

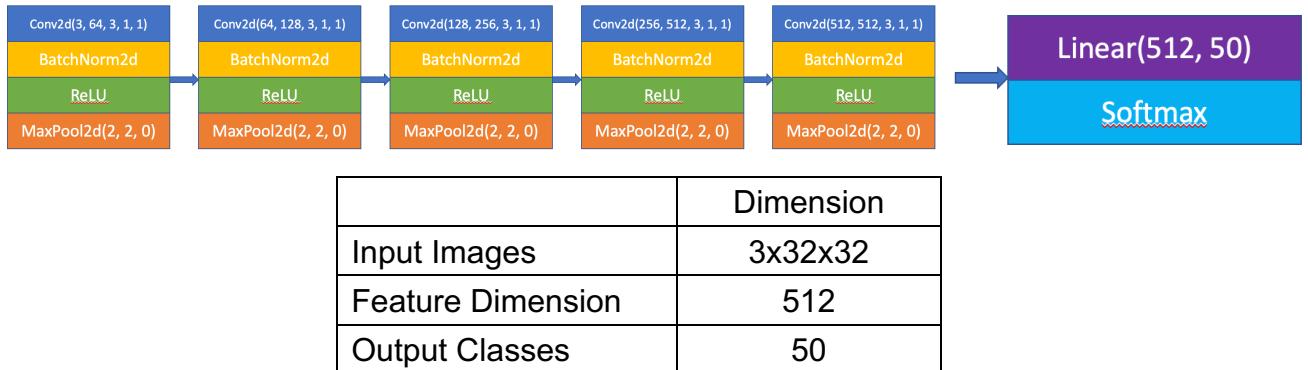
Deep Learning for Computer Vision HW1

R10945019 陳翰儒

Problem1 – Image Classification

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

Method A:

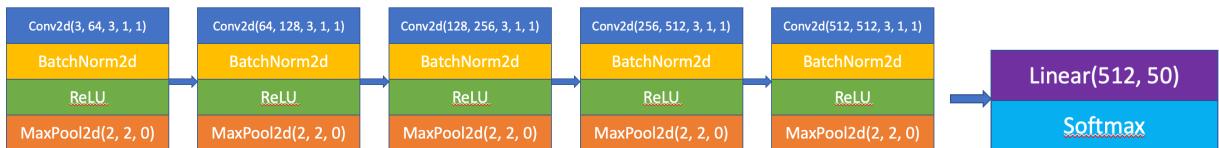


2. Report accuracy of your models (both A, B) on validation set.

	Validation Accuracy
Method A	0.615
Method B	0.903

3. Report your implementation details of model A.

- 前處理：僅 Normalization。
- Augmentation: Random Perspective (distortion_scale=0.6, p=0.5)
- 模型架構：(未使用 pre-trained weight)



由五層的卷積層進行特徵的學習，各層包含二維卷積（filter 數如圖所示）、Batch Normalization、ReLU、Max Pooling。最後 512 維的 feature 會被餵進一層 fully-connected layer 並以 softmax 輸出 50 個類別的機率。

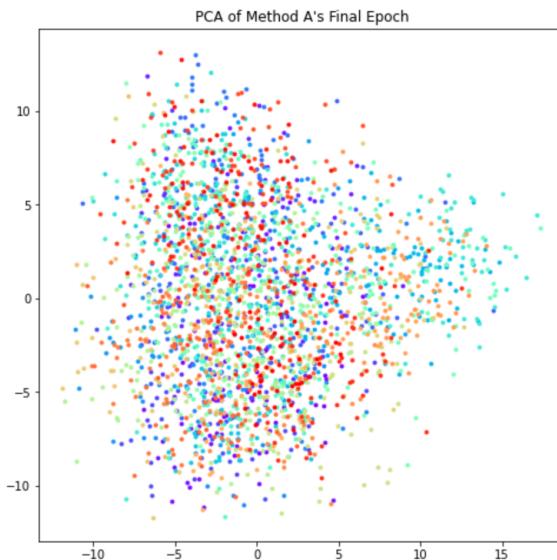
- Hyperparameters:
 - Optimizer: Adam
 - Learning Rate: 0.0001
 - Weight Decay: 0.001
 - Loss Function: cross entropy loss
 - Batch Size: 128
 - Epochs: 50

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

	Method A	Method B
影像解析度	3x32x32	3x96x96
模型架構 – CNN Part	五層普通二維卷積層	ResNeXt101_32x8d 架構
特徵維度	512	2048
參數初始方式	隨機初始	使用在 ImageNet 預訓練過之權重
Epoch 數	50	60
Learning Rate	0.0001	0.00001

