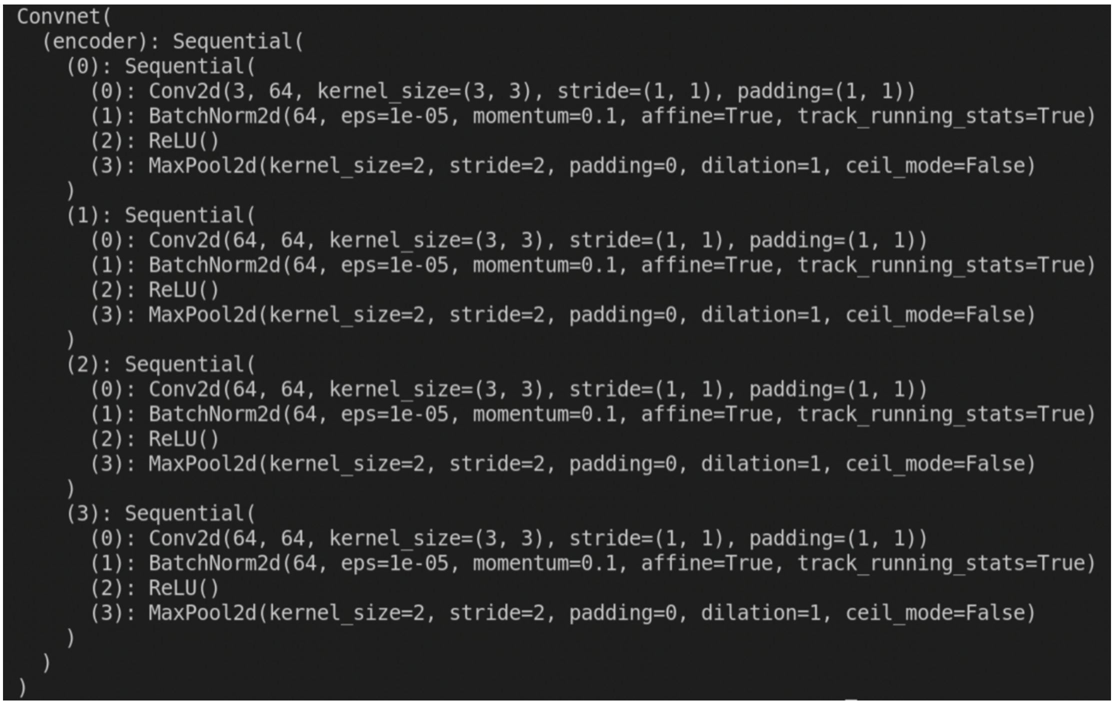
DLCV hw4

R09945021 洪怡庭

- **Problem 1: Few-Shot Learning - Prototypical Network**

1. Describe the architecture & implementation details of your model. (Include but not limited to the number of training episodes, distance function, learning rate schedule, data augmentation, optimizer, and N-way K-shot setting for meta-train and meta-test phase)
   Please report the accuracy on validation set under 5-way 1-shot setting (during inference).



```
Convnet(
  (encoder): Sequential(
    (0): Sequential(
      (0): Conv2d(3, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU()
      (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    )
    (1): Sequential(
      (0): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU()
      (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    )
    (2): Sequential(
      (0): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU()
      (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    )
    (3): Sequential(
      (0): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU()
      (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    )
  )
)
```

Architecture:

使用助教提供的 Conv-4 作為 feature extractor，沒有使用 MLP。

Implementation details:

number of training episodes	600
distance function	Euclidean distance
learning rate schedule	lr_scheduler.StepLR(optimizer, step_size=20, gamma=0.5)
data augmentation	None
optimizer	Adam
meta-train	10 way 1 shot
meta-test	5 way 1 shot

Accuracy: 47.94 +- 0.84 %

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). You should also describe how you design your parametric function。

Distance function	**Val accuracy**
Cosine similarity	36.97 ± 0.80 %
Euclidean distance	46.83 ± 0.82 %
Parametric function	44.18 ± 0.86 %

Parametric function 使用 Manhattan distance，為兩向量相減後取絕對值，結果略差於 Euclidean distance。三種 distance function 準確度比較結果為 Euclidean distance > Parametric function > Cosine similarity，Accuracy 數值如表中所示。

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-shot	**Distance function**	**Val accuracy**
1-shot	Euclidean distance	46.74 ± 0.82 %
5-shot	Euclidean distance	48.11 ± 0.81 %
10-shot	Euclidean distance	48.51 ± 0.76 %

Shot 數值代表每個類別的影像張數，因此 shot 數值越大代表每個類別都有更多訓練資料，準確度應該較高。Accuracy 數值如表中所示，可以看出 1-shot 的表現略差，而 5-shot 及 10-shot 沒有相差太多，但三者的 accuracy 仍有 10-shot > 5-shot > 1-shot 的趨勢，也許 10-shot 需要較多 episode 才能達到收斂。

- **Problem 2: Self-Supervised Pre-training for Image Classification**

1. Describe the implementation details of your SSL method for pre-training the ResNet50 backbone. (Include but not limited to the name of the SSL method you used, data augmentation for SSL, learning rate schedule, optimizer, and batch size setting for this pre-training phase)

SSL method	BYOL
data augmentation for SSL	ColorJitter, RandomGrayscale, Random horizontal flip, GaussianBlur, RandomResizedCrop

learning rate schedule	None
Learning rate	3e-4
optimizer	Adam
batch size	32

2. Following Problem 2-1, please conduct the Image classification on Office-Home dataset as the downstream task for your SSL method. Also, please complete the following Table, which contains different image classification setting, and compare the results.

Setting	Pre-training (Mini-ImageNet)	Fine-tuning (Office-Home dataset)	Classification accuracy on valid set (Office-Home dataset)
A	-	Train full model (backbone + classifier)	0.2266
B	w/ label (TAs have provided this backbone)	Train full model (backbone + classifier)	0.3842
C	w/o label (Your SSL pre-trained backbone)	Train full model (backbone + classifier)	0.4729
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	0.3424
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	0.4163

3. Discuss or analyze the results in Problem 2-2
   i. 五個結果中，A 的準確度最低，可以看出 pretrain 能明顯提升 performance，利用預先訓練好的參數作為模型的初始參數可以幫助模型預先具備較為 general 的知識，使模型有更好的表現也較快收斂。
   ii. 比較使用 supervised 方式 pretrain (B, D)與使用 SSL-BYOL 方式 pretrain(C, E)的結果發現，使用 SSL-BYOL 訓練的模型 fine-tune 後達到比較高的準確度，可見 BYOL 訓練方式確實讓模型學習到擷取重要特徵的能力。
   iii. 比較是否 fix 住 backbone 的結果發現，finetune 時訓練整個 model 可以有比較好的表現，可能因為不同 dataset 仍有不同的特徵需學

習，因此當模型所有參數都能做調整時，能找到更合適的參數。

iv. 由於 backbone 部分已經過 pretrain，finetune 時希望 classifier 部分更新相對快些，因此 finetune 時都有針對各層設置不同 learning rate (A-E 設置相同)。