EAI Lab2 Report

Task 1 – create your own CNN model

1. Print model summary (including parameters) and plot model (5%)

第一個任務中，我使用的模型包含三個convolutional layers，每個layer包含一個Conv2d、BatchNorm2d、ReLU和MaxPool2d，每經過一個convolutional layer，activation的spatial size (both height and width) 會減半、channel數則加倍，最後則是Linear layer輸出10個類別的結果，並加上Dropout做regularization以避免overfitting。

A screenshot of a computer

Description automatically generated

1. Print test accuracy, plot epoch-train accuracy, epoch-val accuracy, epoch-train loss, epoch-val loss (10%)

Training configuration:

* Data normalization with mean and standard deviation calculated from the first image in the training set
* Data augmentation: RandomCorp, RandomHorizontalFlip, and RandomRotation
* Batch size: 128
* Optimizer: Adam
* Base learning rate: 0.001
* Learning rate update strategy: halve per 10 epochs (StepLR)
* Number of epochs: 80

依照上面的配置訓練模型後做testing，準確度達到81.7%，觀察訓練過程中的loss和accuracy顯示，整體訓練過程大致穩定沒有劇烈震盪，且沒有出現training accuracy持續下降但validation accuracy反而上升的狀況，因此沒有overfitting，最後loss和accuracy曲線也趨於平穩收斂。



A graph of loss and an acuity

Description automatically generated with medium confidence

Figure SimpleCNN

1. Describe how do you choose your best model (5%)

在訓練過程中，選擇當前validation accuracy最高的checkpoint存下來。

Task 2 – ResNet 18 implementation

1. Print model summary (including parameters) and plot model (5%)

ResNet18 模型實作如下：

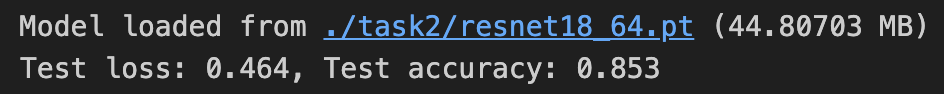
|  |  |
| --- | --- |
|  |  |

1. Print test accuracy, plot epoch-train accuracy, epoch-val accuracy, epoch-train loss, epoch-val loss (10%)

Baseline configuration:

* Data normalization with mean and standard deviation calculated from the first image in the training set
* Data augmentation: RandomCorp, RandomHorizontalFlip, and RandomRotation
* Batch size: 64
* Optimizer: Adam
* Base learning rate: 0.001
* Learning rate update strategy: halve per 10 epochs (StepLR)

依照上述配置訓練模型，準確度達到85.3%，觀察訓練過程中的loss和accuracy顯示，整體訓練過程大致穩定沒有劇烈震盪，雖然沒有出現training accuracy持續下降但validation accuracy反而上升的狀況，不過在訓練後期隨著training loss持續下降，validation卻沒有跟著明顯改善，因此可能有overfitting的狀況。不過因為validation set是從training set隨機選出來的，有做data augmentation，而test set則沒有，因此test accuracy比validation accuracy高屬合理狀況。



A graph of loss and training

Description automatically generated

Figure ResNet-18 baseline

1. Describe how do you choose your best model (5%)

在訓練過程中，選擇當前validation accuracy最高的checkpoint存下來。

1. Experiment on the following and compare the result with baseline

以下將會以第(5)題訓練出來的模型作為baseline，依序探討data normalization、data augmentation、learning rate對模型效能的影響。

1. Input image normalization (15%)

|  |  |  |
| --- | --- | --- |
|  | with data normalization (baseline) | without data normalization |
| Test accuracy | 0.853 (figure 2) | 0.845 (figure 3) |

從loss-accuracy curve中可以發現沒有做data normalization時，在訓練初期validation loss和validation accuracy的震盪較明顯，因此可以推論data normalization可以讓最佳化的過程更加穩定。

A graph of loss and training

Description automatically generated with medium confidence

Figure Without data normalization

1. Data augmentation (15%)

|  |  |  |
| --- | --- | --- |
|  | with data augmentation (baseline) | without data augmentation |
| Test accuracy | 0.853 (figure 2) | 0.783 (figure 4) |

沒有data augmentation時，validation loss從第8個epoch開始就不再隨著training loss下降，發生嚴重的overfitting，即便因為training set變得比較簡單使得training accuracy很快達到90%以上，但validation卻始終達不到80%，因此可以推論data augmentation對於預防模型overfitting以及提升模型的泛化能力有非常顯著的效果。

A graph of loss and training

Description automatically generated

Figure Without data augmentation

1. Different base learning rate and update strategy (15%)

|  |  |  |
| --- | --- | --- |
|  | Step LR (halve every 10 epochs) | Constant LR |
| Base LR = 0.01 | 0.715 (figure 5) | 0.521 (figure 7) |
| Base LR = 0.001 | 0.853 (figure 2, baseline) | 0.797 (figure 8) |
| Base LR = 0.0001 | 0.796 (figure 6) | 0.820 (figure 9) |

下圖5、圖6為使用一樣的learning rate update strategy但改用不同的初始值，可以發現同樣是每10個epoch學習率減半，有較高初始學習率的訓練過程有比較多震盪，且權重更新的步長較大，不容易收斂至最佳的值，因此test accuracy表現得比baseline差，而初始學習率太小時，雖然訓練過程更加平緩，但在更新至最佳權重前就會停下來，會更加依賴權重初始化才能達到更好的準確度。

圖7、圖8和圖9則是固定learning rate，當 learning rate較大時，訓練過程中loss和accuracy曲線震盪較明顯，當 learning rate較小時，訓練過程中loss和accuracy曲線則較為平緩。具體來說，圖7的learning rate 0.01過大，以致於模型無法收斂，自第16個epoch開始validation loss就不再下降，圖8的learning rate 0.001還是過大，自第40個epoch以後validation loss就不再下降，而圖9的learning rate 0.0001則和baseline在訓練後期的learning差不多，但因為前期的步長較小，因此收斂速度不如baseline的配置。

A graph of loss and validation

Description automatically generated with medium confidence

Figure Step LR with inital value 0.01

A graph of loss and validation

Description automatically generated with medium confidence

Figure Step LR with inital value 0.0001

A graph of loss and loss of a person

Description automatically generated with medium confidence

Figure Constant LR 0.01

A graph of loss and validation

Description automatically generated with medium confidence

Figure Constant LR 0.001

A graph of loss and training

Description automatically generated with medium confidence

Figure Constant LR 0.0001

1. What challenges did you encounter, how did you solve them, and what are your thoughts and suggestions regarding this lab? (15%)

Task 1原本想用兩個convolutional layer來訓練，但發現不管怎麼調整learning rate或optimization algorithm，都達不到80%以上的準確度，後來多加了一層convolutional layer後準確度才得到改善。

Task 2 有規定使用 ResNet-18，嘗試了不同的learning rate、optimizer、learning rate scheduler和不同的data augmentation，都難以達到85%以上的準確度，後來仔細再重新看了助教的錄影後才想到可以調整batch size，調整為64後才終於達到85%。

助教給程式碼中train model的部分有縮排上的錯誤，下次可以更正一下。另外，這個lab的目的應該是要教會學生如何調整超參數以訓練出最佳的模型，因此未來的講義或講解影片或許可以說明更多在各種情況下如何判斷應該調整什麼超參數，例如看到validation loss震盪太大可以把learning rate調小，或是當validation loss經過一定epochs後都沒有繼續改善可以做early stop等等。