**VGG阅读总结**

15331416 赵寒旭

## 目录

**引言 2**

**VGG的特点 2**

**论文详细阅读 2**

Abstract 2

1. Introduction 3

2. ConvNet Configurations（卷积网络配置） 3

2.1 Architecture 3

2.2 Configurations 4

2.3 Discussion 4

3. Classification Framework（分类框架） 5

3.1 Training 6

3.2 Testing 7

3.3 Implementation Details 8

4. Classification Experiments（分类实验） 9

4.1 Single Scale Evaluation 9

4.2 Multi-scale Evaluation 9

4.3 Multi-crop Evaluation 10

4.4 ConvNet Fusion（融合） 10

4.5 Comparison with the State of the Art 11

5. Conclusion（总结） 11

**引言**

VGGNet是牛津大学计算机视觉组(Visual Geometry Group)和Google Deepmind公司研究员一起研发的深度卷积神经网络。VGGNet在AlexNet的基础上探索了卷积神经网络的深度与性能之间的关系，通过反复堆叠3\*3的小型 卷积核和2\*2的最大池化层，VGGNet构筑的16~19层卷积神经网络模型取得了很好的识别性能，同时VGGNet的拓展性很强，迁移到其他图片数据上泛化能力很好，而且VGGNet结构简洁，现在依然被用来提取图像特征。

/\*

注：部分理解内容来源于网上的blog，

\*/

**VGG的特点**

**1. 针对网络架构**

1）全部使用的卷积核和的池化核，通过不断加深网络结构来提升性能。

2）使用多个小卷积核串联组成卷积层，和先前的大卷积核相比，拥有同样的感受野，却有着更少的参数，更强的非线性变换，因此有着更强的特征提取能力。

3）使用卷积层，意义在于线性变换，输入通道和输出通道数不变，没有发生降维。

4）LRN（局部响应归一化）层作用不大（并不能提升性能，且会增加计算量），仅在一个网络中使用。

**2. 针对过拟合现象**

1）数据增强

使用Multi-scale的方法，将原始图像缩放到不同尺寸，再随机裁取固定大小的图片，这样可以增加很多数据量，防止模型过拟合。

2）采用multi-crop和dense evaluation配合（两者配合使用，效果比单使用好），可以很好的提升模型的性能。

**3. 针对训练速度**

可以先训练低深度的A网络，再复用A网络的权重初始化后面的几个复杂模型，这样训练收敛的速度更快

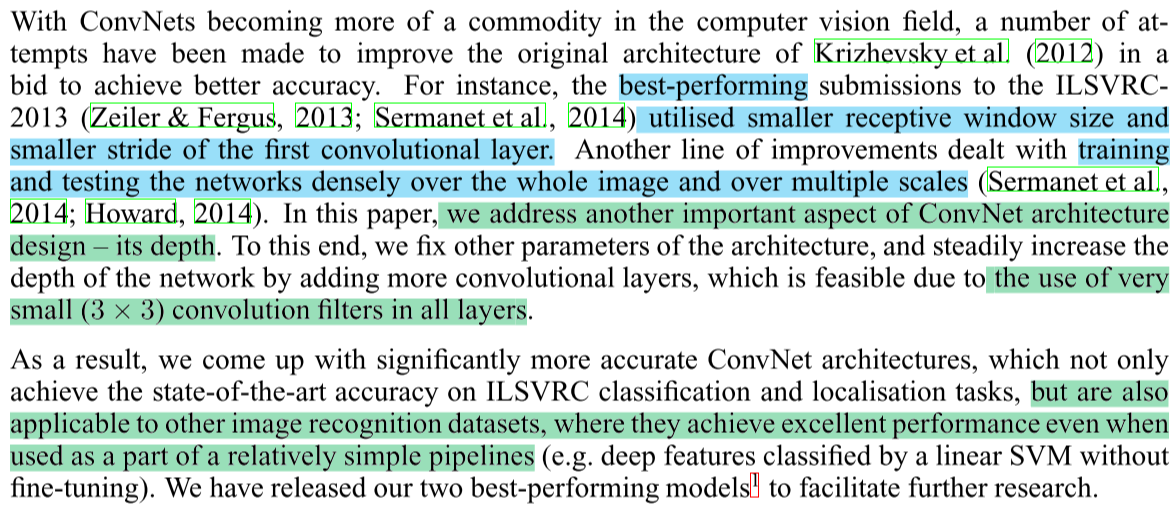
**论文详细阅读**

**Abstract**

研究了卷积网络深度对大型图像识别准确性的影响，主要评估小的卷积核（）同等架构下随着网络深度的增加卷积网络的性能变化，随着网络深度达到16-19层，网络的性能也有着显著的提升。

主要的贡献：展示出网络的深度是算法优良性能的关键部分(使用了非常小的滤波器（）。

**1.** **Introduction**



为在AlexNet原始架构上实现更高的准确率，可以在第一个卷积层上使用更小的感受野和更短的步长，另一种路线是使用多尺度的密集训练和测试网络（后文讨论了Multi-crop和Multi-scale的使用对模型性能的影响）。

