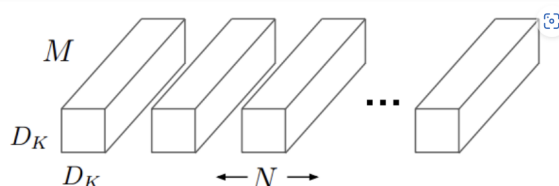


MobileNet

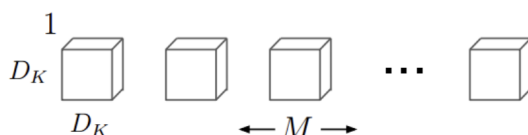
随着深度学习的发展，卷积神经网络变得越来越普遍。当前发展的总体趋势是，通过更深和更复杂的网络来得到更高的精度，但是这种网络往往在模型大小和运行速度上没多大优势。一些嵌入式平台上的应用比如机器人和自动驾驶，它们的硬件资源有限，就十分需要一种轻量级、低延迟（同时精度尚可接受）的网络模型，于是MobileNet应运而生。

网络简介

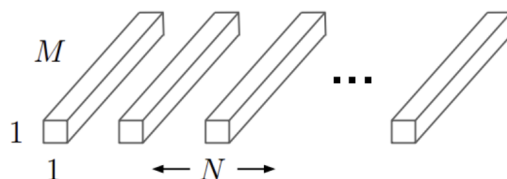
MobileNets模型基于深度可分解的卷积，它可以将标准卷积分解成一个深度卷积和一个点卷积（ 1×1 卷积核）。深度卷积将每个卷积核应用到每一个通道，而 1×1 卷积用来组合通道卷积的输出。后文证明，这种分解可以有效减少计算量，降低模型大小。图2说明了标准卷积是如何进行分解的。



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution <http://blog.csdn.net/>

MobileNets使用了大量的 3×3 的卷积核，极大地减少了计算量（ $1/8$ 到 $1/9$ 之间），同时准确率下降的很少，相比其他的方法确有优势。

网络结构

MobileNets结构建立在上述深度可分解卷积中（只有第一层是标准卷积）。该网络允许我们探索网络拓扑，找到一个适合的良好网络。其具体架构在表1说明。除了最后的全连接层，所有层后面跟了batchnorm和ReLU，最终输入到softmax进行分类。图3对比了标准卷积和分解卷积的结构，二者都附带了BN和ReLU层。按照作者的计算方法，

MobileNets总共28层 ($1 + 2 \times 13 + 1 = 28$) 。

具体网络结构如下图所示。

首先输入的是224x224x3的三通道图像，然后经过若干个3x3、1x1的卷积层提取特征，最终输入到softmax层进行分类。

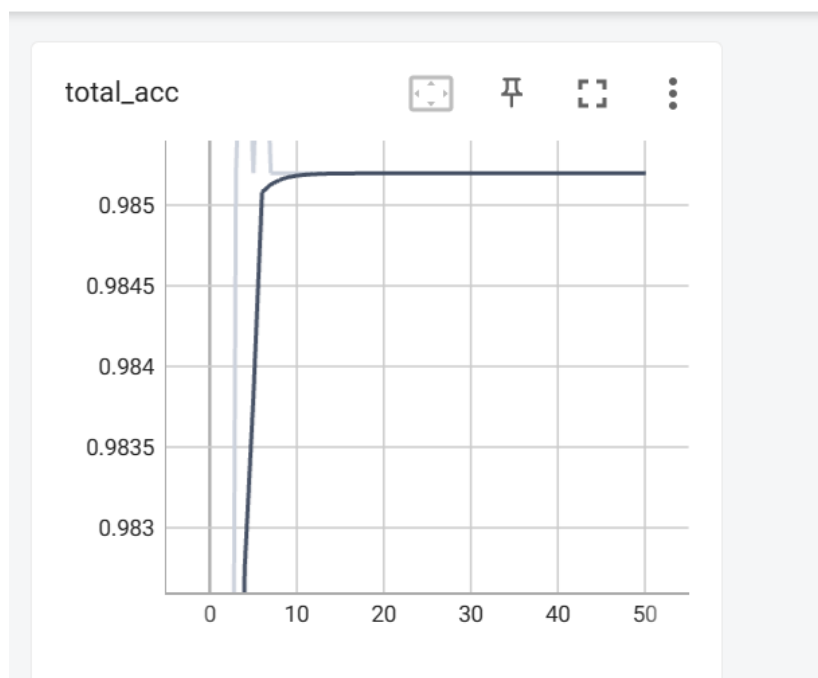
Table 1. MobileNet Body Architecture

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

训练过程

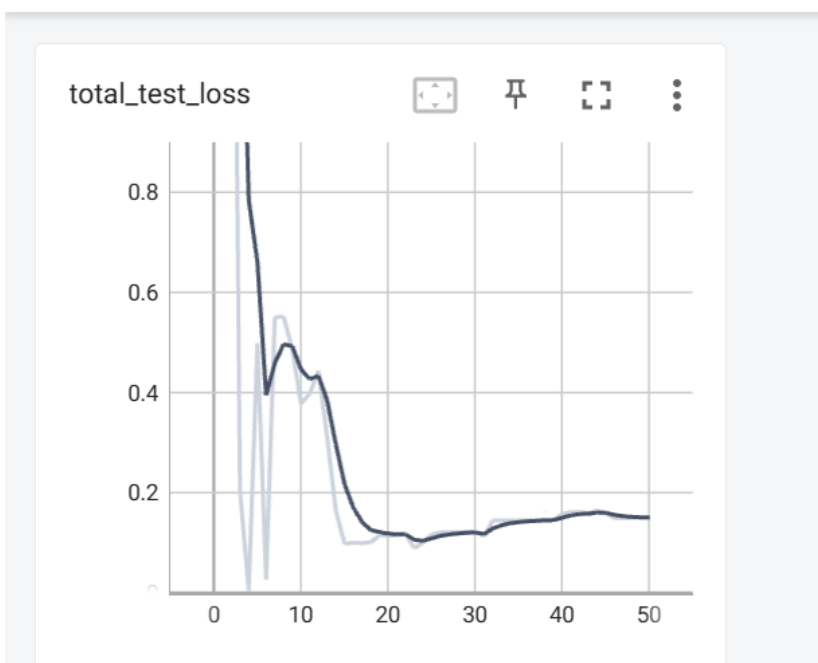
总准确率

total_acc



损失函数

total_test_loss



我们的MobileNet网络一共训练了50个epoch。

根据上图训练的总准确率和损失函数变化过程的曲线所示，可以看到，准确率随着训练轮数的增长而快速增长，最终趋于平稳，达到98.5%左右；损失值则随着训练轮数的增长而整体呈现下降趋势，最终趋于0.15左右。