

第5章

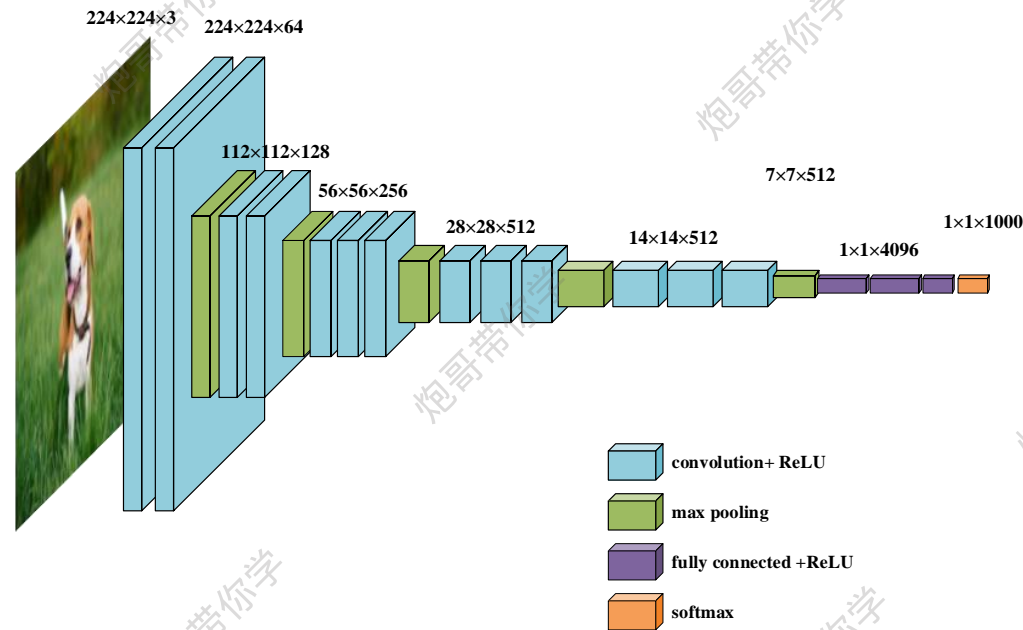
VGG原理

与实战

VGG网络诞生背景

VGGNet是牛津大学计算机视觉组 (Visual Geometry Group) 和谷歌 DeepMind 一起研究出来的深度卷积神经网络，因而冠名为 VGG。VGG是一种被广泛使用的卷积神经网络结构，其在2014年的 ImageNet 大规模视觉识别挑战(ILSVRC -2014)中获得了亚军，不是VGG不够强，而是对手太强，因为当年获得冠军的是GoogLeNet。

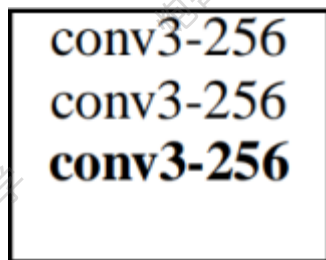
通常人们说的VGG是指VGG-16(13层卷积层+ 3层全连接层)。虽然其屈居亚军，但是由于其规律的设计、简洁可堆叠的卷积块，且在其他数据集上都有着很好的表现，从而被人们广泛使用，从这点上还是超过了GoogLeNet。VGG和之前的AlexNet相比，深度更深，参数更多(1.38亿)，效果和可移植性更好。



VGG网络结构

VGGNet有6种不同结构，我们以通常所说的VGG-16(即图D列)为例。

如图，可以发现，VGG中卷积层是通过block块状形式相连的,block内的卷积层结构相同；block外，block之间通过maxpool连接。图中VGG-16中的一个vgg-block块：



其中conv3-256表示：这是一个卷积层，卷积核尺寸为 3×3 ，通道数为256。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 2: Number of parameters (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

