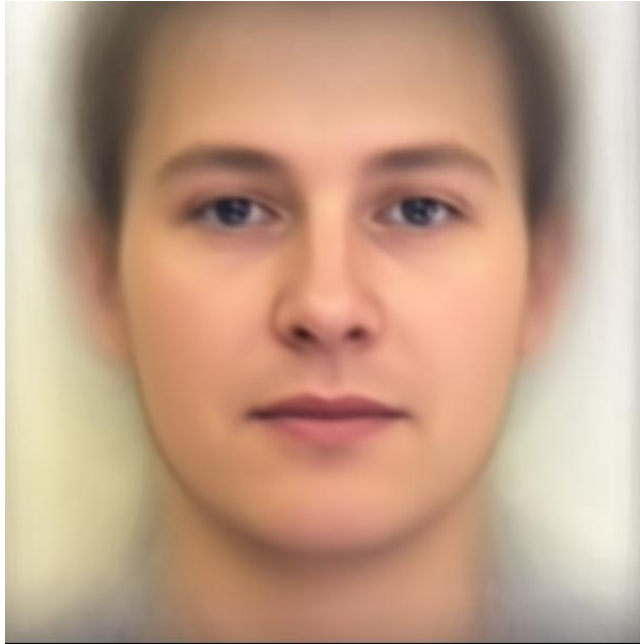


PCA of colored faces

1.



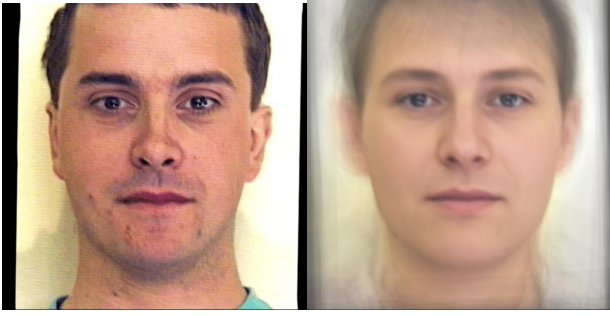
2.



3.



