HW3

r06921062 蘇楷鈞

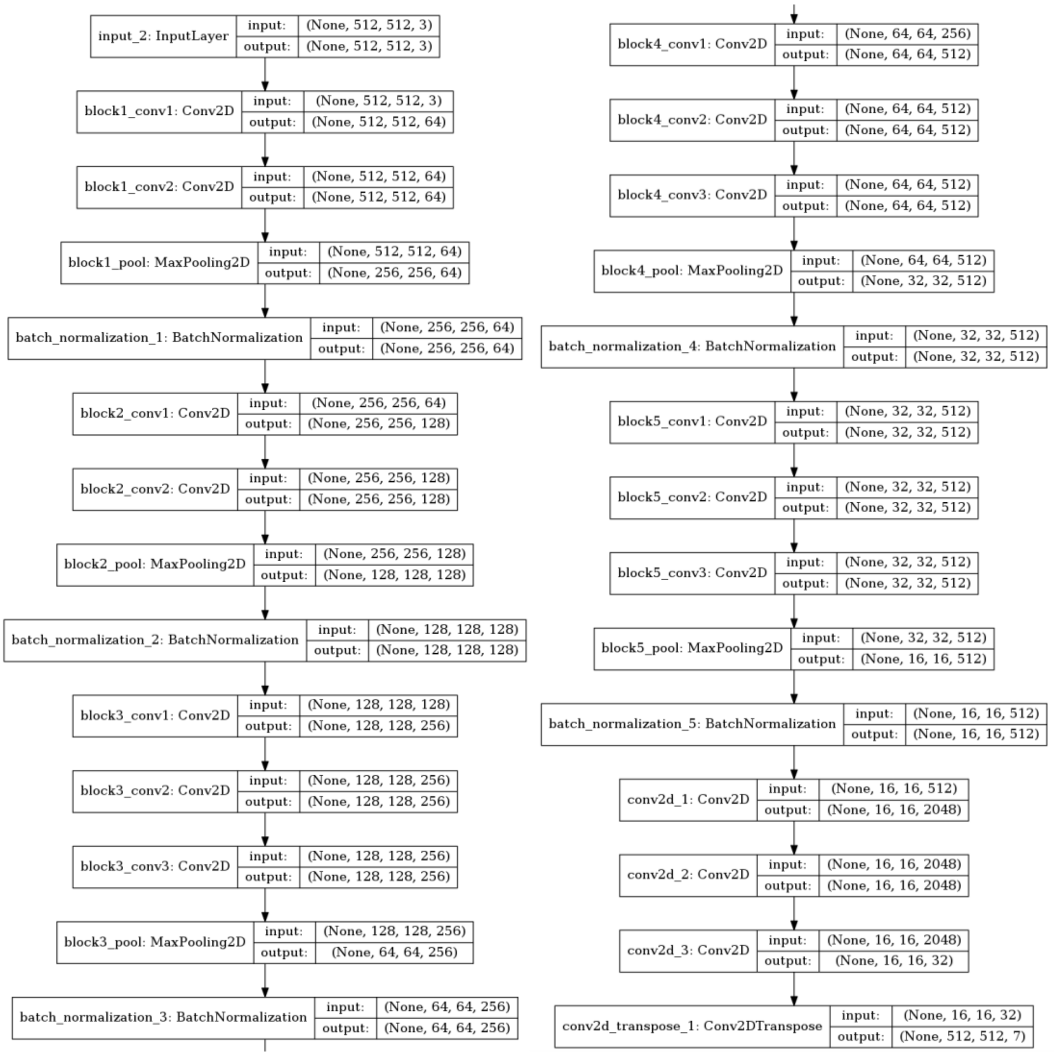
1. (5%) Print the network architecture of your VGG16-FCN32s model.
   - **Training strategy**

     Optimizer：Adam(lr=0.0001 , decay rate=0.9)

     Epoch：20 (Use early stopping and save the best weights by monitor on valid loss)

