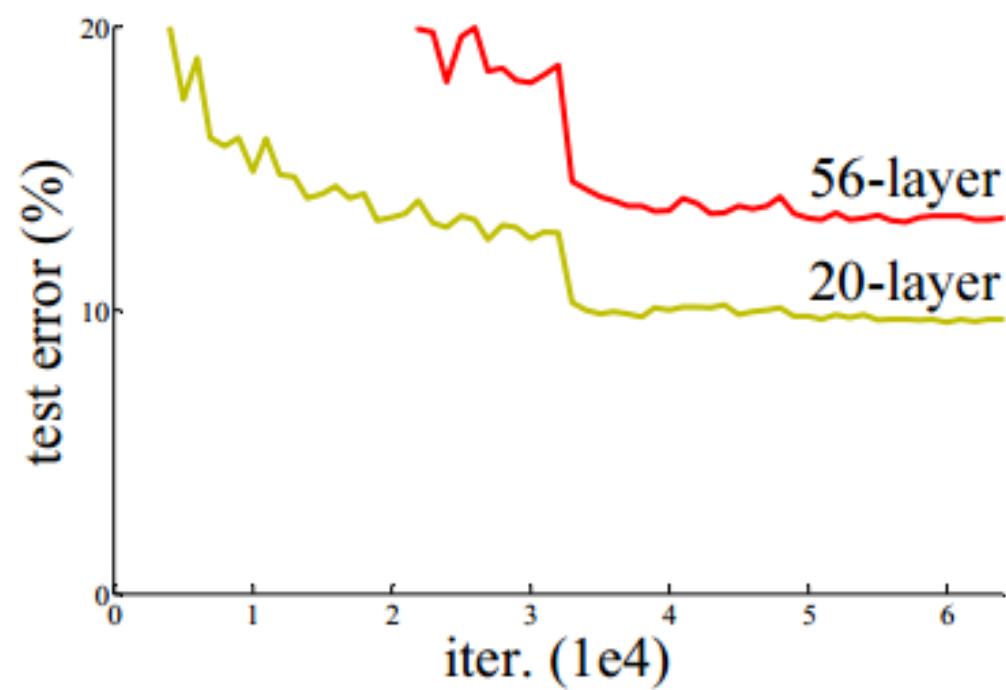
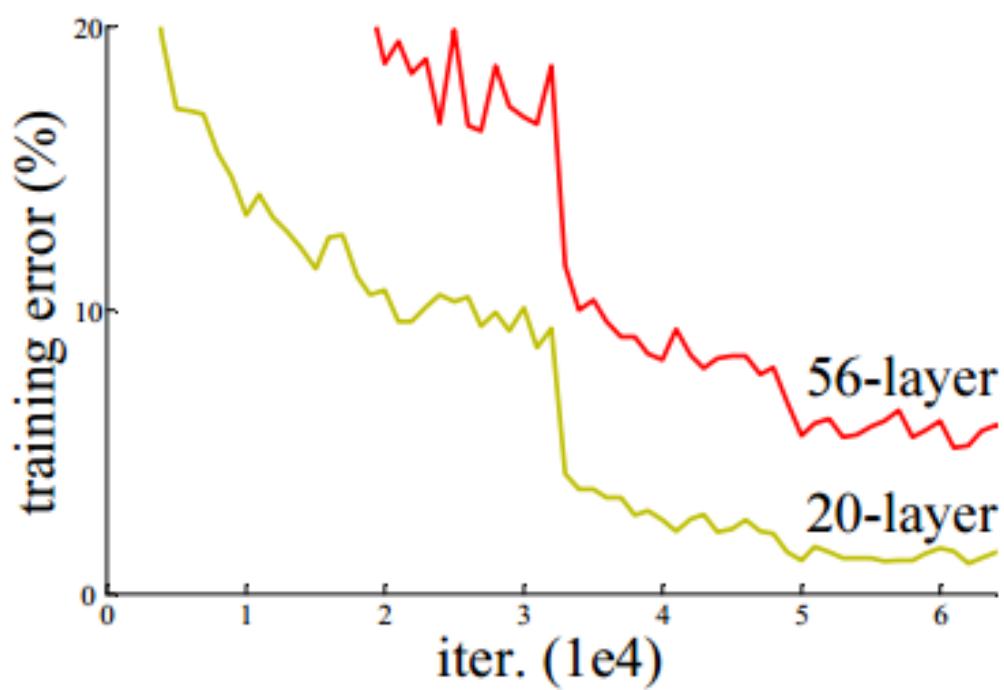


