DLCV HW3
Name: 何吉瑞 Dep.: 電機三 Student ID:

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

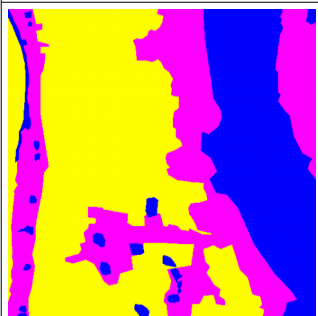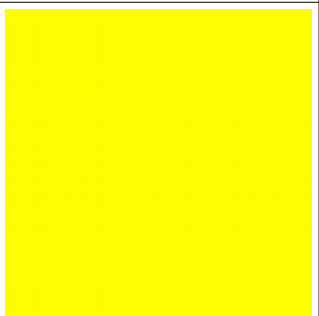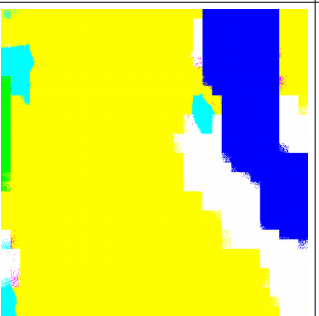| Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 512, 512, 3)	0
block1_conv1 (Conv2D)	(None, 512, 512, 64)	1792
block1_conv2 (Conv2D)	(None, 512, 512, 64)	36928
block1_pool (MaxPooling2D)	(None, 256, 256, 64)	0
block2_conv1 (Conv2D)	(None, 256, 256, 128)	73856
block2_conv2 (Conv2D)	(None, 256, 256, 128)	147584
block2_pool (MaxPooling2D)	(None, 128, 128, 128)	0
block3_conv1 (Conv2D)	(None, 128, 128, 256)	295168
block3_conv2 (Conv2D)	(None, 128, 128, 256)	590080
block3_conv3 (Conv2D)	(None, 128, 128, 256)	590080
block3_pool (MaxPooling2D)	(None, 64, 64, 256)	0
block4_conv1 (Conv2D)	(None, 64, 64, 512)	1180160
block4_conv2 (Conv2D)	(None, 64, 64, 512)	2359808
block4_conv3 (Conv2D)	(None, 64, 64, 512)	2359808
block4_pool (MaxPooling2D)	(None, 32, 32, 512)	0
block5_conv1 (Conv2D)	(None, 32, 32, 512)	2359808
block5_conv2 (Conv2D)	(None, 32, 32, 512)	2359808
block5_conv3 (Conv2D)	(None, 32, 32, 512)	2359808
block5_pool (MaxPooling2D)	(None, 16, 16, 512)	0
fc_1 (Conv2D)	(None, 16, 16, 4096)	8392704
fc_2 (Conv2D)	(None, 16, 16, 4096)	16781312
conv2d_1 (Conv2D)	(None, 16, 16, 7)	28679
conv2d_transpose_1 (Conv2DTr	(None, 512, 512, 7)	200704
reshape_1 (Reshape)	(None, 262144, 7)	0
activation_1 (Activation)	(None, 262144, 7)	0

Total params: 40,118,087
Trainable params: 40,118,087
Non-trainable params: 0

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)
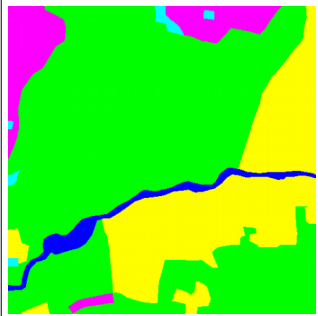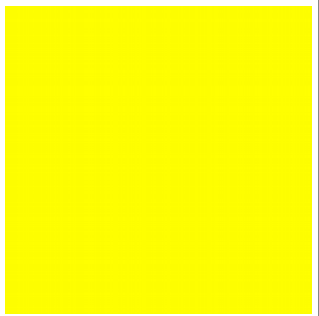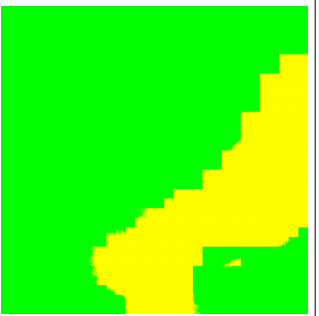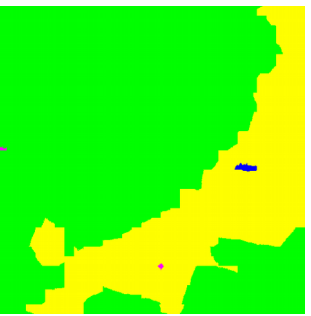
Here shows the result of 1st, 10th, and 40th epoch

