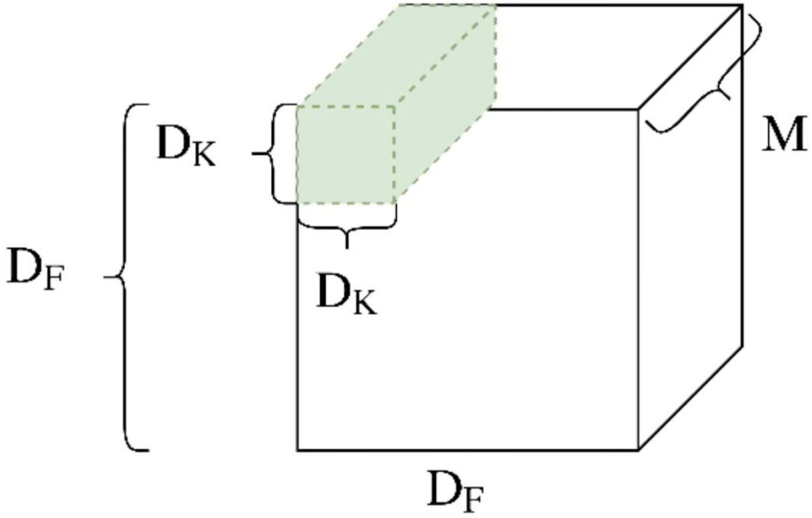
Tema 12 : MobileNet

Convolution



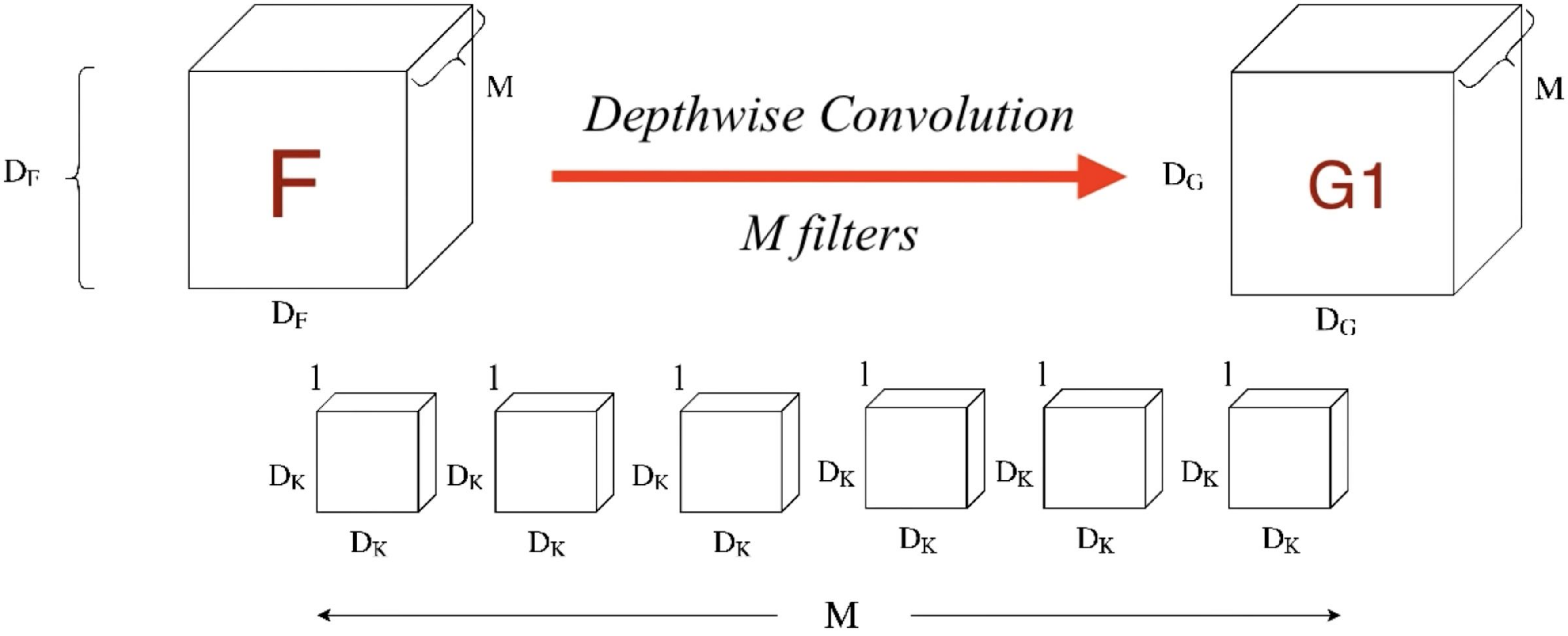
$$\text{Mults once} = \quad D_K^2 \times M$$

