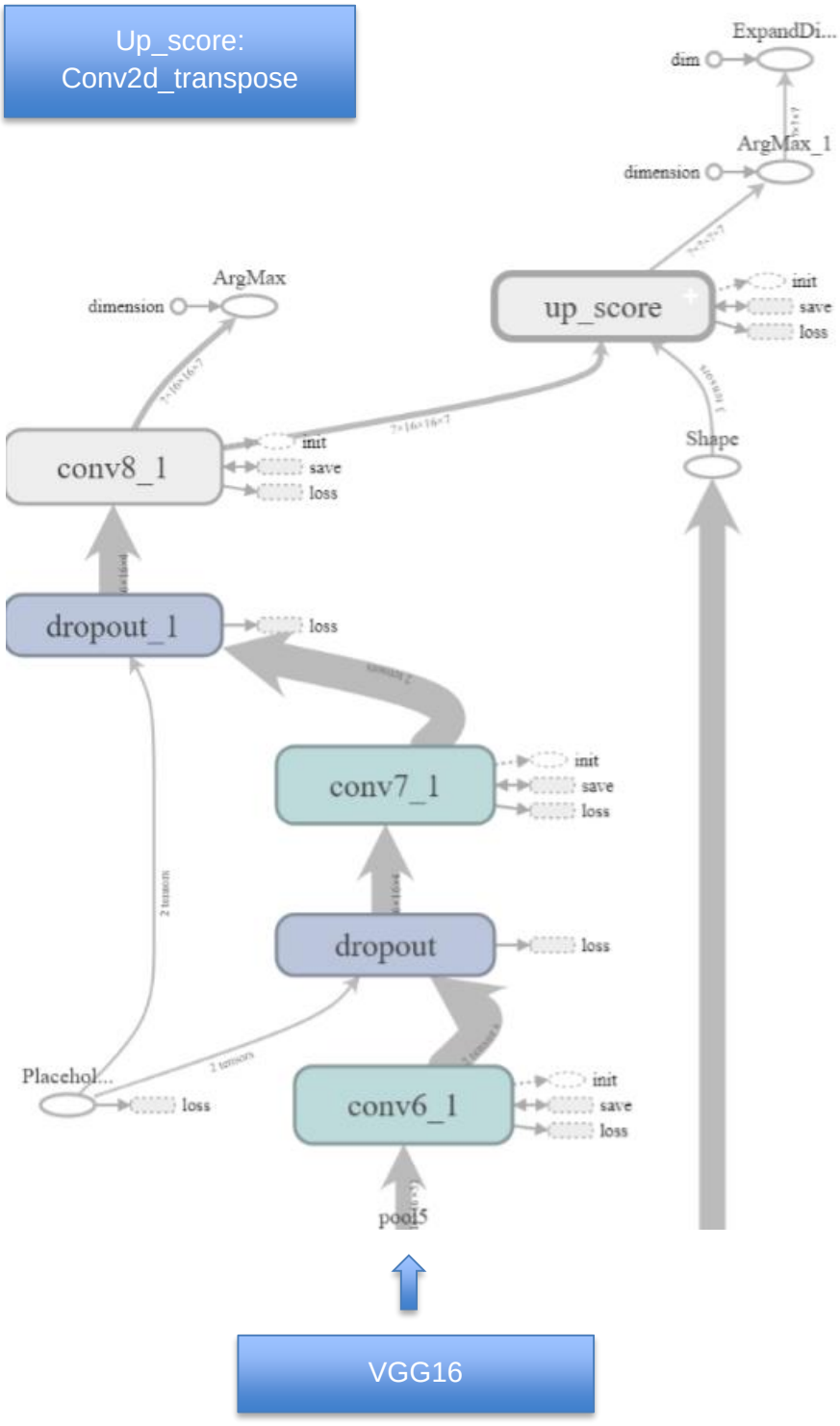
Please use this report template, and upload it in the PDF format. Reports in other forms/formats will result in ZERO point. Reports written in either Chinese or English is acceptable. The length of your report should NOT exceed 6 pages (**excluding bonus**).