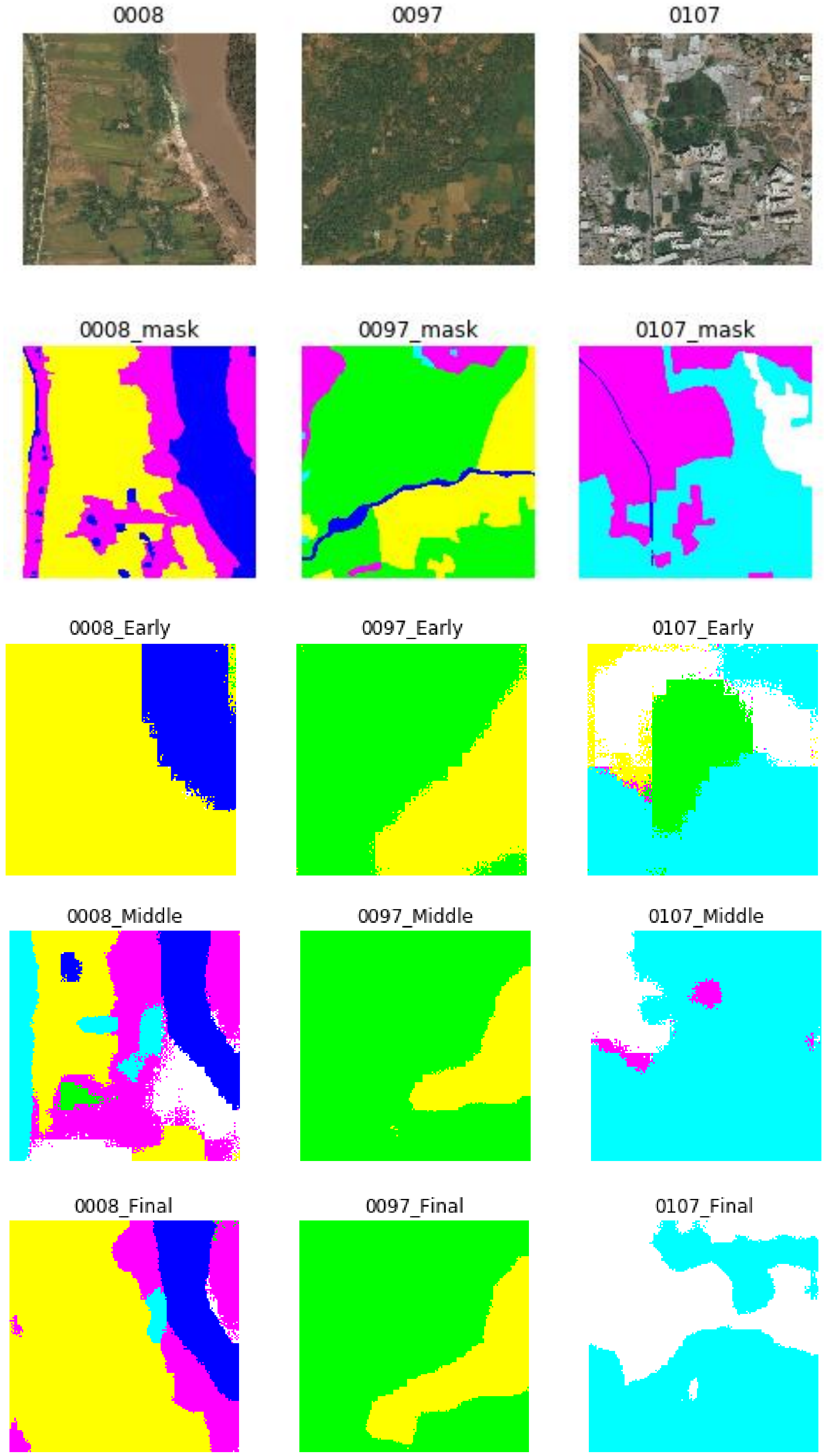
     Batch size：8

     Loss function：Categorical crossentropy
   - **Model Architecture**[1]



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)

   Early：2 epoch , Middle：7 epoch , Final：12 epoch.

