《Deep Residual Learning for Image Recognition》

摘要：

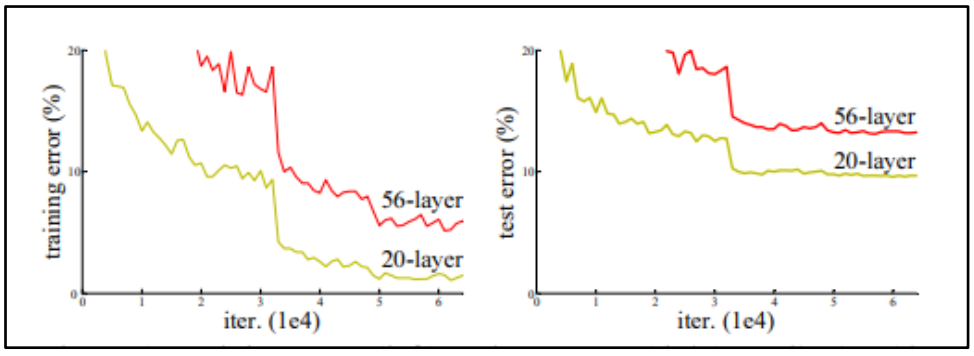
1. 本论文解决的问题：更深的神经网络很难训练，出现**“退化”问题**。
2. 所提出的解决方法：**深度残差学习(Deep Residual Learning)**框架，它使深度神经网络更容易训练。（通过对网络中的层对层输入的残差函数进行学习）
3. 模型的比赛结果：2015 年 ImageNet 分类任务中，以 3.57%的错误率，获得第一名。
4. 研究分析：在 CIFAR-10 数据集上进行了 100 层和 1000 层网络的实验分析。
5. 其他：ImageNet 检测任务，定位任务，COCO 检测和分割任务取得第一。

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：

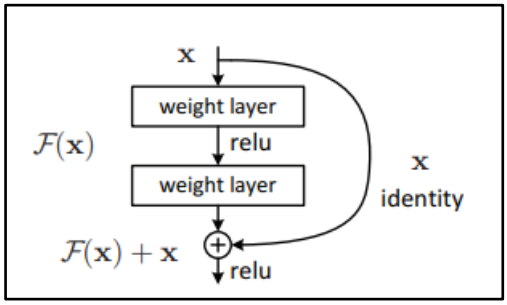


图中，训练误差和测试误差在更深的网络上呈现出更高的误差率。

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：



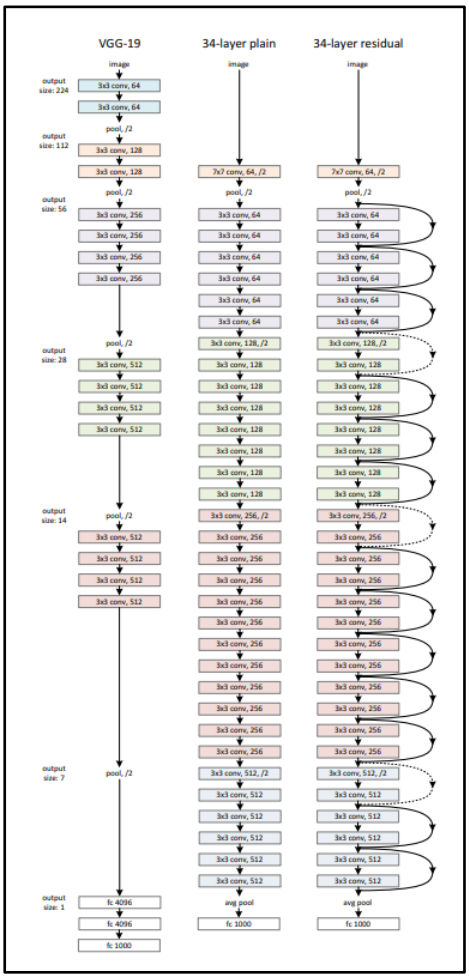
即在原始的层与层连接上增加一个 identity mapping（恒等映射），将原始所需要学的函数$H(x)$转换成$F(x)+x$，其中$F(x)=H(x)-x$。新的非线性网络层用来拟合$F(x)$，也称**残差映射**，最终$H(x)$的结果由$F(x)$和x简单相加得到，相加的处理方式为**跳跃连接**。（这就是整篇论文的核心思想。）

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：

1°——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）；

2°——池化层：3×3 最大池化层，步长为 2（共计 1 层）；

3°——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°——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）；

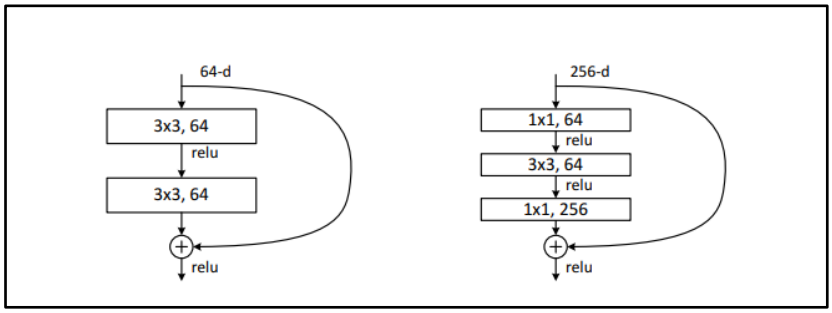
5°——6 个残差连接块：每一个连接块由两层卷积网络组成，卷积核大小为 3×3，卷积核个数为 256（共计 12 层）；

6°——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$1+1+6+8+12+6=34$（层）网络层，而最后的输出结果由全局平均池化层和$softmax$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：

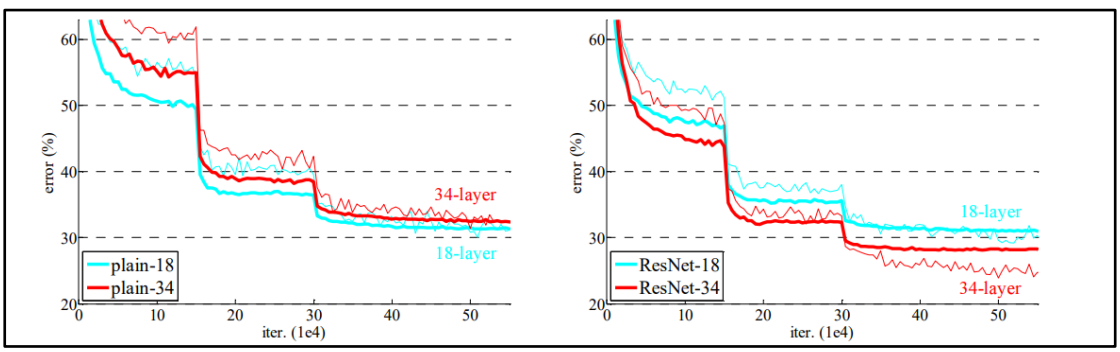
（A）使用恒等映射，如果 residual block 的输入输出维度不一致，则对增加的维度用 0 来填充；（零填充）

（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$F(x)$和x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：



	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	**25.03**

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；

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

method	top-5 err. (**test**)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PReLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

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。