

Machine Learning HW8 Report

學號：B06902049 系級：資工二 姓名：林首志

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%)

這次作業kaggle best solution的架構如下：

Conv2D(30, 3) -> LeakyReLU()

-> DepthwiseBlock(30, 1) -> DepthwiseBlock(30, 1) -> DepthwiseBlock(30, 1)

-> DepthwiseBlock(45, 2) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1)

-> DepthwiseBlock(45, 2) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1)

-> DepthwiseBlock(45, 2) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1) -> DepthwiseBlock(45, 1)

-> GlobalAveragePooling2D -> Dense(7) -> Softmax

其中DepthwiseBlock的定義如下：

DepthwiseBlock(numFilters, strides):

BatchNormalization -> DepthwiseConv2D(3, strides) -> LeakyReLU()

-> BatchNormalization -> Conv2D(numFilters, 1) -> LeakyReLU()

所有Convolution層的padding設置為same。

所有層的參數使用he_uniform初始化，並加上4e-5的L2 regularization。

model使用16bit浮點數儲存，訓練和預測時皆使用32bit浮點數。

原HW3 report的架構如下：

Conv(30, 3) -> ReLU -> BN -> Conv(30, 3) -> ReLU -> BN -> Conv(30, 3) -> ReLU

-> MaxPool(2) -> BN -> Conv(60, 3) -> ReLU -> BN -> Conv(60, 3) -> ReLU -> BN -> Conv(60, 3) -> ReLU

-> MaxPool(2) -> BN -> Conv(90, 3) -> ReLU -> BN -> Conv(90, 3) -> ReLU -> BN -> Conv(90, 3) -> ReLU

-> MaxPool(2) -> BN -> Conv(120, 3) -> ReLU -> BN -> Conv(120, 1) -> ReLU -> BN -> Conv(7, 1)

-> GlobalAveragePool -> Softmax

其中BN為BatchNormalization，Conv, MaxPool, GlobalAveragePool都是2D的版本。

所有Convolution層的padding設置為same。

所有層的參數使用he_uniform初始化，並加上5e-4的L2 regularization。

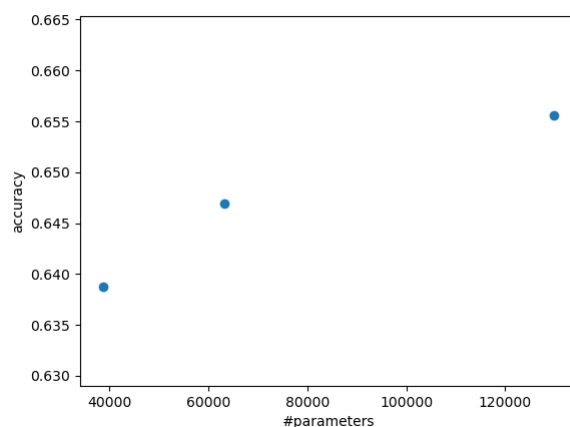
下表是參數量和結果的比較：

	本次作業	HW3
參數量(Keras summary的Total params)	38,647	408,217
準確度(public和private的平均)	0.64321	0.68041

2. 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%)

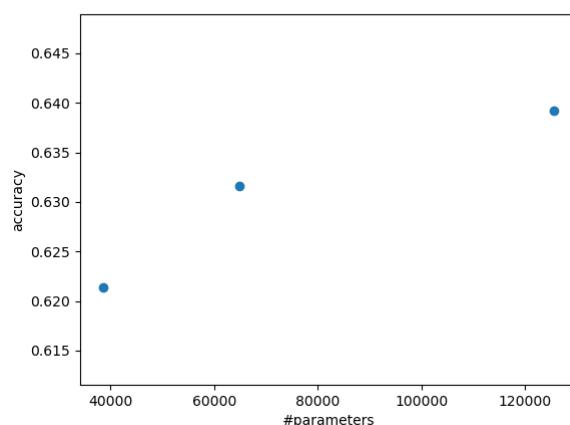
以下是拿kaggle best solution的架構，僅調整每層的filter數量而得出的結果。

(第2, 3題的參數量定義都是Keras summary的Total params，準確度都是public和private的平均)



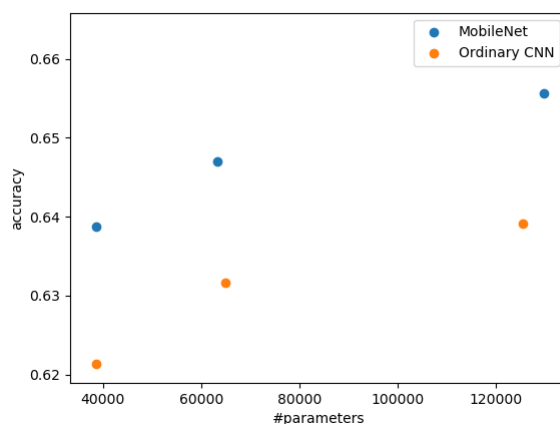
3. 請使用一般CNN的架構，畫出參數量-acc的散布圖 (橫軸為參數量，縱軸為accuracy，且至少3個點，參數量選擇時儘量不要離的太近，結果選擇只要大致收斂，不用train到最好沒關係。) (1%)

以下是將kaggle best solution的Depthwise separable convolution改為一般convolution，減少層數後，變動每層的filter數量而得出的結果。



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%)

以下是將前兩題的結果畫在一起的圖。(有刻意讓一般CNN和MobileNet參數量接近，方便比較。)



可以發現在參數量相當少時，若兩者參數量相當，MobileNet架構的表現會比較好。可能的原因是因為 MobileNet 使用的 Depthwise separable convolution 相對於一般的 Convolution 少了很多的參數量，但是卻能得到差不多的效果（ MobileNet 原論文有做過類似的比較 ）。因此在參數量相同的情況下，MobileNet 可以疊的更深，每層可以加更多的 filters，讓神經網路可以擬合更複雜的函數。