

第 23 课

摘要卡只记大局

本课讲 2016 年的一篇论文。Xi Chen、Diederik Kingma 等科学家想了一个巧办法：故意让“画图的人”看不远，逼电脑把一张图的“大局”记在一小张卡片上。

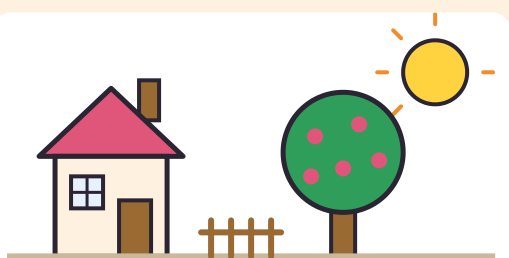
适合 10 岁左右

约 40 分钟

需要：铅笔、白纸、小卡片、2-4 位同伴

1 热身：看一眼，写一张卡

看下面这幅画 5 秒钟，然后用手盖住它。在卡片上写**最多 8 个字**，让没看过画的朋友能大致画出来。



我的卡片：□□□□□□□□

看你的卡片，数一数：

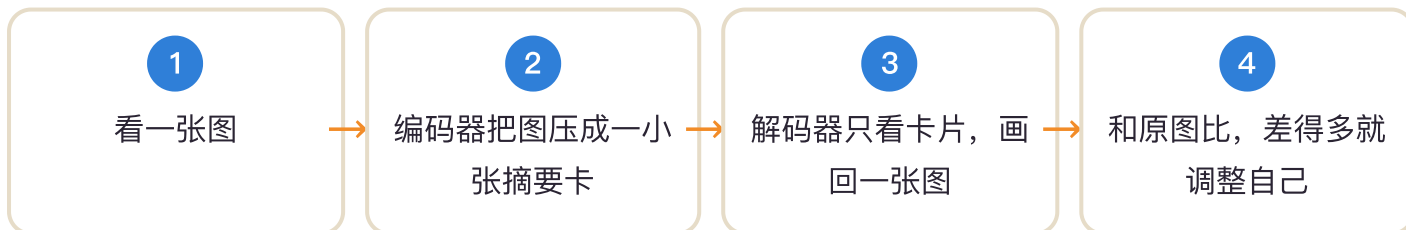
写了“有什么、在哪里”的字：_____ 个

写了“几个苹果、窗户几格”的字：_____ 个

大多数人会写“左房右树有太阳”。8 个字装不下所有东西，你会先写**大局**：有什么，在哪里。苹果几个、篱笆几根，这些**细节**只好丢掉。今天的论文，就是教电脑也这样挑。

2 自动编码器：压成一张摘要卡，再画回来

科学家做过一种程序，专门玩刚才的游戏。它分成两半：



🔄 换一张图，重复几百万次

第 15 课讲过：**编码器**负责写卡片，**解码器**负责照卡片写或画。卡片上写什么，是两个程序练出来的，没人教。这种程序叫**自动编码器**。

和第 4、5 课连起来

卡片比原图小得多，这就是**压缩**。第 5 课的压缩要一字不差地还原。这里的卡片**允许丢东西**，画回来大体像就行。这叫**有损压缩**。可是，**该丢什么，该留什么？**这正是论文要回答的。

3 变分自动编码器：随手抽一张卡，画出新图

2013 年，Kingma 和 Welling 提出了升级版**变分自动编码器**，缩写 VAE。它多了一个本事：**能凭空画出新图**。

秘密在于一条规矩：所有摘要卡都要写得“像样”，长得差不多。随手编一张像样的卡片，解码器就能画出没人见过的新图。可是 2013 年的 VAE 画的图常常**糊糊的**：解码器一下子画出所有的点，点和点不商量。

4 画家太厉害，卡片就没人看

2016 年，科学家做出了很厉害的画法 PixelCNN。它**一个点一个点接着画**，从左到右、从上到下，每画一个新点都先看前面画好的点。这和第 1 课的“猜下一个字”是同一个思路，叫**自回归**。图清楚多了。大家请它来当 VAE 的解码器，结果出了怪事。

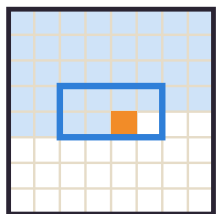
怪事：卡片变成了一张白纸

练着练着，编码器不往卡片上写东西了。解码器也不看卡片，自己照样画得不错。这个毛病叫**后验坍塌**，意思就是：**卡片空了，没人看**。

为什么？论文用第 4、5 课的想法解释：电脑练习时，要让**“卡片 + 画家”的总长度越短越好**。卡片上每写一个字都算长度，还要额外多付一点。画家自己能猜到的，写进卡片反而更费。所以画家越厉害，卡片越空。

5 论文的办法：只给画家一个小窗口

论文作者反过来想：**不让画家看那么远**。他们规定，画家每画一个新点，只能看它**身边一小片**已经画好的点。第 8 课讲过的“小窗口”，在这里又用上了。



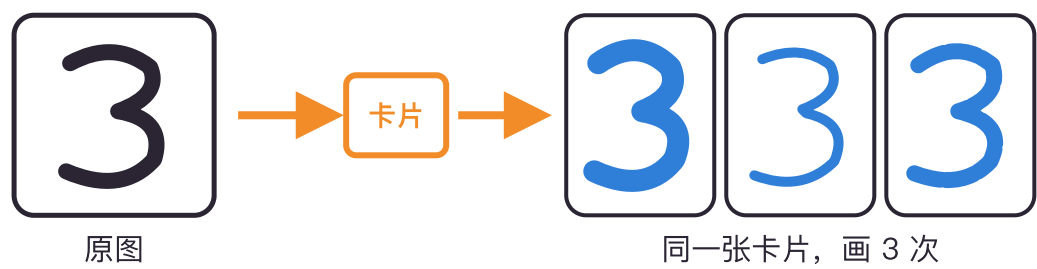
浅蓝格子已经画好，橙色格子正要画。画家只能看到蓝框里的几格。

所以它能把一条线接着画顺，能补上毛边和花纹。可它**不知道整张图画的是什么**。

画家自己能补的，卡片上不用写。画家补不了的，只能写在卡片上：

谁负责	负责什么	比如（手写数字）
摘要卡	大局：画的是什么，整体形状，在哪里	“这是一个 3，写得偏左”
小窗口画家	细节：附近一小片的样子	笔画的粗细，边上的小毛刺

卡片学会了只记大局，丢掉细节。这就是论文名字“变分有损自动编码器”（VLAE）里“有损”的意思。



示意图：3 次画的都是“偏左的 3”，大局一样。笔画粗细和毛边每次不同，这是画家自己补的细节。

6 再加一点：对卡片有更好的预期

VAE 原来要求所有卡片挤成同一种简单样子。作者换了一个更聪明的先验。先验就是“还没看到图，先猜卡片大概长什么样”。

打个比方（示意）：组长收作业，知道大家常写哪几种答案。常见的答案记一个记号，少见的才多写几笔。预期越准，记录越短。电脑越清楚卡片常长什么样，写卡片就越省长度。

小窗口画家加上更好的先验，VLAE 在手写数字、各国文字的手写字母、物体剪影这几个图片考试里压缩得当时最好。在小彩色照片上，它也接近当时最好的程序。压缩得越短，越懂规律。这正是第 4、5 课说的：压缩就是理解。

7 这和今天有什么关系

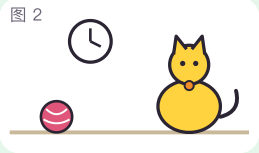
“卡片管大局，画家管细节”的分工今天还在用。很多会画画的 AI 先把图压成小一号的“摘要”，在摘要上定好画什么、放哪里，再由解码器补细节。做法和 VLAE 不同，思路相通。

ChatGPT、Claude 这类聊天 AI 是“猜下一个字”的超强画家，没有摘要卡。给它们硬加一张卡片，卡片常被晾在一边，这篇论文解释了原因。也要诚实地说：今天画图最好的 AI，已经不走 VAE 这条路了。

8 动手：一句话画画

玩法（2 个人，约 12 分钟）：第 4 课比谁说得短，这次看短话留下了什么

- 1. A 当“编码器”，只有 A 看图。B 当“解码器”，先遮住图。
- 2. A 看图 1，在卡片上写一句话，**不超过 10 个字**，不能说话和比画。
- 3. B 只看卡片画出图 1，再对照原图打 ✓ 或 ✕。换图 2，交换角色。



	大局：有哪些东西	大局：各在哪里	大局：大小差不多	细节：小数目（星、鸟）	细节：花纹和装饰
图 1					
图 2					

你会发现：“大局”格大多是 ✓，“细节”格常常是 ✕。10 个字装不下全部，编码器会先写大局。这就是有损压缩。

9 动手：摘要卡写什么

玩法（4 个人，约 15 分钟）

- 1. 一人当“编码器”看雪人图，其他 3 人当“画家”。
- 2. 编码器写 3 张卡，每张**不超过 12 个字**：卡 1 只写细节，卡 2 只写大局，卡 3 两样都写。
- 3. 每位画家只照一张卡画。3 张画并排，和原图比。



卡片	写了什么	画得像吗（1—5 分）
卡 1 细节		
卡 2 大局		
卡 3 都有		

你会发现：只写细节的卡片最没用，画家不知道画的是雪人。只写大局的卡片，画得整体最像。帽子和纽扣，画家不看卡也能补。VLAЕ 的摘要卡学会的就是这个。

10 词语小卡片

自动编码器

先把图压成一小张摘要卡，再照卡片画回来的电脑程序。写卡的叫编码器，画画的叫解码器。

变分自动编码器 (VAE)

卡片都写得“像样”的自动编码器。随手编一张像样的卡片，就能画出新图。

自回归

一个点一个点（或一个字一个字）接着往下，每一步都看前面已经有的。

有损压缩

压缩时允许丢掉一些东西。VLAЕ 的卡片留下大局，丢掉细节。

11 想一想

1. 画家能看整张图时，卡片会变空。为什么画家看得越少，卡片反而写得越多？

2. 如果把画家的小窗口一点点放大，你猜卡片上会发生什么？

3. 你写读书笔记时，会写故事的大局还是每个细节？什么时候细节也值得写？

一句话记住：画家自己能补的，卡片就不写。只给画家一个小窗口，卡片就学会了只记大局。

给家长和老师：本课讲的是 Xi Chen、Diederik P. Kingma、Tim Salimans 等人 2016 年的论文 [Variational Lossy Autoencoder](#) (ICLR 2017)。论文用 Bits-Back 编码的视角解释：自回归解码器（如 PixelCNN）足够强时，VAE 会忽略隐变量（后验坍塌）。作者把解码器的感受野限制在局部，迫使隐变量只编码全局结构，成为丢弃局部纹理的有损表示；再用自回归流作为先验。VLAЕ 在 MNIST、OMNIGLOT、Caltech-101 Silhouettes 上取得当时最好的对数似然，在 CIFAR-10 上达到 2.95 bits/dim。

讲义中的“摘要卡”“画家”“组长收作业”是比喻；“卡片 + 画家的总长度”是对训练目标（ELBO 即平均编码长度）的简化说法，未讲 KL 项和 Bits-Back 细节。热身图、活动图、“同一张卡片画 3 次”的数字 3 都是示意，不是论文原图。想深入可看本站的 [第 23 天中文导读](#)，代码见开源项目 [30u30 第 23 天目录](#)（英文）。