主要藉由增加影像解析度、Pretrained Weight、特殊模型架構(ResNeXt101)來提升模型效能。除了以上表格的差異外，其餘設定皆與 Method A 相同。

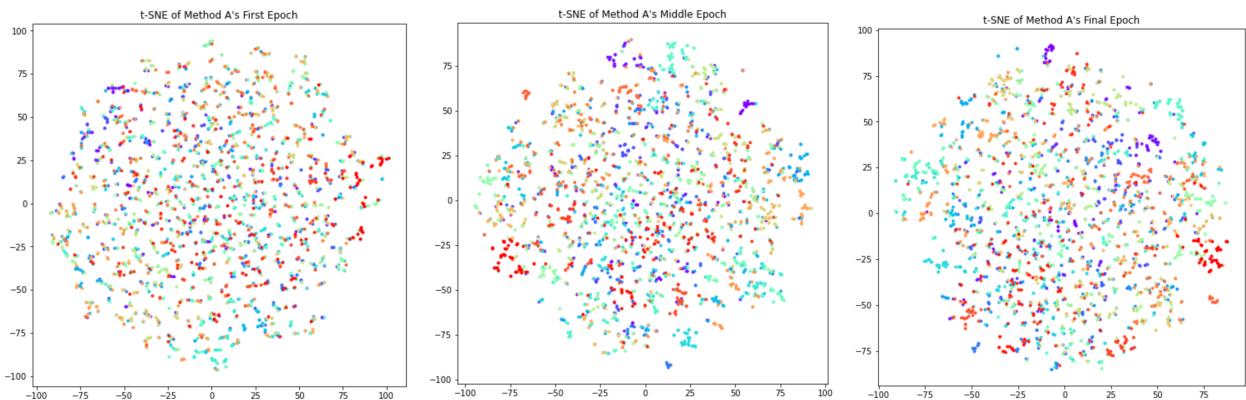
5. Visualize the learned visual representations of model A on validation set by implementing PCA (Principal Component Analysis) on the output of the second last layer. Briefly explain your result of the PCA visualization.



上圖為卷積層所輸出的 512 維特徵進行 PCA 降成二維的結果，每個顏色對應到特定類別。如圖所示，這個模型學出來的特徵基本上都混在一起，只有淺藍色的類別勉強有被分出來在右半部。

而實際上這個模型有嚴重的 overfitting，他的 training accuracy 有 0.889，但 validation accuracy 僅 0.615，故也可預期學到的特徵無法在 validation set 將每個類別分開。

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 last one. Briefly explain the above results.



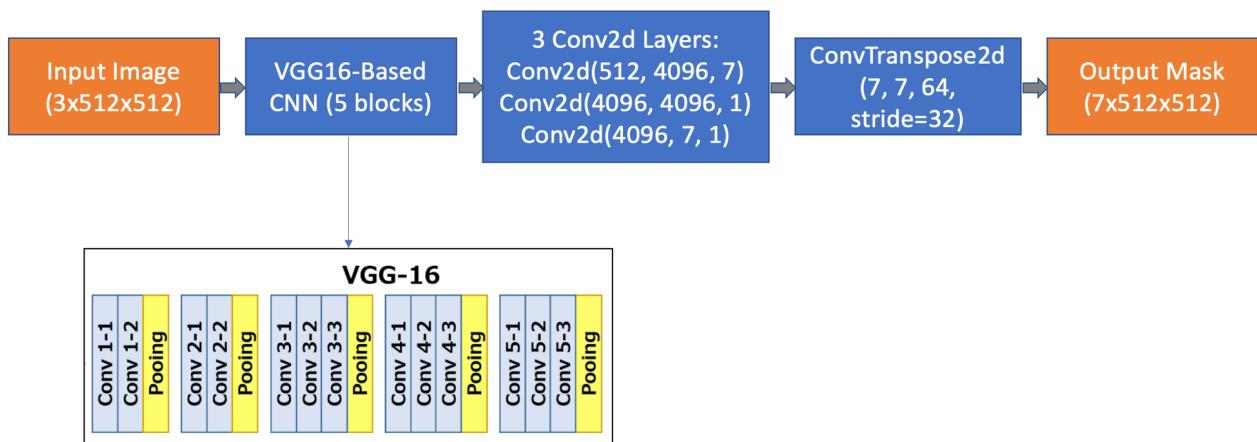
Epoch	1	25	50
Validation Accuracy	0.305	0.581	0.615

