ResNet:2015

Deep Residual Learning for Image Recognition

论文出发点或背景

理论上网络越深对特征的提取会更好，但是有时候会产生一些其他问题比如计算量爆炸、梯度弥散或者梯度爆炸，其中一部分因为BN的引入而得到了解决。

实验发现网络层数增加时，精度达到饱和，之后增加层数会导致网络精度降低，网络发生退化，而且这种问题不是过拟合引起的

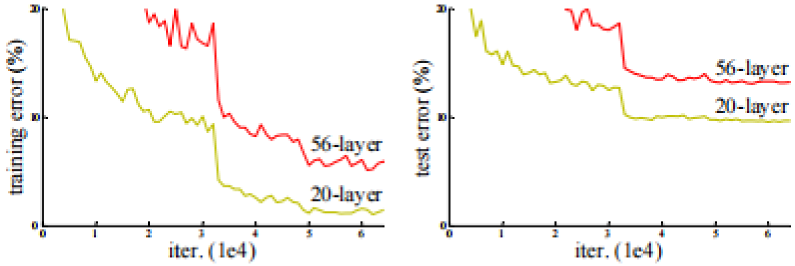


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

论文创新思路

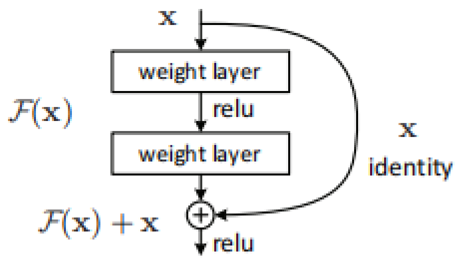


Figure 2. Residual learning: a building block.

引入一个残差结构来解决网络的退化问题，identify是恒等映射，假设无残差结构最后要输出的是$H(x)$，引入残差结构的恒等映射之后就会使得最后一层学习的目标变为了$F(x) = H(x) - x$。网络的输出变为了$F(x) + x$。

如果浅层的解已经是最优的话，那么非线性层的权值就会接近于0，这样整个模块的输出就接近于上一层的输出，这就保证了增加了层数，最起码保证不会让网络的性能变更差

之前的工作：

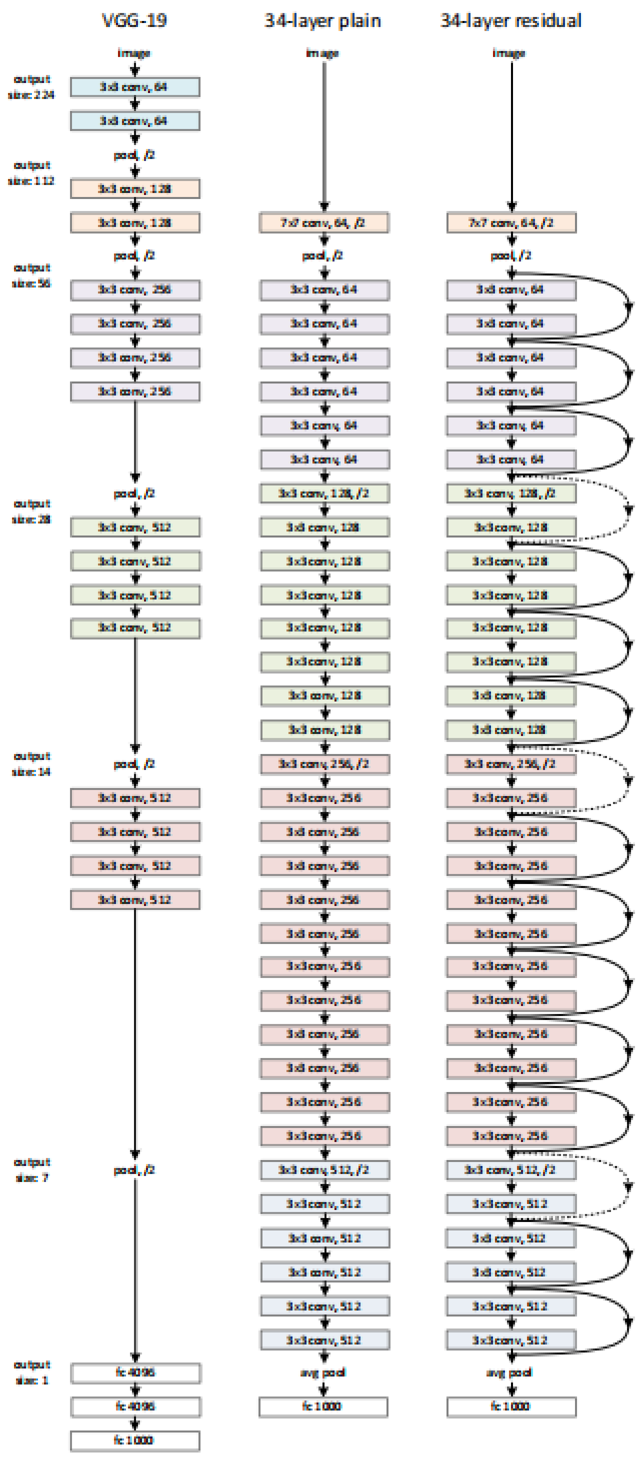
残差表示：在低级视觉和计算机图形学中，为了解决偏微分方程的问题，广泛使用多重网格，将问题分解为多个子问题，每个子问题得到的是粗粒度和细粒度的差值。多重网格的另一种替代方案是分层预处理，依赖于表示两个尺度之间的残差的向量。

