

小窗口认出一只猫

本课讲 2012 年的一篇论文。Krizhevsky、Sutskever 和 Hinton 三位科学家做出了一个看图认东西的程序，后来大家叫它 AlexNet。它让电脑自己从图片里学，在那年最有名的看图比赛里遥遥领先。

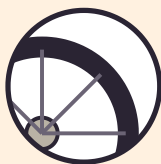
适合 10 岁左右

约 40 分钟

需要：铅笔、尺子、剪刀、硬纸片、2—3 位同伴

1 热身：只看一小块，你能猜出来吗？

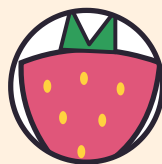
每个圆圈只露出一张图的一小块。猜一猜，整张图画的是什么？



A 是 _____



B 是 _____



C 是 _____

你靠的是小线索：一段圆圆的黑边，一只尖耳朵，几粒小黄点。**大东西是由小花纹拼起来的**。今天这台电脑认图，用的也是这个思路。

2 一场看图认东西的大比赛

2009 年，李飞飞带领的团队做了一个巨大的图片库，叫 **ImageNet**，每张图都标好了答案。大家每年用它办一场比赛，比谁的程序认图最准。

要认的东西	1000 种。光是狗就有上百种。
练习用的图	大约 120 万张。考试用程序没见过的新图。
怎么算对	程序说出 5 个猜测，里面有正确答案就算对。

能猜 5 次，是因为有些东西太像了，比如两种很像的狗。5 个猜测**全都不对**才算错。错的图占多少，叫“**前 5 错误率**”，越小越好。

3 以前的办法：人来写规则

在电脑眼里，一张图是一大片小方格。每个小方格叫一个**像素**，只有一种颜色。电脑只看到一堆数字，不知道哪里是耳朵。2012 年以前，主要靠**人来想好“该找什么”**，比如边缘、角落、纹路。电脑照着人

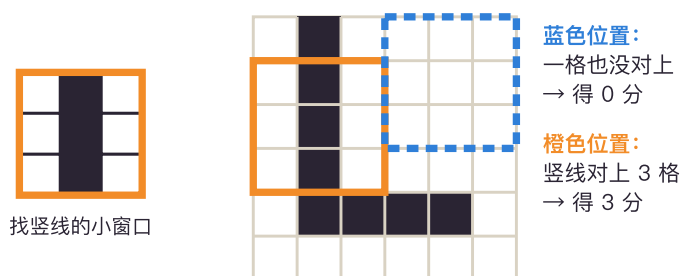
写的步骤去找，再打分。这像一张检查单：

人写的检查单（示意）：有两只尖耳朵吗？有胡须吗？有 4 条腿吗？都有，就是猫。可是猫睡成一团，看不见腿，检查单说“不是猫”。狐狸有尖耳朵和胡须，检查单说“是猫”。图片千变万化，人写的规则总有漏洞。

AlexNet 换了思路：**不告诉电脑该找什么，让它自己从 120 万张图里学**。就像第 1 课的猜字游戏，猜错了就悄悄调整自己。只是这次它猜的是“这张图是什么”。

4 卷积：拿小窗口一格一格扫

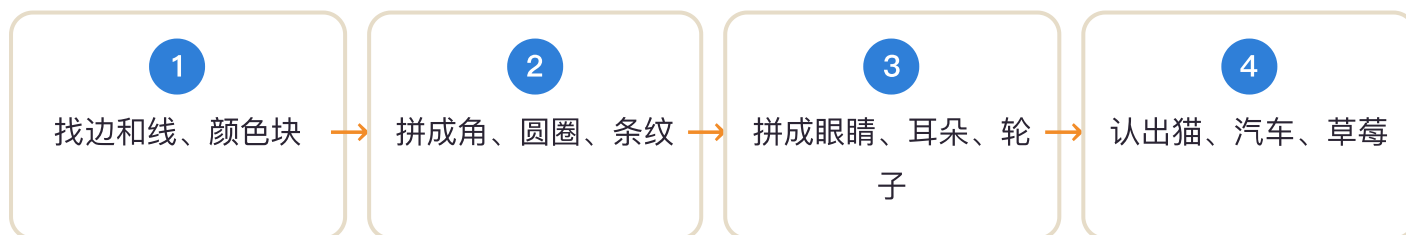
AlexNet 用很多个**小窗口**看图。每个小窗口只找一种小花纹。下面这个 3×3 的小窗口，专找“竖线”：