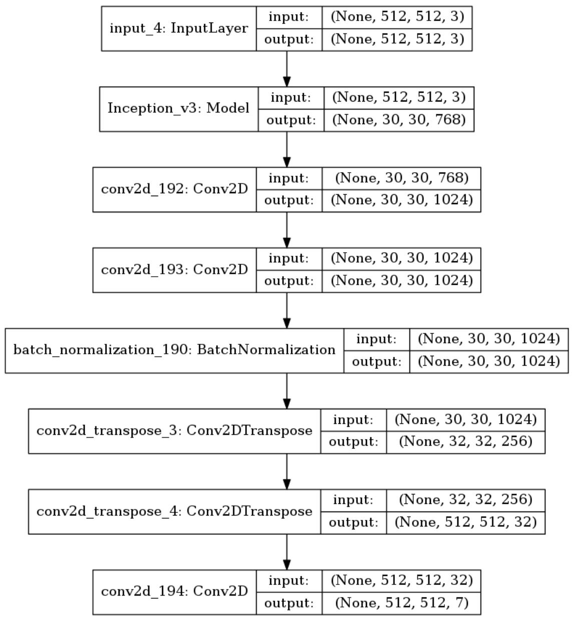

0008
0097
0107
0008_mask
0097_mask
0107_mask
0008_Early
0097_Early
0107_Early
0008_Middle
0097_Middle
0107_Middle
0008_Final
0097_Final
0107_Final

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

- **Training strategy**

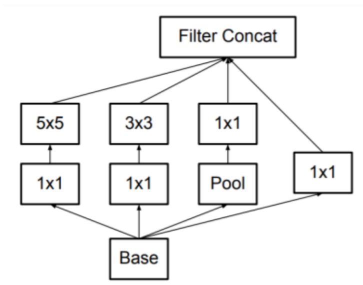
Optimizer：Adam(lr=0.0001 , decay rate=0.9)
Epoch：20(Use early stopping and save the best weights by monitor on valid loss)
Batch size：16
Loss function：Categorical crossentropy

- **Model Architecture**

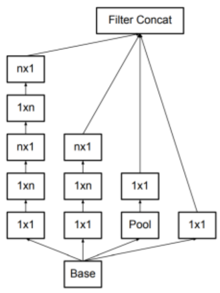


- **Inception_v3 model**[2]

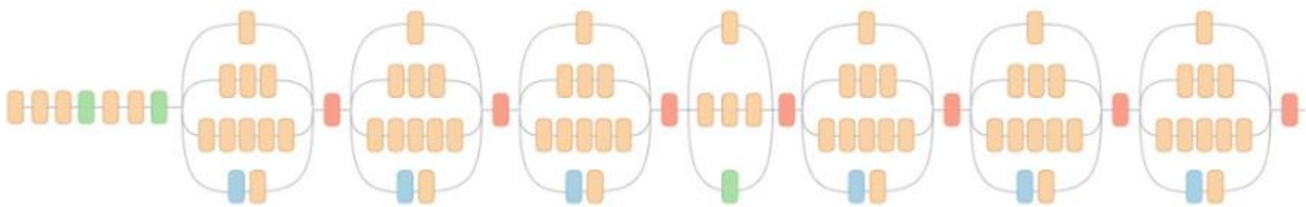
抽出 mixed6 的 feature 接到 FCN16s(圖檔為 inception_v3.png)，模型主要由以下兩種各組成基本的 block(圖(一)、圖(二))在串接而成(圖(三))。[2]



圖(1)