# Deep Residual Learning for Image Recognition

—0801 组会

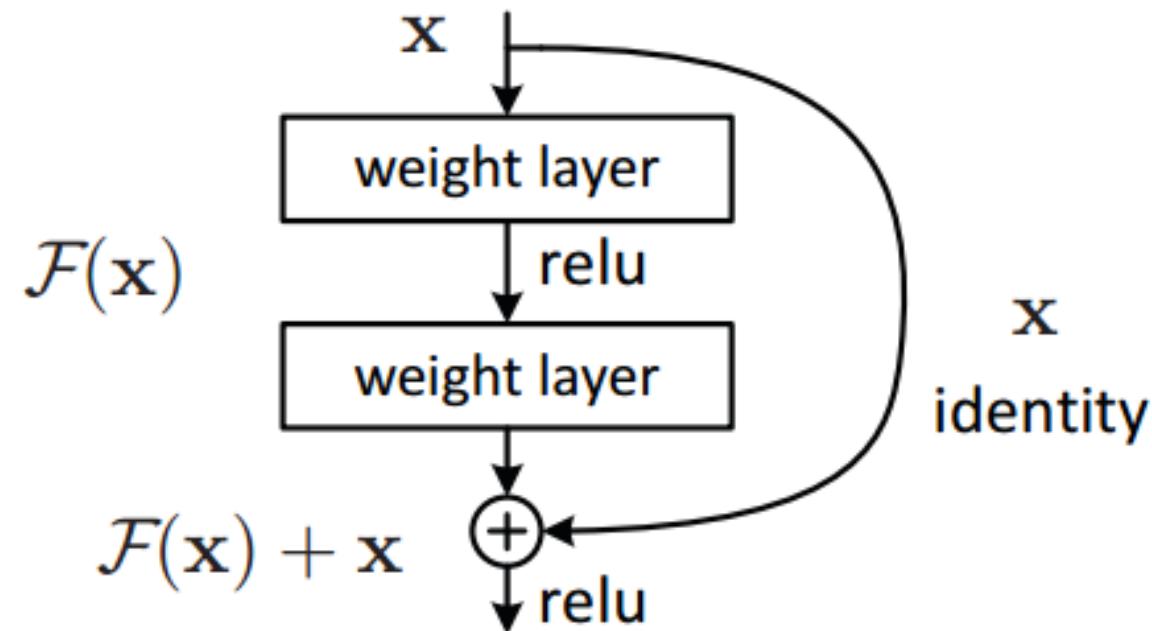
# 动机与背景

- 深度网络出现退化（非过拟合）



# 提出设想

- 浅层模型+恒等映射=深层模型——至少不会比浅层更差
- 神经网络拟合恒等映射十分困难
- 跨层连接——shortcut 进行恒等映射
- 所有层进行残差映射  $H(x) = F(x) + x$
- 不增加额外参数，不增加计算复杂度



# 相关工作——为什么

## ➤ 残差表示

- ✓ 对于 VLAD 和 Fisher Vector，对残差向量编码比对原始向量编码效率更高；
- ✓ 用 Multigrid 解偏微分方程 (Partial Differential Equations, PDE) 时，使用残差向量对于优化更好，收敛速度更快

## ➤ Shortcut连接

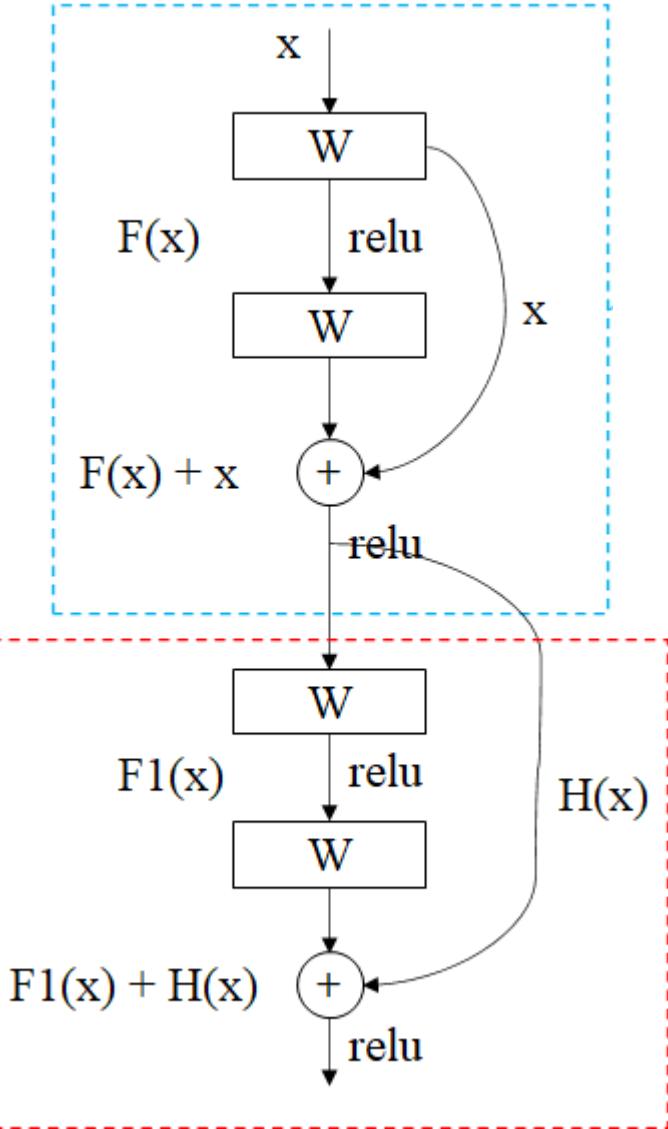
- ✓ Shortcut连接已经经过了很长的一段实践和理论研究过程，证明是有效的；
- ✓ 我们的模型总是学习残差函数，恒等shortcuts从不关闭，是无参数的，所有的信息总是通过的。

# 深度残差学习

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s \mathbf{x}.$$

- (1) 加了残差结构后，给了输入 $\mathbf{x}$ 一个大的选择。若本层神经网络冗余，可以直接走shortcut 跳过这个冗余层，而无需再去拟合参数使得 $H(\mathbf{x})=\mathcal{F}(\mathbf{x})+\mathbf{x}$
- (2) 加了恒等映射后，深层网络至少不会比浅层网络更差。
- (3) 将网络的输出优化为0比将其做一个恒等变换容易得多



