

ResNet

[论文链接](#)

Introduction

分割线上为第一次学ResNet的理解

残差网络使得神经网络加入更多的层最起码**不会变差**。

假如第N层从N-1层什么都没学到，但最起码还是能回到N-1层。你可以带着N-1层的有效知识去N+1层训练。

假如你现在的知识储量为100（抽象一下），你新学了一个知识可能会让你对以前的知识理解更加深，也可能让你知识乱套了。

像我考研时候二重积分学到了一个叫雅可比行列式的东西，这个东西是二重积分换元，

用着用着就只会这个办法了，最基础的什么轮换对称性都手生了，这就是知识乱套了。

残差网络的作用就是就算你没学会新方法，最起码也能保护老方法的能力不变。

再说白点，这一区块的神经网络对训练有用就保留，没用就相当于跳过。

基于这个特性，所以可以做**很深很深**的神经网络。

$$R(x) = F_2(F_1(x)) + x$$

$$R_n(x) = F_{2,n}(F_{1,n}(R_{n-1}(x))) + R_{n-1}(x)$$

$$R_n(x) = F_{2,n}(F_{1,n}(F_{2,n-1}(F_{1,n-1}(\dots + R_1(x)))) + R_{n-1}(x)$$

这里的每一层函数可能代表了一个残差块，残差块一般包含卷积池化批量正则激活函数。

深度神经网络好在可以加很多层把网络变得特别深，然后不同程度的层会得到不同等级的feature，比如低级的视觉特征或者是高级的语义特征。

随着网络越来越深，梯度就会出现爆炸或者消失。20层比56层loss更低（贴图了，比AlexNet好，要学习这点）

这个问题虽然可以通过优化初始化权重和中间规范化层缓解，但也只是收敛。

但是网络深的时候虽然能收敛，但是精度会变差。（离谱的是这种精度变差不是因为过拟合引起的，我认为越深噪声训练的越严重，不必深究，就是变差了。）

Related Work

学科交叉。

Deep Residual Learning

残差函数的引入、动机、优势：说残差函数好。

残差函数基本构建、操作实现、维度匹配灵活：说残差函数好。

和VGG对比：我的层数多，我的训练量小，我的速度比你快，我的结果比你不好。

训练细节：裁剪方式、像素变化、批量大小等流水账。

测试过程：10-crop和多尺度评估：减少空间偏差、增强鲁棒性（认识图像的全局，不外协）

Experiments

介绍了不同批量的卷积参数设计，单纯介绍，不讲解怎么来的，估计也是瞎试的，炼丹。

损失一不动学习率就打一折，现在不可取，早了后期无力，晚了训练噪声。

总的来说，残差函数很好很有用。

残差连接如何处理输入和输出的形状

- 最简单粗暴的：小的添0使形状对应上大的.....
- 其次：输入输出不同时候做投影
- 离谱的：一直做投影

差距微乎其微，作者选择了第二种.....

更加深更多层数的神经网络就是增加通道数（64-256），但是计算量相当于平方通过 1×1 卷积先降维再投影，虽然通道数增加了，但时间差不多。

层数肯定越多越好，但是1000层的神经网络里通常前100层起了99%的作用.....

CVPR限制8页，没有总结.....End.