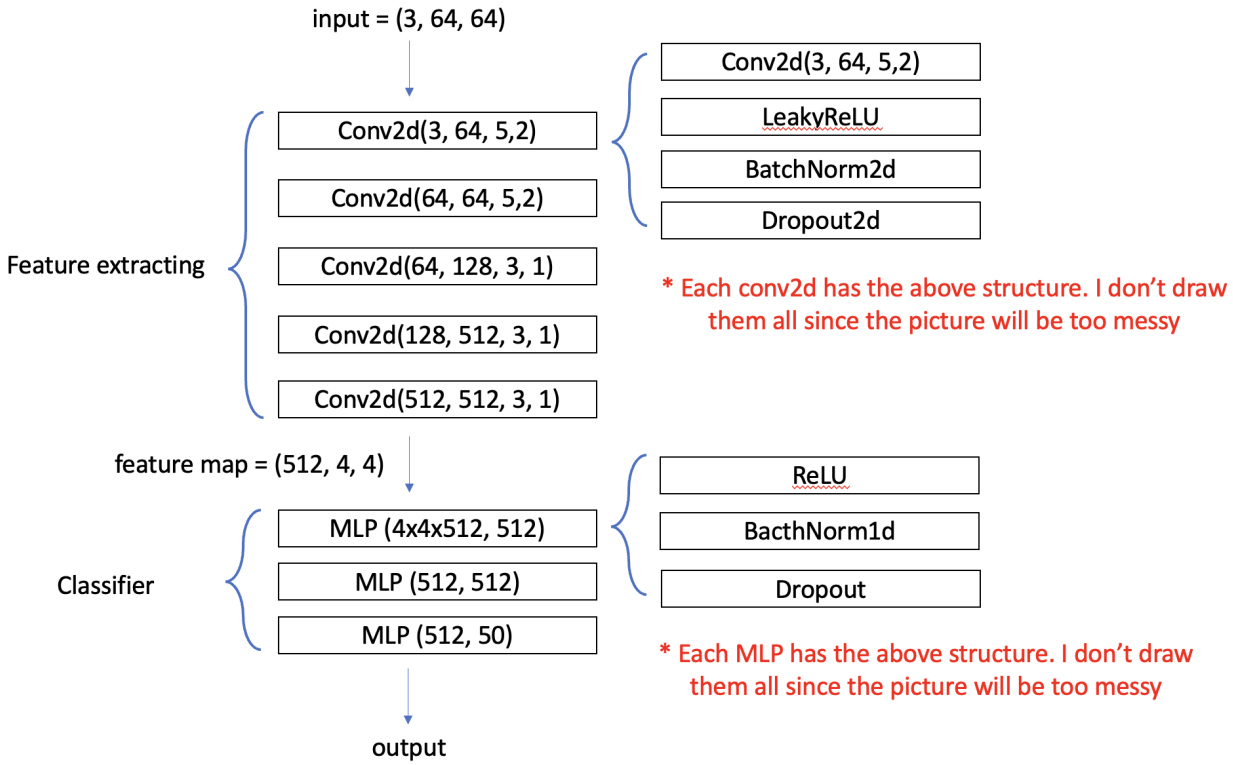
DLCV HW1

R11922A05 資工AI碩一 林聖硯

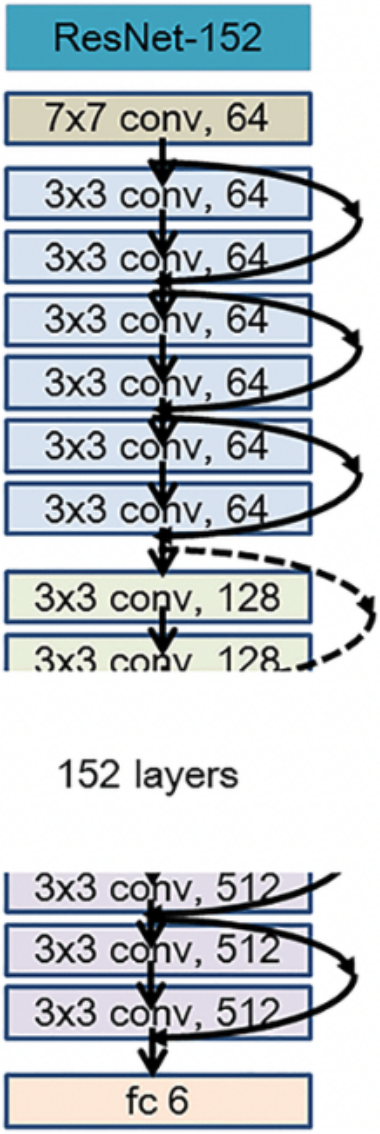
Problem1

1. Draw the network architecture of method A or B

Model A: CNN (designed by myself)



Model B: Resnet152



