

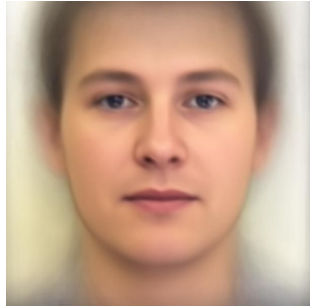
# Machine Learning HW7 Report

學號：b05901063 系級：電機三

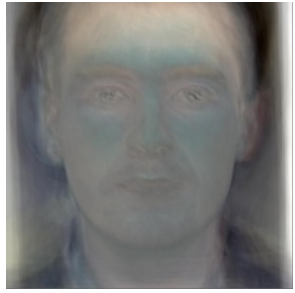
姓名：黃世丞

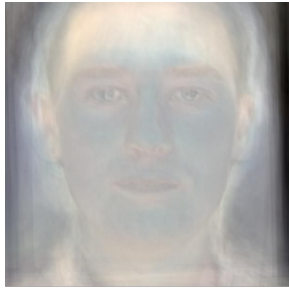

## 1. PCA of color faces:

a. 請畫出所有臉的平均。

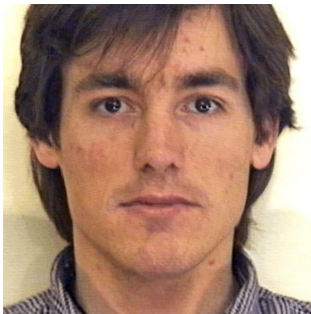
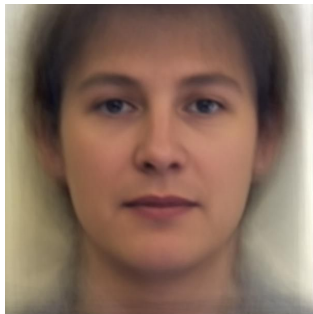
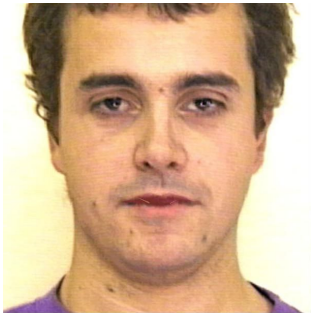
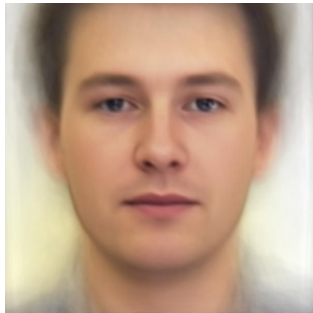



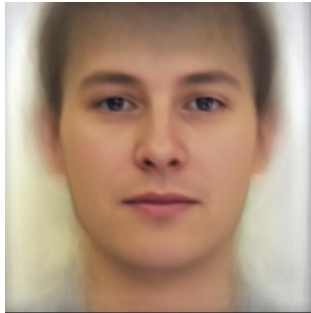

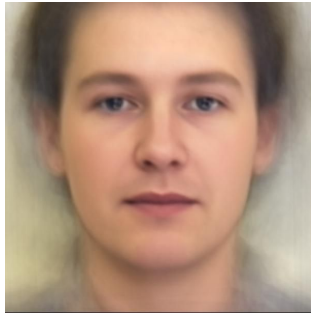

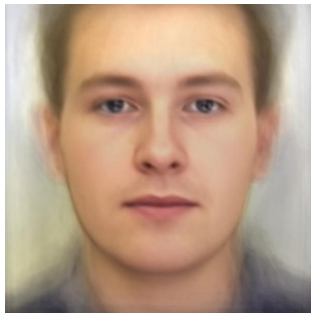
b. 請畫出前五個 Eigenfaces, 也就是對應到前五大 Eigenvalues 的 Eigenvectors。