本文主要研究卷积神经网络架构设计的另一个重要影响因素：深度。

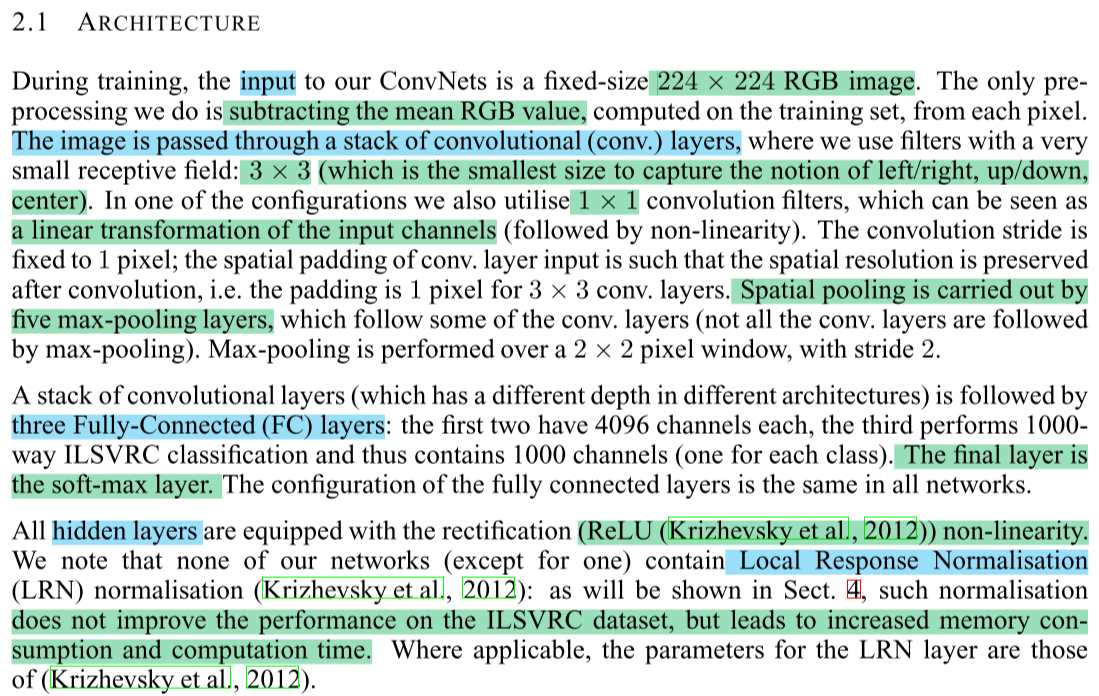
结果：实现了在ILSVRC上更精确的卷积神经网络，在其他数据集上也有很好的表现，迁移性很强。

**2. ConvNet** **Configurations（卷积网络配置）**

为在公平的环境下测量网络深度增加带来的提升，我们所有的卷积网络层结构使用相同的原则来设计。

本节介绍网络配置的细节。

**2.1 Architecture**



1）input：a fixed-size 224 × 224 RGB image

pre-processing：从每个像素减去RGB值的均值

2）The image is passed through a stack of convolutional layers:

使用非常小的的滤波器（是足够捕获一个区域信息的最小的卷积核）

其中一个配置中使用的卷积滤波器（对输入通道的线性变换），这样的卷积不改变输入通道的维度，且可以提高模型的学习能力。

convolution stride：1 pixel

spatial padding：对卷积层输入的扩展要保证卷积操作前后分辨率一致（eg. 的卷积层要扩展1 pixel）

Spatial pooling：由5个最大池化层实现（跟在某些卷积层后面）

Max-pooling：在的窗口进行，步长为2

3）3 Full-Connect layers

4096,4096,1000

最后一层是softmax层。

5）Hidden layers

整流函数：ReLU

6）Local Response Normalisation

除了一个网络之外，我们所有的网络都不含局部响应归一化，它不能提升在ILSVRC数据集上的性能，但会导致内存消耗和计算时间的增加。

7）LRN layer parameters

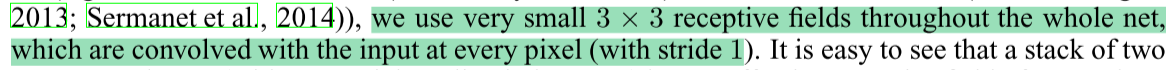
Where applicable, the parameters for the LRN layer are those of (Krizhevsky et al., 2012).

**2.2 Configurations**

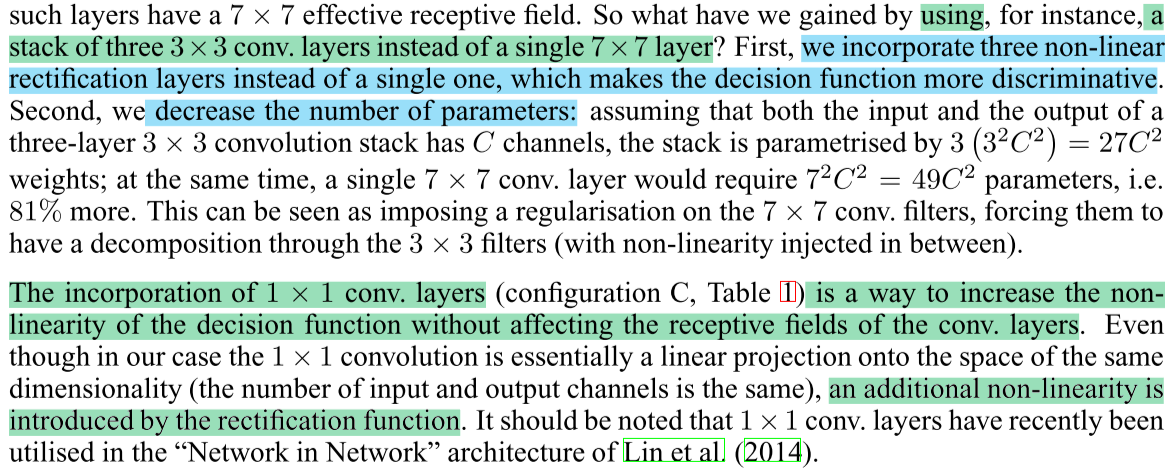
卷积核尺寸很小，卷积核的数目初始为64，每过一次max-pooling层，卷积核数目翻倍，最终增加到512个。

