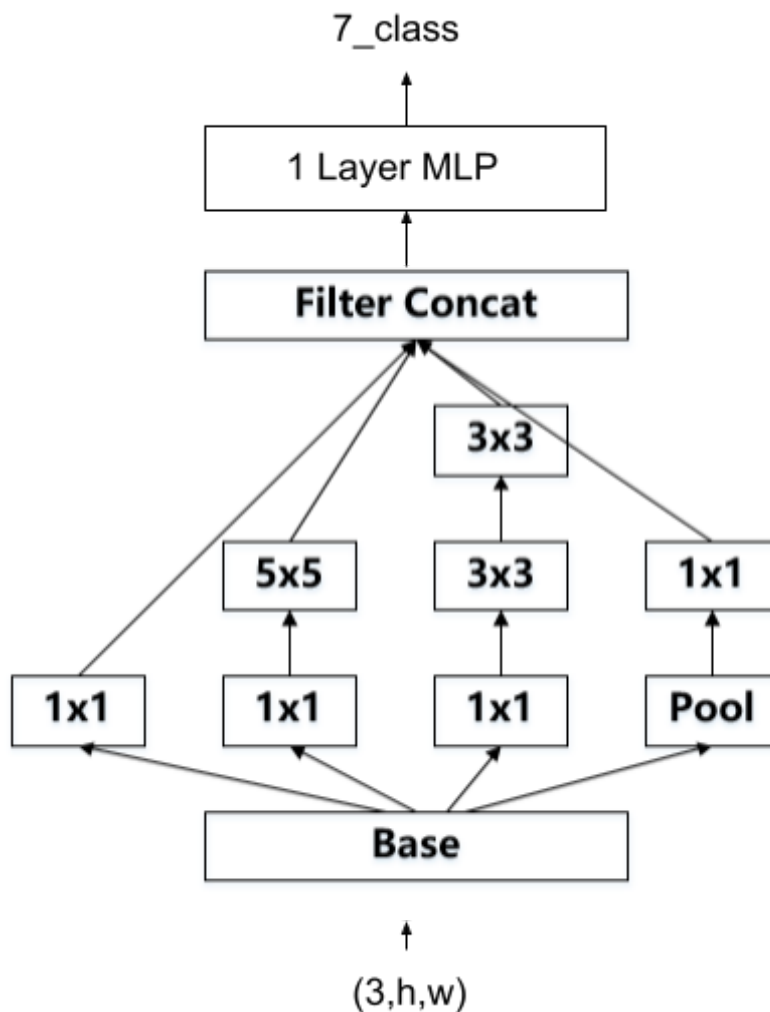


## prob1

1. Model B: inception\_v3



2.

Model A: valid acc 0.6824

```
Training batch index: 2813, train loss: 0.249574, acc: 0.919
0% | 0/2813 [00:00<?, ?it/s] current loss: 0.08226721711084234 valid_acc: 0.6672
best_acc 0.6824
```

Model B: inception\_v3 valid acc=0.8772

```
Training batch index: 1407, train loss: 0.060238, acc: 0.984
current loss: 0.03236865505675989 valid_acc: 0.8716
best_acc 0.8772
```

3.

\*採用三次相同的module, module是由3個convolution結合  
conv(128,256)conv(256,512),conv(512,128), 在最後一層將conv(512,128)改成  
conv(512,512), 在第一層為了結合照片的channel尺寸, 所以是conv(3,128)  
\*在每個convolution前都有做padding避免尺寸變小

