

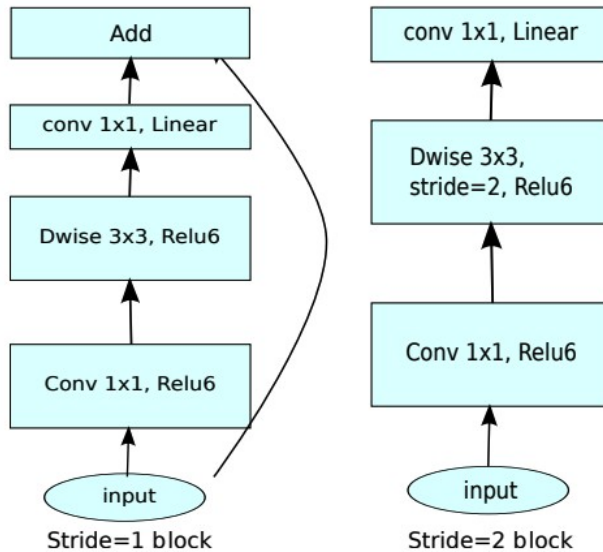
## Machine Learning HW8 Report

學號: B07901069 系級: 電機一 姓名: 劉奇聖

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%)

答: 我在此次作業採用的架構類似 MobileNet v2，其由數個 inverted res block 所組成。

此種 block 有兩類，stride = 1 和 stride = 2，如下圖(Reference: <https://arxiv.org/pdf/1801.04381.pdf>)



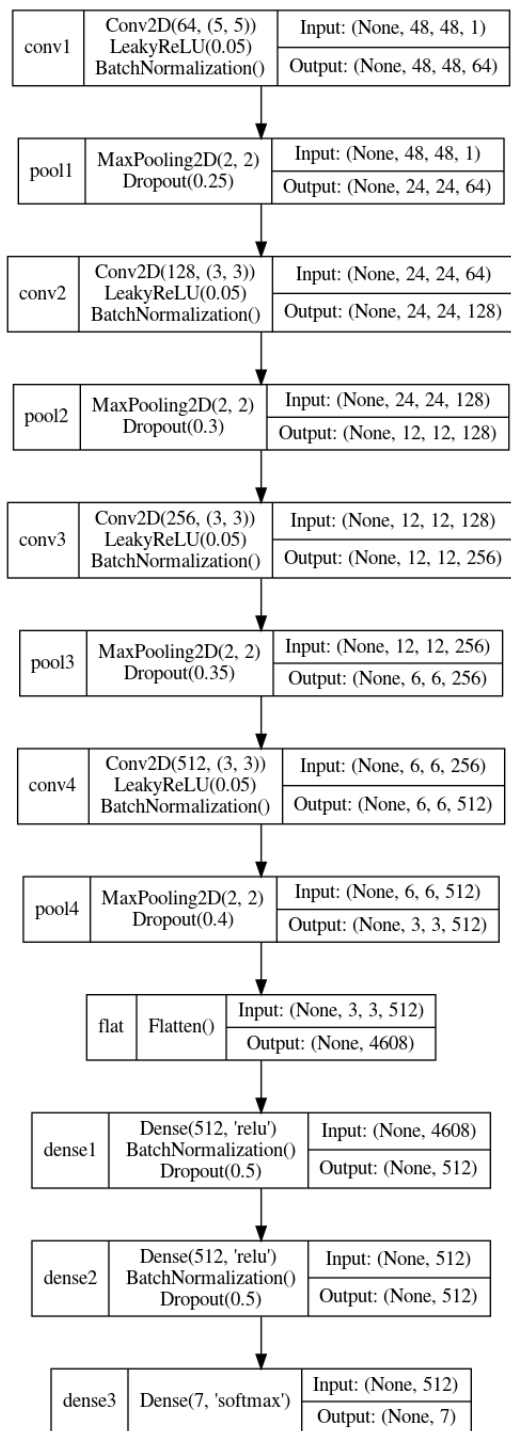
其中第一層的 Conv 1x1 用來做 expansion，此層的 output channels 會是 input channels 的倍數(原架構為 6 倍，我的 model 設為 4 倍)，再通過 Depthwise Conv Layer，再通過 1x1 Conv Layer，若 stride = 1 且 input channels 和 output channels 相等，則將 input 和 output 相加。我的 hw3 和 hw8 的模型架構會在稍後附上，以下先比較參數量與正確率。

