《Deep Residual Learning for Image Recognition》

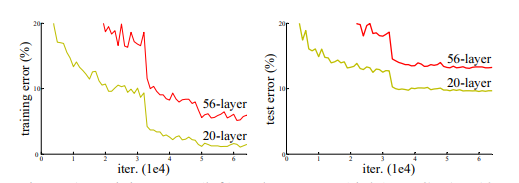
**摘要：**

1. 本论文解决的问题：更深的神经网络很难训练，出现**“退化”问题**。
2. 所提出的解决方法：**深度残差学习(Deep Residual Learning)**框架，它使深度神经网络更容易训练。（通过对网络中的层对层输入的残差函数进行学习）
3. 模型的比赛结果：2015年ImageNet分类任务中，以3.57%的错误率，获得第一名。
4. 研究分析：在CIFAR-10数据集上进行了100层和1000层网络的实验分析。
5. 其他：ImageNet检测任务，定位任务，COCO检测和分割任务取得第一。

**问题背景：**

堆叠更多层数以后的网络（深度网络）是否学习效果更好？

1. 梯度爆炸、梯度消失问题（当时已解决的问题）：可以通过归一化初始化或中间层归一化来解决。
2. “退化”问题——当网络开始收敛时，随着网络深度的增加，准确率趋近饱和，然后迅速下降，而原因并不是过拟合造成的。更深的模型反而会有更高的训练误差，如下图：

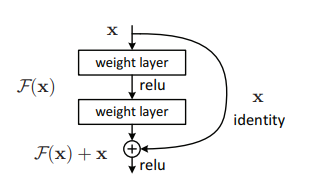


图中，训练误差和测试误差在更深的网络上呈现出更高的误差率。

而作者们认为，随着模型深度的加深，学习能力的增强，更深的模型不应当产生比它更浅的模型更高的错误率。并将此问题归结于**一种优化难题**——即当模型变复杂时，SGD的优化变得更加困难，导致了模型达不到好的学习效果。

**针对“退化”的解决方法：**

“深度残差学习”框架：Residual Learning结构，如下图所示：



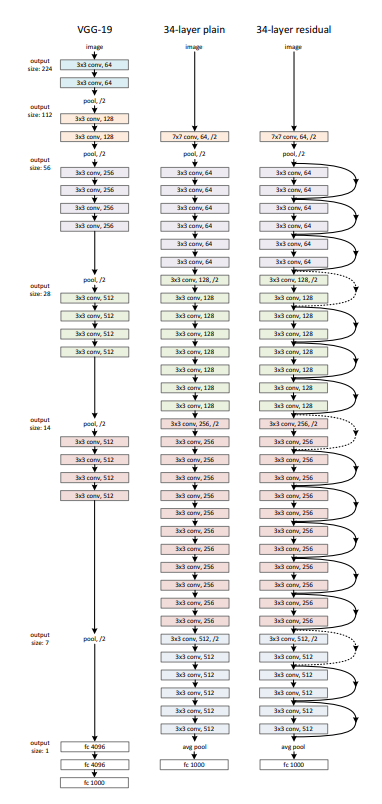
即在原始的层与层连接上增加一个identity mapping（恒等映射），将原始所需要学的函数转换成，其中。新的非线性网络层用来拟合，也称**残差映射**，最终的结果由和简单相加得到，相加的处理方式为**跳跃连接**。（这就是整篇论文的核心思想。）

**关于ResNet，“深度残差学习”的实现：**

1. 设计原则：

（1）对于相同的输出特征图尺寸，卷积层具有相同数量的卷积核；

（2）如果特征图尺寸减半，则卷积核数量加倍，以便保持每层的时间复杂度

2. **总体网络结构**

如左图，以ResNet-34为例：

——卷积层：卷积核大小为7 × 7，卷积核个数为64，步长为2（共计1层）；

——池化层：3 × 3最大池化层，步长为2（共计1层）；

——3个残差连接块：每一个连接块由两层卷积网络组成，卷积核大小为3 × 3，卷积核个数为64（共计6层）；

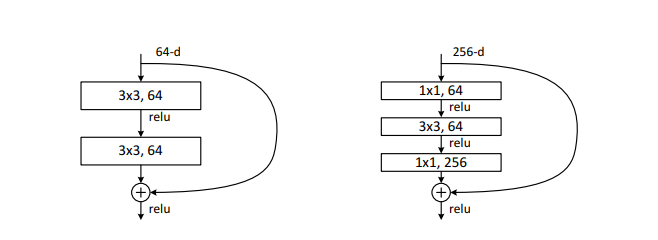
——4个残差连接块：每一个连接块由两层卷积网络组成，卷积核大小为3 × 3，卷积核个数为128（共计8层）；

