Problem 1
classification

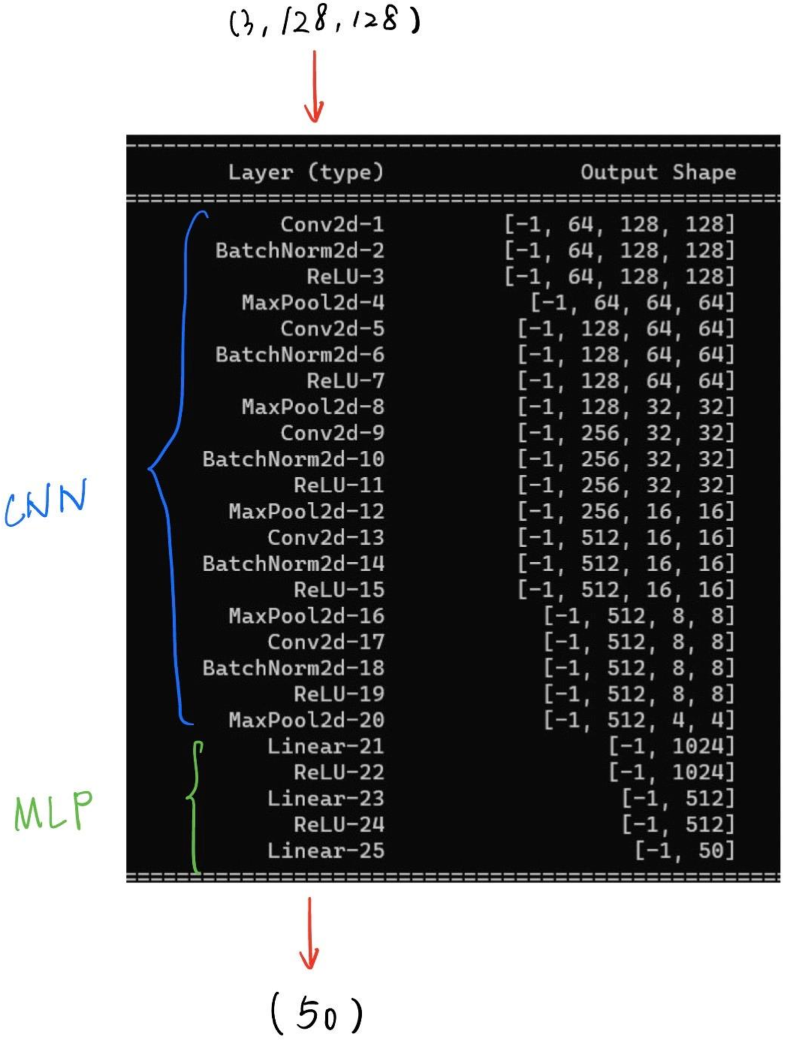
b09901104翁瑋杉

Report (**25%**)

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

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

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



Layer (type)	Output Shape
Conv2d-1	[-1, 64, 128, 128]
BatchNorm2d-2	[-1, 64, 128, 128]
ReLU-3	[-1, 64, 128, 128]
MaxPool2d-4	[-1, 64, 64, 64]
Conv2d-5	[-1, 128, 64, 64]
BatchNorm2d-6	[-1, 128, 64, 64]
ReLU-7	[-1, 128, 64, 64]
MaxPool2d-8	[-1, 128, 32, 32]
Conv2d-9	[-1, 256, 32, 32]
BatchNorm2d-10	[-1, 256, 32, 32]
ReLU-11	[-1, 256, 32, 32]
MaxPool2d-12	[-1, 256, 16, 16]
Conv2d-13	[-1, 512, 16, 16]
BatchNorm2d-14	[-1, 512, 16, 16]
ReLU-15	[-1, 512, 16, 16]
MaxPool2d-16	[-1, 512, 8, 8]
Conv2d-17	[-1, 512, 8, 8]
BatchNorm2d-18	[-1, 512, 8, 8]
ReLU-19	[-1, 512, 8, 8]
MaxPool2d-20	[-1, 512, 4, 4]
Linear-21	[-1, 1024]
ReLU-22	[-1, 1024]
Linear-23	[-1, 512]
ReLU-24	[-1, 512]
Linear-25	[-1, 50]