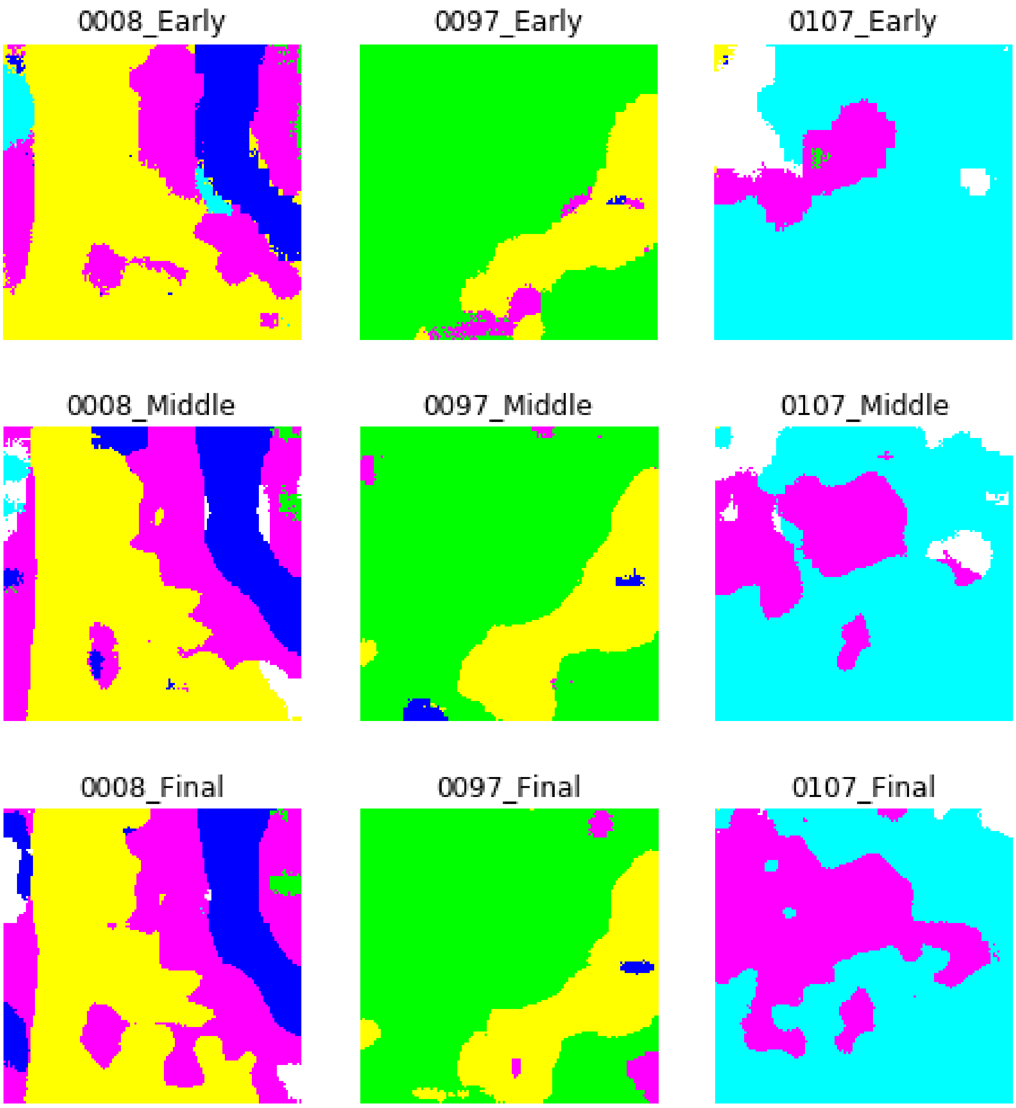
圖(二)



圖(三)

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.

Early：2 epoch , Middle：7 epoch , Final：12 epoch.

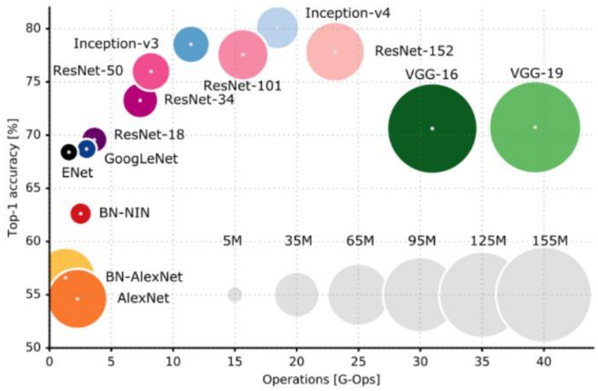


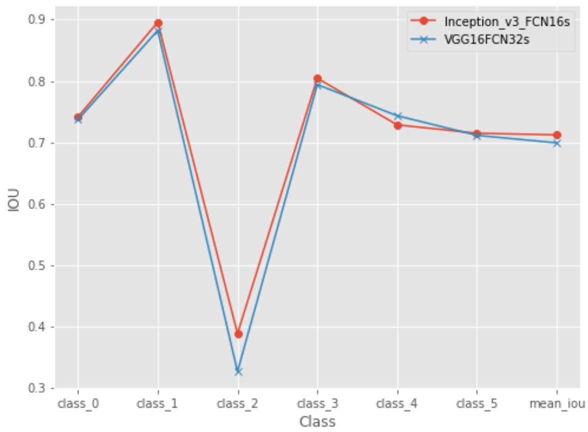
5. (15%) Report mIoU score of both models on the validation set. Discuss the reason why the improved model performs better than the baseline one. You may conduct some experiments and show some evidences to support your reasoning.

