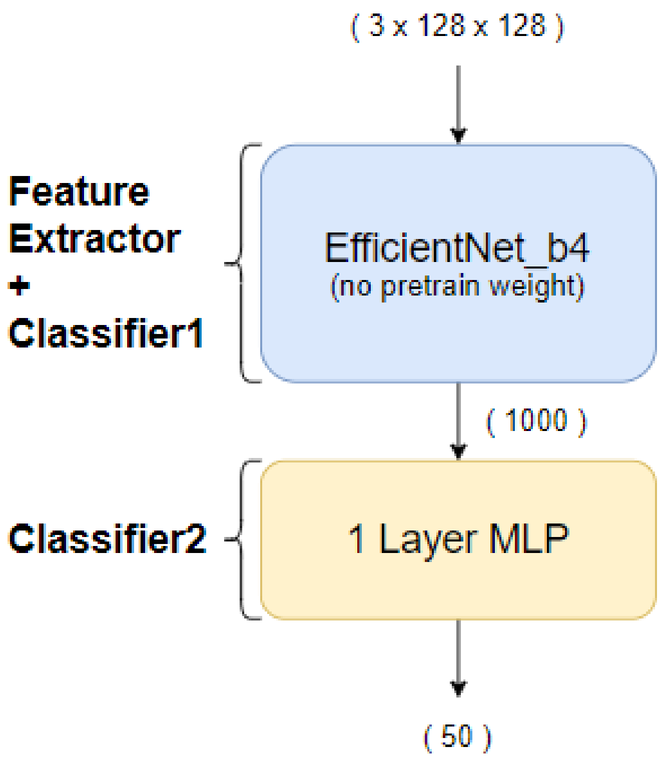
DLCV hw1 電機碩一 R11921100 溫威領

Problem 1: Image Classification

1. Draw the network architecture of method A or B.

Model A



2. Report accuracy of your models (both A, B) on the validation set.

Model A validation accuracy: 71.1%

Model B validation accuracy: 86.84%

3. Report your implementation details of model A.

Optimizer: Adam / learning rate: 1e-4 / weight decay: 1e-4 / with learning rate schedule

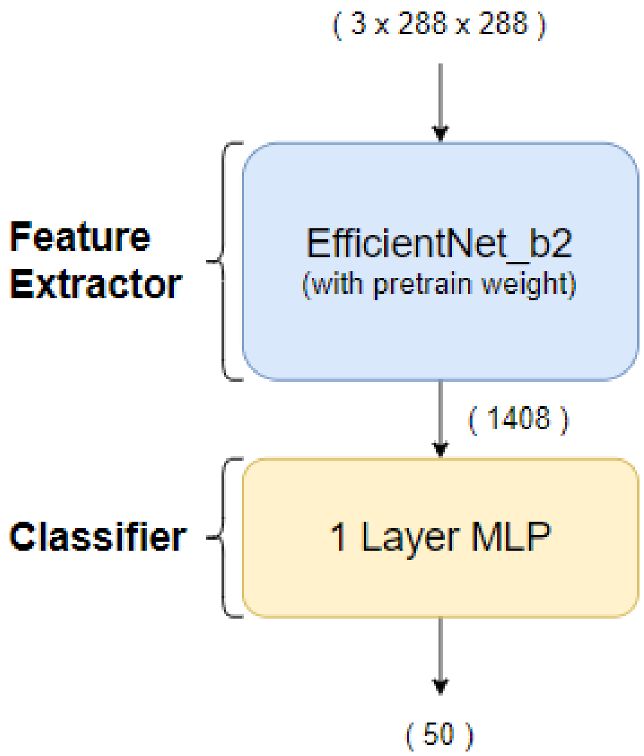
Loss function: cross entropy loss

Input image size: 128x128 / batch size: 64 / with normalized and data augmentation

Data augmentation: random horizon flip, random crop, and auto augment, etc.

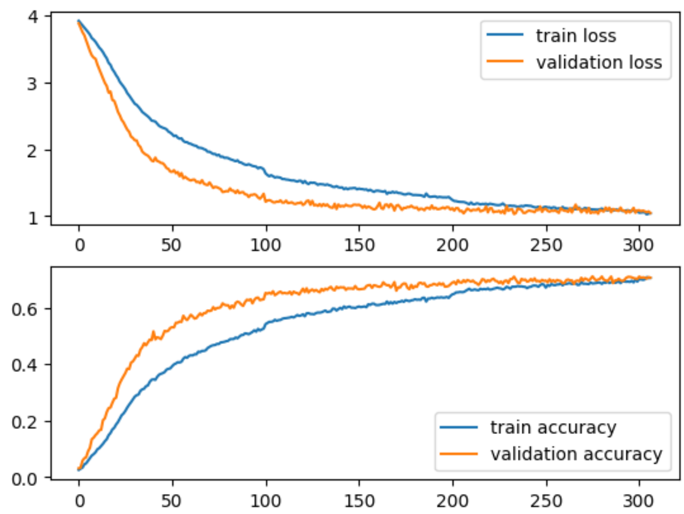
4. Report your alternative model or method in B, and describe its difference from model A.

