Machine Learning HW8 Report

學號：R07943107 系級：電子所碩一 姓名：徐晨皓

1. **請比較你本次作業的架構，參數量、結果和原HW3作業架構、參數量、結果做比較。(1%)**

* 本次作業(HW8)

1. 模型架構

本次作業使用PyTorch實作。Model分為兩部分，CNN與DNN。在CNN部分，4層pointwise convolution layers，6層depthwise convolution layers，以及3層max pooling layers交錯組成。DNN部分使用兩層linear layers。其中activation function使用relu，層與層之間也有使用batch normaliztion。此模型有使用dropout來避免overfitting。

1. 參數量：102,700
2. 結果

Private score: 0.62357

Public score: 0.62412

* 原HW3

1. 模型架構

本次作業使用keras實作。Model分為兩部分，CNN與DNN。首先，由五層convolution layers與三層max pooling layers交錯組成CNN的部分，每層convolution layer都有使用zero padding的技術，讓圖片不會損失原有大小。接著再以三層fully connected layers組成DNN部分，最後再接上output layer。除了output layer的activation function為softmax，convolution layer與dense layer都使用relu作為activation function。此模型有使用dropout來避免overfitting。

1. 參數量：7,255,815
2. 結果

Private score: 0.70103

Public score: 0.71217

* 兩者比較

在模型參數量方面，HW3的約有七百萬，然而MobileNet只有約十萬，相差約70倍。在預測準確率方面，HW3有約0.70的準確率，而MobileNet有約0.62的準確率，相差約8的百分點。由此可見，儘管MobileNet只有非常少量的參數，但準確率沒有大幅下降。

1. **請使用MobileNet的架構，畫出參數量-acc的散布圖（橫軸為參數量，縱軸為accuracy，且至少3個點，參數量選擇時儘量不要離的太近，結果選擇只要大致收斂，不用train到最好沒關係。）(1%)**
2. **請使用一般CNN的架構，畫出參數量-acc的散布圖（橫軸為參數量，縱軸為accuracy，且至少3個點，參數量選擇時儘量不要離的太近，結果選擇只要大致收斂，不用train到最好沒關係。）(1%)**
3. **請你比較題2和題3的結果，並請針對當參數量相當少的時候，如果兩者參數量相當，兩者的差異，以及你認為為什麼會造成這個原因。(2%)**

* 實驗設定

1. 使用的參數量約為10萬、20萬、30萬、50萬。
2. 兩模型都只訓練約30至40個epochs。
3. 兩模型訓練皆有使用data augmentation。

* 兩者差異與原因探討

從下圖可以看出隨著參數的增加，一般CNN的準確率有大幅的提升，然而MobileNet的準確率卻只有緩慢的上升。我認為這是因為當一般CNN有較多的參數時，其convolution的範圍或方式更多元，因此能更偵測到更多的圖片特徵。反觀MobileNet，其使用depthwise convolution因此我認為這樣的filter多樣性不及原來的多樣性，導致MobileNet隨著參數增加，準確率卻沒有大幅增加。

另外，我們可以看到，當參數量約為10萬時，一般CNN準確率只有約0.47，然而MobilneNet確有0.62，這表示MobileNet這樣的結構在參數量少時，也能有不錯的準確率，不像一般CNN參數量減少時，準確率會大受影響。