### [V](http://cn.arxiv.org/abs/1409.1556" \t "_blank)[ery Deep Convolutional Networks for Large-Scale Image Recognition](http://cn.arxiv.org/abs/1409.1556" \t "_blank)

Abstract

本文主要是研究卷积网络深度在大规模图像识别设置中的准确性的影响，主要贡献是使用非常小（3×3）卷积filter，此网络深度较大，为16-19，并表现出显著的性能提升。此网络在一些比赛中表现较好，并且可以很好的推广到其他数据集

Introduction

# ConvNet configurations

# Classification framework

# Classification experiments

# Conclusion

# Introduction

# 卷积网络在大规模图像和视频识别方面表现优异，这得益于大型数据库如Imagenet和高性能计算系统的发展，ImageNet大规模视觉识别挑战赛也发挥了重要作用

# 随着卷积网络的使用日益广泛，有许多对其的改进。例如利用较小的接收窗口尺寸和较小的第一卷积层步长，另一项是在整个图像和多个尺度上密集地训练和测试网络，本文主要讨论网络的深度。

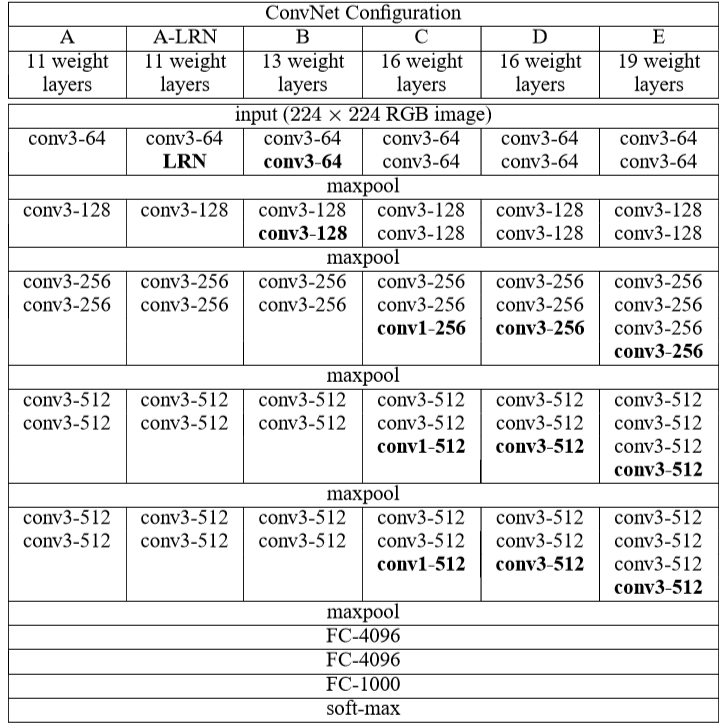
# 2.ConvNet configurations卷积网络配置

# 2.1 Architecture

# 卷积网络的输入为224×224 的RGB图像，预处理是从每个像素中减去在训练集上计算的平均RGB值。使用的filter是3×3，其中一层用的是1×1，卷积步长为1，padding是same，使得卷积后的空间分辨率大小不变。池化层为最大池化，规模为2×2，步长为2

卷积层后有三个全连接层， 前两个每层有4096个通道，第三个包含1000个通道。 最后一层是soft-max层。 全连接层的配置在所有网络中都是相同的。每个隐藏层后都有ReLU非线性函数。本文的网络没有局部响应归一化（LRN）,因为其无法改善性能。

2.2 Configuration

本文的各种卷积网络配置均相同，只有深度不同，从网络A中的11个权重层（8个卷积层和3个全连接层）到网络E中的19个权重层（16个卷积层和3个全连接层）。每层的通道数也很小，从第一层中的64开始，然后在每个最大池层之后增加2倍，直到达到512。

2.3 discussion

2个3×3卷积层连在一起相当于一个5×5的卷积层，三个连在一起相当于单个7×7层。优点是合并了三个非线性层而不是一个非线性形层，这使得决策函数更具有判断性。其次可以减少参数数量。假设输入和输出的通道数都是C，那么3个3×3的filter的参数数量是3\*(3\*3\*c)\*c=27c\*c个参数，而使用一个7×7的filter的参数为(7\*7\*c)\*c=49c\*c个参数。

使用1×1大小的卷积，可以通过修正函数增加额外的非线性，本文的网络在1×1卷积后的输入输出维度没有改变

3. Classification framework

3.1 Training

卷积网络通过使用具有动量的小批量梯度下降优化多项逻辑回归目标来执行训练。批量大小设置为256，动量为0.9。通过权重衰减参数为设置为5e(-4)和前两个完全连接层的dropout正则化（dropout率设置为0.5）对训练进行正则化。学习率初始设定为10e-2，然后当验证设定精度停止改善时，学习率降低10倍。总共，学习率降低了3倍，并且在370K迭代（74个时期）之后学习被停止。

网络权重初始化很重要，因为A很浅可以使用随机初始化训练。先对配置A进行随机初始化。在训练更深的网络时，网络的前四个卷积层和后三个全连接层用配置A进行初始化，中间层采用随机初始化。对于随机初始化，权重从N(0, 0.01)正态分布中进行采样，偏差初始化为0。

