

给深网络修一条近路

本课讲 2015 年的一篇文章。何恺明 (Kaiming He)、张祥雨、任少卿和孙剑四位科学家，给电脑大脑修了一条“近路”。从此，电脑大脑可以盖到 100 多层高，而且越高越聪明。

适合 10 岁左右

约 40 分钟

需要：铅笔、彩笔、便利贴、5 位以上同伴

1 热身：人越多越好吗？

下面 4 件事，人数从 3 个变成 10 个。会更好，还是不一定？圈一圈。

① 拔河：更好 / 不一定

② 搬 100 本书：更好 / 不一定

③ 传话游戏：更好 / 不一定

④ 一幅画轮流重画：更好 / 不一定

传话和重画不一样：每多一个人，就多一次走样的机会。电脑大脑的“层”加多了，也会出这种怪事。

2 电脑大脑是一层一层的

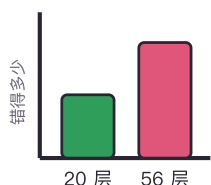
第 8 课讲过 2012 年 ImageNet 认图比赛的冠军。它的“大脑”分成很多层。图片从第一层进去，一层一层往上传，每层多看出一点东西（示意）：



层数越多，我们就说这个网络越深。2012 年那台电脑有 8 层。2014 年，有的网络做到了 19 层、22 层，成绩更好了。大家很自然地想：网络越深，是不是就越聪明？

3 怪事：层越多，练习题越差

何恺明他们用一套 10 类的小图片做实验，比如飞机、汽车、猫和狗。他们分别练了一个 20 层和一个 56 层的网络。



结果，56 层的网络连练习题都做得更差。注意，是练习题，不是考试题。在 ImageNet 的大图片上，34 层的网络也比 18 层的差。

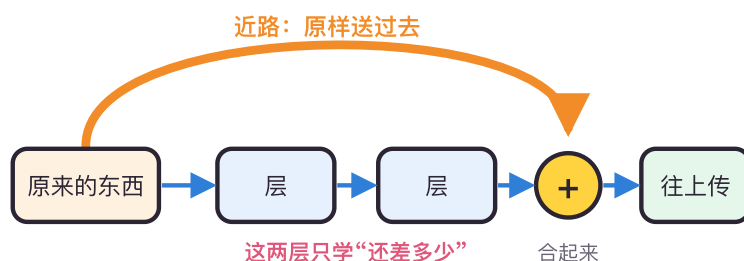
左图是示意，只表示谁高谁低，不是真实的数。

这个毛病叫**退化**：层加多了，反而学不动。它和第 3 课的“死记硬背”不一样。死记硬背是练习题全对、新题出错；退化是**连练习题都做不好**。

奇怪在哪里？56 层本来至少能和 20 层一样好：前 20 层照着 20 层的网络学，多出来的 36 层**什么都不改**就行。可是电脑自己找不到这个办法。每一层都会把东西改一改，“原样往上传”反而很难学。问题不在大脑不够大，而在大脑太难练。

4 ResNet 的办法：修一条近路

四位科学家想了一个很简单的办法：每隔两三层，就修一条**近路**。原来的东西走近路，**原样送**到这几层的后面。然后把两路合在一起，再往上传。



有了近路，原来的东西已经送到后面了。中间这几层不用再从头做一遍，只需要学一件事：**还差多少**。“还差多少”有个名字，叫**残差**。所以这种网络叫**残差网络**，英文简称 ResNet。

像老师改作文：老师不会把你的作文从头重抄一遍，只用红笔标出要改的地方。比如原文写“心里很高心”，红笔只写“心→兴”。原来的作文走近路，红笔改动就是“还差多少”。两样合起来，就是改好的作文。ResNet 里的每几层，就像这支红笔。

5 没用的层，学成“什么都不改”

如果一篇作文已经写得很好，老师的红笔就**一个字也不写**。ResNet 里也一样。如果某几层帮不上忙，它们只要学成“什么都不改”，原来的东西就从近路原样通过。没有近路时，这几层得把东西一点不差地重做一遍，这很难。有了近路，交一张“空白改动”就行，这很容易。所以，有用的层改一点，没用的层不添乱，网络就敢越盖越高了。

6 结果：152 层赢了比赛

有了近路，34 层终于比 18 层好了，退化消失了。科学家一路加到 **152 层**：

8 层（2012）→ 19 层、22 层（2014）→ **152 层**（2015 ResNet）

2015 年，ResNet 在 ImageNet 认图比赛里拿了第一。比赛要认 1000 类东西，电脑可以说 5 个猜测，5 个里有正确答案就算对。2012 年的冠军，1000 张图大约错 153 张（错误率约 15.3%）。ResNet 大约只错 36 张（错误率约 3.57%）。这是几个深网络一起商量出的成绩，其中最深的有 152 层。

