張宇喆 311513058

torchsummay output:

```
            Conv2d-1           [-1, 28, 28, 28]             784
       BatchNorm2d-2           [-1, 28, 28, 28]              56
              ReLU-3           [-1, 28, 28, 28]               0
         AvgPool2d-4           [-1, 28, 14, 14]               0
            Conv2d-5           [-1, 16, 14, 14]             464
              ReLU-6           [-1, 16, 14, 14]               0
 AdaptiveAvgPool2d-7           [-1, 16, 28, 28]               0
            Conv2d-8           [-1, 56, 14, 14]          14,168
       BatchNorm2d-9           [-1, 56, 14, 14]             112
             ReLU-10           [-1, 56, 14, 14]               0
        AvgPool2d-11             [-1, 56, 7, 7]               0
           Conv2d-12            [-1, 136, 7, 7]          68,680
      BatchNorm2d-13            [-1, 136, 7, 7]             272
             ReLU-14            [-1, 136, 7, 7]               0
        AvgPool2d-15            [-1, 136, 3, 3]               0
           Conv2d-16             [-1, 16, 3, 3]           2,192
             ReLU-17             [-1, 16, 3, 3]               0
AdaptiveAvgPool2d-18           [-1, 16, 16, 16]               0
          Dropout-19                 [-1, 1224]               0
           Linear-20                   [-1, 10]          12,250
================================================================
Total params: 98,978
Trainable params: 98,978
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 1.16
Params size (MB): 0.38
Estimated Total Size (MB): 1.55
----------------------------------------------------------------
```

Methods I used:

1.我使用知識蒸餾去做模型壓縮，teacher model 是助教給的 Resnet50，student model 是三層卷積神經網路，除了讓 student model 去學習 teacher 的 soft label，我還有讓 student 去學 teacher 裡面的特徵，和真正的 label。

2.訓練時我每個 epoch 都有做 validation，然後選 accuracy 最好的。

3.訓練時有做 random agumentation，test 時沒有。