

小窗口認出一隻貓

本課講 2012 年的一篇論文。Krizhevsky、Sutskever 和 Hinton 三位科學家做出了一個看圖認東西的程序，後來大家叫它 AlexNet。它讓電腦自己從圖片裡學，在那年最有名的看圖比賽裡遙遙領先。

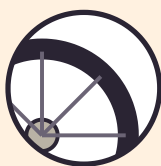
適合 10 歲左右

約 40 分鐘

需要：鉛筆、尺子、剪刀、硬紙片、2—3 位同伴

1 熱身：只看一小塊，你能猜出來嗎？

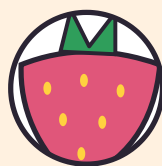
每個圓圈只露出一張圖的一小塊。猜一猜，整張圖畫的是什麼？



A 是 _____



B 是 _____



C 是 _____

你靠的是小線索：一段圓圓的黑邊，一隻尖耳朵，幾粒小黃點。**大東西是由小花紋拼起來的**。今天這臺電腦認圖，用的也是這個思路。

2 一場看圖認東西的大比賽

2009 年，李飛飛帶領的團隊做了一個巨大的圖片庫，叫 **ImageNet**，每張圖都標好了答案。大家每年用它辦一場比賽，比誰的程序認圖最準。

要認的東西	1000 種。光是狗就有上百種。
練習用的圖	大約 120 萬張。考試用程序沒見過的新圖。
怎麼算對	程序說出 5 個猜測，裡面有正確答案就算對。

能猜 5 次，是因為有些東西太像了，比如兩種很像的狗。5 個猜測**全都不對**才算錯。錯的圖佔多少，叫「**前 5 錯誤率**」，越小越好。

3 以前的辦法：人來寫規則

在電腦眼裡，一張圖是一大片小方格。每個小方格叫一個**像素**，只有一種顏色。電腦只看到一堆數字，不知道哪裡是耳朵。2012 年以前，主要靠**人來想好「該找什麼」**，比如邊緣、角落、紋路。電腦照著人

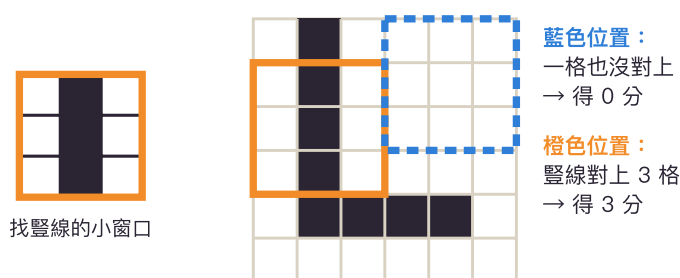
寫的步驟去找，再打分。這像一張檢查單：

人寫的檢查單（示意）：有兩隻尖耳朵嗎？有鬍鬚嗎？有 4 條腿嗎？都有，就是貓。可是貓睡成一團，看不見腿，檢查單說「不是貓」。狐狸有尖耳朵和鬍鬚，檢查單說「是貓」。圖片千變萬化，人寫的規則總有漏洞。

AlexNet 換了思路：**不告訴電腦該找什麼，讓它自己從 120 萬張圖裡學**。就像第 1 課的猜字遊戲，猜錯了就悄悄調整自己。只是這次它猜的是「這張圖是什麼」。

4 卷積：拿小窗口一格一格掃

AlexNet 用很多個**小窗口**看圖。每個小窗口只找一種小花紋。下面這個 3×3 的小窗口，專找「豎線」：



小窗口從左上角開始，一格一格往右挪，一行掃完換下一行。每個位置打一個分。分數高，那裡就有豎線。這種「拿小窗口一格一格掃」的做法，叫**卷積**。用很多層卷積搭起來的程序，叫**卷積神經網絡**。

AlexNet 第一層有 96 種小窗口，每個是 11×11 格。**沒有人告訴它每個窗口該找什麼**。窗口裡的花紋也是「旋鈕」（第 4 課講過），是它練習時自己調出來的。

第一層掃完，結果交給下一層再掃。一層一層往上，小花紋拼成大東西：



AlexNet 一共 8 層：前 5 層用小窗口掃，後 3 層彙總線索，從 1000 種東西里給出答案。層數多就叫「深」，所以這種辦法叫**深度學習**。

5 怎麼練出來：2 塊顯卡和 3 個小竅門

AlexNet 裡有大約 6000 萬個旋鈕。用 120 萬張圖把它們調好，計算量大得嚇人。三位科學家用了 2 塊**遊戲顯卡**（GPU）。

顯卡本來給遊戲畫畫面。電腦的普通「大腦」像一位數學老師，一次做一道題。顯卡像一整個班的同學，每人同時做一道簡單加法。卷積正是大量簡單的加法和乘法，交給顯卡最合適。

當時每塊顯卡只有 3 GB 內存，裝不下整個網絡。他們把網絡劈成兩半，每塊顯卡管一半。2 塊顯卡一起練了 5 到 6 天。他們還用了 3 個小竅門：

ReLU 開關	每個零件後面裝一個簡單開關：沒信號就關，有信號就原樣放過去。學得快好幾倍，一次小測試裡快了約 6 倍。
Dropout（請假法）	第 3 課講過。在最後彙總線索的地方，每次隨機讓一半零件請假，防止死記硬背練習圖。
數據增強（一圖變多題）	把圖左右翻轉，隨機剪下一大塊，或者稍微改改顏色。貓翻過來還是貓，練習題一下子多了很多倍。