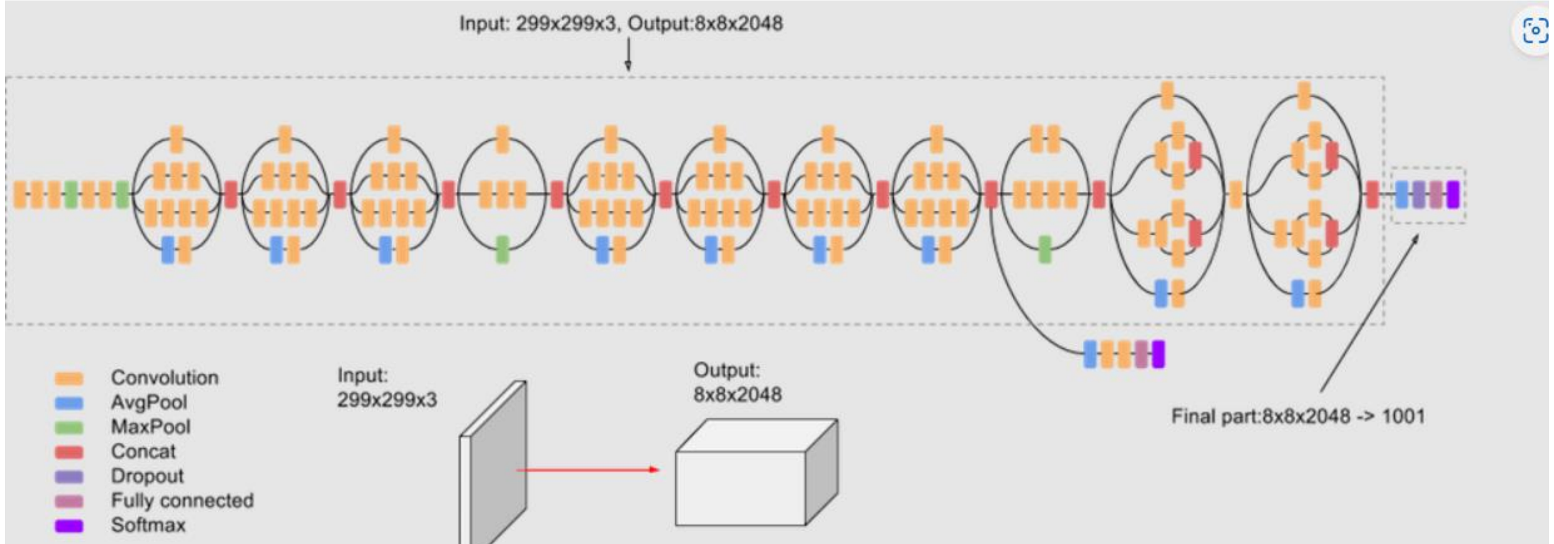
```python
train_tfm = transforms.Compose(
    [
        CIFAR10Policy(),
        transforms.Resize(128),
        # transforms.RandomHorizontalFlip(0.5),
        # transforms.RandomRotation(20),
        # transforms.RandomAffine(degrees=0, translate=(0, 0.5)),
        transforms.ToTensor(),
        transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]),
    ]
)

test_tfm = transforms.Compose(
    [
        CIFAR10Policy(),
        transforms.Resize(128),
        # transforms.Resize(299),
        transforms.ToTensor(),
        transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]),
    ]
)
```

```python
criterion = nn.CrossEntropyLoss()
optimizer = torch.optim.Adam(model.parameters(), lr=1e-5, weight_decay=1e-5)
```

