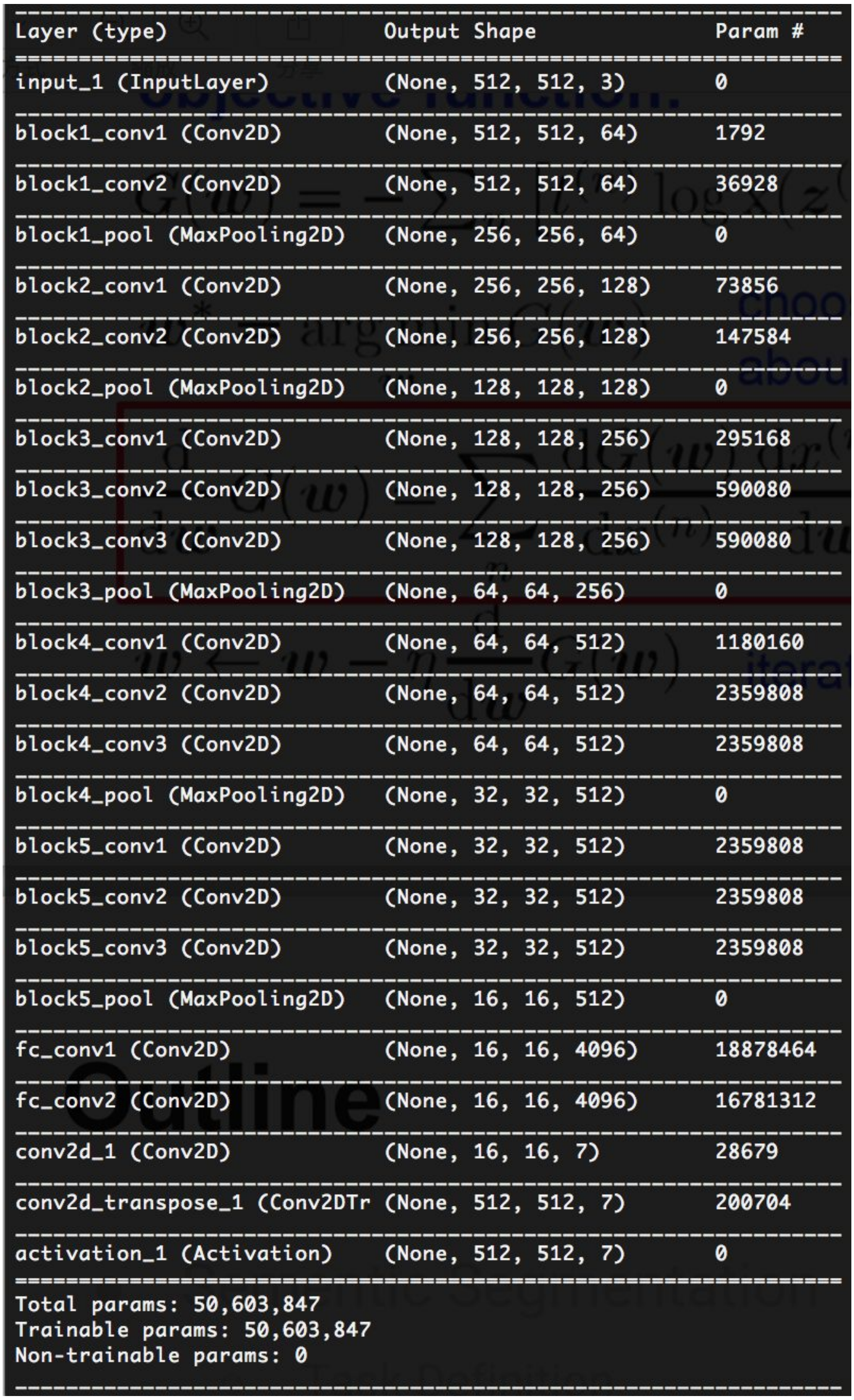
DLCV HW3 Report

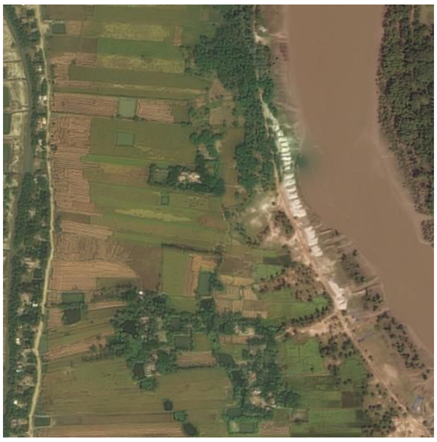
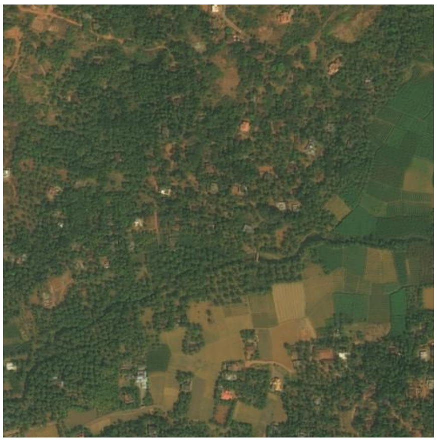
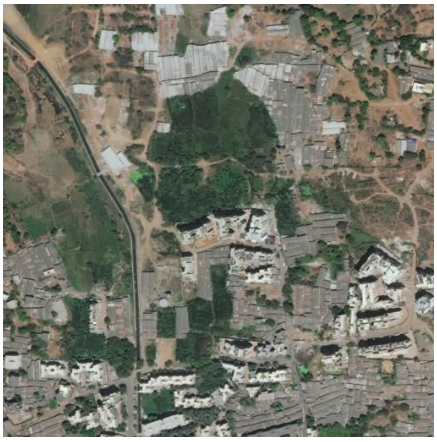
Name: 張景程　Dep.:電機三　Student ID:B04901138

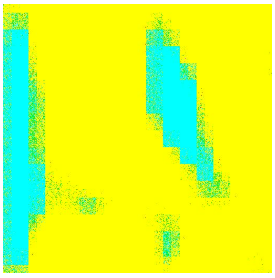
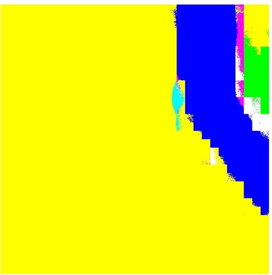
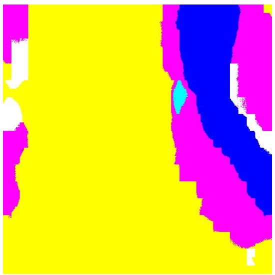
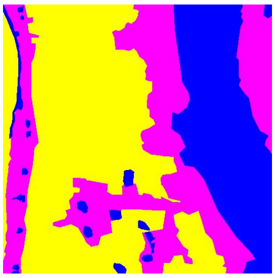