\*每個convolution後都有做batchnormalize及relu

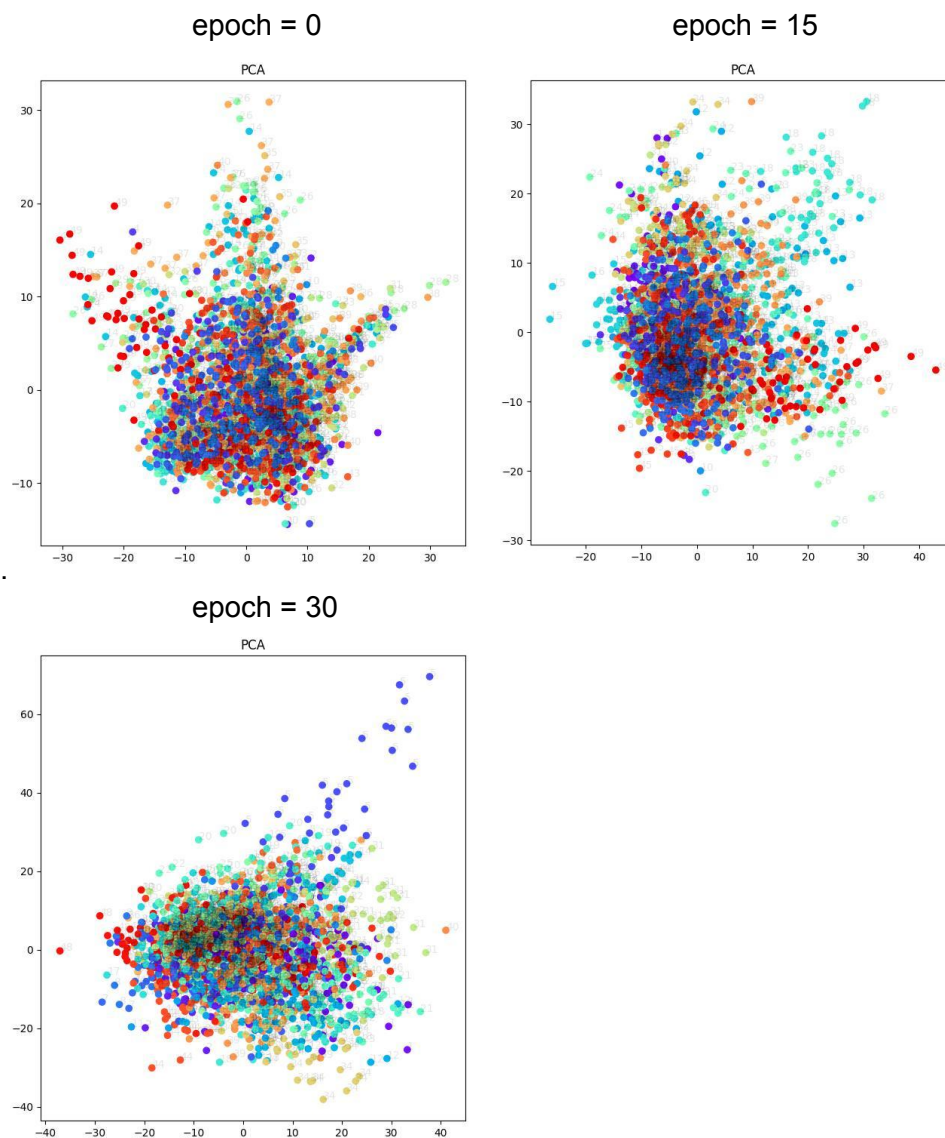
\*每個module做完會有一個maxpooling

\*optimizer採用SGD,learning rate = 0.001,momentum = 0.9

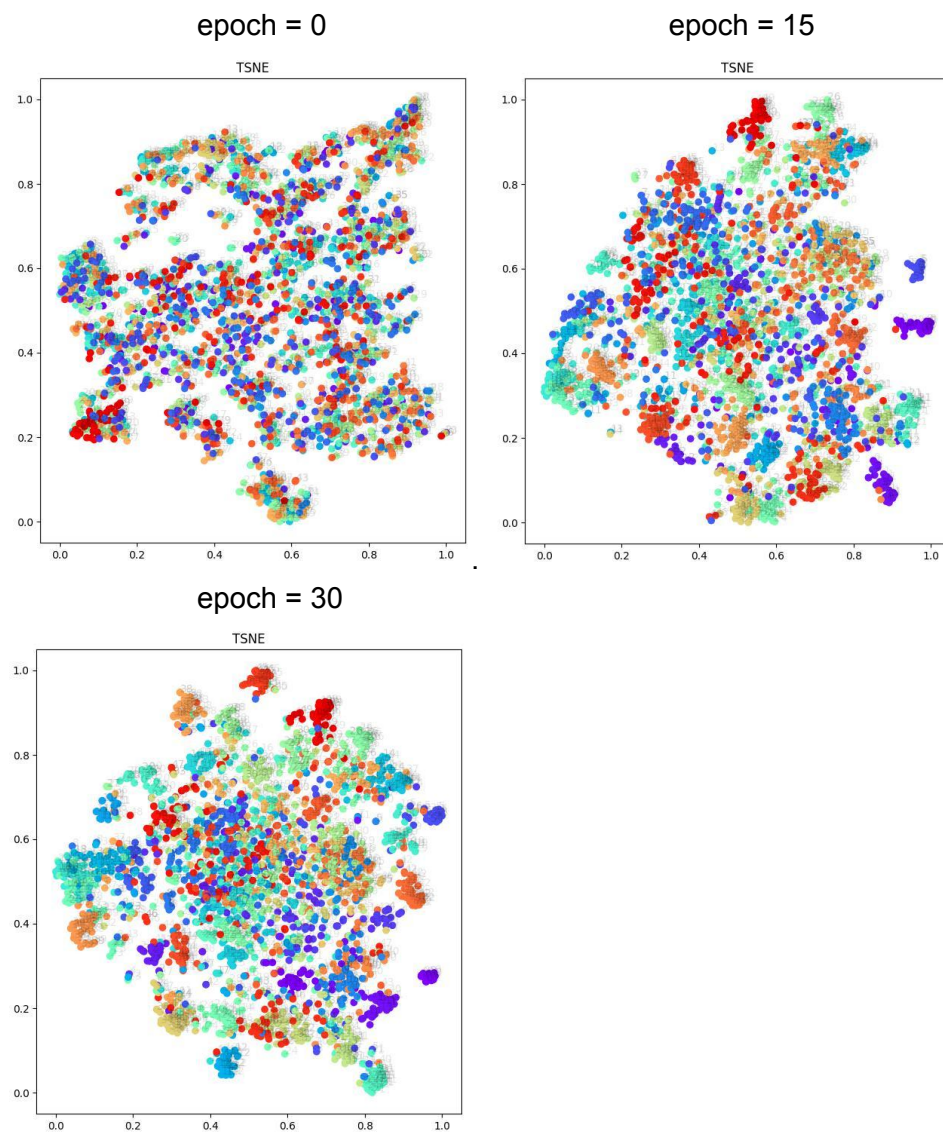
\*loss function採用crossentropy

4.Model B使用INCEPTION\_V3, 其特色包含:label smoothing,factorized 7\*7 convolution,auxiliary classifier, 相較於Model A只用很多convolution+最後幾層分類器來說更加複雜且收斂性更佳, 以及在梯度消失也有處理的方法。

5.從PCA看起來圖片相當混亂, 一方面也可能是由於PCA是線性的所以無法將類別分開, 還有一個原因是我本身訓練的準度就不夠

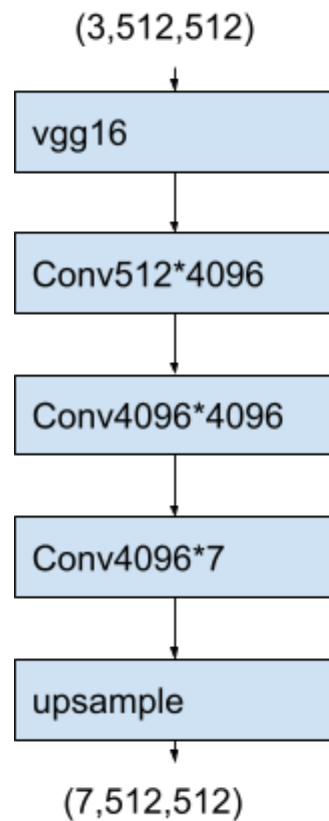


6.T-SNE效果明顯比PCA好，由於TSNE非線性方式取做處理，因此能將每個類別集中且與其他類別分離開來。