Model A, best validation acc = 0.65

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



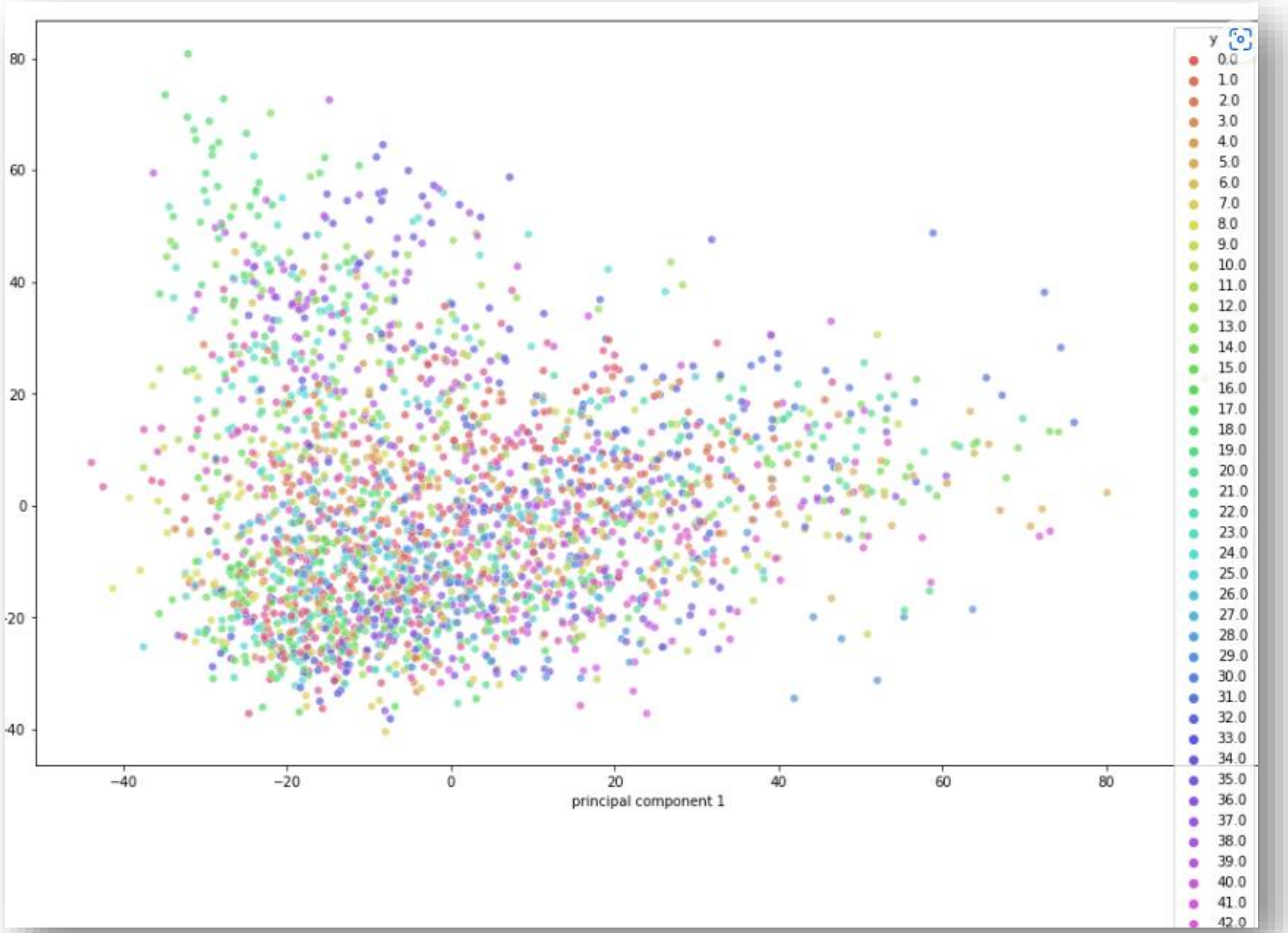
Model B, Inceptionv3 , best validation acc = 0.87 Cite as: arXiv:1512.00567

- Unlike model A , model B avoid drastic down-sampling and bottlenecks, because they might cost some features.

- Model B also has an auxiliary classifier, which model A doesn’t.

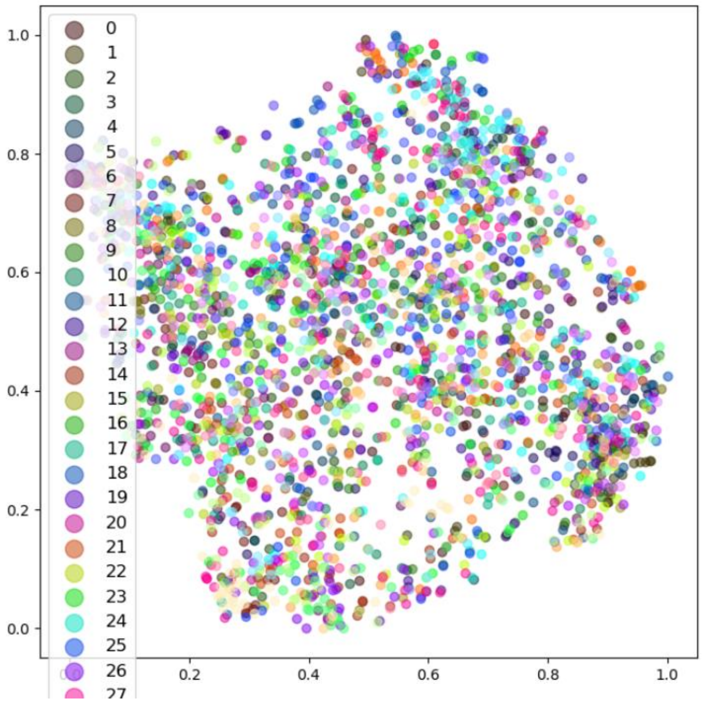
- Using some techniques to reduce parameters, model B is able to develop deeper network than model A, which probably leads to better performance.

