

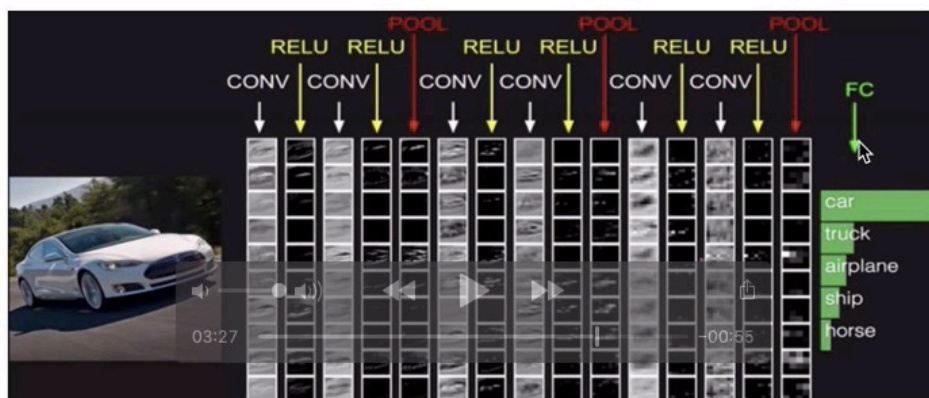
全连接网络问题在于

- 1.需要处理的数据量大，效率低
- 2.图像在维度调整的过程中很难保留原有的特征，导致图像处理的准确率不高

CNN组成

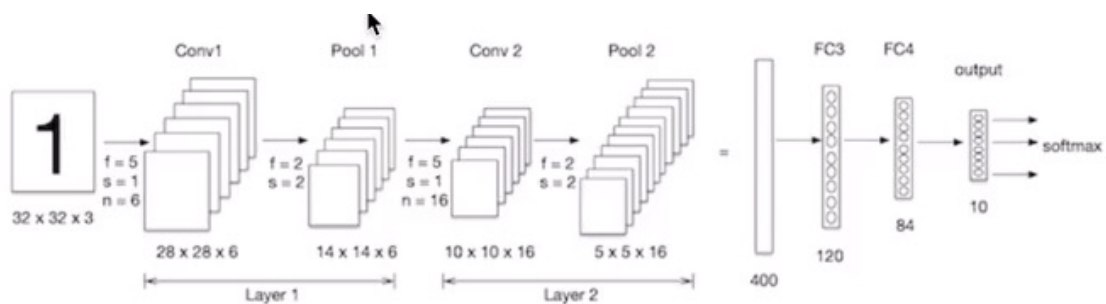


整个CNN网络结构如下图所示：



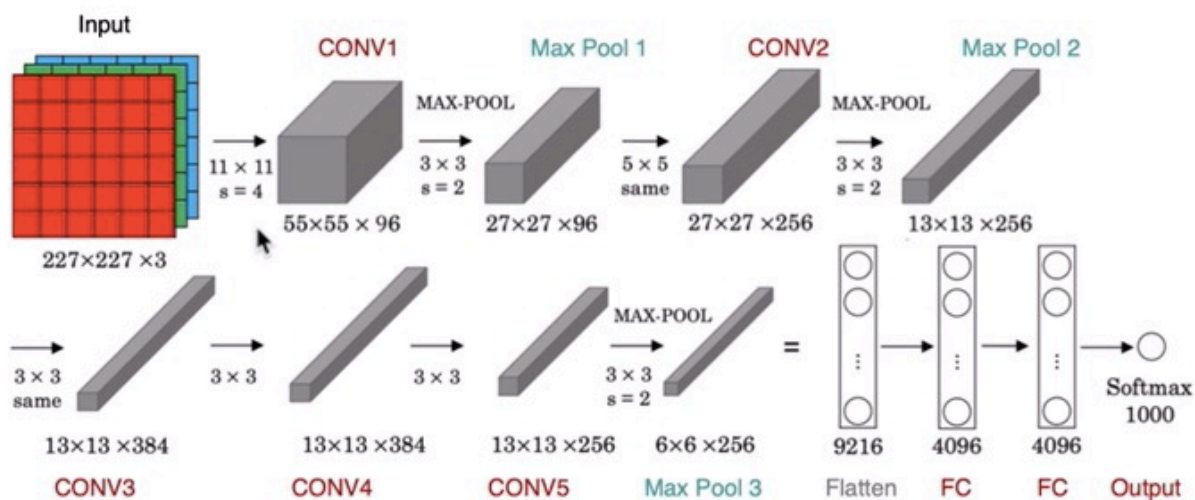
LeNet-5是一个较简单的卷积神经网络