VGG网络结构

经典卷积神经网络的基本组成部分是下面的这个序列：

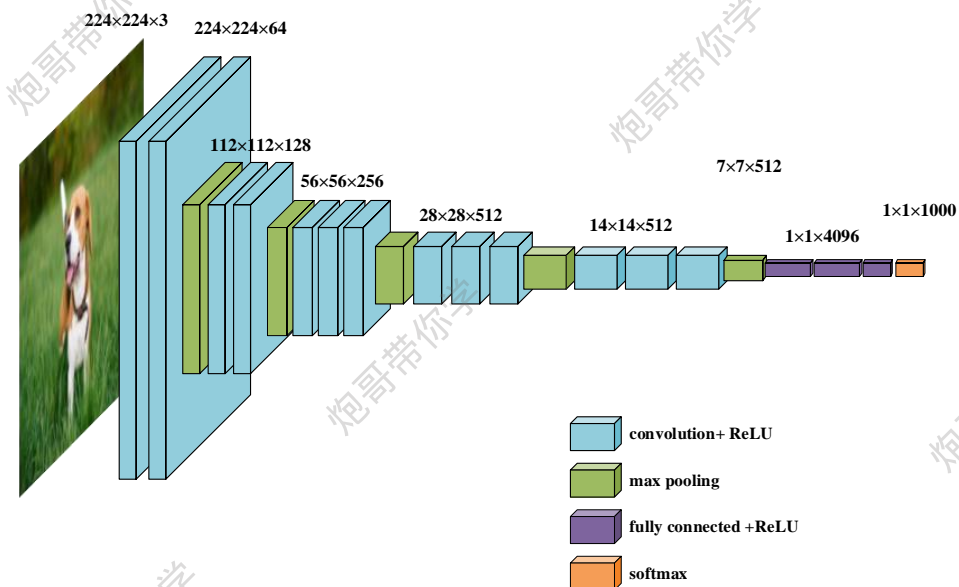
- (1)带填充以保持分辨率的卷积层；
- (2)非线性激活函数，如ReLU；
- (3)池化层，最大池化层。

而一个VGG块与之类似，由一系列卷积层组成，后面再加上用于空间下采样的最大池化层。

而一个VGG块与之类似，由一系列卷积层组成，后面再加上用于空间下采样的最大池化层。

VGG特点：

vgg-block内的卷积层都是同结构的
池化层都得上一层的卷积层特征缩减一半
深度较深，参数量够大
较小的filter size/kernel size



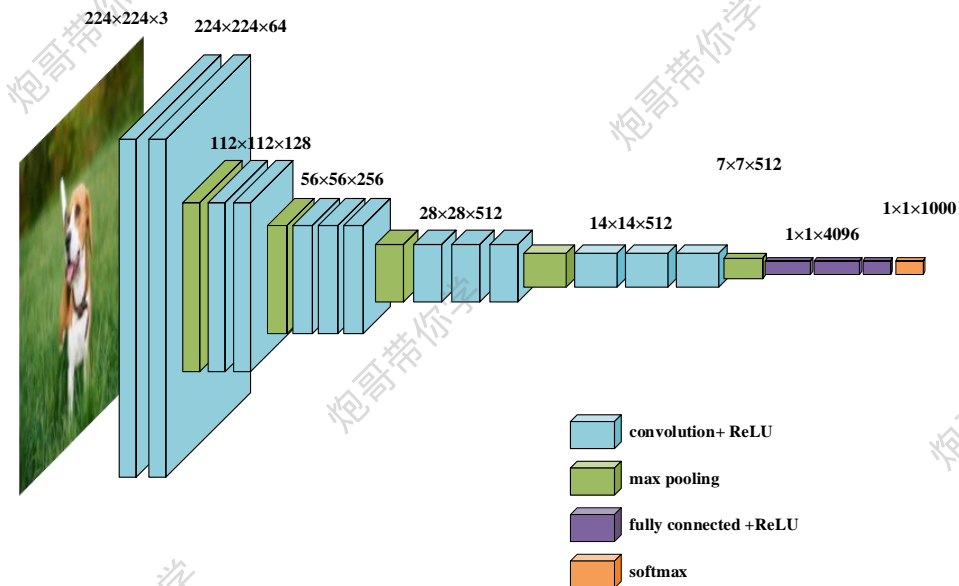
VGG网络结构

1、vgg-block内的卷积层都是同结构的意味着输入和输出的尺寸一样，且卷积层可以堆叠复用，其中的实现是通过统一的size为 3×3 的kernel size + stride1 + padding1实现。

2、maxpool层将前一层(vgg-block层)的特征缩减一半 使得尺寸缩减的很规整，从224-112-56-28-14-7。其中是通过pool size2 + stride2实现。

3、深度较深，参数量够大 较深的网络层数使得训练得到的模型分类效果优秀，但是较大的参数对训练和模型保存提出了更大的资源要求。

4、较小的filter size/kernel size **这里全局的kernel size都为 3×3 ，相比以前的网络模型来说，尺寸足够小。



VGG网络参数详解

第1个vgg block层:

(1) 输入为 $224 \times 224 \times 3$, 卷积核数量为64个; 卷积核的尺寸大小为 $3 \times 3 \times 3$; 步幅为1 (stride = 1), 填充为1 (padding=1); 卷积后得到shape为 $224 \times 224 \times 64$ 的特征图输出。

