

8.SqueezeNet

2016年提出，论文地址[SQUEEZENET: ALEXNET-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE](#)

在不大幅降低模型精度的前提下，最大程度的提高运算速度。提高运算速度有两个可以调整的方向：

- 1.减少可学习参数的数量
- 2.减少整个网络的计算量

这个方向带来的效果是非常明显的：

- 1.减少模型训练和测试时候的计算量，单个step的速度更快；
- 2.减小模型文件的大小，更利于模型的保存和传输；
- 3.可学习参数更少，网络占用的显存更小。

SqueezeNet正是诞生在这个环境下的一个精度的网络，它能够在ImageNet数据集上达到AlexNet近似的效果，但是参数比AlexNet少50倍，结合他们的模型压缩技术 Deep Compression，模型文件可比AlexNet小510倍。

8.1 SqueezeNet的压缩策略

SqueezeNet的模型压缩使用了3个策略：

- 1.将3×3卷积替换成1×1卷积；
- 2.减少3×3卷积输入输出通道数；
- 3.将降采样后置，提升精度但是会增加网络的计算量。

8.2 Fire模块

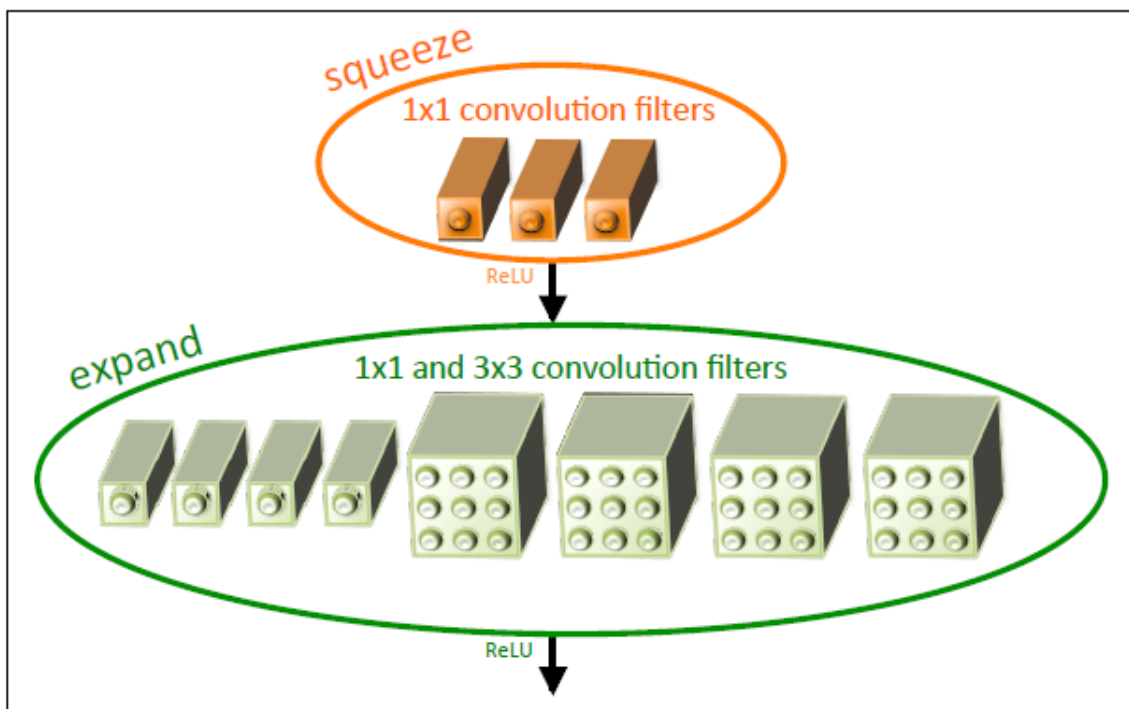


图31 Fire Module

Fire模块由Squeeze和Expand组成，Squeeze由一组连续的1×1卷积组成，Expand由一组连续的1×1卷积核一组连续的3×3卷积concatenate组成，因此3×3卷积需要使用SAME卷积。

8.3 SqueezeNet网络结构

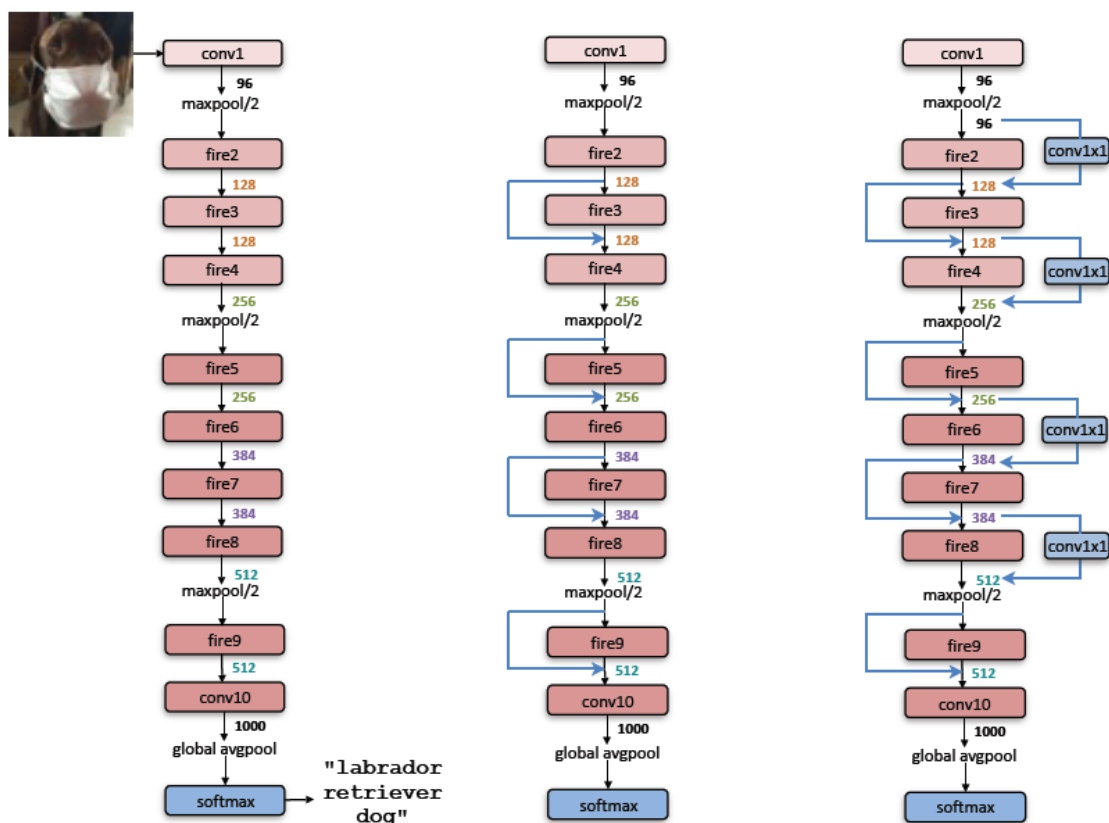


图32 SqueezeNet网络结构

左边是不加shortcut的，中间是加了shortcut的，右侧是shortcut跨有不同Feature Map个数的卷积的。

[pytorch实现](#)