Model B

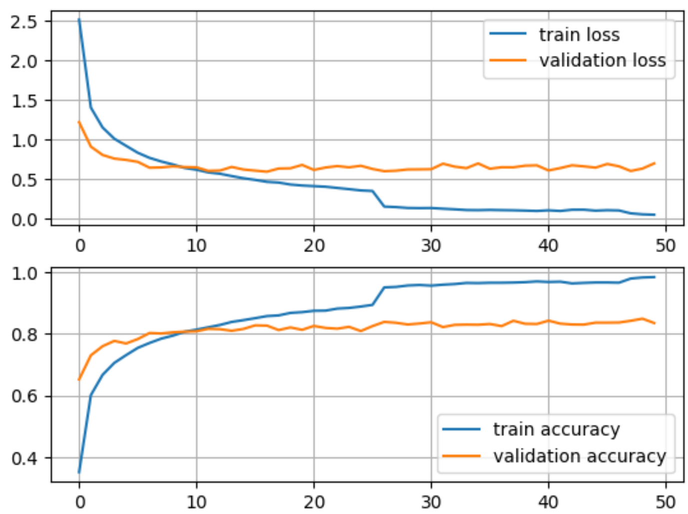


相較於 model A 的 non-pretrain EfficientNet_b4，model B 採用 pretrain 的 EfficientNet_b2，並把輸入圖像大小改為 288x288，會選擇使用 EfficientNet_b2 的原因在於使用 b3、b4 等 overfitting 較明顯，所以降低模型複雜度。

Model A training record

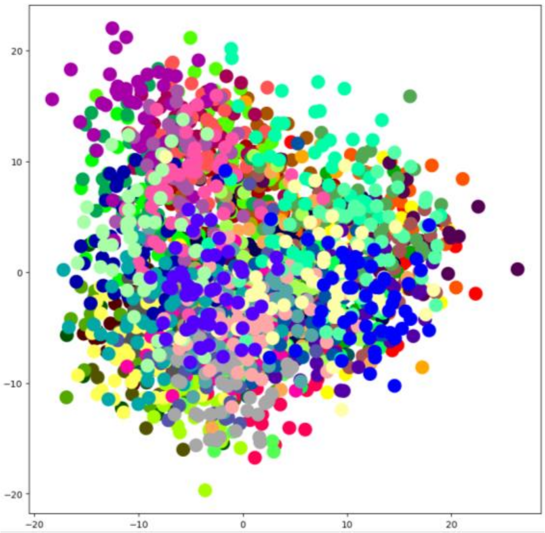


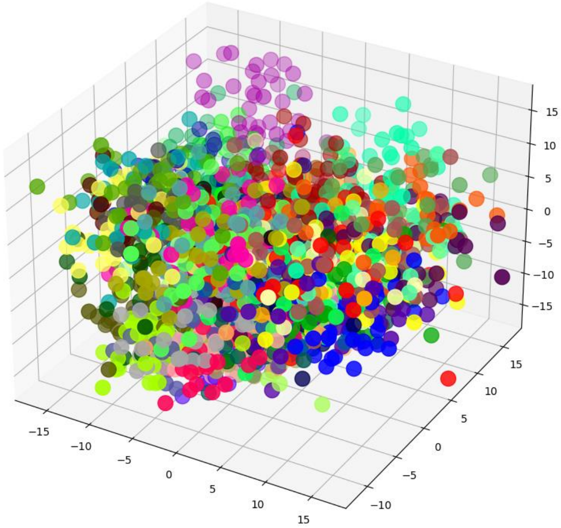
Model B training record



上圖左右分別為 model A 和 model B 訓練過程中所記錄的 training loss、validation loss、training accuracy、validation accuracy。兩者相同處可以發現訓練前期 validation accuracy 皆高於 training accuracy，是因為訓練時在 training data set 上使用 data augmentation 所以較難分類，訓練後期則追上 validation accuracy。兩者相異處可觀察到 model A (non-pretrain)在 epoch 1 的 training 和 validation accuracy 起點相同，而 model B 的 validation accuracy 則高過 training，其中 model A 的訓練過程也較緩慢、長久，在 epoch 10 時，model A 的 training 和 validation accuracy 分別達到約 10%和 20%，此時 model B 都已達到 80%，上述造成兩者相異處的原因皆是因為 model B 使用 pretrain weight 作 fine-tuning，但 model A 無使用。

