

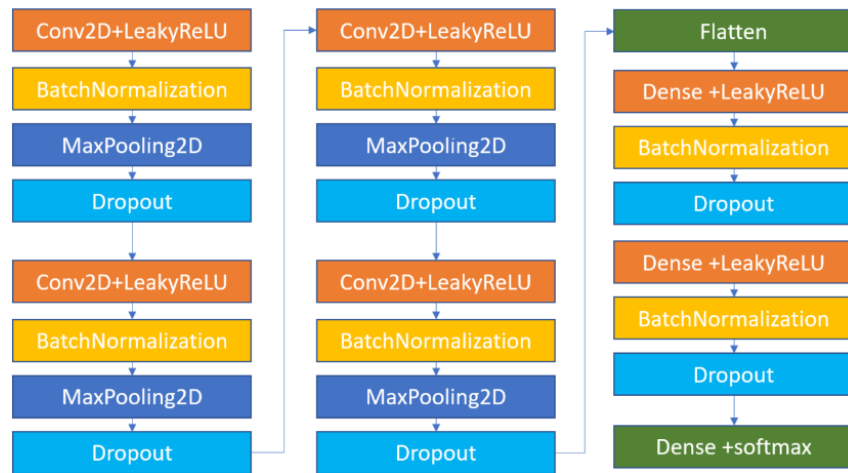
Machine Learning HW8 Report

學號：R07943150 系級：電子碩一 姓名：吳辰鉉

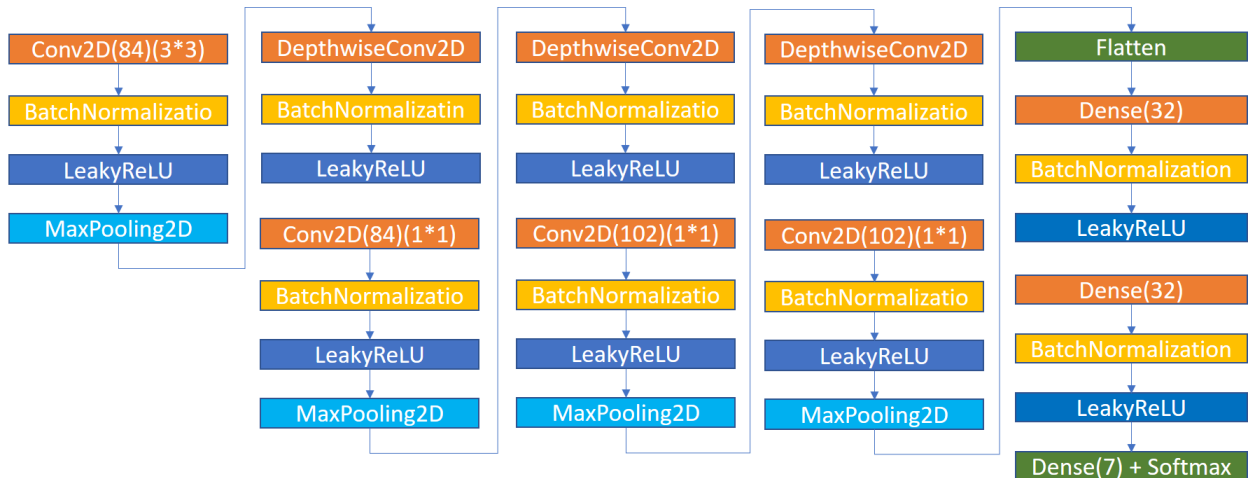
1. 請比較你本次作業的架構，參數量、結果和原 HW3 作業架構、參數量、結果做比較。(1%)

	Total Parameters	Public Score	Private Score
HW3	15033863	0.71719	0.71217
HW8	62955	0.64168	0.64669

CNN 架構:



MobileNet 架構:



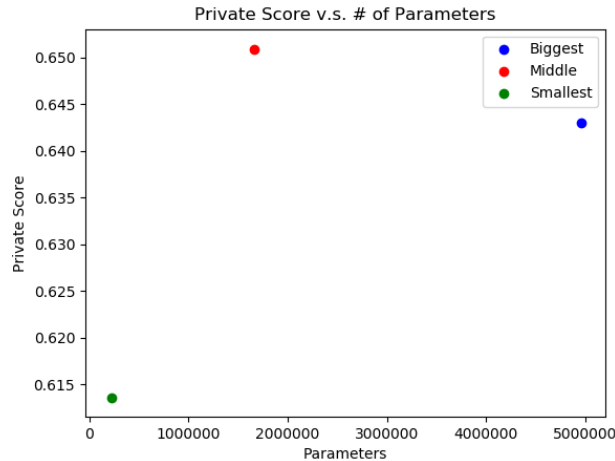
主要差異為

1. 卷積層除了第一層之外皆為 mobile net 的形式(DepthwiseConv2D+Conv2D(1*1))
2. 拿掉 Dropout 層 (因參數已很少，若再 Dropout 會顯著影像準確率)
3. 減小各層參數量

4.改為先經過 BatchNormalization 再過 Activate function，這個順序會比較合理並使得 BatchNormalization 有意義，這部分應該是 HW3 可以改進的地方。

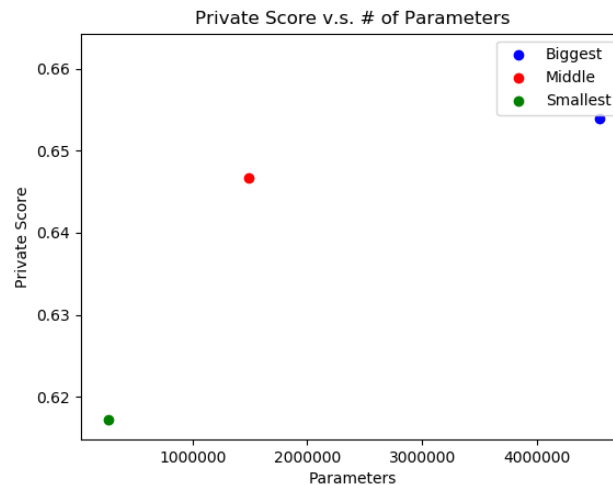
5.MobileNet 採用 knowledge distillation 方式，利用 HW3 預測 dataset 的機率分布來進行訓練，hw3 在 dataset 的準確率約為 96%。

- 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%)



Total Parameters	Validation Score	Public Score	Private Score
4962055	0.66984	0.64279	0.64307
1664007	0.67030	0.64725	0.65087
221319	0.63594	0.61382	0.61354

- 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%)



Total Parameters	Validation Score	Public Score	Private Score
4548607	0.68888	0.66285	0.65394
1489207	0.67402	0.65756	0.64669
264943	0.64082	0.61799	0.61716

4. 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%)

1. 因訓練時間因素兩者都沒有做 data augmentation，所以訓練時準確率較飄，也還有改進空間。

2. 在 data 是隨機切割的情況下，兩張圖的 Validation Score 都明顯高一截，其他同學也有類似的結果，推測是 data 內有重複使得準確率會比 test data 高。

3. 小參數的情況，以下面數據做比較

	Total Parameters	Public Score	Private Score
MobileNet	62955	0.64168	0.64669
CNN	61047	0.59570	0.58790

其中 MobileNet 雖然有被特別調整過，是本次作業最終模型，而 CNN 僅是湊成近似參數的模型，但準確率差異非常大。因模型架構差異，MobileNet 的每個卷積層的 filter 參數都較 CNN 少，故在相同參數量時可以有更多的 filter。而當參數量小時，可推測愈多 filter 能從圖片抽取特徵會比在每個 filter 做計算來的重要，使得 MobileNet 效果較佳。

4. 題 2 題 3 內的數據，因參數量漸增，使得上述兩個方向的差異影響下降，可以看到傳統 CNN 的分數普遍些微高於 MobileNet，推測是當 filter 增加到一定數目後，應被抽取的特徵都已被抽取，剩下的進步空間需要在每個 filter 內做計算，使得 CNN 略為優勢。

5. 但是 MobileNet 主要是勝在訓練速度，根據 reference，參數量相同的情況之下，MobileNet 的訓練速度可以快一個數量級。

Reference:

<https://www.itread01.com/content/1534164982.html>