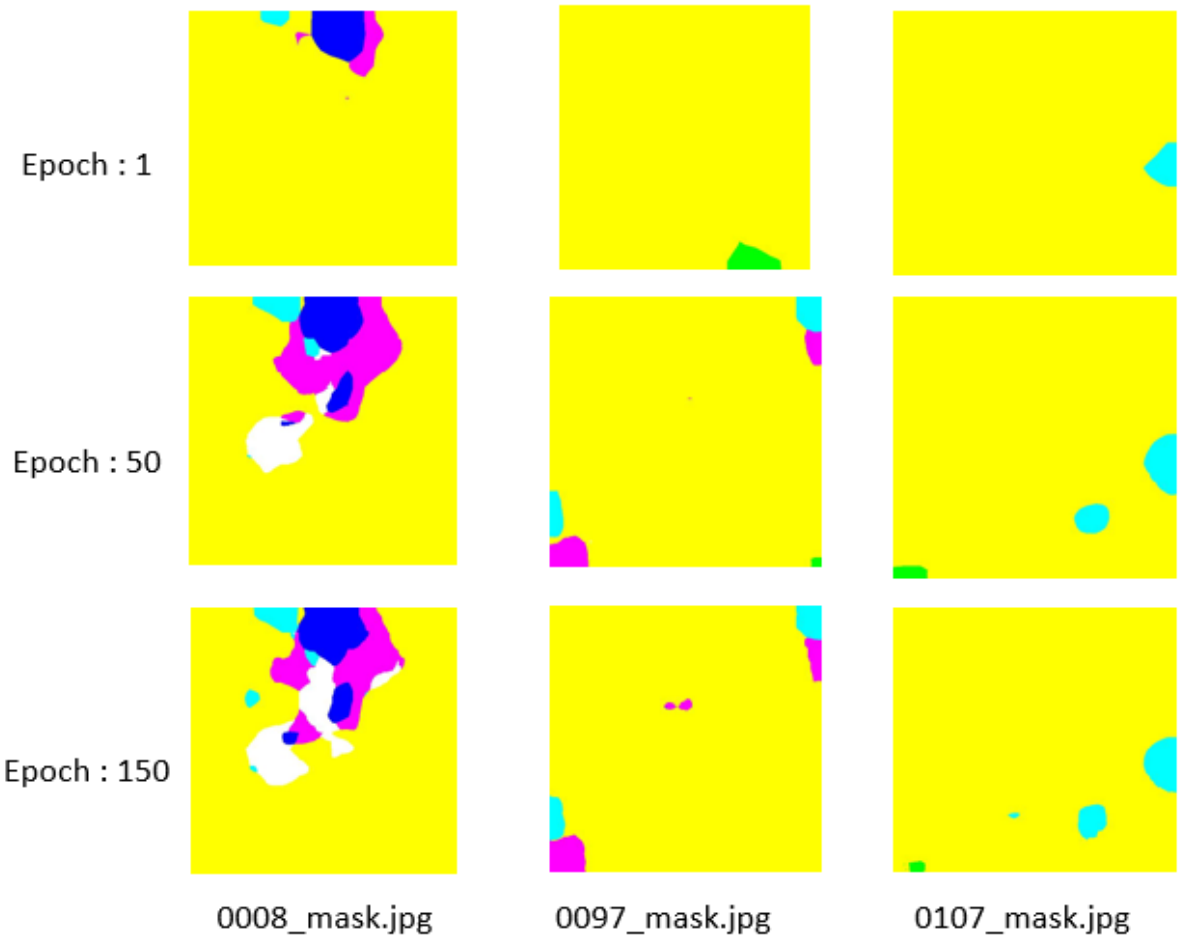
Name: 曾柏偉 Dep.:電信碩一 Student ID:R06942098

