VGG详解

VGG在2014年由牛津大学著名研究组VGG (Visual Geometry Group) 提出，斩获该年ImageNet竞赛中 Localization Task (定位任务) 第一名 和 Classification Task (分类任务) 第二名。

VERY DEEP CONVOLUTIONAL NETWORKS
FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan* **& Andrew Zisserman+**
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk

VGG详解

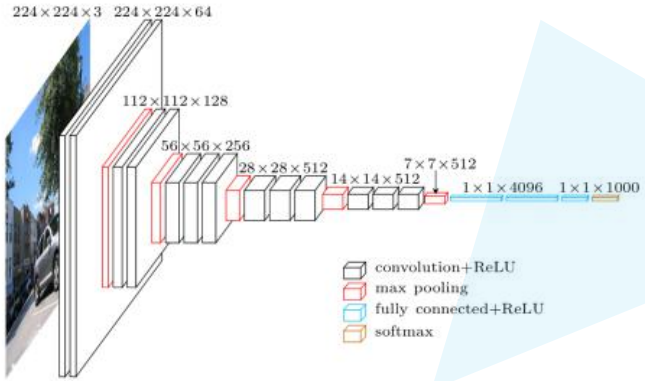
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128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

网络中的亮点：

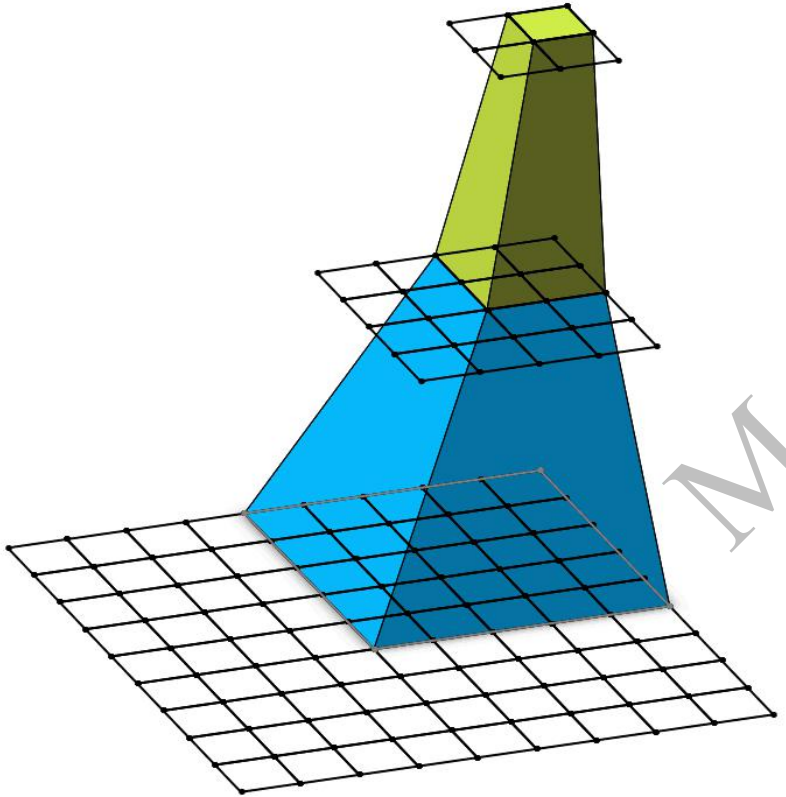
- 通过堆叠多个3x3的卷积核来替代大尺度卷积核（减少所需参数）

论文中提到，可以通过堆叠两个3x3的卷积核替代5x5的卷积核，堆叠三个3x3的卷积核替代7x7的卷积核。

拥有相同的感受野

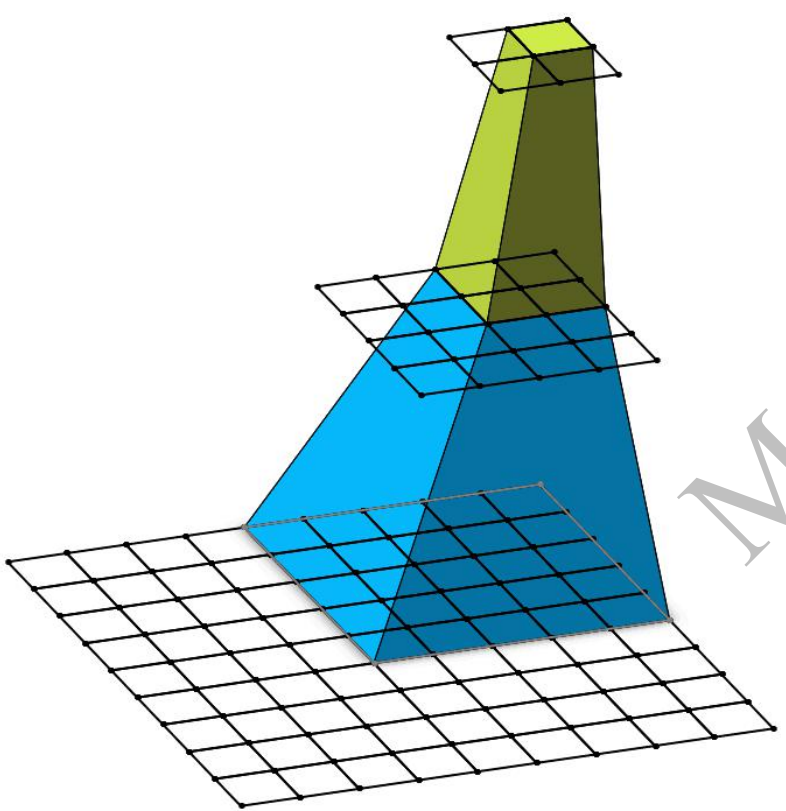


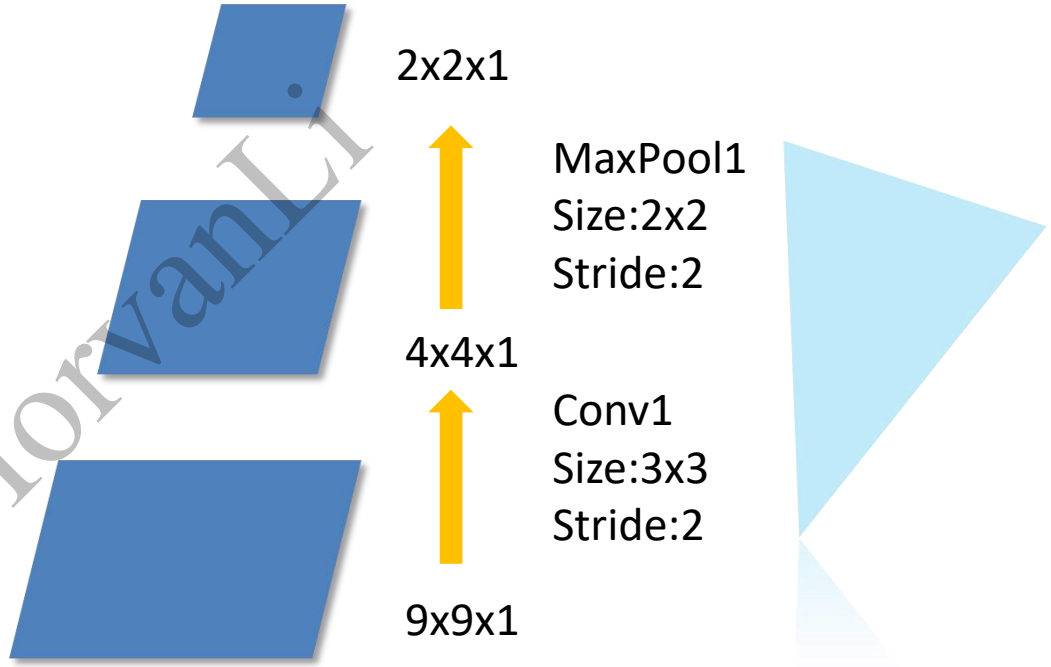
基本概念拓展—CNN感受野



在卷积神经网络中，决定某一层输出结果中一个元素所对应的输入层的区域大小，被称作**感受野**(receptive field)。通俗的解释是，输出feature map上的一个单元对应输入层上的区域大小。

