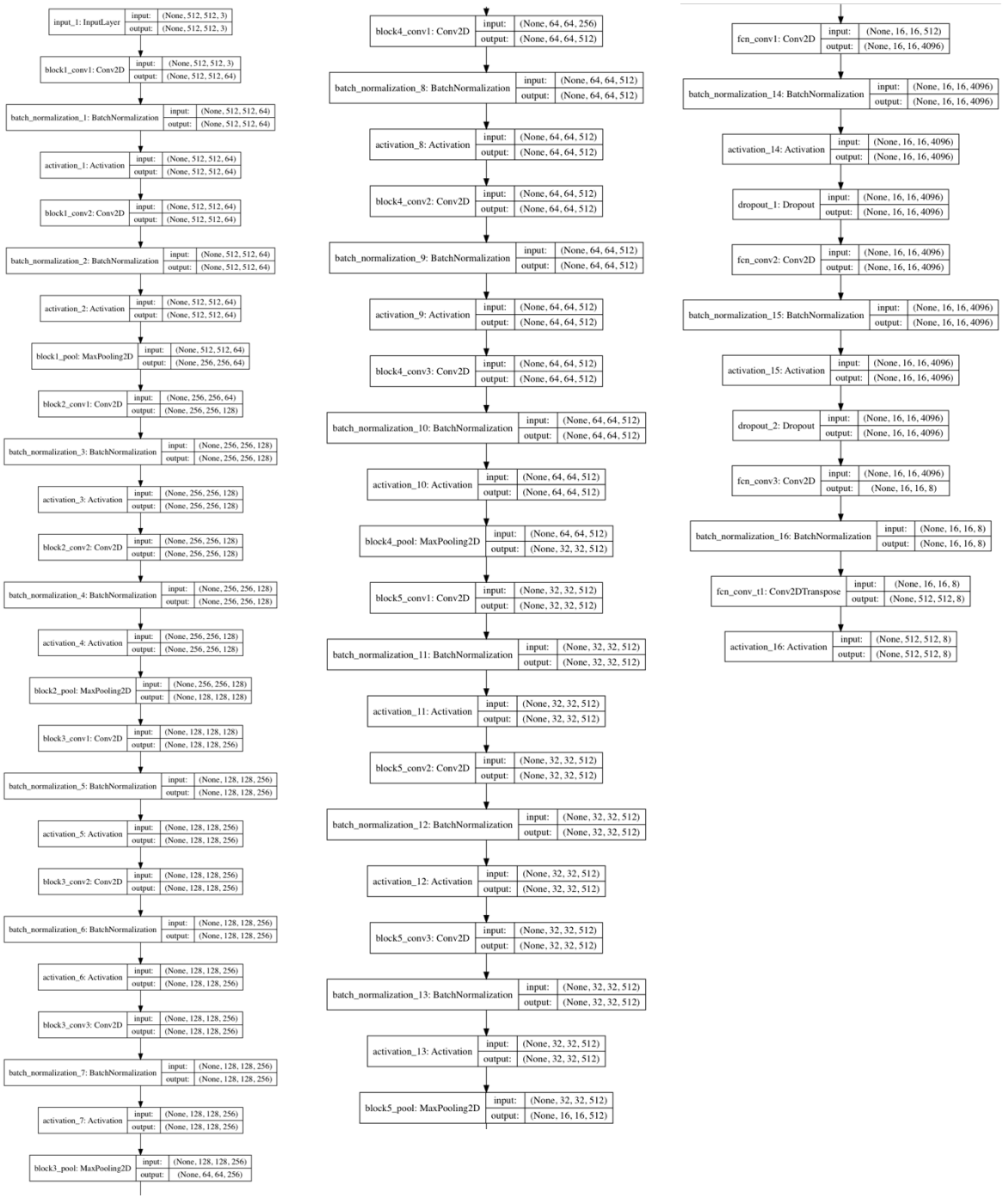
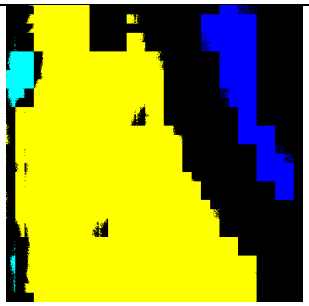
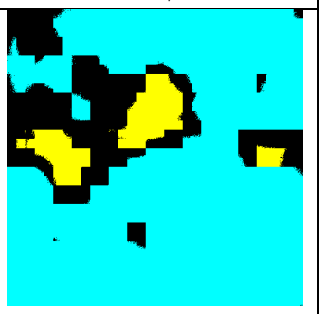
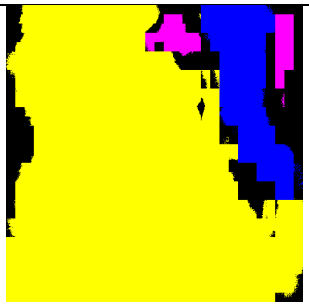
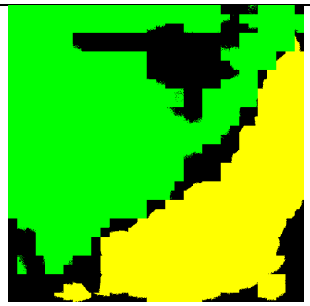
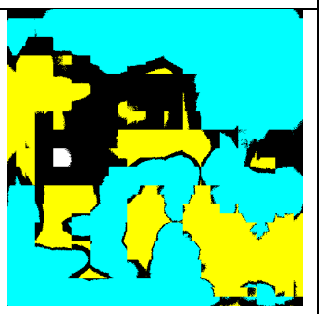
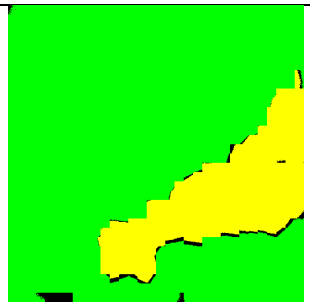
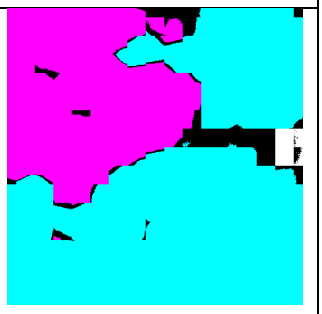
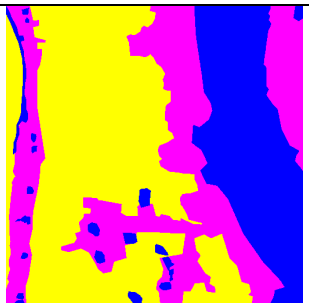
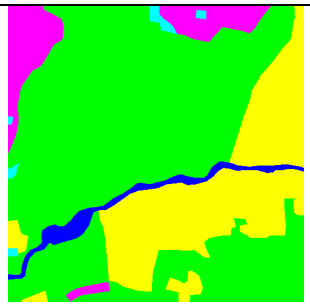
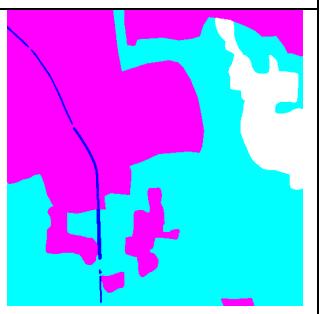
Please use this report template, and upload it in the PDF format. Reports in other forms/formats will result in ZERO point. Reports written in either Chinese or English is acceptable. The length of your report should NOT exceed 6 pages (**excluding bonus**).