楼更高，活却不更多：152 层是 19 层 VGG 的 8 倍高，认一张图要算的量却比 VGG 少。

也不是越高越好：他们在小图片上试了 1202 层。练习题做得很好，新题却比 110 层差一点。作者认为，这回是死记硬背了。

7 这和今天有什么关系

你听说过 ChatGPT、Claude 这些会聊天的 AI 吗？它们用的结构叫 Transformer，第 13、14 课会讲。Transformer 有很多层，**每一层旁边都有一条这样的近路**。

就像全班合写一本作文本：每人读一读，再添一点、改一点，没有人从头重抄。聊天 AI 的每一层，也是读一读，再把自己的“改动”加回去。近路不是聊天 AI 唯一的秘密，但今天几乎所有很深的网络都带着它。

8 动手：传话接力，带不带原纸条

玩法（6 到 8 个人，约 15 分钟，需要 3 张纸条）

- 第 2 课玩过传话链。这次比一比：**队伍变长以后**会怎样。
- 第 1 轮 短队：只用 3 个人。第 1 人看纸条 A，悄悄传话，最后一人大声说出来。
 - 第 2 轮 长队：全部人排成一队，用纸条 B，玩法同上。
 - 第 3 轮 长队 + 近路：用纸条 C，口传照旧。第 1 人还要把纸条 C 从队伍旁边直接递给最后一人。最后一人对照纸条，**只改不一样的地方**，再大声说出来。

A：四只黑猫在十号楼顶吃了十四条小鱼。 B：六只白鹅在七号桥下捡了十七颗石子。
C：九只花狗在三号门口啃了三十根骨头。（5 个要点：几只、动物、几号、在哪、多少）

	第 1 轮 短队	第 2 轮 长队	第 3 轮 长队 + 近路
几个人	3		
对了几个要点			

你会发现：只口传时，人越多，错得越多，就像“退化”。有了直通最后一人的原纸条，队伍再长，也能对回原话。

9 动手：只画改动

玩法（2 组，每组 3 到 4 个人，约 15 分钟，需要白纸、便利贴、彩笔）

- 1. 每组第 1 人画：1 座房子，1 扇门、2 扇窗，1 棵树，1 个太阳，共 5 样东西。其余每人抽一张改动卡：门涂红色、加一朵云、树上加 3 个苹果、什么都不改。
- 2. 重抄组：每人在新纸上把上一人的画整幅重画，再加上自己的改动。旧画收走，只传新画。
- 3. 近路组：原画一直放在桌上。每人只在便利贴上画自己的改动，贴在原画旁边。抽到“什么都不改”的人，贴一张空白便利贴。
- 4. 两组都计时。最后摆出成果，一起数一数。

	一共用了几分钟	原来的 5 样东西还剩几样	改动全都在吗
重抄组			
近路组			

你会发现：重抄组每抄一遍，都可能漏掉东西，还很花时间。近路组的原画一直都在。抽到“什么都不改”的人，在近路组只贴一张白纸，在重抄组却得重画一整遍。

10 词语小卡片

层和深度

电脑大脑一层一层地传消息。层数越多，网络越“深”。

退化

层加多以后，连练习题都做得更差。不是死记硬背，是学不动。

近路（捷径）

跳过几层的一条线，把原来的东西原样送到后面。

残差

“还差多少”。有了近路，中间几层只学这个。用残差的网络叫 ResNet。

11 想一想

1. 56 层的网络连练习题都做不好。这和第 3 课的“死记硬背”有什么不一样？

2. 生活里，还有哪些事情是“只改差多少，不用从头再来”？举一个例子。

3. 1202 层的网络做新题，比 110 层的还差一点。你觉得网络是不是越深越好？为什么？

一句话记住：给每几层修一条近路，让它们只学“还差多少”，网络再深也不会越学越差。

给家长和老师：本课讲的是 Kaiming He、Xiangyu Zhang、Shaoqing Ren、Jian Sun 2015 年的论文 [Deep Residual Learning for Image Recognition](#) (CVPR 2016 最佳论文)。论文发现深网络的“退化”问题：在 CIFAR-10 上，56 层普通网络的训练误差高于 20 层；ImageNet 上 34 层高于 18 层。作者用恒等捷径 (shortcut) 让每 2-3 层只学习残差，解决了这个问题。讲义里的 3.57% 是多个模型集成在 ImageNet 测试集上的 top-5 错误率，单个 152 层模型在验证集上是 4.49%。1202 层网络的测试误差 (7.93%) 高于 110 层 (6.43%)，作者归因于过拟合。

以下是示意或编的例子：“各层看出线条、眼睛”的分工，第 3 节柱状图（只表示高低），作文改错例子，“合写作文本”的比喻，两个活动的句子和画。用“每层都会改一改”解释退化是简化说法，论文的结论是深层普通网络难以优化。想深入可看本站的 [第 9 天中文导读](#)，代码见开源项目 [30u30 第 9 天目录](#) (英文)。

极客教育 jikejiao.com 原创讲义 · 2026-10