(2) 输入为 $224 \times 224 \times 64$, 卷积核数量为64个; 卷积核的尺寸大小为 $3 \times 3 \times 64$; 步幅为1 (stride = 1), 填充为1 (padding=1); 卷积后得到shape为 $224 \times 224 \times 64$ 的特征图输出。

(3) 输入为 $224 \times 224 \times 64$, 池化核为 2×2 , 步幅为2 (stride = 2) 后得到尺寸为 $112 \times 112 \times 64$ 的池化层的特征图输出。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGG网络参数详解

第2个vgg block层:

(1) 输入为 $112 \times 112 \times 64$ ，卷积核数量为128个；卷积核的尺寸大小为 $3 \times 3 \times 64$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $112 \times 112 \times 128$ 的特征图输出。

(2) 输入为 $112 \times 112 \times 128$ ，卷积核数量为128个；卷积核的尺寸大小为 $3 \times 3 \times 128$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $112 \times 112 \times 128$ 的特征图输出。

(3) 输入为 $112 \times 112 \times 128$ ，池化核为 2×2 ，步幅为2 (stride = 2) 后得到尺寸为 $56 \times 56 \times 128$ 的池化层的特征图输出。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGG网络参数详解

第3个vgg block层:

(1) 输入为 $56 \times 56 \times 128$ ，卷积核数量为256个；卷积核的尺寸大小为 $3 \times 3 \times 128$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $56 \times 56 \times 256$ 的特征图输出。

(2) 输入为 $56 \times 56 \times 256$ ，卷积核数量为256个；卷积核的尺寸大小为 $3 \times 3 \times 256$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $56 \times 56 \times 256$ 的特征图输出。

(3) 输入为 $56 \times 56 \times 256$ ，卷积核数量为256个；卷积核的尺寸大小为 $3 \times 3 \times 256$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $56 \times 56 \times 256$ 的特征图输出。

(4) 输入为 $56 \times 56 \times 256$ ，池化核为 2×2 ，步幅为2 (stride = 2) 后得到尺寸为 $28 \times 28 \times 256$ 的池化层的特征图输出。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGG网络参数详解

第4个vgg block层:

(1) 输入为 $28 \times 28 \times 256$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 256$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $28 \times 28 \times 512$ 的特征图输出。

(2) 输入为 $28 \times 28 \times 512$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 512$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $28 \times 28 \times 512$ 的特征图输出。

(3) 输入为 $28 \times 28 \times 512$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 512$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $28 \times 28 \times 512$ 的特征图输出。

(4) 输入为 $28 \times 28 \times 512$ ，池化核为 2×2 ，步幅为2 (stride = 2) 后得到尺寸为 $14 \times 14 \times 512$ 的池化层的特征图输出。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGG网络参数详解

第5个vgg block层:

(1) 输入为 $14 \times 14 \times 512$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 512$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $14 \times 14 \times 512$ 的特征图输出。

(2) 输入为 $14 \times 14 \times 512$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 512$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $14 \times 14 \times 512$ 的特征图输出。

(3) 输入为 $14 \times 14 \times 512$ ，卷积核数量为512个；卷积核的尺寸大小为 $3 \times 3 \times 512$ ；步幅为1 (stride = 1)，填充为1 (padding=1)；卷积后得到shape为 $14 \times 14 \times 512$ 的特征图输出。

(4) 输入为 $14 \times 14 \times 512$ ，池化核为 2×2 ，步幅为2 (stride = 2)后得到尺寸为 $7 \times 7 \times 512$ 的池化层的特征图输出。该层后面还隐藏了flatten操作，通过展平得到 $7 \times 7 \times 512 = 25088$ 个参数后与之后的全连接层相连。

第1~3层全连接层：第1~3层神经元个数分别为4096，4096，1000。其中前两层在使用relu后还使用了Dropout对神经元随机失活，最后一层全连接层用softmax输出1000个分类。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGGNet总结

VGGNet总结

VGGNet通过传统卷积神经网络模型(AlexNet)上的拓展,发现除了较为复杂的模型结构的设计(如GoogLeNet)外,深度对于提高模型准确率很重要,VGG和之前的AlexNet相比,深度更深,参数更多(1.38亿),效果和可移植性更好,且模型设计的简洁而规律,从而被广泛使用。还有一些特点总结如下:

- 1、小尺寸的filter(3×3)不仅使参数更少,效果也并不弱于大尺寸filter如 5×5
- 2、块的使用导致网络定义的非常简洁。使用块可以有效地设计复杂的网络。
- 3、AlexNet中的局部响应归一化作用不大

