Deep Residual Learning for Image Recognition

Abstract

Introduction

Related work

Deep residual learning

Experiment

Abstract

当网络深度增加后，会变得很难训练。本文提出一种残差网络（Residual Network），将网络的层重新组合为新的学习残差网络，本文的经验和证据表明这些残差网络更容易优化，并且可以从显著增加的深度获得准确性

Introduction

深度卷积神经网络[22,21]为图像分类带来了一系列突破[21,50,40]。深度网络自然地以端到端多层方式集成低/中/高级特征[50]和分类器，并且可以通过堆叠层的数量（深度）来丰富特征的“级别”。最近的证据[41,44]揭示了网络深度至关重要，而具有挑战性的ImageNet数据集[36]的主要结果[41,44,13,16]都利用了“非常深”的[41]模型，深度为十六[41]至三十[16]。许多其他非平凡的视觉识别任务[8,12,7,32,27]也从非常深的模型中得到了极大的好处。

在深度显着的推动下，出现了一个问题：学习更好的网络就像堆叠更多层一样容易吗？回答这个问题的一个障碍是消失/爆炸梯度的臭名昭着的问题[1,9]，它从一开始就阻碍了收敛。然而，这个问题在很大程度上通过归一化初始化[23,9,37,13]和中间归一化层[16]来解决，这使得具有数十层的网络能够开始收敛以用于随机梯度下降（SGD）和反向传播[22] ]。

当更深的网络能够开始收敛时，就会出现降级问题：随着网络深度的增加，准确度变得饱和（这可能不足为奇），然后迅速降级。出乎意料的是，这种退化不是由过度配置引起的，如果在适当的深度模型中增加更多的层会导致更高的训练误差，如[11,42]所述，并且我们的实验进行了彻底的验证。图1显示了一个典型的例子。

降级（训练精度）表明并非所有系统都易于优化。让我们考虑一个较浅的架构及其更深的对应物，在其上添加更多层。对于更深层次的模型，存在构造的解决方案：添加的层是身份映射，并且从学习的浅层模型复制其他层。这种构造的解决方案的存在表明，较深的模型不应该产生比较浅的对应物更高的训练误差。但实验表明，我们现有的解决方案无法找到解决方案