MobilenetV1 V2网络详解

传统的卷积神经网络，内存需求大，运算量大。对移动设备和嵌入式设备很不友好。

Vgg16权重大小大概490MB resnet-152权重644MB，基本上不可能在嵌入式设备运行。

目的是服务于我们的生活、

Mobilenet专注于移动端和嵌入式设备的轻量级CNN网络。相比传统神经网络，在准确率小幅降低的情况下，大大减少模型的参数与运算量（相比VGG16减少了0.9%的准确度，但是参数只有VGG的1/32）（ImageNet数据集）

MobilenetV1

Paper：MobileNets ：Efficient Convolutional Neural Networks for Mobile Vision Application

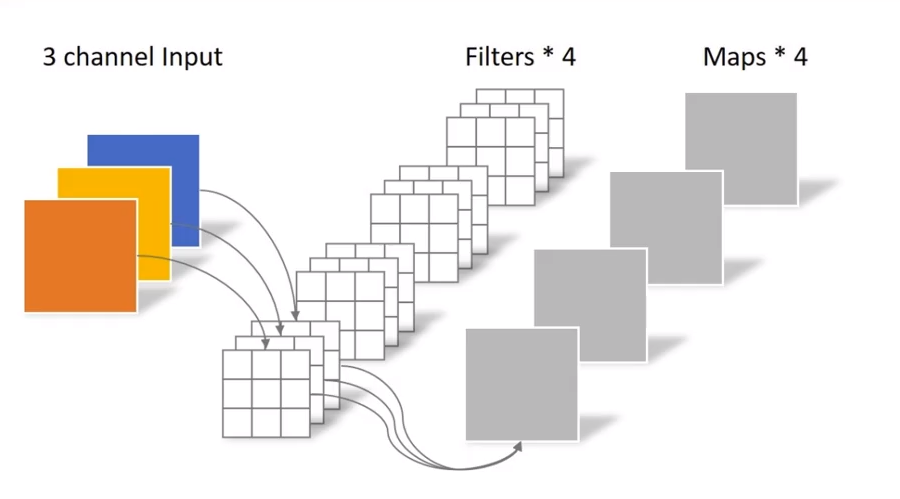
网络中的亮点：

1. Depthwise Convolution
2. 超参数alpha beta

传统卷积：

卷积核channel = 输入特征矩阵channel

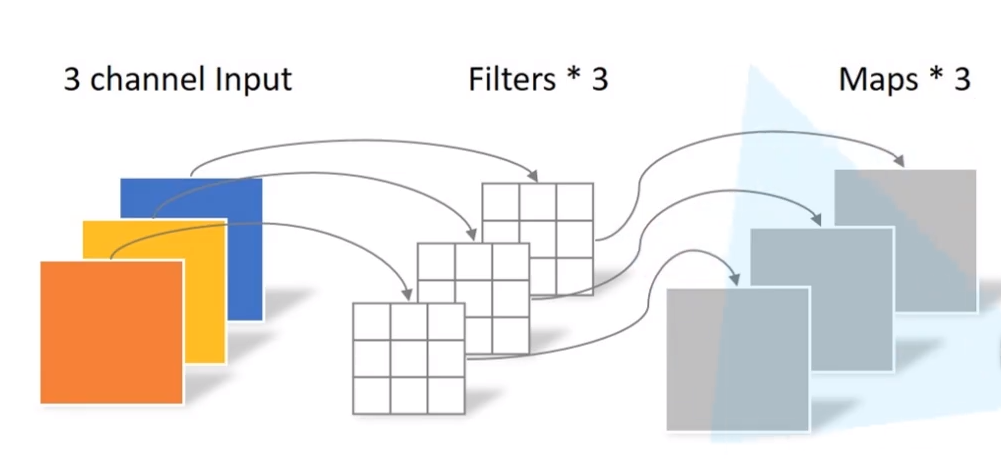
输出特征矩阵channel = 卷积核个数



Depthwise Convolution：

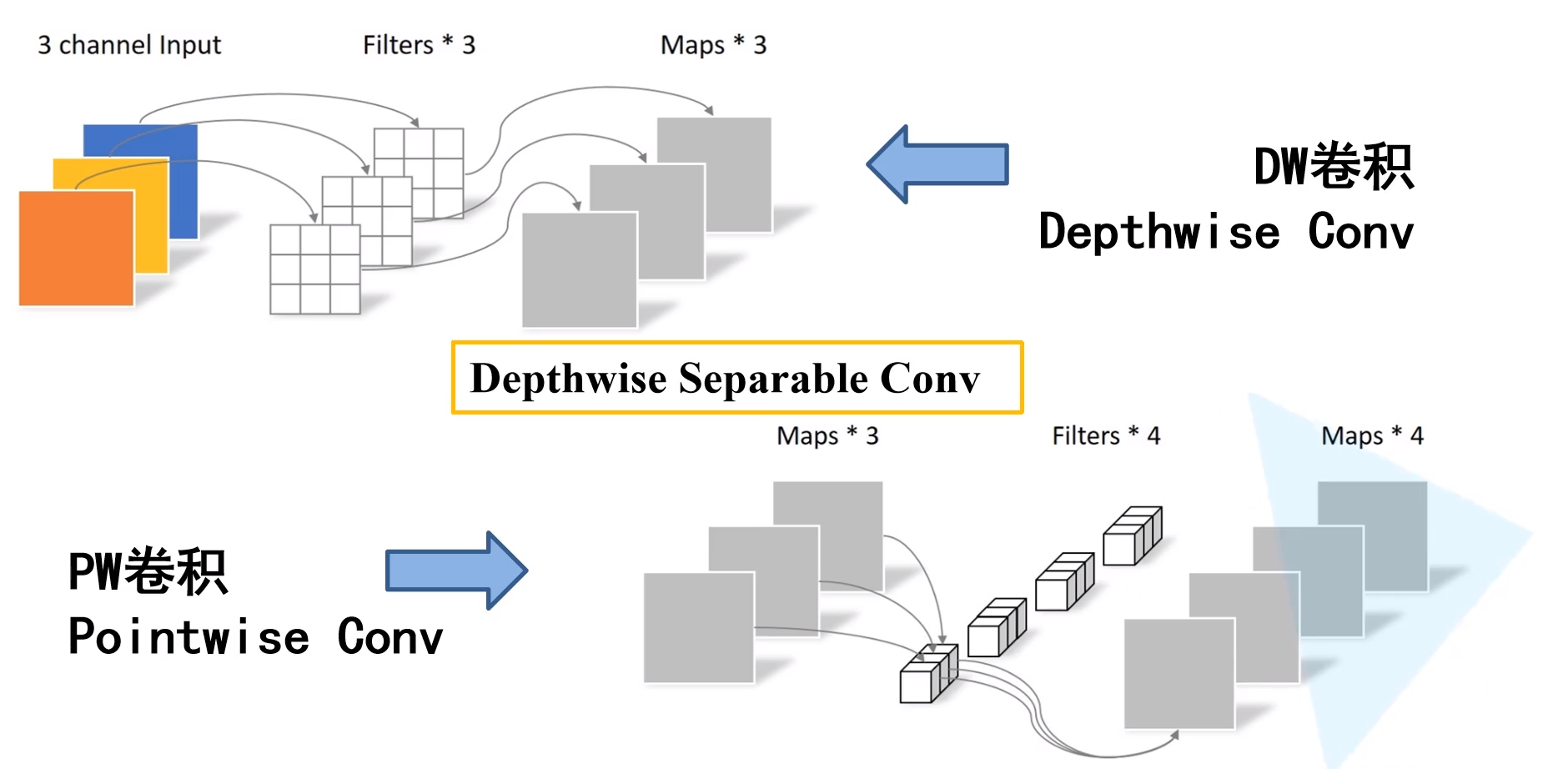
卷积核channel = 1

输入特征矩阵channel = 卷积核个数 = 输出特征矩阵channel



Depthwise Separable Conv:

分为DW卷积（Depthwise Convolution），PW卷积（Pointwise Convolution）



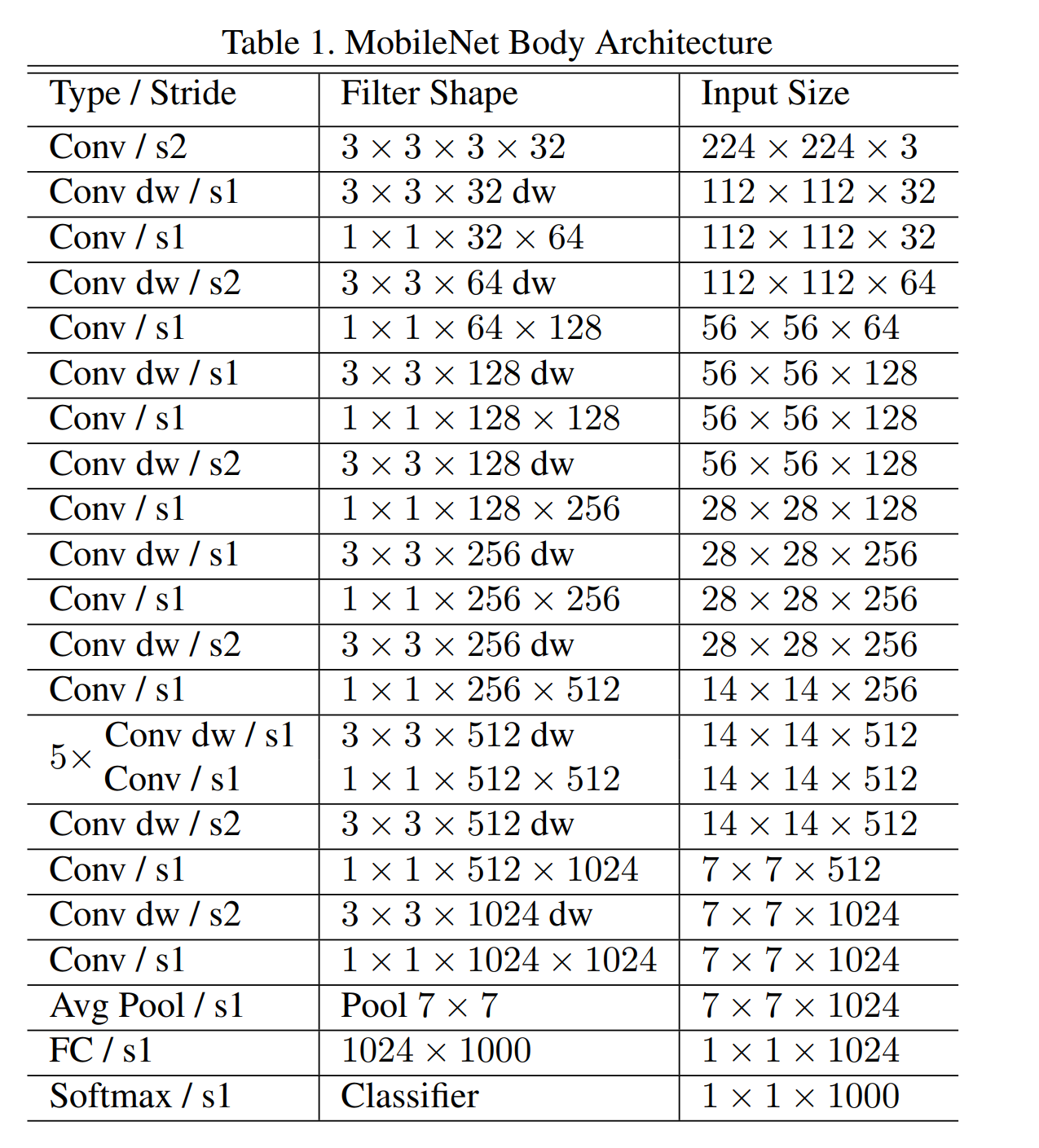
参数量对比：

|  |  |  |
| --- | --- | --- |
|  |  |  |

其中分子为DW+PW卷积操作的参数量，分母为普通卷积操作。为卷积核大小，为输入特征图的大小。为输入特征图的channel，为输出特征图的channel。