5.(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.

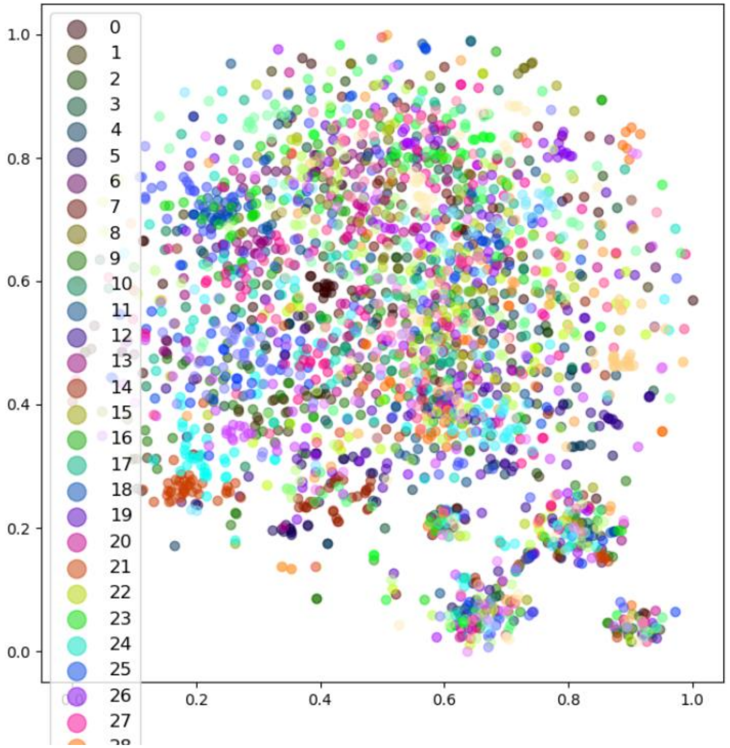


The model has validation acc = 0.65 . We can see that PCA doesn't really form distinct clusters. Data is still uniformly mixed and distributed.

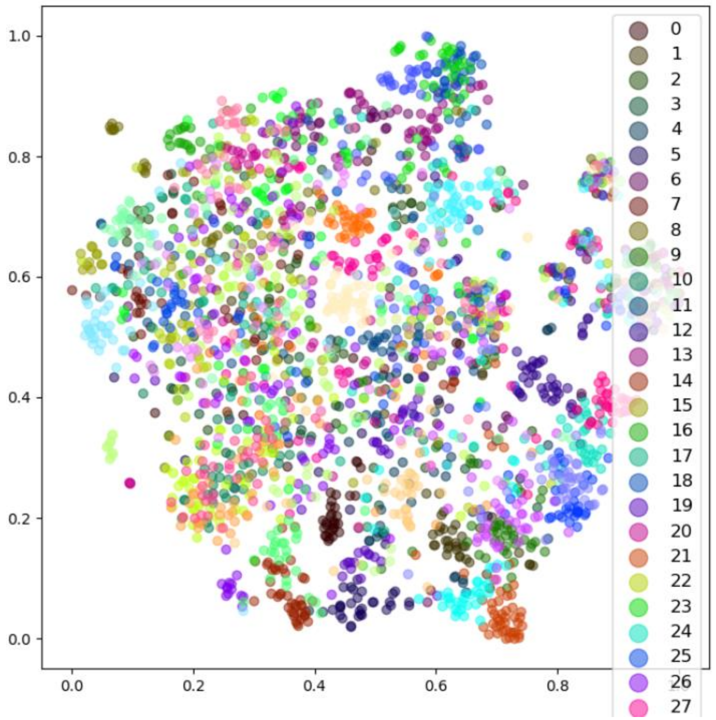
6. (7%) Visualize the learned visual representation of **model A**, again on the output of the second last layer, but using **t-SNE** (t-distributed Stochastic Neighbor Embedding) instead. Depict your visualization from **three different epochs** including the first one and the last one. Briefly explain the above results.



