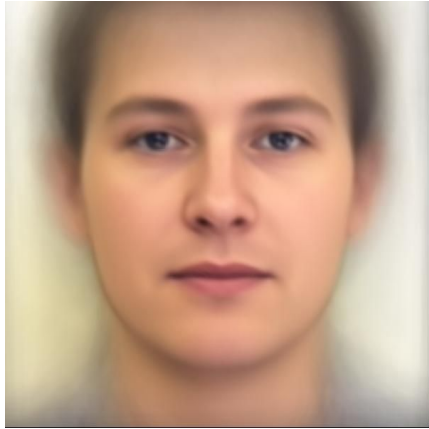


Machine Learning HW7 Report

學號：B05901005 系級：電機三 姓名：賴沂謙

1. PCA of color faces:

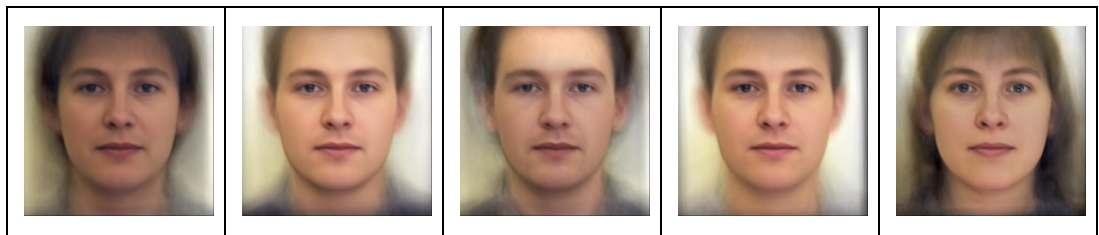
a. 請畫出所有臉的平均。



b. 請畫出前五個Eigenfaces，也就是對應到前五大Eigenvalues的Eigenvectors。



c. 請從數據集中挑出任意五張圖片，並用前五大 Eigenfaces 進行 reconstruction，並畫出結果。



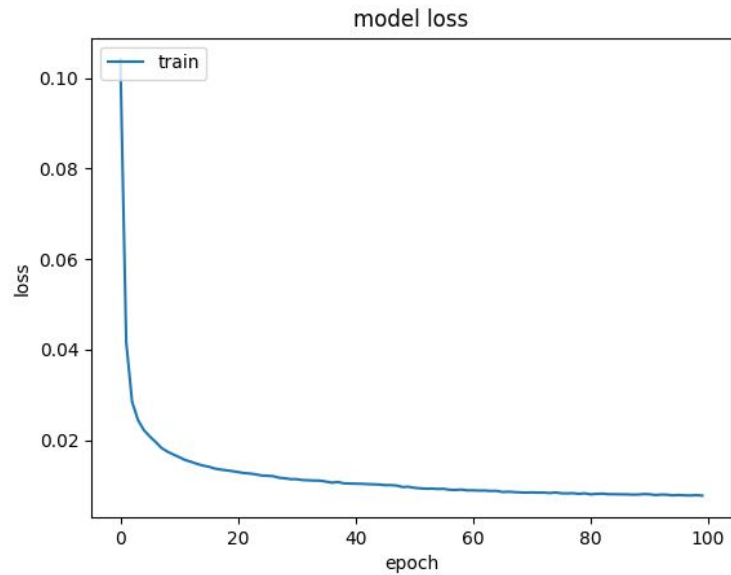
d. 請寫出前五大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入到小數點後一位。

4.1	2.9	2.4	2.2	2.1
-----	-----	-----	-----	-----

2. Image clustering:

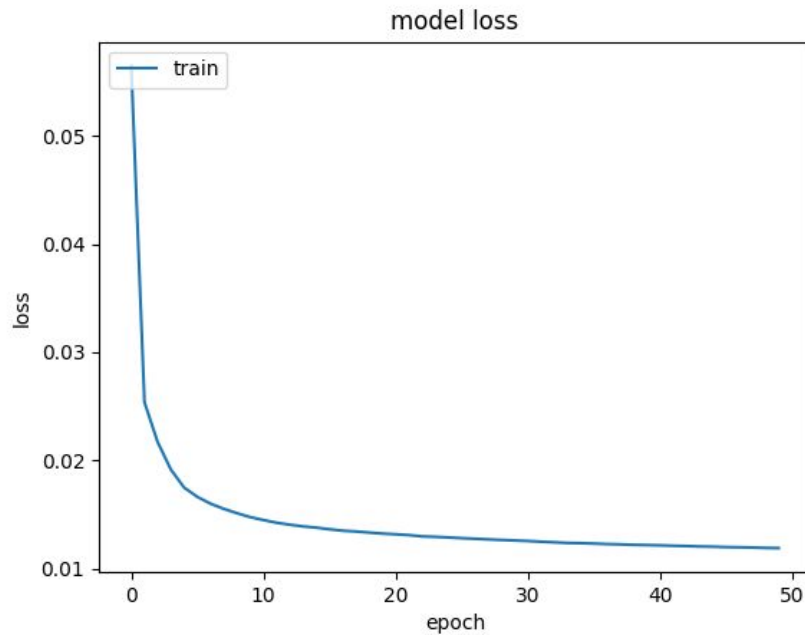
- a. 請實作兩種不同的方法，並比較其結果(reconstruction loss, accuracy)。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

