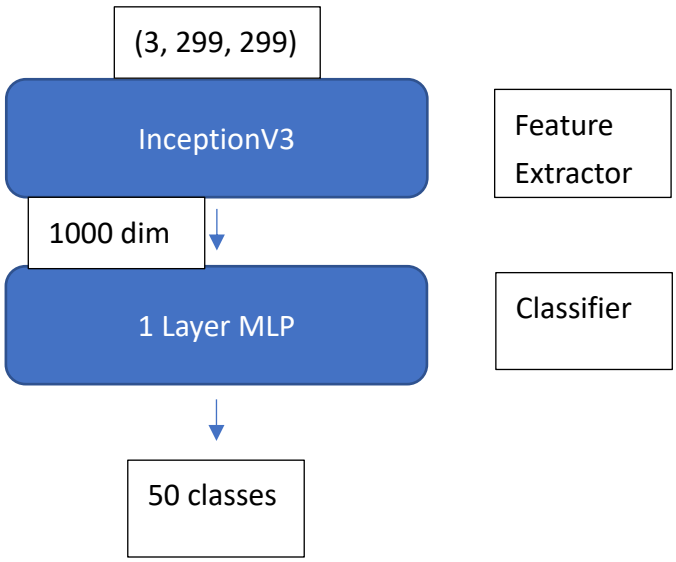
DLCV hw1 Report

B09901062 黃宥翔

Problem1

(2%) Draw the network architecture of method A or B.

Method B



(1%) Report accuracy of your models (both A, B) on the validation set.

A: 0.69

B:0.87

(4%) Report your implementation details of model A.

Optimizer: Adam

Loss function: CrossEntropyLoss

Gradient norm: clip_grad_norm

Some Transforms:

```
transforms.RandomApply(transforms=[transforms.RandomHorizontalFli
p(), transforms.RandomResizedCrop(size=(128, 128))], p = 0.9)

transforms.RandomApply(transforms=[transforms.ColorJitter(brightn
ess=0.2, contrast=0.5, saturation=0.2, hue=0.1),
transforms.RandomEqualize(),
transforms.RandomSolarize(threshold=100.0)], p = 0.4)
```

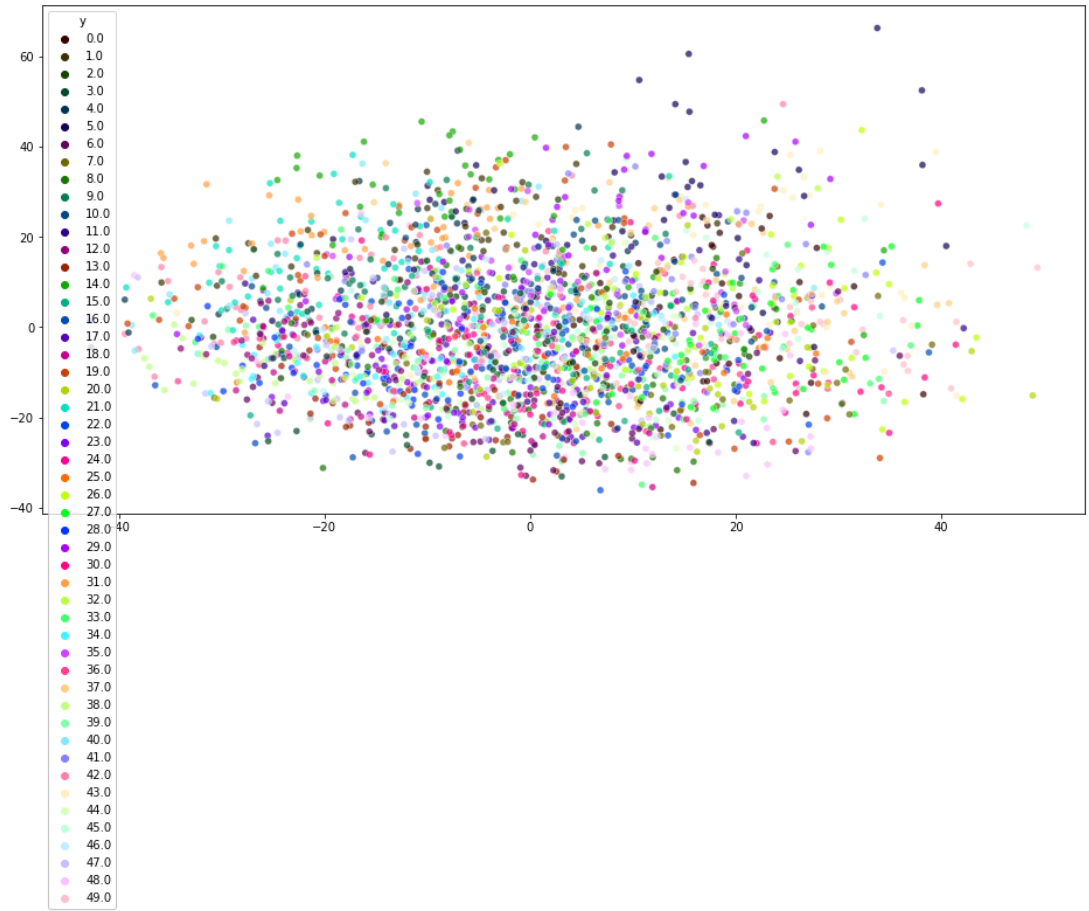
(4%) Report your alternative model or method in B, and describe its difference from model A

