

AOC Paper Reading and Review 1

MobileNets

N26122246 胡家豪

1. Motivation

在影像處理的模型中，多數人都致力於提高模型精確度，而使得模型越來越複雜，導致模型參數與運算增長快速。雖然這樣確實使模型有不錯的 performance，但是卻難以壓低功耗與運算速度。

2. Proposed solution

為了能夠在嵌入式系統或是邊緣系統進行運算，在不犧牲太大準確率的情況下提出了新的架構「MobileNets」。其最核心的方法是將一般常見的 CNN 進行運算分解，拆分成下列兩種運算：

1. Depthwise Convolution

2. Pointwise Convolution

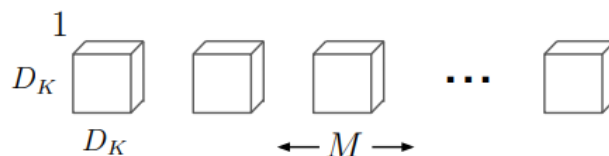
以此種方式進行 Convolution，可以大幅降低運算的參數，以此達到將模型運行在小系統的目的。

3. Evaluation

1. Method Evaluation

1. Depthwise Convolution

在輸入 M channel 中，每一個 channel 都有自己的 filter，且每個 channel 只與自己的 filter 進行運算，如此便會有 M 個 feature maps，這 M 個 maps 即是輸出圖片的 M 個 channel 對應之 maps。

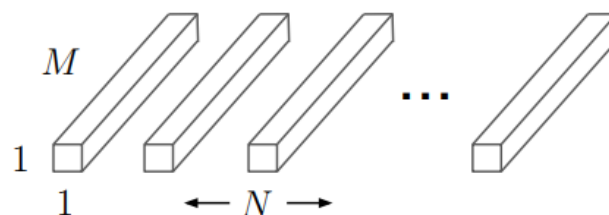


上述 convolution 的運算量為：

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$$

2. Pointwise Convolution

Pointwise Convolution 是一般 convolution 的特例，其代表輸入圖片與為大小只有 1×1 的 kernel 進行 convolution 運算。



上述 convolution 的運算量為：

$$M \cdot N \cdot D_F \cdot D_F$$

綜上所述，所需的總共運算量為

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$$

與原本的 Convolution 進行比較：

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F}$$

可以發現節省的運算量為：

$$\frac{1}{N} + \frac{1}{D_K^2}$$

2. Model Evaluation

使用 ImageNet 的資料集進行訓練與驗證，與一些比較有名的模型，如：「GoogleNet」、「VGG-16」進行比較，可以看到 MobileNet 在沒有犧牲過多的準確率的情況下，降低了至少 3 倍所需要的運算量與 1.6 倍所需要的參數量。

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

ImageNet Accuracy：在 ImageNet 資料庫中進行訓練獲得的準確率，該值越大代表模型預測越準確。

Million Mult-Adds：模型進行的乘加運算，該值越低，代表所需的運算量越少。

Million Parameters：模型參數量，該值越低代表模型所需要的參數越少

4. My analysis

我認為這個方法非常有想法，利用 Depthwise 進行特徵的擷取，再利用 Pointwise 進行特徵的混和，以達到原本的 convolution 的目的，並且可以減少運算量與參數量。

我覺得這篇 paper 未來的研究方向可以朝向硬體邁進，思考如何使用硬體實踐這個方法的加速。

最後，雖然這方法有效，但是作者怎麼知道 pointwise 與 depthwise 可以等於原本方法的 convolution，讓我很好奇。