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

Model	Train Acc.	Valid Acc.
ModelA	60.57%	45.24%
ModelB	95.16%	90.64%

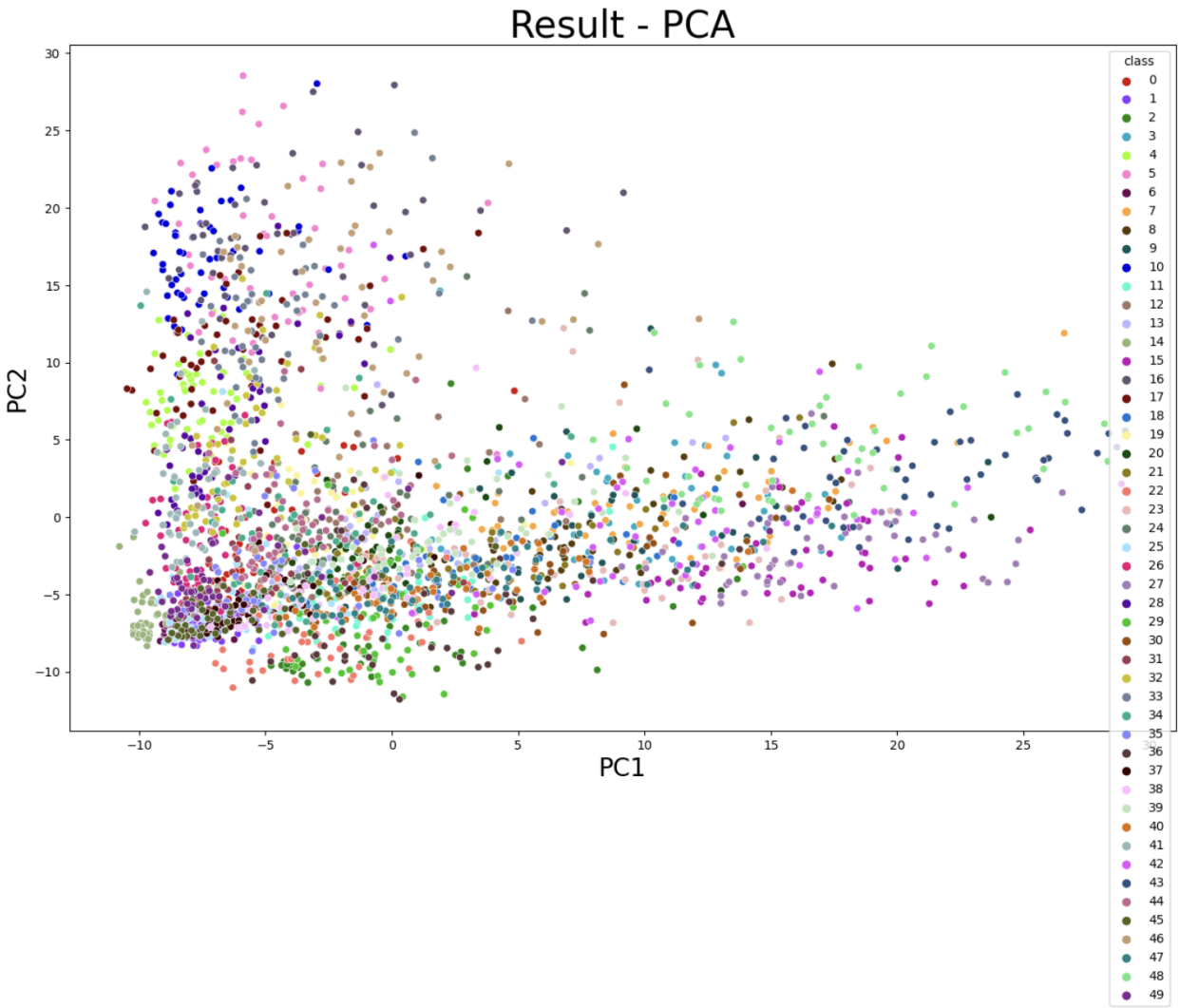
3. Report your implementation details of model A.
   The following setting also applies to model B

