

# Homework3-bonus Report

姓名：劉兆鵬 學號：r07921052 系級：電機碩一

## 1. Cat Resnet50 model

```
cat resnet50_best.* > resnet50_best
```

## 2. (1%) 請說明參數設定方式(a, b 題目以 train.py 的 function 分項說明)？

在這次的作業中，要手做的調整 model 的 hyper-parameter。起初我使用 HTC 所給予的參數直接進行訓練後，所得的結果與 simple baseline 相差甚遠，故我就進行參數逐一設置。

### A) Optimizer 設置：

#### Mobilenetv2:

在實驗中，我首先使用初始的 SGD 來做訓練，發現訓練結果最高只能到 0.6 左右而無法再上去，我想可能是 mobilenet 需要花多一點時間訓練，因為可以看出 training accuracy 是持續在上升的，但因為有限定最多的 Epoch 數量，故無法繼續上升。為了能夠加速訓練的速度，我先設定了 momentum 在 SGD 中，但我發現結果反而是變差的。最後我採用 Adam，Learning rate 設置 0.001，就能夠得到較快速的訓練速度，使得模型能夠提早收斂，而因為有使用 pretrained model，故我在此假設 pretrained 的模型是良好的，故我設置的 learning rate 比較小，避免搞壞了原先訓練好的 pretrained model，又因為使用 Adam 會訓練較快速，容易快速的學習訓練資料的資訊，造成 overfitting 故我加入了 weight decay，從小數值( $10^{-5}$ )逐漸增加直到能夠達到最好的 validation accuracy 則停止。

#### Resnet50:

在 resnet 的實驗中，我也想要效仿 mobilenetv2 的方法，但發現使用 Adam 在 resnet50 上面，反而得到較差的效果，使得模型只能訓練到 0.7 左右。而我將 optimizer 改會 SGD 後，就能夠使模型達到 0.86 的準確度，代表 resnet50 本身就有極高的準確性，使用 SGD 對訓練資料逐步做細微的調整即可達到良好的效果。

### B) before\_epoch 設置:

#### Mobilenetv2 與 Resnet50:

在每一個 epoch 時，都會設定訓練資料的前處理方式以及每次要讀入多少 batch 的資料，在原先什麼都不要設定的情況下，雖然能夠讓 training accuracy 達到很好的效果，但是都會使得模型 overfitting，因為在訓練時模型會過度傾向訓練資料的資料分佈，而使得訓練資料的 loss 變得過低，因為模型調整參數是根據其 loss 的 gradient，所以當 loss 太低時，就會使得模型的參數不再繼續調整，而使得 validation accuracy 不再進步反而逐漸下降。故我使

用了 pytorch 中的 transform 來進行 data augmentation，其中我使用了 random horizontal flip、random crop、random affine、random gray scale，這些資料處理都是以能夠維持圖片本身的特性進行的運算，且由於是 random 的選取前處理的方法，故不會導致模型過度傾向於訓練資料，進而能夠持續提升 validation accuracy。

#### C) before\_epoch 設置：

Mobilenetv2 與 Resnet50:

在每一次的 batch 前的設置，在實驗中，一開始我是使用固定的 learning rate。但是會發現每次訓練到一半，當 validation accuracy 到達最高點後，就會開始下降，我認為是模型已經快要找到 loss 的低點時，因為 learning rate 太大了，所以參數的調整又使得 validation accuracy 降低了，我想這也是 overfitting 的因素之一。所以我就找尋了 learning rate 下降的機制，發現說能夠依據 validation accuracy 的變化來調整 learning rate，但我在每一次訓練的過程中，我發現 validation accuracy 最高的時間點都差不多位於 15 至 20 的 epoch，故我的調整方法是根據目前為第幾個 epoch 來決定我的 learning rate。而其中，我於最新的 ML 文章發現，batch size 也是影響訓練結果的重大關鍵，若模型還沒有訓練起來了應要使用較小的 batch size 讓模型調整參數的次數多一點，而若模型已經 fit 於訓練資料時，應使用較大一點的 batch size 來一次觀察多一點資料。且使用隨著 epoch 增加 batch size 也是一種方式，也就是在訓練初期使用小一點的 batch size 讓模型調整多次一點，並隨著 epoch 的增加，逐漸增加 batch size 數量，文獻上說此效果與逐漸降低 learning rate 是有一樣的效果的。除此之外，我也採用隨著 epoch 增加逐漸降低 learning rate 的做法，將這兩個解果合起來一同使用則能夠達到不錯的效果。另外，因為當訓練到後來，模型依然會 overfitting 於訓練資料，所以我也隨著 epoch 逐漸加強 weight decay 的參數避免模型過度擬合於訓練資料。

#### D) save\_model\_as 設置：

在這個 function 中我則是將每一次訓練 epoch 中，只要是達到最高的 validation accuracy 的模型參數都存起來，因為每次訓練到後來都會造成模型 overfitting 的問題，使得第 50 個 epoch 並非最好的結果，故使用此方法將每一次最高的 validation accuracy 的模型都存起來是最好的作法。

