VggNet

摘要 本文研究了在大规模图片识别中，卷积神经网络的深度对准确率（accuracy）的影响。我们的主要贡献是通过非常小的 3x3 卷积核的神经网络架构全面评估了增加深度对网络的影响，结果表明 16-19 层的网络可以使现有设置的网络性能得到显著提高。这项发现是我们在 2014 年的 ImageNet 比赛中提交方案的基础，我们的团队分别在定位和分类中获得了第一和第二的成绩。我们还证明了此模型可以泛化到其他数据集上，并达到当前最佳水平。我们已经公布了两个性能最佳的卷积神经网络模型，以便深度视觉在计算机视觉中的进一步研究。

1 介绍

卷积神经网络最近在大规模图片和视频识别中取得了重大成功，这可能得益于大型开源图片库，比如 ImageNet，以及高性能计算系统，如 GPU 或大规模分布式集群。特别是 ImageNet 大规模视觉识别挑战（ILSVRC），对深度视觉识别架构的发展起到了重要作用，它为几代大规模图片识别系统——从高维浅层特征编码（ILSVRC-2011 的获胜者）到深层卷积神经网络（ILSVRC-2012 的获胜者）——提供了测试平台。

随着卷积神经网络在计算机视觉领域的应用越来越广，越来越多的人尝试改进 Krizhevsky 等人在 2012 年提出的原始架构，以得到更好的准确率。例如，在 2013 年 ImageNet 大赛中性能最好的改进方案——在第一个卷积层中使用较小的接受域窗口以及较小的步长，另一种改进方案是在整幅图片及多个尺寸上多次训练和测试网络（Sermanet et al.2014；Howard, 2014）。在本文中，我们着眼于卷积神经网络中的另一个方面——深度。为此，我们固定了架构中的其他参数，并通过添加卷积层稳定地增加网络深度。这是可行的，因为我们在每层都使用非常小的 3x3 卷积核。

因此，我们提出了更精确的卷积神经网络架构，不仅在 ILSVRC 分类和定位中取得最好成绩，还在其他图片识别数据集中取得卓越性能，即便只作为简单框架的一部分（如不需要微调的线性 SVM 深度特征分类器）。我们公布了两个最佳性能模型，以便进一步研究。

本文组织结构如下。在第二部分，描述了卷积神经网络的设置。图片分类的训练及评估细节在第三部分中阐述。在 ILSVRC 分类任务中不同设置的比较在第四部分中阐述。在第五部分，总结本文内容。为了完整性，我们还在附录 A 中描述评估了我们在 ILSVRC-2014 中的物体定位系统，并在附录 B 讨论了深度特征在其他数据集上的泛化。最后，在附录 C 中列出了本文的主要修订记录。

2 卷积神经网络的设置

为了公平衡量增加卷积深度对网络的影响，我们所有卷积层的设置均使用与 Ciresan（2011）和 Krizhevsky（2012）相同的设计原则。在这一部分，我们首先描述了卷积神经网络的通用结构，然后详细介绍了评估中具体配置细节。最后描述了我们的模型与先前最好网络的比较。

2.1 架构

在整个训练中，卷积神经网络的输入为固定的 224x224 的 RGB 图片。唯一的预处理是对每个像素减去 ImageNet 训练集中 RGB 的平均值。图片通过一系列 3x3 卷积核（是用来获取上下左右及中心的最小尺寸）的卷积层。在一种配置中，也使用 1x1 的卷积核，这可以看做是输入通道的线性变换（后面接一个非线性变换）。卷积滑动步长固定为 1；卷积层的空间填充（padding）模式为保留原空间分辨率，例如 3x3 的卷积层，padding 为 1。空间池化（pooling）包含 5 个最大池化层，接在部分卷积层后面（不是所有卷积层）。最大池化层使用 2x2 的窗口，滑动步长为 2。

在一系列卷积层（不同架构有不同深度）后为 3 个全连接层（Fully-Connected）：前两个每个含有 4096 个通道，第三个用来给 ILSVRC 进行分类，因此有 1000 个通道（1000 个类）。最后一层使用 softmax。全连接层的设置与所有网络一致。

所有隐藏层都使用 ReLU 非线性激活函数。注意到我们的网络(除了一个)都**不包含**局部响应标准化(LRN)：在第四部分 中会展示，这个标准化并不会提高网络在 ILSVRC 数据集上的性能，反而会增加内存消耗和计算时间。在使用的情况下，LRN 层的参数是（Krizhevsky et al. 2012）的参数。