尽管网络深度很大，网络中的权重数目并不会比有更大卷积层和感受野的更浅的网络多。

**2.3 Discussion**



------

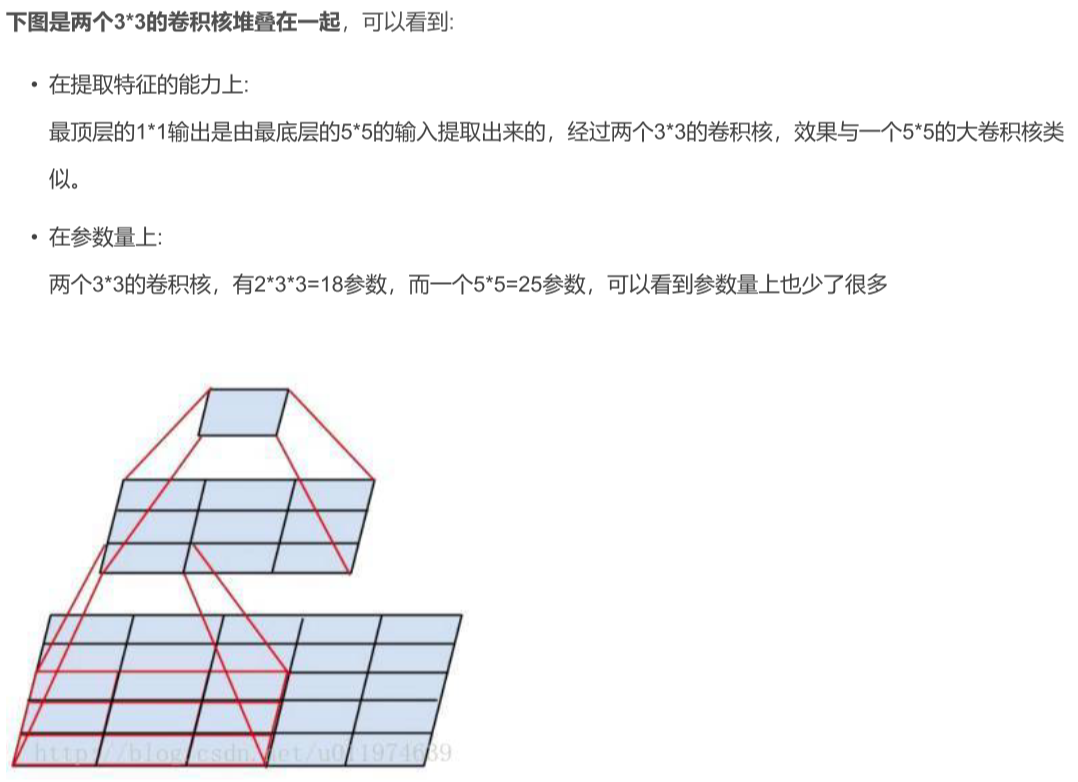


使用多个小卷积核堆叠，感受野与大卷积核相同。

相比于单层，使用三个非线性整流层会使决策函数更有判别力。（多个小卷积核的堆叠会比单个大卷积核有更多激活函数的转换（ReLU），学习能力更强）

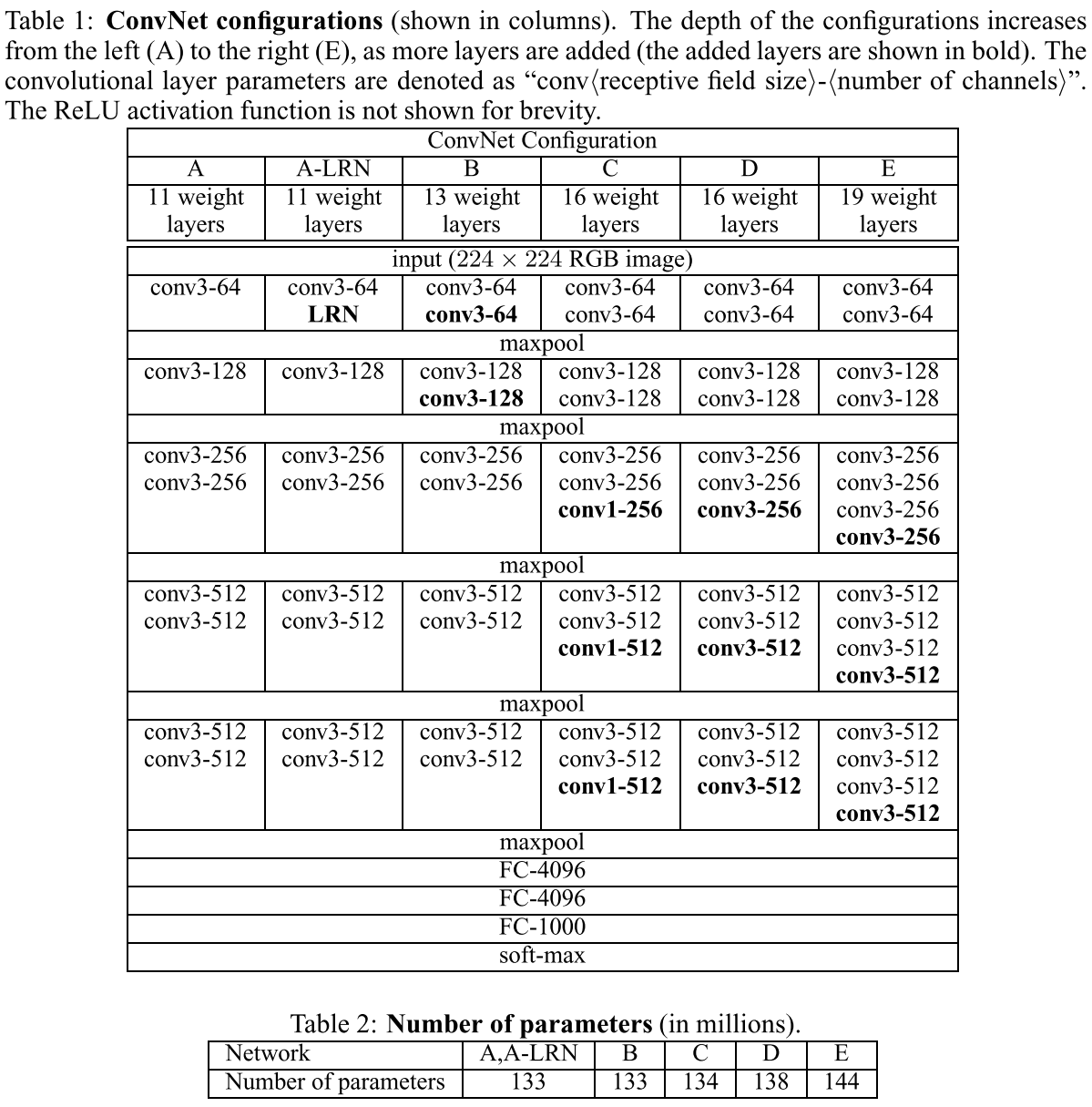
同时，多个小卷积核会有更少的参数。

使用的卷积核，可以在不影响卷积层感受野的情况下增强决策函数的非线性，即使在本文模型使用卷积核是同维度上的线性变换。（卷积后会经过ReLU处理，而ReLU在输入大于0是线性的）

左图是两个3\*3的卷积核堆叠在一起，可以看到:

在提取特征的能力上: 最顶层的1\*1输出是由最底层的5\*5的输入提取出来的，经过两个3\*3的卷积核，效果与一个5\*5的大卷积核类似。

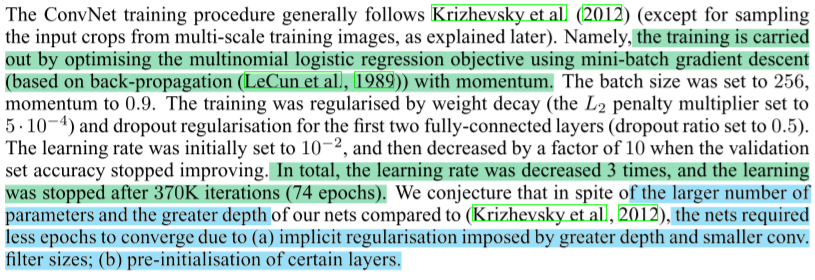
在参数量上:两个3\*3的卷积核，有2\*3\*3=18参数，而一个5\*5=25参数，可以看到参数量上也少了很多