6 結果：一下子領先一大截

2012 年的比賽，AlexNet 拿了第一名。比較一下「前 5 錯誤率」：



換成 100 張考試圖來想：第二名大約 26 張全猜錯，AlexNet 只有大約 15 張。差了大約 11 個百分點，誰也沒法裝作沒看見。

2012 年，3 樣東西剛好湊齊：**大量帶答案的圖**（ImageNet），**夠快的顯卡**，**練得動的深網絡**。之後兩三年，研究看圖的科學家幾乎都改用了深度學習。三位作者裡，Hinton 在第 4 課出現過，Sutskever 是第 3 課的作者之一。Sutskever 後來參與創辦了做出 ChatGPT 的 OpenAI。

7 這和今天有什麼關係

今天很多手機相冊能搜「貓」「海邊」，有的軟件拍一下花就能說出花名。這些看圖本領，都是沿著 AlexNet 這條路發展來的。第 9 課的網絡有 152 層，比它深得多。

今天的聊天 AI 結構和 AlexNet 不一樣，但有幾樣東西傳了下來：用顯卡訓練，用簡單開關，練習時請假，**讓電腦從大量例子裡自己學**。AlexNet 證明了：網絡更大，材料更多，算得更快，電腦就學得更好。

8 動手：小窗口找花紋

玩法（1—2 個人，約 15 分鐘，需要硬紙片、尺子、剪刀、鉛筆）

- 照下圖格子大小，在硬紙片上畫兩個 3×3 方格。一個剪掉**中間一豎列**，叫「豎線卡」；一個剪掉**中間一橫行**，叫「橫線卡」。
- 卡片左上角對準第 1 行第 1 列，數鏤空處有幾個黑格，寫進表裡。再往右挪一格，一行 4 個位置，共 4 行。兩張卡各做一遍。

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

豎線卡				
行 \ 列	1	2	3	4
第 1 行				
第 2 行				
第 3 行				
第 4 行				

橫線卡				
行 \ 列	1	2	3	4
第 1 行				
第 2 行				
第 3 行				
第 4 行				

豎線卡得 3 分的位置在第 _____ 列和第 _____ 列。橫線卡得 3 分的位置在第 _____ 行。

你會發現：豎線卡在豎線上得高分，橫線卡在橫槓上得高分。兩張表合起來，就「看出」了一個 H。AlexNet 第一層就做這件事，只是它的 96 張卡片是自己學出來的。

9 動手：我是分類器

玩法（3—4 個人，約 15 分鐘，需要 6 張小紙片、鉛筆）

- 你們來當「**分類器**」（給圖片分類的程序）。小組寫 3 條認貓規則，全滿足才算「是貓」。

規則 1：_____ 規則 2：_____ 規則 3：_____

- 出題人偷偷畫好下表的 6 張簡筆畫，一張張亮出。其他人**只按規則**判斷，再對答案。

出題人畫什麼	按規則：是貓嗎	真正答案	✓ 或 ✕
1. 正面坐著的花貓		是貓	
2. 睡成一團的貓，看不見腿		是貓	
3. 一隻狐狸		不是貓	
4. 一隻老虎		不是貓	
5. 尖耳朵的小狗		不是貓	
6. 只畫了圓臉的卡通貓		是貓	

- 改一條規則補上漏洞，再判斷一遍。這次錯了幾張？ _____

你會發現：規則寬了，狐狸也算貓；規則嚴了，睡覺的貓不算貓。

10 詞語小卡片

卷積 (小窗口掃描)

小窗口一格一格掃，找一種小花紋。

深度學習

很多層網絡，把小花紋拼成大東西。

顯卡 (GPU)

會同時做很多道簡單題，訓練快。

數據增強 (一圖變多題)

翻轉、剪裁、改顏色，一圖變多題。

11 想一想

1. 比賽為什麼允許猜 5 次？舉出兩種很難分清的東西。

2. AlexNet 把練習圖左右翻轉，卻不上下顛倒。你覺得為什麼？

3. 你學認動物時，是靠背規則，還是靠看了很多例子？舉一個例子。

一句話記住：不用人寫規則，讓電腦拿小窗口自己從大量圖片裡學，它就能認出貓和車。

給家長和老師：本課講的是 Alex Krizhevsky、Ilya Sutskever、Geoffrey E. Hinton 2012 年在 NIPS (今 NeurIPS) 發表的論文 [ImageNet Classification with Deep Convolutional Neural Networks](#)。網絡有 5 個卷積層、3 個全連接層、約 6000 萬參數，在 2 塊 GTX 580 (各 3 GB) 上訓練 5 到 6 天。ILSVRC-2012 的 top-5 錯誤率為 15.3%，第二名為 26.2%。「快約 6 倍」指 CIFAR-10 上 4 層網絡 ReLU 對 tanh 的對比；dropout 用在前兩個全連接層，概率 0.5。

真實的卷積核是一組可正可負的數，「數黑格」只是示意。論文展示了第一層學到的邊緣和顏色斑塊；「再拼成眼睛、輪子」來自後續可視化研究 (如 2013 年 ZFNet)。熱身小孔圖、柯基犬的 5 個猜測、認貓檢查單、「一個班同學」比喻、H 字格子圖和分類器遊戲都是示意。小窗口活動答案：豎線卡第 1–3 行都是 3、1、1、3，第 4 行是 3、0、0、3；橫線卡第 2 行是 2、3、3、2，其餘各行都是 1。詳見 [第 8 天中文導讀](#)，代碼見 [30u30 第 8 天目錄](#) (英文)。