输入二维图像，经过两次卷积-池化，再经过全连接层，最后用softmax分类作为输出层。



-----图片分类

AlexNet



该网络的特点是：

AlexNet包含8层变换，有5层卷积和2层全连接隐藏层，以及1个全连接输出层

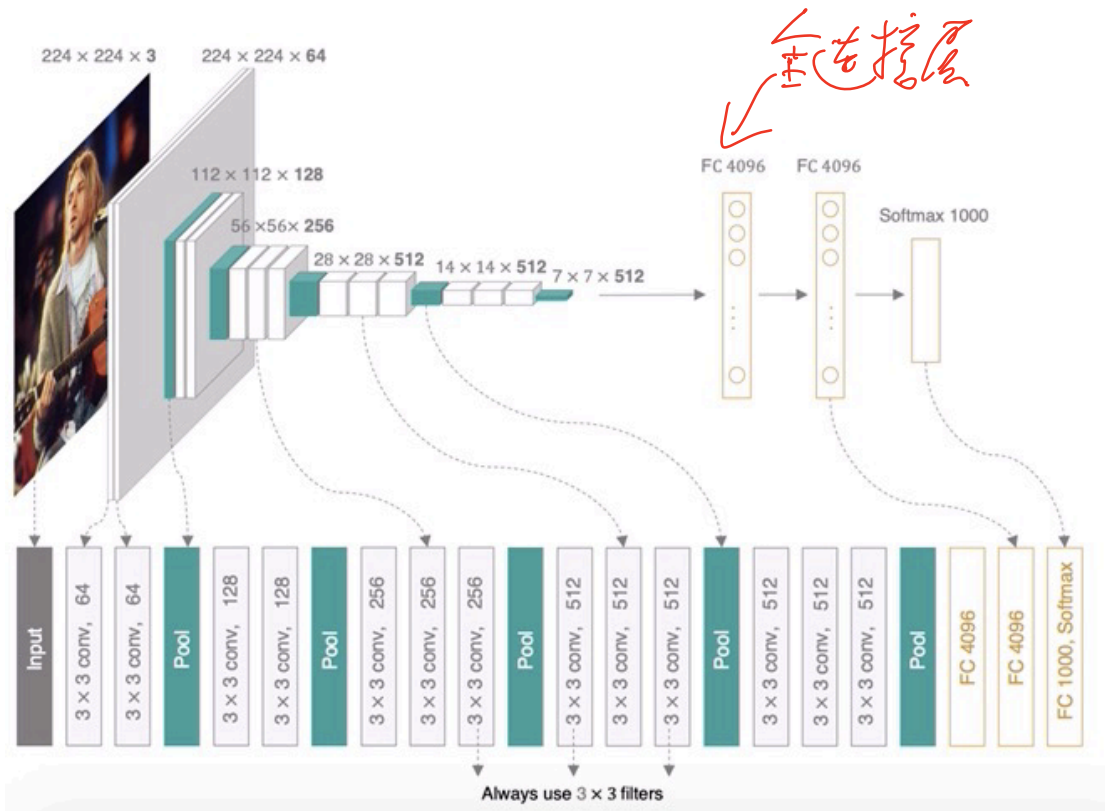
AlexNet第一层中的卷积核形状是 $11 \times 11$ 。第二层中的卷积核形状减小到 $5 \times 5$ ，之后全采用 $3 \times 3$ 。所有的池化层窗口大小为 $3 \times 3$ 、步幅为2的最大池化。

