Name: 廖宜倫 Dep.:電機四 Student ID:B03901001

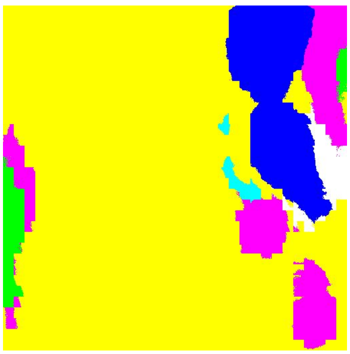
1. (5%) Print the network architecture of your VGG16-FCN32s model.

Name	Layer(Size, Filter)
block1_conv1	Conv2D(64, 3x3)
block1_conv2	Conv2D(64, 3x3)
block1_pool	MaxPool2D(, 2x2)
block2_conv1	Conv2D(128, 3x3)
block2_conv2	Conv2D(128, 3x3)
block2_pool	MaxPool2D(, 2x2)
block3_conv1	Conv2D(256, 3x3)
block3_conv2	Conv2D(256, 3x3)
block3_conv3	Conv2D(256, 3x3)
block3_pool	MaxPool2D(, 2x2)
block4_conv1	Conv2D(512, 3x3)
block4_conv2	Conv2D(512, 3x3)
block4_conv3	Conv2D(512, 3x3)
block4_pool	MaxPool2D(, 2x2)
block5_conv1	Conv2D(512, 3x3)
block5_conv2	Conv2D(512, 3x3)
block5_conv3	Conv2D(512, 3x3)
block5_pool	MaxPool2D(, 2x2)
7x7conv	Conv2D(4096, 7x7)
dropout_1	Dropout(0.5)
fc_conv1	Conv2D(4096, 1x1)
dropout_2	Dropout(0.5)
fc_conv2	Conv2D(7, 1x1)
up32	Conv2DTranspose (7, (64,64), stride=(32,32))
Output (512, 512, 7)	

2. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

0008_sat:

early

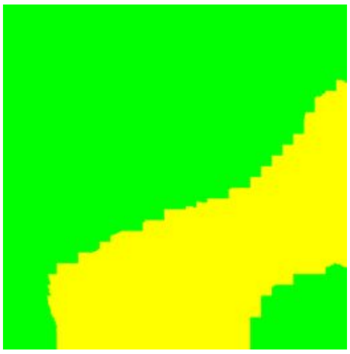
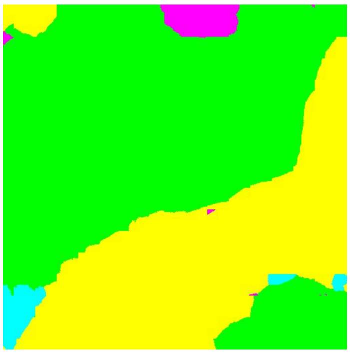
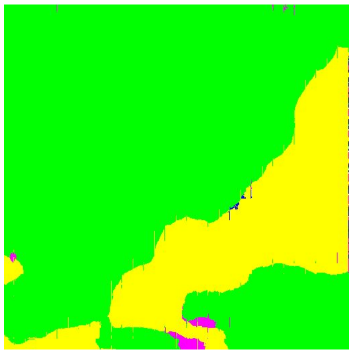


middle

final

0097_sat:

early middle final

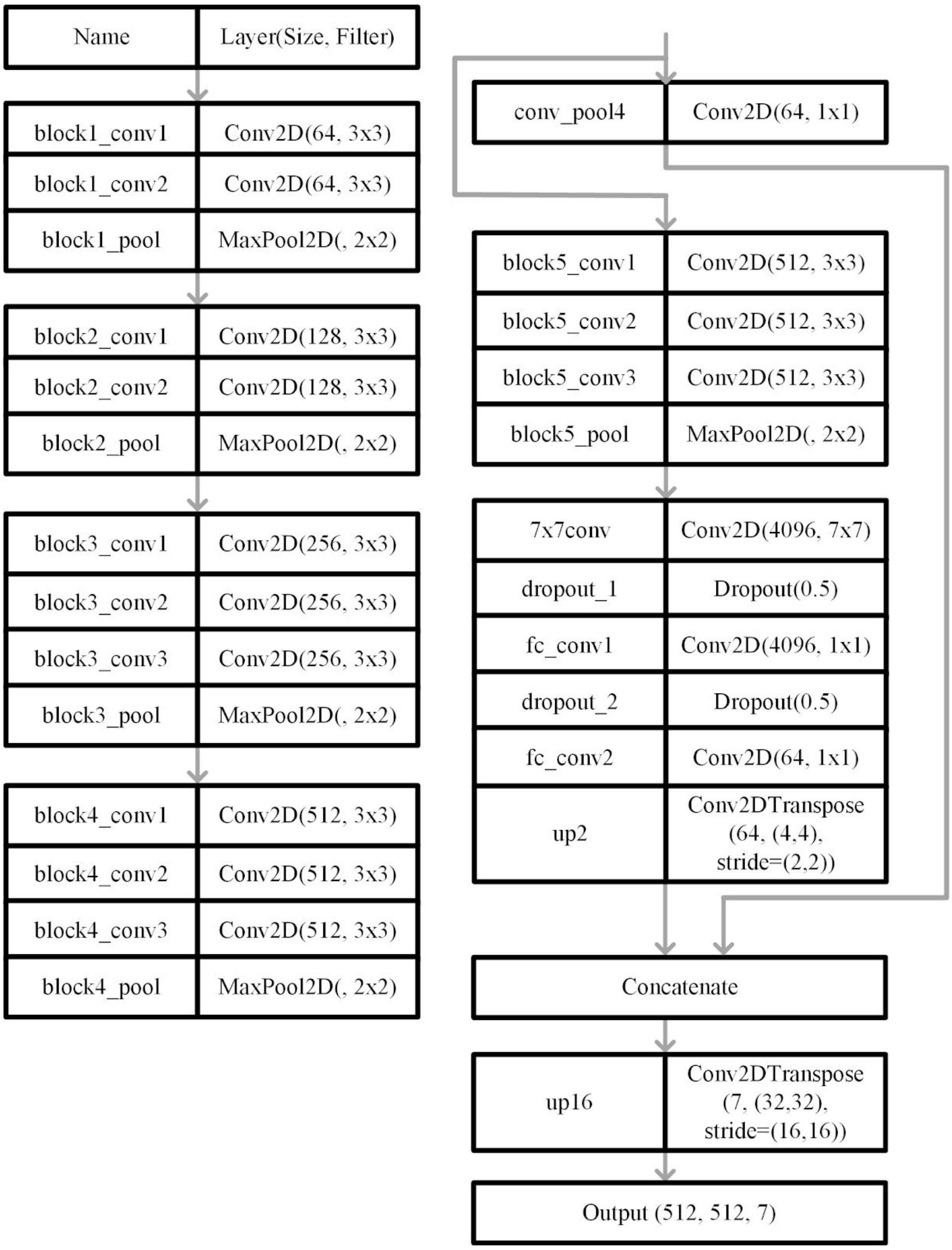
0107_sat:

early middle final



3. (15%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model.

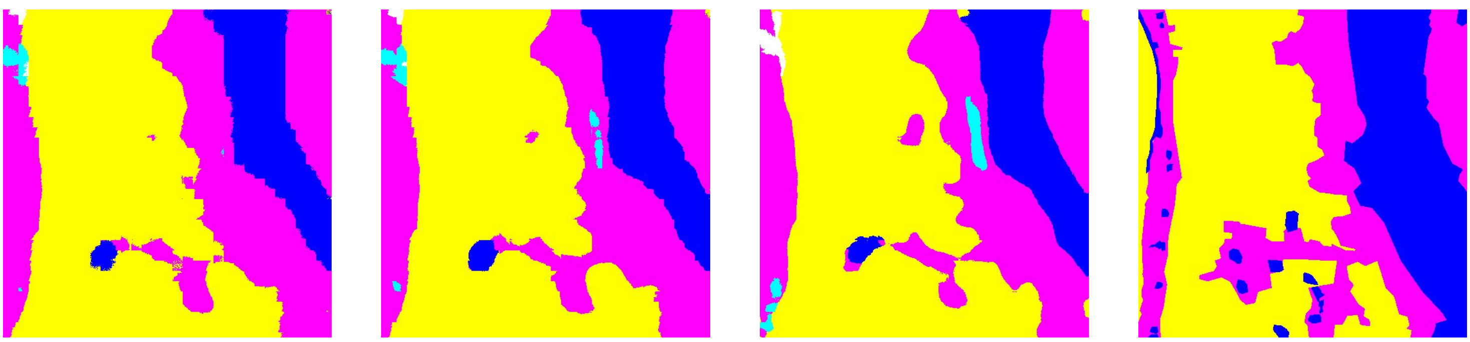
Use 16s structure:



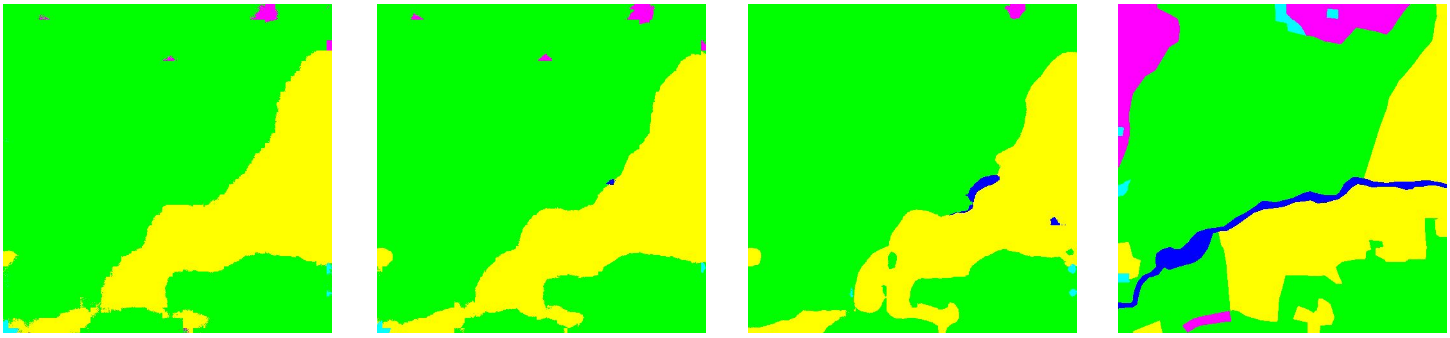
4. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training process of this improved model.

0008_sat:

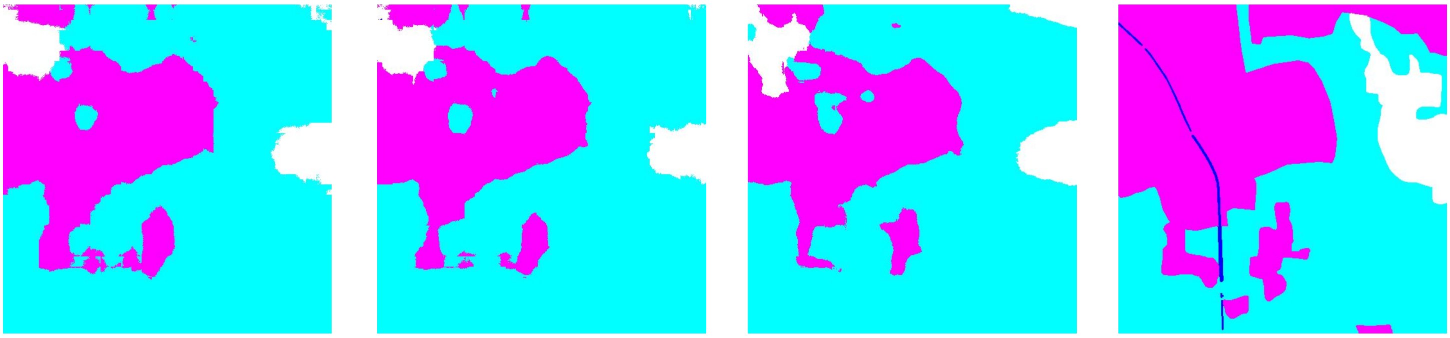
early middle final GT



0097_sat:
early middle final GT



0107_sat:
early middle final GT



5. (15%) Report mIoU score of both models on the validation set. Discuss the reason why the improved model performs better than the baseline one. You may conduct some experiments and show some evidences to support your discussion.

Baseline model:

```
class #0 : 0.74018
class #1 : 0.87071
class #2 : 0.32914
class #3 : 0.80258
class #4 : 0.69369
class #5 : 0.68525

mean_iou: 0.686924
```

Improved model:

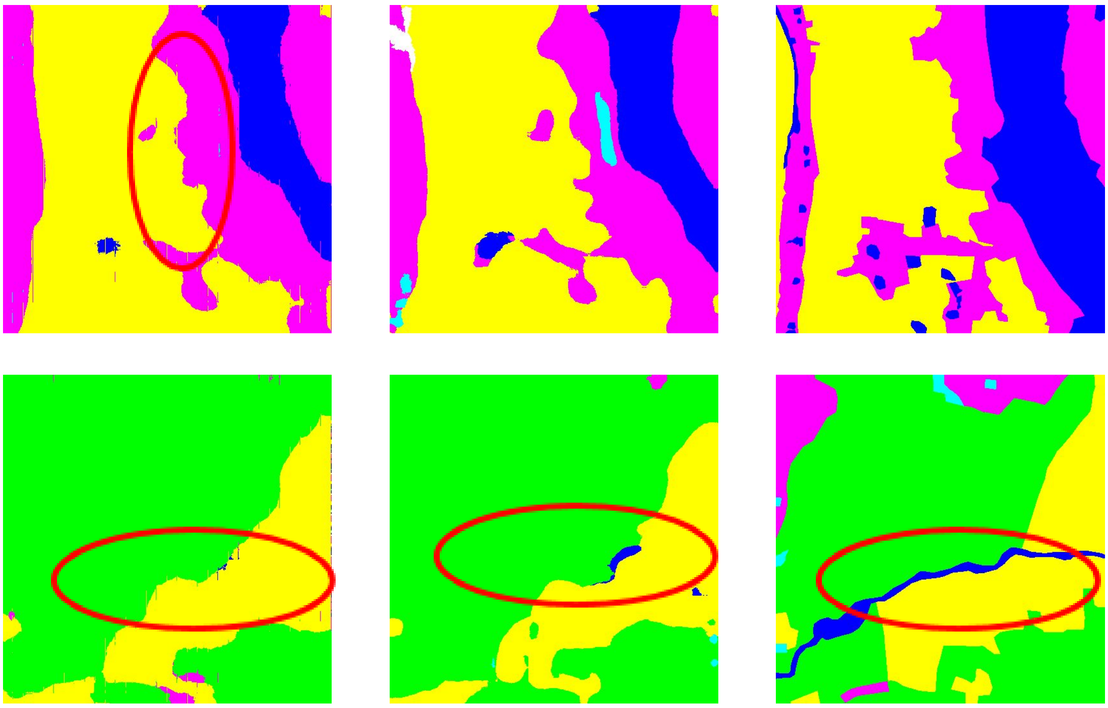
```
class #0 : 0.75550
class #1 : 0.87793
class #2 : 0.33290
class #3 : 0.81124
class #4 : 0.72379
class #5 : 0.69069

mean_iou: 0.698676
```

In the baseline model, input images are fed to a series of layers without any skip paths. the feature maps are compressed into 16-by-16 and then upsampled by 32 times directly. As the number of layers is large, some fine features of input image may be lost in the last few layers. As shown in 2., the predicted masks of the baseline model consist of only majority colors.

Some colors or classes that are not dominant are ignored. For example, the groundtruth mask of 0097_sat consists of a deep blue strip. Since this strip occupies relatively small space, it is simply ignored in the predicted mask. This problem slightly improves when using the improved model. Colors accounting for small portions are more likely to be identified.

The improved model first up-samples 2 times and then up-samples different layers of feature maps by 16 times. As it combines features maps of greater size and consider features near input images, some fine details can be preserved. The predictions suggest that the predicted masks have more winding shapes and more fine structures. Some differences are shown:

baseline | improved | GT



6. (5%) [bonus] Calculate the result of d/dw G(w):

objective function:

$$G(\boldsymbol{w}) = -\sum_n \left[t^{(n)} \log \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w}) + (1 - t^n) \log\left(1 - \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w})\right)\right] \geq 0$$

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} G(\boldsymbol{w})$$ choose the weights that minimise the network's surprise about the training data

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{w}} G(\boldsymbol{w}) = \sum_n \frac{\mathrm{d}G(\boldsymbol{w})}{\mathrm{d}x^{(n)}} \frac{\mathrm{d}x^{(n)}}{\mathrm{d}\boldsymbol{w}} = -\sum_n (t^{(n)} - x^{(n)}) \boldsymbol{z}^{(n)}$$ = prediction error x feature

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{w}} G(\boldsymbol{w})$$ iteratively step down the objective (gradient points up hill) 39

for a certain n:

$$\frac{dG}{dx^n}\frac{dx^n}{dw} = \frac{dG}{dw}$$

$$G = -(t^n \log x^n + (1-t^n)\log(1-x^n))$$

$$\frac{dG}{dx^n} = -\left[t^n \frac{1}{x^n} + (1-t^n)\frac{1}{1-x^n}(-1)\right]$$

$$= -\left(\frac{t^n - x^n}{(1-x^n)\, x^n}\right)$$

Let $\alpha = w^T z$, $x =$ sigmoid(α), $\frac{dx}{d\alpha} = (1-x)x$

$$\frac{dx^n}{dw} = \frac{dx^n}{d\alpha}\frac{d\alpha}{dw}$$

$$= x^n(1-x^n)z^n$$

$$\therefore \frac{dG}{dw} = \frac{-(t^n - x^n)}{\cancel{(1-x^n)}\cancel{x^n}}\cancel{x^n}\cancel{(1-x^n)}z^n$$

$$= -(t^n - x^n)z^n$$

Summing over n: $\frac{d}{dw}G = -\sum_n (t^n - x^n) z^n$