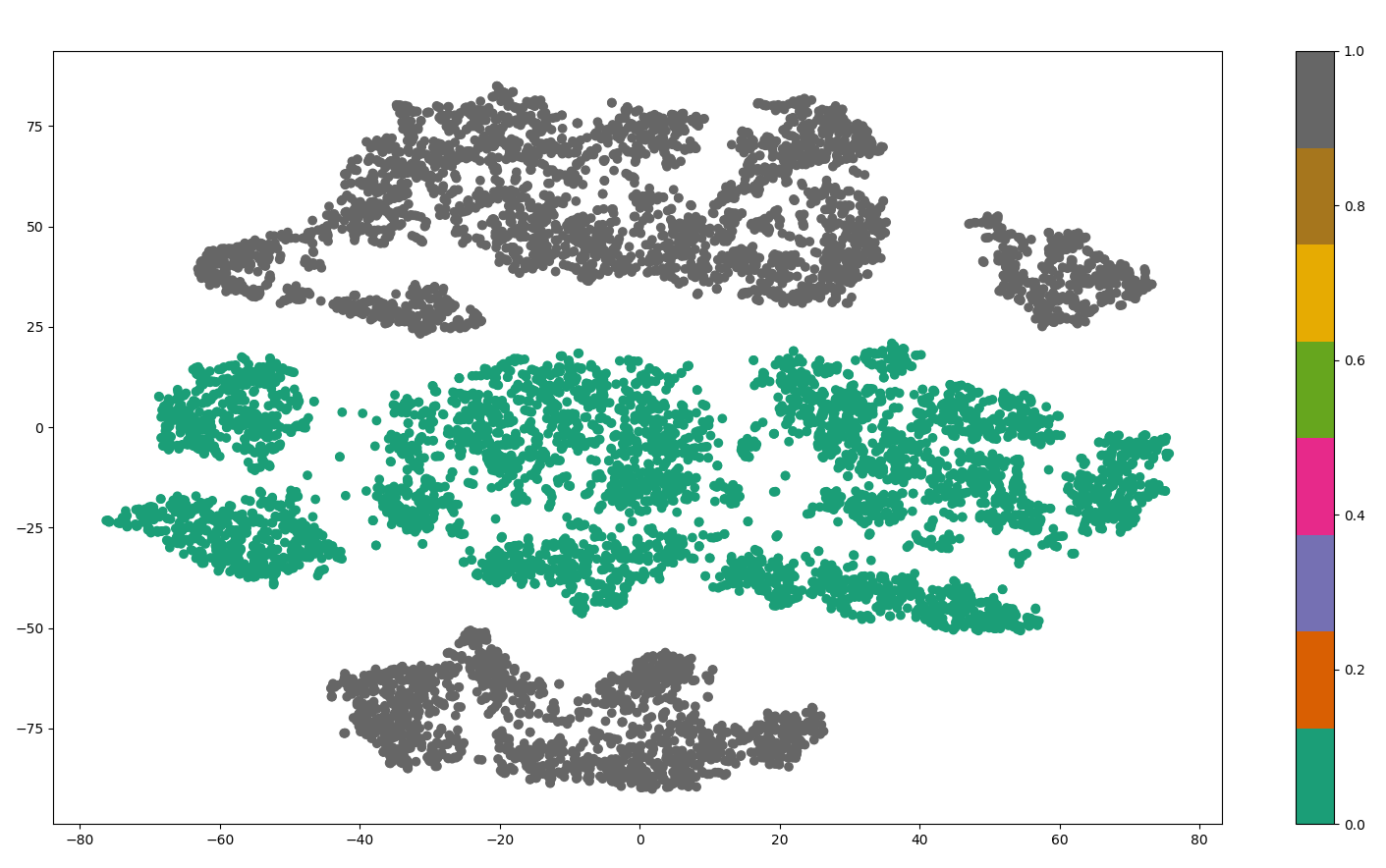
Autoencoder with encoding layer = 128 : f-score = 0.80193



* 1. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。



* 1. (.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。



可以發現，預測之結果非常準確，兩張圖看起來近乎相同。