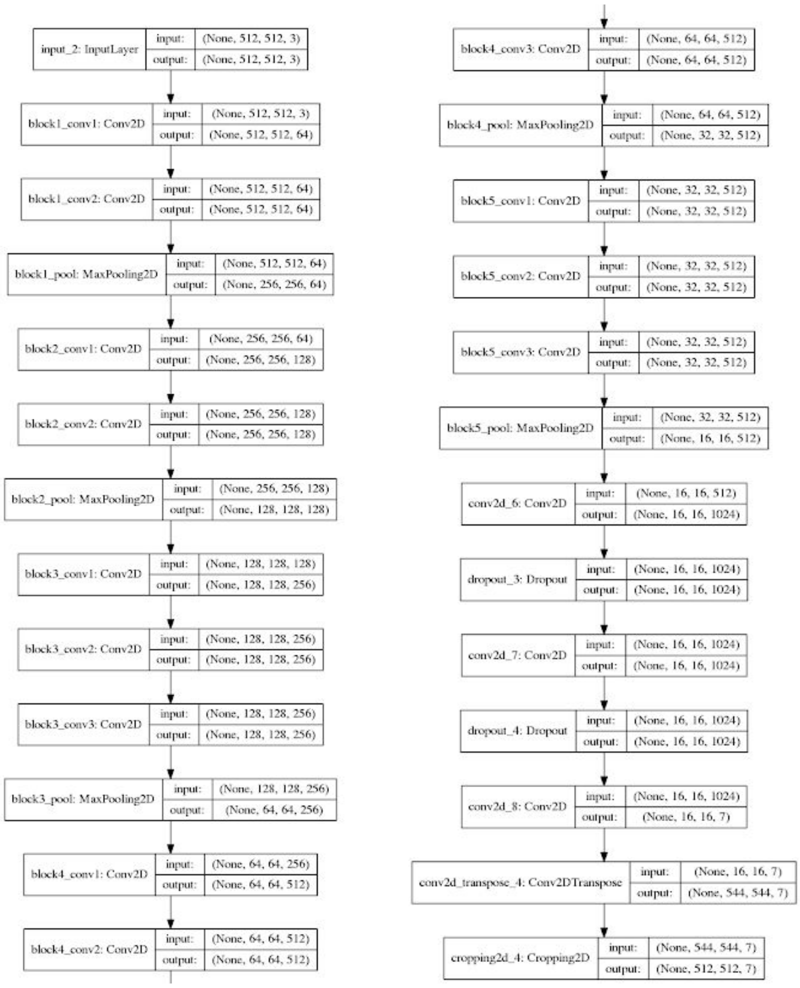
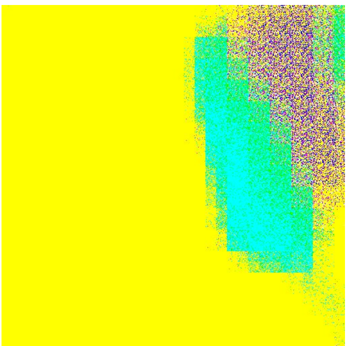
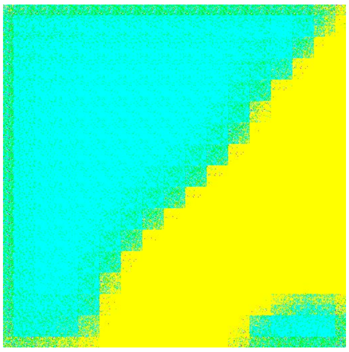
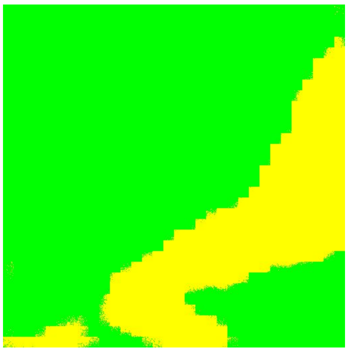
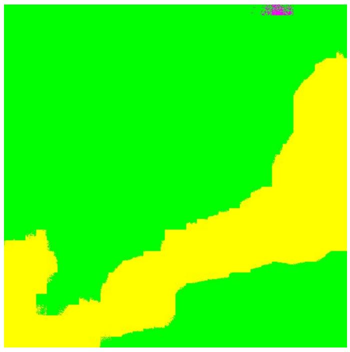
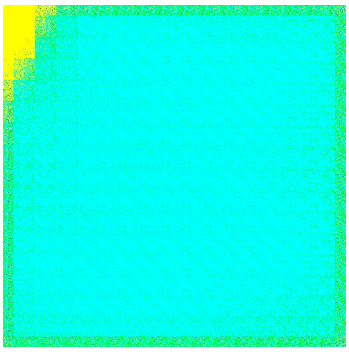
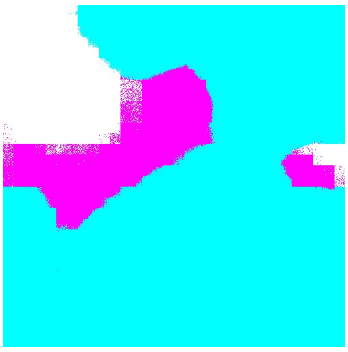
Name: 詹少宏 Dep.:電信碩一 Student ID:R06921017

