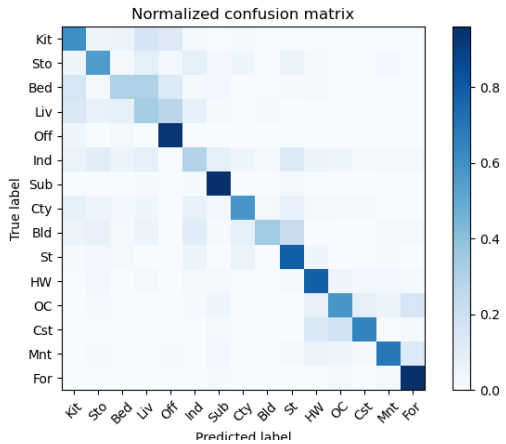
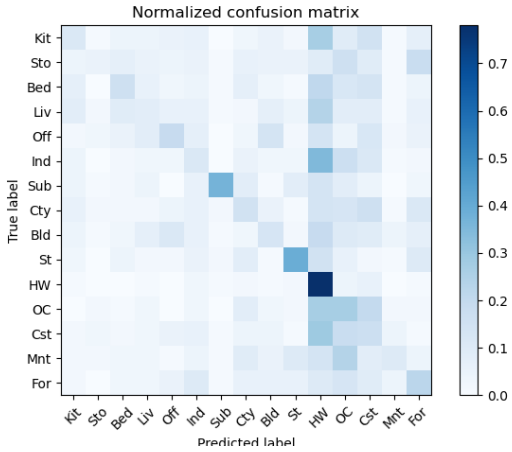
Computer Vision HW2 Report

Student ID: R11528025
Name: 劉容綺

Part 1. (10%)

• Plot confusion matrix of two settings. (i.e. Bag of sift and tiny image) (5%)
Ans:

Bag of shift	tiny image
	

• Compare the results/accuracy of both settings and explain the result. (5%)
Ans:

Accuracy	
tiny image	0.2186
Bag of sift	0.642

如預期，Tiny Image 的準確性遠低於 Bag of SIFT，因為 Tiny Image 特徵僅捕獲非常基本的圖像特徵，例如顏色分佈和亮度，而 Bag of SIFT 特徵捕獲更詳細和有鑑別力的特徵，例如紋理和局部形狀信息。在產生 test_image_feats.pkl 與 train_image_feats.pkl 的過程非常漫長，可能是我的寫法或電腦性能的問題，也有時出現 RAM 記憶體不足的情況，後來我透過提高 build_vocabulary 的 step sample，調整 vocab size 和 batch size 來提高準確度。

Part 2. (25%)

• Report accuracy of both models on the validation set. (2%)
Ans:

Accuracy	
MyNet	0.85

ResNet18	0.9184

• Print the network architecture & number of parameters of both models. What is the main difference between ResNet and other CNN architectures? (5%)
Ans:

MyNet:

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 32, 32, 32]             896
       BatchNorm2d-2         [-1, 32, 32, 32]              64
              ReLU-3         [-1, 32, 32, 32]               0
            Conv2d-4         [-1, 32, 32, 32]           9,248
       BatchNorm2d-5         [-1, 32, 32, 32]              64
              ReLU-6         [-1, 32, 32, 32]               0
         MaxPool2d-7         [-1, 32, 16, 16]               0
            Conv2d-8         [-1, 64, 16, 16]          18,496
       BatchNorm2d-9         [-1, 64, 16, 16]             128
             ReLU-10         [-1, 64, 16, 16]               0
           Conv2d-11         [-1, 64, 16, 16]          36,928
      BatchNorm2d-12         [-1, 64, 16, 16]             128
             ReLU-13         [-1, 64, 16, 16]               0
        MaxPool2d-14           [-1, 64, 8, 8]               0
           Conv2d-15          [-1, 128, 8, 8]          73,856
      BatchNorm2d-16          [-1, 128, 8, 8]             256
             ReLU-17          [-1, 128, 8, 8]               0
           Conv2d-18          [-1, 128, 8, 8]         147,584
      BatchNorm2d-19          [-1, 128, 8, 8]             256
             ReLU-20          [-1, 128, 8, 8]               0
        MaxPool2d-21          [-1, 128, 4, 4]               0
           Linear-22                [-1, 512]         524,800
      BatchNorm1d-23                [-1, 512]           1,024
             ReLU-24                [-1, 512]               0
           Linear-25                 [-1, 10]           5,130
================================================================
Total params: 819,594
Trainable params: 819,594
Non-trainable params: 0
```

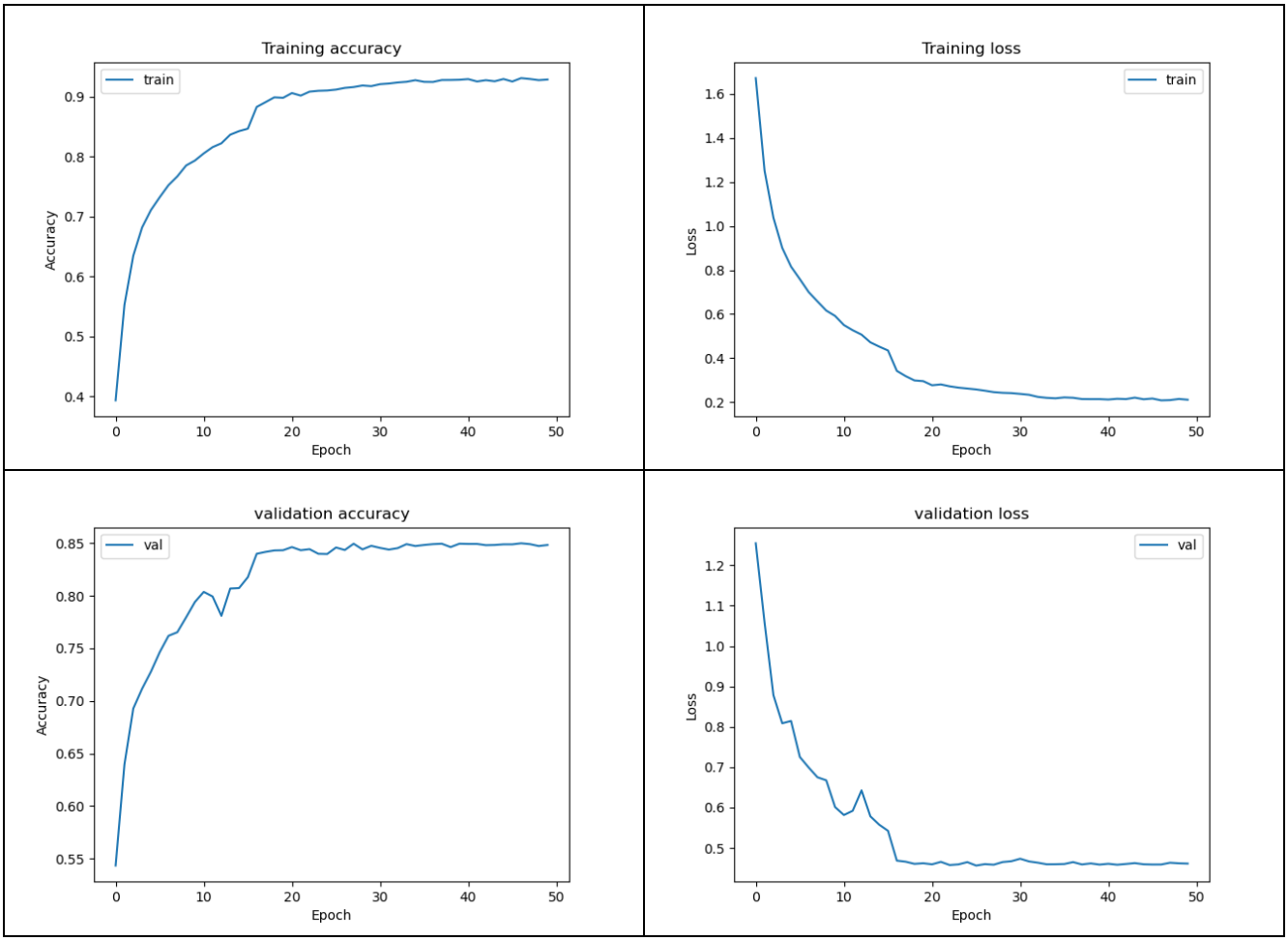
ResNet:

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
ResNet18                                 --                        --
├─Sequential: 1-1                        [1, 512, 1, 1]            --
|    └─Conv2d: 2-1                       [1, 64, 32, 32]           1,728
|    └─BatchNorm2d: 2-2                  [1, 64, 32, 32]           128
|    └─ReLU: 2-3                         [1, 64, 32, 32]           --
|    └─MaxPool2d: 2-4                    [1, 64, 32, 32]           --
|    └─Sequential: 2-5                   [1, 512, 1, 1]            --
|    |    └─BasicBlock: 3-1              [1, 64, 32, 32]           75,136
|    |    └─BasicBlock: 3-2              [1, 64, 32, 32]           73,856
|    |    └─BasicBlock: 3-3              [1, 128, 16, 16]          230,144
|    |    └─BasicBlock: 3-4              [1, 256, 8, 8]            919,040
|    |    └─BasicBlock: 3-5              [1, 512, 4, 4]            3,673,088
|    └─AdaptiveAvgPool2d: 2-6            [1, 512, 1, 1]            --
|    └─Linear: 2-7                       [1, 10]                   5,130
==========================================================================================
Total params: 4,903,114
Trainable params: 4,903,114
Non-trainable params: 0
Total mult-adds (M): 238.21
==========================================================================================
```