First epoch , acc = 0.1



acc = 0.45



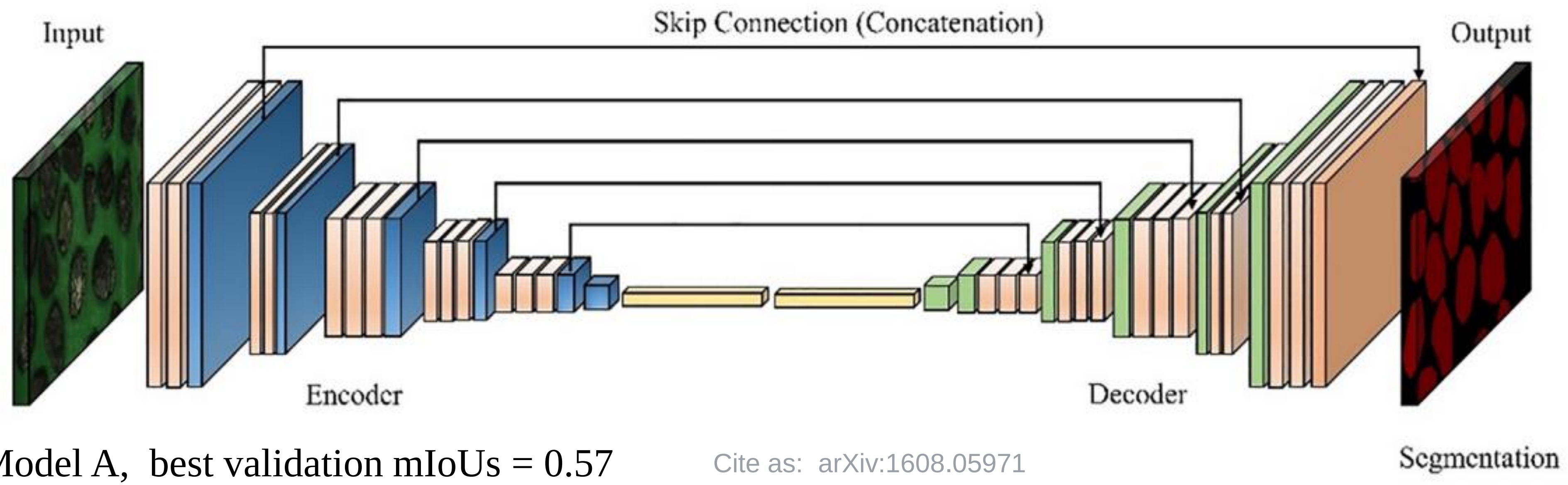
last epoch , acc = 0.67

For the first epoch, data is very uniformly distributed. In the middle stage, we observe clusters , which means the model is classifying, although not exactly right. For the last epoch, clusters are getting more distinct, and color in clusters are generally same.

Problem 2
segmentation

b09901104翁瑋杉

1. **(5%)** Draw the network architecture of your VGG16-FCN32s model (model A).
2. **(5%)** Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model.
3. **(3%)** Report mIoUs of two models on the validation set.



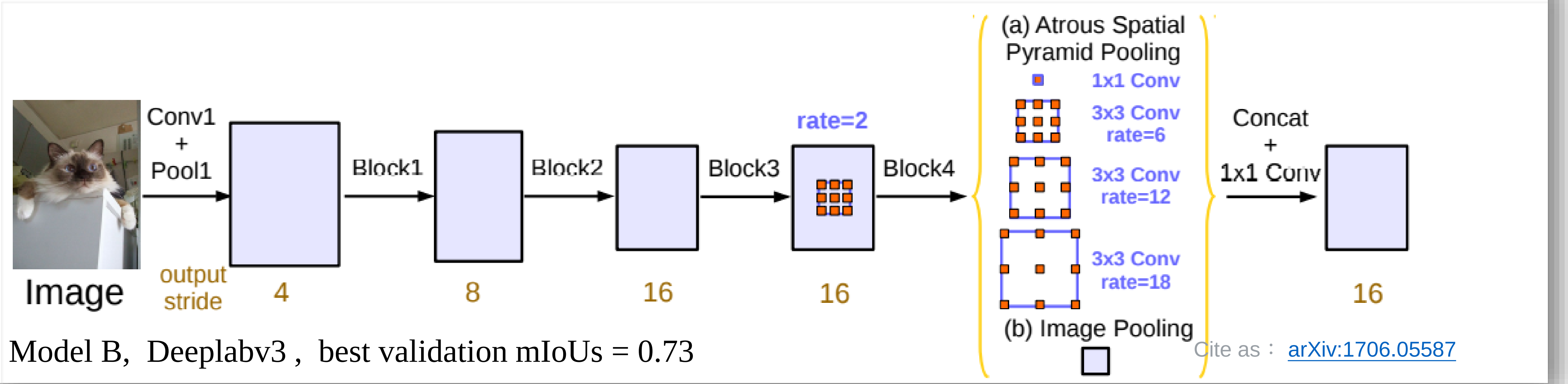
Model A, best validation mIoUs = 0.57

Cite as: arXiv:1608.05971

- Model A adds skip connection between encoder and decoder.