小窗口从左上角开始，一格一格往右挪，一行扫完换下一行。每个位置打一个分。分数高，那里就有竖线。这种“拿小窗口一格一格扫”的做法，叫**卷积**。用很多层卷积搭起来的程序，叫**卷积神经网络**。

AlexNet 第一层有 96 种小窗口，每个是 11×11 格。**没有人告诉它每个窗口该找什么**。窗口里的花纹也是“旋钮”（第 4 课讲过），是它练习时自己调出来的。

第一层扫完，结果交给下一层再扫。一层一层往上，小花纹拼成大东西：



AlexNet 一共 8 层：前 5 层用小窗口扫，后 3 层汇总线索，从 1000 种东西里给出答案。层数多就叫“深”，所以这种办法叫**深度学习**。

5 怎么练出来：2 块显卡和 3 个小窍门

AlexNet 里有大约 6000 万个旋钮。用 120 万张图把它们调好，计算量大得吓人。三位科学家用了 2 块**游戏显卡**（GPU）。

显卡本来给游戏画画面。电脑的普通“大脑”像一位数学老师，一次做一道题。显卡像一整个班的同学，每人同时做一道简单加法。卷积正是大量简单的加法和乘法，交给显卡最合适。

当时每块显卡只有 3 GB 内存，装不下整个网络。他们把网络劈成两半，每块显卡管一半。2 块显卡一起练了 5 到 6 天。他们还用了 3 个小窍门：

ReLU 开关	每个零件后面装一个简单开关：没信号就关，有信号就原样放过去。学得快好几倍，一次小测试里快了约 6 倍。
Dropout（请假法）	第 3 课讲过。在最后汇总线索的地方，每次随机让一半零件请假，防止死记硬背练习图。
数据增强（一图变多题）	把图左右翻转，随机剪下一大块，或者稍微改改颜色。猫翻过来还是猫，练习题一下子多了很多倍。

6 结果：一下子领先一大截

2012 年的比赛，AlexNet 拿了第一名。比较一下“前 5 错误率”：



换成 100 张考试图来想：第二名大约 26 张全猜错，AlexNet 只有大约 15 张。差了大约 11 个百分点，谁也没法装作没看见。

2012 年，3 样东西刚好凑齐：**大量带答案的图** (ImageNet)，**够快的显卡**，**练得动的深网络**。之后两三年，研究看图的科学家几乎都改用了深度学习。三位作者里，Hinton 在第 4 课出现过，Sutskever 是第 3 课的作者之一。Sutskever 后来参与创办了做出 ChatGPT 的 OpenAI。

7 这和今天有什么关系

今天很多手机相册能搜“猫”“海边”，有的软件拍一下花就能说出花名。这些看图本领，都是沿着 AlexNet 这条路发展来的。第 9 课的网络有 152 层，比它深得多。

今天的聊天 AI 结构和 AlexNet 不一样，但有几样东西传了下来：用显卡训练，用简单开关，练习时请假，**让电脑从大量例子里自己学**。AlexNet 证明了：网络更大，材料更多，算得更快，电脑就学得更好。

