

給深網絡修一條近路

本課講 2015 年的一篇文章。何愷明 (Kaiming He)、張祥雨、任少卿和孫劍四位科學家，給電腦大腦修了一條「近路」。從此，電腦大腦可以蓋到 100 多層高，而且越高越聰明。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、彩筆、便利貼、5 位以上同伴

1 熱身：人越多越好嗎？

下面 4 件事，人數從 3 個變成 10 個。會更好，還是不一定？圈一圈。

① 拔河：更好 / 不一定

② 搬 100 本書：更好 / 不一定

③ 傳話遊戲：更好 / 不一定

④ 一幅畫輪流重畫：更好 / 不一定

傳話和重畫不一樣：每多一個人，就多一次走樣的机会。電腦大腦的「層」加多了，也會出這種怪事。

2 電腦大腦是一層一層的

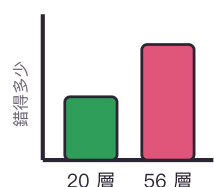
第 8 課講過 2012 年 ImageNet 認圖比賽的冠軍。它的「大腦」分成很多層。圖片從第一層進去，一層一層往上傳，每層多看出一點東西（示意）：



層數越多，我們就說這個網絡越深。2012 年那臺電腦有 8 層。2014 年，有的網絡做到了 19 層、22 層，成績更好了。大家很自然地想：網絡越深，是不是就越聰明？

3 怪事：層越多，練習題越差

何愷明他們用一套 10 類的小圖片做實驗，比如飛機、汽車、貓和狗。他們分別練了一個 20 層和一個 56 層的網絡。



結果，56 層的網絡連練習題都做得更差。注意，是練習題，不是考試題。在 ImageNet 的大圖片上，34 層的網絡也比 18 層的差。

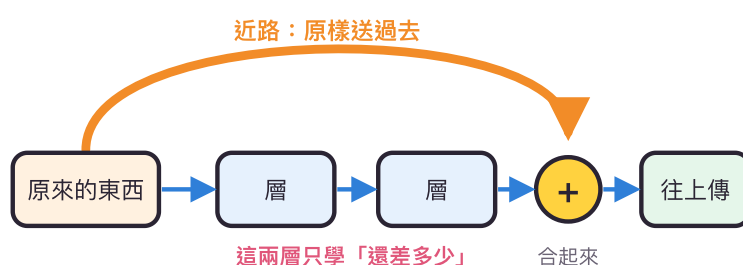
左圖是示意，只表示誰高誰低，不是真實的數。

這個毛病叫**退化**：層加多了，反而學不動。它和第 3 課的「死記硬背」不一樣。死記硬背是練習題全對、新題出錯；退化是**連練習題都做不好**。

奇怪在哪裡？56 層本來至少能和 20 層一樣好：前 20 層照著 20 層的網絡學，多出來的 36 層**什麼都不改**就行。可是電腦自己找不到這個辦法。每一層都會把東西改一改，「原樣往上傳」反而很難學。問題不在大腦不夠大，而在大腦太難練。

4 ResNet 的辦法：修一條近路

四位科學家想了一個很簡單的辦法：每隔兩三層，就修一條**近路**。原來的東西走近路，**原樣**送到這幾層的後面。然後把兩路合在一起，再往上傳。



有了近路，原來的東西已經送到後面了。中間這幾層不用再從頭做一遍，只需要學一件事：**還差多少**。「還差多少」有個名字，叫**殘差**。所以這種網絡叫**殘差網絡**，英文簡稱 ResNet。

像老師改作文：老師不會把你的作文從頭重抄一遍，只用紅筆標出要改的地方。比如原文寫「心裡很高心」，紅筆只寫「心→興」。原來的作文走近路，紅筆改動就是「還差多少」。兩樣合起來，就是改好的作文。ResNet 裡的每幾層，就像這支紅筆。

5 沒用的層，學成「什麼都不改」

如果一篇作文已經寫得很好，老師的紅筆就**一個字也不寫**。ResNet 裡也一樣。如果某幾層幫不上忙，它們只要學成「什麼都不改」，原來的東西就從近路原樣通過。沒有近路時，這幾層得把東西一點不差地重做一遍，這很難。有了近路，交一張「空白改動」就行，這很容易。所以，有用的層改一點，沒用的層不添亂，網絡就敢越蓋越高了。

6 結果：152 層贏了比賽

有了近路，34 層終於比 18 層好了，退化消失了。科學家一路加到 **152 層**：

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

2015 年，ResNet 在 ImageNet 認圖比賽裡拿了第一。比賽要認 1000 類東西，電腦可以說 5 個猜測，5 個裡有正確答案就算對。2012 年的冠軍，1000 張圖大約錯 153 張（錯誤率約 15.3%）。ResNet 大約只錯 36 張（錯誤率約 3.57%）。這是幾個深網絡一起商量出的成績，其中最深的有 152 層。