Ans:

In model B, I use pretrained model – InceptionV3 as feature extractor. The others are the same as model A.

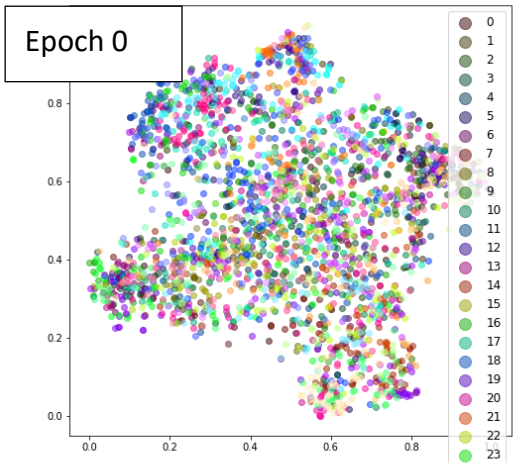
The InceptionV3 model acts as a great feature extractor compared with the CNN model A, because the inception architecture reduces the param so that the model can build deeper.

(7%) Visualize the learned visual representations of **model A** on the **validation set** by implementing **PCA** (Principal Component Analysis) on the output of **the second last layer**. Briefly explain your result of the PCA visualization.

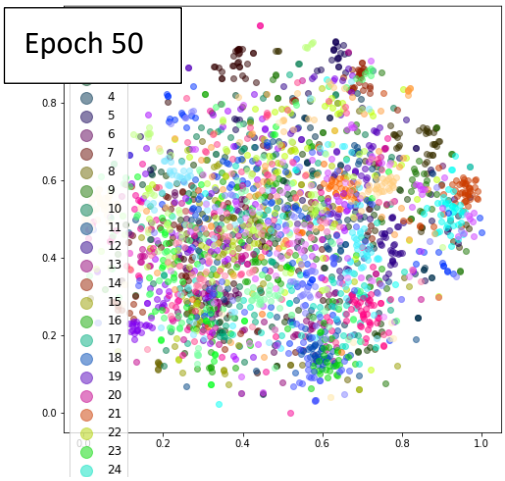


The clustering is not obvious. Part of it may be caused by the poor performance of model A.

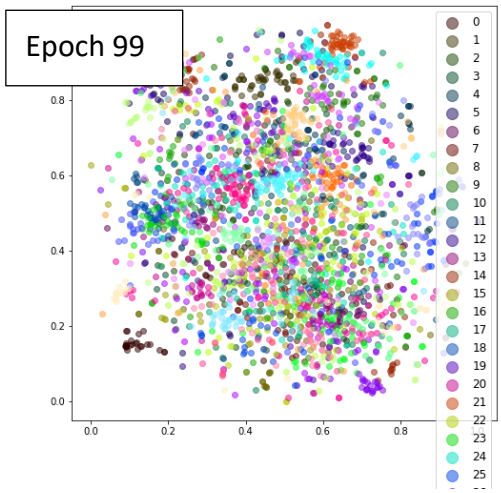
(7%) Visualize the learned visual representation of **model A**, again on the output of the second last layer, but using **t-SNE** (t-distributed Stochastic Neighbor Embedding) instead. Depict your visualization from **three different epochs** including the first one and the last one. Briefly explain the above results.



