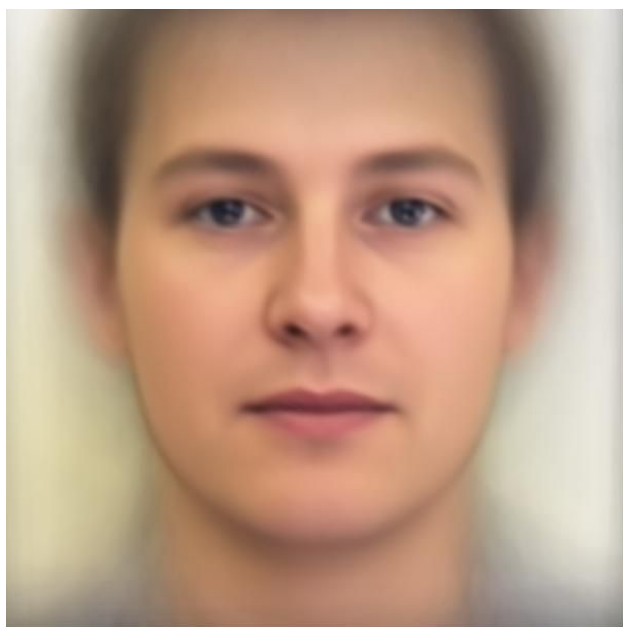


A. PCA of colored faces

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。



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。

First eigenface	Second eigenface	Third eigenface	Fourth eigenface
4.1%	2.9%	2.4%	2.2%

B. Image clustering

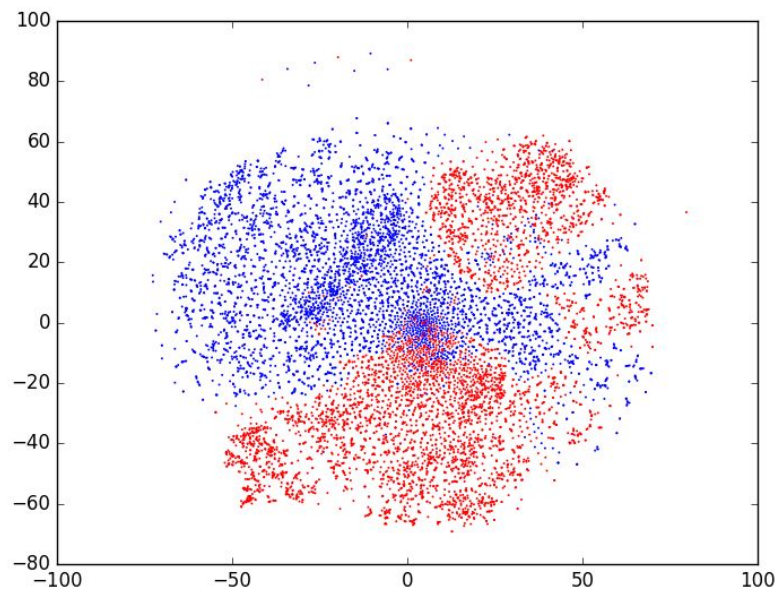
B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

autoencoder	PCA
Avg acc : 0.98967	Avg acc : 0.99998

這部份分別使用了PCA以及autoencoder來做降維的動作，cluster方法則都是使用Kmeans；autoencoder使用了三層encoder和三層decoder，由原本的784-dim，降到128, 64, 32-dim再慢慢的decode回來，運用多層的

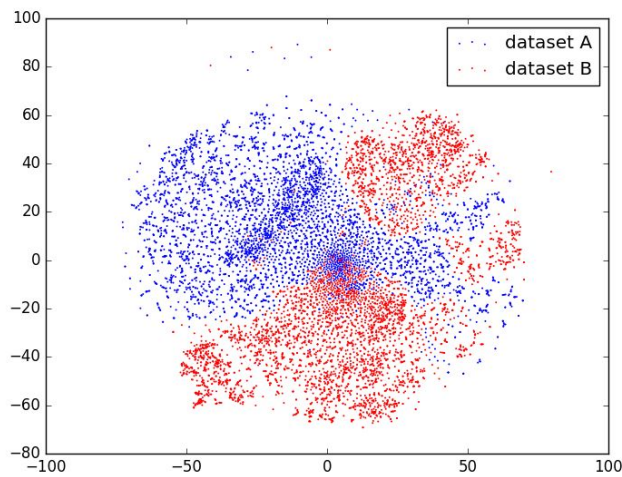
encoder可以有較多的參數，若是只使用單層，則acc會大幅下降（avg acc 0.50064）；PCA則是principle components設為500有較好結果，若調低或是增加principle components的個數都會影響acc，雖然PCA在降到比較高dim有好表現，但如果是比較低的dim則多層encoder有較好表現。

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

經過比較後，發現預測的結果和真實的答案有非常相似的結果，而在 visualize 裡一樣是透過PCA的方式去做降維，再用Kmeans做clustering，在 kaggle private 上也有接近1的acc，所以預測結果和答案相似是可以預期的。



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)

(因為hw3 kaggle無法做late submission, 所以比較val_loss)

Reference:

<https://towardsdatascience.com/ensembling-convnets-using-keras-237d429157eb>

Without ensemble	0.28789
With ensemble	0.24177

這次的ensemble實作是透過將三個model的output透過Average()取merge layer, 而這三個model是以hw3為基礎, 改變參數數量以及架構的結果, 那經過ensemble之後也可以發現確實有比單一model的val_loss有所降低。