——6个残差连接块：每一个连接块由两层卷积网络组成，卷积核大小为3 × 3，卷积核个数为256（共计12层）；

——3个残差连接块：每一个连接块由两层卷积网络组成，卷积核大小为3 × 3，卷积核个数为512（共计6层）；

以上总计（层）网络层，而最后的输出结果由全局平均池化层和的1000维度的全连接层得到。

3. 各连接块的实现细节；



如图所示，对于低维度特征（如64×64），采用两层残差结构；对于高纬度特征（如256×256），采用三层残差结构，也称为“bottleneck”**（说明：三层残差结构的方法主要用在构建更深层的神经网络上）**。

4. 针对不同维度的卷积层的残差连接；

**shortcut方法**，提出了三种方式：

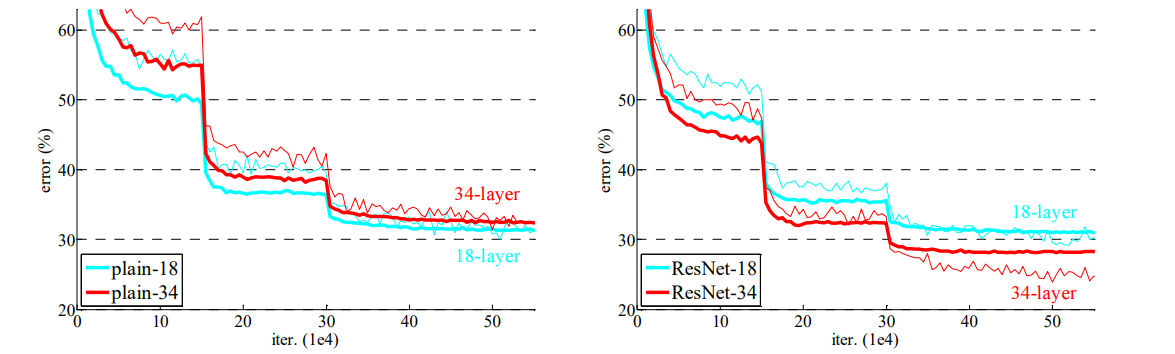
（A）使用恒等映射，如果residual block的输入输出维度不一致，则对增加的维度用0来填充；（零填充）

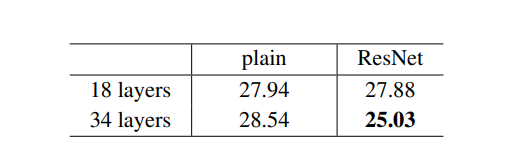
（B）在residual block输入输出维度一致时使用恒等映射，不一致时使用线性投影以保证维度一致；即使用1×1的卷积核来匹配和两者的维度；

（C）对所有的residual block均使用线性投影。

在对上述三个方法进行实验分析后，发现虽然C的效果好于B的效果好于A的效果，但是差距很小，因此线性投影并不是必需的。（作者建议使用B方式）

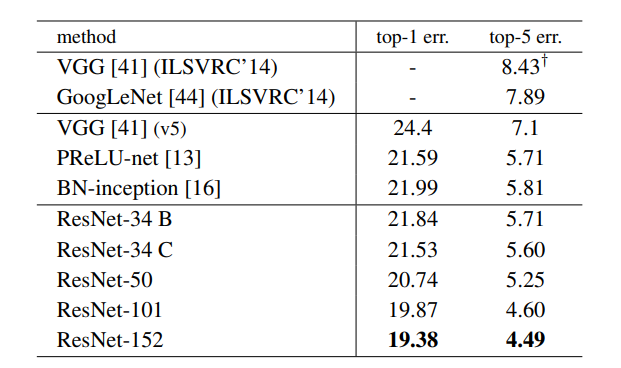
**实验结果与分析：**

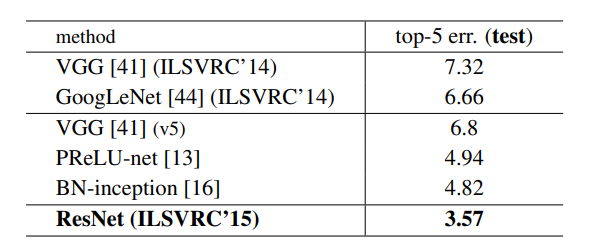




通过对比可以看到：

1. ResNet网络的层数越深，训练误差越小，间接证明“退化”问题可以通过残差学习得到解决；
2. 与plain-34网络相比，训练误差下降了3.5%，且随着网络深度的不断增加，网络性能进一步提高；
3. 与palin-18/34网络相比，残差网络收敛速度更快；





通过对比可以看到，随着网络深度的不断增加，错误率不断下降，同时在训练过程中也没有出现退化现象，且在单个模型上取得4.49%的错误率。而在ImageNet 2015的比赛上，通过对比6个不同的模型，取得了3.57%错误率的最优成绩。

**结论：**

**ResNet解决了网络训练退化的问题，找到了可以训练更深网络的办法（残差连接），如今也成为了深度学习中最重要的一种模型。**