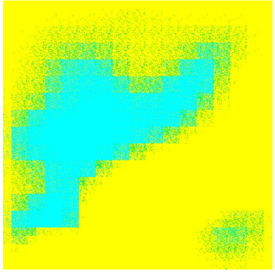
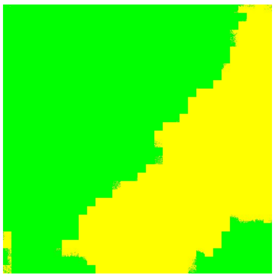
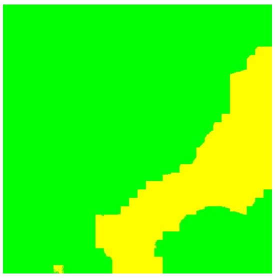
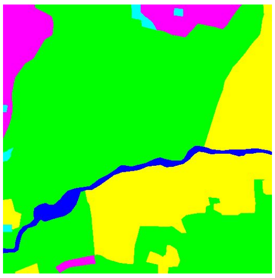
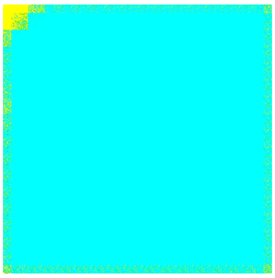
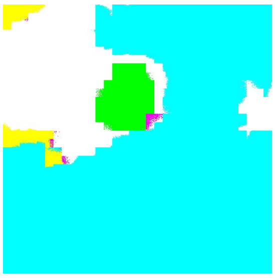
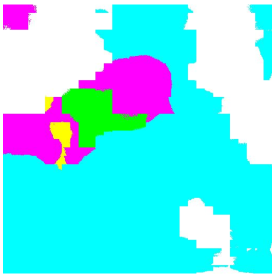
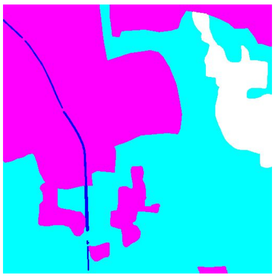
1. (5%) Print the network architecture of your VGG16-FCN32s model.