8 动手：小窗口找花纹

玩法（1—2 个人，约 15 分钟，需要硬纸片、尺子、剪刀、铅笔）

- 照下图格子大小，在硬纸片上画两个 3×3 方格。一个剪掉**中间一竖列**，叫“竖线卡”；一个剪掉**中间一横行**，叫“横线卡”。
- 卡片左上角对准第 1 行第 1 列，数镂空处有几个黑格，写进表里。再往右挪一格，一行 4 个位置，共 4 行。两张卡各做一遍。

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

竖线卡				
行 \ 列	1	2	3	4
第 1 行				
第 2 行				
第 3 行				
第 4 行				

横线卡				
行 \ 列	1	2	3	4
第 1 行				
第 2 行				
第 3 行				
第 4 行				

竖线卡得 3 分的位置在第 _____ 列和第 _____ 列。横线卡得 3 分的位置在第 _____ 行。

你会发现：竖线卡在竖线上得高分，横线卡在横杠上得高分。两张表合起来，就“看出”了一个 H。AlexNet 第一层就做这件事，只是它的 96 张卡片是自己学出来的。

9 动手：我是分类器

玩法（3—4 个人，约 15 分钟，需要 6 张小纸片、铅笔）

- 你们来当“**分类器**”（给图片分类的程序）。小组写 3 条认猫规则，全满足才算“是猫”。

规则 1：_____ 规则 2：_____ 规则 3：_____

- 出题人偷偷画好下表的 6 张简笔画，一张张亮出。其他人**只按规则**判断，再对答案。

出题人画什么	按规则：是猫吗	真正答案	✓ 或 ✕
1. 正面坐着的花猫		是猫	
2. 睡成一团的猫，看不见腿		是猫	
3. 一只狐狸		不是猫	
4. 一只老虎		不是猫	
5. 尖耳朵的小狗		不是猫	
6. 只画了圆脸的卡通猫		是猫	

- 改一条规则补上漏洞，再判断一遍。这次错了几张？ _____

你会发现：规则宽了，狐狸也算猫；规则严了，睡觉的猫不算猫。

10 词语小卡片

卷积（小窗口扫描）

小窗口一格一格扫，找一种小花纹。

深度学习

很多层网络，把小花纹拼成大东西。

显卡（GPU）

会同时做很多道简单题，训练快。

数据增强（一图变多题）

翻转、剪裁、改颜色，一图变多题。

11 想一想

1. 比赛为什么允许猜 5 次？举出两种很难分清的东西。

2. AlexNet 把练习图左右翻转，却不上下颠倒。你觉得为什么？

3. 你学认动物时，是靠背规则，还是靠看了很多例子？举一个例子。

一句话记住：不用人写规则，让电脑拿小窗口自己从大量图片里学，它就能认出猫和车。

给家长和老师：本课讲的是 Alex Krizhevsky、Ilya Sutskever、Geoffrey E. Hinton 2012 年在 NIPS（今 NeurIPS）发表的论文 [ImageNet Classification with Deep Convolutional Neural Networks](#)。网络有 5 个卷积层、3 个全连接层、约 6000 万参数，在 2 块 GTX 580（各 3 GB）上训练 5 到 6 天。ILSVRC-2012 的 top-5 错误率为 15.3%，第二名为 26.2%。“快约 6 倍”指 CIFAR-10 上 4 层网络 ReLU 对 tanh 的对比；dropout 用在前两个全连接层，概率 0.5。

真实的卷积核是一组可正可负的数，“数黑格”只是示意。论文展示了第一层学到的边缘和颜色斑块；“再拼成眼睛、轮子”来自后续可视化研究（如 2013 年 ZFNet）。热身小孔图、柯基犬的 5 个猜测、认猫检查单、“一个班同学”比喻、H 字格子图和分类器游戏都是示意。小窗口活动答案：竖线卡第 1-3 行都是 3、1、1、3，第 4 行是 3、0、0、3；横线卡第 2 行是 2、3、3、2，其余各行都是 1。详见 [第 8 天中文导读](#)，代码见 [30u30 第 8 天目录](#)（英文）。