AlexNet将sigmoid激活函数改成了ReLU激活函数，使计算更简单，网络更容易训练

AlexNet通过dropOut来控制全连接层的模型复杂度。

AlexNet引入了大量的图像增强，如翻转、裁剪和颜色变化，从而进一步扩大数据集来缓解过拟合。

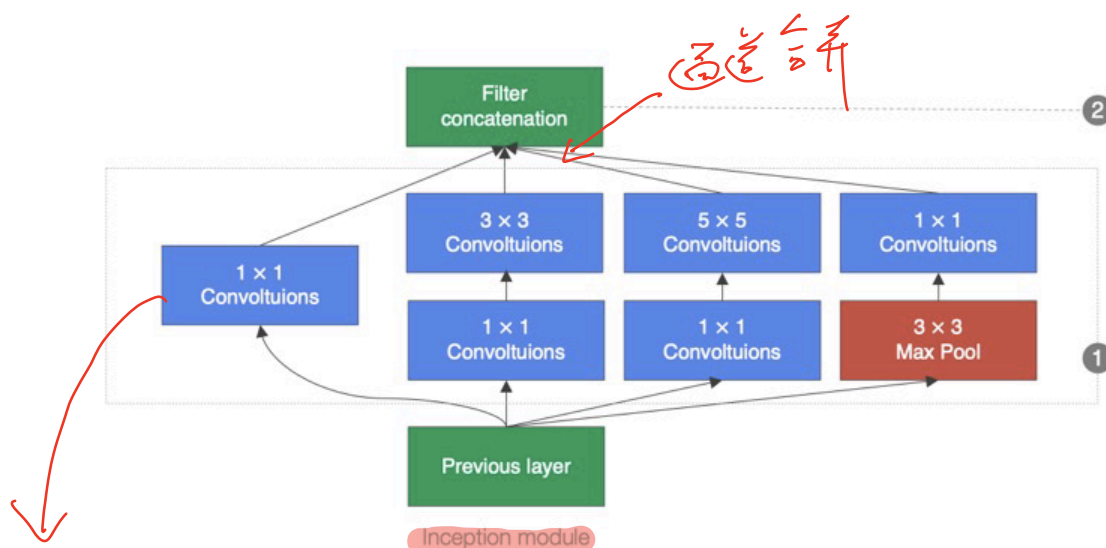
## VGG网络



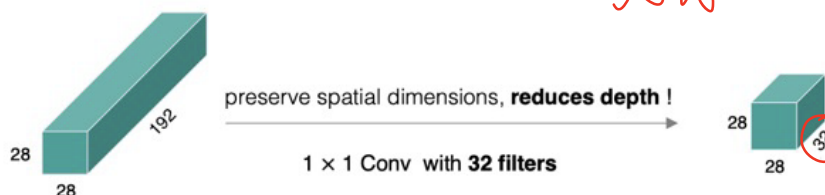
VGGNet使用的都是3x3的小卷积核，通过不断加深网络来提升性能。可以通过重复使用简单的基础块来构建深度模型

## GoogLeNet

GoogLeNet和AlexNet/VGGNet这类依靠加深网络结构的深度的思想不完全一样。GoogLeNet在加深度的同时做了结构上的创新，引入了一个叫做Inception的结构来代替之前的卷积加激活的经典组件。GoogLeNet在ImageNet分类比赛上的Top-5错误率降低到了6.7%。。



1x1 卷积:



对原先的192个通道整合

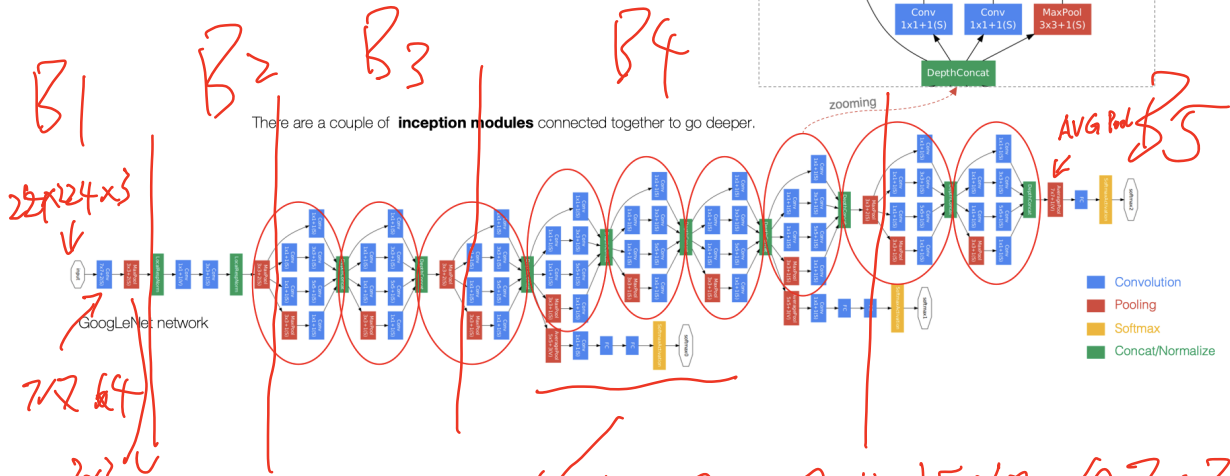
它的作用主要是:

实现跨通道的交互和信息整合

卷积核通道数的降维和升维，减少网络参数

激活函数都用的是 Relu  
输出用的是 Softmax

11/11/2017



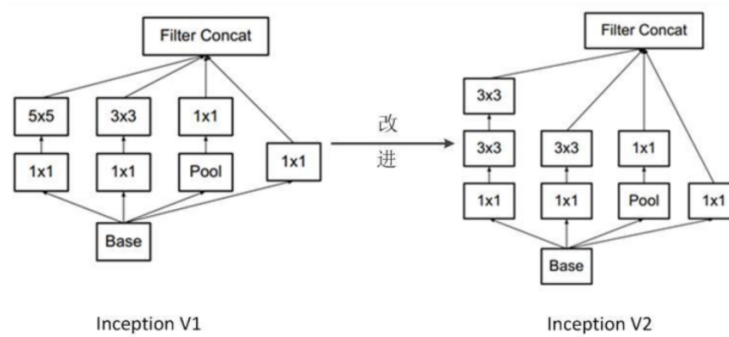
分为5个模块. 每个模块之间使用步幅为2的 $3 \times 3$ 最大池化层来减小输出高宽.

增加辅助分类器, 发现网络中间层具有强的识别能力, 为了利用中间层抽象的特征, 加入分类器

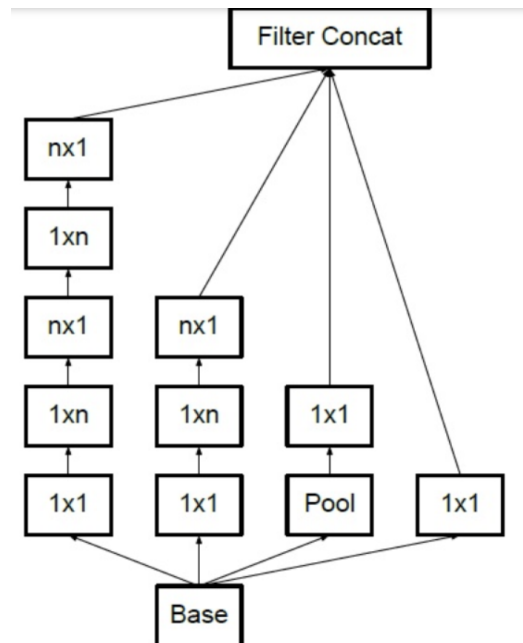
注意!  
B5后面不是全连接层, 而是 Global Average Pooling.  
GAP: 将特征图每一通道中所有像素值相加后求平均.

