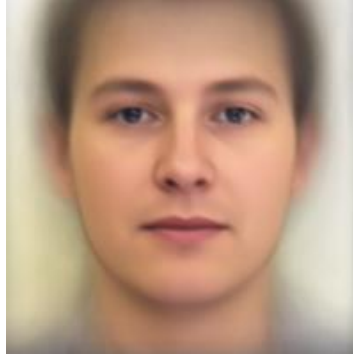


#### HW4

學號: b04901009 系級: 電機三 姓名: 林孟瑾

##### 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。

由左往右各為第 0 個, 第 1 個, 第 2 個, 第 3 個 eigenface。



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。

由左往右各為 0. jpg , 12. jpg, 234. jpg , 356. jpg 的 reconstruction



A. 4. (.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。

第 0 個 eigenface	第 1 個 eigenface	第 2 個 eigenface	第 3 個 eigenface
4.1%	3.0%	2.4%	2.2%

##### B. Image clustering

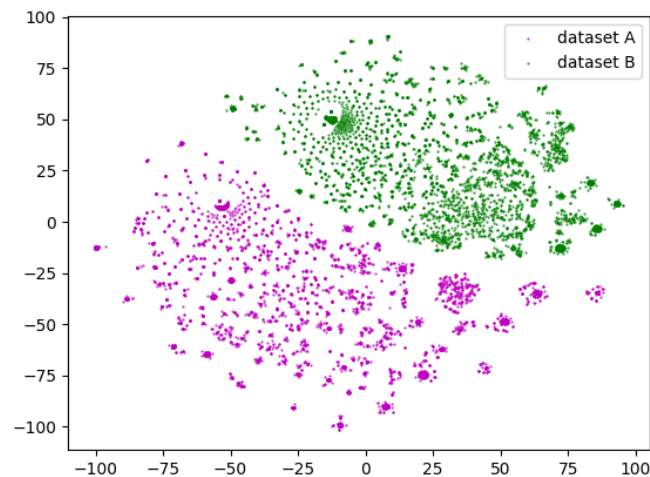
B. 1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的

降維方法或不同的 cluster 方法都可以算是不同的方法)

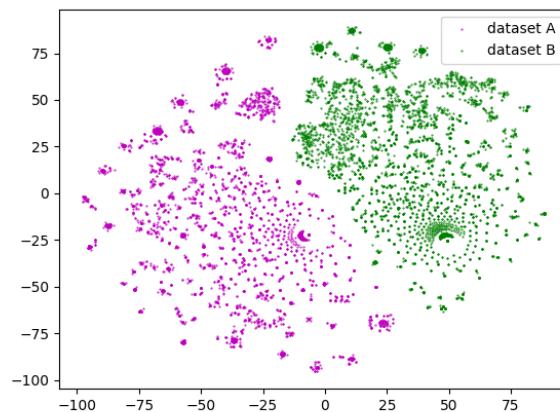
	Private score	Public score
Autoencoder+kmeans	0.99890	0.99896
Autoencoder+ Cosine similarity	0.81572	0.81637

使用不同的 cluster 方法之後，由 private score 和 public score 可以發現，Cosine similarity 的結果準確度較差，因為 threshold 的選取可能會影響準確度，且 kmeans 的 cluster 效果較佳。

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



B. 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。



基本上，和自己預測的 label 幾乎差不多，表示 model 準確率還不錯。兩張圖中，點

分佈的情形有點不太一樣。

### C. Ensemble learning

- C.1. (1.5%) 請在 hw1/hw2/hw3 的 task 上擇一實作 ensemble learning，請比較其與未使用 ensemble method 的模型在 public/private score 的表現並詳細說明你實作的方法。（所有跟 ensemble learning 有關的方法都可以，不需要像 hw3 的要求硬塞到同一個 model 中）

我使用 hw3 實作 ensemble learning，方法為 voting。事先 train 好五個 model，predict 的時候，如果多數的 model 都認為某一張照片屬於某一類，那就將此一 predict 的結果存起來（也就是所有 model predict 某一圖片出來的情況中，看哪一 label 最多，就將此圖片歸為那一類 label）。如此一來，便能提升預測的準確率，也能避免某一 model overfit 的情況。由 private score 和 public score 可以看出，有 ensemble 後 score 明顯提升，且 private 和 public 之間相差不大，表示 predict 結果較不會有 overfit 的情況。而 validation accuracy 的部分可以看出，有 ensemble 的 validation loss 較低，準確率較高。

	private	public
有 ensemble	0.71273	0.71440
沒 ensemble	0.68960	0.70075

	validation accuracy
有 ensemble	0.9725521613431328
沒 ensemble	0.8470166149987809