

摘要卡只記大局

本課講 2016 年的一篇論文。Xi Chen、Diederik Kingma 等科學家想了一個巧辦法：故意讓「畫圖的人」看不遠，逼電腦把一張圖的「大局」記在一小張卡片上。

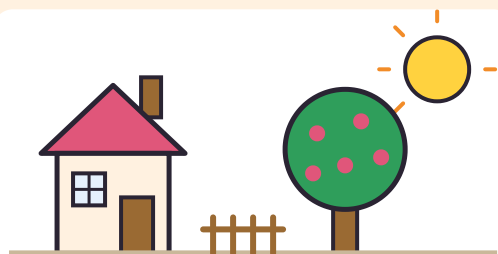
適合 10 歲左右

約 40 分鐘

需要：鉛筆、白紙、小卡片、2-4 位同伴

1 熱身：看一眼，寫一張卡

看下面這幅畫 5 秒鐘，然後用手蓋住它。在卡片上寫**最多 8 個字**，讓沒看過畫的朋友能大致畫出來。



我的卡片：□□□□□□□□

看你的卡片，數一數：

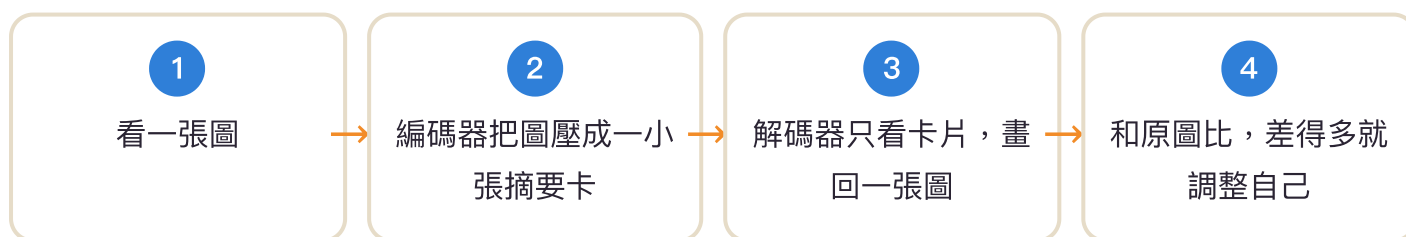
寫了「有什麼、在哪裡」的字：_____ 個

寫了「幾個蘋果、窗戶幾格」的字：_____ 個

大多數人會寫「左房右樹有太陽」。8 個字裝不下所有東西，你會先寫**大局**：有什麼，在哪裡。蘋果幾個、籬笆幾根，這些**細節**只好丟掉。今天的論文，就是教電腦也這樣挑。

2 自動編碼器：壓成一張摘要卡，再畫回來

科學家做過一種程序，專門玩剛才的遊戲。它分成兩半：



↻ 換一張圖，重複幾百萬次

第 15 課講過：**編碼器**負責寫卡片，**解碼器**負責照卡片寫或畫。卡片上寫什麼，是兩個程序練出來的，沒人教。這種程序叫**自動編碼器**。

和第 4、5 課連起來

卡片比原圖小得多，這就是**壓縮**。第 5 課的壓縮要一字不差地還原。這裡的卡片**允許丟東西**，畫回來大體像就行。這叫**有損壓縮**。可是，**該丟什麼，該留什麼？**這正是論文要回答的。

3 變分自動編碼器：隨手抽一張卡，畫出新圖

2013 年，Kingma 和 Welling 提出了升級版**變分自動編碼器**，縮寫 VAE。它多了一個本事：**能憑空畫出新圖**。

秘密在於一條規矩：所有摘要卡都要寫得「像樣」，長得差不多。隨手編一張像樣的卡片，解碼器就能畫出沒人見過的新圖。可是 2013 年的 VAE 畫的圖常常**糊糊的**：解碼器一下子畫出所有的點，點和點不商量。

4 畫家太厲害，卡片就沒人看

2016 年，科學家做出了很厲害的畫法 PixelCNN。它**一個點一個點接著畫**，從左到右、從上到下，每畫一個新點都先看前面畫好的點。這和第 1 課的「猜下一個字」是同一個思路，叫**自迴歸**。圖清楚多了。大家請它來當 VAE 的解碼器，結果出了怪事。

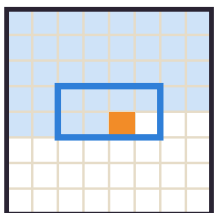
怪事：卡片變成了一張白紙

練著練著，編碼器不往卡片上寫東西了。解碼器也不看卡片，自己照樣畫得不錯。這個毛病叫**後驗坍塌**，意思就是：**卡片空了，沒人看**。

為什麼？論文用第 4、5 課的想法解釋：電腦練習時，要讓「**卡片 + 畫家**」的**總長度越短越好**。卡片上每寫一個字都算長度，還要額外多付一點。畫家自己能猜到的，寫進卡片反而更費。所以畫家越厲害，卡片越空。