跳跃连接的方式：

在Highway networks中通过具有门控函数的支路进行跳连，但是我们的残差网络是通过恒等映射实现的跳连，所以没有计算量，是Highway Network的一种特殊情况。

论文方法的大概介绍

（1）受到VGGNet的启发，卷积层大多都是3×3卷积核并且遵循两个设计规则：对于相同尺寸的输出特征图大小，各层有着相同数量的卷积核；如果特征图尺寸缩小一半，卷积核的数量就要加倍，以保证每层的时间复杂度，我们通过步幅为2的卷积层进行降采样，该网络以一个全剧平均池化层核一个softmax的全连接层结束，中间的加权层层数为34层



（2）左边的设计的基础模块是为了让特征图减半的同时，图像的通道数逐渐增加，右边是为了降低网络的时间复杂度设计出来的模块，先1×1卷积降维，然后3×3卷积提取特征，最后1×1卷积升维。

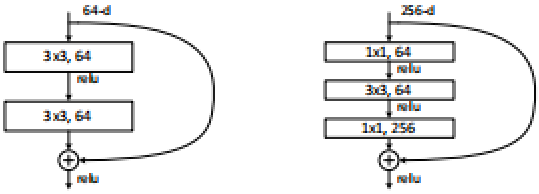


Figure 5. A deeper residual function $\mathcal{F}$ for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a "bottleneck" building block for ResNet-50/101/152.

实际效果

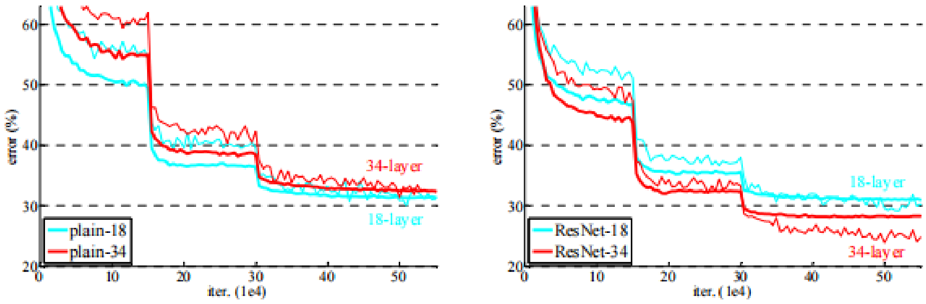


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	**25.03**

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

Table 3. Error rates (%, **10-crop** testing) on ImageNet validation. VGG-16 is based on our test. ResNet-50/101/152 are of option B that only uses projections for increasing dimensions.

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

method	top-5 err. (**test**)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PReLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

Table 5. Error rates (%) of **ensembles**. The top-5 error is on the test set of ImageNet and reported by the test server.

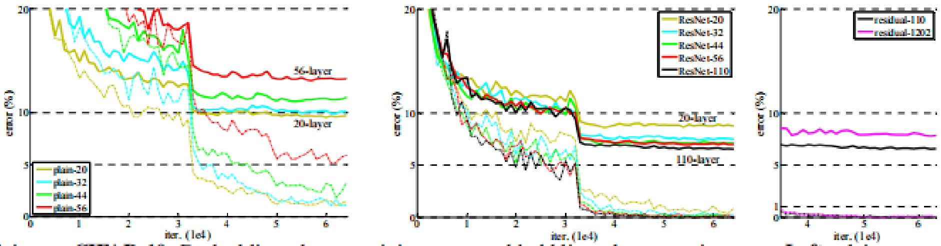


Figure 6. Training on **CIFAR-10**. Dashed lines denote training error, and bold lines denote testing error. **Left**: plain networks. The error of plain-110 is higher than 60% and not displayed. **Middle**: ResNets. **Right**: ResNets with 110 and 1202 layers.

training data	07+12	07++12
test data	VOC 07 test	VOC 12 test
VGG-16	73.2	70.4
ResNet-101	**76.4**	**73.8**

Table 7. Object detection mAP (%) on the PASCAL VOC 2007/2012 test sets using **baseline** Faster R-CNN. See also Table 10 and 11 for better results.

metric	mAP@.5	mAP@[.5, .95]
VGG-16	41.5	21.2
ResNet-101	**48.4**	**27.2**

Table 8. Object detection mAP (%) on the COCO validation set using **baseline** Faster R-CNN. See also Table 9 for better results.

个人对这篇论文的理解

1.残差模型之前在Highway Network中也有提到，ResNet可以看作是Highway Network的一个特殊情况，去掉了门控模块，通过恒等映射的方式在优化时会变得更加容易，同时Highway Network的假设空间相比ResNet更复杂一点，因为Highway Network需要寻找最适合数据的超参数、

2.出发点还是在加深网络的深度的时候，我的网络表现如果不能表现更好，最起码不能比浅层差，所以通过恒等映射保证了浅层的效果得以保留，之后通过对主干支路学习到的加权得到$F(x)$，如果浅层效果已经最优了，那么$F(x)$应该是趋向于0的

3.作者尝试了优化一千层以上的残差网络，最后发现还不如100多层网络的表现，认为这时候是出现了一定的过拟合，需要大量的数据加更强的正则化方式