MobileNet:2016

Efficient Convolutional Neural Networks for MObile Vision Applications

Google Inc.

论文背景

近几年的创新主要是将神经网络做的更深更复杂进而获得更高的准确率，然而忽略了网络的大小和速度。在现实生活中，经常需要将网络部署在一些算力有限的设备上，因而对网络的大小和推理速度要求较高

之前的解决思路要么是压缩预训练模型，要么就是直接训练一个小型网络

论文创新点

提出了一种在移动设备和嵌入式设备上面部署的神经网络叫做MobileNet，使用的是深度可分离卷积，引入了两个超参数进一步减少参数

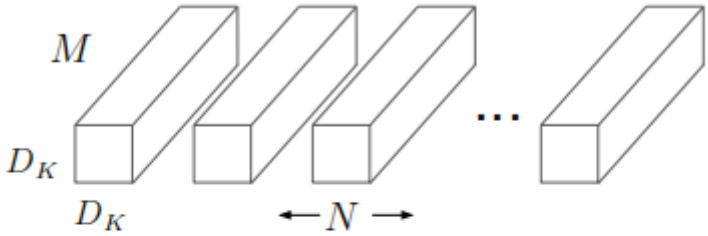
论文方法介绍

深度可分离卷积：一种分解普通卷积的方式，将一个标准卷积分解为深度卷积和逐点卷积，深度卷积对每个输入通道运用一个卷积核，之后使用一个逐点卷积组合深度卷积的输出。

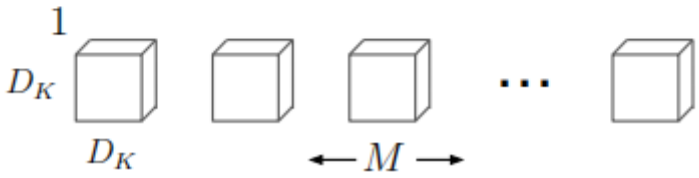
标准卷积的运算量$D_k \cdot D_k \cdot M \cdot N \cdot D_F \cdot D_F$,

深度卷积的运算量$D_k \cdot D_k \cdot M \cdot D_F \cdot D_F$

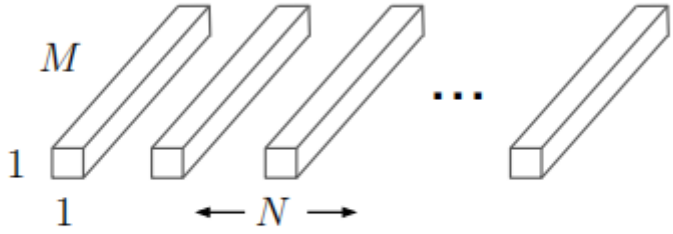
逐点卷积的运算量$M \cdot N \cdot D_M \cdot D_N$



(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.

除了第一层，MobileNet是一个全卷积网络;除了最后一个全连接层没有非线性激活函数函数，除了每一个层都在后面接了一个BN和ReLU

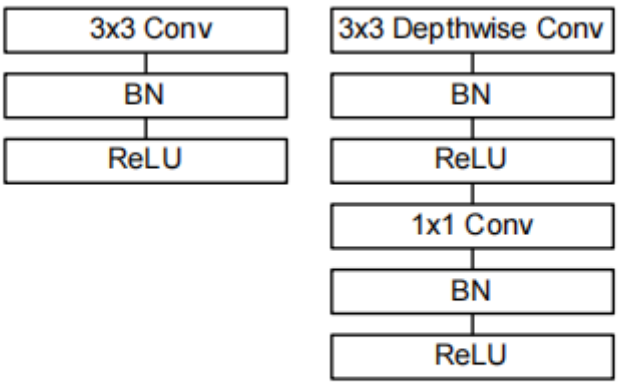


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

训练时引入一个参数α，称作宽度系数，作用是在每一层均匀细化一个网络

分辨率系数ρ

$$D_K \cdot D_K \cdot \alpha M \cdot \rho D_F \cdot \rho D_F + \alpha M \cdot \alpha N \cdot \rho D_F \cdot \rho D_F \quad (7)$$