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



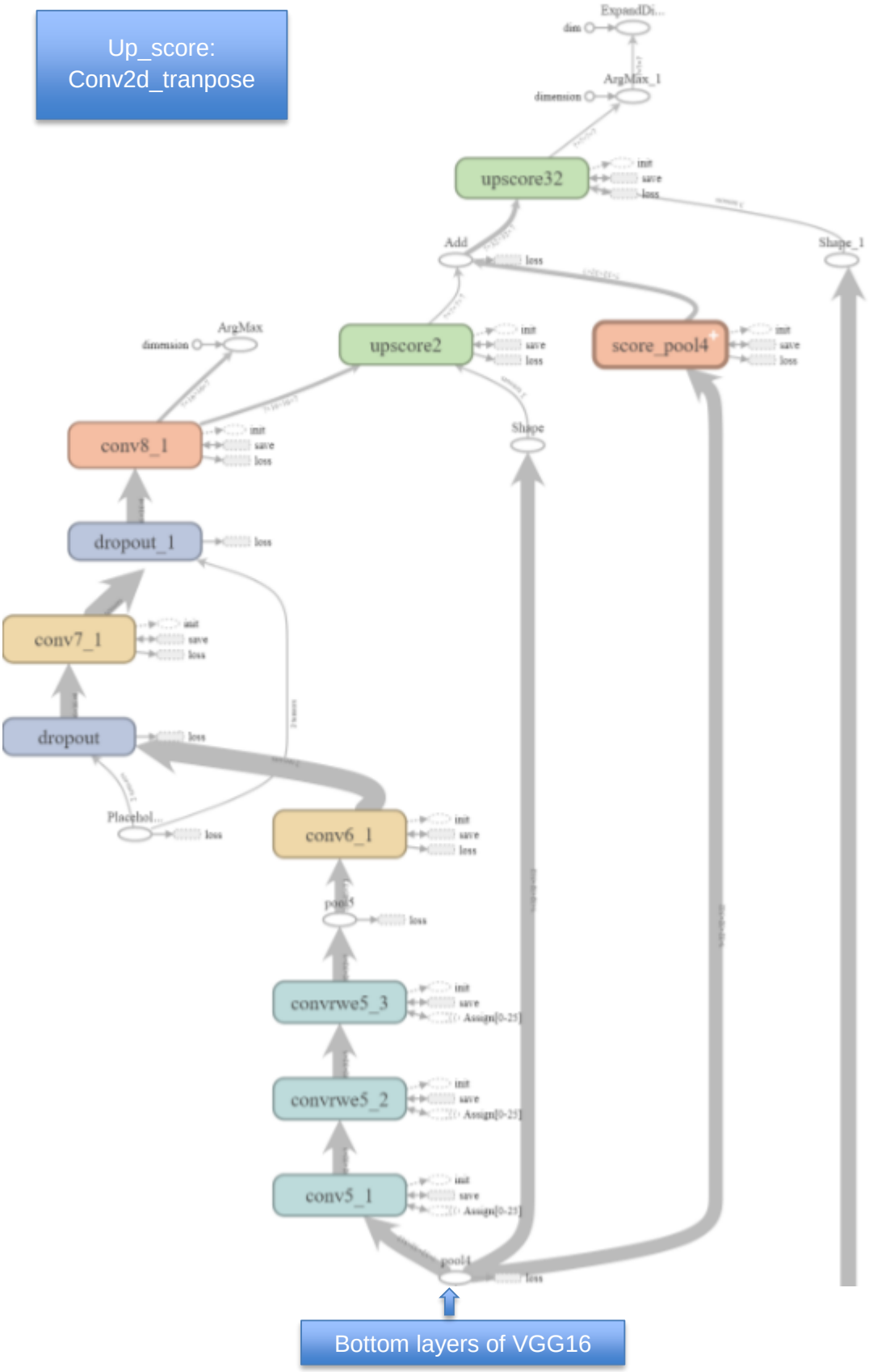
2. (10%) Show the predicted segmentation mask of validation/0008_sat.jpg, validation/0097_sat.jpg, validation/0107_sat.jpg during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

   Ans : I have training for 150 epochs , so I choose epochs of 0 ,50 ,150 and plot the mask image to answer this question.