In early stage, there is no clustering.



Howerver, in mid stage, clusters start to form, like the orange part in the right and the purple part in the left-lower part.



In the late stage, clustering is more obvious. Like purple, yellow, black, pink, orange, brown….

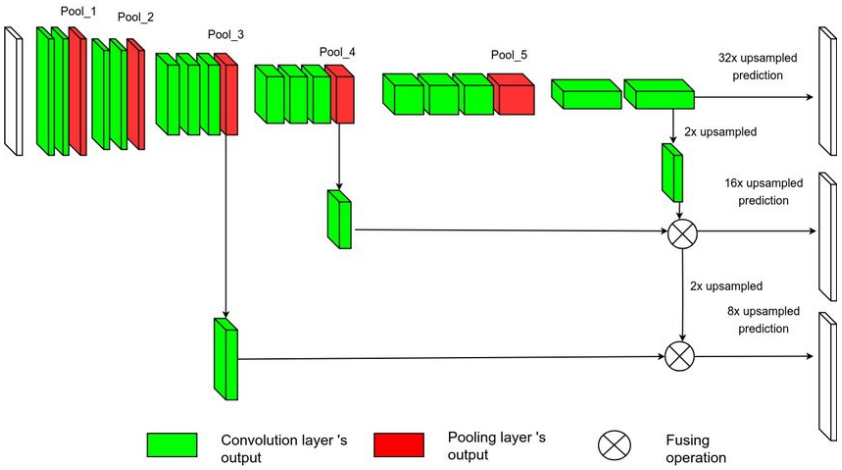
This shows that the feature extractor is actually growing as the training stage moves on, since the similar features are gathered together.

Problem2

(5%) Draw the network architecture of your VGG16-FCN32s model (model A).

From

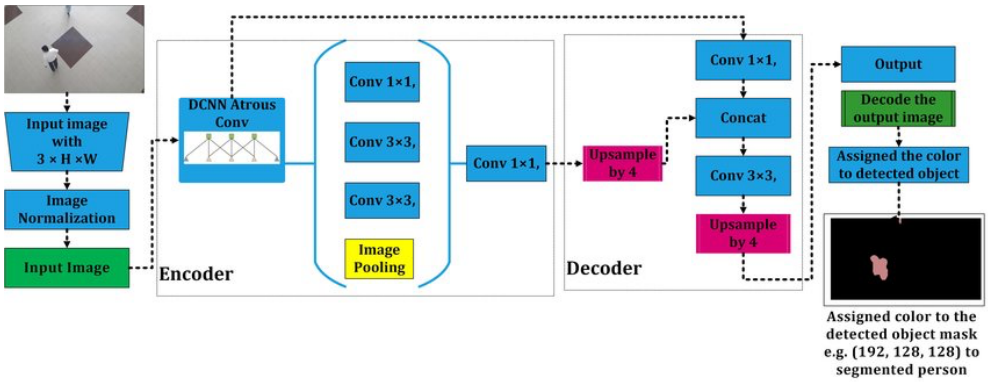
Aerial Image Semantic Segmentation Using Neural Search Network Architecture: 12th International Conference, MIWAI 2018, Hanoi, Vietnam, November 18–20, 2018, Proceedings



(5%) Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model.

From

Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images



Main difference:

- Atrous Spatial Pyramid Pooling (ASPP) avoids information loss in downsampling
- Instead of VGG, the backbone changes to ResNet

(3%) Report mIoUs of two models on the validation set.
A:0.43
B:0.75

(7%) Show the predicted segmentation mask of "validation/0013_sat.jpg", "validation/0062_sat.jpg", "validation/0104_sat.jpg" during the early, middle, and the final stage during the training process of the improved model.