```
_________________________________________________________________
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_conv1 (Conv2D)            (None, 16, 16, 4096)      18878464  
_________________________________________________________________
fc_conv2 (Conv2D)            (None, 16, 16, 4096)      16781312  
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 16, 16, 7)         28679     
_________________________________________________________________
conv2d_transpose_1 (Conv2DTr (None, 512, 512, 7)       200704    
_________________________________________________________________
activation_1 (Activation)    (None, 512, 512, 7)       0         
=================================================================
Total params: 50,603,847
Trainable params: 50,603,847
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)

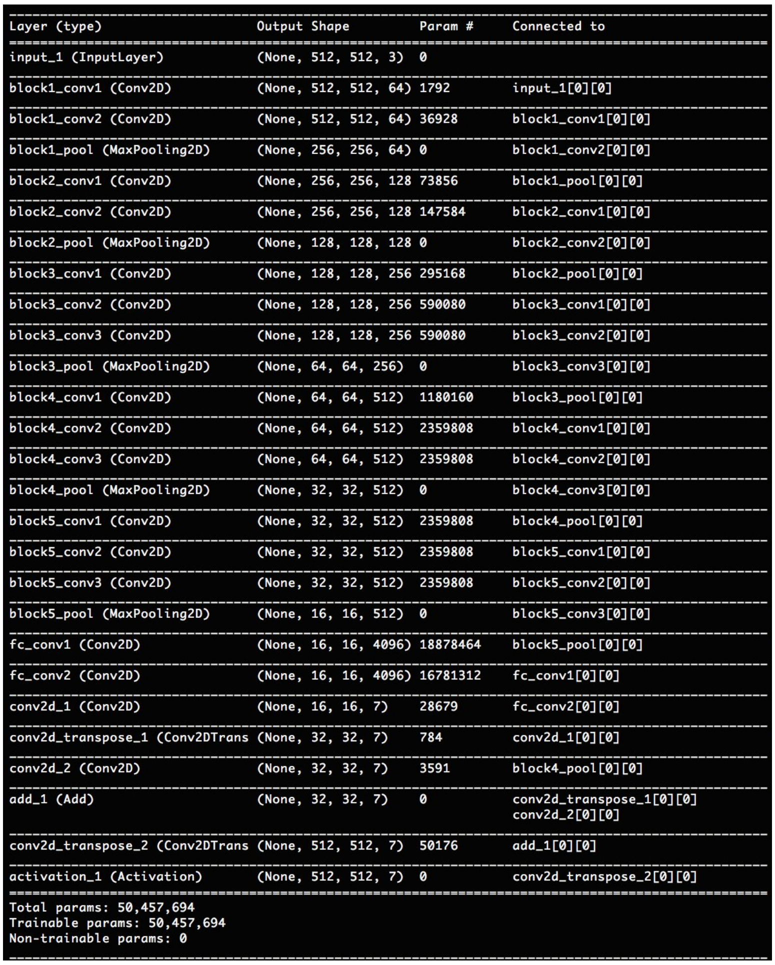
0008_sat.jpg	0097_sat.jpg	0107_sat.jpg
		

	epoch 2	epoch10	final epoch	ground truth
0008_ mask.jpg				
0097_ mask.jpg				
0107_ mask.jpg				

(final epoch指的是training時earlystopping前的model，epoch 1的時候model還沒學到什麼東西，整張圖幾乎都同個顏色，因此這裡以epoch 2的output來做比較，Q4比較不同model的部分也同理)

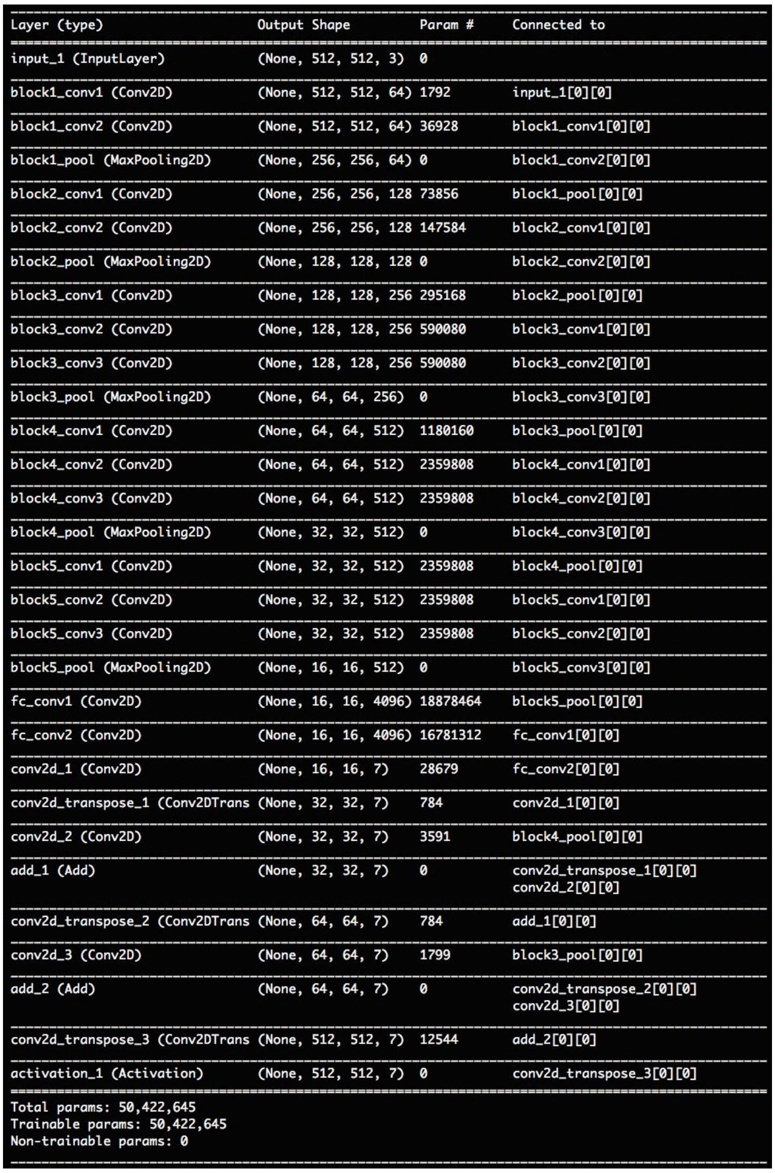
3. (15%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model.

