Deep Learning for Computer Vision
Homework 4

R10522606 曾柏翔

Problem 1

1. Describe the architecture & implementation details of your model.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 64, 84, 84]           1,792
       BatchNorm2d-2           [-1, 64, 84, 84]             128
              ReLU-3           [-1, 64, 84, 84]               0
         MaxPool2d-4           [-1, 64, 42, 42]               0
            Conv2d-5           [-1, 64, 42, 42]          36,928
       BatchNorm2d-6           [-1, 64, 42, 42]             128
              ReLU-7           [-1, 64, 42, 42]               0
         MaxPool2d-8           [-1, 64, 21, 21]               0
            Conv2d-9           [-1, 64, 21, 21]          36,928
      BatchNorm2d-10           [-1, 64, 21, 21]             128
             ReLU-11           [-1, 64, 21, 21]               0
        MaxPool2d-12           [-1, 64, 10, 10]               0
           Conv2d-13           [-1, 64, 10, 10]          36,928
      BatchNorm2d-14           [-1, 64, 10, 10]             128
             ReLU-15           [-1, 64, 10, 10]               0
        MaxPool2d-16             [-1, 64, 5, 5]               0
          Convnet-17                 [-1, 1600]               0
           Linear-18                  [-1, 800]       1,280,800
             ReLU-19                  [-1, 800]               0
          Dropout-20                  [-1, 800]               0
           Linear-21                  [-1, 800]         640,800
             ReLU-22                  [-1, 800]               0
          Dropout-23                  [-1, 800]               0
           Linear-24                  [-1, 400]         320,400
================================================================
```

將給定的 Convnet 的輸出 1600 維通過 MLP，最終輸出 400 維的 tensor 作為 prototypr。將 query 和 support 使用 Euclidean distance 計算距離後，透過 softmax 找出預測值最大的 class。

Epoch	200	**Episodes**	600
Optimizer	Adam		
Learning rate	1e-4	**Learning rate schedule**	40epoch*0.9
Meta-train	10-way 1-shot	**Meta-test**	5-way 1-shot

Acuracy : 46.14 ± 0.93 %

2. When meta-train and meta-test under the same 5-way 1-shot setting, please report and discuss the accuracy of the prototypical network using 3 different distance function (i.e., Euclidean distance, cosine similarity and parametric function).

	Euclidean distance	Cosine similarity	Parametric function
Accuracy	44.50 ± 0.90 %	44.28 ± 0.91 %	45.62 ± 0.88 %

Parametric function 將 prototype 跟 query 一起輸入兩層 MLP，輸出得到一個兩者相似程度的 Score。

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1             [-1, 1, 1, 400]         320,400
              ReLU-2             [-1, 1, 1, 400]               0
           Dropout-3             [-1, 1, 1, 400]               0
            Linear-4               [-1, 1, 1, 1]             401
================================================================
```

可以看到三者的分數其實不會差太多，推測 Cosine 是因為有+-1 的限制，因此導致判斷距離時無法太精確，有點類似四捨五入的概念。而 Parametric 則是兩者一起輸入 Multilayer Perceptron，進行相似性比較，而非透過相減捨去的方式，因此能更精確地呈現出兩者的距離。

3. When meta-train and meta-test under the same 5-way K-shot setting, please report and compare the accuracy with different shots. (K=1, 5, 10)

	K = 1	K = 5	K = 10
Accuracy	44.50 ± 0.90 %	48.55 ± 0.89%	49.28 ± 0.87%

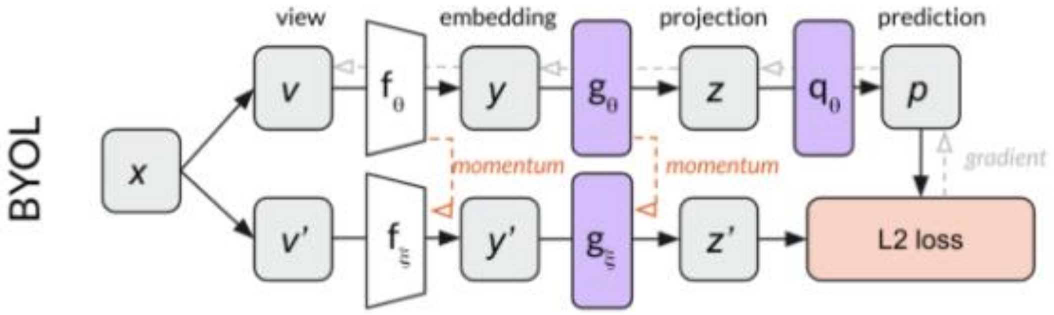
結果符合老師上課所說的，當 K 越大時，精確度能夠越高。這也符合理論，也就是 5 種類別中，每一個類別可以判斷的張數越多，相對的精確度也會越高。也好比人類的行為，能夠有越多的資料來判斷，也就能推斷出越準確的答案。

Problem 2

1. Describe the implementation details of your SSL method for pre-training the ResNet50 backbone.

所使用的方法是 BYOL，架構與參數設定如下圖與表，BYOL 分為三個階段：特徵抽取得到 y、特徵投影得到 z、最後通過預測得到 p，再丟到 loss function 中，計算出 loss 大小。

batch size	32	**transform**	Resize (128,128)
optimizer	Adam		CenterCrop (128)
learning rate	5e-4		ToTensor ()
			expand (3, -1, -1)



2. Following Problem 2-1, please conduct the Image classification on Office-Home dataset as the downstream task for your SSL method.

Setting	Pre-training (Mini-ImageNet)	Fine-tuning (Office-Home dataset)	Mean classification accuracy on valid set (Office-Home dataset)
A	-	Train full model (backbone + classifier)	2.72%
B	w/ label (TAs ha ve provided this backbone)	Train full model (backbone + classifier)	17.81%
C	w/o label (Your SSL pre-trained backbone)	Train full model (backbone + classifier)	38.12%
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	21.26%
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	40.39%

3. Discuss or analyze the results in Problem 2-2

可以明顯看到 Part A，是在沒有任何 pretrain 的狀況下，其結果 2.72% 幾乎可以說是用猜的。再來是 Part B (17.81%)對應 Part D (21.26%)及 Part C (38.12%)對應 Part E (40.39%)，可以看到當 Fine-tuning 為 Fix the backbone 時的 accuracy 都有比較高的現象，猜測是因為 backbone 在 pretrain 時已經訓練好了，因此在使用時只需要將 optimizer 的專注力放在 Train classifier 即可，若是也將 backbone 一起 optimizer，就會導致模型一方面要優化 backbone，另一方面又要想辦法 classifier，不能說一定會比較差，但會增加 model 的負擔。

Reference

[1] Few Shot Learning

https://youtu.be/UkQ2FVpDxHg

[2] Prototypical Net

https://reurl.cc/q1y64n

[3] Self-Supervised Pre-training

https://github.com/lucidrains/byol-pytorch

[4] BYOL

https://reurl.cc/mG6ppG