**理论上**普通的卷积计算量是DW+PW的八到九倍。

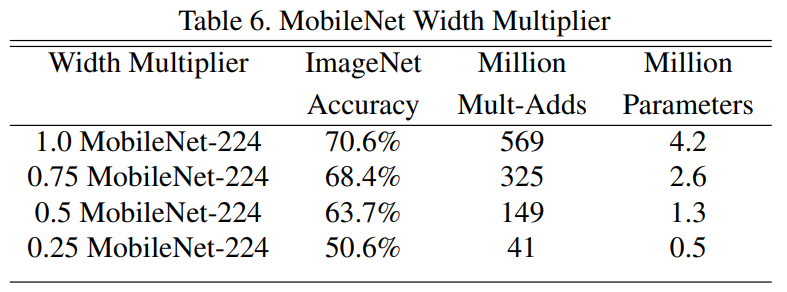
mobileNetV1结构



使用Width Multiplier: alpha

可以根据参数建立Thinner Models

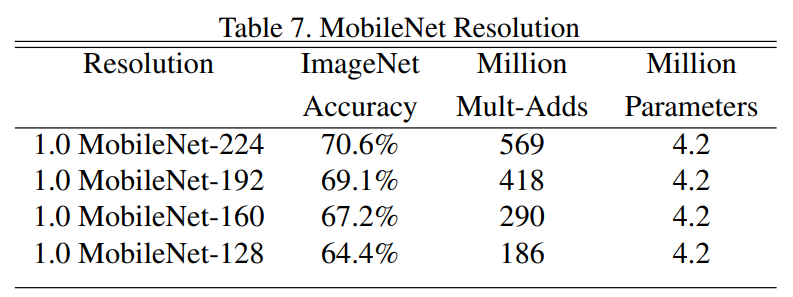
效果如下表



使用Resolution Multiplier: beta