(1)VGG16-FCN16s :



```
__________________________________________________________________________________________________
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_conv1 (Conv2D)               (None, 16, 16, 4096) 18878464    block5_pool[0][0]
__________________________________________________________________________________________________
fc_conv2 (Conv2D)               (None, 16, 16, 4096) 16781312    fc_conv1[0][0]
__________________________________________________________________________________________________
conv2d_1 (Conv2D)               (None, 16, 16, 7)    28679       fc_conv2[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, 512, 512, 7)  50176       add_1[0][0]
__________________________________________________________________________________________________
activation_1 (Activation)       (None, 512, 512, 7)  0           conv2d_transpose_2[0][0]
==================================================================================================
Total params: 50,457,694
Trainable params: 50,457,694
Non-trainable params: 0
__________________________________________________________________________________________________
```

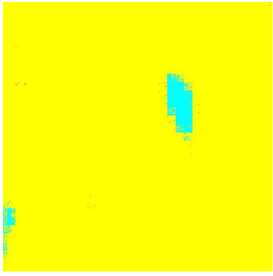
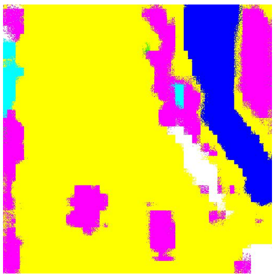
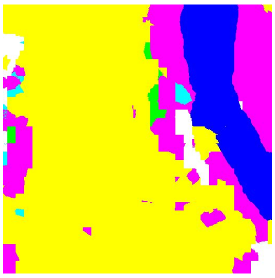
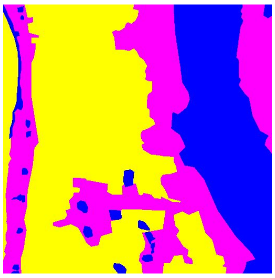
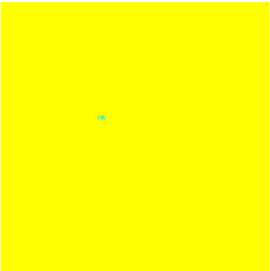
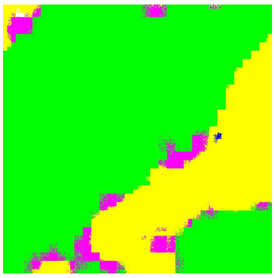
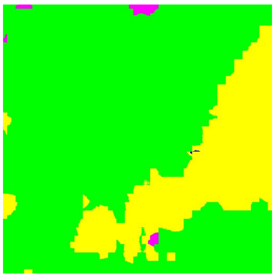
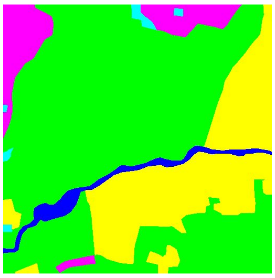
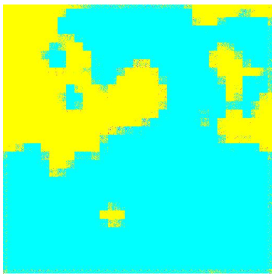
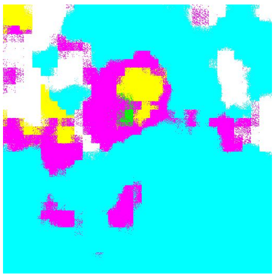
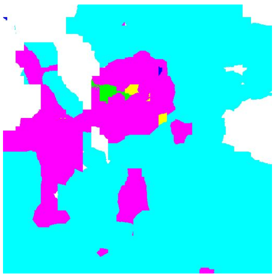
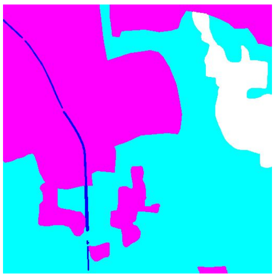
(2) VGG16-FCN8s :



