
MobileNets Notes

1. Depthwise Separable Convolutions (深度可分离卷积)

设输入输出的特征图 channel 维度都是 C ，卷积核的大小是 f ，输出特征图的长宽都是 F ，则经过一个标准的卷积操作这个卷积层的参数量是： $f*f*C*C$ ，卷积操作的计算量是： $F*F$ (一个卷积核在输入的特征图上扫过的次数)* $f*f*C$ (一次卷积操作的计算量)* C (输出特征图的维度)

对于深度可分离卷积, 计算方法是先对输入的每个维度的特征图进行卷积, 这样操作之后, 输出特征图的 channel 等于输入特征图的 channel, 然后对得到的特征图进行 $1*1$ 卷积操作, 融合所有 channel 上的信息, 上述操作的参数量： $f*f*C + 1*1*C*C$ ，卷积操作的计算量是： $F*F*f*f*1 + 1*1*C*C$

下图是标准卷积核深度可分离卷积的对比

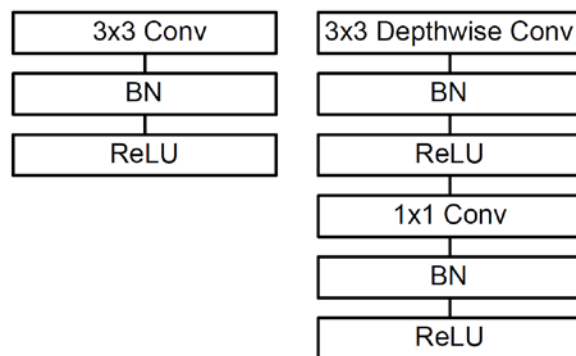


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

2. Architecture Baseline

基本的 MobileNets 结构包含 28 层，网络结构如下：

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5×	Conv dw / s1	$3 \times 3 \times 512$ dw
	Conv / s1	$1 \times 1 \times 512 \times 512$
	Conv dw / s2	$3 \times 3 \times 512$ dw
	Conv / s1	$1 \times 1 \times 512 \times 1024$
	Conv dw / s2	$3 \times 3 \times 1024$ dw
	Conv / s1	$1 \times 1 \times 1024 \times 1024$
	Avg Pool / s1	Pool 7×7
	FC / s1	1024×1000
	Softmax / s1	Classifier

3. Width Multiplier: Thinner Models

为了降低模型复杂度，可以考虑将每个卷积层 channel 的数量减少，这个参数 α 就是将 baseline 中的 channel 数量降到原来的 α 倍

Table 6. MobileNet Width Multiplier

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

4. Resolution Multiplier: Reduced Representation

另外一种降低模型复杂度的方法是将每层特征图缩小，一个较为直接的方法就是直接降低输入图片的分辨率

Table 7. MobileNet Resolution				
Resolution	ImageNet	Million	Million	
	Accuracy	Mult-Adds	Parameters	
1.0 MobileNet-224	70.6%	569	4.2	
1.0 MobileNet-192	69.1%	418	4.2	
1.0 MobileNet-160	67.2%	290	4.2	
1.0 MobileNet-128	64.4%	186	4.2	