# 实验对比

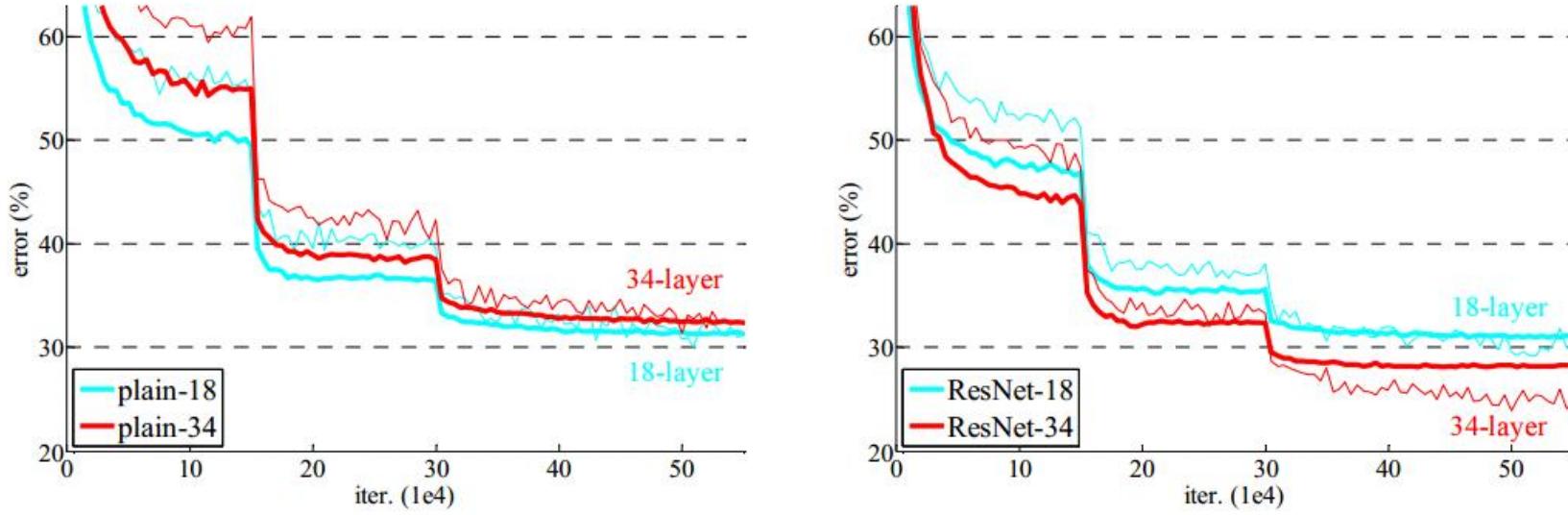


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

- 34层的 ResNet 比18层的 ResNet 好2.8%。
- 34层的 ResNet 显示出相当低的训练误差，并且可以推广到验证数据
- 与普通同类相比，34层ResNet 将top-1的误差减少了3.5%，验证了残差学习在极深度系统上的有效性
- 18层的普通/残余网络相对准确，但18层的 ResNet 收敛得更快
- 当网络“不太深”时，当前的 SGD 求解器仍然能够找到普通网络的良好解

# 恒等 vs 映射 Shortcuts

- ①零填充快捷方式用于增加维度，且所有快捷方式都是无参数的；
- ②投影快捷键用于增加维数，其他快捷键为恒等；
- ③所有的捷径都是投影。

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	<b>21.43</b>	<b>5.71</b>

Table 3. Error rates (%), **10-crop** testing) on ImageNet validation. VGG-16 is based on our test. ResNet-50/101/152 are of option B that only uses projections for increasing dimensions.

③>②>①，但差别较小

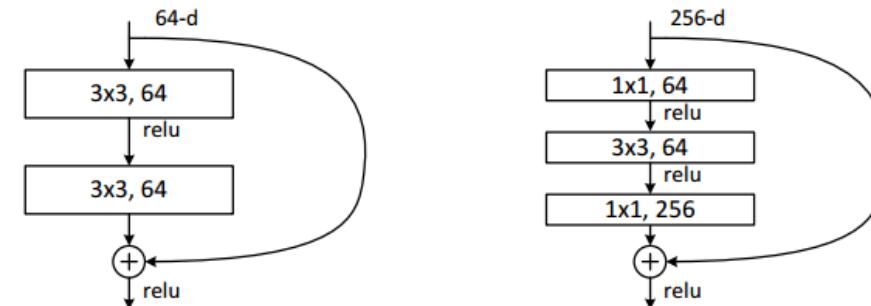


Figure 5. A deeper residual function  $\mathcal{F}$  for ImageNet. Left: a building block (on  $56 \times 56$  feature maps) as in Fig. 3 for ResNet-34. Right: a “bottleneck” building block for ResNet-50/101/152.