

残差网络 (Residual Networks, ResNets)

1. 什么是残差 (residual) ?

“残差在数理统计中是指实际观察值与估计值（拟合值）之间的差。”“如果回归模型正确的话， 我们可以将残差看作误差的观测值。”

更准确地，假设我们想要找一个 x ，使得 $f(x) = b$ ，给定一个 x 的估计值 x_0 ，残差 (residual) 就是 $b - f(x_0)$ ，同时，误差就是 $x - x_0$ 。

即使 x 不知道，我们仍然可以计算残差，只是不能计算误差罢了。

2. 什么是残差网络 (Residual Networks, ResNets) ?

在了解残差网络之前，先了解下面这个问题。

Q1：神经网络越深越好吗？（Deeper is better?）

A1：如图 1 所示，在训练集上，传统神经网络越深效果不一定越好。而 [Deep Residual Learning for Image Recognition](#) 这篇论文认为，理论上，可以训练一个 shallower 网络，然后在这个训练好的 shallower 网络上堆几层 identity mapping（恒等映射）的层，即输出等于输入的层，构建出一个 deeper 网络。这两个网络（shallower 和 deeper）得到的结果应该是一模一样的，因为堆上去的层都是 identity mapping。这样可以得出一个结论：**理论上，在训练集上，Deeper 不应该比 shallower 差，即越深的网络不会比浅层的网络效果差。**但为什么会出现图 1 这样的情况呢，随着层数的增多，训练集上的效果变差？这被称为**退化问题 (degradation problem)**，原因是随着网络越来越深，训练变得原来越难，网络的优化变得越来越难。理论上，越深的网络，效果应该更好；但实际上，由于训练难度，过深的网络会产生退化问题，效果反而不如相对较浅的网络。而残差网络就可以解决这个问题的，残差网络越深，训练集上的效果会越好。（测试集上的效果可能涉及过拟合问题。过拟合问题指的是测试集上的效果和训练集上的效果之间有差距。）

公告

昵称：wuliytTaotao
园龄：1年
粉丝：18
关注：9
[+加关注](#)

<		2019年3月						>		
日	一	二	三	四	五	六				
24	25	26	27	28	1	2				
3	4	5	6	7	8	9				
10	11	12	13	14	15	16				
17	18	19	20	21	22	23				
24	25	26	27	28	29	30				
31	1	2	3	4	5	6				

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

随笔分类(40)

deep learning(8)
Git(1)
LaTeX(1)
Linux(3)
Mac(3)
machine learning(14)
Mendeley(1)
python(5)
Server
Shadowsocks(1)
sklearn(1)
tensorflow(2)

随笔档案(26)

2019年3月 (2)
2019年1月 (2)
2018年12月 (1)
2018年10月 (1)
2018年9月 (4)
2018年8月 (12)
2018年7月 (4)

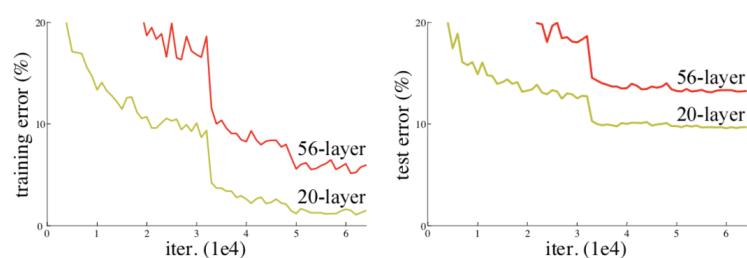


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

图 1 不同深度的传统神经网络效果对比图

(“plain” network指的是没有使用 shortcut connection 的网络)

残差网络通过加入 shortcut connections，变得更加容易被优化。包含一个 shortcut connection 的几层网络被称为一个残差块（residual block），如图 2 所示。

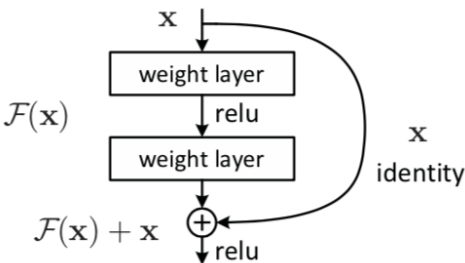


图 2 残差块

2.1 残差块 (residual block)

如图 2 所示， x 表示输入， $F(x)$ 表示残差块在第二层激活函数之前的输出，即 $F(x) = W_2\sigma(W_1x)$ ，其中 W_1 和 W_2 表示第一层和第二层的权重， σ 表示 ReLU 激活函数。（这里省略了 bias。）最后残差块的输出是 $\sigma(F(x) + x)$ 。

当没有 shortcut connection（即图 2 右侧从 x 到 \oplus 的箭头）时，残差块就是一个普通的 2 层网络。残差块中的网络可以是全连接层，也可以是卷积层。设第二层网络在激活函数之前的输出为 $H(x)$ 。如果在该 2 层网络中，最优的输出就是输入 x ，那么对于没有 shortcut connection 的网络，就需要将其优化成 $H(x) = x$ ；对于有 shortcut connection 的网络，即残差块，最优输出是 x ，则只需要将 $F(x) = H(x) - x$ 优化为 0 即可。后者的优化会比前者简单。这也是残差这一叫法的由来。