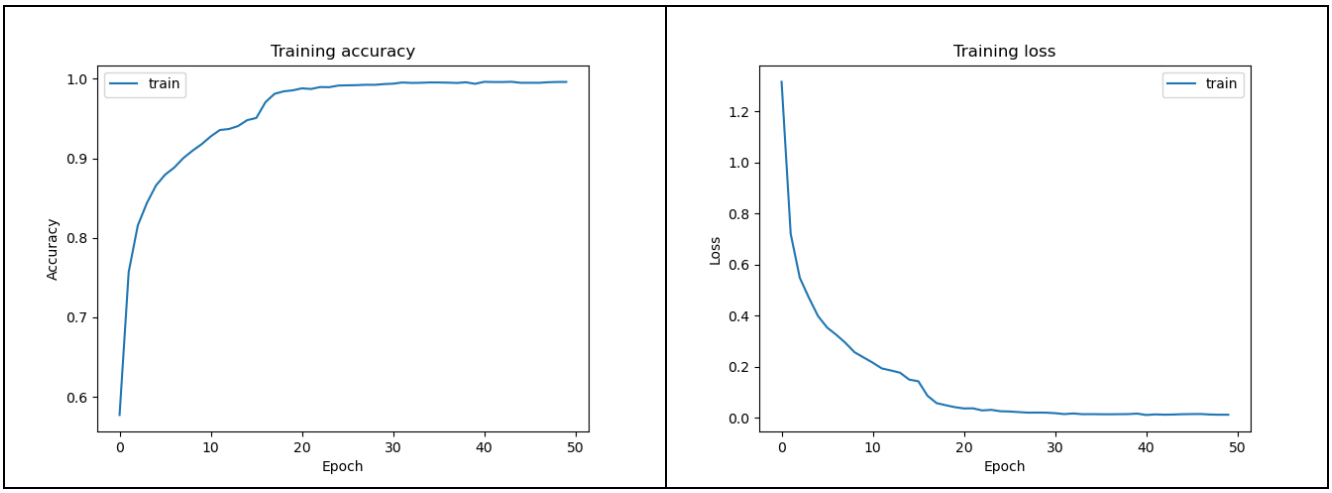
• Plot four learning curves (loss & accuracy) of the training process (train/validation) for both models. Total 8 plots. (8%)