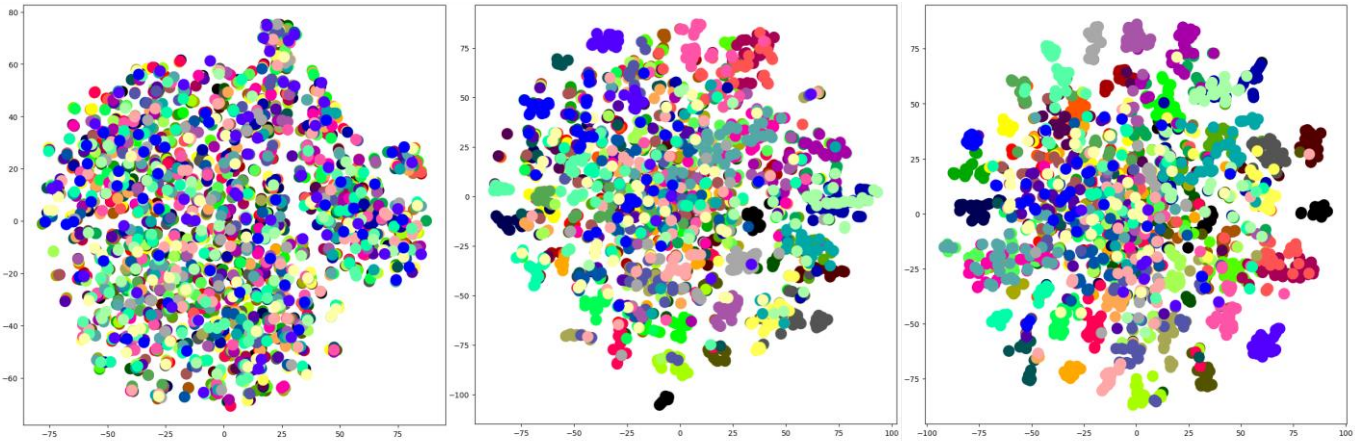
5. Visualize the learned visual representations of model A on the validation set by implementing PCA (Principal Component Analysis) on the output of the second last layer. Briefly explain your result of the PCA visualization.





原先 1000 維特徵經由 PCA 降維，而相同類別的圖像被標示為相同顏色並作圖如上。從上圖可觀察到相同顏色(即相同類別)的點大部分都聚集在同一區塊，使得各圖像特徵可以更容易達到分類。

6. Visualize the learned visual representation of model A, again on the output of the second last layer, but using t-SNE (t-distributed Stochastic Neighbor Embedding) instead. Depict your visualization from three different epochs including the first one and the last one. Briefly explain the above results.

