



人工智能与深度学习

4-5 ResNet

王中雷

经济学院、王亚南经济研究院、邹至庄经济研究院

厦门大学

(第一版)

内容摘要

1. 研究动机

2. 网络结构

研究动机

1. 在 ILSVRC 2015 的 ImageNet 分类任务中获得第一名
2. 理论上讲，训练误差通常会随着网络层数的增加而降低
3. 然而，实际情况并非如此

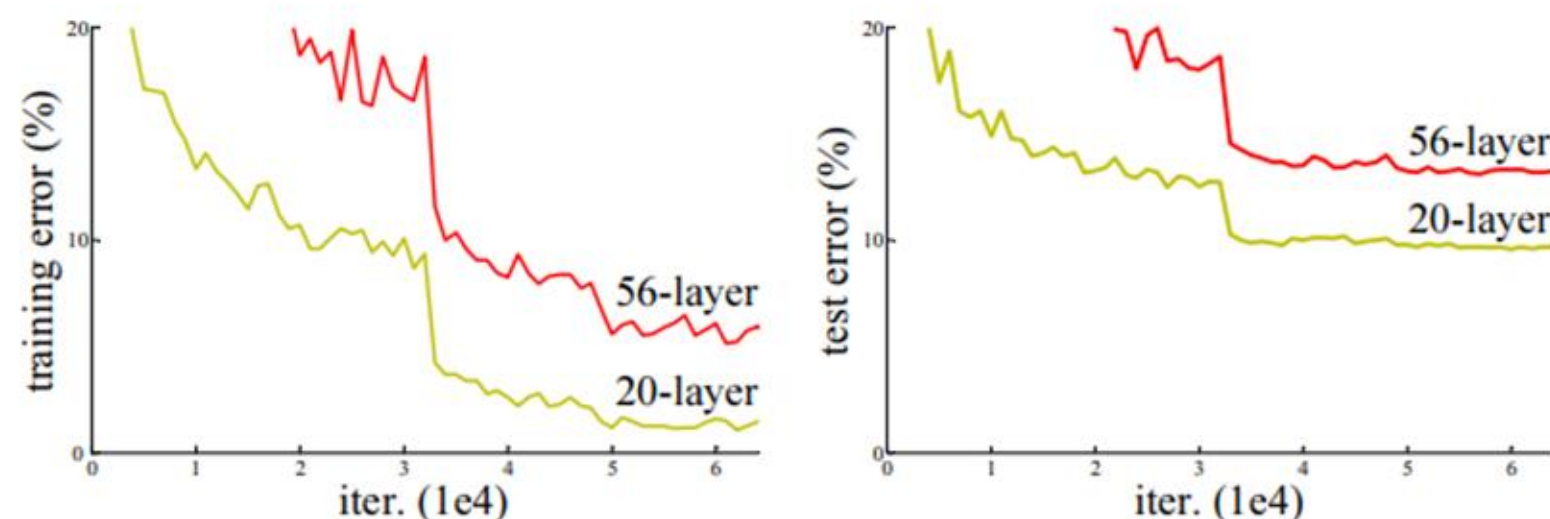


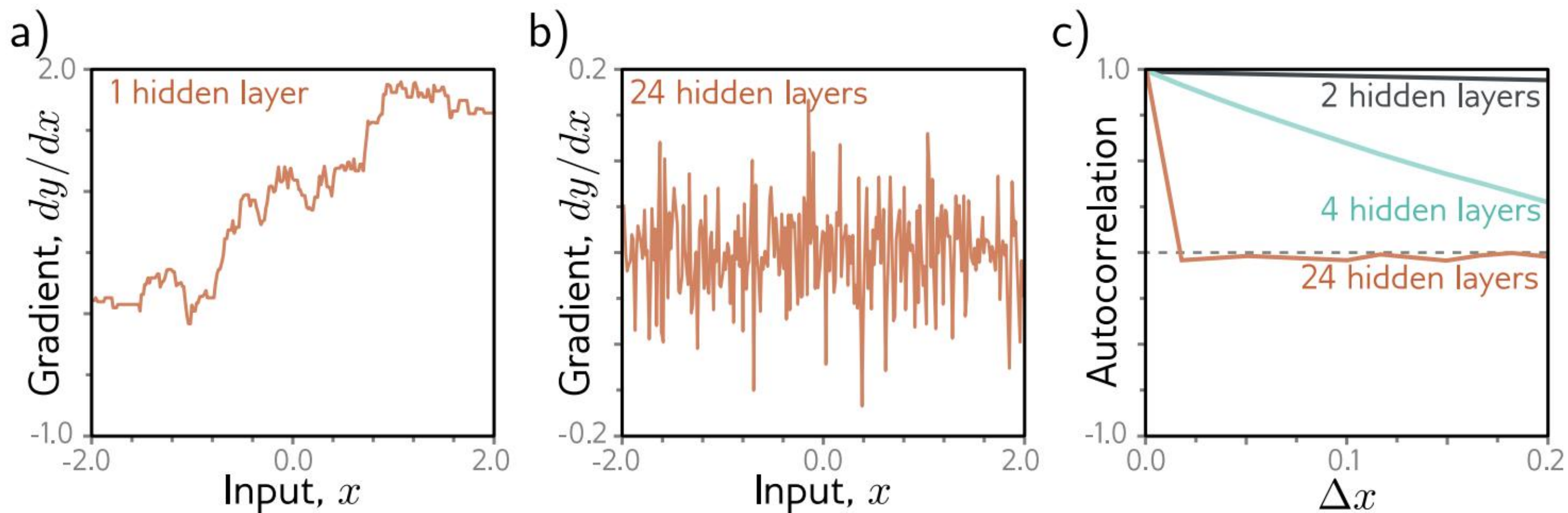
Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

研究动机

1. 一般而言，复杂模型往往会对训练样本进行过拟合，导致在测试集上表现较差
2. 然而，过拟合往往会降低训练误差
3. 因此，这种退化现象主要源于深层网络的优化困难，而不是过拟合所导致的泛化问题
4. 这一现象尚未得到完全统一的理论解释；一种可能的解释与初始化时的梯度结构有关
 - 当我们稍微修改浅层参数时，梯度的表现往往会出乎意料
 - 利用合理的初始化，我们可以避免缓解消失或者梯度爆炸的问题
 - 然而，基于梯度的更新，往往要求很小的步长（学习率）
 - ▷ 在实际中，步长（学习率）是一个（给定的）数值
 - ▷ 当代价函数表现得像重重叠叠的小山峰时，参数更新就可能出问题
 - 这一解释得到了只有一维的输入和输出的神经网络数值实验的支持

研究动机

1. 以下图片来自于 Prince (2024) 的图 11.3



研究动机

1. 我们得到如下结论：

- 对于浅层神经网络模型，模型输出对输入的偏导数变化**较为平滑（缓慢）**
- 然而，对于深层神经网络模型，往往**不那么平滑（变化较为剧烈）**

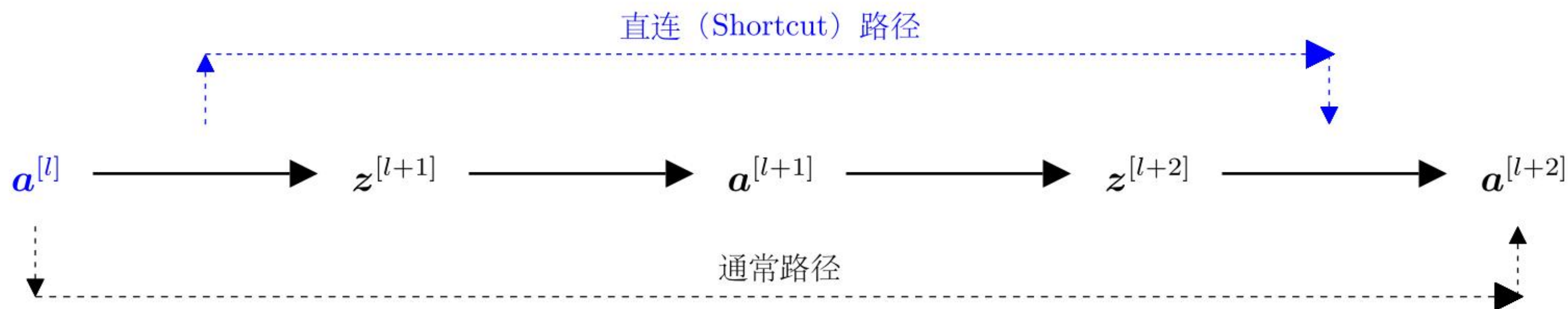
2. 这种现象可以通过梯度间的“自相关函数”解释

- 对于浅层网络，相近输入点处的梯度通常具有较强的相关性
- 随着网络加深，这种相关性会**迅速衰减**

3. 这被称作：**梯度破碎（shattered gradients）**现象

4. 残差连接可以减缓梯度相关性随网络深度的衰减，从而缓解梯度破碎问题

网络结构



$$z^{[l+1]} = \mathbf{W}^{l+1} \cdot a^{[l]} + b^{[l+1]} \quad a^{[l+1]} = \sigma^{[l+1]}(z^{[l+1]}) \quad z^{[l+2]} = \mathbf{W}^{l+2} \cdot a^{[l+1]} + b^{[l+2]} \quad a^{[l+2]} = \sigma^{[l+2]}(z^{[l+2]})$$

1. 直连 (Shortcut) 路径

$$a^{[l+2]} = \sigma^{[l+2]}(z^{[l+2]} + a^{[l]})$$

2. 也就是说, $z^{[l+2]}$ 学习的是完整映射与“残差块输入” $a^{[l]}$ 之间的残差修正

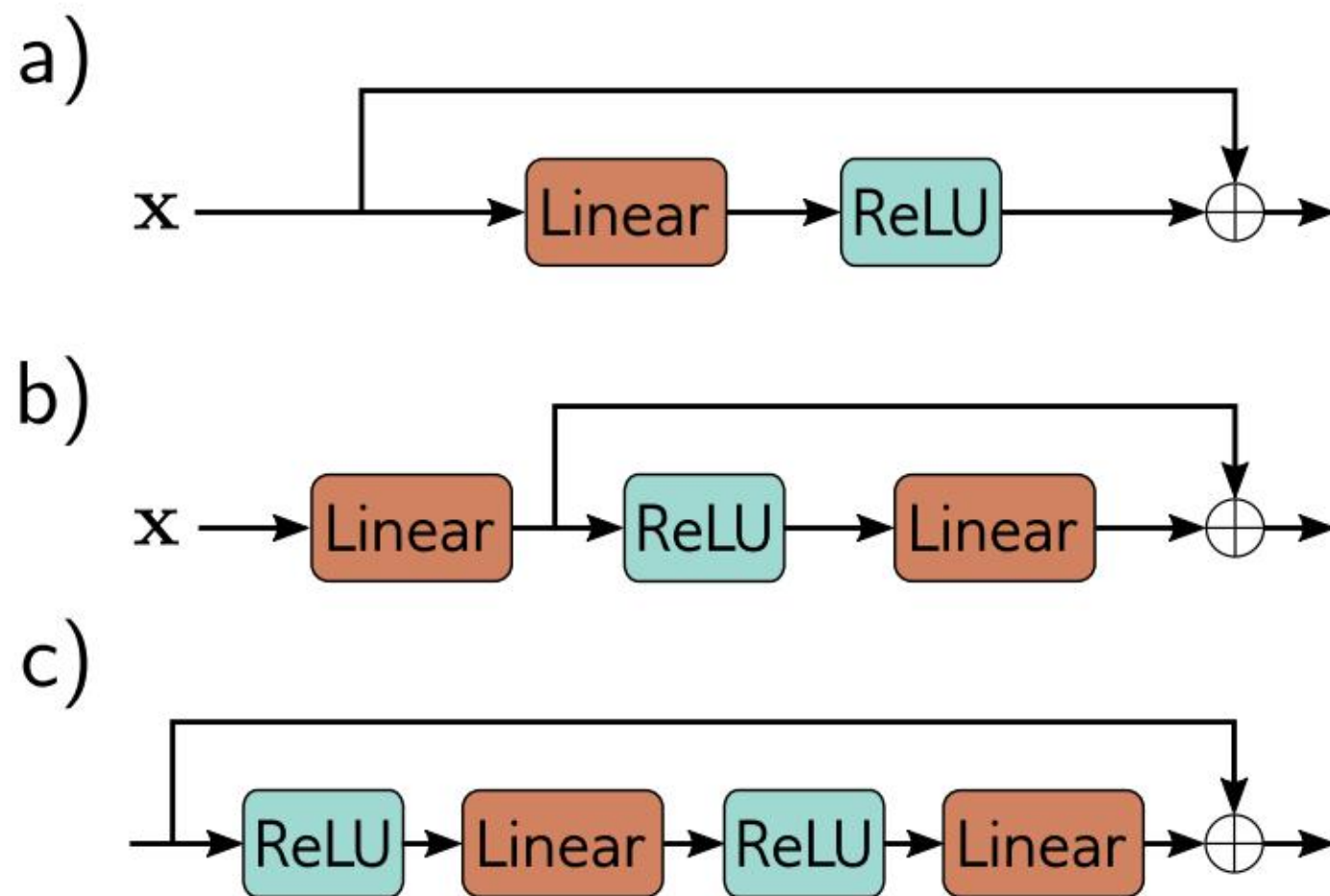
备注

1. 更一般地，向量 $\mathbf{z}^{[l+2]}$ 和 $\mathbf{a}^{[l]}$ 的维度可以不同
2. 维度不一致时，可以引入投影矩阵 $\mathbf{W}_s^{[l+2]}$ 进行匹配

$$\mathbf{a}^{[l+2]} = \sigma^{[l+2]}(\mathbf{z}^{[l+2]} + \mathbf{W}_s^{[l+2]} \cdot \mathbf{a}^{[l]})$$

备注

1. 下图来自于 Prince (2024) 的图 11.5



备注

1. ReLU 激活函数的输出是非负的
2. 如果直连路径在 ReLU 之后，由于 ReLU 的输出非负，它只能对直连路径的值进行非负修正
3. 因此，直连路径主要存在于两个线性变换之间

备注

1. 直连路径有助于训练更深的神经网络
2. 但是，模型的训练可能出现（指数级）的梯度消失或者梯度爆炸
3. 我们可利用批量归一化（batch normalization），缓解这个问题带来的影响

课程总结

1. ResNet 通过直连路径，将学习任务改写成“保留原有信息 + 学习增量修正”
2. 当激活函数是 ReLU 时，我们要注意直连路径的起点和终点
3. 对于深层神经网络而言，ResNet 并不能解决梯度的所有问题