**3. Classification Framework（分类框架）**

本节介绍卷积神经网络训练和评估的细节。

**3.1 Training**

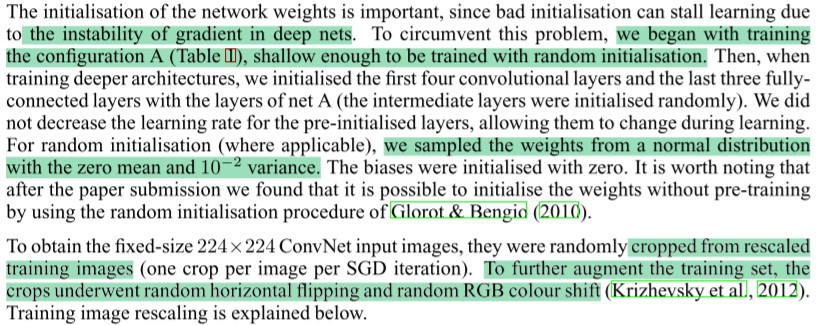


训练过程基本按照AlexNet的步骤进行（除了sampling the input crops from multi-scale training images）使用基于反向传播的小批量梯度下降来优化多项逻辑回归目标进行训练。

尽管与AlexNet相比，网络的参数数量更多，网络深度更大，但网络达到收敛需要的迭代次数更少，原因在于：

(a) 由更大的网络深度和更小的卷积层滤波器尺寸带来的隐式正则化。

(b) 某些层的预初始化

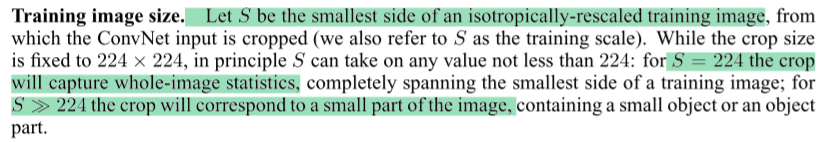


网络权重的初始化很重要，因为深层网络中的梯度不稳定，初始化不好可能会导致学习停滞。

为此先训练一个深度较浅的网络A，再利用已经训练好的A网络权值初始化深度较深的网络。（迁移学习）

初始化权重从均值为0方差为0.01的正态分布中采样。

为了获得固定大小的的卷积网络输入图像，他们从重新缩放的训练集图像中随机裁剪（每个SGD迭代的每张图像进行一次裁剪）。增强训练集的方法：随机水平翻转和RGB颜色偏移。