上圖為卷積層所輸出的 512 維特徵進行 t-SNE 降成二維的結果，每個顏色對應到特定類別。如圖所示，這個模型學出來的特徵基本上都混在一起。

而實際上這個模型除了有嚴重的 overfitting 外，他各個 epoch 的 accuracy 僅 0.305,、0.581、0.615，皆不盡理想，故也可預期學到的特徵無法在 validation set 將每個類別分開。

Problem2 – Semantic Segmentation

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

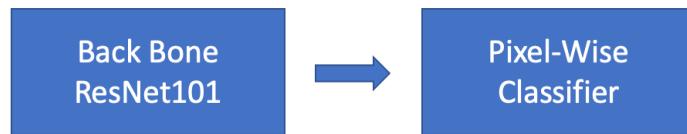


Reference: TA's HW1 Slide

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

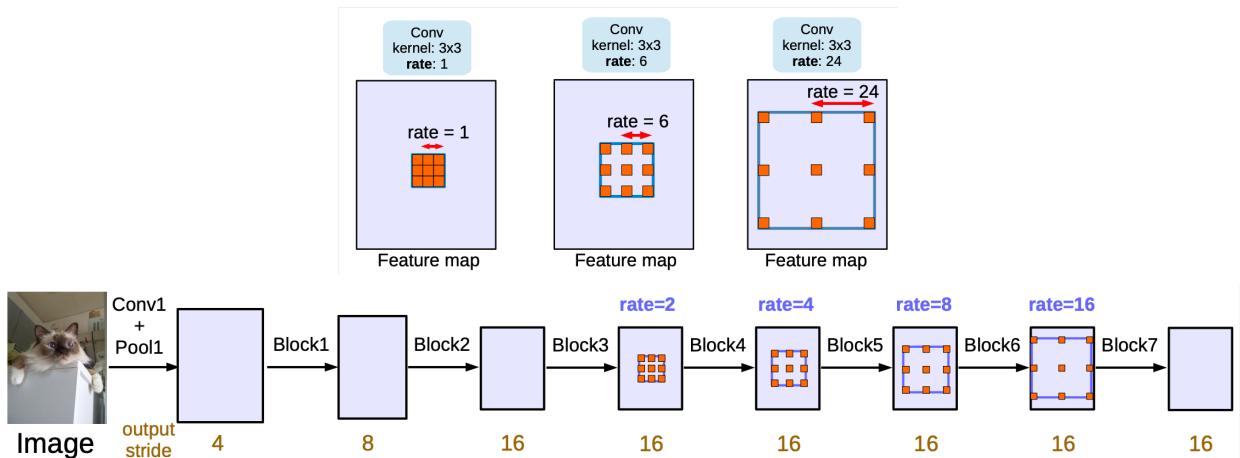
我使用 ensemble 的方式將七個模型各自的輸出取加權平均得到最終結果，這七個模型包含了 DeepLabv3 – ResNet101、TernausNet (a VGG-based U-Net)、ResNet50-based U-Net 三種架構，並採用不同的 learning rate, loss function, augmentation methods 進行訓練。

DeepLabv3 – ResNet101



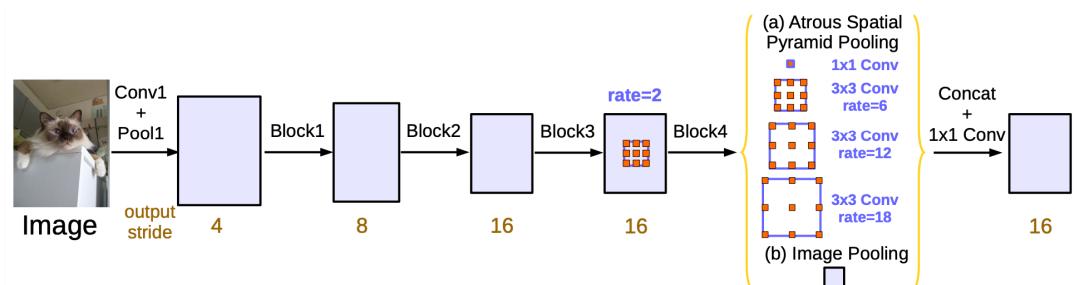
此模型主要有 Back Bone 進行特徵的學習，再由 Pixel-Wise Classifier 產生 Mask。其與 VGG16-FCN32s 的差別包括：

- ResNet101 有使用 pretrained weights。
- Multi-grid atrous(dilated) convolution：在維持相通 kernel size 的情況下增加模型的視野，如下圖。同時在模型內設計多種不同的 rate，使 feature map 維持一定大小，有利於 semantic segmentation。