- batch size: 48
- Learning rate setting
  - optimizer: SGD
  - learning rate: 0.0001
  - momentum: 0.9

    - weight decay: 0.00005
- Loss function: cross entropy loss
- Data augmentation
    - resize to 96 x 96
    - CenterCrop to 64 x 64
    - RandomVerticalFlip
    - RandomRotation
    - Normalization using image net's mean and std

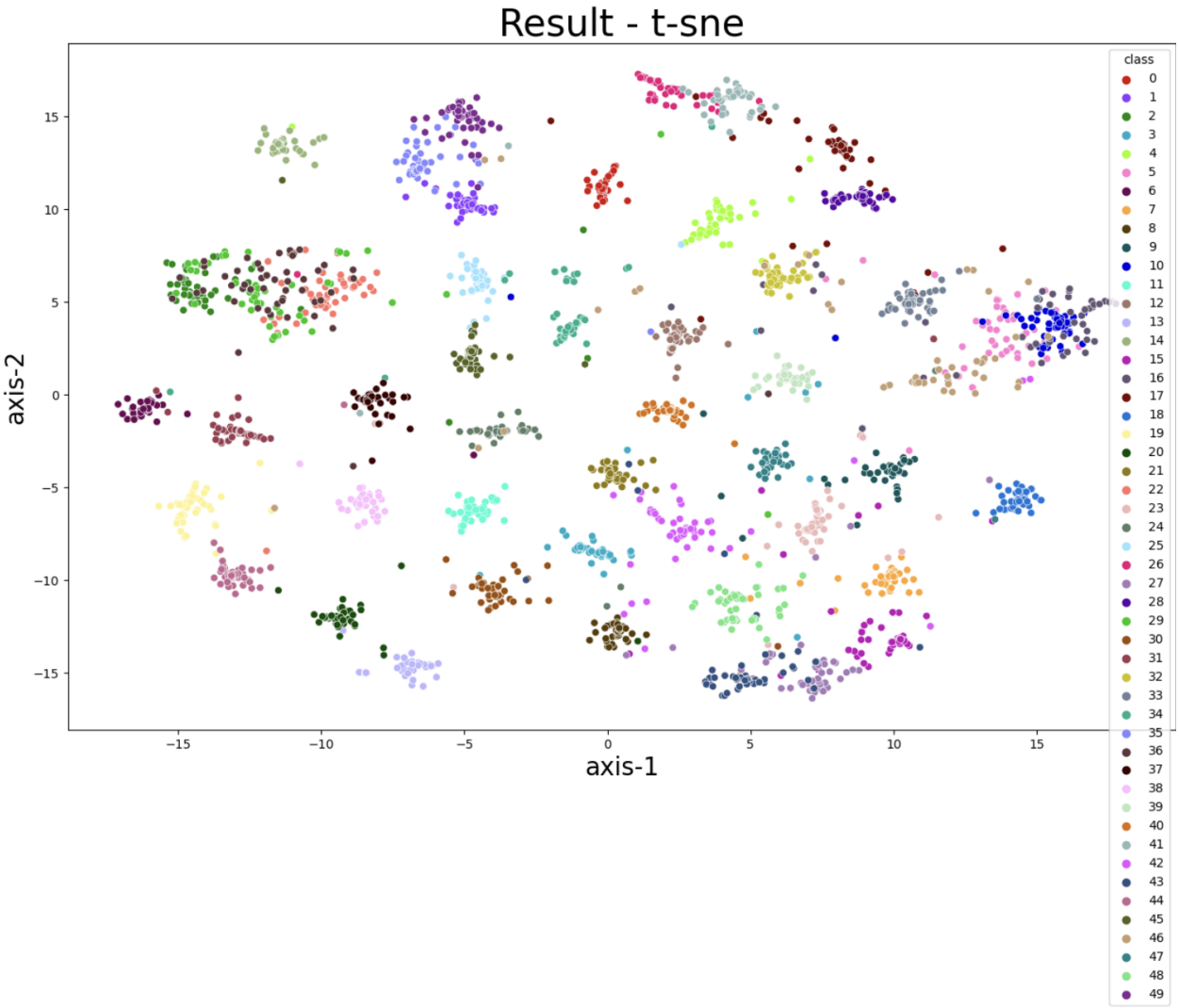
4. Report your alternative model or method in B, and describe its difference from model A.
   The second model that I use is Resnet 152, pre-trained on ImageNet-1k at resolution 224x224. ResNet (Residual Network) is a convolutional neural network that uses the concepts of residual learning and skip connections. This enables to train much deeper models. In addition, the pretrained weight is from a large dataset. Thus, it is possible to get a better result with this pretrained model.