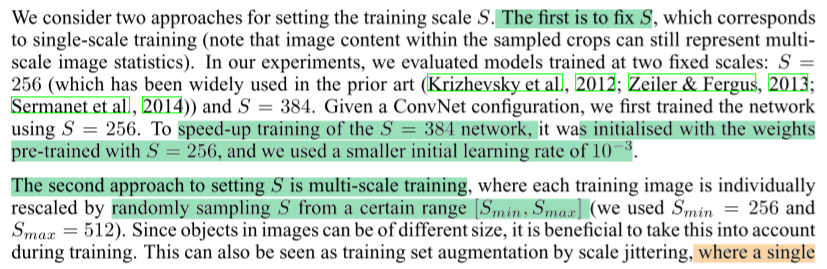
设S是各向同性重新缩放的训练图像的最小一侧长度（也称S为训练尺度）。

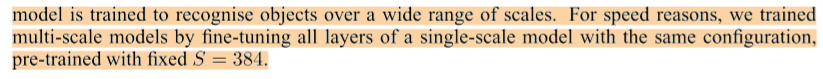
虽然剪裁大小固定为224×224，但原则上S可以取不小于224的任何值：

，剪裁将捕获整幅图像统计数据，完全跨越训练图像的最小侧;

，剪裁将对应于图像的一小部分，包含一个小物体或物体的一部分。

部分。





我们考虑设置训练尺度S的两种方法：

(1) 使用固定的输入图片尺寸S

论文评估了以两个固定尺度训练的模型：S=256和S=384

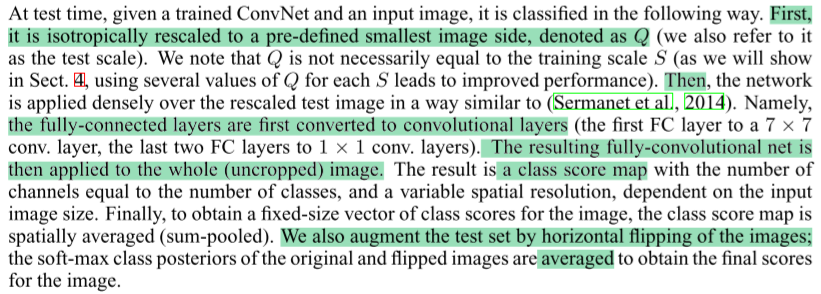
给定一个卷积神经网络配置，我们首先使用S=256来训练网络。为了加速S=384的训练，网络权重被初始化为S=256的预先训练的权重，并且使用0.001的较小的初始学习速率。

(2) 使用变长的输入尺寸S，

每个训练图像通过在区间内随机取样S而单独重新缩放。

由于图像中的物体可能具有不同的大小，因此在训练中考虑这一点是有益的。这也可以看作是通过缩放抖动来增强训练集，其中单个模型被训练以识别范围广泛的物体。出于速度的原因，我们通过使用相同配置对单尺寸模型的所有图层进行微调来训练多尺度模型，预先训练使用固定的尺寸S=384

**3.2 Testing**



首先将输入图像各向同性地缩放到预定义的最小图像边长，表示为Q（测试尺度）。

Q不一定等于S。

完全连接层首先被转换成卷积层，将所得到的全卷积网络应用于整个（未裁剪的）图像。

结果是一个分类分数映射，通道数量等于分类数量以及一个可变的空间分辨率（取决于输入图像的大小）

为获得图像分类分数的固定大小向量，分类分数映射是空间平均的（sum-pooled）

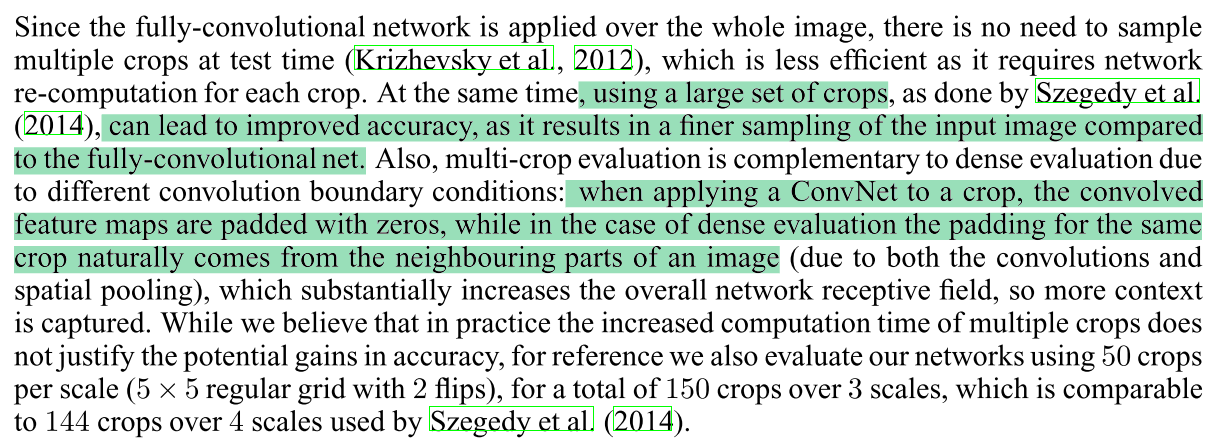
我们还通过水平翻转图像来增加测试集;对原始图像和翻转图像的softmax类后代进行平均以获得图像的最终得分。（多个输入在softmax层做平均输出）