	總參數量	Kaggle public score	Kaggle private score
hw3	4,184,839	0.71412	0.71440
hw8	102,709	0.70548	0.69601

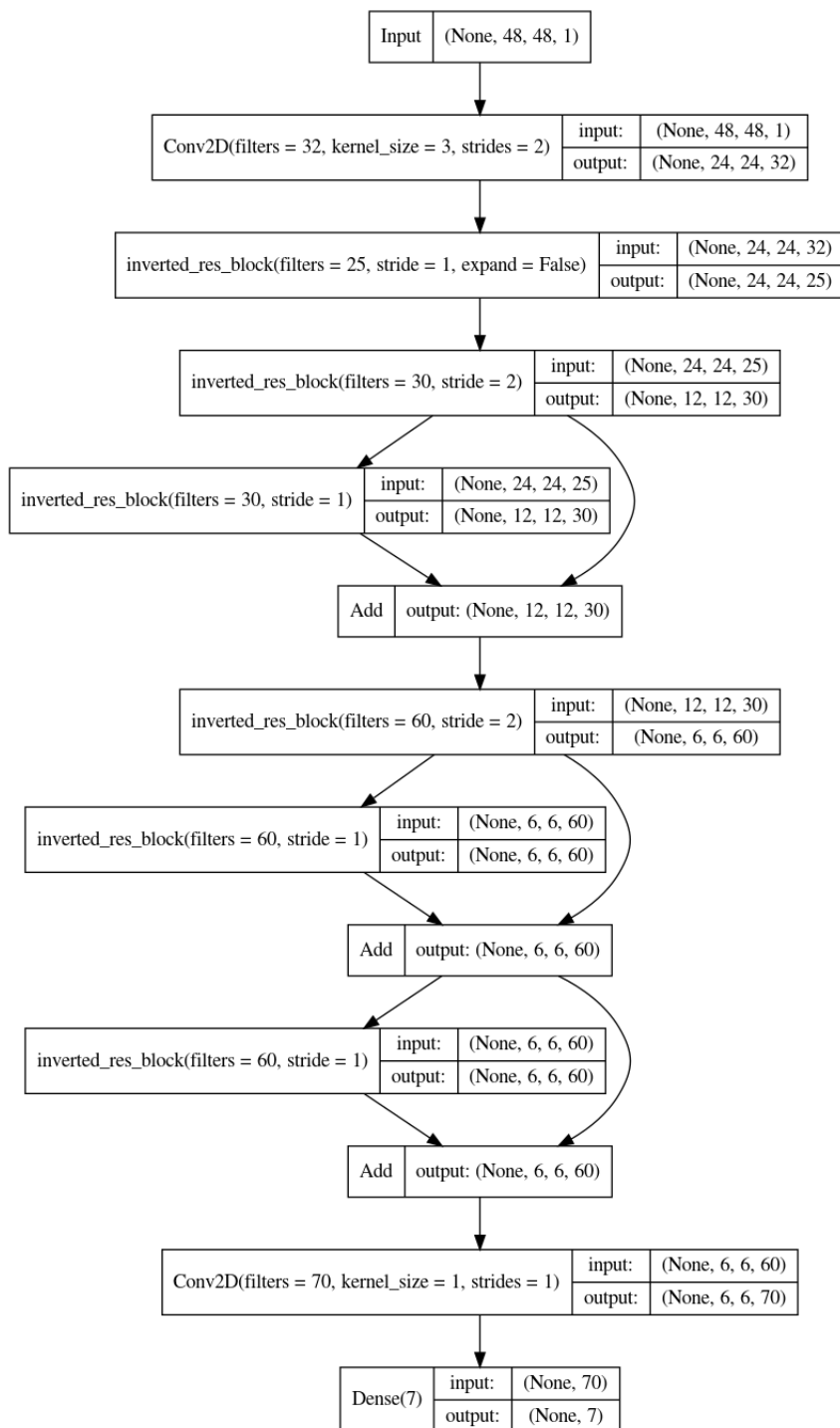
以上呈現的 hw3 在 kaggle 上的分數為 5 個 model ensemble 的結果，總參數量則是挑選這 5 個 model 中的其中一個作為代表。本次作業我除了使用 MobileNet v2 之外，我多訓練了 3 個大的 model，加上 hw3 的 5 個 model，共 8 個 model 當作 teacher models，實作 knowledge distillation，並用將參數轉為 float16 用 npz 存起來。以上的 hw8 在 kaggle 上的分數為應用了這些技術之後的結果。