- Model A is CNN-based model and has encoder-decoder structure. Down-sampling in encoder extracts features; Up-sampling in decoder restore image feature maps.

1. **(5%)** Draw the network architecture of your VGG16-FCN32s model (model A).
2. **(5%)** Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model.
3. **(3%)** Report mIoUs of two models on the validation set.



Model B, Deeplabv3 , best validation mIoUs = 0.73

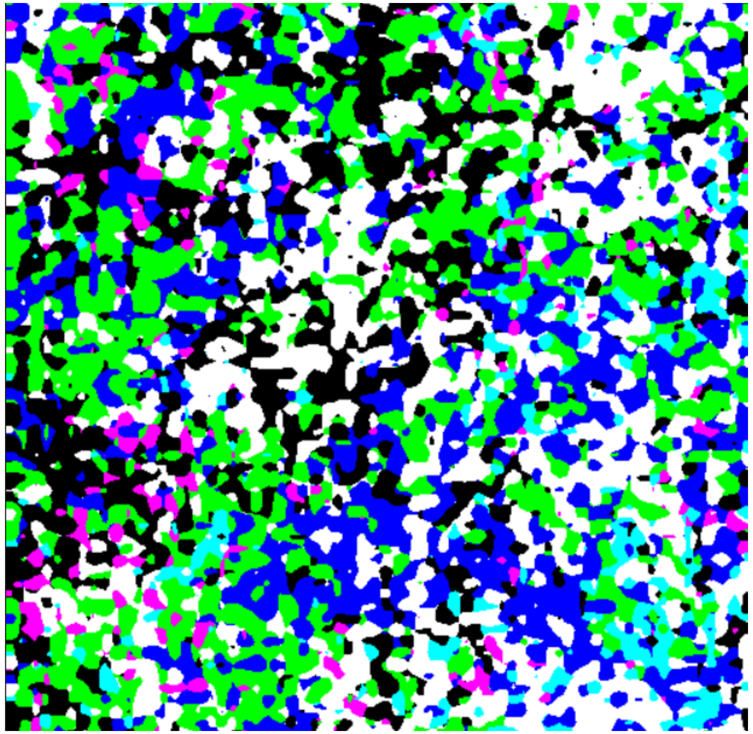
- Model B makes use of Pyramid Pooling Module and Atrous Pyramid Pooling Module to improve the feature extraction ability and avoid information loss in down-sampling, which could happen in model A.

- Model B design cascade and parallel structure for atrous convolution and utilize multiple atrous rates.

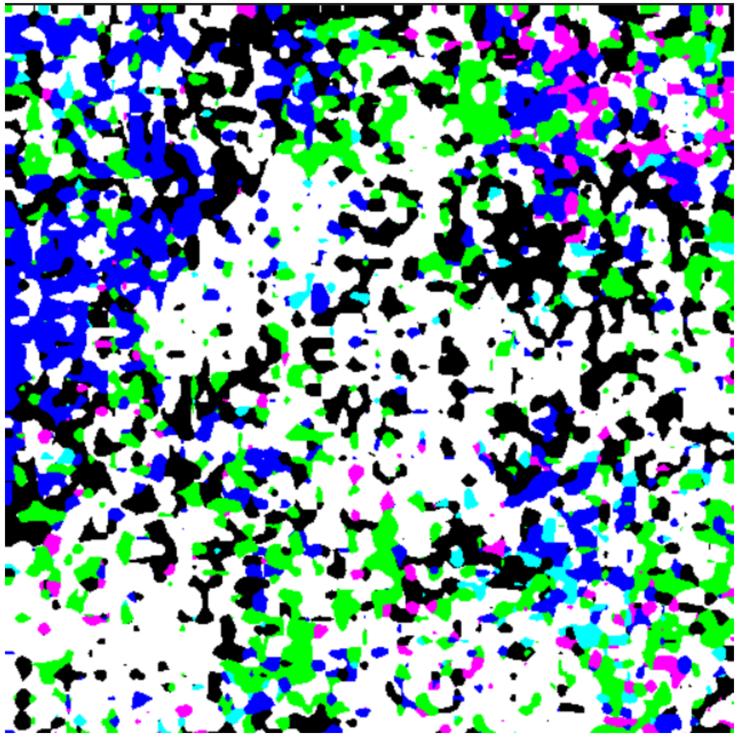
- The backbone of model changes from VGG-16 to ResNet.

(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.