validation/0008_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

validation/0097_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

validation/0107_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

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

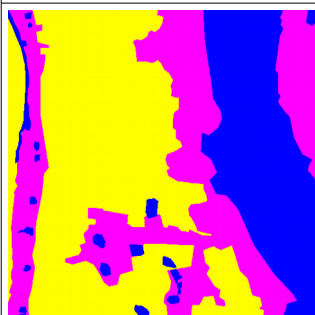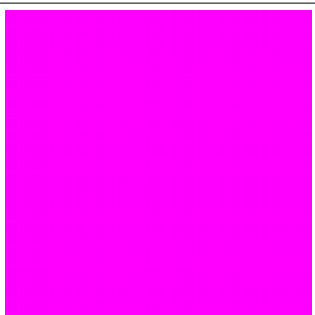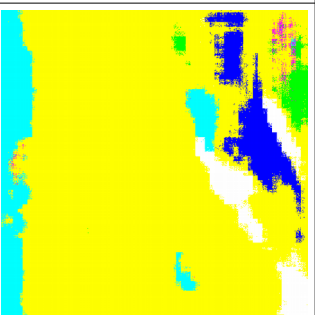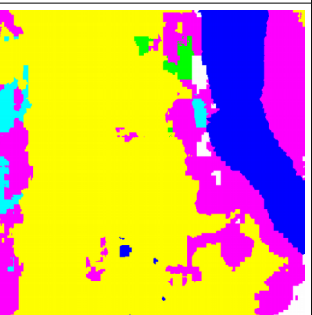I implemented VGG16-FCN8 model to improve the performance

```
Layer (type)                     Output Shape         Param #     Connected to
==================================================================================================
input_1 (InputLayer)             (None, 512, 512, 3)  0

block1_conv1 (Conv2D)            (None, 512, 512, 64) 1792        input_1[0][0]

block1_conv2 (Conv2D)            (None, 512, 512, 64) 36928       block1_conv1[0][0]

block1_pool (MaxPooling2D)       (None, 256, 256, 64) 0           block1_conv2[0][0]

block2_conv1 (Conv2D)            (None, 256, 256, 128 73856       block1_pool[0][0]

block2_conv2 (Conv2D)            (None, 256, 256, 128 147584      block2_conv1[0][0]

block2_pool (MaxPooling2D)       (None, 128, 128, 128 0           block2_conv2[0][0]

block3_conv1 (Conv2D)            (None, 128, 128, 256 295168      block2_pool[0][0]

block3_conv2 (Conv2D)            (None, 128, 128, 256 590080      block3_conv1[0][0]

block3_conv3 (Conv2D)            (None, 128, 128, 256 590080      block3_conv2[0][0]

block3_pool (MaxPooling2D)       (None, 64, 64, 256)  0           block3_conv3[0][0]

block4_conv1 (Conv2D)            (None, 64, 64, 512)  1180160     block3_pool[0][0]

block4_conv2 (Conv2D)            (None, 64, 64, 512)  2359808     block4_conv1[0][0]

block4_conv3 (Conv2D)            (None, 64, 64, 512)  2359808     block4_conv2[0][0]

block4_pool (MaxPooling2D)       (None, 32, 32, 512)  0           block4_conv3[0][0]

block5_conv1 (Conv2D)            (None, 32, 32, 512)  2359808     block4_pool[0][0]

block5_conv2 (Conv2D)            (None, 32, 32, 512)  2359808     block5_conv1[0][0]

block5_conv3 (Conv2D)            (None, 32, 32, 512)  2359808     block5_conv2[0][0]

block5_pool (MaxPooling2D)       (None, 16, 16, 512)  0           block5_conv3[0][0]

fc_1 (Conv2D)                    (None, 16, 16, 4096) 8392704     block5_pool[0][0]

fc_2 (Conv2D)                    (None, 16, 16, 4096) 16781312    fc_1[0][0]

conv2d_1 (Conv2D)                (None, 16, 16, 7)    28679       fc_2[0][0]

conv2d_transpose_1 (Conv2DTrans  (None, 32, 32, 7)    784         conv2d_1[0][0]

conv2d_2 (Conv2D)                (None, 32, 32, 7)    3591        block4_pool[0][0]

add_1 (Add)                      (None, 32, 32, 7)    0           conv2d_transpose_1[0][0]
                                                                  conv2d_2[0][0]

conv2d_transpose_2 (Conv2DTrans  (None, 64, 64, 7)    784         add_1[0][0]

conv2d_3 (Conv2D)                (None, 64, 64, 7)    1799        block3_pool[0][0]

add_2 (Add)                      (None, 64, 64, 7)    0           conv2d_transpose_2[0][0]
                                                                  conv2d_3[0][0]

conv2d_transpose_3 (Conv2DTrans  (None, 512, 512, 7)  12544       add_2[0][0]

reshape_1 (Reshape)              (None, 262144, 7)    0           conv2d_transpose_3[0][0]

activation_1 (Activation)        (None, 262144, 7)    0           reshape_1[0][0]
==================================================================================================
Total params: 39,936,885
Trainable params: 39,936,885
Non-trainable params: 0
```