关于这一段的解释：

在测试阶段中，作者把测试图片的最短边设为Q(Q和训练过程中图片的最短边S不一定相等)，而且可以对同一张测试图片rescale成不同的大小，即多个scale的同一张图片都在网络中进行测试，这样可以提升测试效果。在测试阶段，作者参考Sermanet等人的做法，把网络的连接层转换为多个卷积层进行处理。

（可以把全连接层中的每个节点看作是大小的feature map，若有n个节点，看作是的n通道的feature maps）

以VGG-16为例：对于的输入，最后一层卷积可得输出为，如后层是一层含4096个神经元的全连接层，则可用卷积核为的全局卷积来实现这一全连接运算过程。



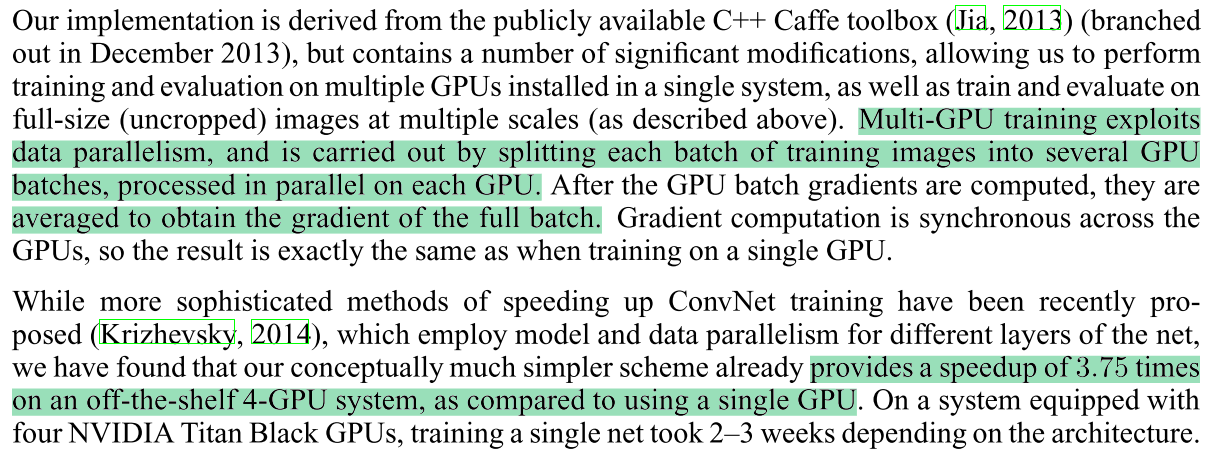
全卷积网络应用于整个图片之后，不需要像AlexNet中那样对一个测试图片进行多次裁剪（低效，需要网络为每个crop重新计算一遍）

AlexNet需要对不同crop分别进行计算，而VGG中可以一次计算获取同一张图片不同区域的分类结果。

同时，把图片切成大量的crop，可以提高准确率。

在做卷积操作时，边界做零扩展，而在multi-crop评估的情况下，图像扩展来自其相邻部分。

**3.3 Implementation Details**

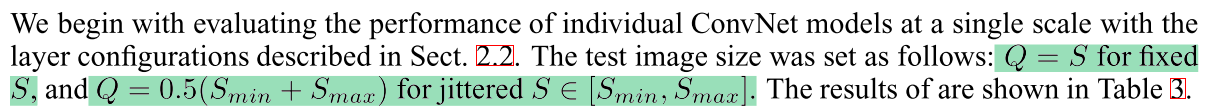


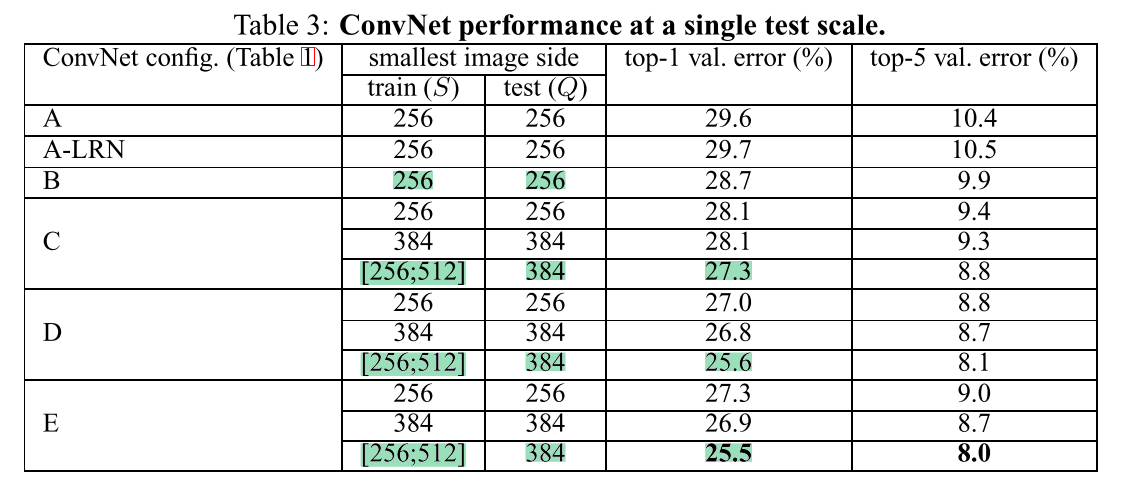
实现基于caffe，做了一定的修改，允许在单系统的多核cpu上进行训练，用多GPU并行计算训练图像的每个batch，当所有GPU都计算完成之后，求所有batch得到梯度的平均值。

