使用综合卷积神经网络对大规模图像进行识别

白稹 青海民族大学 galijiangzhi@163.com

摘要

在这项工作中，我主要整合了自 2010 年以来，对大规模图像识别任务具有重大提升的卷积神经网络模型，主要包括在 ImageNet 大规模视觉挑战识别赛中取得优秀成绩的 AlexNet，vgg，GoogLeNet，ResNet。这些网络结构从各个方面解决了卷积神经网络参数数量多，模型特征提取结果表现差，模型网络退化等问题，我通过参考这些模型的论文，设计了一个综合这些模型的优点的网络，其中训练结果最好的一个版本被我命名为 PhantomNet，这个网络的深度在我做完实验后我会给出。

1.引言

自 2012 年 AlexNet 卷积神经网络问世以来，使用卷积神经网络进行图像识别的效果在以惊人的速度进步，这一现象不仅仅是因为计算机性能的提升，更是因为深度学习技术的迅速发展和广泛应用。卷积神经网络作为一种经典的深度学习模型，在图像识别领域展现出了强大的特征提取和分类能力，使得其成为了图像识别领域的主流技术。除了 AlexNet 之外，Vgg、GoogLeNet、ResNet 等一系列卷积神经网络模型的提出和不断改进也为图像识别的进步贡献良多。举例来说：在 2014 年，vgg 团队提出了“使用多个小卷积核堆叠替代大卷积核”的思路，减少了大感受野带来的参数量暴涨问题，同年，GoogLeNet 首次提出的 inception 结构解决了不同大小卷积核对特征提取不全面的问题，同时，该团队根据新加坡国立大学发表的“network in network”论文中的方法，在网络中加入 1*1 的卷积减少网络参数的数量，使得网络大小较 vgg 的网络缩小了十二倍，而且更加精确。

值得注意的是，这些网络都有很好的效果，但是这些网络之间同样也缺少了一些关联性，比如在 ResNet 中，团队没有考虑使用多个尺寸的卷积核并行提取特征，而是选择在保留当前特征图的同时，对当前特征图进行优化，将优化前后的特征图进行相加输出，又或者，在以优化算法效率为目标之一的 GoogLeNet 中，团队也没有考虑使用多个小卷积核来代替大卷积核。因此，我萌生了创造一个综合这些主流网络优点的网络的想法。

在本文中我将重点讲述一个集多种网络结构与一身的高效深度卷积神经网络，相比经典的卷积神经网络模型，该模型的参数量更少，特征提取手段更多，同时也在一定程度上解决了网络退化的问题，这得益于将残差模块和 inception 模块进行结合。同时我也很重视模型的效率，我认为这是非常重要的，就像

GoogLeNet 中提到的，模型的价值是投入实际应用，而不是成为学术奇观。

2.相关工作

小卷积核叠加：在卷积神经网络结构中，小卷积核叠加是一种通过叠加小卷积核以提高感受野的的方法，图 1 展示了使用两个 3x3 卷积核与一个 5x5 卷积核在感受野上的区别，使用多个较小的 3x3 卷积核堆叠来模拟大卷积核的效果在 vgg 网络中被证明是可行的，同时根据"参数数量=输入维度 x 卷积核面积 x 输出维度"公式可以算出，这种方法在可以显著减少参数数量。

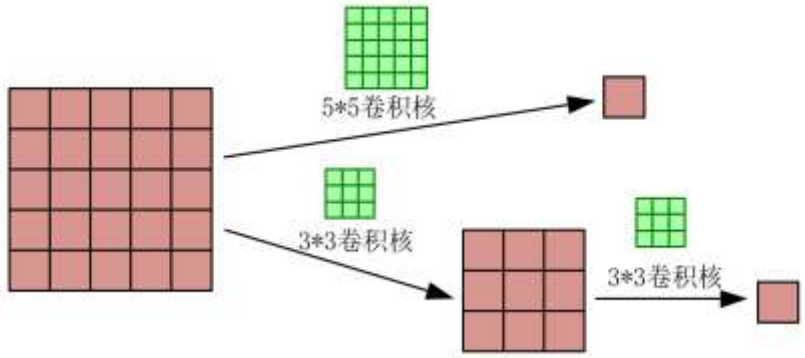


图 1.使用多个 3x3 卷积核在特征提取上与 5x5 卷积核的区别

网络中的网络：该方法是由新加坡国立大学研究人员提出的一种用于提高神经网络表征能力的方法，原文中指出，通过在每次卷积之后添加线性层，以增加模型的非线性，图 2 展示了该方法的实现过程，在实际应用中，可以通过添加 1x1 卷积层来实现。我在模型中大量使用了这种结构。除了增加模型的非线性外，使用该结构还有降维的作用，通过降维可以减少模型的参数和计算量，为后续增加模型的宽度和深度做铺垫。

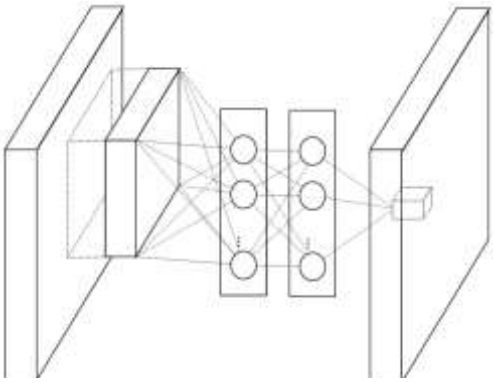


图 2. 通过在卷积层后添加线性层增加非线性

Inception 结构：Inception 架构的主要思想是找出如何在卷积视觉网络中逼近和覆盖一个最优的局部稀疏结构，GoogLeNet 团队在这个问题上使用的实现方式是通过多个不同尺寸的卷积核对图像进行特征提取，同时在特征提取之前增加 1x1 卷积核以减少参数数量和计算量，希望在能捕获到更多的特征的同时

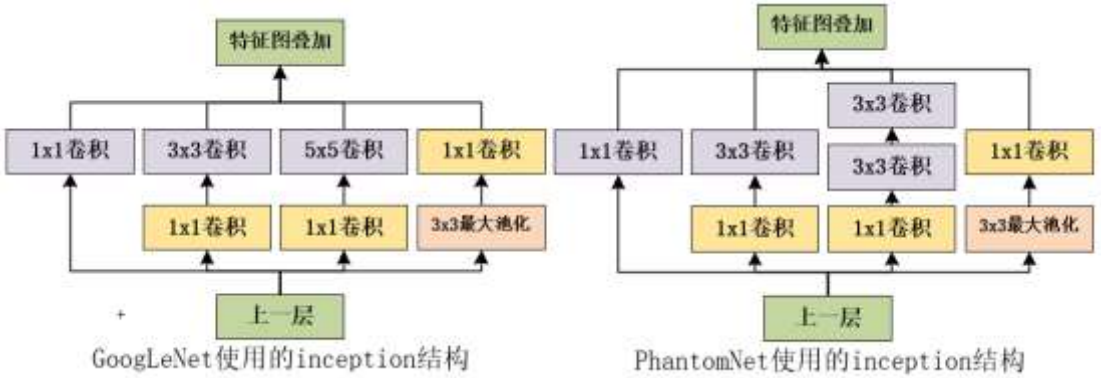


图 3. inception 结构

减少计算量，团队把这个捕获特征的结构成为 inception 结构，我根据小卷积核叠加理论对该结构进行了改进（如图 3 所示），使用多个小卷积核叠加来代替不同尺寸的卷积核，已达到进一步减少参数数量的目的。

残差学习：残差学习指通过捷径连接将输入跳过残差模块，和残差模块的结果一起输出（如图 4 所示），在模型训练过程中，残差模块主要负责在输入的基础上对输入进行优化，使用残差学习在神经网络领域可以可以防止梯度消失并且创建恒等映射，在传统神经网络结构中，恒等映射是很难做到的，同时残差网路有别于传统的串联网络，将残差网络结构展开可以发现，该网络结构为一个串并联结构，经过验证在该结构中，少数模块出现问题对结果的影响不大。

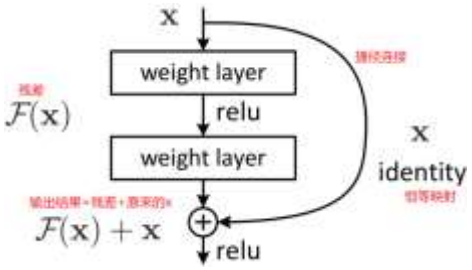


图 4. 残差结构

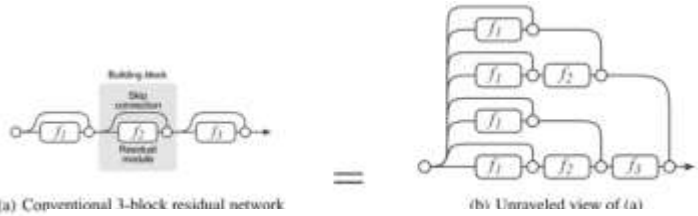


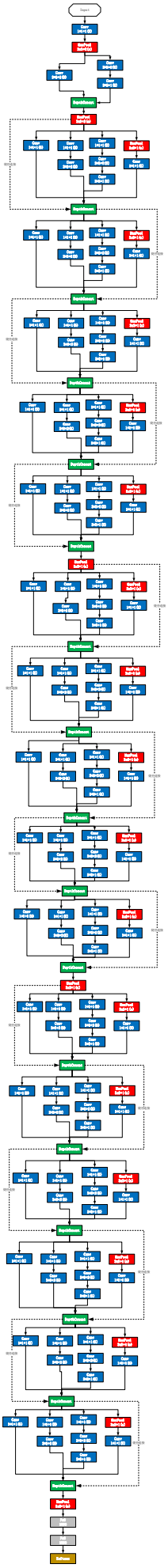
图 5. 残差结构展开

3.Phantom Net

当前阶段模型在 imagenet 数据集上的训练还未完成在这里，为了演示，我要描述模型在 mnist 数据集上的实验情况。

Type	Patch size/ stride	Output size	#1x1	#3x3 reduce	#3x3	#3x3->3x3 reduce	#3x3->3x3	Pool proj	Params
Convolution	1x1/1	224x224x3							0.03K
Max pool	3x3/2	112x112x3							
Convolution(2a)	3x3/2	56x56x64							1.7k
Convolution(2b)	3x3/2	56x56x64							1.7k
Convolution(2b)	3x3/1	56x56x64							18k
将 2a 和 2b 两条线的输出叠加在一起，输出出 56*56*64									
Max poll	3x3/2	28x28x128							
ResInception 集		28x28x256	32	48	64	8	16/16	16	92k
Max pool	3x3/2	14x14x256							
ResInception 集		28x28x256	32	48	64	8	16/16	16	92k
Max pool	3x3/2	7x7x256							
ResInception 集		28x28x256	32	48	64	8	16/16	16	92k
Max pool	3x3/2	4x4x256							
Fc0		1x1x1000				0			
Depout(50%)									
Fc1		1x1x10							10k

图表 1 PhantomNet



所有的卷积，包括 Inception 模块内部的卷积都使用 Relu 激活函数。网络的感受野大小为 224x224，采用灰度图像进行标准化。“#3x3reduce”和“#3x3->3x3reduce”代表在残差 inception 模块中单 3x3 卷积和双 3x3 堆叠卷积之前使用的降维层中使用多个 1x1 滤波器数量，“PoolProj”代表在残差 inception 模块中，内置最大池化投影层中 1x1 滤波器数量，这些降维层也使用 Relu 激活函数，整体网络结构如表 1 所示。

ResInception 集表示多个 ResInception 模块的堆叠，在后续模型命名中我会使用三个 ResInception 集中所有的模块数量为模型命名，在 mnist 数据集中使用的模型就是 PhantomNet6，表示该模型包含六个 ResInception 层。本文提供了 ResInception16 的模型结构图，该模型现在正在用于 imagenet2012 数据的训练。

4.训练方法

我的网络训练环境在 pytorch 环境，使用小批量随机梯度下降算法和 0.5 的动量，学习率调度方式主要包含以下两种：第一种是根据输出的损失情况，手动调整学习率，第二种是固定的学习率调度（每 5 个批次将学习率下降 20%），

分层梯度下降： ResInception 模块可以抽象成传统残差模块，即 y=F(x)+x, 当我们添加一个新的 ResInception 时，对权重 F 进行 0 初始化，这样我们就可以得到 y=x,此时对权重 F 进行梯度下降，尽可能保证了 F(x)对于模型的作用是正面的。

5.MNIST 分类的设置和结果

MNIST 数据集是图像识别常用的数据集只有，训练集包含 6 万张图像，测试集包含 5 万张图像，每个图象都与一个数字关联。我使用 PhantomNet6 模型对 MNIST 数据集进行训练，采用小批量随机梯度下降（对一部分层采用了分层梯度下降），准确率最高到达了 99.46，当然这不是一蹴而就的，最初准确率只有 99.01，经过微调之后才得到了比较好的结果。在小数据集上的效果已经接近 ResNet，与其他经典模型的对比如表 2。

Method	**Top-1 acc**	**model size**
EfficientNet	99.55	20 MB
ResNet	99.50	44 MB
PhantomNet	99.46	4.6MB
DenseNet	99.45	32 MB
GoogLeNet	99.30	92 MB
MobileNet	99.25	14 MB
VGG	99.20	552 MB
AlexNet	98.75	240 MB

图表 2 PhantomNet 模型与传统模型在 mnist 数据集上的表现

6.ImageNet 分类的设置和结果

7.结论

就目前实验结果来说，使用基于 nin 理论构建的 ResInception 模块是改进计算机视觉神经网络的一种可行方法，这种模块的优势在于，它比传统的 res 模块的拟合能力更强，且比传统的 inception 模块收敛更快且参数更少。但是与传统的单分支模块网络相比，inception 模块的收敛速度还是慢一些。虽然该模型的 16 版本还没有完成在 imageNet 上的训练工作，但是我提供的方法是可行的，尽管这些优化方法都不是我提出来的。

10 致谢

在此我要感谢张长宏老师在专业课上对我提供的指导和帮助。同时还要感谢青海民族大学计算机学院为我提供的算力支持。

11 参考文献