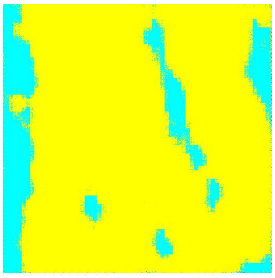
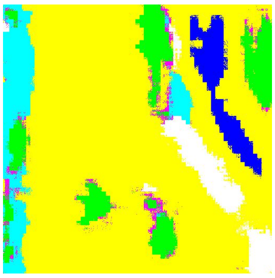
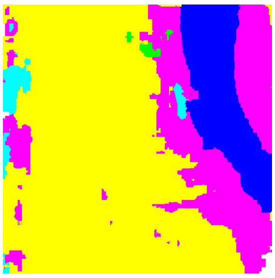
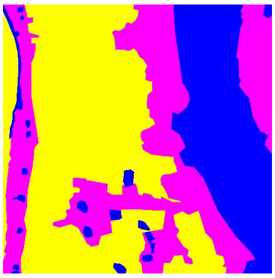
```
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_1 (InputLayer)            (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_conv1 (Conv2D)               (None, 16, 16, 4096) 18878464    block5_pool[0][0]
__________________________________________________________________________________________________
fc_conv2 (Conv2D)               (None, 16, 16, 4096) 16781312    fc_conv1[0][0]
__________________________________________________________________________________________________
conv2d_1 (Conv2D)               (None, 16, 16, 7)    28679       fc_conv2[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]
__________________________________________________________________________________________________
activation_1 (Activation)       (None, 512, 512, 7)  0           conv2d_transpose_3[0][0]
==================================================================================================
Total params: 50,422,645
Trainable params: 50,422,645
Non-trainable params: 0
__________________________________________________________________________________________________
```

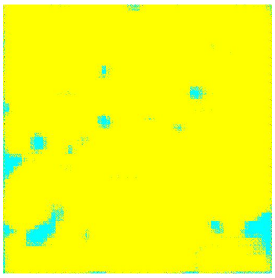
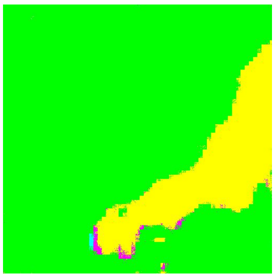
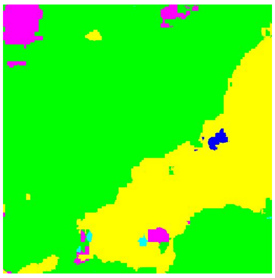
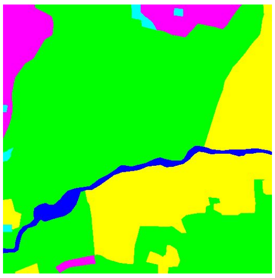
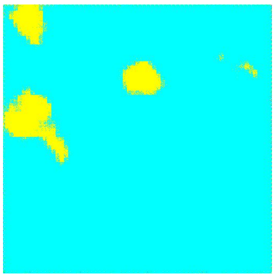
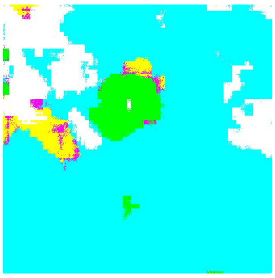
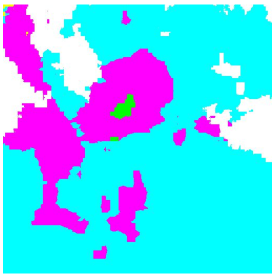
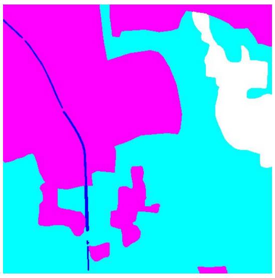
4. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training process of this improved model.

VGG16-FCN16s :

	epoch 2	epoch 10	final epoch	ground truth
0008_ mask.jpg				
0097_ mask.jpg				
0107_ mask.jpg				

VGG16-FCN8s :

	epoch 2	epoch 10	final epoch	ground truth
0008_ mask.jpg				

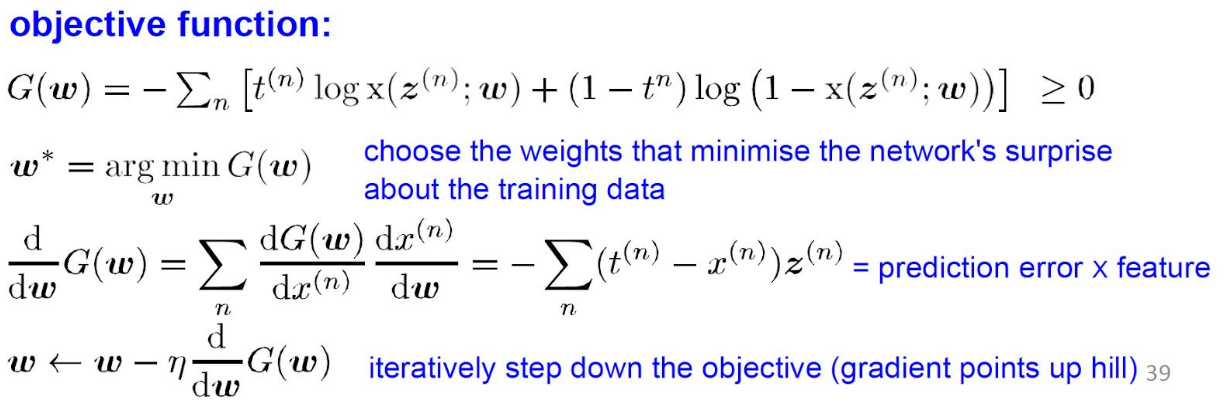
0097_ mask.jpg				
0107_ mask.jpg				

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.

	FCN_32s	FCN_16s	FCN_8s
class #0	0.71774	0.74861	0.75908
class #1	0.87164	0.86890	0.87564
class #2	0.25933	0.29304	0.34733
class #3	0.74692	0.76598	0.78900
class #4	0.70425	0.74807	0.73500
class #5	0.65134	0.64684	0.64123
mean_IoU	0.658988	0.678574	0.691213

由上面第四小題的output圖片比較可以看出，和baseline的32s model的結果相比，16s和8s的結果好上許多，而8s的又比16s稍好一些，從mean_IoU值來看也是相同結果，推測是因為16s和8s的model架構中都有將前面幾層還沒經過MaxPooling壓縮的layer接到後面來，因此可以保存更多圖片的細節，使得在做segmentaion時可以將細微的地方切得更好，從圖片也可觀察到，比起FCN_32s，這兩個improved model 的結果多了更多的細節，像是一些零星的有色碎塊，而不是整大片都相同的顏色，這都會使得mean_IoU的值有所提升。

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



