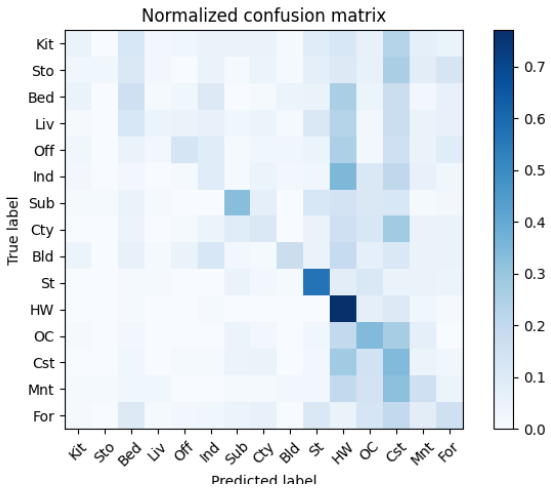
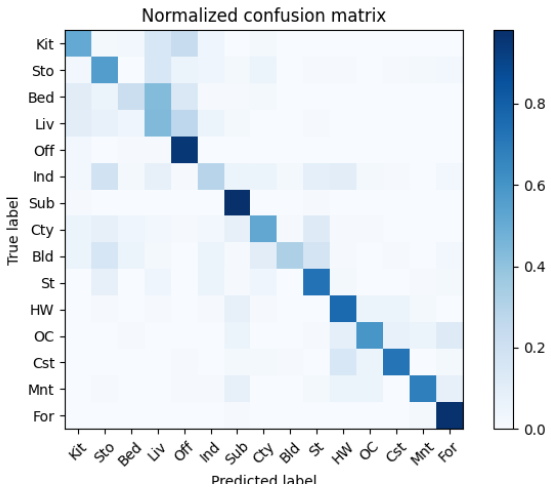
Computer Vision HW2 Report

Student ID: R12K41025
Name: 杜冠廷

Part 1. (10%)

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

Tiny Image	Bag of SIFT
	

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

Accuracy	
Tiny Image	0.23
Bag of SIFT	0.616

As expected, the performance of Tiny Image is worse than that of Bag of SIFT because the Tiny Image approach can only capture very basic image features, such as color distribution and brightness. In contrast, Bag of SIFT can extract more detailed and discriminative features, such as textures, edges, and local shapes. The process of building the vocabulary and generating the vocab.pkl file takes a significant amount of time. To reduce this, we limited the vocabulary construction to only 1/10 of the images from the training set and extracted a maximum of 80 feature points per image.

Part 2. (25%)

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

Ans:

Accuracy	
MyNet	0.7464
ResNet18	0.9232

• 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, 64, 224, 224]           1,792
              ReLU-2         [-1, 64, 224, 224]               0
            Conv2d-3         [-1, 64, 224, 224]          36,928
              ReLU-4         [-1, 64, 224, 224]               0
         MaxPool2d-5         [-1, 64, 112, 112]               0
            Conv2d-6        [-1, 128, 112, 112]          73,856
              ReLU-7        [-1, 128, 112, 112]               0
            Conv2d-8        [-1, 128, 112, 112]         147,584
              ReLU-9        [-1, 128, 112, 112]               0
        MaxPool2d-10          [-1, 128, 28, 28]               0
           Conv2d-11          [-1, 256, 28, 28]         295,168
             ReLU-12          [-1, 256, 28, 28]               0
           Conv2d-13          [-1, 256, 28, 28]         590,080
             ReLU-14          [-1, 256, 28, 28]               0
        MaxPool2d-15            [-1, 256, 7, 7]               0
          Flatten-16                [-1, 12544]               0
           Linear-17                [-1, 1024]      12,846,080
             ReLU-18                [-1, 1024]               0
           Linear-19                 [-1, 512]         524,800
             ReLU-20                 [-1, 512]               0
           Linear-21                  [-1, 10]           5,130
================================================================
Total params: 14,521,418
Trainable params: 14,521,418
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.57
Forward/backward pass size (MB): 160.23
Params size (MB): 55.39
Estimated Total Size (MB): 216.20
----------------------------------------------------------------
```

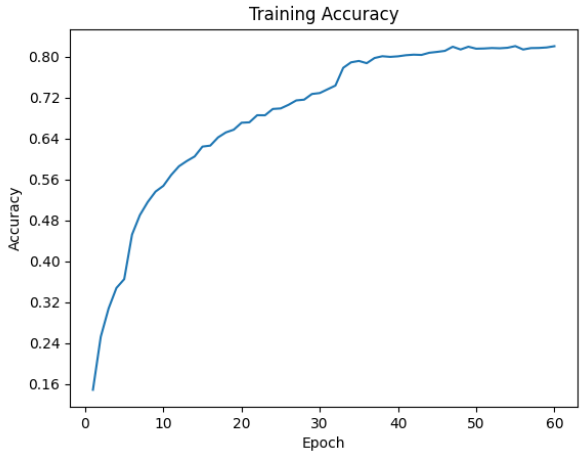
ResNet18

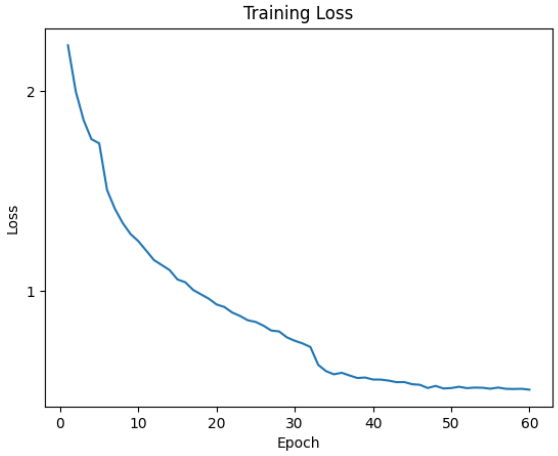
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 64, 114, 114]           1,728
       BatchNorm2d-2         [-1, 64, 114, 114]             128
              ReLU-3         [-1, 64, 114, 114]               0
         MaxPool2d-4           [-1, 64, 57, 57]               0
            Conv2d-5           [-1, 64, 57, 57]          36,864
       BatchNorm2d-6           [-1, 64, 57, 57]             128
              ReLU-7           [-1, 64, 57, 57]               0
            Conv2d-8           [-1, 64, 57, 57]          36,864
       BatchNorm2d-9           [-1, 64, 57, 57]             128
             ReLU-10           [-1, 64, 57, 57]               0
       BasicBlock-11           [-1, 64, 57, 57]               0
            Conv2d-12           [-1, 64, 57, 57]          36,864
       BatchNorm2d-13           [-1, 64, 57, 57]             128
             ReLU-14           [-1, 64, 57, 57]               0
            Conv2d-15           [-1, 64, 57, 57]          36,864
       BatchNorm2d-16           [-1, 64, 57, 57]             128
             ReLU-17           [-1, 64, 57, 57]               0
       BasicBlock-18           [-1, 64, 57, 57]               0
            Conv2d-19          [-1, 128, 29, 29]          73,728
       BatchNorm2d-20          [-1, 128, 29, 29]             256
             ReLU-21          [-1, 128, 29, 29]               0
            Conv2d-22          [-1, 128, 29, 29]         147,456
       BatchNorm2d-23          [-1, 128, 29, 29]             256
            Conv2d-24          [-1, 128, 29, 29]           8,192
       BatchNorm2d-25          [-1, 128, 29, 29]             256
             ReLU-26          [-1, 128, 29, 29]               0
       BasicBlock-27          [-1, 128, 29, 29]               0
            Conv2d-28          [-1, 128, 29, 29]         147,456
       BatchNorm2d-29          [-1, 128, 29, 29]             256
             ReLU-30          [-1, 128, 29, 29]               0
            Conv2d-31          [-1, 128, 29, 29]         147,456
       BatchNorm2d-32          [-1, 128, 29, 29]             256
             ReLU-33          [-1, 128, 29, 29]               0
       BasicBlock-34          [-1, 128, 29, 29]               0
            Conv2d-35          [-1, 256, 15, 15]         294,912
       BatchNorm2d-36          [-1, 256, 15, 15]             512
             ReLU-37          [-1, 256, 15, 15]               0
            Conv2d-38          [-1, 256, 15, 15]         589,824
       BatchNorm2d-39          [-1, 256, 15, 15]             512
            Conv2d-40          [-1, 256, 15, 15]          32,768
       BatchNorm2d-41          [-1, 256, 15, 15]             512
             ReLU-42          [-1, 256, 15, 15]               0
       BasicBlock-43          [-1, 256, 15, 15]               0
            Conv2d-44          [-1, 256, 15, 15]         589,824
       BatchNorm2d-45          [-1, 256, 15, 15]             512
             ReLU-46          [-1, 256, 15, 15]               0
            Conv2d-47          [-1, 256, 15, 15]         589,824
       BatchNorm2d-48          [-1, 256, 15, 15]             512
             ReLU-49          [-1, 256, 15, 15]               0
       BasicBlock-50          [-1, 256, 15, 15]               0
            Conv2d-51            [-1, 512, 8, 8]       1,179,648
       BatchNorm2d-52            [-1, 512, 8, 8]           1,024
             ReLU-53            [-1, 512, 8, 8]               0
            Conv2d-54            [-1, 512, 8, 8]       2,359,296
       BatchNorm2d-55            [-1, 512, 8, 8]           1,024
            Conv2d-56            [-1, 512, 8, 8]         131,072
       BatchNorm2d-57            [-1, 512, 8, 8]           1,024
             ReLU-58            [-1, 512, 8, 8]               0
       BasicBlock-59            [-1, 512, 8, 8]               0
            Conv2d-60            [-1, 512, 8, 8]       2,359,296
       BatchNorm2d-61            [-1, 512, 8, 8]           1,024
             ReLU-62            [-1, 512, 8, 8]               0
            Conv2d-63            [-1, 512, 8, 8]       2,359,296
       BatchNorm2d-64            [-1, 512, 8, 8]           1,024
             ReLU-65            [-1, 512, 8, 8]               0
       BasicBlock-66            [-1, 512, 8, 8]               0
AdaptiveAvgPool2d-67            [-1, 512, 1, 1]               0
           Linear-68                  [-1, 10]           5,130
           ResNet-69                  [-1, 10]               0
================================================================
Total params: 11,173,962
Trainable params: 11,173,962
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.57
Forward/backward pass size (MB): 67.01
Params size (MB): 42.63
Estimated Total Size (MB): 110.21
----------------------------------------------------------------
```