GoogLeNet是以InceptionV1为基础进行构建的，所以GoogLeNet也叫做InceptionNet,在随后的几年里，研究人员对GoogLeNet进行了数次改进，就又产生了InceptionV2，V3,V4等版本。

在InceptionV2中将大卷积核拆分为小卷积核，将V1中的 $5 \times 5 \times 5$ 的卷积用两个 $3 \times 3 \times 3$ 的卷积替代，从而增加网络的深度，减少了参数。



在Inception V3中将 $n \times n$ 卷积分割为 $1 \times n$ 和 $n \times 1$ 两个卷积，例如，一个的 $3 \times 3$ 卷积首先执行一个 $1 \times 3$ 的卷积，然后执行一个 $3 \times 1$ 的卷积,这种方法的参数量和计算量都比原来降低。

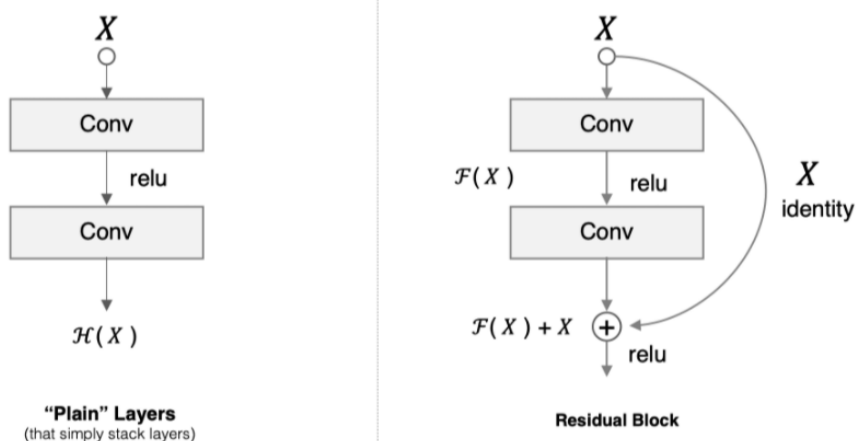


## ResNet

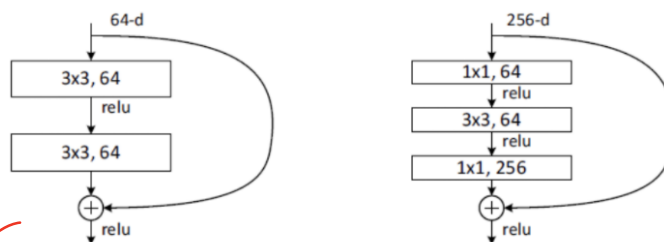
网络越深，获取的信息就越多，特征也越丰富。但是在实践中，随着网络的加深，优化效果反而越差，测试数据和训练数据的准确率反而降低了。

