**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。

1. **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。

1. **Evaluation**
2. 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。

一張含有 圖表, 字型, 設計 的圖片

自動產生的描述

上述convolution的運算量為：



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kernal進行convolution運算。

一張含有 行, 寫生, 圖表, 白色 的圖片

自動產生的描述

上述convolution的運算量為：



綜上所述，所需的總共運算量為



與原本的Convolution進行比較：

一張含有 文字, 字型, 行, 印刷術 的圖片

自動產生的描述

可以發現節省的運算量為：

一張含有 符號, 字型, 數字, 印刷術 的圖片

自動產生的描述

1. 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。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

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

1. **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。