非常深的卷积网络对于大规模图像识别

# 摘要

本文的贡献调研了大规模图像识别中卷积网络深度的影响，使用3\*3的小卷积核加深深度。表明了性能的提升到16到19层显著怎加了性能。在ImageNet上实现取得了第一第二的成绩（定位和分类），同时对其他数据集具有良好的泛化性能。

# 介绍

介绍卷积神经网络的优越性，从方法上站住脚。然后提出他人的贡献。一方面减小卷积核另一方面多尺度检测，作者则主要讨论了深度对卷积神经网络的影响。

结果，作者提出了一个优秀的网络架构，再ImagNet取得了良好的结构，并且可以在其他数据集取得良好的成绩（可以作为SVM的fine\_tune）。

论文架构：

1.讨论神经网络的配置

2.讨论图像分类训练的细节

3.在ImageNet上的最实验

4.总结论文

5.介绍在其他数据集上的泛化

# 卷积网络配置

为了公平体现增加卷积网络深度带来的提升，本文采用了与。。。相同的配置。首先描述通用架构，再讨论细节，最后做比较。

## 架构

在训练期间，输入到卷积网络的是一个固定大小地224\*224RGB图像。

唯一的预处理是对每个像素值减去在训练集上的均值。

然后使用3\*3的卷积核来进行卷积处理，在其中一个构型里使用了1\*1的卷积核，可以看作线性操作。卷积核步长为1，做1padding，保持分辨率不变。空间池化层2\*2最大池化，步长为2，紧跟卷积层。后面跟着全连接层，4096通道，最后1000通道，softmax，配置相同。

所有层使用ReLU激活函数，不进行归一化（LRN）操作，这种操作为增加内存和计算时间。

## 配置

不同的配置只是深度不同，从A到E命名。开始通道数较少，后来通道数增加，每个池化层后增加一倍。

尽管深度更深，但权重的数量并没有比（更大卷积，层宽和感受野）的更浅的网络多。

## 讨论

可以看到相对于以前的网络作者用3\*3的卷积核代替了大卷积核，步长为1，2层3\*3代替5\*5的感受野。3层7\*7的感受野。

原因：首先，我们将3个非线性层用1个代替，着增加了决策函数的区分能力。其次减少了参数的数目。这里作者将其看作对于7\*7卷积的正则化，3个3\*3的卷积核是27C方个参数，一个7\*7卷积层是49C方的参数，比较来说，这里通过加深卷积深度提高解析能力，同时强制减少了参数的数目（正则化）。

1\*1的卷积核是增加非线性的方法，在不改变感受野的情况下。这种架构有所使用。

其他人也使用了小卷积核的方法，但他们没有我们的网络深，同时没在ImageNet数据集上有进行测试。Goodfellow使用11层的网络对街道号码进行识别取得了良好的效果。 展示了增加深度可以有一个更好的表现。2014年GooleNet取得了一个良好的ImageNet分类效果，他们也使用很深的网络，除了3\*3的卷积核外，他们也使用了1\*1和5\*5的卷积核，他们的网络结构更复杂，在第一层中特征图的减少更剧烈，以减少计算数量，我们的网络比他们的好。

# 分类框架

前一部分，我们展示了我们网络配置的细节。在这部分，我们描述了我们用网络做份时是如何训练和如何探索的。

## 训练

卷积神经网络的训练一般遵循链接的要求，除了采样做多尺度训练。也就是训练时使用小批量多维度逻辑回归（动量的形式）。Batch\_size256,动量0.9，训练时通过权重衰减进行正则化（L2是5\*10的负四次方）和dropout后面的全连接层（0.5）。学习率初始10的负二次方，当验证集效果停止提升时除以10.总的来说，学习率要衰减3倍，在370k个迭代后停止学习（74个epoch）。我们推测尽管我们的网络有更大的深度和更大的参数，网络需要更少的epoch是由于（a）更大深度和更小卷积核的正则化效果（b）某些层的预处始化。

网络的初始化很重要，一个糟糕的初始化会阻碍深度网络里梯度的稳定性。为了解决这个问题我们从A开始使之可以通过一个随机的初始化来进行训练。然后，当训练结构加深时，我们用已经初始化的4个全连接层和后面的三个全连接层来指导训练中间随机初始化的层来进行训练。我们不降低预初始化层的学习速度，允许它们在学习过程中发生变化。对于随机初始化（在适用的情况下），我们从正态分布中抽取权重，使用0均值和10的2负方差。

在交论文的时候，发现可能不初始化权重，直接随机初始化也是可能的，见链接

为了获得固定大小的224\*224的卷积网络输入图像，我们对训练图像通过裁剪的方法进行缩放处理（每次迭代使用一种裁剪），为了进一步增强数据集，我们进行了随机水平翻转和随机RGB颜色偏移，见链接。

训练图像：让S指的是一个等轴缩放的训练图像的最小边（称S为训练范围）。当裁剪范围被固定到了224\*224的时候，原则上S可以取任何不小于224的值，对于S等于224，裁剪范围将取整个图像的统计范围，对于S远大于224，这个裁剪将获得图像包含一个小物体或是物体一小部分。

作者使用了两种方式来进行处理，一是设置训练范围固定为S，这对应与单一尺度的训练（注意这里图像内容进行裁剪后依然可以代表图像的统计性特征，在我们的实验中我们探索了两个固定的范围，S=256,S=384,为了加速训练，我们首先用S=256的网络进行训练，然后用S=256训练好的特征初始化权重，用一个10的负三次方的初始化学习速率。

第二种是多尺度学习，从一个范围内随机抽取一个S对图像进行单独的缩放[Smin,Smax](这里S取256，S取512)，因为目标在图像中可能有不同大小，因此这里在训练期间这样做是有意义的。这可以通过一个按照比例调整的数据集，模型可以被训练来识别范围广泛的物体，由于速度原因，我们训练了多尺度模型通过对具有相同配置的单尺度模型的所有层进行微调，预训练的固定S=384。

## 测试

在测试时，给定一个训练的卷积神经网络和一个输入图像，它以如下的方式分类。让Q指的是一个等轴缩放的训练图像的最小边（称Q为测试范围），我们注意到Q并不一定等于培训规模S（正如我们将在第4节中所展示的那样，使用几个Q值可以提高性能），网络以一种紧密的形式应用于测试图像，相似于链接。 全连接层被转换为卷积层（第一个全连接层被转换为7\*7的卷积层，后两个全连接层用1\*1的卷积层代替）。这个全卷积网络被应用到整个未裁剪的图像中。结果时一个类别分数和类别数目相等以及一个可变空间分辨率。最后为了获取一个固定size的类别分数对于图像，分类得分映射被应用于一个空间平均（总池化），我们采用水平翻转的形式增强数据集，对原始图像和翻转图像的softmax分类得分的平均获得图像的最终得分。

因为全卷积网络覆盖整幅图像，因此不需要进行裁剪，如果当网络要求对每个裁剪进行重新计算时很低效（例如，似乎faster rcnn就是认为一个小范围的图像可以代表总体链接）。同时进行大量的裁剪似乎可以提高准确度，相对全卷积网络而言，这里是对输入图像更好的采样，链接。多裁剪评估与密集的评估是相互补充的。当使用裁剪的时候，卷积特征图会padding0，而在密集评估的情况下，padding自然来源与图像的相邻部分。 这里作者感觉这两者互相补，此外，多重剪切对于密集探索是一种互补，有助于探索不同卷积的边界条件的影响，在这种条件下，剪切是用0做pad，密集探索使用一个图像邻近区域做pad（由于卷积核空间池化），相当于增加了增加了网络的全局感受野，可以捕捉更多背景信息。

而我们相信实践中多重切割带来的收益不能与增加的计算时间带来的负担相称，作为参考我们也探索了使用50种切割范围的方法作为参考，使用了类似上面论文的方法，两种方法切割范围相当。

这里的意思一个人似乎是全卷积网络应用时覆盖整幅图像，没有必要在测试时对图像重新裁剪，这种方式很低效。这样就可以使用原图像与周围图像的信息，不用补0。但另一个人认为，为了提高准确率，应该使用一个大规模的裁剪，得到一些子图像，与全卷积网络相比这样是原图像更好的采样，似乎这里的意思是对原图像进行裁剪后判断一个和得分更好，因为这种采样可以帮助一些受到阻挡的物体得到一个正确的分类，但似乎会使更多反例被识别进去，这里似乎取决于原始分布，经过阻挡的物体是否能表现一个物体的特征呢？？？？，而且这种实现方式裁剪后就是裁剪后，不能再加信息了0，但似乎可以折中，对与图像通过活动窗口的方法进行全卷积，同时利用周围的信息，但效率会不会太低了，这里具体方式需要看论文，整体有待商榷，似乎需要一种既可以代替全连接层的方法，又可以实现全卷积的效果的运算来进行这种处理，这部分呢感觉跟我以前遇到的问题一样什么样的样本应该算是正样本，将控制裁剪的范围是为超参数，就像文中所做的一样，多范围比较。

## 实施细节

这里是土豪的方法，多GPU并行计算，实施不起啊，需要2到3周的时间，我的天哪。

# 分类实验

数据集：ImageNet 1.3万幅（如果用M来表示就会引起歧义，这里需要知识库的支持）训练集，50K验证集，100K测试集。这里使用两种指标来衡量性能，top-1error,top-5error.前面是一个多分类错误率了，衡量的是分类错误的比例，后面是ImageNet的主要指标，判断只要前五个分类概率中有正确概率计算正确。大多数实验中用验证集做测试集。

## 单一尺度检验

使用前面的单一范围图像进行检验，Q=S和Q=0.5(Smin+Smax).j结果如表3.

这里使用局部归一化没有提高效果（LRN），因此更深架构里也没有用。

随着深度的增加，错误率降低了，这里显示1\*1卷积层加一层没有加一层3\*3卷积层表现好，这证明了额外的非线性确实有帮助，但空间信息更重要，即带有更大的感受野。模型的错误率随着更深到19层而达到饱和，但更深的模型对更大的数据集可能有益。这里显示用小卷积核代替大卷积核是有效果的。

最后使用[Smin, Smax]的方法效果比单一尺度下的训练效果更好，即使测试时使用同一尺度的图像。这证实了通过数据集的扩充可以实际上捕捉多尺度图像的信息。

## 多尺度探索

测试时在多个重新缩放的图像上进行测试（不同的Q），对测试结果进行平均,考虑到训练和测试尺寸上的巨大差异会导致性能的下降，用固定S训练的模型，在三个尺寸上进行测试，Q={S-32,S,S+32},同时训练时候进行不同尺度训练的网络，S在[Smin,Smax]内的，Q在{Smin,0.5(Smin+Smax),Smax}上进行评估。这里显示测试时的缩放范围会带来更好的性能，多尺度图像分类取平均。

## 多尺度剪切探索

这里作者用多尺度剪切的方法和密集探测的方法进行对比，发现多尺度剪切的方法效果缺失好一些，同时也证明了两种方法确实是互补的，如上所述，我们假设这是由于不同的卷积条件造成的。

## 卷积模型融合

这里作者考虑了近几年在ImageNet上表现良好的好的模型方法，对自己训练表现最好的模型进行了融合，平均了模型的softmax得分。在提交结果的时候，作者只训练了单一尺度模型，和通过fine\_tune全连接层的方法获得了一个多尺度模型D，由此获得了7个模型的融合。

作者虽然没有脱离LeCun的卷积架构，但从本质上改进了它，通过增加网络深度。

# 总结

我们的工作表明了深度的的重要性对于大规模图像分类任务。深度对于对于分类正确率是有益的，在卷积网络的架构上只增加深度。附录里我们展示了我们的方法对于一系列数据集的泛化性能都很良好，与不那么深的图像表示效果相当或更好。证实了深度的重要性。

全文虽然主题是深度，这也是作者的重要发现，但所涉及的内容对于许多实验都有指导意义，这提示我们写论文时要紧扣自己发现的核心，并且通过巧妙的实验设计证明。