

學號：B05705053 系級：資管三 姓名：蔡涵如

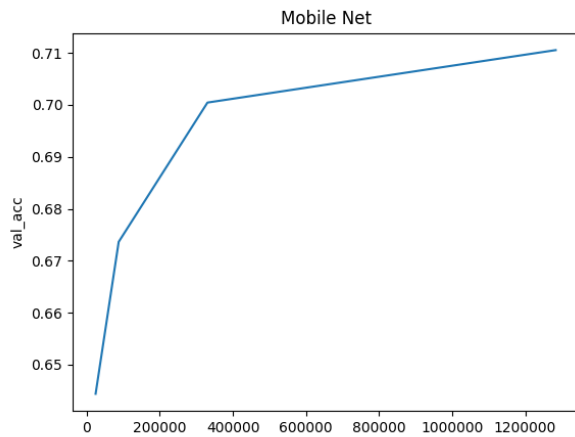
1. 請比較你本次作業的架構，參數量、結果和原HW3作業架構、參數量、結果做比較。(1%)

model架構為mobileNet，加了knowledge distillation的方式用hw3的ensemble model下去做train，效果大概從acc: 0.644 到 0.67左右。

	Hw8	Hw3
架構	conv_dw( 64, 128, 2), conv_dw(128, 128, 1), conv_dw(128, 128, 2), conv_dw(128, 256, 1), nn.AvgPool2d(4),	Conv2D(256) BatchNormalization()  Conv2D(256) BatchNormalization() MaxPooling2D()  Conv2D(512) BatchNormalization() MaxPooling2D()  Conv2D(256) BatchNormalization() MaxPooling2D()  Conv2D(256) BatchNormalization() MaxPooling2D()  Conv2D(256) BatchNormalization() MaxPooling2D() model.add(Conv2D(256))
參數量	24231	3,618,055
Public acc	0.67177	0.69629
Private acc	0.66369	0.69378

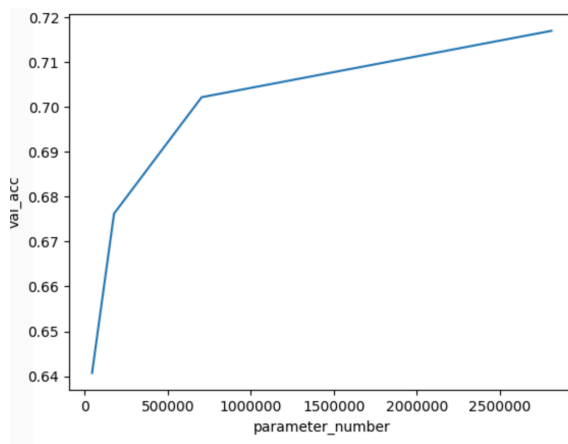
2. 請使用MobileNet的架構，畫出參數量-acc的散布圖（橫軸為參數量，縱軸為accuracy，且至少3個點，參數量選擇時儘量不要離的太近，結果選擇只要大致收斂，不用train到最好沒關係。）(1%)

使用knowledge distillation 下去train，效果比原本用cnn單個model正確率，還要好，而且越多餐數目前看來是越好的情況，總共用了四個點當作實驗的數據，各train了200個epochs，還沒完全到收斂的狀態。



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%)

使用跟mobileNet同樣的架構只是把所有depth wise的block都換成一般的CNN block，以下是實驗結果採用五個實驗數據點,一樣使用knowledge distillation的方式下去做training。



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%)

CNN 參數	CNN acc	Mobile Net 參數	Mobile Net Acc
44759	0.6407175200278650	28497	0.627133
177063	0.6762452107279690	143265	0.673633
704327	0.7021943573667710	635457	0.698711
2809479	0.7169975618251480	2670465	0.7016596

由上面可以看出來其實參數看到在差不多的參數底下還是CNN的model表現會稍微好一些，我想應該是因為CNN的model會使用同一個filter會同時兼顧到比較多的channel，而mobilenet每一個filter只會針對一個channel去抽取feature，由此可知CNN的generalize的能力還是比mobilenet來的好，由此實驗結果，可以推斷，作業如果用cnn來做，在相同參數量的情況下cnn的分數應該會比較好。