2.2 设置

本文所评估的卷积神经网络的设置在表 1 列出，每列一个。接下来我们称他们为（A-E）。所有配置都遵循 2.1 所述的通用设计，只有深度不同：从网络 A 的 11 层（8 个卷积层 3 个全连接层）到网络 E 的 19 层（16 个卷积层 3 个全连接层）卷积层的宽度（通道数）非常小，从第一层的 64 开始，每个最大池化层后增加 1 倍，直到 512。

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					

表 1：网络设置（按列显示）。网络的深度从 A 到 E 依次增加，增加的层用加粗显示。卷积层的参数用“conv<接受域大小>-<通道数>”来表示，为了简洁，Relu 激活函数没有显示

表 2 给出了每个设置的参数数目。尽管网络很深，但是网络的权重数目并没有一个更浅但是卷积层更宽和接受域更大的网络权重数目大（sermanet et al., 2014 有 144M 的权重）。

Table 2: **Number of parameters** (in millions).

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

表 2：参数数量（百万）

2.3 讨论

本文网络的设置与 ILSVRC-2012 好 ILSVRC-2013 大赛中的前几名完全不同。没有在第一个卷积层使用大的接受域（如 11x11 的卷积核，滑动步长为 4（Krizhevsky et al. 2012），或者 7x7 的卷积核，滑动步长为 2（Zeiler&Fergus，2013；Sermanet et al. 2014）），我们在整个网络使用 3x3 的卷积核，与每个像素值进行卷积（步

长为 1）。很明显，两个 3x3 卷积层（中间没有池化层）相当于 5x5 的接受域；三个这样的层相当于 7x7 的接受域。那么用三个 3x3 的卷积层代替一个 7x7 的卷积层有什么好处呢？首先，我们包含三个非线性修正层而非单一层，这使决策函数更具有区分性。其次，我们减少了参数数量：假设一个含有三层 3x3 卷积层堆叠的输入和输出都包含 C 个通道的网络，权重数量为 $3(3^2C^2)=27C^2$; 而一个 7x7 的卷积层，需要 $7^2C^2=49C^2$ 个权重参数，相对增加了 81%，这相当于在 7x7 的滤波器上加了一个正则化，迫使它们通过 3x3 的滤波器进行分解（中间有非线性的加入）。

1x1 卷积层的加入（表 1 中的 C）是一种为决策增加非线性因素的方式，不影响卷积层接受域。尽管在这里，1x1 的卷积实质上是相同空间维度的线性投影（输入和输出通道相同），但是修正函数引入了非线性因素。值得注意的是 1x1 卷积层最近被 Lin 等人（2014）用在“Network in Network”结构中。

小尺寸的卷积滤波器之前被 Ciresan（2011）等人用过，但是他们的网络深度远小于我们，并且他们没有在大规模 ILSVRC 数据集上做评估。Goodfellow 等人（2014）在识别街景数字的任务中使用了深度卷积神经网络（11 层），展示了增加深度带来的优越性能。GoogLeNet（2014），在 ILSVRC-2014 的识别任务中获得了最佳表现，虽然与我们的网络不同，但是相似的是都基于很深的卷积网络（22 层）以及很小的卷积滤波器（除了 3x3，他们还使用了 1x1 和 5x5 的滤波器）。但是他们的网络拓扑比我们的更复杂，而且为了减少计算量，特征图的空间分辨率在第一层衰减的很严重。在第 4.5 部分将展示我们的模型在单一网络分类中准确率优于 GoogLeNet。

3 分类框架

前面的部分我们介绍了网络设置的细节。这一部分，我们将详细描述分类卷积神经网络的训练与评估。

3.1 训练

卷积神经网络的训练过程与 Krizhevsky 等人（2012）（除了多尺寸训练数据样本的裁剪，后面会介绍）的一样。就是说，通过用包含动量的小批量梯度下降（基于反向传播）做多项式逻辑回归的优化器来对模型进行训练。批次大小为 256，动量为 0.9，通过权值衰减（L_2 惩罚因子设置为 $5*10^{-4}$）和对前两个全连接层进行 dropout（比率 0.5）实现正则化。学习率初始化为 0.01，当验证集准确率不提升时以 10 倍速率衰减（除以 10）。总的来说，学习率会衰减 3 次，然后训练次数为 370K（74 代）。我们猜想，尽管与 Krizhevsky（2012）等人的网络相比，我们的网络参数更多，深度更深，但是却需要更少的 epoch 次数来收敛，因为（1）深度及更小的滤波器数量隐式增强了正则化；（2）某些层执行了预初始化。

网络权重的初始化很重要，由于深度网络梯度下降的不稳定性，不好的初始化会阻碍学习。为了规避这个问题，我们从训练网络 A（表 1）开始，它足够浅，能

用随机初始化。然后，当训练更深网络结构时，我们用网络 A 的权重初始化前四个卷积层和后三个全连接层（中间层随机）。对预初始化层，不降低学习率，允许他们在学习过程中改变。对于随机初始化，我们从 0 均值和 0.01 方差的正态分布中取值。偏差初始化为 0。值得注意的是，我们发现可以用 Glorot&Bengio(2010)中的随机初始化程序来对权重进行初始化，而不需要进行预训练。

为了得到固定的 224x224 的 RGB 输入图片，我们随机从经过尺寸缩放的训练集图片中进行裁剪（每张图的每次 SGD 迭代时裁剪一次）。为了进一步对训练集数据进行增强，被裁剪图片将进行随机水平翻转及 RGB 颜色转换。训练图片的尺寸缩放将在后面阐释。

训练集图片尺寸 令 S 为各向同性缩放的训练图像最小边，卷积神经网络的输入就是从中裁剪的（S 也称为训练尺寸）。裁剪尺寸固定为 224x224，原则上 S 可以取任何大于等于 224 的值：若 S=224，裁剪图像将使用整个图像的统计信息，完全涵盖训练图像的最小边；若 S>>224，裁剪图像就会取图像的一小部分，包含一个很小的对象或对象的一部分。

我们考虑使用两种方式来设置训练尺寸 S。第一种是固定 S，针对单尺寸图片的训练。（注意，裁剪的样本图像内容仍然能够代表多尺寸图片的统计信息）在实验中，评估了两种固定尺寸的训练模型：S=256（在之前研究中广泛使用）和 S=384。给一个卷积神经网络，首先用 S=256 训练。为了加速 S=384 的训练，使用在 S=256 上的预训练权重来初始化权重，并且使用较小的初始学习率 0.001。

第二种设置 S 的方式是使用多尺寸图像训练，即每个训练图片的尺寸是$[S_{min}, S_{max}]$之间的随机数（这里使用 S_{min}=256,S_{max}=512）。由于图像中的对象可能大小不一，所以训练中采用这种方式是有利的。这可以看作是一种尺寸不定(scale jittering)的训练集数据增强，使得一个单一模型能够识别各种尺寸的对象。考虑到速度，我们使用与微调后的 S=384 的单一尺寸预训练模型相同设置的模型，来训练多尺寸模型。

3.2 测试

在测试时，给定一个训练后的卷积神经网络及一张输入图片，用以下方式进行分类。首先，各向同性缩放成预定义的最小边，设为 Q（也称为测试尺寸，注意 Q 不需要等于训练尺寸 S（将在第 4 部分解释），每个 S 使用多个 Q 可以提高性能）。然后，根据 Sermanet 的方法将网络密集应用在测试图片上，也就是说，全连接层先转化为卷积层（第一个全连接层转为 7x7 的卷积层，后两个转化为 1x1 的卷积层）。再将这样得到的全卷积网络运用在整幅图像上(未裁切的)。输出是一个分类得分图，通道数与类别数先沟通呢个，空间分辨率依赖于输入图片的尺寸。最后，为了得到固定尺寸的分类得分向量，将分类得分图进行空间平均化（求和——池化）。我们同样使用水平翻转对测试图像进行增强；在原始图像和翻转图像上的 soft-max 分类概率的平均值作为这幅图像的最终得分。

由于测试阶段将全卷积网络用在了整个图像，因此不需要对图像进行多个裁切采样（Krizhevsky2012），因为网络对每个裁切的重新计算会使效率降低。但是，使用大量裁切图像可以提高准确率，如同 Szegedy 等人的网络，因为和全卷积网络相比，它能生成关于输入图像更好的采样。同样的，由于不同的卷积边界条件，多重裁切评估与密集评估是互补的：对一个裁剪图片使用卷积网络，卷积得到的特征图被 0 填充，而密度评估中，相同裁切图的填充自然而然来自于图片的相邻像素（由于卷积和空间池化），大大增加了网络整体的接受域，所以更多上下午信息被获取。尽管我们认为在实践中多尺寸裁切图像增加的计算时间并不能证明其具有更高准确率的潜质，但是为了参考，我们依然在评估时对每个尺寸使用了 50 张裁切图像(5×5 个规则网格以及水平翻转)，3 种尺寸一共 150 张裁切图像，这和 Szegedy 等人的网络中使用 4 种尺寸一共 144 张裁切图像是可比的。

3.3 实现细节

我们的实现使用开源的 C++ Caffe 工具箱（Jia，2013）（2013 年 12 月的分支），但是进行了一些重新修改，允许我们用同一个系统的多个 GPU 训练和评估模型，以及对全尺寸（未裁剪）图片的多种缩放（上文提到的）进行训练评估。GPU 批量梯度下降计算完成后，取平均数作为所有批次的梯度。梯度计算在多个 GPU 间是并行计算的，所以结果与在单个 GPU 上训练是一样的。

虽然最近提出了更复杂的加速卷积网络训练的方法（Krizhevsky2014），它在网络不同层上用模型和数据并行计算，但是我们发现我们的方法更简单，且在 4 个 GPU 系统上的速度相对于单 GPU 提升了 3.75 倍，在 NVIDIA Titan Black GPU 上，训练单个网络需要 2~3 周的时间。

4 分类实验

数据集 在本章，我们讲述了卷积神经网络在 ILSVRC2012 数据集上的分类结果（被用在 ILSVRC2012——2014 挑战赛上）。数据集包含 1000 个类别，被分为三部分：训练集（1.3M 张图片），验证集（50K 张图片），测试集（100K 张图片，没有标签）。分类性能使用两个办法评估：top-1 和 top-5 error。前者是一个多类分类错误率，即错误分类图像的比例；后者是在 ILSVRC 上的主要评估标准，即真实类别不在 top-5 预测类别之中的图像的比例。

对于大部分实验，我们使用验证集作为测试集。某些实验也在测试集上进行，并提交给官方 ILSVRC 服务器作为“VGG”团队参加 ILSVRC-2014 竞赛。

4.1 单一尺寸测试数据评估

我们从评估在单一尺度上使用第 2.2 中配置的独立卷积网络模型的性能开始。测试集图片大小如下设置：对于固定的 S，Q=S，对于变动的 $S\in[S_{min}, S_{max>}]$，$Q=0.5(S_{min} + S_{max>})$。结果如表 3 中。

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**

表 3 单一尺寸上的卷积网络性能

首先，注意使用局部相应标准化网络（A-LRN）的性能并没有比未用标准化层的 A 高。因此我们没有在更深的网络结构上使用标准化操作（B-E）。

其次，我们发现分类的错误率随着卷积层的增加而减少：从 11 层的 A 到 19 层的 E。注意，尽管深度相同，配置 C（包含 3 个 1x1 卷积层）没有配置 D（使用 3x3 卷积层）性能好，这意味着添加非线性层的确有用（C 比 B 好），但是使用卷积获取空间上下文信息更有用（D 比 C 好）。当深度达到 19 层时，错误率达到饱和，但是更大的数据集使用更深的模型会更好。我们也用网络 B 与一个 5x5 的浅卷积网络（派生自 B 但是将 3x3 卷积层换成了一个 5x5 卷积层，与 2.3 种所述接受域相同）进行了比较，浅层网络的 top-1 错误率比 B（在中心裁剪图像上）高了 7%，证明了小滤波器的神剧卷积网络比大滤波器的浅层网络性能更好。

最后，训练时尺寸变化(S ∈ [256;512]) 的性能比固定最小边(S = 256 or S = 384)的性能要好，尽管测试时使用的是单一尺寸。这证明训练集通过变化尺寸来进行数据增强的确能获取更多尺寸的图片统计信息。