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



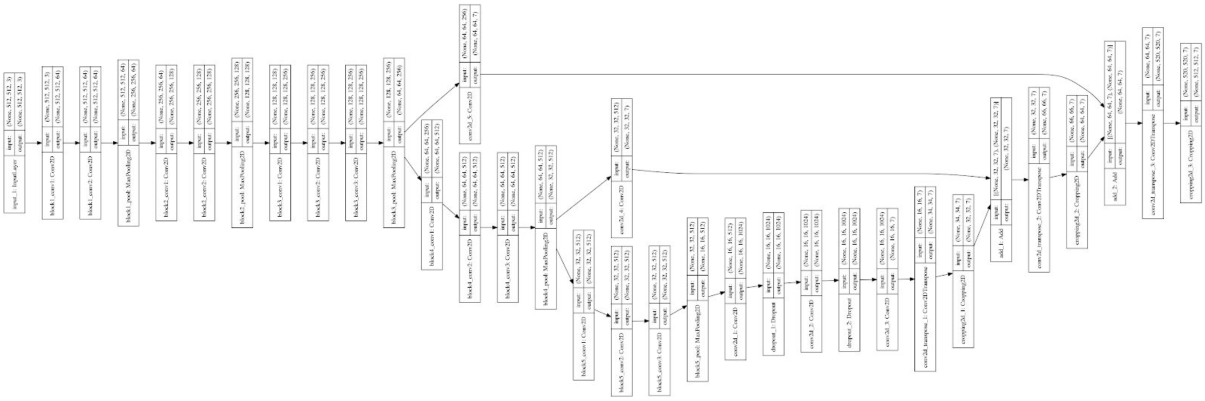
2. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

	epochs = 1	epochs = 10	epochs = 20
0008			

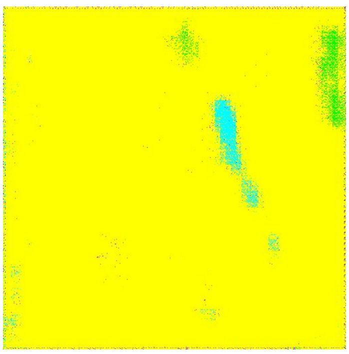
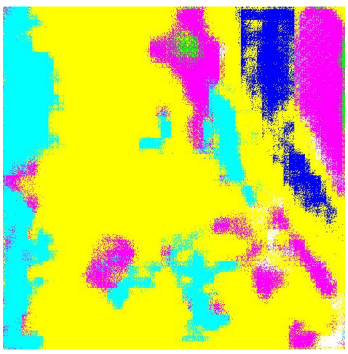
0097			
0107			

