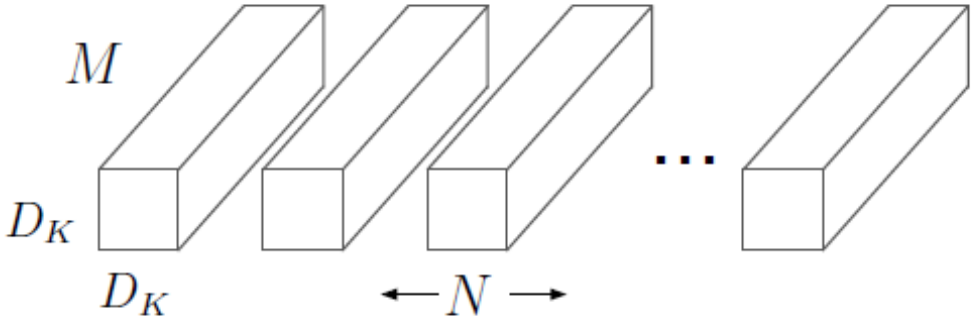
MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

> Howard et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv 2017.

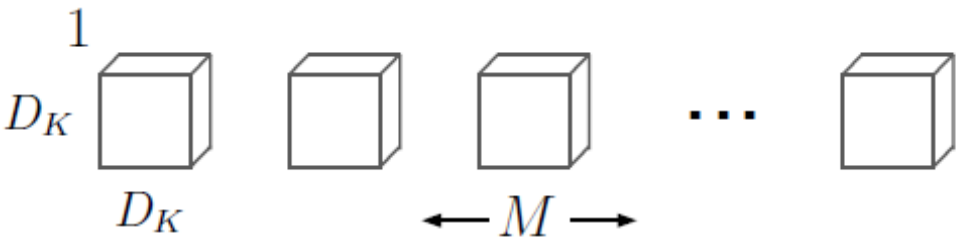
Introduction

This paper proposed a compressing and accelerating method with trading of accuracy.

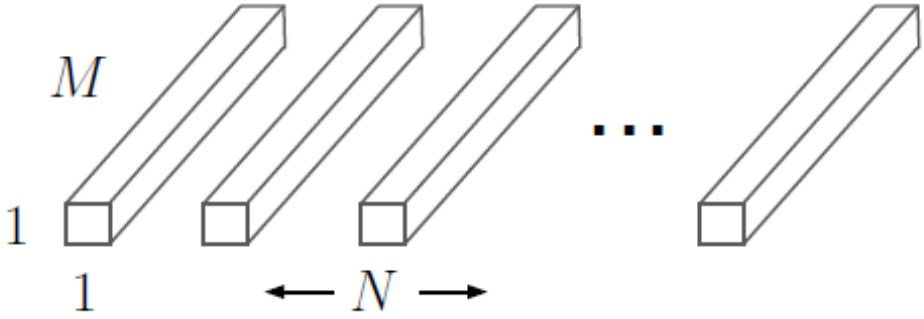
Method



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figure 2. The standard convolutional filters in (a) are replaced by two layers: depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter.

Basic

Replace **standard convolution filters** with combination of **depthwise convolution filters** and **pointwise convolution filters** and the time complexity can be reduce:

- Original Time Complexity: $D_K^2 NMD_F^2$
- Modified Time Complexity: $(D_K^2 + N)MD_F^2$

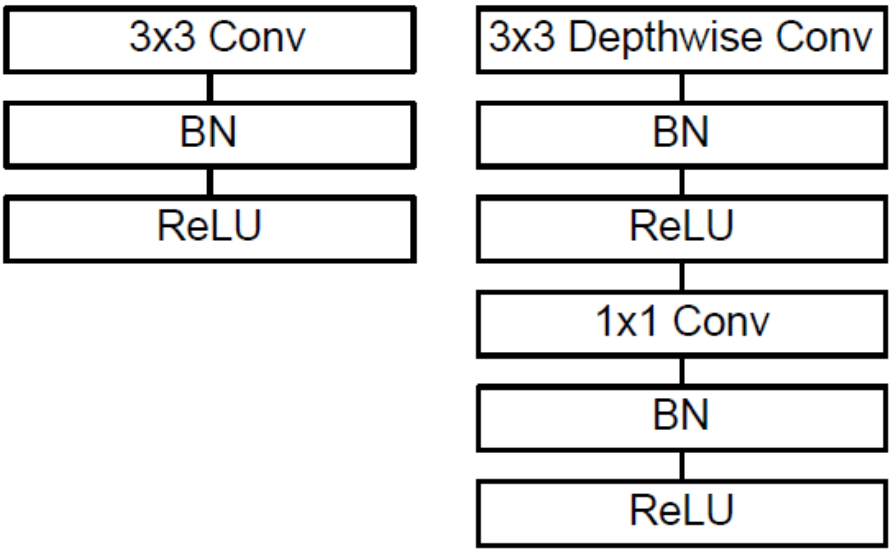


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

Result

Model	ImageNet Accuracy	Mult-Add(Million)	Parameters(Million)
MobileNet	70%	569	4.2
GoogleNet	69%	1550	6.8
VGG16	71%	15300	138