特徵值編號(從大到小)	特徵臉
0	
1	
2	

3	
4	

- 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。

圖片編號	原圖	重構圖片
1		
10		

22		
37		
72		

- 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。

特徵值編號	0	1	2	3	4
百分比(%)	4.1	2.9	2.4	2.2	2.1

## 2. Image clustering:

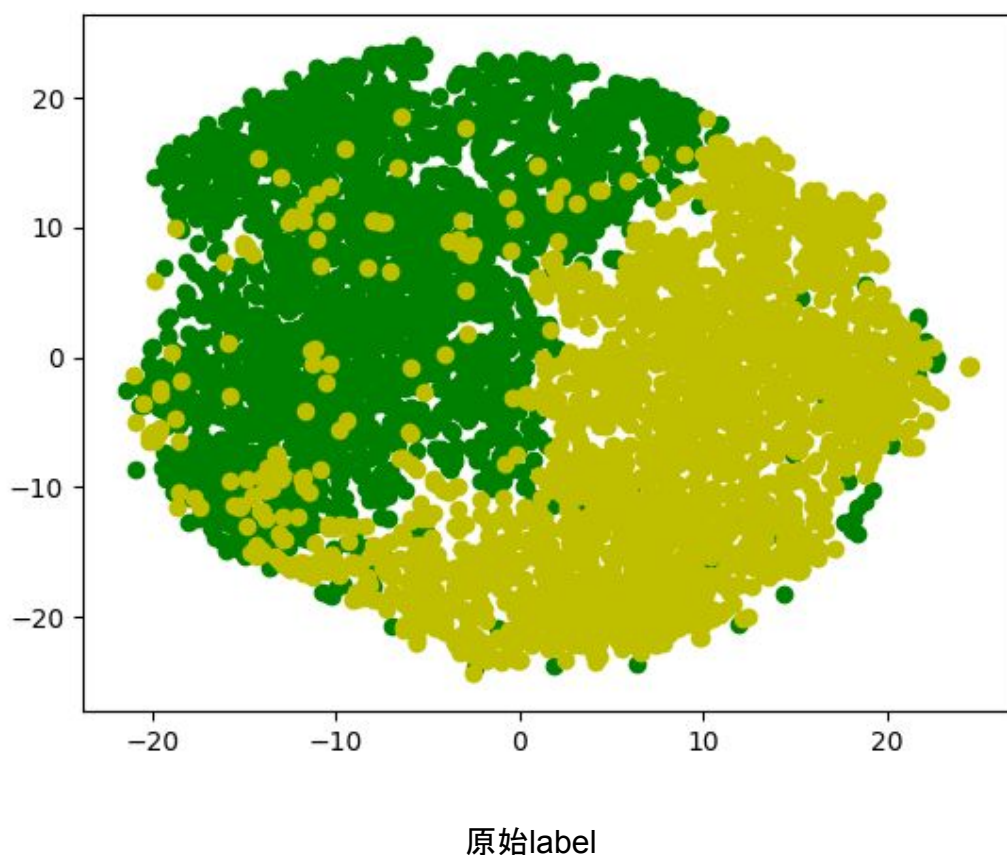
- a. 請實作兩種不同的方法，並比較其結果(reconstruction loss, accuracy)。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

