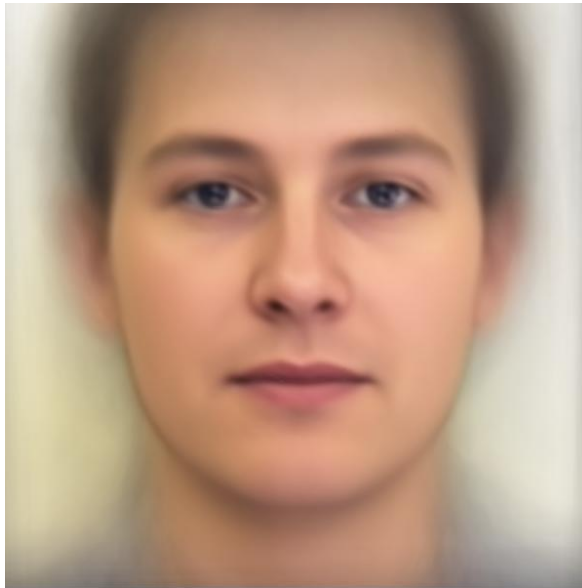


## A. PCA of colored faces

A.1. (.5%) 請畫出所有臉的平均。



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



A.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。

Select = [8, 14, 19, 93]



A.4. (.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。

4.1%, 2.9%, 2.4%, 2.2%

## B. Visualization of Chinese word embedding

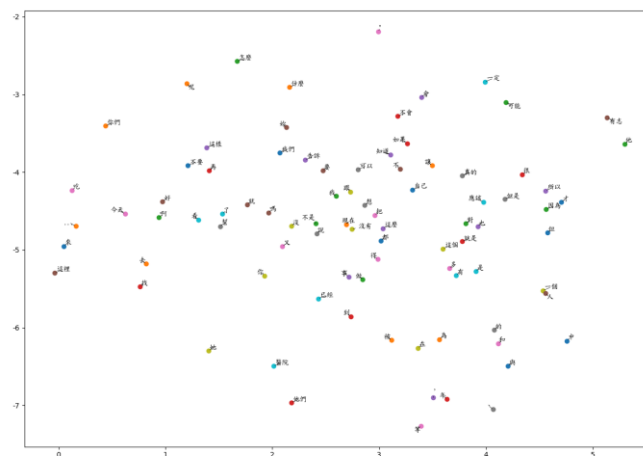
B.1. (.5%) 請說明你用哪一個 word2vec 套件，並針對你有調整的參數說明那個參數的意義。

word2vec 套件，word\_dim = 555, min\_count = 5555

word\_dim 負責計算代表一個詞的 dim

min\_count 負責計算重複幾次我才把這個字納入我的字典庫裡

B.2. (.5%) 請在 Report 上放上你 visualization 的結果。



B.3. (.5%) 請討論你從 visualization 的結果觀察到什麼。

我的 min\_count = 5555，我覺得可以在小一點，不然只存到一些慣用語，例如：我、我們、等，所以之後降到二維之後，其實分的沒有太好

## 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)

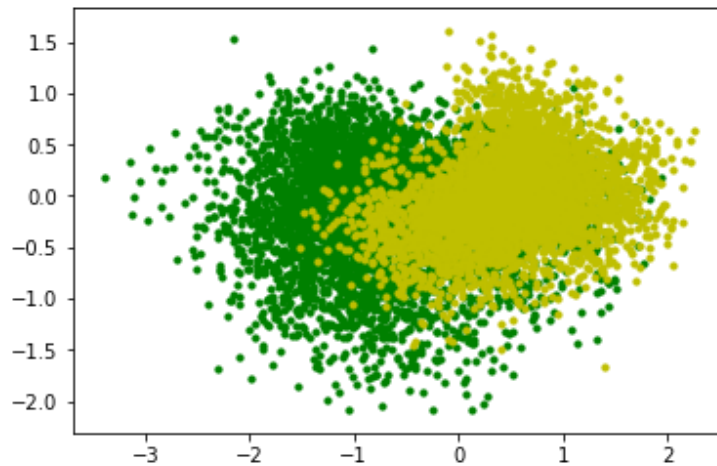
我去做 DNN 架構下，中間層的維度不同去比較差異

中間那層，我使用 128 維去做解釋，Score = 0.99631

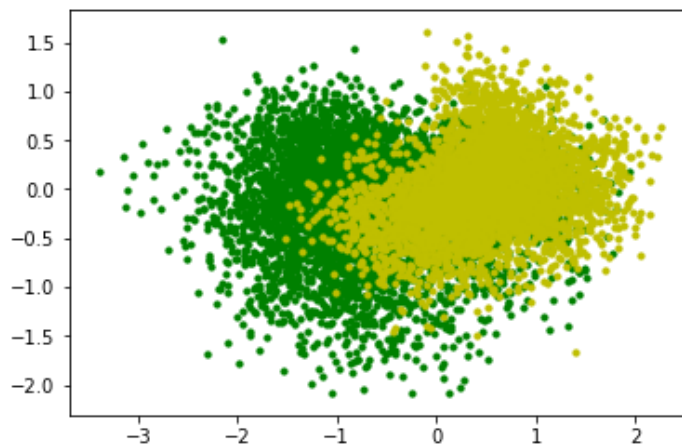
中間那層，我使用 256 維去做解釋，Score = 0.06477

所以用更少的維度去解釋，比較好!!

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



C.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。



幾乎一模一樣! 我取 feature 的前 2 維/128 維