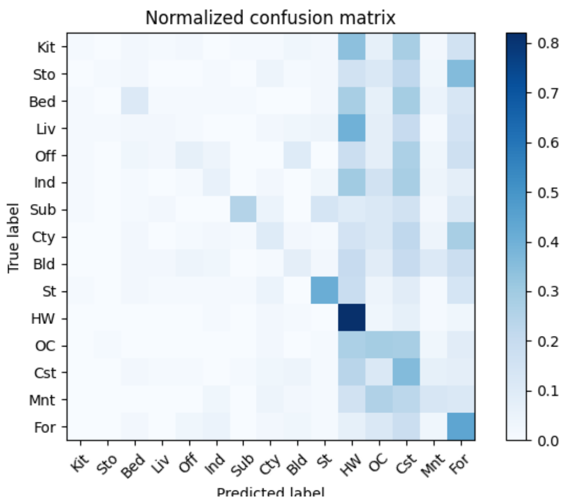
Computer Vision HW2 Report

Student ID: R10522606
Name:曾柏翔

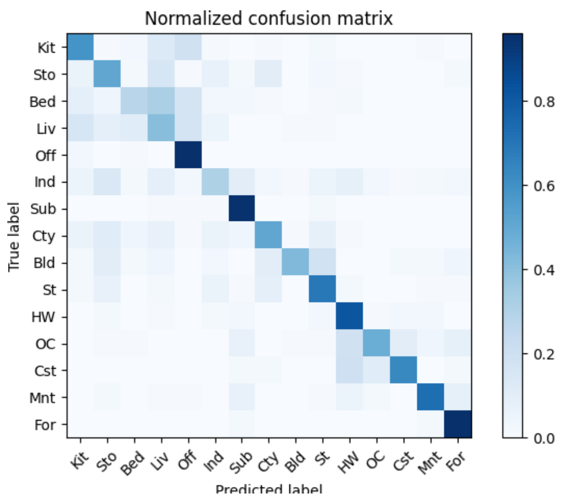
Part 1. (10%)

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

Ans:



Tiny image



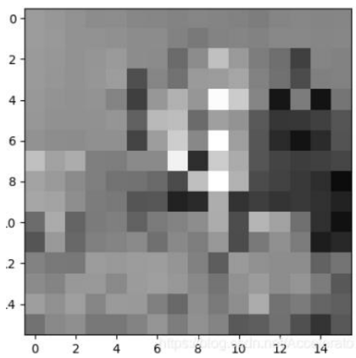
Bag of sift

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

Ans:

Tiny image : **0.21**、Bag of sift : **0.62**

Tiny image 我是透過 cv2.resize 的方式來將圖片變為 16 x 16，再通過 flatten()來提取特徵，這方法簡單且速度快，但相對的精確度就不高，經過調整一些參數後，仍卡在 20%左右，下圖為 resize 示意圖。



Bag of sift 首先是建立 vocabulary，透過 dsift()得到 descriptors，但這邊只要得到代表性的 sample 即可，因此只取出一半的 descriptors，進入 kmeans(vocab_size=400)中，得到 vocabulary。而 bag of sift 也透過 dsift()得到 descriptors，再利用 distance.cdist()找最相近的 vocabulary，最後透過 histogram 的方式得到 features。

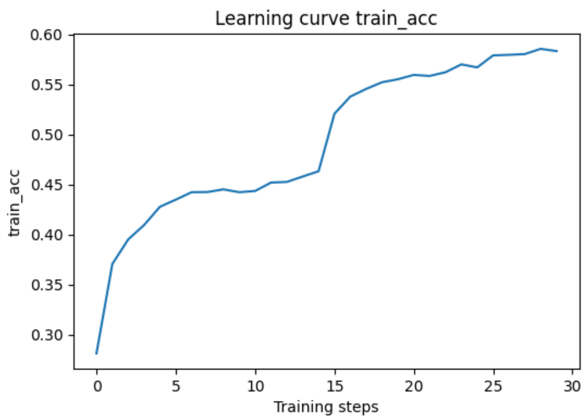
兩種方法最終丟進 KNN 中進行 predicts，經過嘗試後，得到 k＝6 有最好的精確度，以及 City-block 距離效果比 Euclidean 距離好。

