

Machine Learning HW7 Report

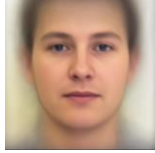
學號：B05203050

系級：化學三

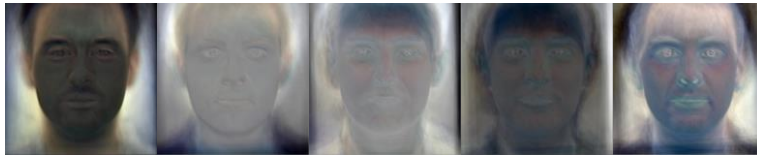
姓名：陳品翰

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。

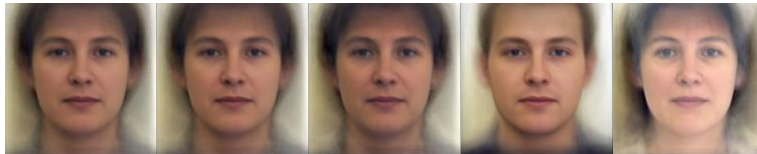


- c. 請從數據集中挑出任意五張圖片，並用前五大 Eigenfaces 進行 reconstruction，並畫出結果。

原圖：



Reconstructed:

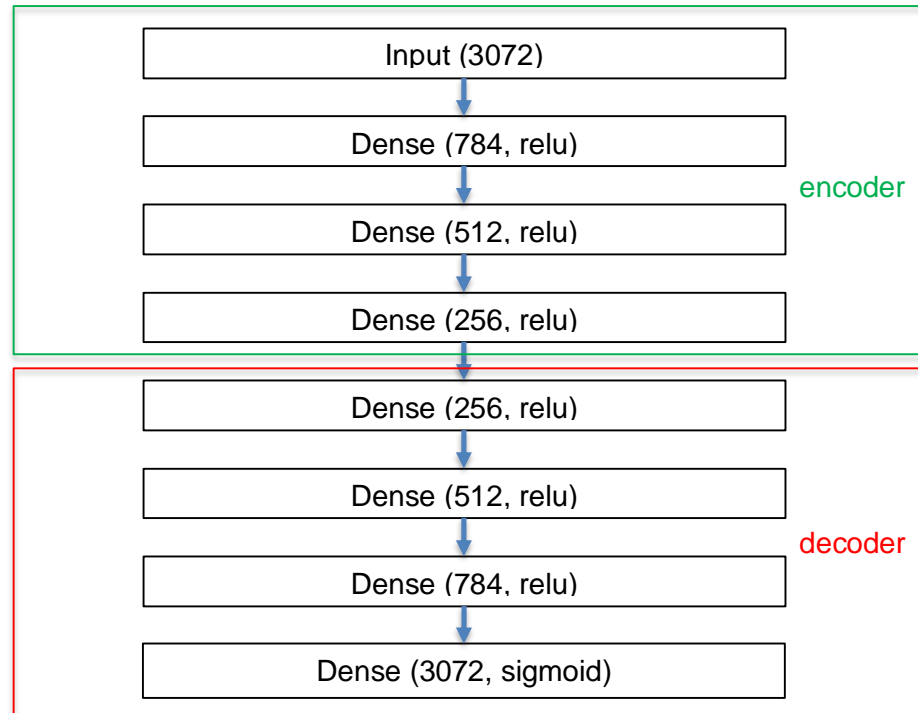


- d. 請寫出前五大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入到小數點後一位。

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)
分別使用 CNN、DNN 作為 autoencoder，CNN 架構見 2-c，DNN 架構：



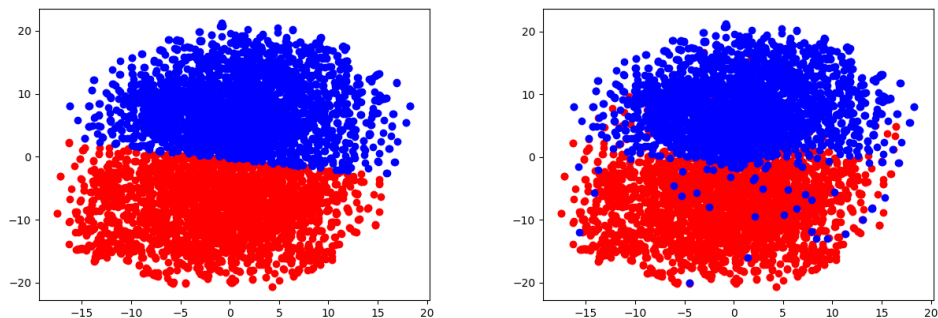
並同樣以 PCA、TSNE 降維，以 Kmeans 做 clustering，得到結果如下：