方法一：用dnn架構的autoencoder + pca



結果：accuracy (public) = 0.96597, loss如上圖

方法二：用cnn架構的autoencoder + tSNE



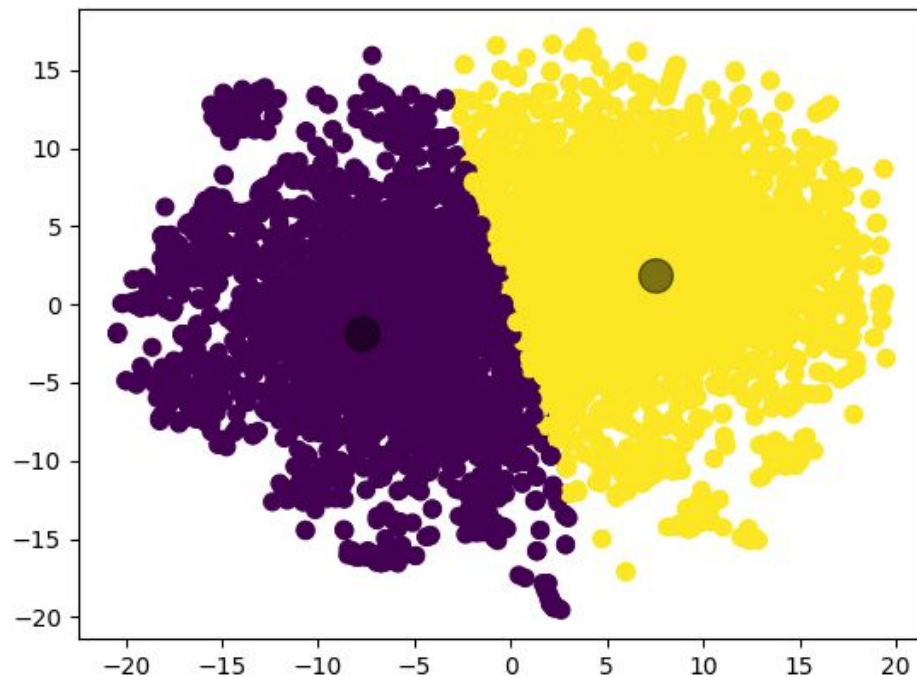
結果：accuracy (public) = 0.92780, loss如上圖

- b. 預測 visualization.npy 中的 label, 在二維平面上視覺化 label 的分佈。
(用 PCA, t-SNE 等工具把你抽出來的 feature 投影到二維, 或簡單的取前兩維 2 的 feature)

其中visualization.npy 中前 2500 個 images 來自 dataset A, 後 2500 個 images 來自 dataset B, 比較和自己預測的 label 之間有何不同。

用我的model降維之後, 經過t-SNE再畫出來如下圖。

label預測的正確率為94.14%



- c. 請介紹你的model架構(encoder, decoder, loss function...), 並選出任意32張圖片, 比較原圖片以及用decoder reconstruct的結果。

Autoencoder架構 : (loss function=MSE)

```

input_dim = 3072
# Encoder Layers
autoencoder.add(Dense(3072, input_shape=(input_dim,), activation='relu'))
autoencoder.add(Dense(1024, activation='relu'))
autoencoder.add(Dense(512, activation='relu'))
autoencoder.add(Dense(256, activation='relu'))

# Decoder Layers
autoencoder.add(Dense(512, activation='relu'))
autoencoder.add(Dense(1024, activation='relu'))
autoencoder.add(Dense(3072, activation='relu'))
autoencoder.add(Dense(input_dim, activation='sigmoid'))

autoencoder.compile(optimizer='adam', loss='mse')
history = autoencoder.fit(train_data, train_data, epochs=100, batch_size=256)

```

Original:



Reconstruct:

