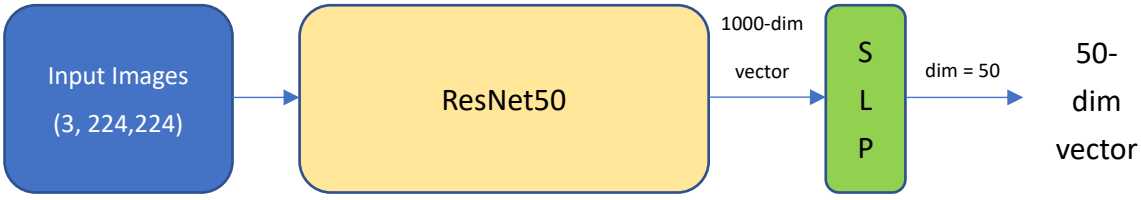
DLCV HW1

B08901165 電機四 南策昇

Problem 1.

1. Model structures of model A and model B are the same.



2. Accuracy: A: 71.92% B: 87.28%

3. Implementation details of model A:

Optimizer: Adam with lr = 0.00005 and pytorch default hyperparameters.

Loss function: Cross entropy

In training part, I first doubled the number of data by flipping horizontally. And, for each image, both flipped and original, I resized it to 256x256 and then implemented some transforms, details as below, in pytorch package.

Transforms:

ColorJitter(brightness = 0.3, contrast = 0.3, saturation = 0.3, hue = 0.1),

RandomRotation(5),

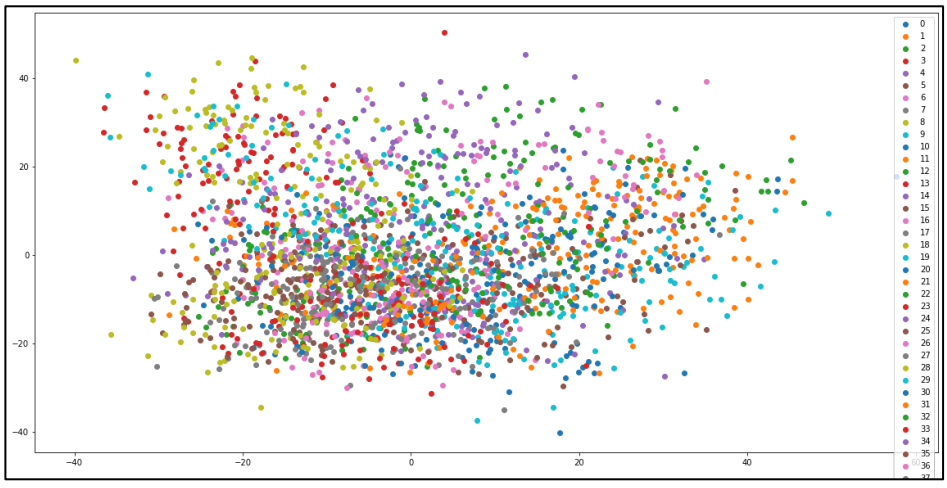
RandomPerspective(p = 0.7)

Finally, images were resized to 224x224 and normalized with mean = (0.485, 0.456, 0.406) and std = (0.229, 0.224, 0.225) for each dimension.

The batchsize was 32 and the model was trained until it stopped improving in the validation set for five epochs.

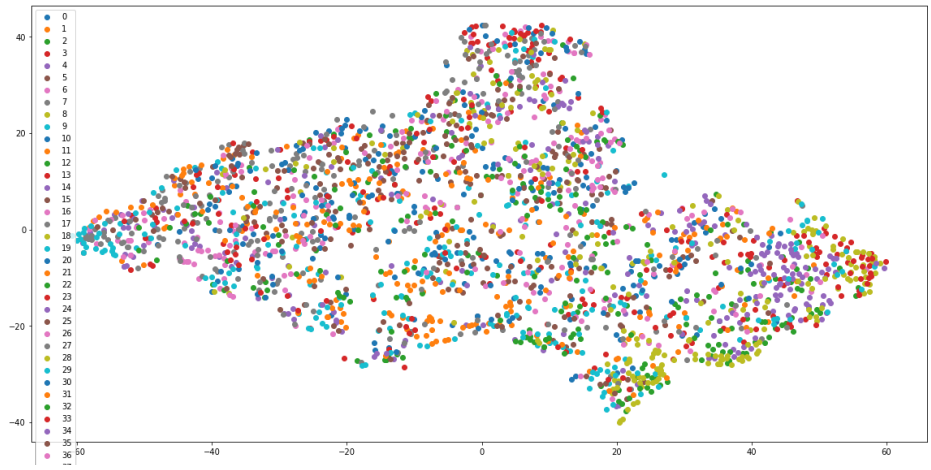
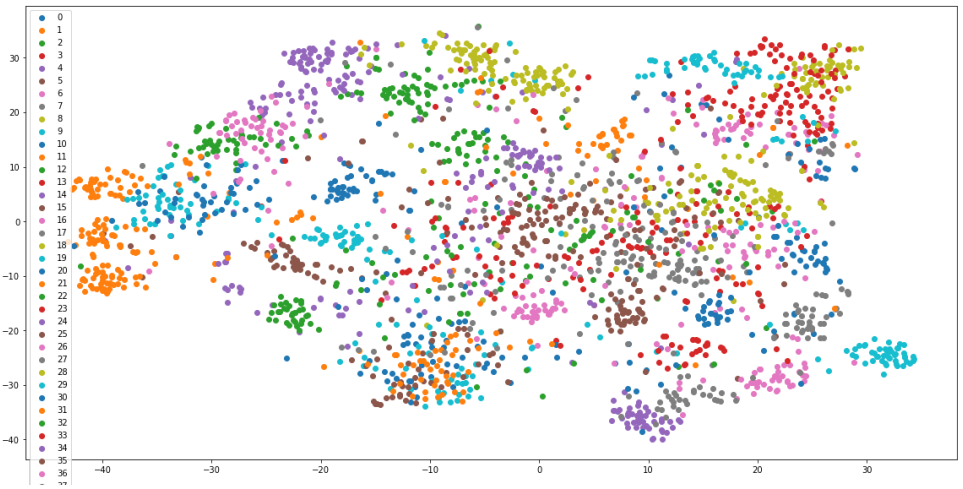
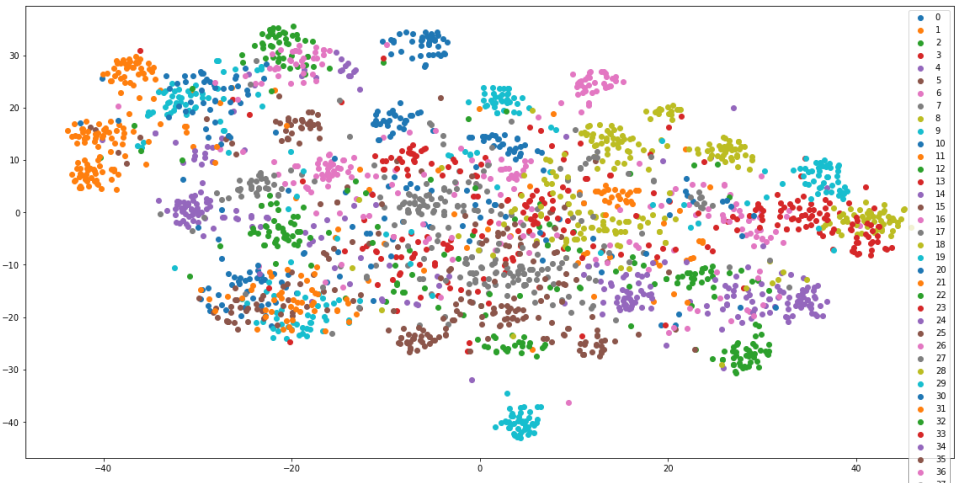
4. The transforms and hyperparameters of model B are all same as those for model A except that model B uses pretrained model of resnet50.

