

## DLCV #HW4

姓名：魏資碩

學號：R10944006

Problem 1-1, you will have to implement the prototypical network as your baseline model to perform 5-way 1-shot classification.

1. (20%) 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).

- The accuracy should be the same as your model performance in P.8.
- TAs will run your code to verify the performance.

**準確度：46.80 +- 0.87 %**

設計模型的時候，原本feature extractor 輸出維度是1600維度，這是因為 $(out\_channel * train\_way * train\_way) = 1600$ 。我將out\_channel輸出改為32，所以輸出維度變為原本的一半，也就是800。另外，我仿造MLP原本的設計，來設計我後面的模型結構，主要是使用兩層全連接層，並搭配ReLU作為 activation function。第一層輸出維度是512，第二層輸出維度是128。另外，使用Euclidean distance來做距離衡量機制。

另外，訓練上面，我train way 以及 test way 都設 5，並沒有特別使用甚麼技巧。而選擇 adam 作為optimizer，learning rate則是設計在0.001，同時也有設計使用 scheduler來調整learning rate大小。另外，我沒有使用data argumentation的技巧，最後準確度大概是46.80。

For Problem 1-2 and 1-3, you will do some experiments about different distance function and different K shot settings.

2. (20%) 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.

- Parametric: use a learnable model to measure distance between two features

	Euclidean distance	cosine similarity	parametric function
val acc	46.80 +- 0.87 %	35.99 +- 0.84 %	40.76 +- 0.91%

這裡我都使用5 way 1 shot設定。並比較不同distance衡量方式在訓練上的優劣。

其中，準確度最好的是使用Euclidean distance所做的結果，次好的是使用cosine similarity，而最差的是使用parametric function。

另外 parametric function, 的部分，我設計上面很簡單。為了不將資訊丟失，所以我將兩個向量concat，而不是使用相減。另外，結構主要使用兩層的神經網路層，第一層將1200縮減至512，第二層將512縮減至5。但是我沒有使用activation function。

在這項 parametric function 實驗中，我的準確度達到大約 40.76，我覺得是比預期的低上許多的，因為我預期的是 parametric function 比其他兩個高，但是這裡卻只排第二高。我想可能的原因是我的神經網路架構設計的不是太好，維度對應上面有些混亂，所以造成特徵擷取的完整性被破壞。

另一方面來說，cosine similarity 我在訓練時的準確度有高達 52 左右，但是測試時只有 35.99，我覺得非常值得思考原因。這差距實在是太大了。我覺得可能的原因是 cosine similarity 在對應變化大的資料集上面比較沒有這麼大的彈性，以至於測試時準確度很低。

最後 Euclidean distance 準確度就中規中矩，也很穩定的表現，跟訓練時期差不多，所以我最後選擇 Euclidean distance 作為主要使用。

3.(10%) 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)

	5 way 1 shot	5 way 5 shot	5 way 10 shot
val acc(20epoch 取最佳)	0.43 +- 0.77 %	0.62 +- 0.84 %	0.66 +- 0.72 %

這裡我採用 problem1-1 的模型結構，並都選用 Euclidean distance，並使用不同的 shot 數來做實驗，同時都訓練 20 個 epoch，以解省訓練時間。準確度來說 5 way 10 shot 的準確度比 5 way 5 shot 的準確度高，而 5 way 1 shot 準確度則是最低。我認為這蠻直覺的，因為 shot 越多表示訓練資料數量越多，也就表示越容易學習。因此 5 way 10 shot 訓練起始準確度，也就是第一個 epoch 就高達 0.46，而 5 way 5 shot 訓練起始準確度則是只有 0.44，最後 5 way 1 shot 訓練起始準確度則最低，只有 0.29。

而在二十個 epoch 內，準確度最高的是 0.68，5 way 10 shot，而準確度最低的是 5 way 1 shot，只有 0.43。

#### Problem 2: Self-Supervised Pre-training for Image Classification (50%)

1. (10%) 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)

我使用助教所提供的 BYOL 連結來訓練 resnet 50 作為 backbone。learning rate 大約設置在 3e-4。並使用 adam 作為 optimizer。但是我沒有設計 scheduler 來修正我們 learning rate，所以最後可能是因為 learning rate 有些太大，所以 loss 收斂大致在 0.29 左右。

另外，我也遵循助教所說的，將圖片 resize 至 128 的大小再輸入，並使用 normalize 來控制圖片的數值。這些設計上面我相信對 leanring 上面一定有幫助。

此外，我沒有使用 data argumentation 的相關技術，只用原本的圖片跑。

當然，由於我是使用 rtx3080ti 跑，batch 設計在 64 左右，大概花費了 10g 的 gpu 記憶體，所以大概 7 小時左右，就卡住不收斂了，

最後，做個總結。loss 大致在 0.29。

2. (10%) 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.

- Please report the classification accuracy on validation set.
- TAs will run your code to verify the performance of the setting C in the Table below.
- The architecture of the classifier need to be consistent across all settings.

setting	pre_training(miniNet)	Fine-tuning (Office-Home dataset)	Classification accuracy on valid set (Office-Home dataset)
A	-	Train full model (backbone + classifier)	0.2477
B	w/ label (TAs have provided this backbone)	Train full model (backbone + classifier)	0.308
C	w/o label (Your SSL pre-trained backbone)	Train full model (backbone + classifier)	<b>0.495</b>
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	0.241
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	0.392

3. (10%) Discuss or analyze the results in Problem 2-2

整體準確度來說，最高的是C再來是E, B, D，而最差的是A。

設定A來說準確度最高大約在0.22左右，而B的準確度大約在0.3左右，而C的準確度最高，有到0.49左右，而D的話大約0.24，最後E的話大約0.39。

其實這個結果還蠻符合我的預期的，有使用SSL的方式做pretrain的效果最好，而助教給的pretrain weight則是大概次好，最差的是沒有使用pretrain wieght的A。另一方面來說，允取整個模型訓練的C 與限制訓練的E相較來說，準確度更高，另外，B的準確度也比D高。所以我覺得整體來說，還蠻符合預期的。另外，我在訓練D、E的時候觀察到，訓練準確度其實就不是太好，大概只有0.4、0.5左右而已。這也就表示訓練上面，模型其實受限於backbone的影響，所以準確度被限制在一定範圍，不能學習更好。想當然，val acc的準確度又會更低。