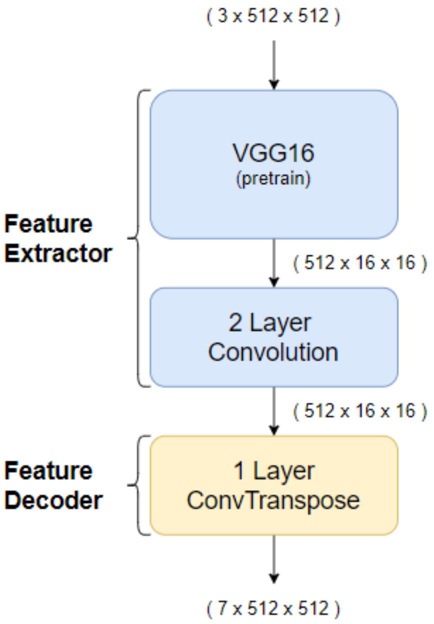


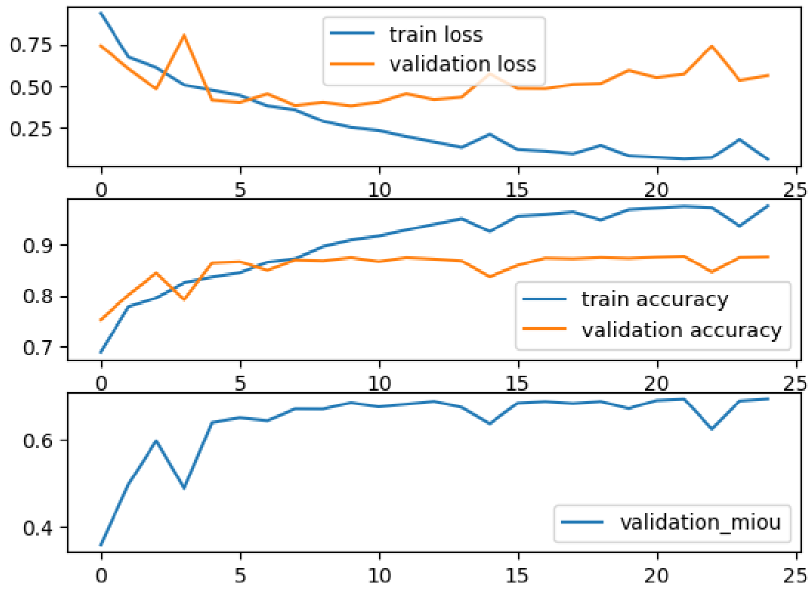
validation accuracy: @epoch1 = 1.6%, @epoch50 = 55.7%, @epoch300 = 71.1%

經由 t-SNE 降維並作圖如上，左圖因 epoch 1 未完成訓練，所以會看到不同顏色點散亂無區分。訓練 50 epochs 後有 56%的正確率，作圖後如中間的圖，可觀察到有部分相同顏色點開始聚集在一塊，並有粗略分類能力。繼續訓練到 300 epochs 後達 71%正確率，從其右圖中可明顯看到外圈相同顏色點都緊密聚集，而內部較亂但也有些微顏色區分，推測是較難分類的類別，因此正確率無法近 100%。