$$\text{Mults per Kernel} = D_G^2 \times D_K^2 \times M$$

$$\text{Mults N Kernels} = N \times D_G^2 \times D_K^2 \times M$$

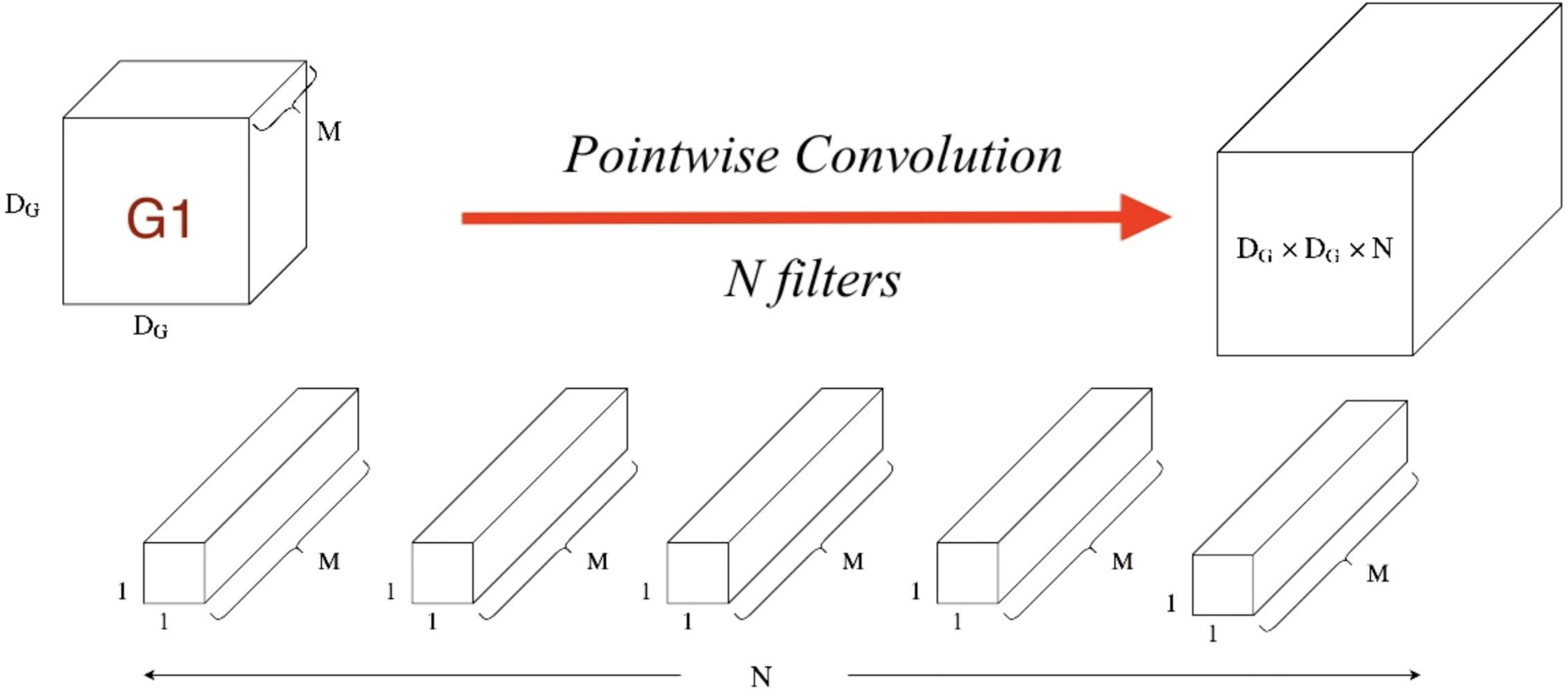
Depthwise Separable Convolution

1. Depthwise Convolution: Filtering Stage



Depthwise Separable Convolution

2. Pointwise Convolution: Filtering Stage



$$\text{Mults once} = D_K^2$$

$$\text{Mults 1 Channel} = D_G^2 \times D_K^2$$

$$\text{DC Mults} = M \times D_G^2 \times D_K^2$$

$$\text{Mults once} = M$$

$$\text{Mults 1 Kernel} = D_G \times D_G \times M$$

$$\text{PC Mults} = N \times D_G \times D_G \times M$$

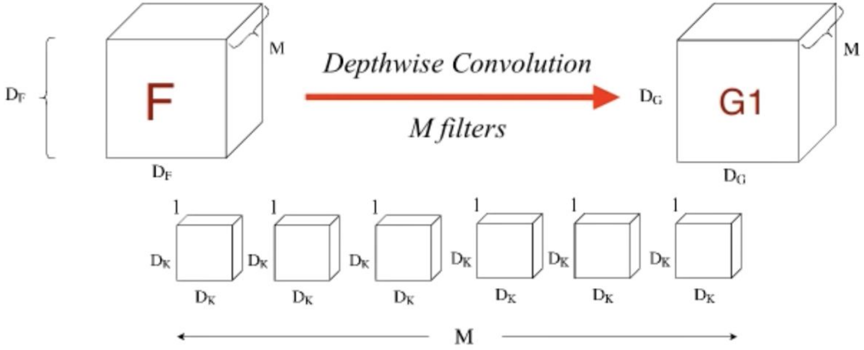
$$\text{Total} = \text{DC Mults} + \text{PC Mults}$$

$$M \times D_G^2 \times D_K^2 + N \times D_G^2 \times M$$

$$M \times D_G^2 \left(D_K^2 + N \right)$$

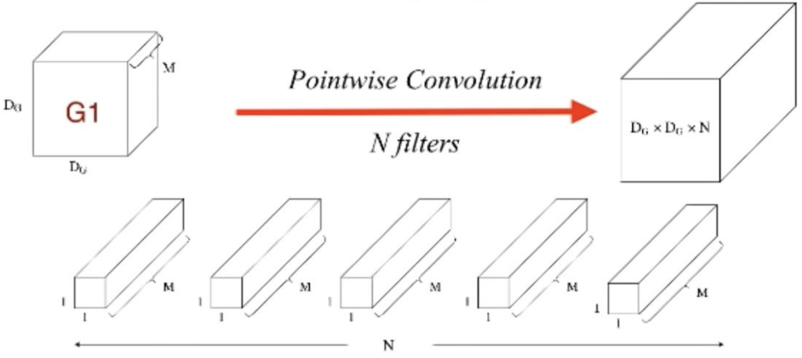
Depthwise Separable Convolution

1. Depthwise Convolution: Filtering Stage



Depthwise Separable Convolution

2. Pointwise Convolution: Filtering Stage