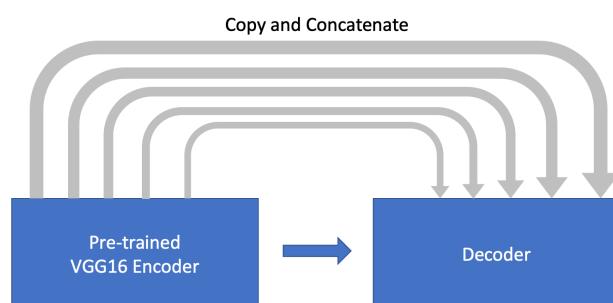


Reference: Chen, Liang-Chieh, et al. "Rethinking atrous convolution for semantic image segmentation." *arXiv preprint arXiv:1706.05587*(2017).

- Atrous spatial pyramid pooling：將不同 rate 的 atrous convolution 結果 concat 在一起（如下圖），使該 feature map 包含不同 scale 的資訊。之後利用 kernel size 為 1 的 filter 進行卷積運算得到最終 output。



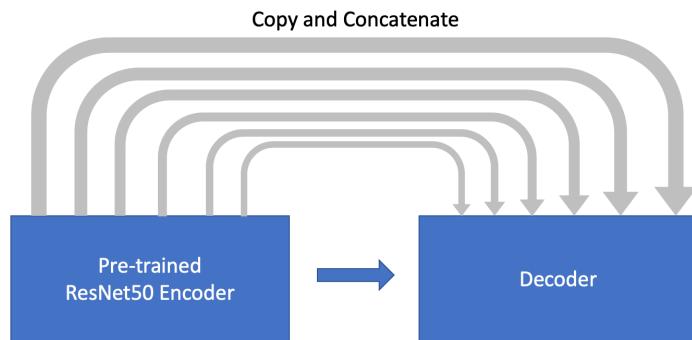
TernausNet (VGG-Based U-Net)



此模型使用 VGG16 作為 encoder，並在 decoder 的地方放入 Transpose Convolution 使其大小還原為原本大小，並在最後做 1×1 Convolution 得到最終 output。其與 VGG16-FCN32s 的差別包括：

- VGG16 有使用 pretrained weights。
- 在 VGG16 的部分會將每個 block 出來的 feature map 接到 decoder 中對稱對應的 feature map，維持在 encoder 中學到的特徵資訊。

ResNet50-Based U-Net

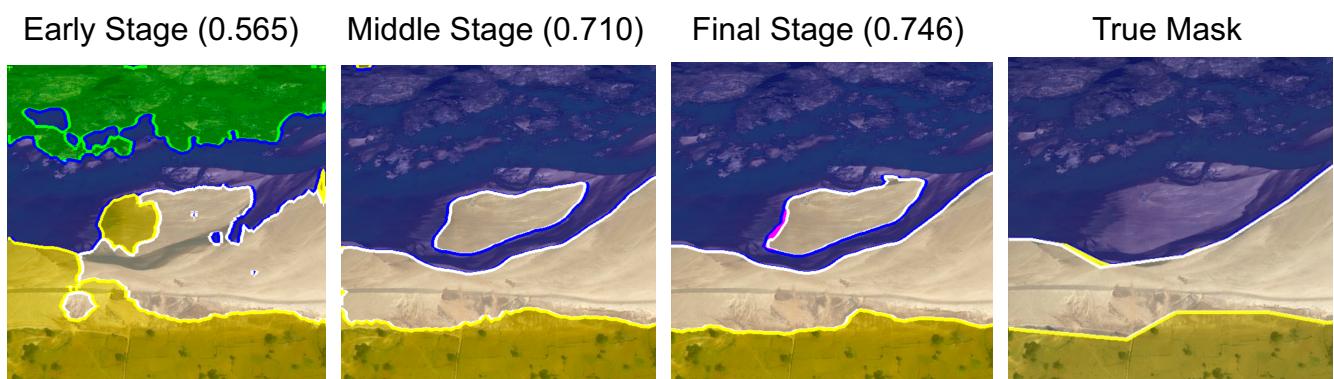


其構造與 TernausNet 大致相同，差別為使用 pretrain 過的 ResNet50 作為 encoder。

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

	Validation mIoUs
Method A	0.103
Method B	0.746

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.





Note. 括號內為 validation mIoU 結果。

Note. 因為我使用 ensemble，故我這邊的 early, middle stage 也是用那七個模型各自的 early, middle stage 做 ensemble。