objective function:

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

$$\boldsymbol{w}^{*} = \arg\min_{\boldsymbol{w}} G(\boldsymbol{w})$$

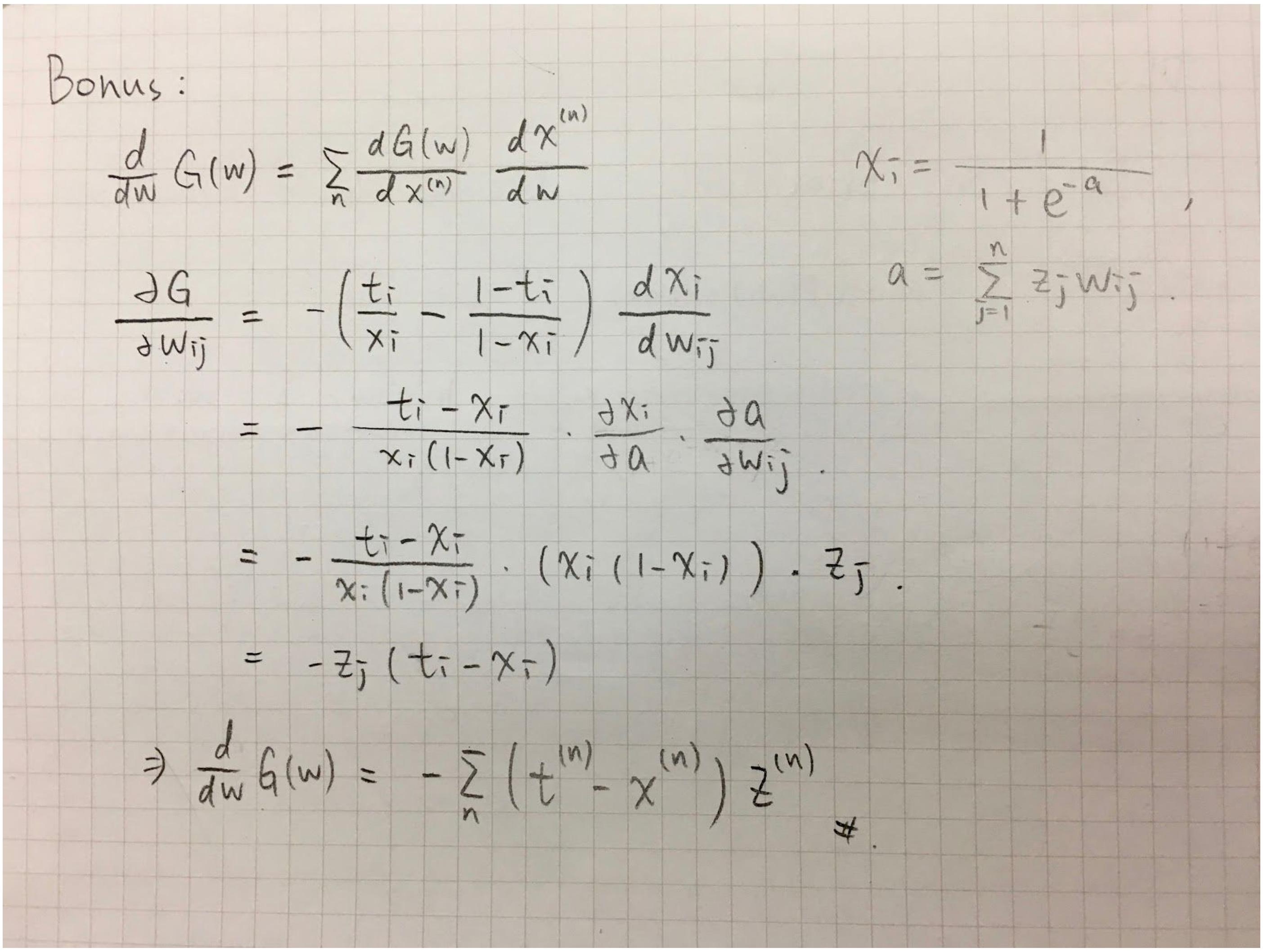
choose the weights that minimise the network's surprise about the training data

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

= 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



Bonus:

$$\frac{d}{dw} G(w) = \sum_n \frac{dG(w)}{dx^{(n)}} \frac{dx^{(n)}}{dw} \qquad x_i = \frac{1}{1+e^{-a}},$$

$$a = \sum_{j=1}^{n} z_j w_{ij} .$$

$$\frac{\partial G}{\partial w_{ij}} = -\left(\frac{t_i}{x_i} - \frac{1-t_i}{1-x_i}\right) \frac{dx_i}{dw_{ij}}$$

$$= - \frac{t_i - x_i}{x_i(1-x_i)} \cdot \frac{\partial x_i}{\partial a} \cdot \frac{\partial a}{\partial w_{ij}} .$$

$$= - \frac{t_i - x_i}{x_i(1-x_i)} \cdot (x_i(1-x_i)) \cdot z_j .$$

$$= -z_j(t_i - x_i)$$

$$\Rightarrow \frac{d}{dw} G(w) = -\sum_n \left(t^{(n)} - x^{(n)}\right) z^{(n)}$$

#.