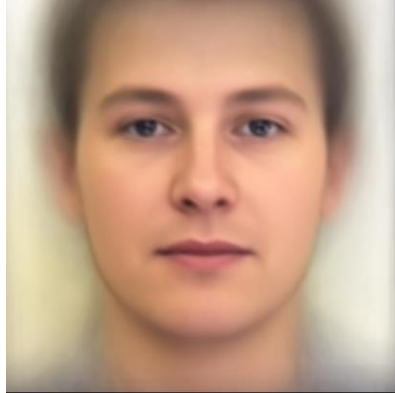


A. PCA of colored faces

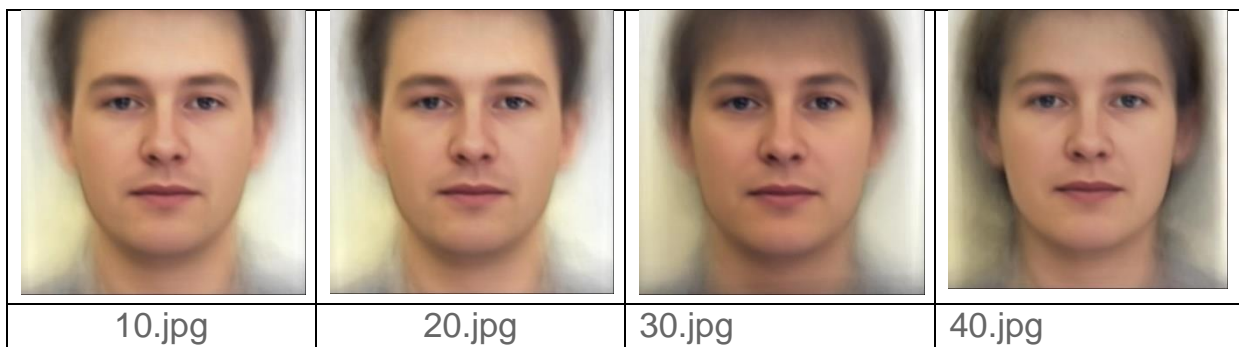
A.1. (.5%) 請畫出所有臉的平均。



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 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。

第一大:4.1%

第二大:3.0%

第三大:2.4%

第四大:2.2%

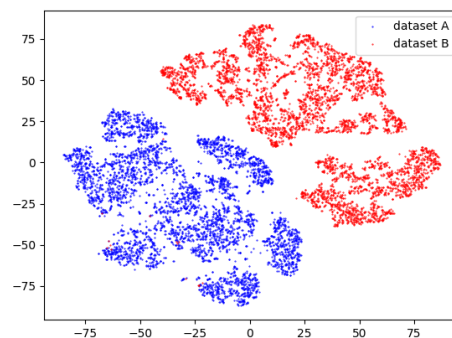
B. Image clustering

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)

Method	autoencoder	PCA
Accuracy(kaggle)	0.99447	1.0

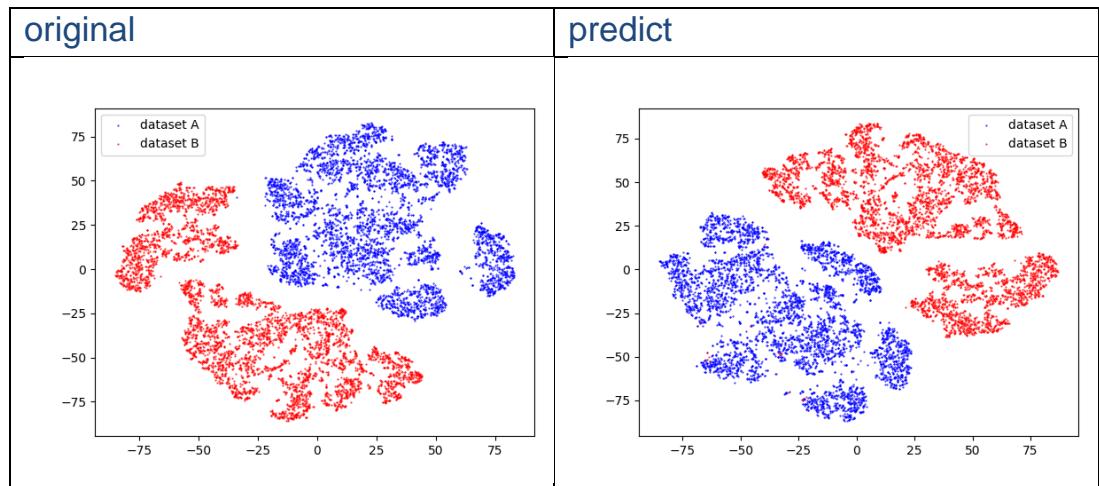
Autoencoder 跟 PCA 最大的不同為 autoencoder 找出來的基底不是線性獨立的,所以不一定能 **minimize reconstruction error**

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。



基本上我覺得分的都還蠻正確的,這題我原本是用 **pca** 先降到 400 維,再用 **tsne** 降到 2 維做 **visualize**,但是發現這樣會有很多不同分類的點混在一起,於是我改用 **autoencoder** 的方式,從 32 維降到 2 維,這樣再 **visualize** 的效果會比較好,雖然由圖中還是可以看出有些 **predict** 後應該是藍色的點沒有被分好,但是已經達到還不錯的分類效果。

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)

我的想法很簡單,就是拿作業三 **kaggle** 上成績最好的 3 個 **model** 做 **voting**,其中 3 個 **model** 和經過 **voting** 之後的 **model** 在 **kaggle** 上的表現如下

	Public	Private
Model1	0.66341	0.68013
Mdoel2	0.69183	0.67679

Model3	0.69657	0.68793
Voting result	0.70911	0.70576

這個表現讓我大為吃驚,因為我在原本的 CNN model 中視過了各種方法,acc 都沒有辦法超過 0.69,但是現在選了 3 個最好的 model 做 voting,降讓準確率提升了 1%!!!

我 voting 的方式是如果今天三個 model 都預測到不同的數字的話,就用原本表現最好的 model predict 出來的結果,如果 2 個一樣,另一個不一樣的話,就採多數決,選用兩個一樣的當作 predict 出來的結果。