

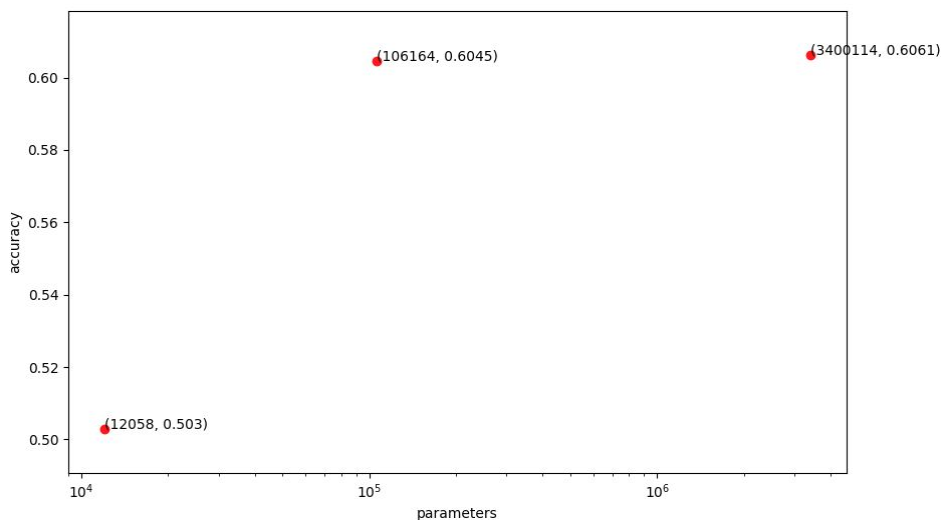
學號：B04507025 系級：電機四 姓名：韓秉勳

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%)

	hw8	hw3
架構	以mobilenet 的 depthwise 與 pointwise convolution 去疊這次的 model, channel 從8->16->32->64, 並在做完兩組depth跟pointwise conv時做max_pooling, 最後拉平後過linear layer (2304->29->7), data有transform旋轉, crop, flip 以及 normalize, weight 的bit也只取一半做quantization	用VGG16疊成, 最後用三層fully-connected layer連接。data有transform旋轉, flip 以及 normalize, 並用5個model ensemble。
參數量	106164	14801110
結果(kaggle accuracy)	0.62607 (epoch=261)	0.69406(epoch=70)

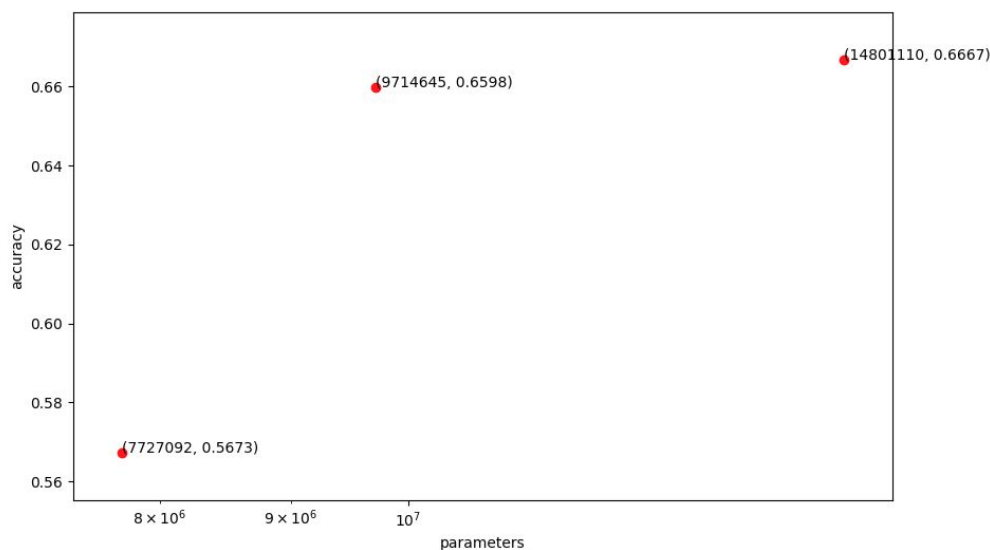
2. 請使用MobileNet的架構，畫出參數量-acc的散布圖（橫軸為參數量，縱軸為accuracy，且至少3個點，參數量選擇時儘量不要離的太近，結果選擇只要大致收斂，不用train到最好沒關係。）(1%)

以下圖x軸都以log scale 表示，以validation作為accuracy，以把參數點拉開。我這邊取的三個點參數量都是不同數量級（每個都差至少10 倍）可以看到參數最少的準確率最差，大約5成，但其他兩個準確率差不多。



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%)

以下圖x軸都以log scale 表示，以validation作為accuracy，可以看到參數最少的點比例上並沒有少太多參數，但準確率差非常多。



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%)

從實驗結果來看，CNN的參數量為7727092時，準確率比參數量9714645少了非常多，但比例上並未減少太多參數（仍為10的6次方），因此若將cnn參數量壓到10的5次方（此為mobilenet目前使用的參數106164的數量級，accuracy 0.60左右），相信accuracy一定無法達到mobilenet的0.60。

從理論角度來看，由於cnn是對所有channel做convolution，若要降低參數需要減少channel數量，可能是stride拉大，或maxpooling 要多做幾次，但這麼一來自然會丟掉許多特徵值，最後預測結果自然不會太好。mobilenet是直接做 depthwise seperable convolution，以3\*3 kernel來看參數量就已經比標準convolution少9倍的參數，但整體的feature與標準convolution取出來的不會相差太多。故若比較在少參數量時哪種方法留下的特徵比較多，mobilenet應該能保留夠多特徵，故準確率也較有可能比較高。