Please use this report template, and upload it in the PDF format. Reports in other forms/formats will result in ZERO point. Reports written in either Chinese or English is acceptable. The length of your report should NOT exceed 6 pages (**excluding bonus**).

Name: 鄭閎 Dep.:電機三 Student ID:B04901155

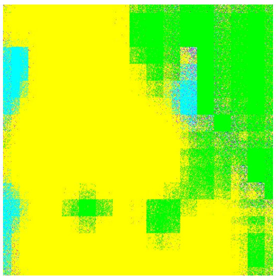
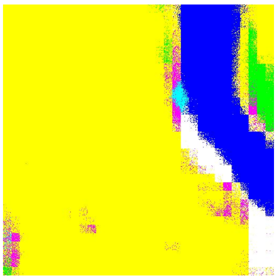
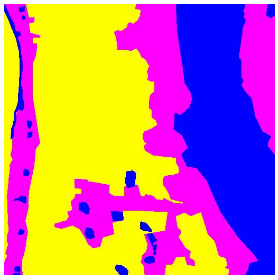
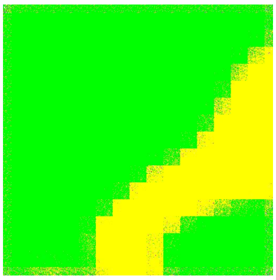
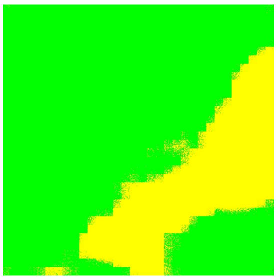
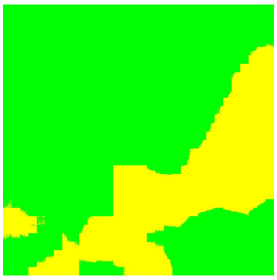
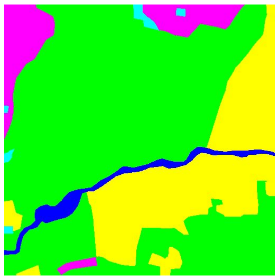
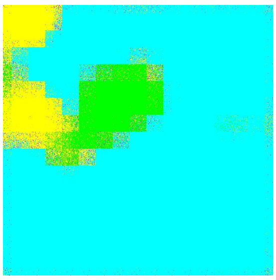
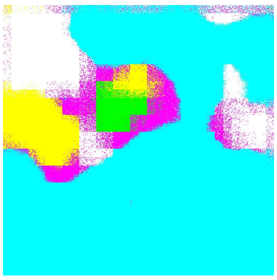
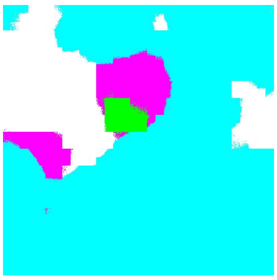
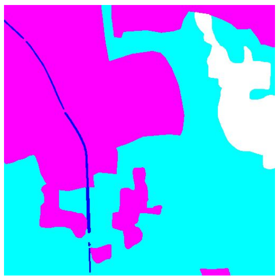
1. (5%) Print the network architecture of your VGG16-FCN32s model.

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
fc1 (Conv2D)	(None, 16, 16, 4096)	8392704
fc2 (Conv2D)	(None, 16, 16, 4096)	16781312
conv2d_1 (Conv2D)	(None, 16, 16, 7)	28679
conv2d_transpose_1 (Conv2DTr	(None, 512, 512, 7)	200704

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)

