

# Homework 3 Bonus Report

資工四 B04902131 黃郁凱

November 25, 2018

## 1 實驗比較及最後選擇此值的原因

### 1. learning rate

調整 learning rate 在 validation 的時候就可以看出大約在 0.001 會是穩定遞增的，然而這樣正確率只能達到 80% 出頭，原因是只能 train 五十個 epoch，使得學習率慢的模型尚未收斂到局部最低點訓練就已經結束了，為了解決此問題，我將 mobilenetv2 的學習率調大十倍，雖然較不穩定，但可以在短時間內達到更好的 performance；而 resnet50 的學習率只調大五倍，原因是十倍實在太大了，validation 的正確率有過大的震盪，使得模型時常跳開原本已經收斂到的不錯的局部最小值。

### 2. pretrain

pretrain 對於很多模型都有益處，原因是先讓模型看過類似類型的訓練資料，有機會收斂到更好的局部最佳值；就好比下過象棋再去學西洋棋，兩者雖無直接關連，但卻能很快就上手。本次作業中，在沒有加上 pretrain 的參數情況下，只有幾千張的訓練資料是 train 不動龐大的 resnet50 模型參數，在訓練階段就可以看出嚴重的 overfitting 現象。

### 3. batch size

不像 learning rate 或 pretrain 有這麼大的影響，batch size 大概設在 32 左右 performance 都不會有太大差異，Yunn LeCun 曾說過 batchsize 不要大過 32，也有一番的解釋<https://arxiv.org/abs/1804.07612>。

### 4. dropout

Dropout 可以減緩 overfitting 的發生，因此適當的 dropout rate 可以有效抵抗這次訓練資料少的情況，大約在 0.5 可以得到不錯的 performance。

### 5. batch norm momemtum

由於每次批次的訓練資料不同，可能會有不穩定的 normalization 結果，加上 momemtum 有助於減緩這事情的發生，上網查過別人的 momemtum rate 大約設在 0.9 左右，因此我採用 0.9 的值來訓練，得到不錯的成效。

### 6. Data Preprocessing (image transform)

這部分的處理可以說是訓練的畫龍點睛，當模型已經達到一定程度的水準，加上 augmentation 可以有效增加訓練的 performance。我有做了水平翻轉、色彩飽和度，以及隨機裁切，做完這些以後，testing 正確率可以從 86% 進步到 87% 多。