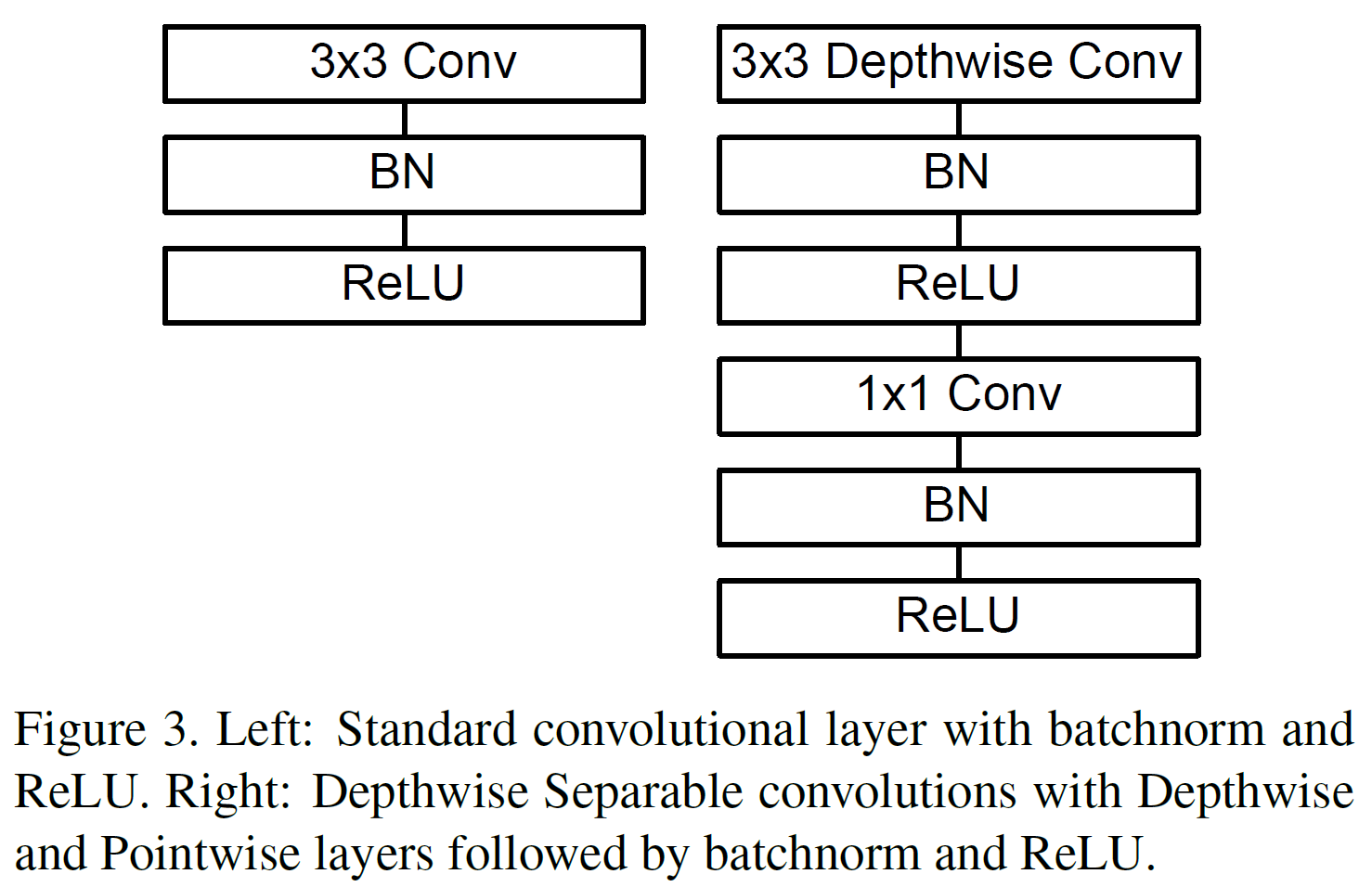
MobileNets Notes

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)