实际效果

Table 4. Depthwise Separable vs Full Convolution MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Table 5. Narrow vs Shallow MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.75 MobileNet	68.4%	325	2.6
Shallow MobileNet	65.3%	307	2.9

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

Table 7. MobileNet Resolution

Resolution	ImageNet Accuracy	Million Mult-Adds	Million 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

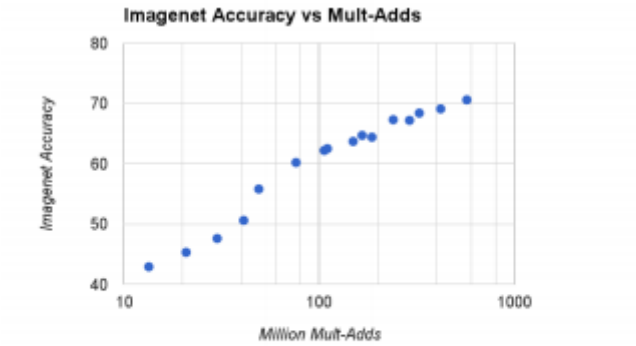


Figure 4. This figure shows the trade off between computation (Mult-Adds) and accuracy on the ImageNet benchmark. Note the log linear dependence between accuracy and computation.

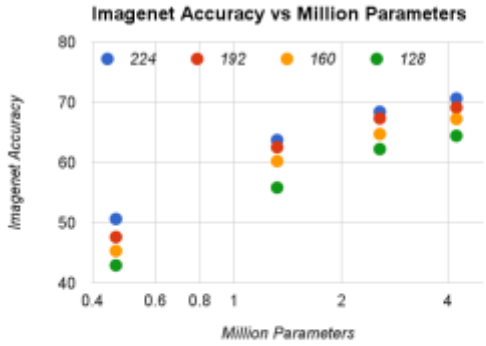


Figure 5. This figure shows the trade off between the number of parameters and accuracy on the ImageNet benchmark. The colors encode input resolutions. The number of parameters do not vary based on the input resolution.

Table 8. MobileNet Comparison to Popular Models

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

Table 9. Smaller MobileNet Comparison to Popular Models

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.50 MobileNet-160	60.2%	76	1.32
Squeezenet	57.5%	1700	1.25
AlexNet	57.2%	720	60

Table 10. MobileNet for Stanford Dogs

Model	Top-1 Accuracy	Million Mult-Adds	Million Parameters
Inception V3 [18]	84%	5000	23.2
1.0 MobileNet-224	83.3%	569	3.3
0.75 MobileNet-224	81.9%	325	1.9
1.0 MobileNet-192	81.9%	418	3.3
0.75 MobileNet-192	80.5%	239	1.9

Table 11. Performance of PlaNet using the MobileNet architecture. Percentages are the fraction of the Im2GPS test dataset that were localized within a certain distance from the ground truth. The numbers for the original PlaNet model are based on an updated version that has an improved architecture and training dataset.

Scale	Im2GPS [7]	PlaNet [35]	PlaNet MobileNet
Continent (2500 km)	51.9%	77.6%	79.3%
Country (750 km)	35.4%	64.0%	60.3%
Region (200 km)	32.1%	51.1%	45.2%
City (25 km)	21.9%	31.7%	31.7%
Street (1 km)	2.5%	11.0%	11.4%

个人理解

之前在做视频分类任务的时候了解过一些拆分卷积的思想，比如P3D将3D卷积拆分成了时间域的和空间域的，减少了参数以及降低了训练难度。这种思路在数字信号处理之前也接触过（具体记不清是什么算法了）。引入缩放因子的思想也在后面的PVT中减少SA模块计算量引入过。

针对这篇文章的话，就是对普通卷积做了一个分解，深度卷积去主要进行提特征，但是忽略了通道之间的信息交互，引入逐点卷积来对通道之间进行信息交互，两者是互补的。然后就是引入衰减因子，这个我觉得其实很常见，毕竟除了算法创新以外，最直接的就是对数据动刀了。