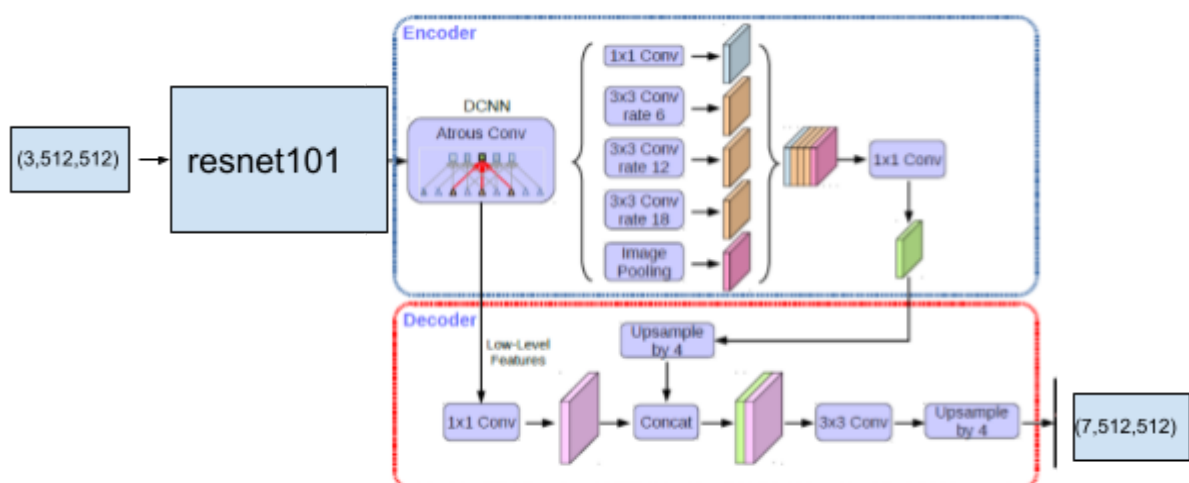
# prob2

## 1. Model A: vgg16+fcn32



## 2. Model B: resnet101+deeplabv3

backbone採用resnet101與model A 的vgg16不同，encoder及decoder的部份fcn是以三個convolution組成，而deeplabv3則採用較為複雜的方法組成。然而本來期待model B能有更好的表現，但由於時間的關係，沒能找到較好的參數，所以在結果上反而是model A的MIOU較高。



3.

model A	model B
class #0 : 0.74086	class #0 : 0.62339
class #1 : 0.88328	class #1 : 0.82317
class #2 : 0.34309	class #2 : 0.24231
class #3 : 0.80713	class #3 : 0.66605
class #4 : 0.70031	class #4 : 0.42509
class #5 : 0.68350	class #5 : 0.63369
mean_iou: 0.693027	mean_iou: 0.568951

4.