进行Reduced Representation

减少特征图的分辨率，进行运算量的压缩，效果如下表：



但是在后续实验中发现，Depthwise部分的卷积和容易废掉，也就是卷积核大部分参数为零。

MobileNetV2

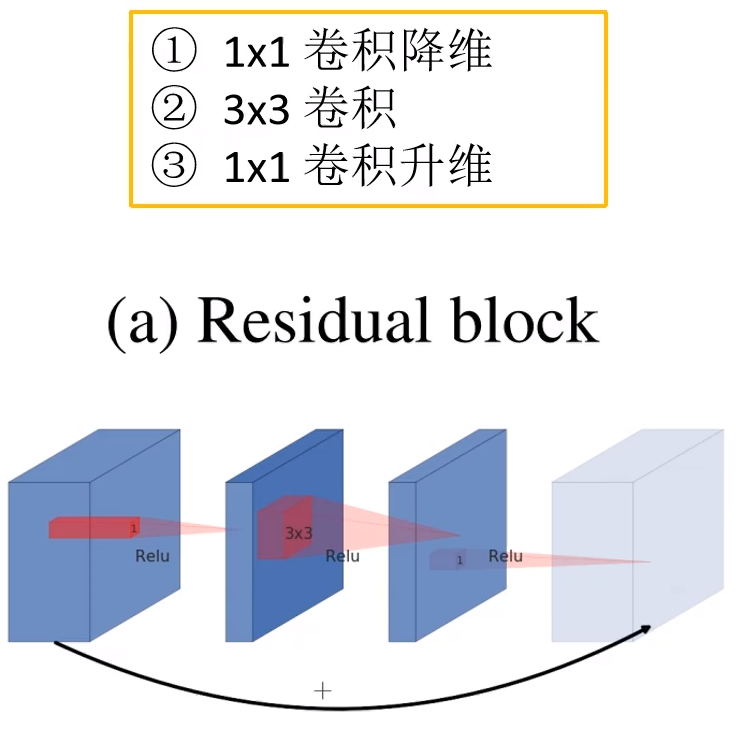
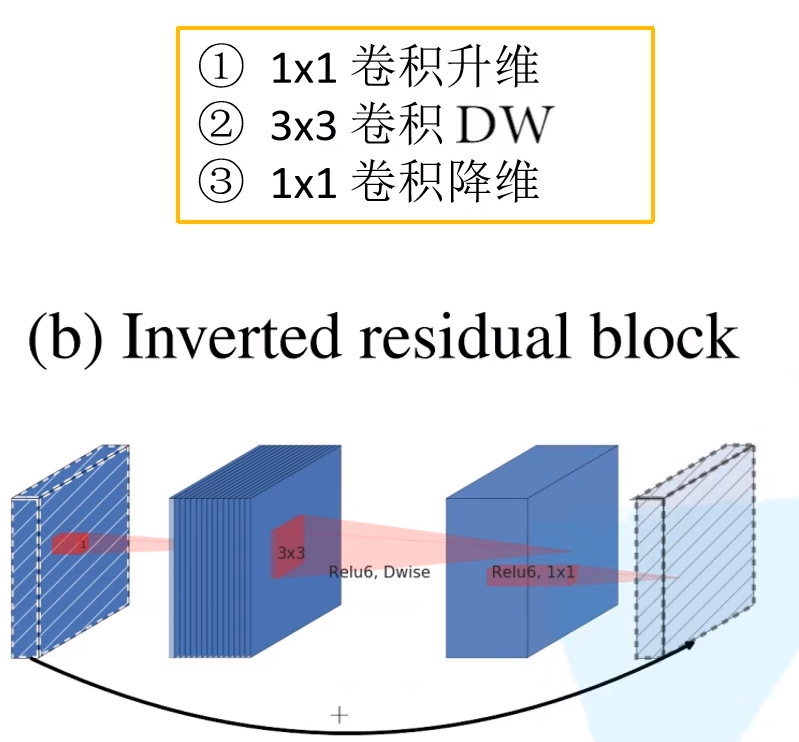
由google团队在2018年提出，相比mobilenetV1，准确率更高，模型更小。

