3.VGG

2014年由牛津大学研究组Visual Geometry Group提出，论文地址Very Deep Convolutional Networks for Large-Scale Image Recognition

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					

图3 VGG网络结构

文章亮点：通过堆叠多个3×3卷积核来代替大尺度卷积核（减少所需参数）。

论文中提到：堆叠2个3×3的卷积核代替5×5的卷积核，堆叠3个3×3的卷积核代替7×7的卷积核，它们拥有相同的感受野。

基本概念拓展：CNN感受野

在CNN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）。

$$F(i) = (F(i+1) - 1) \times S + ks$$

例如，$F = 1$，经过3个3×3的卷积核后感受野为7×7

$Conv3 \times 3(3) : F = (1-1) \times 1 + 3 = 3$
$Conv3 \times 3(2) : F = (3-1) \times 1 + 3 = 5$
$Conv3 \times 3(1) : F = (5-1) \times 1 + 3 = 7$

pytorch实现