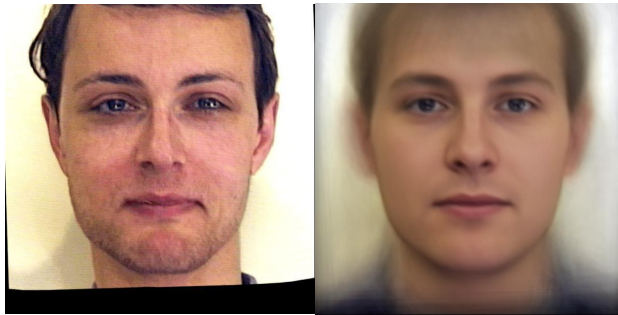
id = 15



id = 178



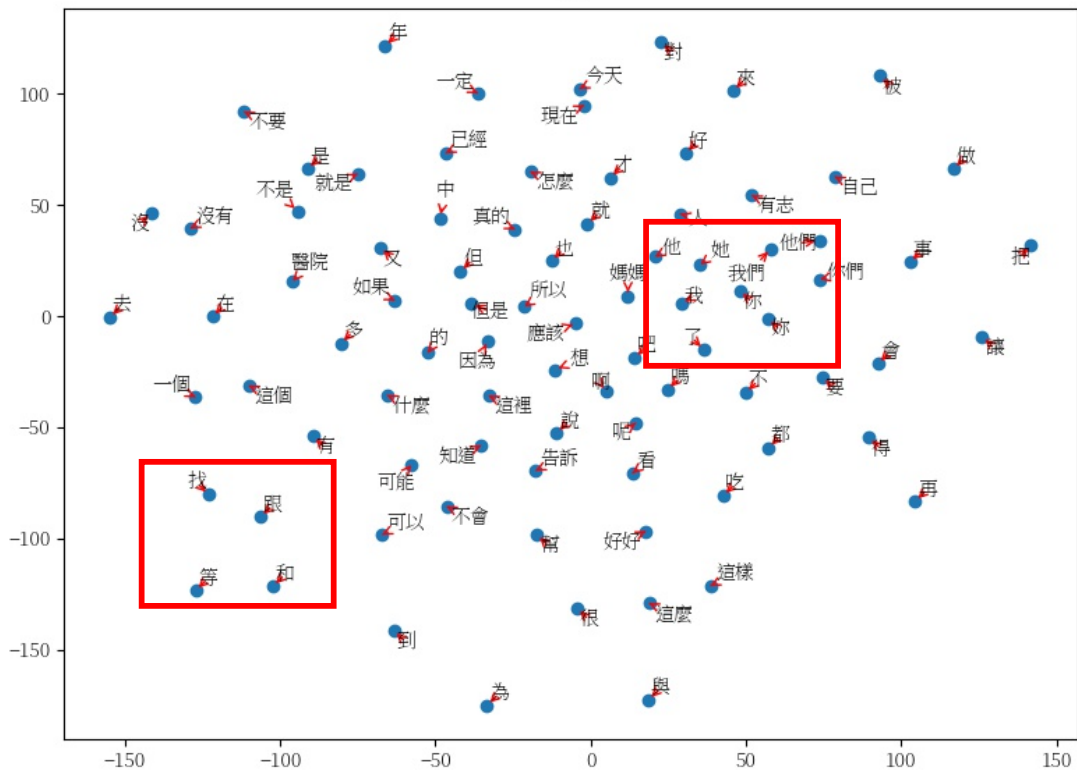
id = 287



4. 前四大 Eigen face 透過 $\frac{S_i}{\sum S_i}$ 運算後得到各自佔的比例為 4.1%, 2.9%, 2.4% 以及 2.2%，比起做在其他 dataset 上的 eigenvalue 的比例似乎少了很多，可能表示人臉百百種，並沒有一個所謂的基本臉。

Visualization of Chinese word embedding

1. 我使用 gensim 的 word2vec model，我僅將 embedding vector size 調整到 200，其他維持 package default。
- 2.



3. 我覺得效果並沒有像市面上常常看到的 $\text{king} = \text{queen} - \text{male} + \text{female}$ 這樣的明顯。但是依然可以看到諸如和、跟、找這三個單詞算是意義蠻相近的，而確實在圖上他們也在較為靠近的位置。或是他、妳、你們等等的也頗為靠近

Image clustering

1.

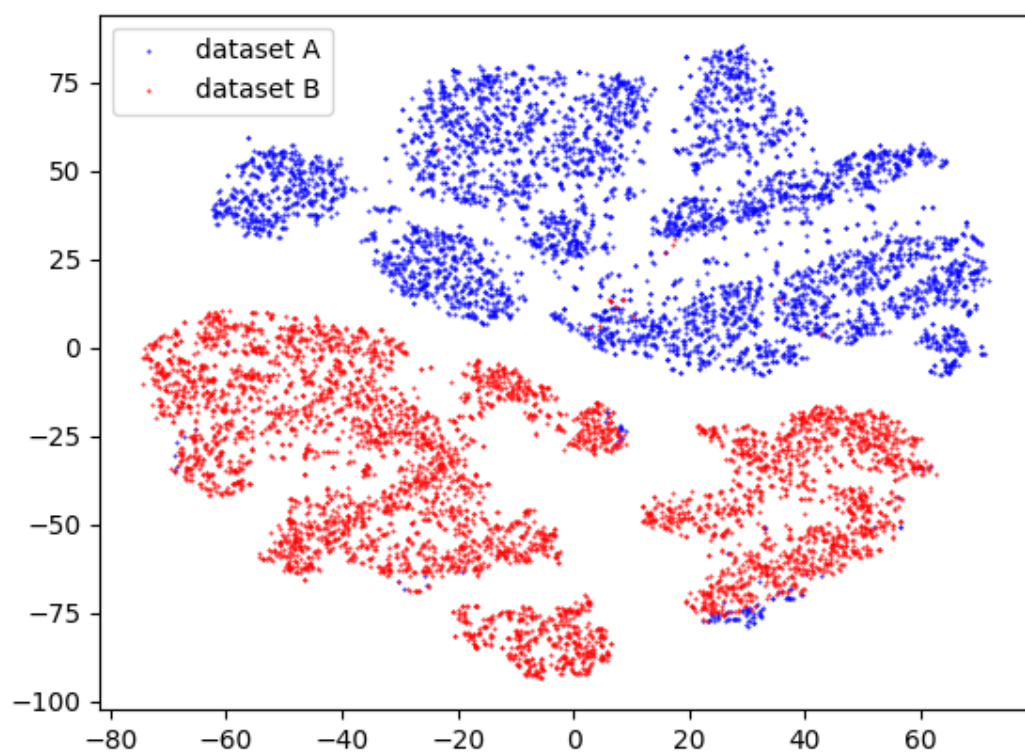
我試了兩種 model，一個是 convolution auto-encoder，結構如下左