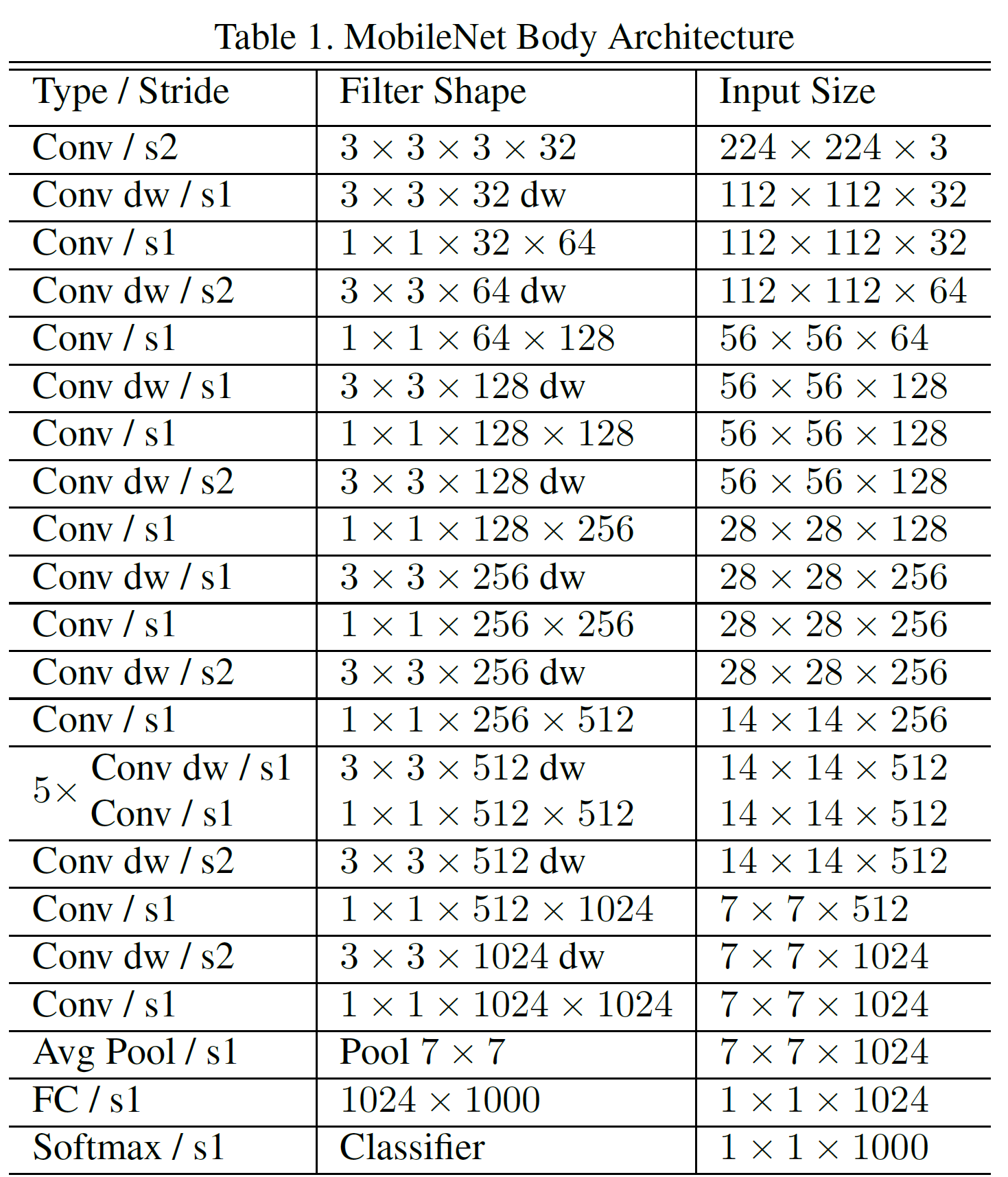
对于深度可分离卷积，计算方法是先对输入的每个维度的特征图进行卷积，这样操作之后，输出特征图的channel等于输入特征图的channel，然后对得到的特征图进行1\*1卷积操作，融合所有channel上的信息，上述操作的参数量：f\*f\*C + 1\*1\*C\*C，卷积操作的计算量是：F\*F\*f\*f\*1 + 1\*1\*C\*C

下图是标准卷积核深度可分离卷积的对比



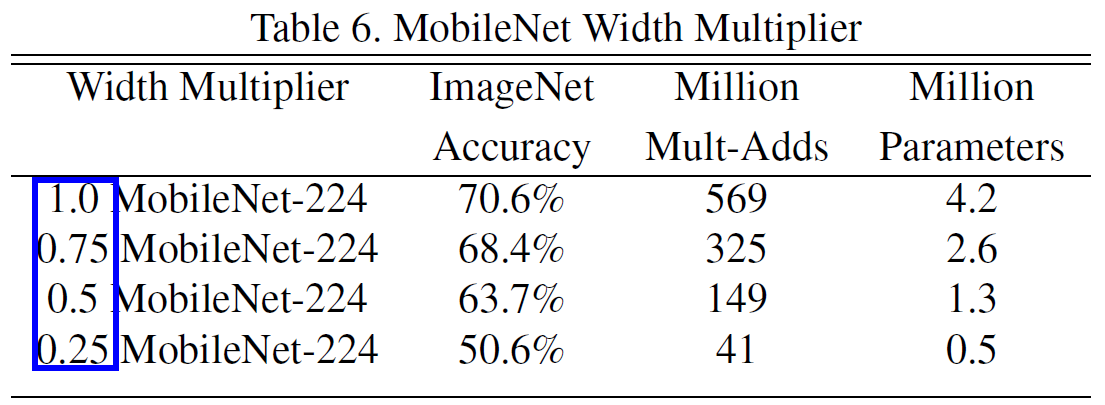
1. Architecture Baseline

基本的MobileNets结构包含28层，网络结构如下：



1. 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



1. 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