5. 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.



From the picture, we could observe that the points aren't effectively seperated by PCA. Most of the classes are mixed together severely in the graph since PCA does dimension reduction in a linear fashion.

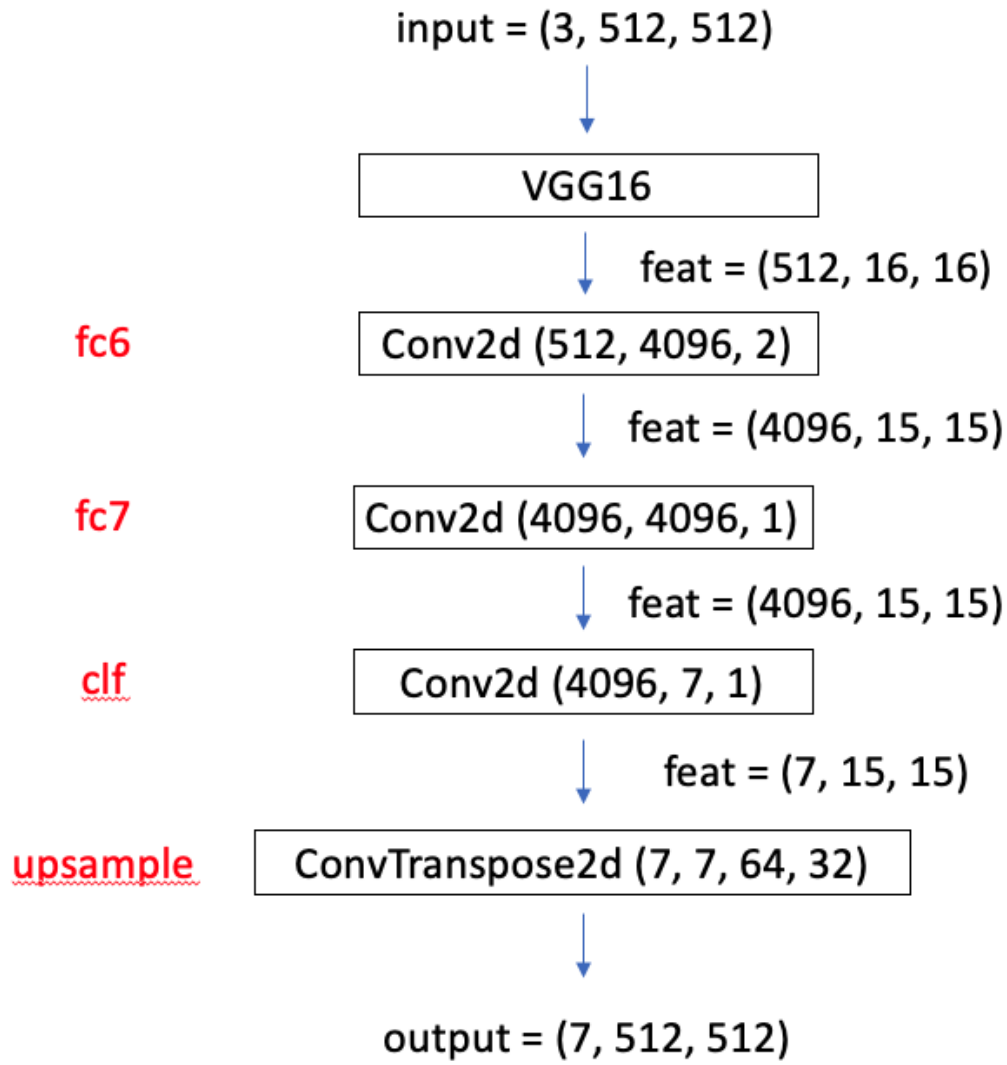
6. Visualize the learned visual representation of model A, again on the output of the second last layer, but using t-SNE (t-distributed Stochastic Neighbor Embedding) instead. Depict your visualization from three different epochs including the first one and the last one. Briefly explain the above results.