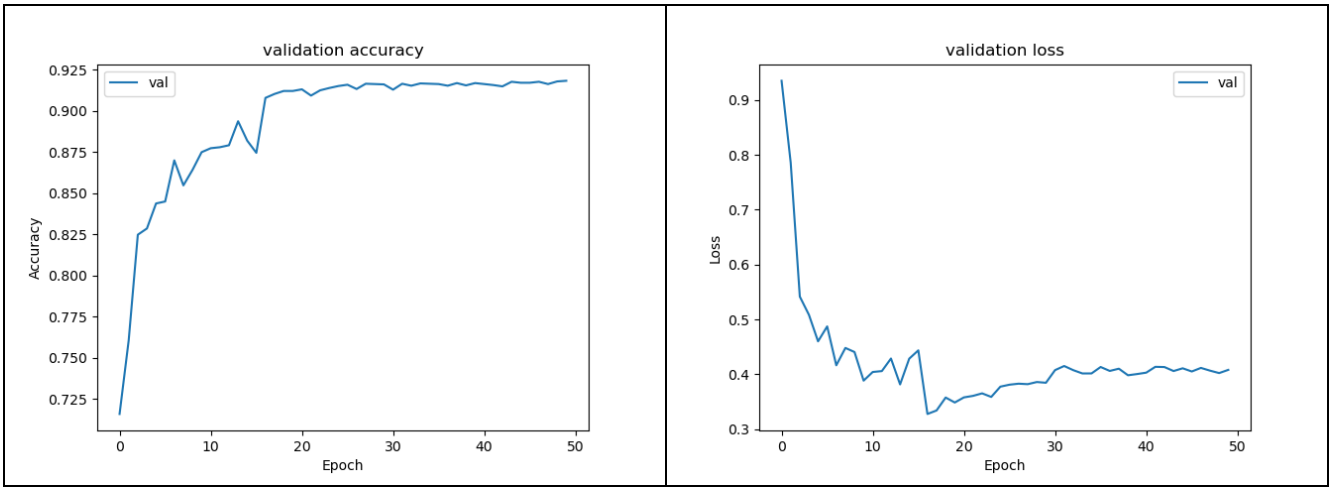
Ans:

MyNet:


Training accuracy
train
Accuracy
Epoch
Training loss
train
Loss
Epoch
validation accuracy
val
Accuracy
Epoch
validation loss
val
Loss
Epoch

ResNet18:


Training accuracy
train
Accuracy
Epoch
Training loss
train
Loss
Epoch



• Briefly describe what method do you apply on your best model? (e.g. data augmentation, model architecture, loss function, etc) (10%)

Ans:

- Batch Normalization 的使用：我在每個 convolutional & fully connected layers 之後都有 Batch Normalization。這可以幫助規範化每一層的輸入，使訓練過程更加穩定並減 overfitting 的可能性。

- ReLU Activation 的使用：我在每個 convolutional & fully connected layer 之後 ReLU 激活函數。這允許模型學習數據中更複雜和非線性的關係，這有助於提高其準確性。

- Max Pooling 的使用：我在每兩個 convolutional layer 後使用 Max Pooling layer,這有助於對特徵圖進行下採樣，降低數據的空間維度，並使模型對輸入數據的變化更加穩健。

- Depth of the Network：該模型有 3 convolutional layers & 2 fully connected layers。這種網絡深度允許模型學習更複雜和層次化的數據表示，從而提高其準確性。我一開始的 model 並沒有達到 accuracy baseline,主要是靠這個方法達成。