	reconstruction loss (validation)	accuracy
CNN	0.545	0.95259
DNN	0.595	0.87534

可見使用 CNN 做為 autoencoder 效果較佳。

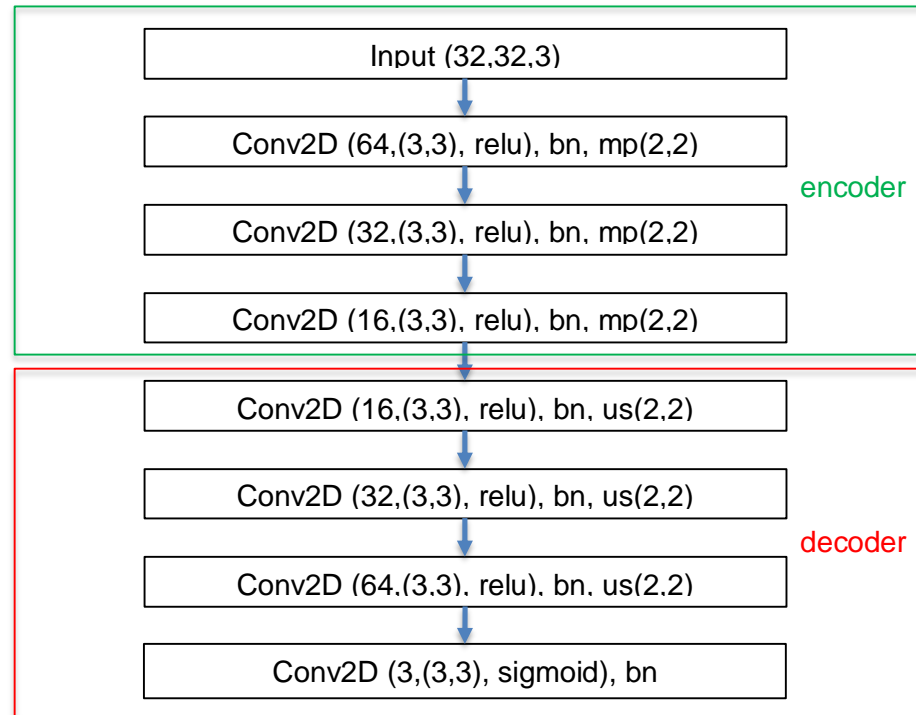
- b. 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。
(用 PCA, t-SNE 等工具把你抽出來的 feature 投影到二維，或簡單的取前兩維 2 的 feature)

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。



由 PCA, TSNE 降至二維後繪製於平面上，左為 Kmeans 預測之 label，右為真實 label，可見 clustering 效果很好。誤差應來自降維後，不同 label 之數據相近，因此可見少數藍點混在紅點中，導致判斷錯誤。

- 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。

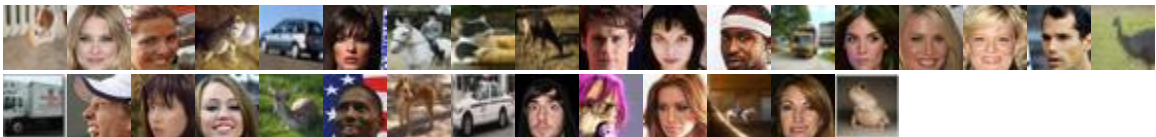


(bn: batchnormalization, mp: maxpooling2D, us: upsampling2D)

Optimizer: adadelta, Loss function: sigmoid

Encode 後的維度：(16,4,4)，flatten 成 (256) 後，經由 PCA ($n_components = 2$) 以及 TSNE ($n_components = 2$) 降維後，以 Kmeans ($n_clusters = 2$) 做 clustering。

原圖：



Reconstructed:

