# 摘要

本工作中我们研究了卷积神经网络网络深度对于大图片识别准确率的影响。我们的主要贡献是对使用很小的3\*3卷积核的越来越深的神经网络做了彻底的评估，它显示在现有技术配置下，将网络层数增加到16-19层，可以带来明显的改进。这些发现是我们在2014年ImageNet挑战赛的本地化和分类任务中**取得第一和第二的**基础。我们还展示了我们的表示可以扩展到其它数据集上，并能获得最先进的结果。我们公开了两个性能最好的ConvNet模型，以便于对在计算机视觉领域使用这种 **深度表示型的方法** 做进一步研究。

# 引言

卷积网络进来在大尺寸图片和视频领域取得了相当的成功，这归功于公开的图片库，高性能的计算机系统，gpu或者分布式集群等。特别的，ImageNet挑战赛对这些进步起到了重要的作用，它担任了几代图片分类系统的试验台。从2010的 high-dimensional shallow feature en

codings 到2012的 deep ConvNets (Krizhevsky et a，2012) 。随机卷积网络成为cv领域的常用品之后，大量的基于原始架构的尝试被提出，改进了准确率。例如，2013的最佳论文，在第一层使用更小的接受窗口和步长。另一种改进线路是，在整个图片和不同尺寸缩放图片上，不断的做训练和测试。本文中，我们提出另一种重要的设计方向，**网络的深度。**为了实现这个目的，我们固定网络的其它参数，然后使用卷积层稳定的增加网络的深度，这个通过使用很小的3\*3的卷积核变得可行。

结果，我们在ConvNet架构上得到了显著的准确度提示，不只是ImageNet挑战赛上，还包括其它的数据集上。并且即使他作为其它架构的管道的一部分，也能取得优秀的成果。

我：总结也就是说 **小核+大深度的方法， 效果强+适应度广**

# ConvNet配置

为了在公平的条件下测量增加网络层数对网络的提升，我们卷积网络的所有层都使用同一种原则进行设计，这个受到Ciresan 等的启发。这这一节中，我们首先在2.1中描述我们的ConvNet的通用布局，然后在2.2中详细描述评估中所用到的特殊设置。最后在2.3中讨论我们的设计选择，并和当前已有的例子进行比较。

2.1 架构

训练过程中，输入固定为 224\*224 的RGB图像。只有一种预处理，就是对每个像素，减去整个数据集的RGB的均值。图像被传递到卷积栈中，卷积栈中使用3\*3的卷积核，这是能捕捉到中间+上下左右概念的最小的卷积核了。在另一种配置中，我们还使用过1\*1的卷积核，这可以被视为一种对输入通道的线性转换。卷积步长是1。增加padding来保持输入图片的尺寸，因此3\*3卷积和对应的padding = 1。空间池化由5个最大池化层组成，它放在某些卷积层后面，但不是所有卷积层后面都有。池化窗口大小为2\*2 步长为2.

卷积栈后面放置全连接层，全连接第一层通道数4096 第三层支持1000个分类因此有1000个通道。最后一层是soft-max层。全连接层的配置和所有其它网络的配置一样。所有隐藏层使用RELU函数来非线性化。我们注意到，除了一个层之外，所有其它的层都不包含局部响应正则化。因为，如第四节中所示，这样的正则化不能提升模型的性能，反而导致了内存的增加和时间的消耗。

2.2 配置

本文中说评估的ConvNet的配置，总结在了Table1中，一列代表一种。下面我们使用它们的代号（A-E）来指代它们。所有配置都遵循上面2.1所述的通用设计。不同的仅仅是深度，从11个权重层到19个权重层。卷积层的宽度，也即通道数相当小，从第一层的64开始，然后在每个池化层之后以2倍的方式递增，直到512.表格2，我们报告了每种配置的参数数量，尽管深度很大，我们的网络的权重数量并没有比其他的更浅的使用更大的卷积层的模型大。

2.3 讨论

我们的模型配置和2012、2013挑战赛中表现很好的模型有很大的不同。不是在第一层使用很大的感受野，比如11\*11 + stride 4 ，7 \* 7 + stride 2，我们在整个网络中都使用3\*3的网络。很容易看到2个3\*3卷积栈就相当于一个5\*5的卷积栈，3个3\*3的相当于一个7\*7的。所以为什么我们不用3个3\*3的代替一个1个7\*7的呢。**首先，我们使用3个非线性的整流层而不是一个，这使得决策函数具有更强的识别能力。第二，这样降低了参数的数量，具体计算...。**这可以看做一种对7\*7卷积过滤器的正则化，强迫他们结构为3\*3的过滤器。合并1\*1的卷积层是在不影响卷积层感受野的情况下，增加决策函数非线性的一种方法。