网络输入要求为224×224，需要在重新缩放的训练集上进行随机裁剪

为了数据增强data augmentation，对裁剪区域进行随机水平翻转（horizontal ﬂipping）和随机RGB颜色转换（shift）

Training image rescaling

对训练图片进行缩放，在图片上随机提取224x224窗口，进行训练。设各向同性重新缩放的训练图像最小边为S，本文网络要求输入图像是224×224，S=224时，会直接以全图进行训练，当S>>224时，裁剪后的图像只是原图一小部分。设置S的方法有两个：

一种方法是固定S，文中训练S=256和384两个分类器

第二种方法是通过从特定范围[Smin，Smax]随机采样S来单独地重新调整每个训练图像。即每次数据输入的时候，每张图片被重新缩放，缩放的短边S随机从 [Smin，Smax]中选择一个，文中用的是[256,512]。通过尺寸抖动（scale jittering）也可以认为是数据增强

3.2 testing

dense ConvNet evaluation with multiple-crop evaluation

测试阶段步骤：1，对输入图像各向同性地重缩放到一个预定义的最小图像边的尺寸Q; 2. 网络密集地应用在重缩放后的测试图像上。也就是说全连接层转化为卷积层（第一个全连接层转化为7\*7的卷积层，后两个全连接层转化为1\*1的卷积层） ，然后将转化后的全连接层应用在整张图像上。结果就是一个类别分数图(class score map)，其通道数等于类别数量，依赖于图像尺寸

3对类别分数图进行空间平均来获得固定尺寸的类别分数向量

3.3 Implementation Detail

使用多GPU加速

4. Classification Experiment

单尺度评估 Single scale evaluation

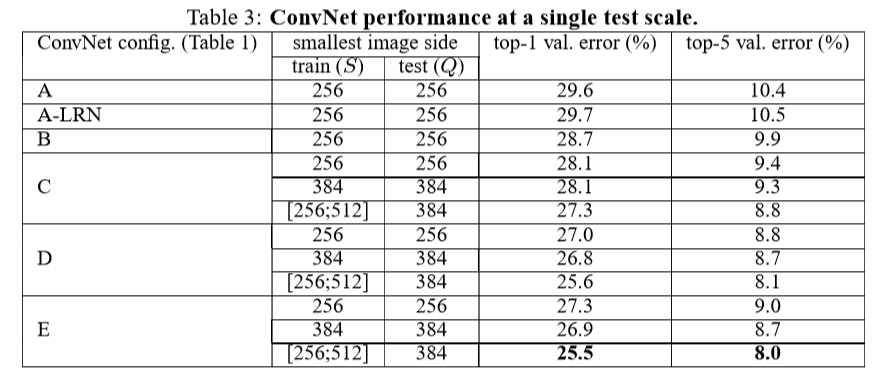
LRN没有改善网络，所以在后面的网络中去掉了LRN

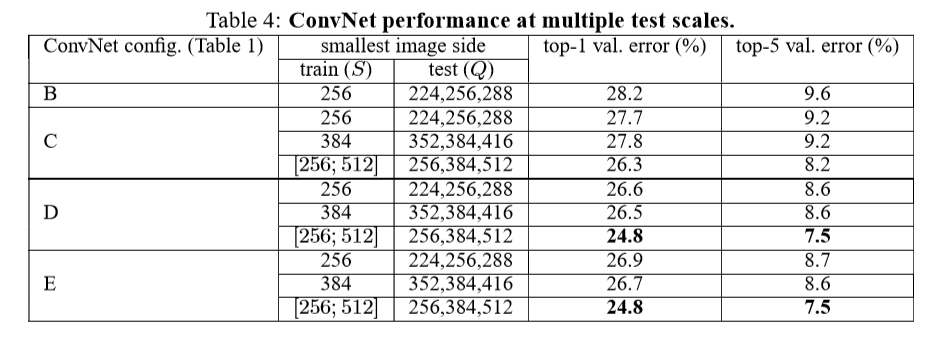
网络越深，分类误差越小

使用1×1卷积增加额外的非线性有所帮助（C比B好），但是不如使用3×3卷积（D比C好）。

B中有两个3×3的卷积。将两个卷积换成一个5×5卷积后的网络与B比较，发现B的效果更好，表明小的filter效果比大的好

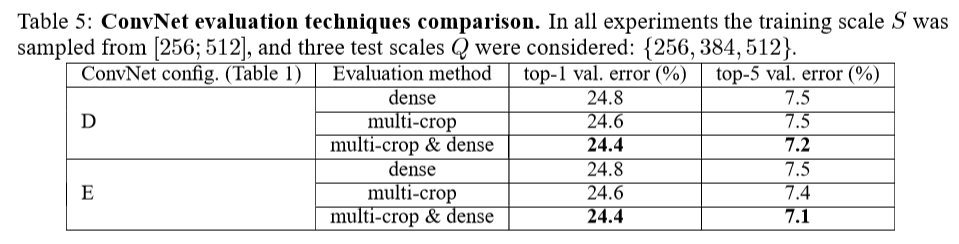
采用图像尺度抖动(scale jittering)可以改善图像分类效果





多尺度测试 multi-scale evaluation

测试尺寸抖动性能更好



多裁剪(multi-crop)评估比起密集(dense)评估，效果更好。而且两者具有互补作用，结合两种方式，效果更好。

卷积网络融合ConvNet Fusion

如果结合多个卷积网络的sofamax输出，分类效果会更好。   
先是我们结合7个网络实现测试误差7.3%。之后，结合最好的两个模型(D&E)并使用密集评估(dense evaluation)，测试误差降低到7.0%，而使用密集评估和多裁剪评估相结合，测试误差为6.8%。最好的单一模型验证误差为7.1%。

# 