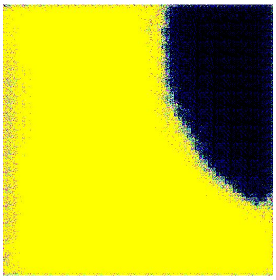
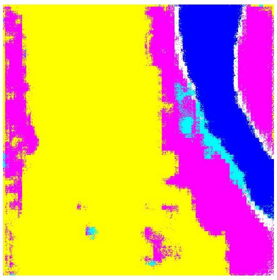
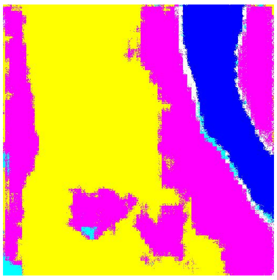
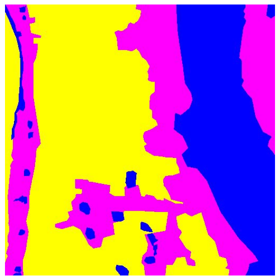
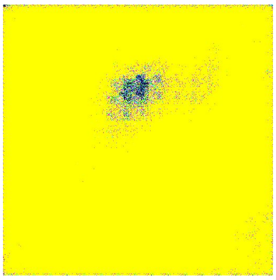
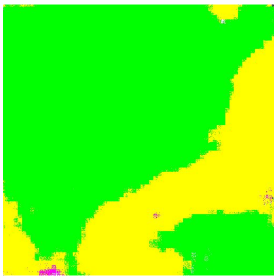
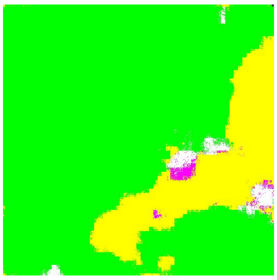
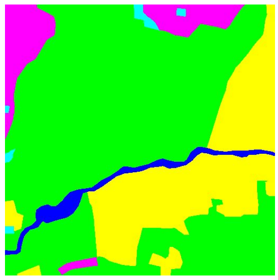
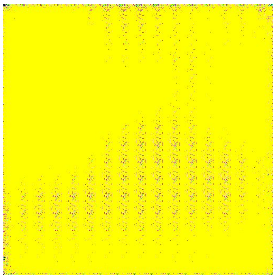
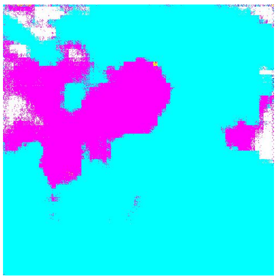
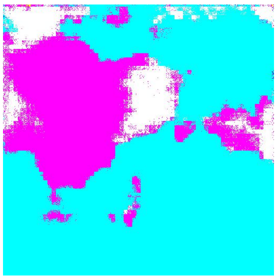
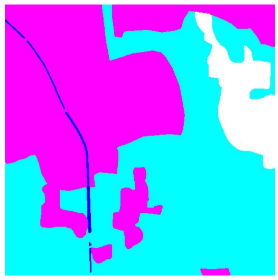
		

epoch 1	epoch 10	epoch 20	ground truth
			
			
			

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

```
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]
__________________________________________________________________________________________________
fc1 (Conv2D)                     (None, 16, 16, 4096) 102764544   block5_pool[0][0]
__________________________________________________________________________________________________
fc2 (Conv2D)                     (None, 16, 16, 4096) 16781312    fc1[0][0]
__________________________________________________________________________________________________
conv2d_1 (Conv2D)                (None, 16, 16, 7)    28679       fc2[0][0]
__________________________________________________________________________________________________
conv2d_transpose_1 (Conv2DTrans  (None, 32, 32, 7)    784         conv2d_1[0][0]
__________________________________________________________________________________________________
16sforward (Conv2D)              (None, 32, 32, 7)    3591        block4_pool[0][0]
__________________________________________________________________________________________________
add_1 (Add)                      (None, 32, 32, 7)    0           conv2d_transpose_1[0][0]
                                                                   16sforward[0][0]
__________________________________________________________________________________________________
conv2d_transpose_2 (Conv2DTrans  (None, 64, 64, 7)    784         add_1[0][0]
__________________________________________________________________________________________________
8sforward (Conv2D)               (None, 64, 64, 7)    1799        block3_pool[0][0]
__________________________________________________________________________________________________
add_2 (Add)                      (None, 64, 64, 7)    0           conv2d_transpose_2[0][0]
                                                                   8sforward[0][0]
__________________________________________________________________________________________________
conv2d_transpose_3 (Conv2DTrans  (None, 512, 512, 7)  12544       add_2[0][0]
==================================================================================================
Total params: 134,308,725
Trainable params: 134,308,725
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.

epoch 1	epoch 10	epoch 20	ground truth
			
			
			

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.

model	32s	8s
class 0	0.74281	0.73002
class 1	0.87113	0.87970
class 2	0.19848	0.27704
class 3	0.79061	0.79446
class 4	0.70357	0.74200
class 5	0.65533	0.66272
mIoU	0.660320	0.680990

將前面還沒因為pooling壓小而解析度下降的資料接到後面來可以幫助segmentation在細節的地方切得更好，由上面的圖片中可看出8s比起32s的圖片找到更多零碎小面積的色塊。這是我覺得兩者表現差異的原因。

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

objective function:

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

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

$$x_i = \frac{1}{1+e^{-s_i}} \qquad s_i = \sum_{j=1} z_j w_{ji}$$

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

$$\frac{\partial G}{\partial x_i} = \frac{-t_i}{x_i} + \frac{1-t_i}{1-x_i} = \frac{x_i - t_i}{x_i(1-x_i)}$$

$$\frac{\partial x_i}{\partial s_i} = x_i(1-x_i) \qquad \frac{\partial s_i}{\partial w_{ji}} = z_j$$

$$\frac{\partial G}{\partial w_{ji}} = (x_i - t_i) z_j \Rightarrow \frac{dG}{dw} = -\sum_n (t^{(n)} - x^{(n)}) z^{(n)}$$