Comparison Standard Vs. Depthwise

$$\frac{No.Mults\ in\ Depthwise\ Separable\ Conv}{No.Mults\ in\ Standard\ Conv} = \frac{M \times D_G^2\ (\ D_K^2 + N)}{N\ \times D_G \times D_G \times D_K \times D_K \times M}$$

$$\frac{No.\ Mults\ in\ Depthwise\ Separable\ Conv}{No.\ Mults\ in\ Standard\ Conv} = \frac{D_K^2 + N}{(\ D_K^2 \times N)} = \frac{1}{N} + \frac{1}{D_K^2}$$

$$N\ = 1{,}024 \qquad D_K = 3$$

$$\frac{No.\ Mults\ in\ Depthwise\ Separable\ Conv}{No.\ Mults\ in\ Standard\ Conv} = \frac{1}{1024} + \frac{1}{3^2} = 0.112$$

Comparacion modelo clásico vs Implementación MobileNets

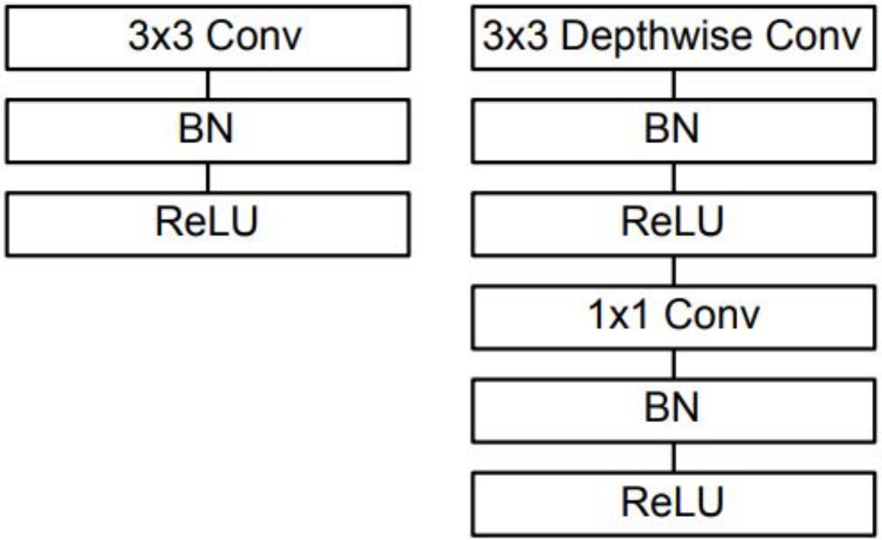


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

Table 14. MobileNet Distilled from FaceNet

Model	1e-4 Accuracy	Million Mult-Adds	Million Parameters
FaceNet [25]	83%	1600	7.5
1.0 MobileNet-160	79.4%	286	4.9
1.0 MobileNet-128	78.3%	185	5.5
0.75 MobileNet-128	75.2%	166	3.4
0.75 MobileNet-128	72.5%	108	3.8