4.2 多尺寸测试数据评估

评估了卷积网络模型在单一尺度上的性能之后，我们现在来评估在测试阶段使用尺寸抖动的效果。先在多个尺寸的测试数据上运行模型（多个 Q 值），然后计算每个类概率的平均值。考虑到训练尺寸与测试尺寸的差异太大会导致性能下降，模型使用固定的 S 训练，通过 3 个接近训练集的测试集尺寸评估，：Q={S-32，S， S+32}。同时，训练时的尺寸波动使测试时能使用更大范围尺寸的图像，所以使用 S ∈ [S_{min};S_{max}]训练的模型用更大范围的 Q 来评估，Q={S_{min}, 0.5(S_{min}, S_{max}), S_{max}}。

结果如表 4，表明在测试时图片尺寸波动会使性能更好（对比表 3 中单一尺寸的结果）。与之前相同，最深的配置（D 和 E）表现的最好，并且训练时尺度波动比固定最小边 S 表现更好。我们在验证集上最好的单一网络模型错误率为 24.8%

（top-1）7.5%（top5），在表 4 种加粗。在测试集上，配置 E 达到了 7.3%的 top-5 错误率。

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**

表 4 多尺寸测试集上的卷积网络表现

4.3 多裁剪评估

Table5 中我们对密集卷积网络评估和多重裁切评估进行了比较(见 Sect 3.2)。我们同样还评估了两种技术通过计算两者 soft-max 输出平均值的互补结果。可以看出，使用多重裁切比密集评估的效果略好，并且两种方法是完全互补的，因为两者组合的效果比每一种都要好。根据以上结果，我们假设这是由对于卷积边界条件的不同处理方法造成的。

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**

表 5 卷积网络评估技术比较。所有实验中，S 来源于[256,512]，三个测试尺寸 Q 为{256， 384， 512}

4.4 卷积网络融合

到目前为止，我们评估了独立卷积网络模型的性能。这一部分的实验，我们将通过计算多个模型 soft-max 分类概率的平均值来对它们的输出进行组合。由于模型的互补性，性能得到了改善，这也用在 2012 (Krizhevsky et al., 2012) 和 2013(Zeiler & Fergus, 2013; Sermanet et al., 2014)的 ILSVRC 的最佳结果中。

结果如表 6。在 ILSVRC 比赛中我们进训练了单一尺寸网络和多尺寸网络 D（仅仅微调了全连接层而非所有层）。7 个模型组合结果在 ILSVRC 中测试的错误率为 7.3%。提交后，我们考虑禁用两个最好表现的多尺寸模型（D 和 E）进行组合，

使用密集评估时错误率减少到 7.0%，使用密集和多裁剪评估时错误率为 6.8%。作为参考，我们的最佳单一模型错误率为 7.1%（E，表 5）。

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**

表 6 多卷积网络融合结果

4.5 与业界最好结果的比较

最后，我们在表 7 与业界最好结果进行了比较。在 2014 年的 ILSVRC 比赛的分类任务中，我们的 VGG 团队取得了第二名的成绩，使用了 7 个模型组合的测试错误率，为 7.3%，提交后，使用 2 个模型的组合，将错误率降低到了 6.8%。

从表 7 可以看出，我们的深度卷积神经网络比在 ILSVRC-2012 和 ILSVRC-2013 中成绩最好的模型效果明显要好。我们的结果与分类任务的冠军旗鼓相当(GoogLeNet 为 6.7%的错误率)，并且明显比 ILSVRC-2013 的冠军 Clarifai 的表现好得多，它使用外部训练数据时的错误率为 11.2%，而不使用外部数据时为 11.7%。更标志性的是，我们最佳的结果是通过对两个模型的组合——这明显比大多数 ILSVRC 参赛模型要少。在单一网络性能上，我们的模型取得了最好的结果(7.0%的测试错误率)，比单一的 GoogLeNet 低 0.9%。值得注意的是，我们并没有摒弃经典的卷积网络框架，并通过显著增加深度对它的性能进行了提升。

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	-

表 7 与 ILSVRC 分类任务中的最佳结果比较。我们的方法叫 VGG，只显示了未使用外部训练数据的结果

5 结论

本文评估了深度卷积网络（到 19 层）在大规模图片分类中的应用。结果表明，深度有益于提高分类的正确率，通过在传统的卷积网络框架中使用更深的层能够在 ImageNet 数据集上取得优异的结果。附录中，展示了我们的模型可以很好的泛化到更多数据集种，性能达到甚至超过了围绕较浅深度的图像表达建立的更复杂的识别流程。我们的实验结果再次确认了深度在视觉表达中的重要性。