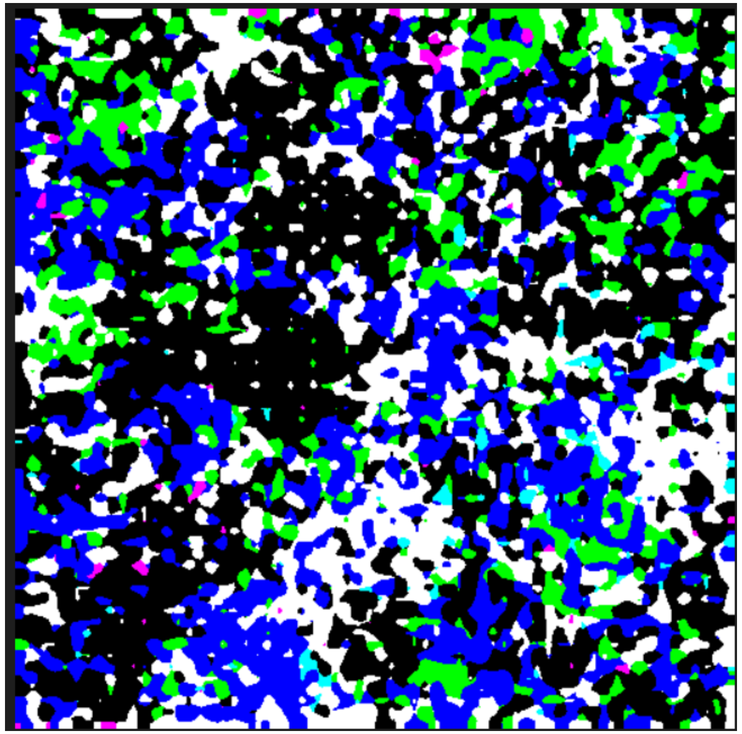
Early stage



0013_sat.png

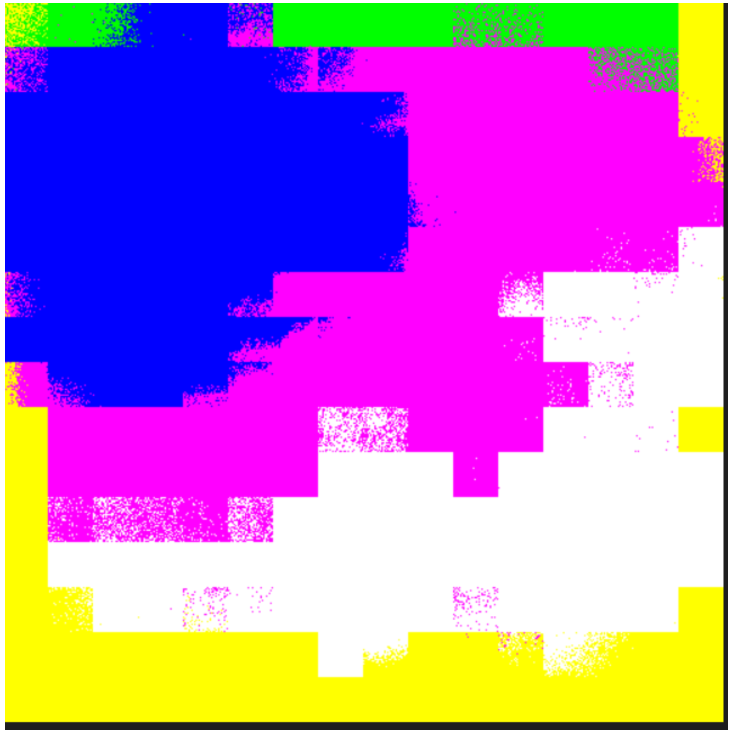


0062_sat.png



0104_sat.png

Middle stage

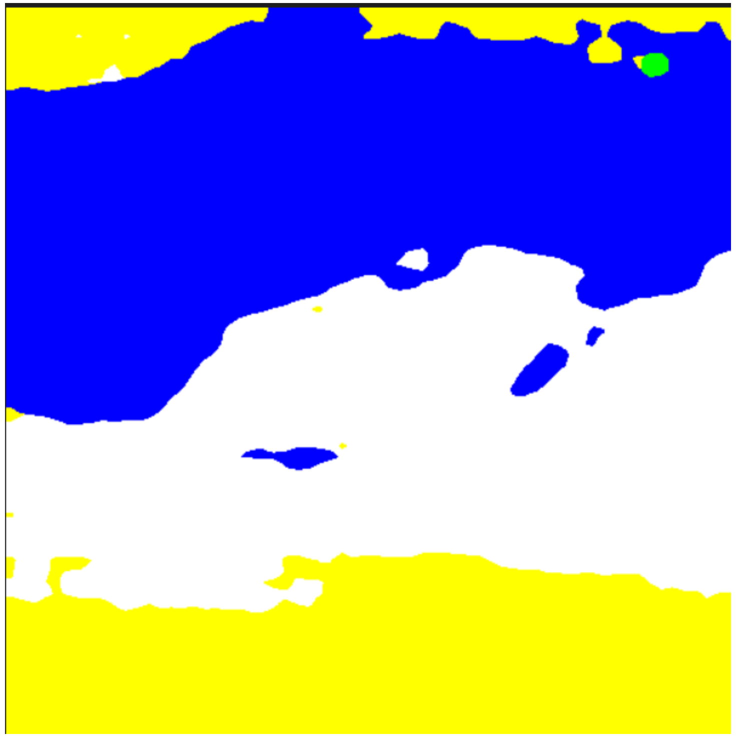


