

學號：b05901063 系級：電機三 姓名：黃世丞

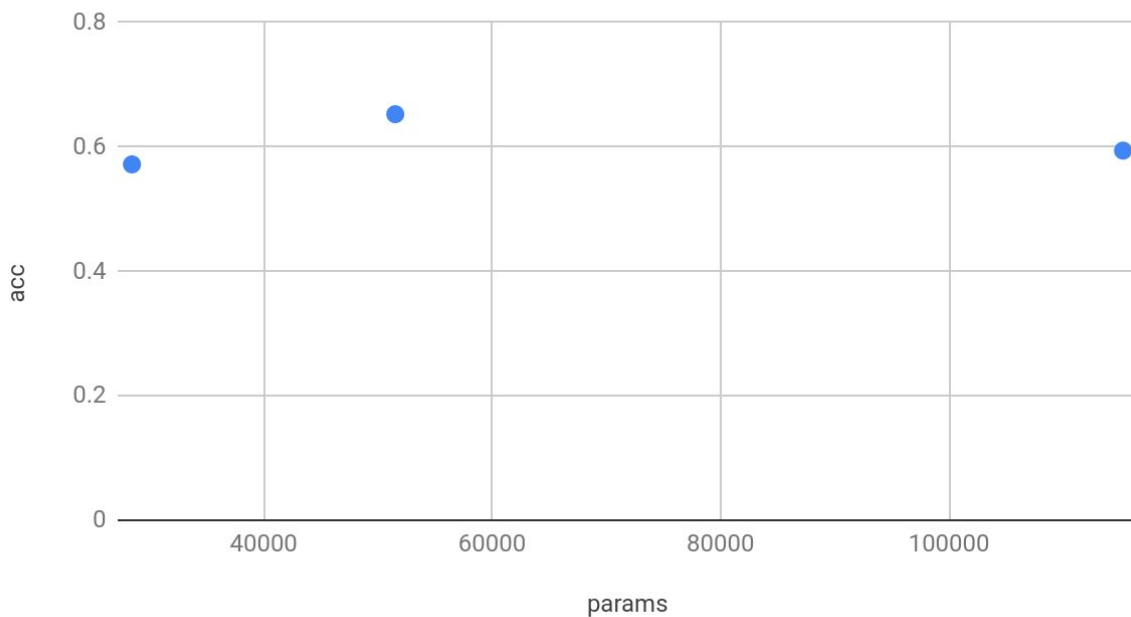
1. 請比較你本次作業的架構，參數量、結果和原HW3作業架構、參數量、結果做比較。(1%)

	HW3	HW8
參數量	6,137,287	51,527
架構	仿VGG的CNN架構	仿Mobilenet的架構
分數(private/public)	0.65867/0.66592	0.65449/0.65254

hw3的參數量是hw8的超過一百萬倍，但是表現上只有提昇大約1%的準確率

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%)

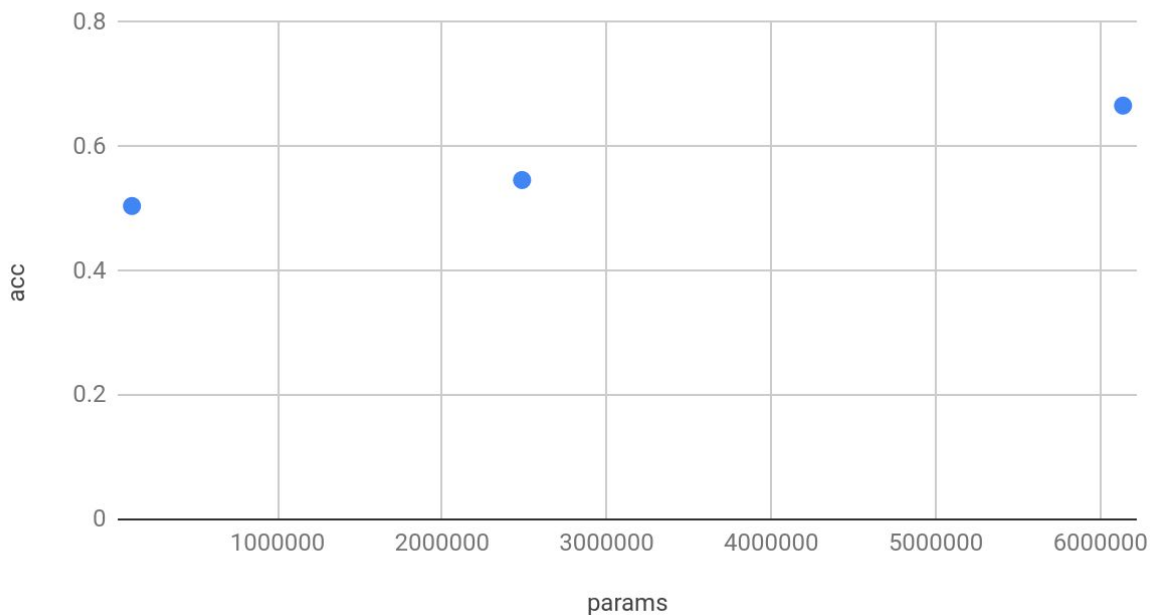
縱軸：acc，橫軸：params



補充說明：參數量11萬的那個model並沒有收斂(大約20幾epoch之後一直卡住我就停掉了，因為要拼strong這個model顯然太大)，而參數量5萬的model是有訓練到收斂的(120多epochs)，而且參數量5萬的model事實上20幾epoch時acc只有0.57上下，所以參數量11萬的model訓練久一點應該是能夠更高的

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%)

縱軸：acc，橫軸：params



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%)

在參數量降到很低(10萬左右)時，mobilenet的優勢就顯現出來了，performance完全贏過cnn，不過這樣的結果是顯而易見的，畢竟mobilenet提出的目的就是為了在保持performance的條件下盡可能縮減參數量，事實證明他很成功。

從個人的角度來分析的話，從運算的定義來看，要產生同樣數量的feature map，mobilenet的運算量和參數量肯定是少很多的，因為depthwise convolution的filter是和上一張feature map的filter一對一做運算，而不是對每個filter做運算。而後面接的convolution的kernel大小只有1x1，也大幅減少了參數量和運算量。而表現方面，坦白講現在再分析就有種馬後炮的嫌疑了，畢竟深度學習都是實驗數據說話，在做實驗之前沒都說不準這方法會不會有用。只能說depthwise convolution的filter能夠成功抽取feature map每個channel的特徵，再透過1x1的convolution和其他channel做線性疊加，效果和原本的CNN kernel對每個channel取特徵再加起來類似。