基本概念拓展—CNN感受野



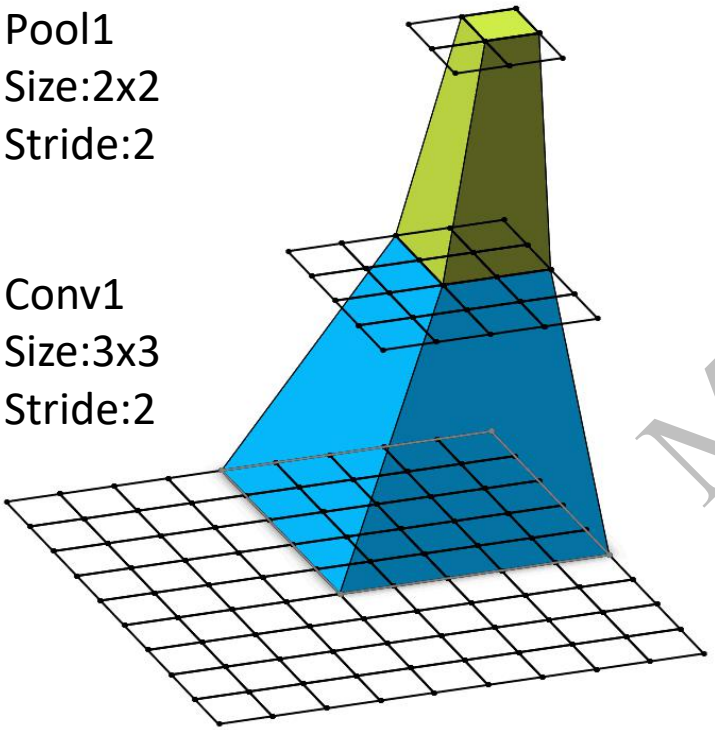


$$out_{size} = \left(in_{size} - F_{size} + 2P\right) / S + 1$$

基本概念拓展—CNN感受野

Pool1
Size:2x2
Stride:2

Conv1
Size:3x3
Stride:2



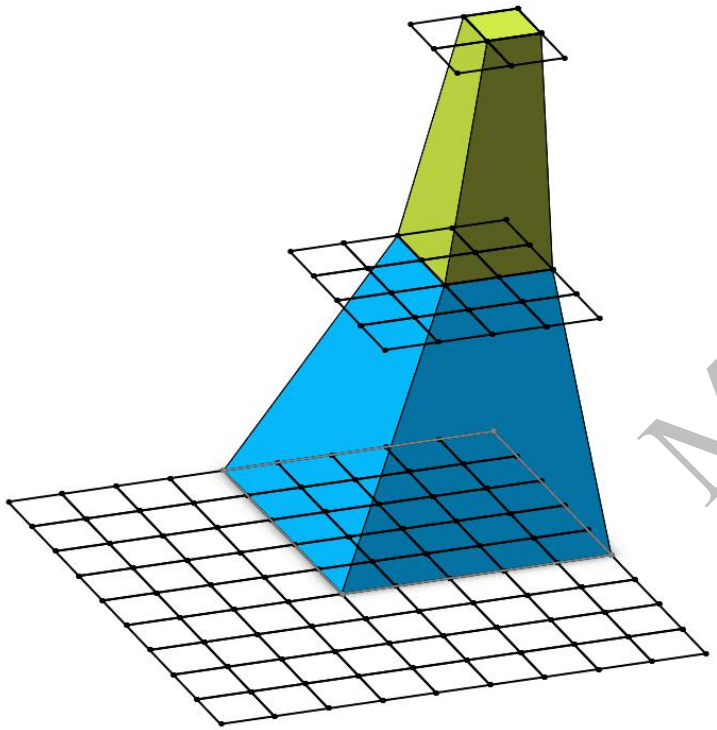
感受野计算公式：

$$F(i) = (F(i+1) - 1) \times Stride + Ksize$$

F(i)为第i层感受野，
Stride为第i层的步距，
Ksize为卷积核或池化核尺寸

Feature map: F = 1
Pool1: F = (1 - 1) x 2 + 2 = 2
Conv1: F = (2 - 1) x 2 + 3 = 5

基本概念拓展—CNN感受野



感受野计算公式：

$$F(i) = (F(i+1) - 1) \times Stride + Ksize$$

F(i)为第i层感受野，
Stride为第i层的步距，
Ksize为卷积核或采样核尺寸

Feature map: F = 1
Conv3x3(3): F = (1 - 1) x 1 + 3 = 3
Conv3x3(2): F = (3 - 1) x 1 + 3 = 5
Conv3x3(1): F = (5 - 1) x 1 + 3 = 7

VGG详解

论文中提到，可以通过**堆叠两个3x3的卷积核替代5x5的卷积核**，
堆叠三个3x3的卷积核替代7x7的卷积核。

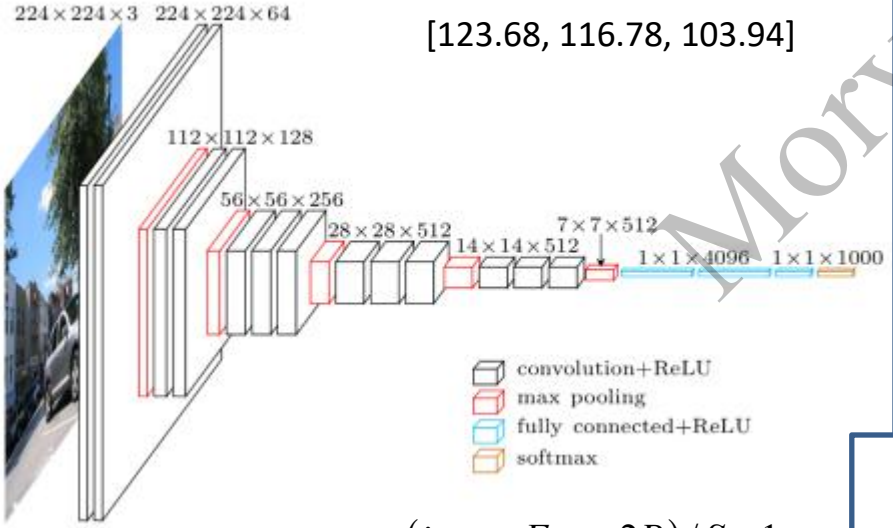
使用7x7卷积核所需参数，与堆叠三个3x3卷积核所需参数(假设输入输出channel为C)

$$7\times7\times C\times C=49C^2$$

$$3\times3\times C\times C+3\times3\times C\times C+3\times3\times C\times C=27C^2$$

VGG详解

- conv的stride为1，padding为1
- maxpool的size为2，stride为2



[123.68, 116.78, 103.94]

$$out_{size} = \left(in_{size} - F_{size} + 2P\right)/S + 1$$

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**
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**
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**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					