5. PCA plot



因為 PCA 是線性降維，能做的事情比較有限，所以當類別稍多或者複雜的時候，就沒有辦法呈現的很好。

6. t-SNE plots

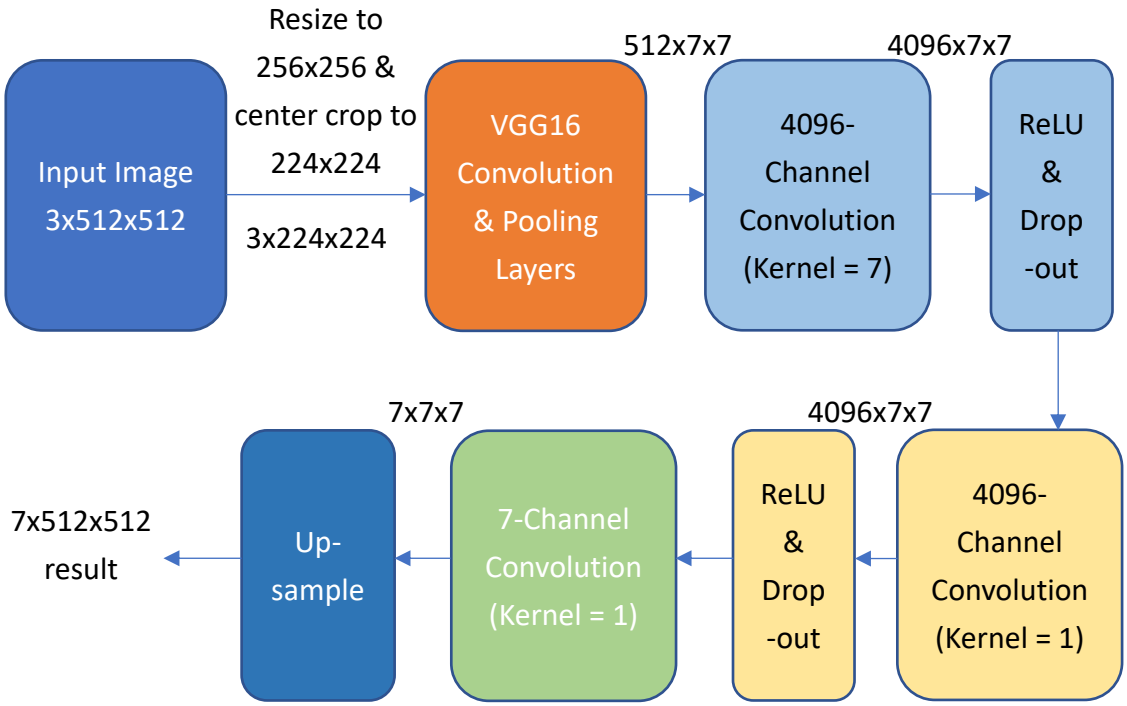
Epoch	Plot
1	
19	
34	

此為 perplexity = 50 的圖。t-SNE 是非線性的降維，所以在這個 50 個圖像分類的 case 裡面，表現的明顯比 PCA 好。另外隨著 Epoch 數增加，t-SNE 分群也做

得越好，顯示 model 是有在逐漸學習的。最後，由於 t-SNE 有 stochastic 的成分在，所以每一次跑出來的結果都不太一樣。

Problem 2.

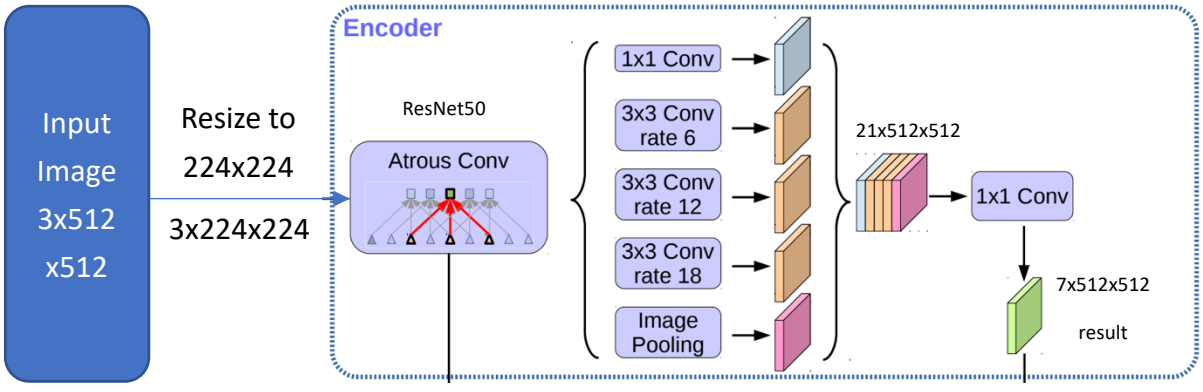
1. VGG16-FCN32



For data augmentation, each image had the probability of 0.5 to be flipped and and a uniform probability to rotate [0, 90, 180, 270] degrees. Besides, auto-contrast was applied and sharpness was raised by a factor of 2.

I used Adam with lr = 0.00005 as the optimizer and cross entropy as the loss function. The model was trained until it had stopped improving for 5 epochs with respect to mIoU.

2. DeepLab v3 with pretrained weights



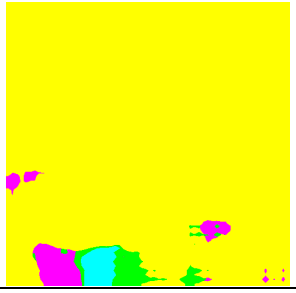
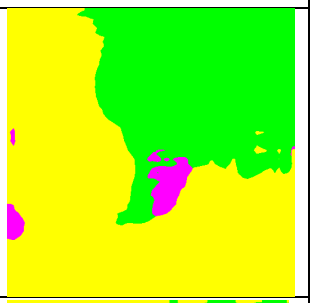
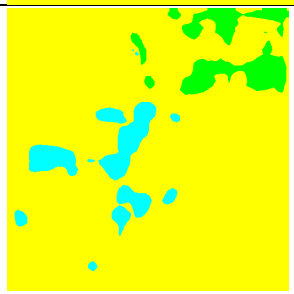
Source: [1802.02611v3.pdf (arxiv.org)]

In DeepLab v3 model with ResNet50 as the backbone, the method ASPP with different rates is used to capture any possible size of features. In this way, it's expected to generate more precise semantic segmentation images. Besides, ResNet is considered better than VGG in general cases, so it's also a factor to enhancing the performance.

For data augmentation, most were same as model A while sharpness and auto-contrast were both turned off since they seemed to affect the model in a negative way. The optimizer was still Adam but the lr was tuned from 0.00005 to 0.00003 and finally to 0.00001 once the model stopped improving. Cross entropy was used as the loss function.

3. mIoU: A: 0.6463 B: 0.7551

4.

	0013_sat	0062_sat	0104_sat
Early			
Middle			
Final