2.2 残差网络举例

图 3 最右侧就是就是一个残差网络。34-layer 表示含可训练参数的层数为 34 层，池化层不含可训练参数。图 3 右侧所示的残差网络和中间部分的 plain network 唯一的区别就是 shortcut connections。这两个网络都是当 feature map 减半时，filter 的个数翻倍，这样保证了每一层的计算复杂度一致。

文章分类(7)

Linux(2)
machine learning(1)
python(1)
Server(2)
Shadowsocks(1)

文章档案(4)

2018年9月 (2)
2018年8月 (2)

相关链接

GitHub

积分与排名

积分 - 18897
排名 - 28923

最新评论

1. Re:贝叶斯深度学习 (bayesian deep learning)
thanks，之前看到深度网络中的采样一直没明白到底是在做什么，你这个文章给了我一个理解的突破口！
--买白菜不用券
2. Re:【机器学习之数学】01
导数、偏导数、方向导数、梯度写得好，帮我厘清了多元函数的方向导数和梯度的概念
--eryar
3. Re:【机器学习之数学】01
导数、偏导数、方向导数、梯度这书不错，看目录正是我想找的
--会长
4. Re:漫谈“采样” (sampling)
博客园的一股清流，一看你就是学数学的~哈哈
--溪源More
5. Re:【神经网络基础】为什么神经网络选择了“深度”？
@wuliytTaotao那肯定超级耗时。感觉就是在搞科研。...
--kiba518

阅读排行榜

1. 残差网络 (Residual Networks, ResNets) (5454)

ResNet 因为使用 identity mapping, 在 shortcut connections 上没有参数, 所以图 3 中 plain network 和 residual network 的计算复杂度都是一样的, 都是 3.6 billion FLOPs.