Name: 黃宇平 **Dep.:**電信碩一 **Student ID:**R06942065

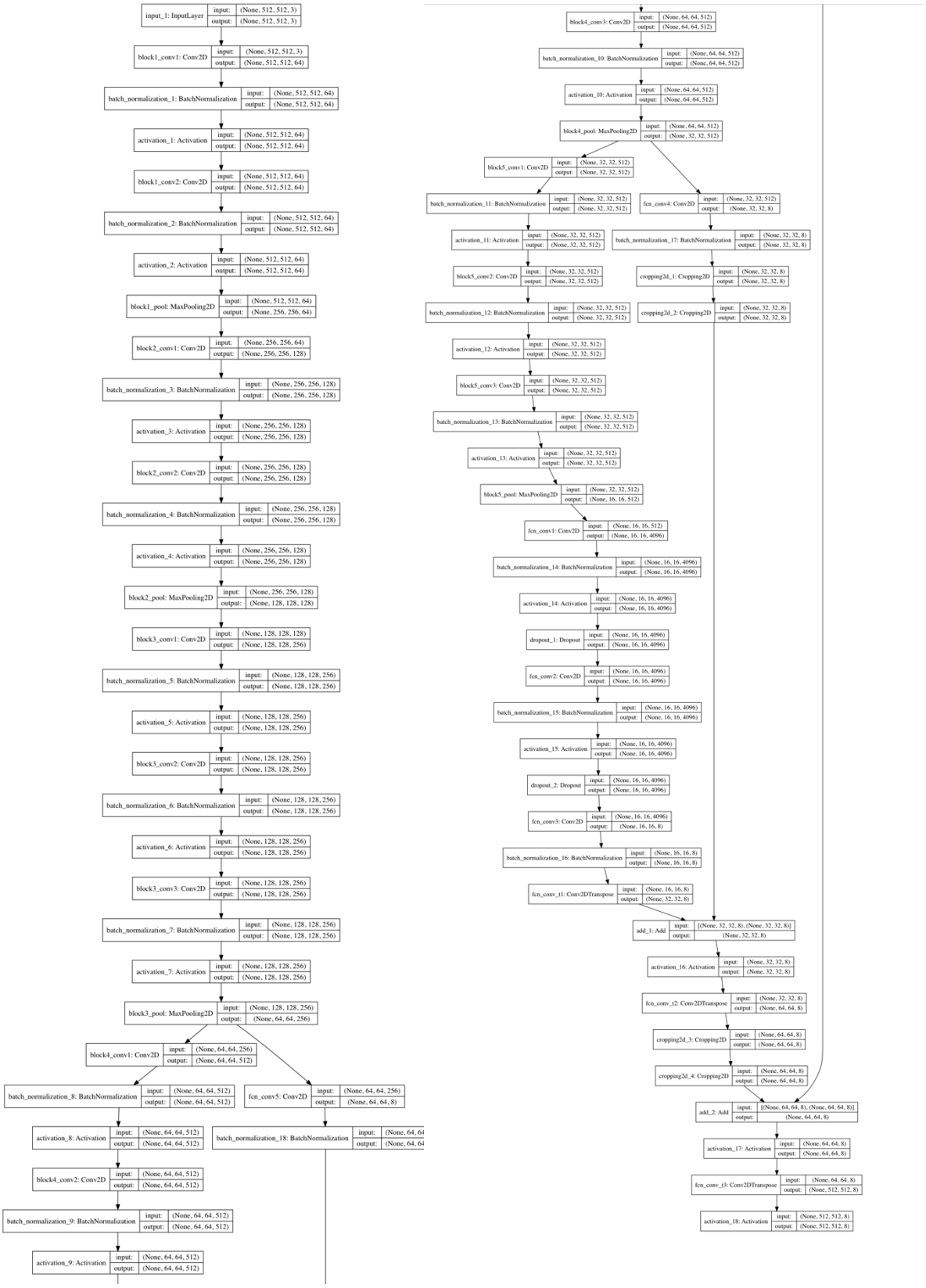
1. (5%) Print the network architecture of your VGG16-FCN32s model.