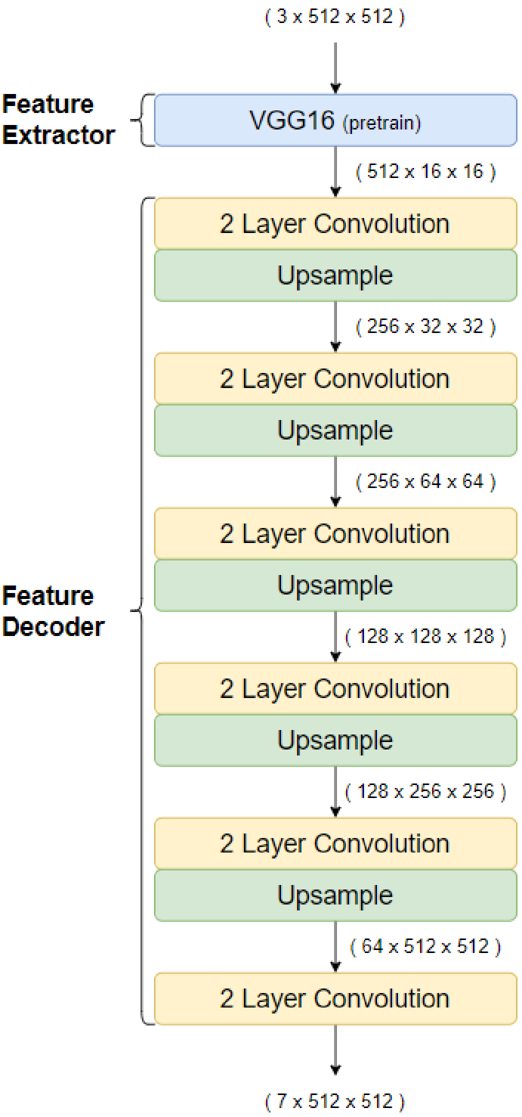
Problem 2: Semantic Segmentation

1. Draw the network architecture of your VGG16-FCN32s model (model A).

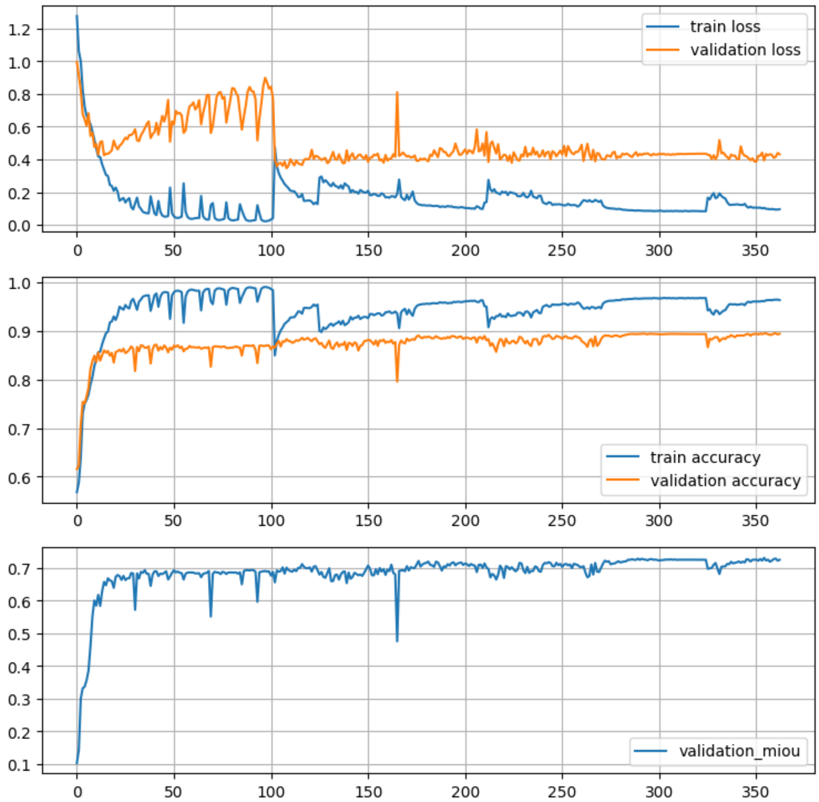


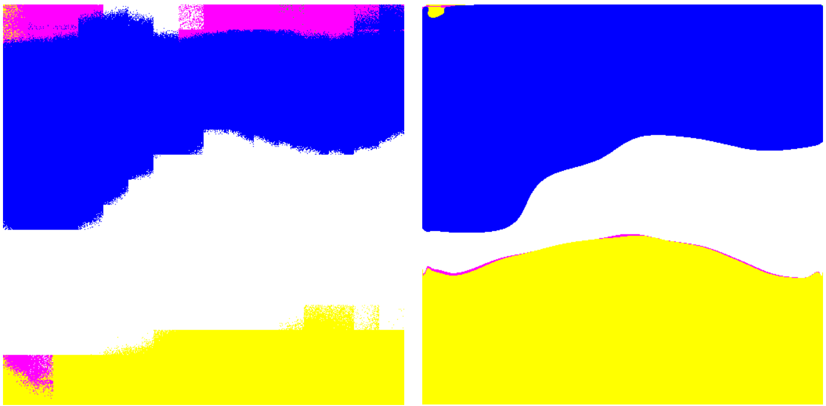


2. Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model.



Model B 選擇使用(convolution layer + bilinear upsample)*5 來取代 model A 的 ConvTranspose layer，類似於 SegNet 的架構。並在訓練過程中陸續加入 data augmentation，將 image 和 mask 同時 flip 或 crop，使 miou 獲得些微的提升。





validation/0013_sat.jpg

prediction @early epoch

上圖分別為 Model A (FCN32)和 Model B (improved)對原圖進行影像分割的圖片，左邊為 model A 有明顯的方塊狀且邊緣有噴霧感，而右圖 model B 則邊緣平滑，由此觀察到兩種模型分割出來的圖片有些許差異。

3. Report mIoUs of two models on the validation set.

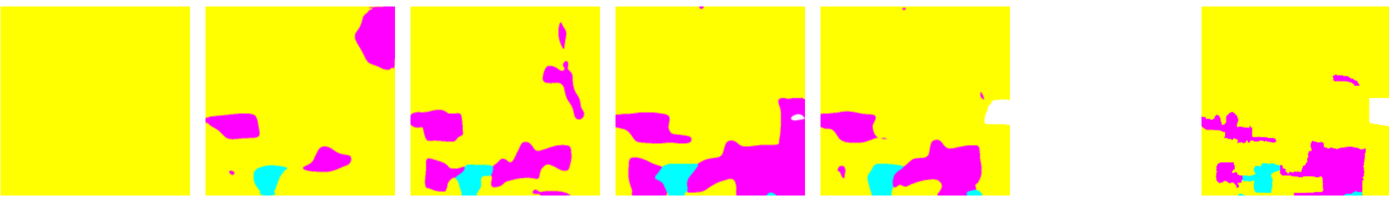
Model A validation mIoU: 69.4%

Model B validation mIoU: 73.15%

4. Show the predicted segmentation mask of “validation/0013_sat.jpg”, “validation/0062_sat.jpg”, “validation/0104_sat.jpg” during the early, middle, and the final stage during the training process of the improved model.

validation/0013_sat.jpg @epoch 1, 8, 101, 286, 363 Truth mask

validation/0062_sat.jpg @epoch 1, 15, 101, 201, 363 Truth mask



validation/0104_sat.jpg @epoch 1, 9, 101, 201, 363 Truth mask