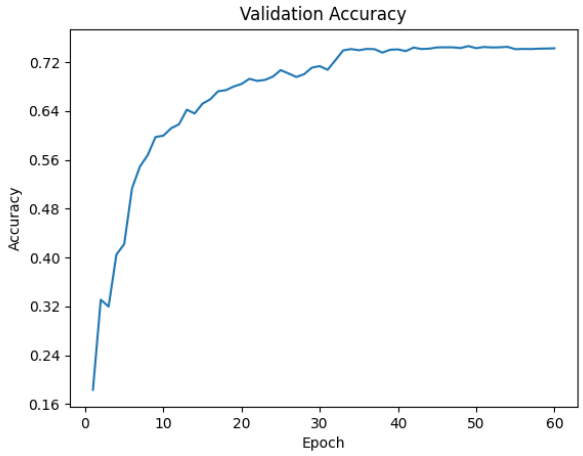
• Plot four learning curves (loss & accuracy) of the training process (train/validation) for both models. Total 8 plots. (8%)

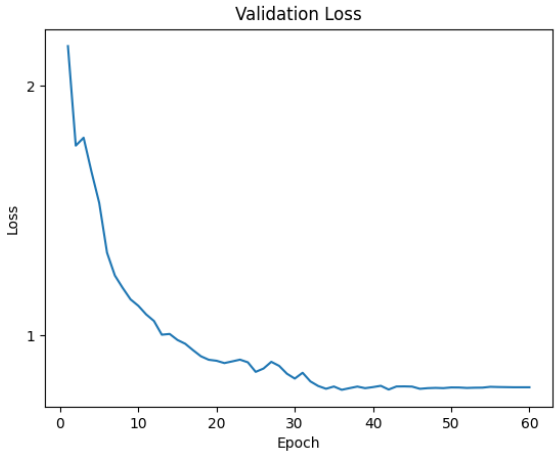
Ans:

MyNet

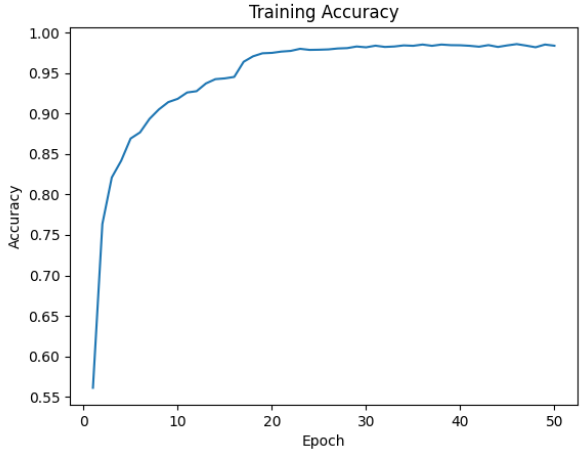


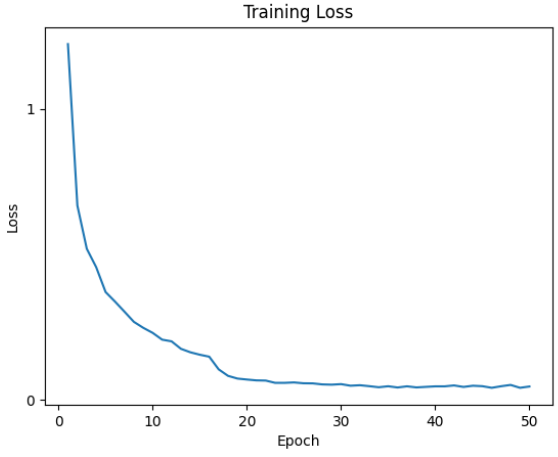


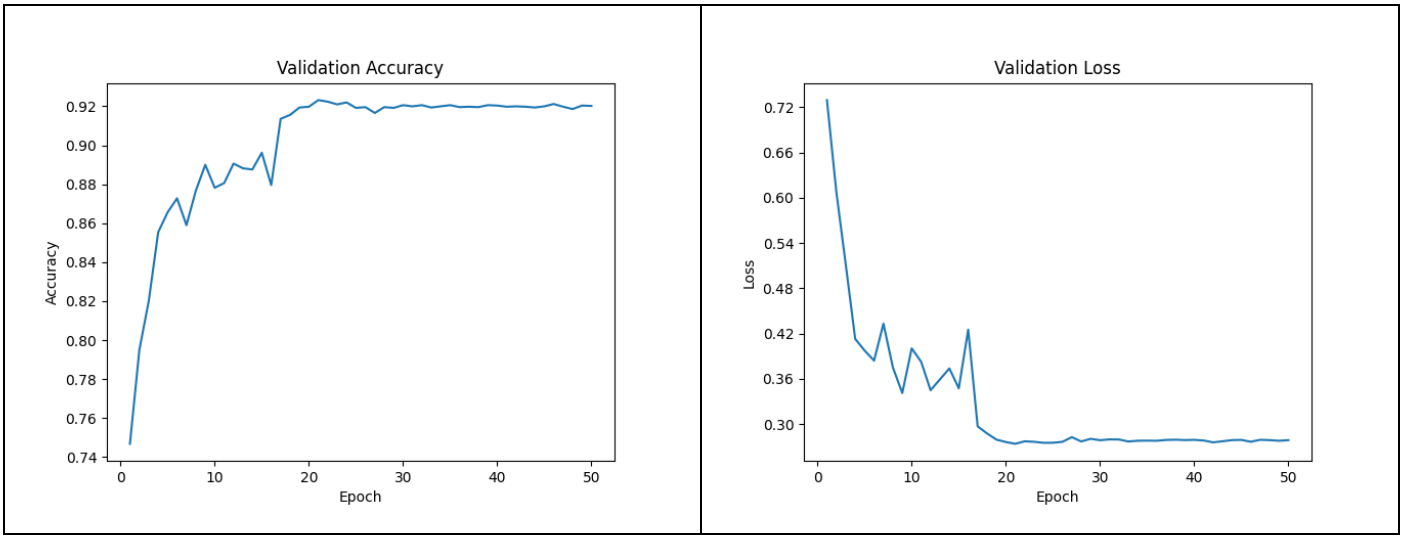




ResNet18







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

Since ResNet18 is designed for classifying the large-scale dataset, I resized the images on CIFAR-10 to 224×224 to prevent excessive information loss caused by the Max Pooling layer. Additionally, due to the limited amount of training data, I applied several data augmentation techniques to prevent overfitting:

1. transforms.RandomResizedCrop(224, scale=(256/480, 1.0))
2. transforms.RandomHorizontalFlip()
3. transforms.RandomRotation(10)
4. transforms.RandomGrayscale()`