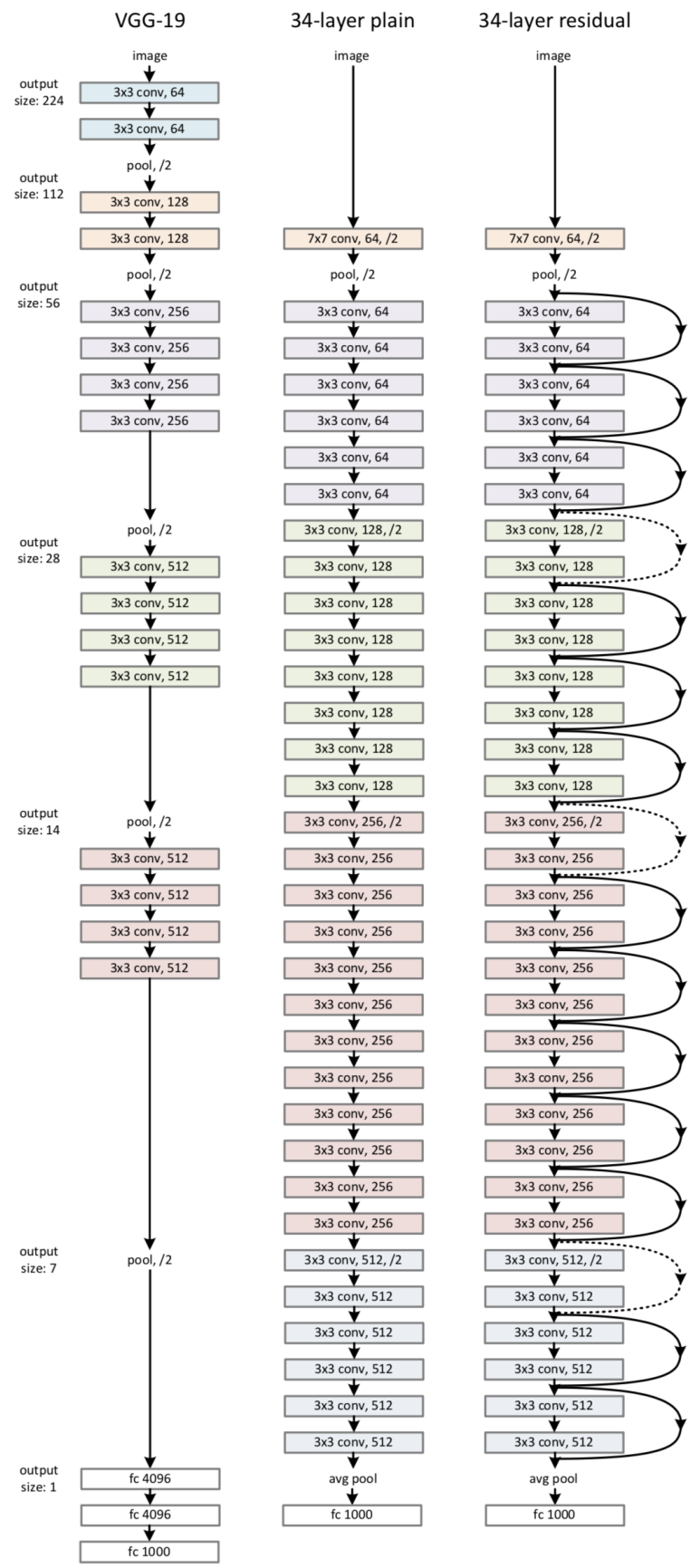


图 3 VGG-19、plain network、ResNet

- 2. 理解滑动平均(exponential moving average)(3925)
- 3. 【python3基础】相对路径, \', \./, \../(3293)
- 4. 经典卷积神经网络结构——LeNet-5、AlexNet、VGG-16 (1148)
- 5. 【TensorFlow基础】tf.add 和 tf.nn.bias_add 的区别(996)
- 6. 【神经网络基础】为什么神经网络选择了“深度”? (956)
- 7. 贝叶斯深度学习 (bayesian deep learning) (915)
- 8. Mac查看和杀死后台进程(605)
- 9. 关于pip安装时提示"pkg_resources.DistributionNotFound"错误(605)
- 10. 【机器学习基础】熵、KL散度、交叉熵(502)

评论排行榜

- 1. 【神经网络基础】为什么神经网络选择了“深度”? (4)
- 2. 【机器学习之数学】01 导数、偏导数、方向导数、梯度 (2)
- 3. 贝叶斯深度学习 (bayesian deep learning) (1)
- 4. 漫谈“采样” (sampling) (1)

推荐排行榜

- 1. 【机器学习之数学】01 导数、偏导数、方向导数、梯度 (5)
- 2. 漫谈“采样” (sampling) (3)
- 3. 理解滑动平均(exponential moving average)(2)
- 4. 贝叶斯深度学习 (bayesian deep learning) (2)
- 5. 【神经网络基础】为什么神经网络选择了“深度”? (1)
- 6. 卷积神经网络 (Convolutional Neural Network, CNN) (1)

残差网络可以不是卷积神经网络，用全连接层也可以。当然，残差网络在被提出的论文中是用来处理图像识别问题。

2.3 为什么残差网络会work?

我们给一个网络不论在中间还是末尾加上一个残差块，并给残差块中的 weights 加上 L2 regularization (weight decay)，这样图 1 中 $F(x) = 0$ 是很容易的。这种情况下加上一个残差块和不加之前的效果会是一样，所以加上残差块不会使得效果变得差。如果残差块中的隐藏单元学到了一些有用信息，那么它可能比 identity mapping (即 $F(x) = 0$) 表现的更好。

"The main reason the residual network works is that it's so easy for these extra layers to learn the identity function that you're kind of guaranteed that it doesn't hurt performance. And then lot of time you maybe get lucky and even helps performance, or at least is easier to go from a decent baseline of not hurting performance, and then creating the same can only improve the solution from there."

References

[残差 百度百科](#)

[Residual \(numerical analysis\) - Wikipedia](#)

[Course 4 Convolutional Neural Networks by Andrew Ng](#)

[He, K., Zhang, X., Ren, S., & Sun, J. \(2015\). Deep Residual Learning for Image Recognition, 1–9. Retrieved from <http://arxiv.org/abs/1512.03385>](#)

作者: [wuliytTaotao](#)

出处: <https://www.cnblogs.com/wuliytTaotao/>

本作品采用[知识共享署名-非商业性使用-相同方式共享 4.0 国际许可协议](#)进行许可，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接。

分类: deep learning , machine learning

标签: machine learning , deep learning

好文要顶

关注我

收藏该文



[wuliytTaotao](#)

关注 - 9

粉丝 - 18

[+加关注](#)

0

0

« 上一篇: [【人工智能神经网络基础】为什么神经网络选择了“深度”?](#)

» 下一篇: [【机器学习基础】熵、KL散度、交叉熵](#)

发表评论

昵称:

评论内容:

B

提交评论

退出

订阅评论

[Ctrl+Enter快捷键提交]

- 【幸运】99%的人不知道我们有可以帮你薪资翻倍的秘笈！
- 【推荐】超50万C++/C#源码：大型实时仿真组态图形源码
- 【推荐】百度云“猪”你开年行大运，红包疯狂拿
- 【推荐】55K刚面完Java架构师岗，这些技术你必须掌握

相关博文:

· ad-hoc networks

· 卷积神经网络Convolutional Neural Networks

· 计算机科学类优秀书籍推荐 Networks (网络)

· Bayesiannetworks

· Communication Networks

- 最新新闻:
- 优化至死：多数人都在努力过上坏日子

· 苹果CEO库克：感谢中国打开了大门，我们的未来密不可分

· 美国拟调查科技公司收集数据行为 谷歌FB亚马逊压力大增

· “非洲之王”传音控股闯关，能否受科创板青睐？

· 马斯克律师反驳SEC：产量推文谈不上特斯拉“重大披露”

» 更多新闻...