**4. Classification Experiments（分类实验）**

**4.1 Single Scale Evaluation**

单一尺度，Q值固定

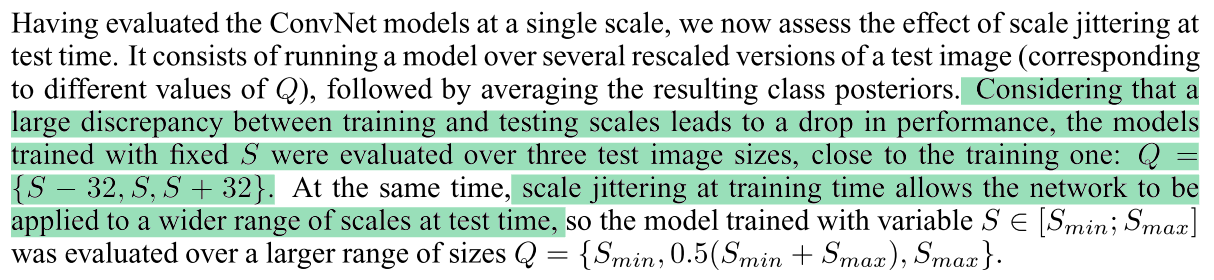




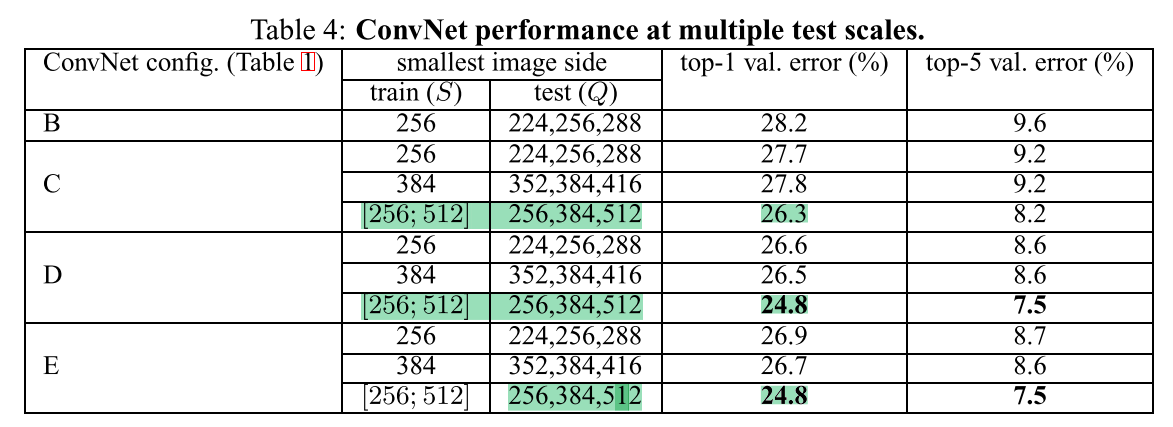
结论：scale jittering下有更低的错误率。

**4.2 Multi-scale Evaluation**

取多个Q值

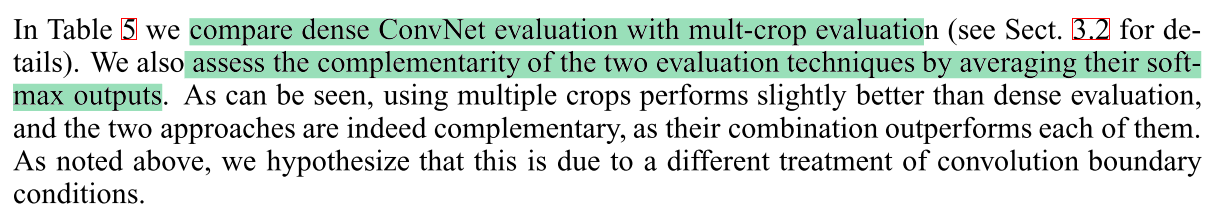


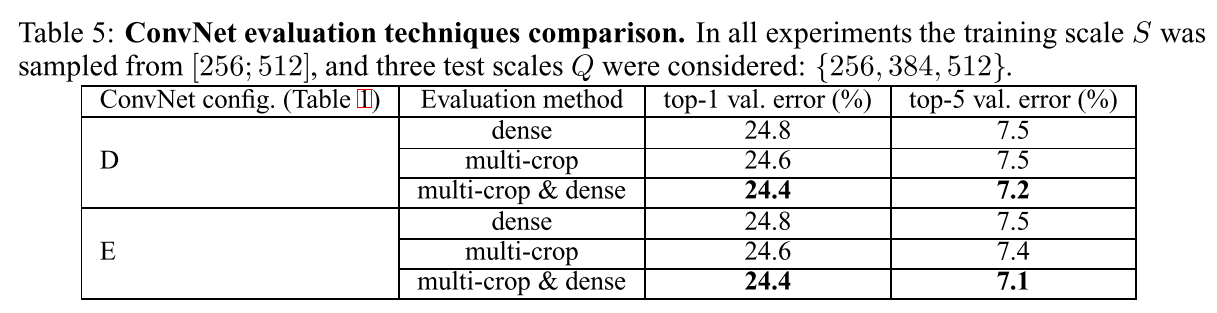
考虑到由训练和测试的尺度不同导致性能上的巨大差异: 模型使用固定的训练输入尺寸S来评估三种尺寸 的测试图片，即测试输入图片尺寸Q取 。 同时，我们可以对训练尺寸S做尺寸抖动，抖动范围为，相对的测试图片的大小就变为 . 在表格中，我们可以看见带尺寸抖动的模型是要优于不带抖动的。