0013_sat.png

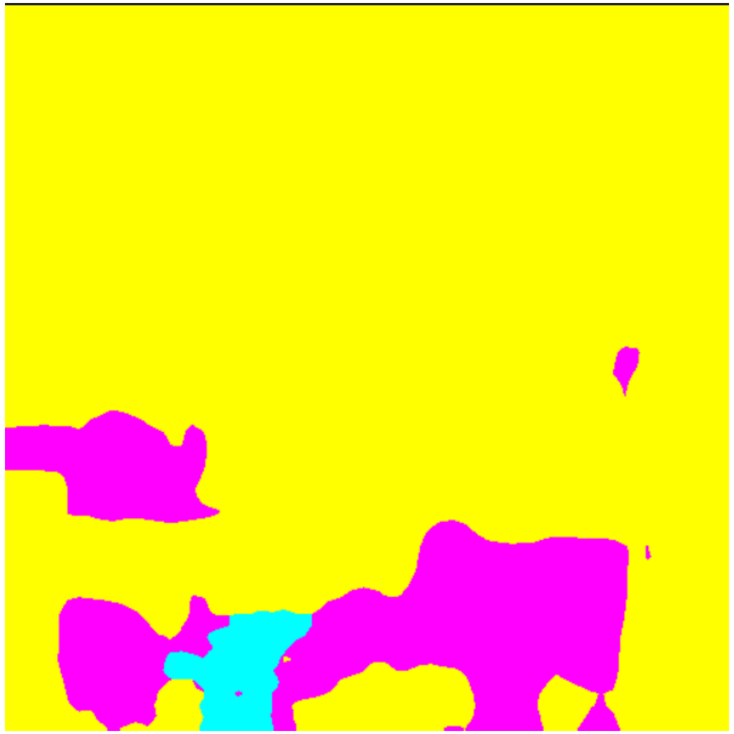
0062_sat.png

0104_sat.png

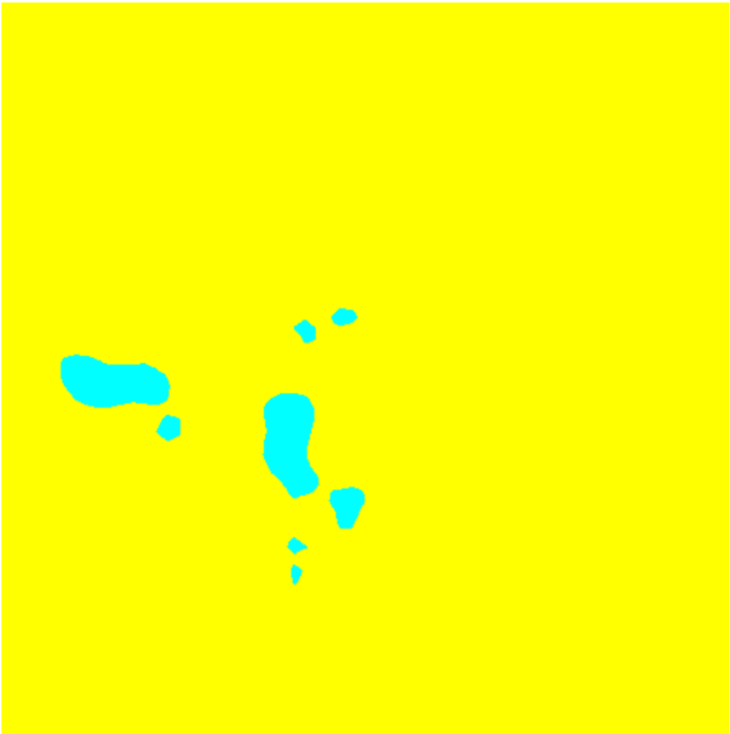
Final stage



0013_sat.png



0062_sat.png



0104_sat.png