The result of T-SNE is much better that that of PCA. We could clearly see that classes are seperated from each other. However, there are still some points mixing together since the pictures of these classes are similar and my model couldn't effectively distinguish one from another.

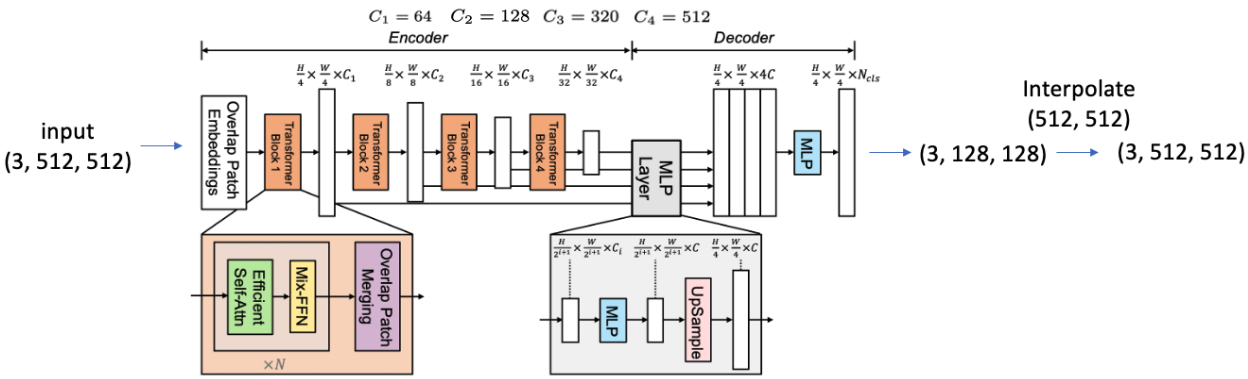
Problem2

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



2. Draw the network architecture of the improved model (model B)

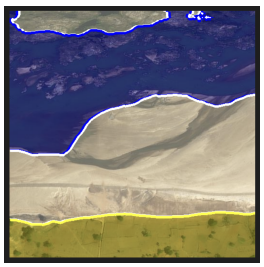
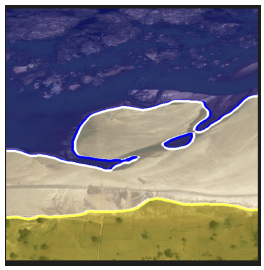
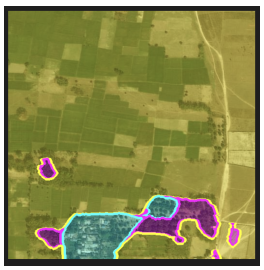
The second model that I used is SegFormer (b1-sized).



3. Report mIoUs of two models on the validation set.

Model	Valid mIoUs
ModelA	67.53%
ModelB	74.26%

4. 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.

pic	Epoch 1	Epoch 10	Epoch 20
0013_sat.jpg			
0062_sat.jpg			
0104_sat.jpg