下面分別是我 hw3 和 hw8 的 model 架構。

## hw3 架構

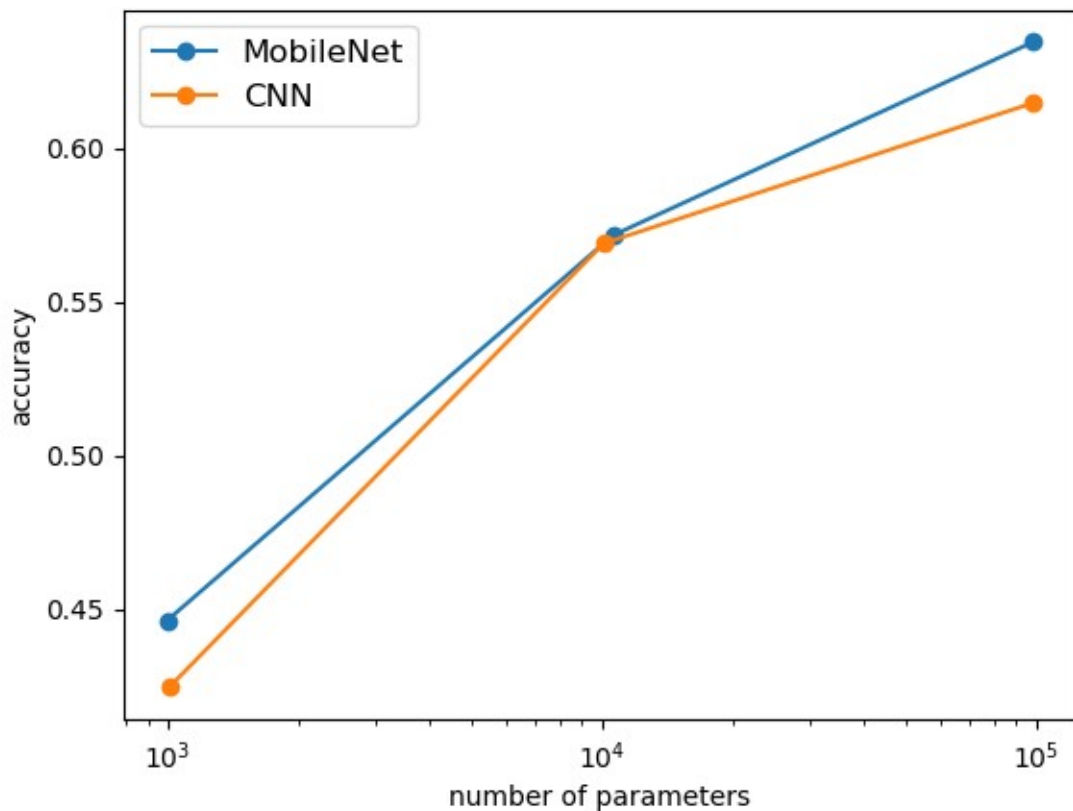


## hw8 架構



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%)

以下的 MobileNet 我使用的是 MobileNet v1，參數皆用 float32 的方式儲存。下面的圖將 MobileNet 和 CNN 畫在一起以便比較，且橫軸取對數刻度，縱軸的正確率為上傳 Kaggle 後將 public 和 private 的分數平均而得。由圖可發現在參數小於十萬且參數量大致相同時 MobileNet 的正確率皆稍高於 CNN，且正確率隨著參數量增加而提高。



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%)

此題的圖已畫在上題，和 MobileNet 的圖疊在一起以方便比較。

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%)

在參數量皆很少且兩者參數量相當時，MobileNet 正確率皆較 CNN 高，例如參數量大約只有 1000 時 MobileNet 的正確率約比 CNN 高了 2%。我推測會造成此差異的原因是在同參數量的情況下，CNN 的 filter 的數目較 MobileNet 少很多，因此 model 抽取 feature 的能力便會下降很多，導致最後的 dense layer 分類的能力下降。