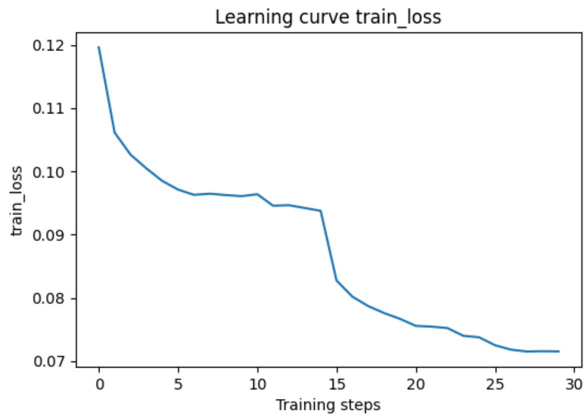
Part 2. (35%)

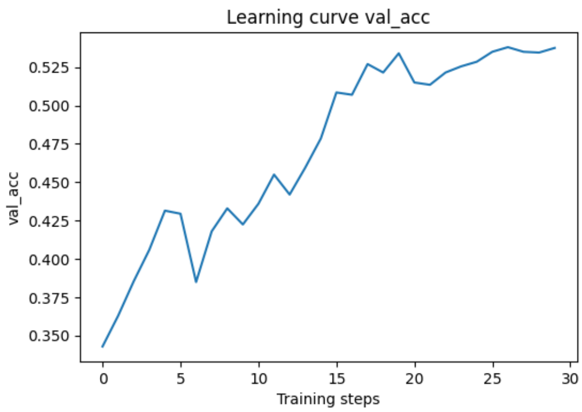
• Compare the performance on residual networks and LeNet. Plot the learning curve (loss and accuracy) on both training and validation sets for both 2 schemes. 8 plots in total. (20%)

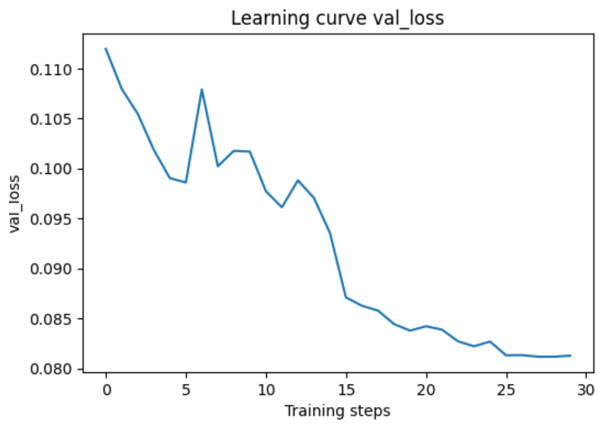
Ans:

LeNet

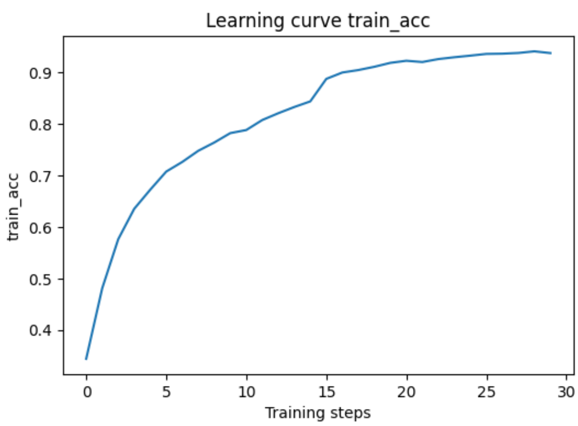


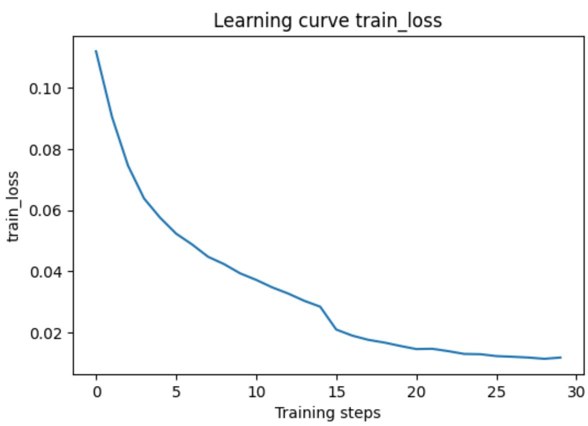


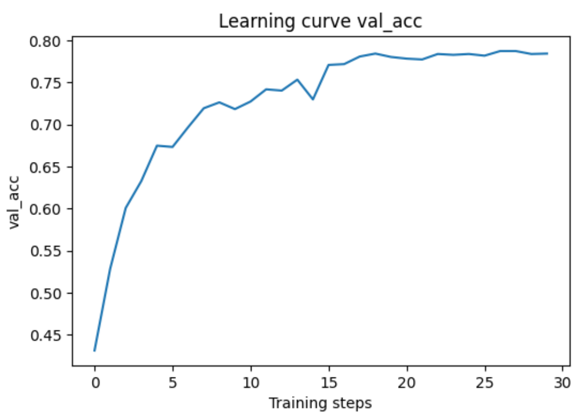


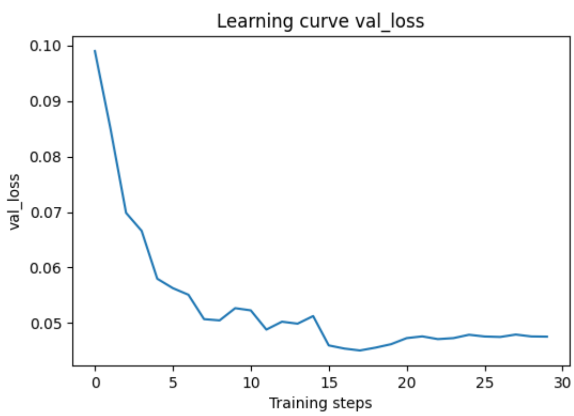


ResNet









• Attach basic information of the model you use including model architecture and number of the parameters. (5%)

Ans:

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 64, 32, 32]           1,792
            Conv2d-2           [-1, 64, 32, 32]          36,928
       BatchNorm2d-3           [-1, 64, 32, 32]             128
              ReLU-4           [-1, 64, 32, 32]               0
    residual_block-5           [-1, 64, 32, 32]               0
            Conv2d-6           [-1, 64, 32, 32]          36,928
         MaxPool2d-7           [-1, 64, 16, 16]               0
       BatchNorm2d-8           [-1, 64, 16, 16]             128
              ReLU-9           [-1, 64, 16, 16]               0
           Conv2d-10           [-1, 64, 16, 16]          36,928
      BatchNorm2d-11           [-1, 64, 16, 16]             128
             ReLU-12           [-1, 64, 16, 16]               0
   residual_block-13           [-1, 64, 16, 16]               0
           Conv2d-14          [-1, 128, 16, 16]          73,856
        MaxPool2d-15            [-1, 128, 8, 8]               0
      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
   residual_block-21            [-1, 128, 8, 8]               0
           Linear-22                  [-1, 512]       4,194,816
             ReLU-23                  [-1, 512]               0
           Linear-24                   [-1, 10]           5,130
================================================================
Total params: 4,534,858
Trainable params: 4,534,858
Non-trainable params: 0
----------------------------------------------------------------
```

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

Ans:

LeNet : **0.52** 、 ResNet : **0.79**

Data augmentation	RandomHorizontalFlip RandomVerticalFlip RandomRotation(degrees=5)	**Optimizer**	SGD
		Scheduler	MultiStepLR
		Epoch	30
Loss function	CrossEntropy	**Batch**	16

首先輸入一層 Conv2d 後，接著一層 residual block(Conv2d + BatchNorm2d + ReLU)，再通過一層 cnn_layer2(Conv2d + MaxPool2d + BatchNorm2d + ReLU)，再將前面的 residual block 接進來，接著通過 cnn_layer4(Conv2d + MaxPool2d + BatchNorm2d + ReLU)，再照前面的方式接上 residual block，最後通過 fully connected。

其中加上 max pooling 跟 batch normalization 有效地防止 overfitting 與 vanishing gradient，並對精確度有所提升。