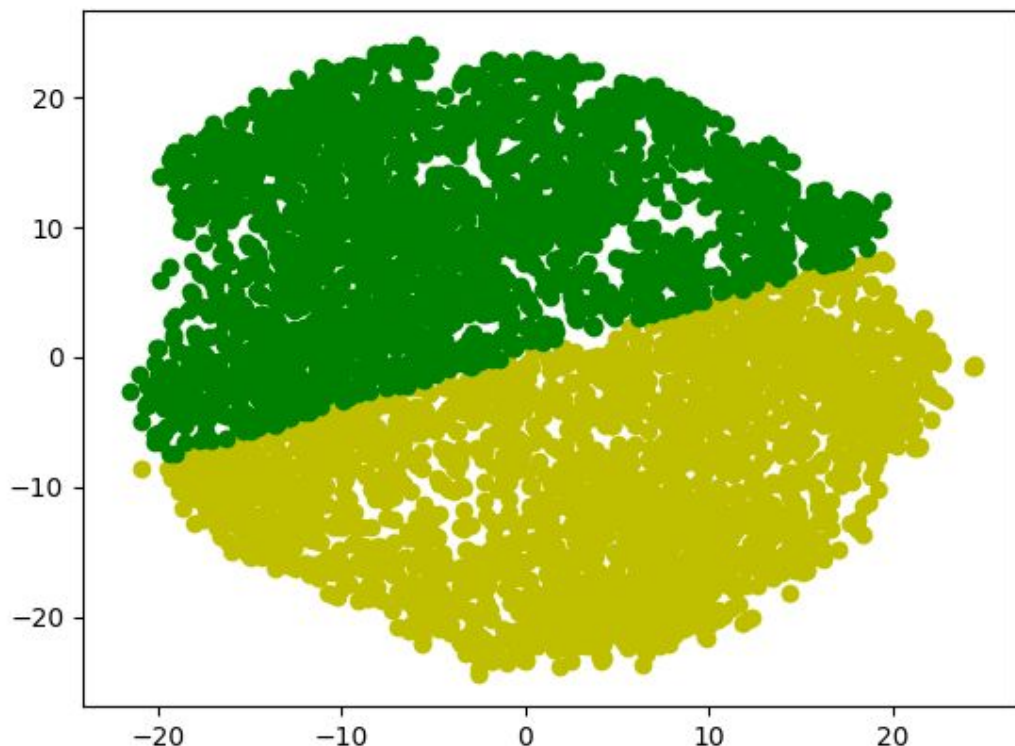
	private	public
t-SNE	0.98651	0.98644
PCA	0.87003	0.87031

t-SNE和PCA後都做Kmeans，t-SNE的效果明顯比PCA好得多。

- 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。

原始的label散佈圖上可以看到大致上2個label會分開成兩個大區塊，但也有一些零散的點是散佈在對面的區塊裡。  
預測的label是用Kmeans切成兩塊，大致上和實際label的趨勢符合，但很顯然那些零散分佈在對面區塊的圖片會被分入錯誤的類別。

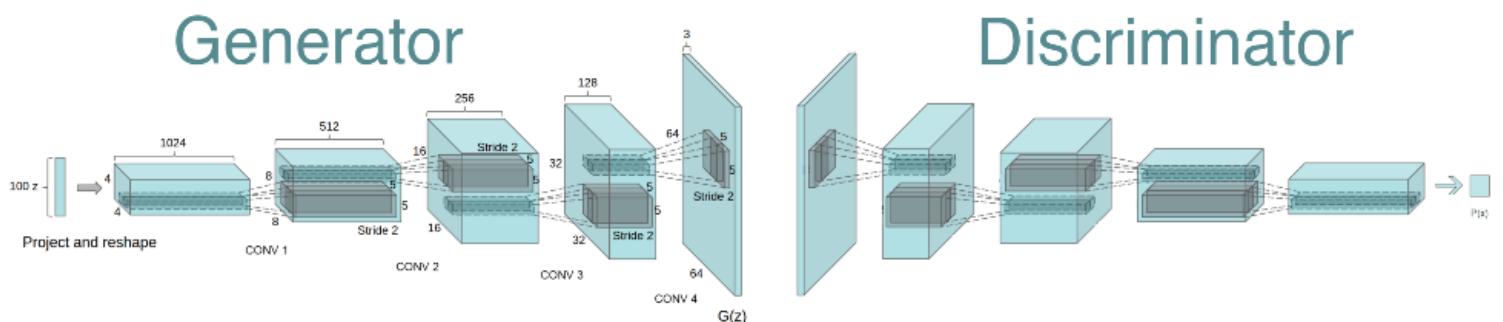




Kmeans預測的label

- 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。

仿造DCGAN的discriminator和generator, decoder仿造generator的結構, encoder仿造discriminator的結構, 但把discriminator最後一層的dense輸出改成encoded\_dim(128)。DCGAN結構如附圖：



Loss function使用mean square error, 並且訓練過程加入augmentation

原圖	重構	原圖	重構	原圖	重構	原圖	重構