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 28, 28, 1)	0
conv2d_1 (Conv2D)	(None, 28, 28, 16)	160
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 16)	0
conv2d_2 (Conv2D)	(None, 14, 14, 8)	1160
max_pooling2d_2 (MaxPooling2D)	(None, 7, 7, 8)	0
conv2d_3 (Conv2D)	(None, 7, 7, 8)	584
max_pooling2d_3 (MaxPooling2D)	(None, 4, 4, 8)	0
conv2d_4 (Conv2D)	(None, 4, 4, 8)	584
up_sampling2d_1 (UpSampling2D)	(None, 8, 8, 8)	0
conv2d_5 (Conv2D)	(None, 8, 8, 8)	584
up_sampling2d_2 (UpSampling2D)	(None, 16, 16, 8)	0
conv2d_6 (Conv2D)	(None, 14, 14, 16)	1168
up_sampling2d_3 (UpSampling2D)	(None, 28, 28, 16)	0
conv2d_7 (Conv2D)	(None, 28, 28, 1)	145

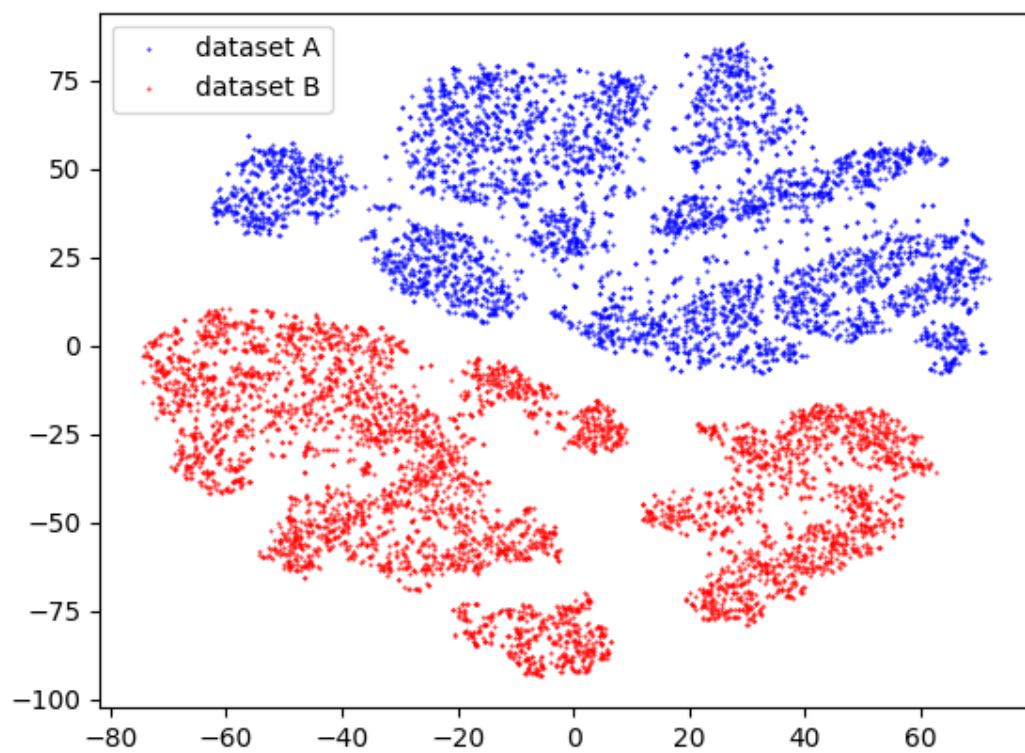
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 784)	0
dense_1 (Dense)	(None, 128)	100480
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 64)	2112
dense_5 (Dense)	(None, 128)	8320
dense_6 (Dense)	(None, 784)	101136
Total params: 222,384		
Trainable params: 222,384		
Non-trainable params: 0		

另一種則是一般的 auto-encoder，如上右，結果會發現 CAE 完全爛掉，連 simple baseline 都過不了，一樣都是跑 50 個 epoch。

2.



3.



可以看到較明顯的錯誤分佈在圖的右下角有一塊，以及一些零星的散落在圖的中央。但是因為這 **model** 的 **F-score** 可以到 **99.1** 左右，因此錯誤僅算是少部分。