Paper：MobileNetV2：Inverted Residuals and Linear Bottlenecks

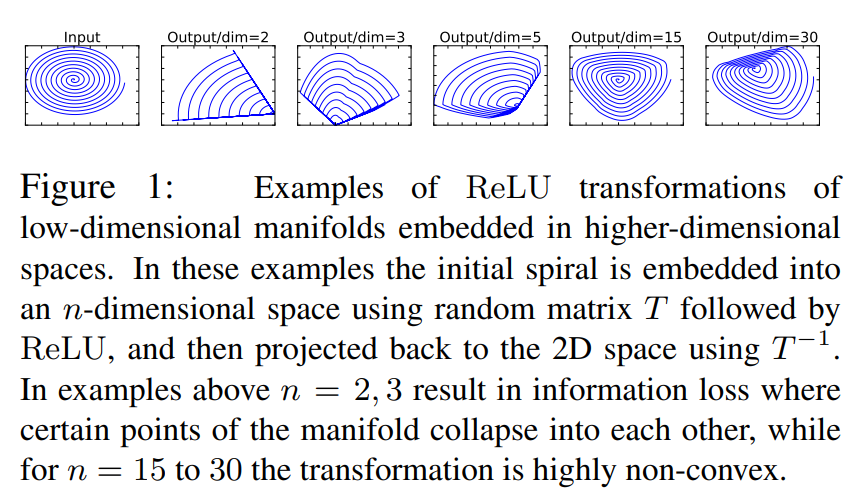
亮点：

1. Inverted Residuals（倒残差结构）
2. Linear Bottlenecks

Inverted Residuals

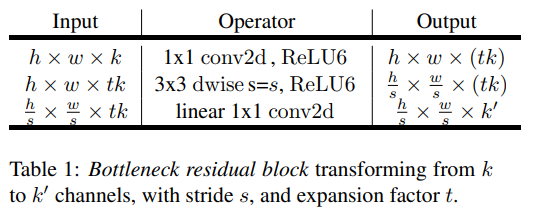
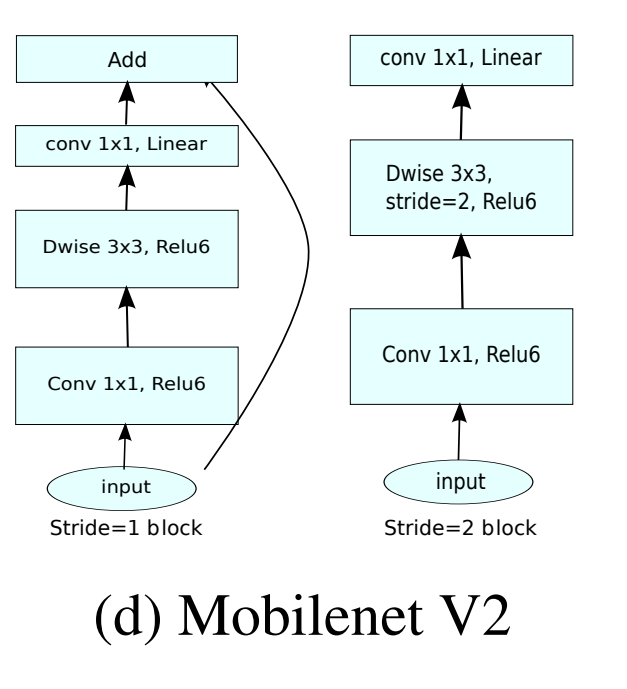
 

Linear Bottlenecks

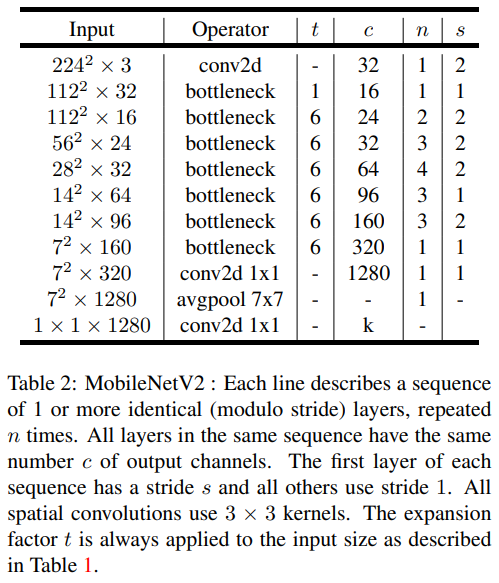


在通过实验之后，作者发现，在较低维度的情况下使用relu6函数进行激活会造成信息的丢失，在高维情况下信息则会更好的保留。由于倒残差结构的特点是两头薄中间纬度高，因此在中间两层使用ReLU6函数进行激活，在降维操作后使用线性激活函数。

MobileNetV2 block 图：



MobileNetV2 Architecture：



t是扩展因子

c是输出特征矩阵的channel

n是bottleneck的重复次数

s是步距（针对第一层，其他为1）

ps：其中短连接层因为每一bottleneck第一层要进行一次channel变换，因此shortcut的channel要对齐。