Table 13. COCO object detection results comparison using different frameworks and network architectures. mAP is reported with COCO primary challenge metric (AP at IoU=0.50:0.05:0.95)

Framework Resolution	Model	mAP	Billion Mult-Adds	Million Parameters
SSD 300	deeplab-VGG	21.1%	34.9	33.1
	Inception V2	22.0%	3.8	13.7
	MobileNet	19.3%	1.2	6.8
Faster-RCNN 300	VGG	22.9%	64.3	138.5
	Inception V2	15.4%	118.2	13.3
	MobileNet	16.4%	25.2	6.1
Faster-RCNN 600	VGG	25.7%	149.6	138.5
	Inception V2	21.9%	129.6	13.3
	Mobilenet	19.8%	30.5	6.1

Table 12. Face attribute classification using the MobileNet architecture. Each row corresponds to a different hyper-parameter setting (width multiplier α and image resolution).

Width Multiplier / Resolution	Mean AP	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	88.7%	568	3.2
0.5 MobileNet-224	88.1%	149	0.8
0.25 MobileNet-224	87.2%	45	0.2
1.0 MobileNet-128	88.1%	185	3.2
0.5 MobileNet-128	87.7%	48	0.8
0.25 MobileNet-128	86.4%	15	0.2
Baseline	86.9%	1600	7.5

Things to Remember

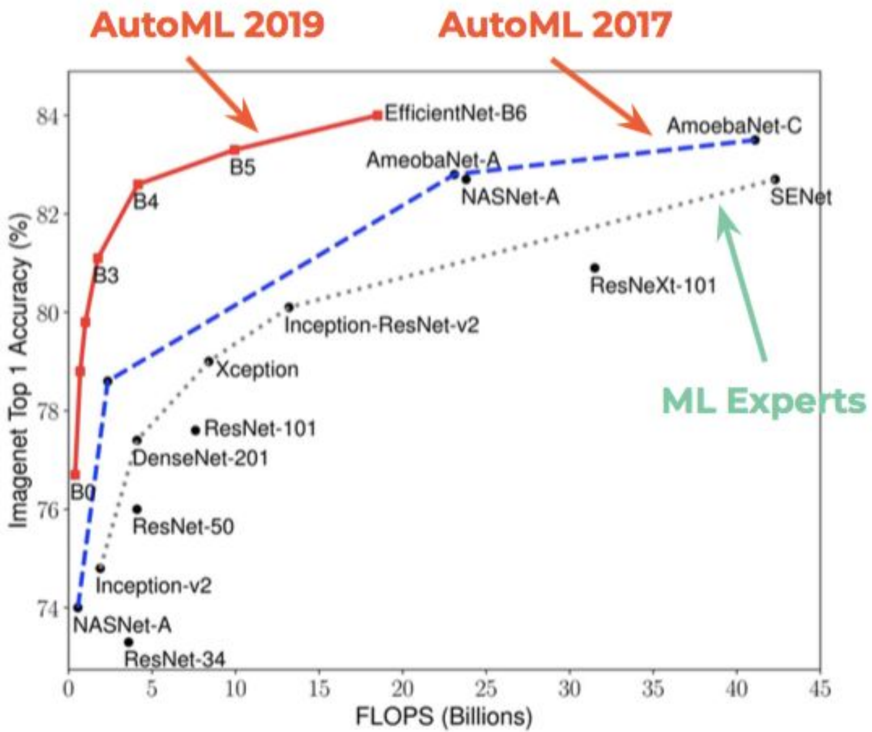
1. Depthwise Sep. Conv. reduces computation time, parameters

2. Depthwise Sep. Conv. = Depthwise Conv + Poinwise Conv

3. Used in recent architectures (MultiModel Nets, Xception, MobileNets).

Better Models, Across Multiple Modalities/Domains

Image Recognition



Google

Tan *et al*. EfficientNet: Rethinking Model Scaling for Deep Convolutional Neural Networks, ICML 2019, arxiv.org/abs/1905.11946

Referencias

- MobileNets : https://arxiv.org/pdf/1704.04861.pdf