结论：多尺度测试比单尺度结果要好，且使用scale jittering有助于提高模型的性能。

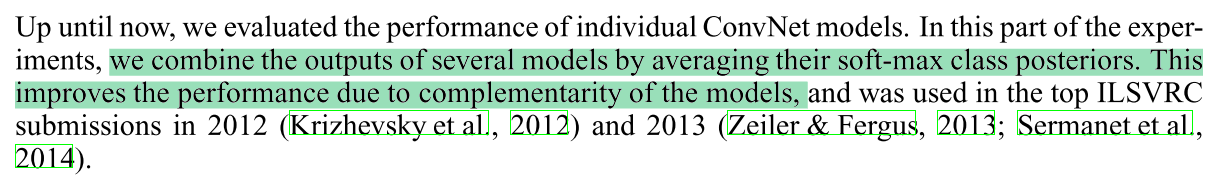
**4.3 Multi-crop Evaluation**





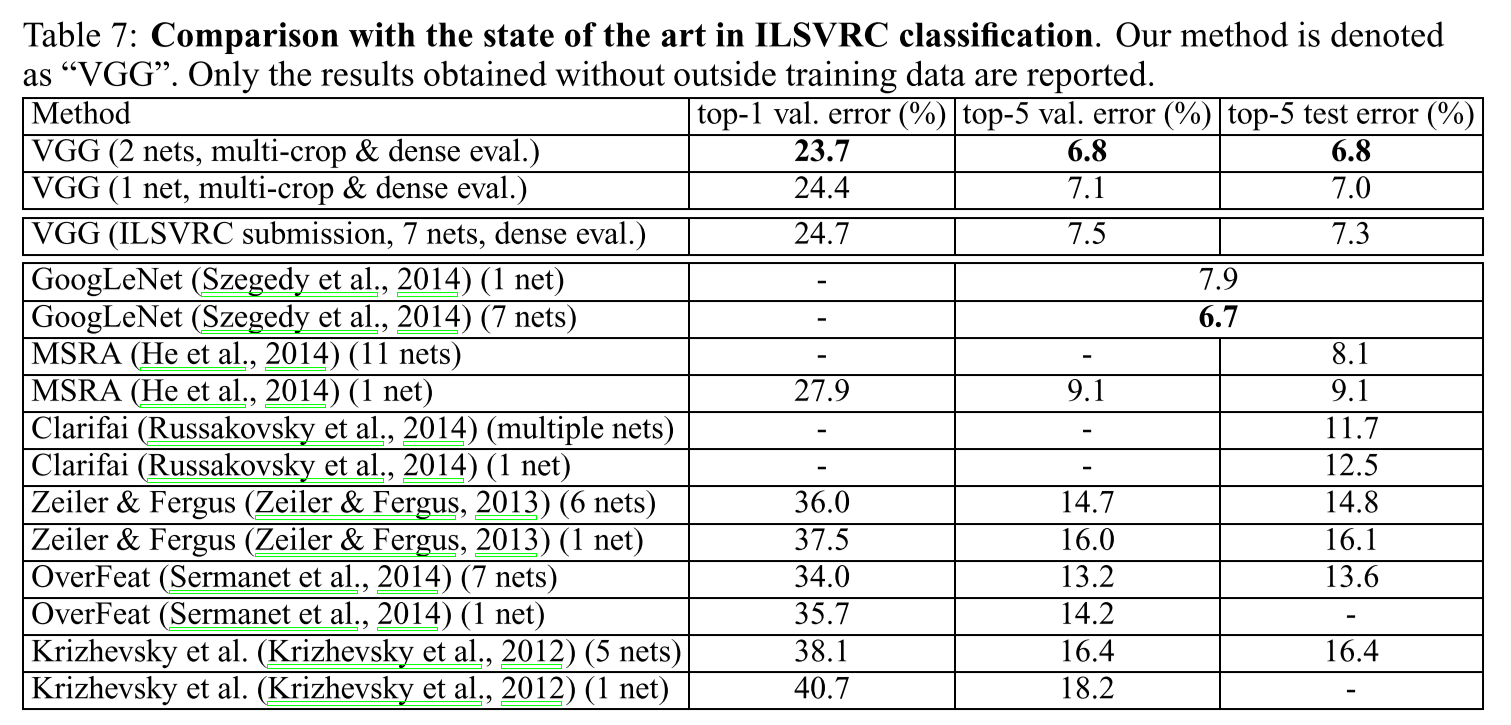
结论：单独使用multi-crop evaluation 要比单独使用dense evaluation效果好，两个方法同时使用时，要比单独使用任一种都好。

**4.4 ConvNet Fusion（融合）**



结论：将几个模型的soft-max分类策略的输出求平均后再用于识别，这样可以提高最后的性能。

**4.5 Comparison with the State of the Art**



结论：挺好

**5. Conclusion（总结）**

我们评估了非常深的卷积网络（多达19个权重层）用于大规模图像分类。已经证明，增加深度有利于分类准确性的提升，通过大幅增加深度，可以使用传统的ConvNet架构实现更好的性能。

附录中，展示了模型应用于更广泛的任务和数据集的情况，可以和围绕深度图像表示形成的更复杂的识别流水线的结果相当或者更优。

综上：实验的结果证实了深度在卷积神经网络中的重要性。

**A Localisation**

**A.1 Localisation ConvNet**

**A.2 Localisation Experiments**

**B Generalisation of Very Deep Features**