

VGGNet:2014

Vert Deep CONvolutional Networks For Large-Scale Image Recognition

University of Oxford

摘要:

在本文中我们研究了卷积神经网络的深度在大型数据集上对其精度的影响，我们的主要贡献就是**使用非常小的3×3卷积核来增加网络的深度进而对网络的性能进行研究，基于此方法可以将网络的深度推广到16-19层**，对现有技术可以有显著改进。

论文出发点或背景:

在此之前，在ILSVRC-2013中表现最好的网络模型使用了更小接受窗口大小和更小的步长在第一个卷积层上，另外一个提升是通过在整张图片上多尺度密集地训练和测试网络。而本文的改进主要是基于神经网络的另外一个特点——深度。

论文创新思路:

使用非常小的3×3卷积核来增加网络的深度进而对网络的性能进行研究，基于此方法可以将网络的深度推广到16-19层

为了对比增加深度的公平性，我们所有的网络层配置都使用相同的设计原则

3×3的卷积核可以通过堆叠增加其感受野，两个3×3等同于一个5×5的感受野，三个等同于一个7×7卷积核的感受野。多个小卷积核的好处就是相对于一个大卷积核在其后面增加的多个非线性激活函数可以增强网络的非线性，同时也可以降低参数。

1×1卷积核的使用：增加非线性，1×1卷积核本质上是对输入做了一个线性变换，之后经过了非线性的激活函数，因而增加了网络的非线性。

论文方法的大概介绍:

我们调整了网络中的其他参数，并通过增加卷积层的个数进而增加网络的深度，这也被证明是可行的，因为我们在所有层上都使用的是3×3的小型卷积核。

为了对比增加深度的公平性，我们所有的网络层配置都使用相同的设计原则

预处理：[R,G,B]--->[R-mean_R,G-mean_G,B-mean_B]

所有的隐藏层都带有一个非线性激活函数ReLU

提到LRN并没有提高性能，反而导致了内存消耗和计算时间的增加。

权重随机初始化为零均值和具有 10^{-2} 方差的正态分布，偏置被初始化为0

使用了随机裁剪、随即水平翻转和随机RGB变化。

多尺度裁剪

密集评估：将全连接层改为卷积层

预训练：先训练一个浅层网络，然后用得到的权重去初始化更深层次的网络的层的参数

caffe，4块gpu上分布式训练

实际效果：

虽然之前有人使用过小型卷积核，但是他们的网络深度比我们低，还没有在大型数据集上测试过

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

capturing multi-scale image statistics.

Table 3: **ConvNet performance at a single test scale.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
A	256	256	29.6	10.4
A-LRN	256	256	29.7	10.5
B	256	256	28.7	9.9
C	256	256	28.1	9.4
	384	384	28.1	9.3
	[256;512]	384	27.3	8.8
D	256	256	27.0	8.8
	384	384	26.8	8.7
	[256;512]	384	25.6	8.1
E	256	256	27.3	9.0
	384	384	26.9	8.7
	[256;512]	384	25.5	8.0

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	24.8	7.5
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	24.8	7.5

Table 5: **ConvNet evaluation techniques comparison.** In all experiments the training scale S was sampled from [256; 512], and three test scales Q were considered: {256, 384, 512}.

ConvNet config. (Table 1)	Evaluation method	top-1 val. error (%)	top-5 val. error (%)
D	dense	24.8	7.5
	multi-crop	24.6	7.5
	multi-crop & dense	24.4	7.2
E	dense	24.8	7.5
	multi-crop	24.6	7.4
	multi-crop & dense	24.4	7.1

Table 6: **Multiple ConvNet fusion results.**

Combined ConvNet models	Error		
	top-1 val	top-5 val	top-5 test
ILSVRC submission			
(D/256/224,256,288), (D/384/352,384,416), (D/[256;512]/256,384,512) (C/256/224,256,288), (C/384/352,384,416) (E/256/224,256,288), (E/384/352,384,416)	24.7	7.5	7.3
post-submission			
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), dense eval.	24.0	7.1	7.0
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop	23.9	7.2	-
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop & dense eval.	23.7	6.8	6.8

Table 7: **Comparison with the state of the art in ILSVRC classification.** Our method is denoted as “VGG”. Only the results obtained without outside training data are reported.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	23.7	6.8	6.8
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-	7.9	-
GoogLeNet (Szegedy et al., 2014) (7 nets)	-	6.7	-
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	27.9	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-

个人理解：

- 1.实验证明LRN对模型的准确率提升没用
- 2.网络的初始化十分重要，可以通过预训练方式来防止深层的网络训练失败、
- 3.网络的深度十分重要，卷积神经网络中，随着特征图空间上信息的减少，通道数会相应增加，将低层次的空间信息转换为了抽象的语义信息

4.对比之前使用的大卷积核和大步长，小卷积核和小步长更能减少空间信息的损失，同时小卷积模块的堆叠可以等效大卷积模块的感受野