- **Comment**

會選擇以 Inception v3 做為比較是因為再 Imagenet 上，Inception v3 參數量較少、執行運算量、準確率來說都較 VGG16 高，所以猜測抓出來的特徵應該較能表示原圖資訊。

從以上觀察兩個模型預測出的 mask 來說，雖然 Inception v3 參數量少了非常多，但較能預測出更多細節，其可能是因為 Inception 的架構下有諸多不同形狀的 CNN（像是 3x1,1x3,3x3 等 kernel size)，因此可特別抓出不同方向間的關係，如果從這角度看 VGG16，要抓出這些特徵，勢必需要更大的維度去尋找（假使方正的 kernel size 要抓出橫向的，就可能變成橫向的 kernel 有值，其餘等於 0)；以 mean IOU 來看，還是較 VGG16FCN32s 高一個百分點，從各類別來看(右圖)，在 class 2 下，Inception v3 有 0.38848，VGG16 只有 0.32664，贏了 6.3 個百分點，而 class 4 則輸了 1.5 個百分點，總體來說較 VGG16 更能有效預測各類別。





	VGG16FCN32s	Inception v3 FCN16s
Variable counts	71,282,279	17,714,471
Mean IOU	0.699188	0.712273

6. Bonus(下頁)

$$G(w) = -\Sigma_n [t^{(n)} \log x(z^{(n)}; w) + (1-t^{(n)}) \log(1 - x(z^{(n)}; w))]$$

$$x(z^{(n)}; w) = \frac{1}{1+e^{-y^{(n)}}}, \quad y^{(n)} = w^T z^{(n)} + b$$

$$\frac{dG(w)}{dw} = \frac{dG(w)}{dx(y)} \times \frac{dx(y)}{dw}$$

$$= \frac{dG(w)}{dx(y)} \times \frac{dx(y)}{dy} \times \frac{dy}{dw} \quad \text{代入 note ① ②}$$

$$= -\Sigma_n [\frac{t^{(n)}}{x(z^{(n)};w)} + (1-t^{(n)}) \frac{-1}{1 - x(z^{(n)}; w)}] \times x(z^{(n)}; w)(1 - x(z^{(n)}; w)) z^{(n)}$$

$$= -\sum_n \frac{t^{(n)} - x(z^{(n)};w)}{x(z^{(n)};w)\,(1 - x(z^{(n)};w))} \times x(z^{(n)};w)\,(1 - x(z^{(n)};w))\, z^{(n)}$$

$$= -\sum_n \left(t^{(n)} - x(z^{(n)};w)\right) z^{(n)}$$

note: ① $\frac{d \log t}{dt} = \frac{1}{t}$

② $$\frac{d\,x(y)}{dy} = \frac{d\,x(y)}{de^{-y}} \times \frac{de^{-y}}{dy}$$

$$= \frac{-1}{(1+e^{-y})^2} \times (-e^{-y})$$

$$= \frac{1}{1+e^{-y}} \times \frac{e^{-y}}{(1+e^{-y})}$$

$$= x(y) \times (1 - x(y))$$

Reference

[1] keras-FCN. https://github.com/aurora95/Keras-FCN，另外有與林家慶學長討論 FCN32s 架構。

[2] Rethinking the Inception Architecture for Computer Vision. Christian Szegedy Google Inc. , Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna University College London, Dec,11,2015.

[3] 圖(三) http://ju.outofmemory.cn/entry/315667

[4] An Analysis of Deep Neural Network Models forRethinking the Inception Architecture for Computer Vision Practical Applications.　Alfredo Canziani, Adam Paszke, Eugenio Culurciello, 14 Apr 2017 .