

会跳格的小窗口

本课讲 2015 年的一篇论文，2016 年在一个大会上正式发表。科学家 Fisher Yu 和 Vladlen Koltun 让电脑的小窗口“跳着看”，给图片上每一个点都标对名字。

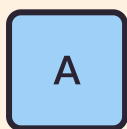
适合 10 岁左右

约 40 分钟

需要：铅笔、方格纸、卡纸、剪刀、一张旧杂志照片、一位同伴

1 热身：这一小块是什么？

下面 4 个小方块，是从 4 张照片里剪下来的。它们可能是什么？每块写出 2 种猜测。



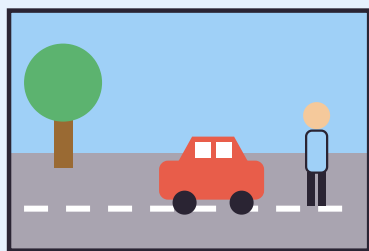
A: _____ B: _____

C: _____ D: _____

蓝色的 A，可能是天空，也可能是大海、蓝墙、蓝衣服。只看一小块，**谁也说不准**。要是能看到它周围的东西，就容易多了。

2 比“认出来”更难的事：给每个点涂颜色

第 8 课里，电脑学会了看一张图，说出“这是一只猫”。整张图只要一个答案。还有一件更难的事：给图上**每一个小点**都标出它属于什么，是天空、马路、人还是车。这件事叫**语义分割**。“语义”就是“意思”，“分割”就是“分成一块一块”。



原来的照片

■ 天空 ■ 马路 ■ 树 ■ 车 ■ 人



电脑要交的“涂色图”

示意图：右边每种东西只用一种颜色，车轮也算车。这很像填涂色书。可是照片上没有“1 号涂蓝色”这样的号码，电脑得自己判断。

3 只看一小块，分不清

仔细看上一节的照片。那个人穿着一件**浅蓝色的衣服**，和天空的颜色一模一样。

电脑看到的范围	它会怎么想
只看衣服上的一个点	“一片浅蓝……是天空吧？”（猜错了）
看到点的周围一小圈	“还是浅蓝，有一条黑边。说不准。”
看到周围一大片	“上面有头，下面有腿，这是人的衣服！”（猜对了）

所以，要标对一个点，电脑必须看到它周围很大一片。一个点能看到的周围范围，我们叫它视野，科学家叫它“感受野”。视野越大，越不容易认错。

4 老办法：缩小再放大

第 8 课讲过，电脑看图用一个红窗口扫描。窗口一次只看一小块，比如 3×3 共 9 个格子。窗口小，视野就小。以前的程序有一个办法：把图片一次次缩小。图变小了，同样大的窗口就盖住了更多内容，就像站远一点看画。

原图：细杆子、车轮

缩小

再放大：细节没了

可是缩小会丢东西。最后要给每个点涂颜色，图又得放大回来。丢掉的细节，放大也找不回来。

示意图：格子越黑，东西越多。缩小时，4 个格子并成 1 个。细杆子变成又宽又淡的带子，车轮也看不出来了。

想看得远，就要缩小。想保住细节，就不能缩小。两位作者想两样都要。

5 新办法：让窗口跳着看

他们的办法是：窗口还是 9 个格子，但格子之间留出空隙，跳着看。“间隔”表示从一个格子到下一个格子要走几步。

间隔 1：格子挨着
看到 3×3 一片

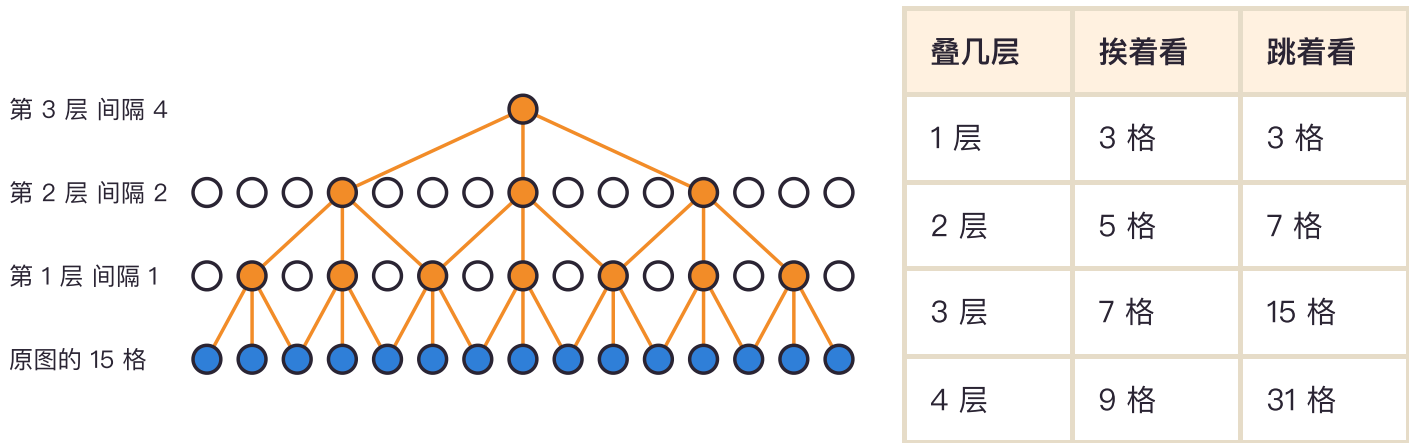
间隔 2：中间空 1 格
看到 5×5 一片

间隔 4：中间空 3 格
看到 9×9 一片

橙色是窗口真正看的 9 个格子，蓝框是它盖住的范围。

三个窗口都只看 9 个格子，花的力气一样多。间隔越大，盖住的范围越大。这种跳着看的窗口，叫空洞卷积。“卷积”就是第 8 课那种小窗口扫描，“空洞”就是窗口里留了洞。