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
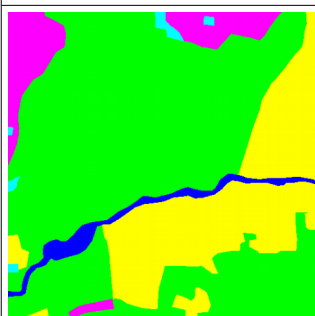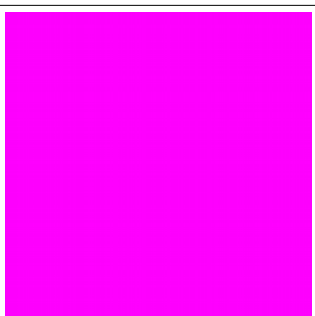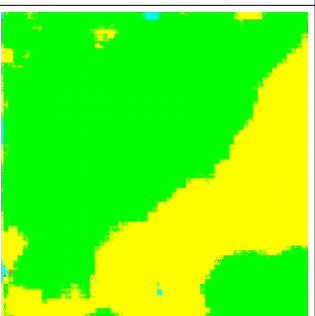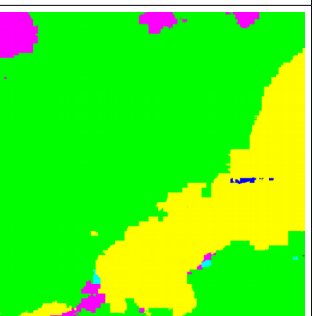
final stage during the training process of this improved model.

The raw image was shown in problem 2. Here I just show the predicted segmentation mask.
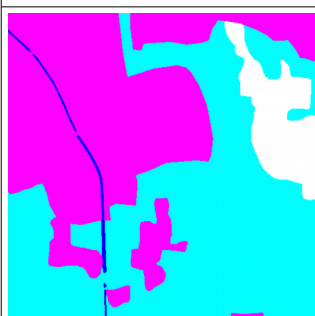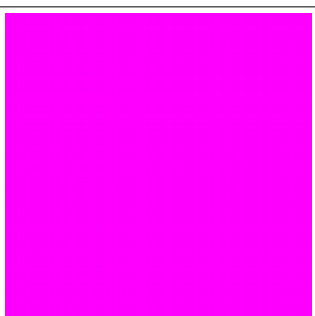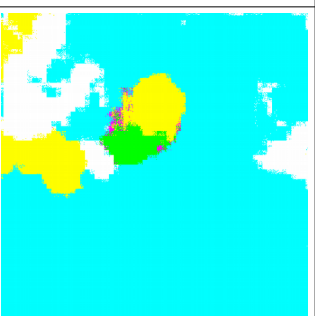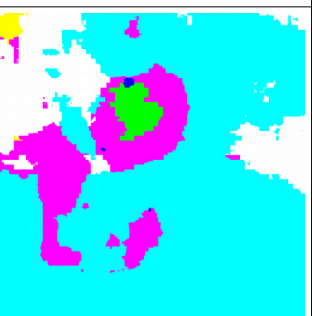
validation/0008_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

validation/0097_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

validation/0107_sat.jpg

Ground truth	1st epoch	10th epoch	40th epoch

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(VGG16-FCN32)

```
class #0 : 0.72855
class #1 : 0.87541
class #2 : 0.27936
class #3 : 0.77845
class #4 : 0.72438
class #5 : 0.64602

mean_iou: 0.672028
```

improved model(VGG16-FCN8)

```
class #0 : 0.76043
class #1 : 0.87120
class #2 : 0.34998
class #3 : 0.79988
class #4 : 0.73222
class #5 : 0.63884

mean_iou: 0.692094
```

The reason may be that the improved model is able to predict the mask in more detail because the number of maxpooling layers is less than that of baseline model by two. In the figures, it can be observed that the resolution in the mask of improved model is higher than that of baseline model.

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

DL CV bonus

b045D7009 電機三 佶瑝

$$x_i = \frac{1}{1+e^{-s_i}}, \quad \text{where} \quad S_i = \sum_j W_{ji} \delta_j$$

$$\frac{\partial G}{\partial w_{ji}} = \frac{\partial G}{\partial x_i} \cdot \frac{\partial x_i}{\partial w_{ji}}$$

$$= \frac{\partial G}{\partial x_i} \cdot \frac{\partial x_i}{\partial s_i} \cdot \frac{\partial s_i}{\partial w_{ji}}$$

$$= \left(\frac{-t_i}{x_i} + \frac{1-t_i}{1-x_i} \right) \left(x_i(1-x_i) \right) \delta_j$$

$$= (x_i - t_i) \delta_j$$

$$\therefore \frac{\partial G}{\partial w} = -\sum_n \left(t^{(n)} - x^{(n)} \right) \delta^{(n)}$$