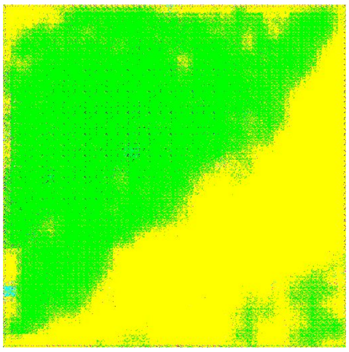
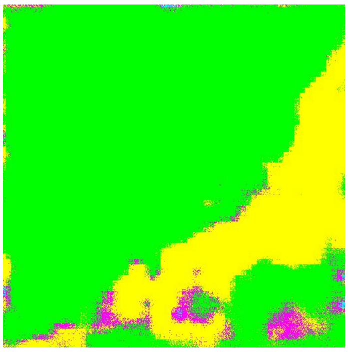
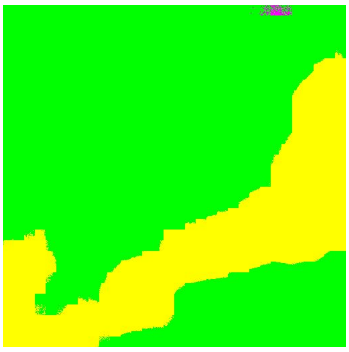
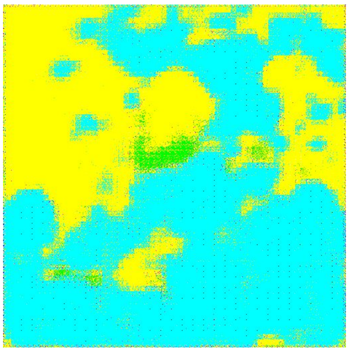
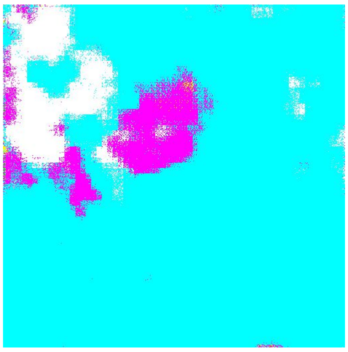
3. (15%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model.

Ans: 採用 vgg16-fcn8 的架構



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.

	epochs = 1	epochs = 10	epochs = 20
0008			

0097			
0107			

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.
   Ans:

	VGG16-FCN32	Improved Model (VGG16-FCN8)
Class 0	0.75959	0.75733
Class 1	0.88065	0.88696
Class 2	0.29925	0.35147
Class 3	0.80002	0.80257
Class 4	0.74475	0.73440
Class 5	0.68908	0.69435
mIoU	0.695557	0.704513

Discuss: FCN8 會提取不同層的feature，可以利用低維的特徵（例如線條辨識）或高維的特徵（如形狀辨識），進而有更多features 來做分類，效果類似pyramid。相較於FCN32 只以高維卷積產生的特徵為分類依據，FCN8 更可以結合低解析度與高解析度的特徵將像素分類，故能有更佳的 mean IoU。

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

Sol:

Set $g^{(n)} = t^{(n)} \cdot \log x^{(n)} + (1-t^{(n)}) \cdot \log(1-x^{(n)}) \Rightarrow \begin{cases} x^{(n)}(z^{(n)}; w) = \dfrac{1}{1+e^{-w^T z^{(n)}}} \\ G(w) = -\sum_n g^{(n)} \end{cases}$

$$\frac{d}{dw} G(w) = (-1) \cdot \sum_n \frac{d}{dw} g^{(n)} = (-1) \cdot \sum_n \left(\frac{d g^{(n)}}{d x^{(n)}}\right) \cdot \left(\frac{d x^{(n)}}{dw}\right)$$

①

$$\frac{d g^{(n)}}{d x^{(n)}} = t^{(n)} \cdot \frac{1}{x^{(n)}} + (1-t^{(n)}) \cdot \frac{1}{1-x^{(n)}} \cdot (-1) = \frac{t^{(n)}(1-x^{(n)}) + (1-t^{(n)}) \cdot x^{(n)} \cdot (-1)}{x^{(n)}(1-x^{(n)})}$$

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

② Let's set $u^{(n)}(z^{(n)}; w) = 1 + e^{-w^T z^{(n)}} \in$ Scalar. $\Rightarrow \begin{cases} e^{-w^T z^{(n)}} = u^{(n)} - 1 \\ x^{(n)} = \dfrac{1}{u^{(n)}} \end{cases}$

$$\therefore \frac{d x^{(n)}}{dw} = \frac{d x^{(n)}}{d u^{(n)}} \cdot \frac{d u^{(n)}}{dw} = \frac{-1}{(u^{(n)})^2} \cdot (-z^{(n)}) e^{-w^T z^{(n)}} = z^{(n)} \cdot \frac{1}{u^{(n)}} \cdot \frac{1}{u^{(n)}} (u^{(n)} - 1)$$

$$= z^{(n)} \cdot x^{(n)} \cdot (1 - x^{(n)})$$

$$\therefore \frac{d}{dw} G(w) = (-1) \sum_n \cdot \left[\frac{t^{(n)} - x^{(n)}}{\cancel{x^{(n)}}\,\cancel{(1-x^{(n)})}}\right] \cdot z^{(n)} \cdot \cancel{x^{(n)}} \cdot \cancel{(1-x^{(n)})} = -\sum_n (t^{(n)} - x^{(n)})\, z^{(n)}$$

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

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

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