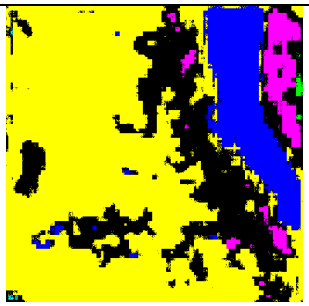
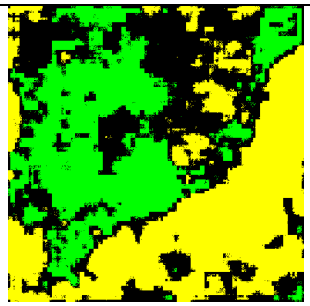
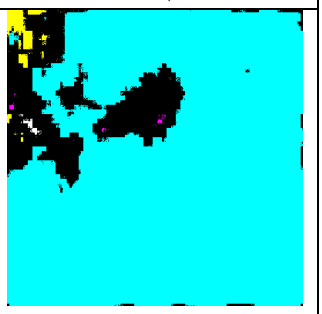
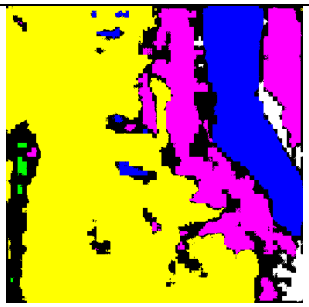
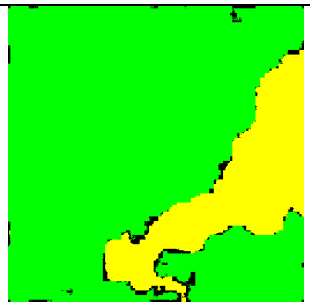
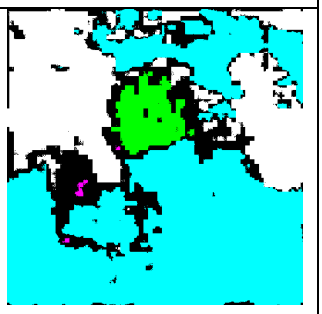
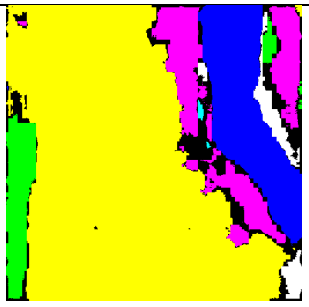
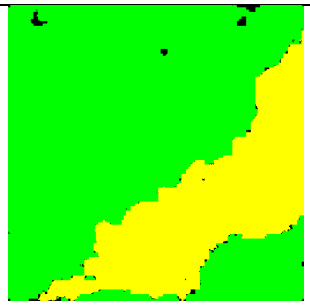
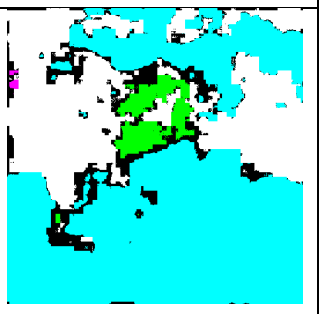
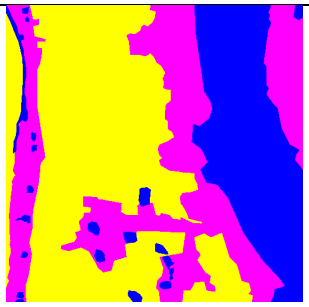
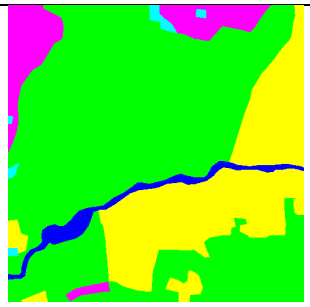
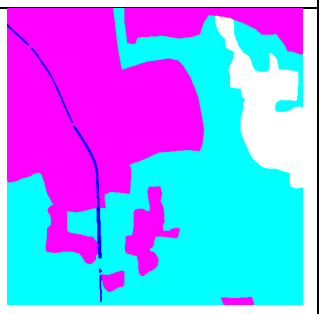
2. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

mask_id	validation/0008	validation/0097	validation/0107
10th epoch			
30th epoch			
50th epoch			
ground truth			

3. (15%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model.
我實作的是VGG16-FCN8的model，架構如下：



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.

mask_id	validation/0008	validation/0097	validation/0107
10th epoch			
30th epoch			
50th epoch			
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.

	Baseline(VGG16-FCN32)	Best(VGG16-FCN8)
#epochs	50	50
mIoU	0.599	0.650

FCN8 model和 FCN32的差別只在於多了兩個skip connections，這說明了skip connections能讓 Fully Convolution Network 更順利地把不同scale的資訊一起看待。除了FCN-8，我還實作了FCN16(也就是把FCN-8的block3接往後面的skip connection移除)，並比較其Performance：

	VGG16-FCN32	VGG16-FCN8	VGG16-FCN16
#epochs	50	50	50
mIoU	0.599	0.650	0.622

由上表，我發現在這個task上有skip connections的network結果都比沒有的好。

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