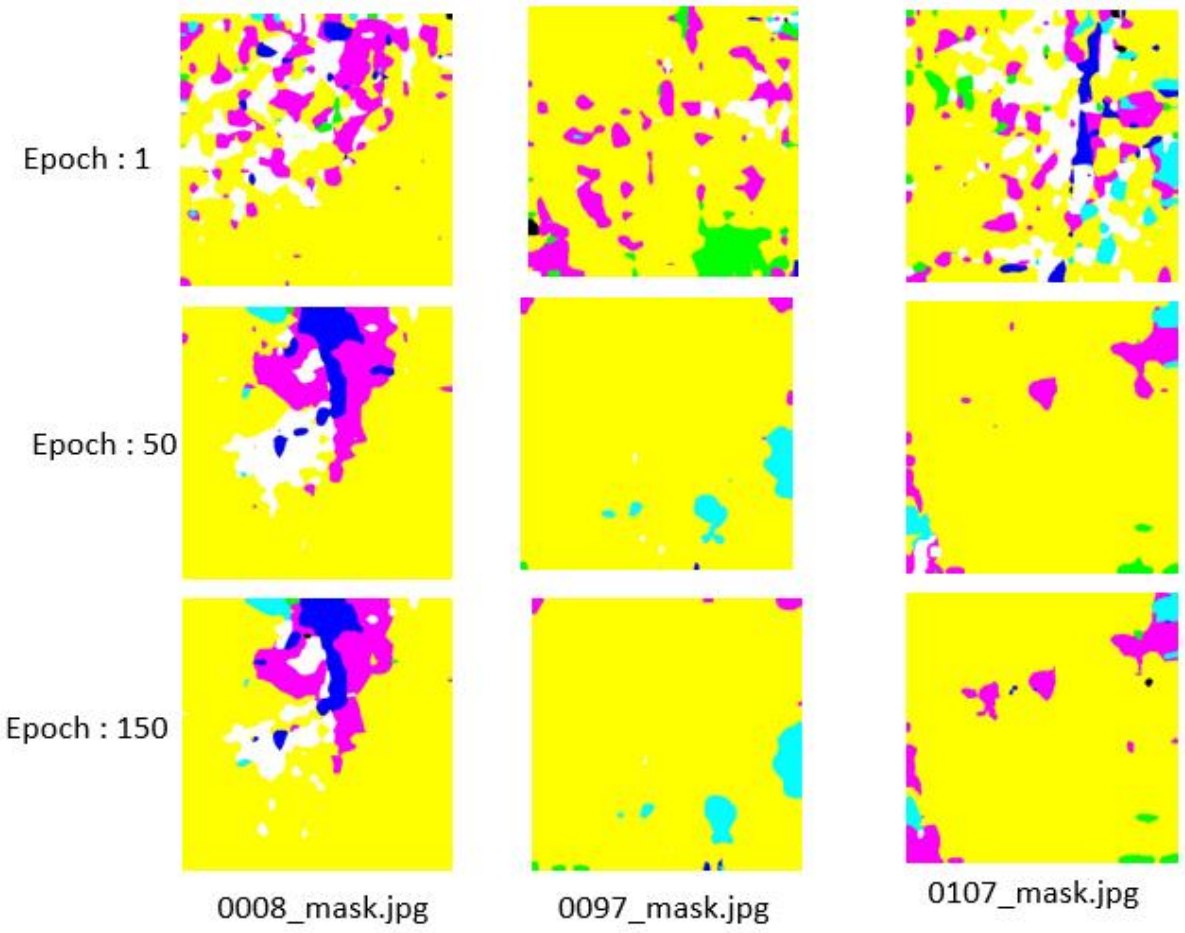

3. (15%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model.

   Ans : I try both VGG16-FCN16s and VGG16-FCN8s. VGG-FCN16s can reach the baseline but the performance of VGG16-FCN16s is better than 32S.




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.

Ans : I have training for 150 epochs , so I choose epochs of 0 ,50 ,150 and plot the mask image to answer this question.




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: VGG-F32c : 0.645742 , VGG-F16c : 0.649150 , VGG-F8c:0.63816. All of them are trained with 150 epochs but different model structure. Obviously , F16s is better than F32s , the reason of result may be F16s consider the features of top layer of vgg16 , and add them to the following convolution layers .It consider the roughly embedding of image into training , so it is better than the simpler model VGG-F32s. On the other hands, F8c consider more features of VGG16 bottom layer, its performance is the worst in these three models.

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

For a logistic regression problem :

Output probability : $\sigma(w^T x)$ , $x$ : $training\ data$ , $w$ : $weight\ matrix$

$$\frac{d}{dw_i} \sigma(w^T x) = \sigma(w^T x)(1 - \sigma(w^T x)) e^{W^T x} w_i$$

$$\frac{d}{dw_i} G(w) = -\sum_n t^n \frac{1}{\sigma(w^T x^n)} \sigma(w^T x^n)\left(1 - \sigma(w^T x^n)\right) e^{w^T x^n} w_i + (1 - t^n) \frac{-1}{1 - \sigma(w^T x^n)} \sigma(w^T x^n)\left(1 - \sigma(w^T x^n)\right) e^{w^T x^n} w_i$$

and by simple operation , get :

$$-\sum_n (t^n - \sigma(w^T x^n) e^{w^T x^n} w_i$$