用于图像识别的深度残差网络

摘要：

深度神经网络很难训练。我们提出了一种残差学习框架来减轻那些本质上只是比以前所使用的更深的网络的训练。我们明确的将各层重新表示为学习关于输入曾的残差函数，而不是学习与其无关的函数。我们提供了综合的经验证据表明残差网络易于优化，并能从显著增加的深度中获得准确性。在ImageNet数据集上我们评估了一个152层的残差网络——比VGG网络深8倍，但仍然具有更低的复杂度。这些残差网络的集合在ImageNet测试集上达到了3.57%误差。该结果也赢得了ILSVRC2015分类任务的第一名。我们也提出了在CIFAR-10上使用100层和1000层的分析。

表示的深度对视觉识别任务至关重要。仅仅根据我们深度的表示，我们在COCO物体检测数据集上获得了28%的相对改进。深度残差网络是我们在ILSVRC以及COCO2015比赛中提交结果的基础，在其上我们赢得了ImageNet检测，ImageNet定位，COCO检测以及COCO分割任务的第一名。

1. 简介

深度卷积神经网络致使了图片分类中一系列的突破。深度网络自然的整合了低/中/高级的特征，分类器以一种端到端多层的方式并且特征的“水平”可以被众多堆叠的层（深度）所补足。最近的证据表明网络的深度至关重要并致使在ImageNet数据集上的结果都研究“非常深”的模型，深度从16到30不等。许多其他不平常的视觉识别任务也同样从深度模型中受益匪浅。

受到深度重要性的驱使，一个问题出现：*是否学习更好的网络和堆叠更多的层一样简单？*回答该问题的一个阻碍就是臭名昭著的梯度消失/爆炸问题，该问题妨碍了最开始的收敛。这个问题