樓更高，活卻不更多：152 層是 19 層 VGG 的 8 倍高，認一張圖要算的量卻比 VGG 少。

也不是越高越好：他們在小圖片上試了 1202 層。練習題做得很好，新題卻比 110 層差一點。作者認為，這回是死記硬背了。

7 這和今天有什麼關係

你聽說過 ChatGPT、Claude 這些會聊天的 AI 嗎？它們用的結構叫 Transformer，第 13、14 課會講。Transformer 有很多層，**每一層旁邊都有一條這樣的近路**。

就像全班合寫一本作文本：每人讀一讀，再添一點、改一點，沒有人從頭重抄。聊天 AI 的每一層，也是讀一讀，再把自己的「改動」加回去。近路不是聊天 AI 唯一的秘密，但今天幾乎所有很深的網絡都帶著它。

8 動手：傳話接力，帶不帶原紙條

玩法（6 到 8 個人，約 15 分鐘，需要 3 張紙條）

第 2 課玩過傳話鏈。這次比一比：**隊伍變長以後**會怎樣。

- 1. **第 1 輪 短隊：**只用 3 個人。第 1 人看紙條 A，悄悄傳話，最後一人大聲說出來。
- 2. **第 2 輪 長隊：**全部人排成一隊，用紙條 B，玩法同上。
- 3. **第 3 輪 長隊 + 近路：**用紙條 C，口傳照舊。第 1 人還要把紙條 C 從隊伍旁邊直接遞給最後一人。最後一人對照紙條，**只改不一樣的地方**，再大聲說出來。

A：四隻黑貓在十號樓頂吃了十四條小魚。 B：六隻白鵝在七號橋下撿了十七顆石子。
C：九隻花狗在三號門口啃了三十根骨頭。（5 個要點：幾隻、動物、幾號、在哪、多少）

	第 1 輪 短隊	第 2 輪 長隊	第 3 輪 長隊 + 近路
幾個人	3		
對了幾個要點			

你會發現：只口傳時，人越多，錯得越多，就像「退化」。有了直通最後一人的原紙條，隊伍再長，也能對回原話。

9 動手：只畫改動

玩法（2 組，每組 3 到 4 個人，約 15 分鐘，需要白紙、便利貼、彩筆）

1. 每組第 1 人畫：1 座房子，1 扇門、2 扇窗，1 棵樹，1 個太陽，共 5 樣東西。其餘每人抽一張改動卡：門塗紅色、加一朵雲、樹上加 3 個蘋果、什麼都不改。
2. 重抄組：每人在新紙上把上一人的畫整幅重畫，再加上自己的改動。舊畫收走，只傳新畫。
3. 近路組：原畫一直放在桌上。每人只在便利貼上畫自己的改動，貼在原畫旁邊。抽到「什麼都不改」的人，貼一張空白便利貼。
4. 兩組都計時。最後擺出成果，一起數一數。

	一共用了幾分鐘	原來的 5 樣東西還剩幾樣	改動全都在嗎
重抄組			
近路組			

你會發現：重抄組每抄一遍，都可能漏掉東西，還很花時間。近路組的原畫一直都在。抽到「什麼都不改」的人，在近路組只貼一張白紙，在重抄組卻得重畫一整遍。

10 詞語小卡片

層和深度

電腦大腦一層一層地傳消息。層數越多，網絡越「深」。

退化

層加多以後，連練習題都做得更差。不是死記硬背，是學不動。

近路（捷徑）

跳過幾層的一條線，把原來的東西原樣送到後面。

殘差

「還差多少」。有了近路，中間幾層只學這個。用殘差的網絡叫 ResNet。

11 想一想

1. 56 層的網絡連練習題都做不好。這和第 3 課的「死記硬背」有什麼不一樣？

2. 生活裡，還有哪些事情是「只改差多少，不用從頭再來」？舉一個例子。

3. 1202 層的網絡做新題，比 110 層的還差一點。你覺得網絡是不是越深越好？為什麼？

一句話記住：給每幾層修一條近路，讓它們只學「還差多少」，網絡再深也不會越學越差。

給家長和老師：本課講的是 Kaiming He、Xiangyu Zhang、Shaoqing Ren、Jian Sun 2015 年的論文 [Deep Residual Learning for Image Recognition](#) (CVPR 2016 最佳論文)。論文發現深網絡的「退化」問題：在 CIFAR-10 上，56 層普通網絡的訓練誤差高於 20 層；ImageNet 上 34 層高於 18 層。作者用恆等捷徑 (shortcut) 讓每 2-3 層只學習殘差，解決了這個問題。講義裡的 3.57% 是多個模型集成在 ImageNet 測試集上的 top-5 錯誤率，單個 152 層模型在驗證集上是 4.49%。1202 層網絡的測試誤差 (7.93%) 高於 110 層 (6.43%)，作者歸因於過擬合。

以下是示意或編的例子：「各層看出線條、眼睛」的分工，第 3 節柱狀圖（只表示高低），作文改錯例子，「合寫作文本」的比喻，兩個活動的句子和畫。用「每層都會改一改」解釋退化是簡化說法，論文的結論是深層普通網絡難以優化。想深入可看本站的 [第 9 天中文導讀](#)，代碼見開源項目 [30u30 第 9 天目錄](#) (英文)。

極客教育 jikejiao.com 原創講義 · 2026-10