148 | 计算机视觉领域的深度学习模型（三）：ResNet

2018-09-12 洪亮劼

AI技术内参

进入课程 >

讲述：初明明
时长 05:36　大小 2.57M

今天我们继续来讨论经典的深度学习模型在计算机视觉领域应用。今天和你分享的论文是《用于图像识别的深度残差学习》（Deep Residual Learning for Image Recognition）[1]。这篇论文获得了 CVPR 2016 的最佳论文，在发表之后的两年间里获得了超过 1 万 2 千次的论文引用。

论文的主要贡献

我们前面介绍 VGG 和 GoogleNet 的时候就已经提到过，在深度学习模型的前进道路上，一个重要的研究课题就是**神经网络结构究竟能够搭建多深。**

这个课题要从两个方面来看：第一个是现实层面，那就是如何构建更深的网络，如何能够训练更深的网络，以及如何才能展示出更深网络的更好性能；第二个是理论层面，那就是如何

真正把网络深度，或者说是层次度，以及网络的宽度和模型整体的泛化性能直接联系起来。

在很长的一段时间里，研究人员对神经网络结构有一个大胆的预测，那就是更深的网络架构能够带来更好的泛化能力。但是要想真正实现这样的结果其实并不容易，我们都会遇到哪些挑战呢？

一个长期的挑战就是**模型训练时的梯度“爆炸”（Exploding）或者“消失”（Vanishing）**。为了解决这个问题，在深度学习研究刚刚开始的一段时间，就如雨后春笋般爆发出了很多技术手段，比如“线性整流函数”（ReLu），“批量归一化”（Batch Normalization），“预先训练”（Pre-Training）等等。

另外一个挑战是在 VGG 和 GoogleNet 的创新之后，大家慢慢发现**单纯加入更多的网络层次其实并不能带来性能的提升**。研究人员有这样一个发现：当一个模型加入到 50 多层后，模型的性能不但没有提升，反而还有下降，也就是模型的准确度变差了。这样看，好像模型的性能到了一个“瓶颈”。那是不是说深度模型的深度其实是有一个限度的呢？

我们从 GoogleNet 的思路可以看出，网络结构是可以加深的，比如对网络结构的局部进行创新。而这篇论文，就是追随 GoogleNet 的方法，在网络结构上提出了一个新的结构，叫“**残差网络**”（Residual Network），简称为 **ResNet**，从而能够把模型的规模从几层、十几层或者几十层一直推到了上百层的结构。这就是这篇文章的最大贡献。

从模型在实际数据集中的表现效果来看，ResNet 的错误率只有 VGG 和 GoogleNet 的一半，模型的泛化能力随着层数的增多而逐渐增加。这其实是一件非常值得深度学习学者振奋的事情，因为它意味着深度学习解决了一个重要问题，突破了一个瓶颈。

论文的核心方法

那这篇论文的核心思想是怎样的呢？我们一起来看。

我们先假设有一个隐含的基于输入 x 的函数 H。这个函数可以根据 x 来进行复杂的变换，比如多层的神经网络。然而，在实际中，我们并不知道这个 H 到底是什么样的。那么，传统的解决方式就是我们需要一个函数 F 去逼近 H。

而这篇文章提出的“残差学习”的方式，就是不用 F 去逼近 H，而是去逼近 H(x) 减去 x 的差值。在机器学习中，我们就把这个差值叫作“残差”，也就是表明目标函数和输入之间的差距。当然，我们依然无法知道函数 H，在实际中，我们是用 F 去进行残差逼近。

F(x)=H(x)-x，当我们把 x 移动到 F 的一边，这个时候就得到了残差学习的最终形式，也就是 F(x)+x 去逼近未知的 H。

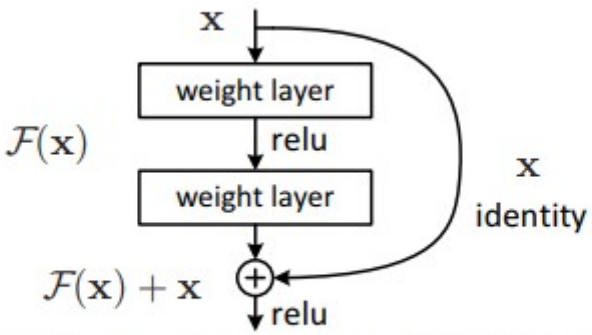

Figure 2. Residual learning: a building block.

我们引用论文中的插图来看这个问题，就会更加直观。（图片来源：https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf ）

在这个公式里，外面的这个 x 往往也被称作是“捷径”（Shortcuts）。什么意思呢？有学者发现，在一个深度神经网络结构中，有一些连接或者说层与层之间的关联其实是不必要的。我们关注的是，**什么样的输入就应当映射到什么样的输出**，也就是所谓的“等值映射”（Identity Mapping）。

遗憾的是，如果不对网络结构进行改进，模型无法学习到这些结构。那么，构建一个从输入到输出的捷径，也就是说，从 x 可以直接到 H（或者叫 y），而不用经过 F(x)，在必要的时候可以强迫 F(x) 变 0。也就是说，捷径或者是残差这样的网络架构，在理论上可以帮助整个网络变得更加有效率，我们希望算法能够找到哪些部分是可以被忽略掉的，哪些部分需要保留下来。

在真实的网络架构中，作者们选择了在每两层卷积网络层之间就加入一个捷径，然后叠加了 34 层这样的架构。从效果上看，在 34 层的时候 ResNet 的确依然能够降低训练错误率。于是，作者们进一步尝试了 50 多层，再到 110 层，一直到 1202 层的网络。最终发现，在 110 层的时候能够达到最优的结果。而对于这样的网络，所有的参数达到了 170 万个。

为了训练 ResNet，作者们依然使用了批量归一化以及一系列初始化的技巧。值得一提的是，到了这个阶段之后，作者们就放弃了 Dropout，不再使用了。

小结

今天我为你讲了一篇经典论文，提出了 ResNet，残差网络这个概念，是继 VGG 和 GoogleNet 之后，一个能够大幅度提升网络层次的深度学习模型。

一起来回顾下要点：第一，我们总结归纳了加深网络层次的思路以及遇到的挑战；第二，我们讲了讲残差网络的概念和这样做背后的思考以及在实际应用中的一些方法。

最后，给你留一个思考题，从 AlexNet 到 VGG、GoogleNet，再到 ResNet，除了网络深度加深以外，模型进化过程中是否还有一些地方也让你有所感触？

欢迎你给我留言，和我一起讨论。

参考文献

1. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. **Deep Residual Learning for Image Recognition**. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

上一篇　147 | 计算机视觉领域的深度学习模型（二）：VGG & GoogleNet

下一篇　149 | 计算机视觉高级话题（一）：图像物体识别和分割

精选留言 (3)

写留言

sky
2018-09-12

1

我还有个大胆地猜想，在几何领域，有保角映射和等距离映射这样的反应几何特性的映射，如果我想要神经网络提高对这些特征的识别，是否可以把输入做保角映射或者等距离映射，然后作为残差网络的捷径

展开

sky
2018-09-12

1

我能不能这样理解，resnet的捷径其实就是给网络加了一个线性因子，resnet其实就是线性和非线性的组合达到了这样的效果，其实我还是不太明白作者为什么回想到用去逼近残差，逼近残差在其他地方有类似的应用吗

展开

Andy
2018-09-15

老师 为什么层数多了之后就不用dropout了呢?

展开