一层比一层跳得远。只跳一次，中间的格子会漏掉。所以他们把窗口叠成好几层，间隔一层翻一倍：1、2、4、8.....下面用一排格子来看（示意图）：



最上面那 1 个点，通过 3 层，连到了原图一整排 15 个格子，一个也没漏。右边的表里，挨着看每多一层只看 2 格，跳着看每多一层视野差不多翻一倍。而且**图片一直没有缩小**，细节都还在。

6 上下文模块：专门看大范围的小组

“上下文”就是一个东西周围的情况。作者用跳着看的窗口，搭了一个专门看大范围的小组，叫**上下文模块**。它一共 8 层，中间几层的间隔是 1、2、4、8、16。程序先用“认图部分”给每个点一个初步答案，上下文模块再看看周围一大片，改掉错的。比如，认图部分把衣服上的点标成了“天空”。上下文模块一看，这个点上面是头，下面是腿，就改成“人”（我们编的例子）。

- **刚装上时什么都不改。**上下文模块一开始原样交出答案，再慢慢学会改错，就不会把对的答案改坏。
- **拿掉两次缩小。**作者从一个认图程序里拿掉两次“缩小”，换成跳着看的窗口。只做这一步，就比当时两种有名的方法标得更准。加上上下文模块，又准了一些。在城市街道的照片上，它也好用。

7 这和今天有什么关系

“跳着看”很快跑到了别的地方。声音是一长串很密的数字。2016 年，WaveNet 程序用“跳着听”生成说话的声音，间隔也是一层翻一倍。2020 年以后，有些读长文章的 AI 也借用了“跳着看”。

说实话，今天给图片涂颜色的 AI，很多已经改用“注意力”（第 13 课会讲），每个点能一下子看到整张图。不过，跳着看的窗口还在一些程序里用着。它的道理也没过时：**不一定要每个格子都看，聪明地跳着看，也能看得又远又清楚。**

8 动手：做三张跳格卡片

玩法（1—2 人，约 15 分钟）

1. 在 3 张卡纸上各画 9 个小方孔，排成 3 行 3 列，每个孔和方格纸的 1 格一样大。A 的孔挨着（间隔 1），B 的孔之间空 1 格（间隔 2），C 的孔之间空 3 格（间隔 4）。请大人帮忙剪孔。
2. 把卡片放在方格纸上，透过每个孔点一个点。拿开卡片，数一数点了几个点，最外面的点围成的正方形一边有几格。
3. 先猜间隔 8 的卡片 D 会盖住多大，再做一张验证。

卡片	间隔	点了几个点	盖住的范围（几格 × 几格）
A	1		
B	2		
C	4		

你会发现：每张卡片都只点 9 个点，间隔越大，盖住的范围越大。点和点中间漏掉的格子，就是电脑要一层层叠起来的原因。

9 动手：只露一小块，猜猜是什么

玩法（2—3 人，约 10 分钟）

1. 准备 2 张旧杂志照片和 3 张白纸。白纸中间各剪一个方孔：小孔 1 厘米，中孔 3 厘米，大孔 6 厘米。
2. 出题人选照片上的一个点，用小孔的纸盖住照片，孔对准这个点。猜的人说：这个点是什么东西的一部分？
3. 换中孔、大孔再猜，孔对准同一个点。最后拿开纸核对。换照片，换人出题。

照片	小孔猜	中孔猜	大孔猜	正确答案
1				
2				

你会发现：看到的周围越多，越容易猜对。电脑给每个点涂颜色，也需要很大的视野。

10 词语小卡片

语义分割

给图片上每一个点都标出它属于什么，就像给涂色书的每一块涂对颜色。

视野（感受野）

一个点能看到的周围范围。视野越大，越不容易认错。

空洞卷积

格子之间留空、跳着看的小窗口。花一样的力气，看到更大一片。

上下文模块

用跳着看的窗口搭成的小组，专门看大范围，帮忙改掉标错的点。

11 想一想

1. 热身时，哪个小方块最难猜？要看到周围的什么东西，你才能猜对？

2. 一辆会自己开的车在看路。为什么只说“前面有人”不够，还要知道每个点是马路还是人？

3. 跳着看会漏掉中间的格子。为什么一层一层叠起来，就不怕漏了？试着用第 5 节的图说一说。

一句话记住：窗口跳着看，一层比一层跳得远，图片不用缩小，也能看到一大片。

给家长和老师：本课讲的是 Fisher Yu 和 Vladlen Koltun 的论文 [Multi-Scale Context Aggregation by Dilated Convolutions](#) (arXiv 2015, ICLR 2016)。论文用空洞卷积（dilated convolution）让感受野随层数指数增长，同时不降分辨率、不增参数。它设计了可接在分割网络后面的上下文模块（空洞率 1、1、2、4、8、16、1，再接 1×1 卷积，恒等初始化），并去掉 VGG-16 最后两个池化层作为前端。前端在 Pascal VOC 2012 上已优于 FCN-8s 和 DeepLab，加上下文模块后再提升，在 CamVid、KITTI、Cityscapes 上同样有效。

街道小图、蓝衣服的例子、“缩小再放大”的格子图都是示意。第 5 节用一排格子演示感受野 3、7、15、31，原文是二维 3×3 卷积，边长相同。“认图部分”指前端，“缩小”指池化和步长。讲义未写具体分数。想深入可看本站的 [第 11 天中文导读](#)，代码见开源项目 [30u30 第 11 天目录](#)（英文）。