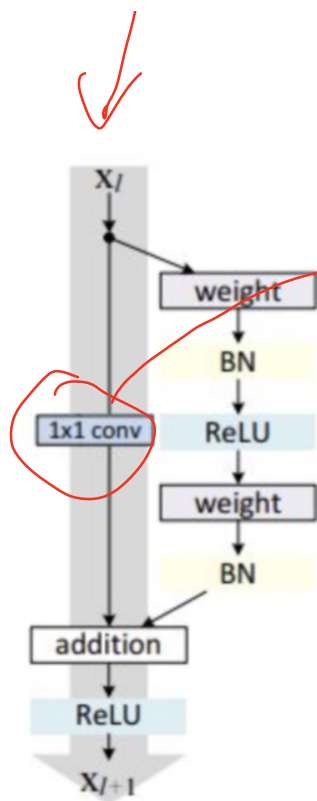
### 残差块

假设  $F(x)$  代表某个只包含有两层的映射函数， $x$  是输入， $F(x)$  是输出。假设他们具有相同的维度。在训练的过程中我们希望能够通过修改网络中的  $w$  和  $b$  去拟合一个理想的  $H(x)$  (从输入到输出的一个理想的映射函数)。也就是我们的目标是修改  $F(x)$  中的  $w$  和  $b$  逼近  $H(x)$ 。如果我们改变思路，用  $F(x)$  来逼近  $H(x)-x$ ，那么我们最终得到的输出就变为  $F(x)+x$  (这里的加指的是对应位置上的元素相加，也就是element-wise addition)，这里将直接从输入连接到输出的结构也称为shortcut，那整个结构就是残差块，ResNet的基础模块。



ResNet沿用了VGG全 $3\times 3$ 卷积层的设计。残差块里首先有2个有相同输出通道数的 $3\times 3$ 卷积层。每个卷积层后接BN层和ReLU激活函数，然后将输入直接加在最后的ReLU激活函数前，这种结构用于层数较少的神经网络中，比如ResNet34。若输入通道数比较多，就需要引入 $1\times 1$ 卷积层来调整输入的通道数，这种结构也叫作瓶颈模块，通常用于网络层数较多的结构中。

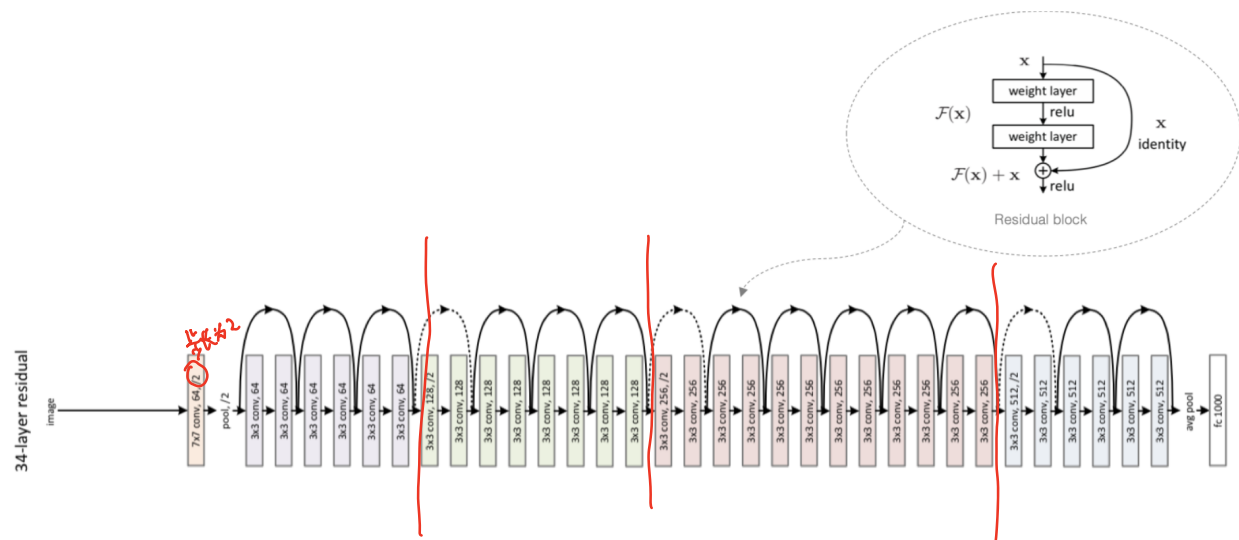




→ 目的是降维  
调整通道数。



ResNet模型构成图



图像增强