5 論文的辦法：只給畫家一個小窗口

論文作者反過來想：**不讓畫家看那麼遠**。他們規定，畫家每畫一個新點，只能看它**身邊一小片**已經畫好的點。第 8 課講過的「小窗口」，在這裡又用上了。



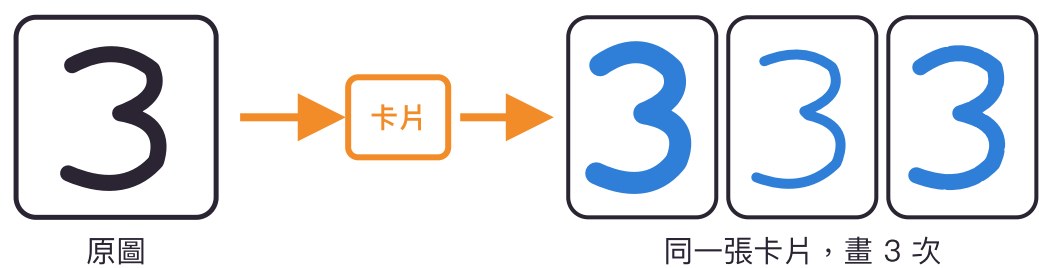
淺藍格子已經畫好，橙色格子正要畫。畫家只能看到藍框裡的幾格。

所以它能把一條線接著畫順，能補上毛邊和花紋。可它**不知道整張圖畫的是什麼**。

畫家自己能補的，卡片上不用寫。畫家補不了的，**只能寫在卡片上**：

誰負責	負責什麼	比如（手寫數字）
摘要卡	大局：畫的是什麼，整體形狀，在哪裡	「這是一個 3，寫得偏左」
小窗口畫家	細節：附近一小片的樣子	筆畫的粗細，邊上的小毛刺

卡片學會了**只記大局，丟掉細節**。這就是論文名字「變分有損自動編碼器」（VAE）裡「有損」的意思。



示意圖：3 次畫的都是「偏左的 3」，大局一樣。筆畫粗細和毛邊每次不同，這是畫家自己補的細節。

6 再加一點：對卡片有更好的預期

VAE 原來要求所有卡片擠成同一種簡單樣子。作者換了一個更聰明的**先驗**。先驗就是「還沒看到圖，先猜卡片大概長什麼樣」。

打個比方（示意）：組長收作業，知道大家常寫哪幾種答案。常見的答案記一個記號，少見的才多寫幾筆。預期越準，記錄越短。電腦越清楚卡片常長什麼樣，寫卡片就越省長度。

小窗口畫家加上更好的先驗，VAE 在手寫數字、各國文字的手寫字母、物體剪影這幾個圖片考試裡壓縮得**當時最好**。在小彩色照片上，它也接近當時最好的程序。壓縮得越短，越懂規律。這正是第 4、5 課說的：**壓縮就是理解**。

7 這和今天有什麼關係

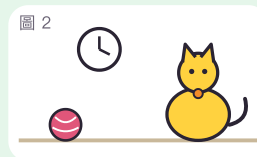
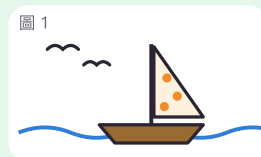
「卡片管大局，畫家管細節」的分工今天還在用。很多會畫畫的 AI 先把圖壓成小一號的「摘要」，在摘要上定好畫什麼、放哪裡，再由解碼器補細節。做法和 VLAЕ 不同，思路相通。

ChatGPT、Claude 這類聊天 AI 是「猜下一個字」的超強畫家，沒有摘要卡。給它們硬加一張卡片，卡片常被晾在一邊，這篇論文解釋了原因。也要誠實地說：今天畫圖最好的 AI，已經不走 VAE 這條路了。

8 動手：一句話畫畫

玩法（2 個人，約 12 分鐘）：第 4 課比誰說得短，這次看短話留下了什麼

1. A 當「編碼器」，只有 A 看圖。B 當「解碼器」，先遮住圖。
2. A 看圖 1，在卡片上寫一句話，**不超過 10 個字**，不能說話和比畫。
3. B 只看卡片畫出圖 1，再對照原圖打 ✓ 或 ✕。換圖 2，交換角色。



	大局：有哪些東西	大局：各在哪裡	大局：大小差不多	細節：小數目（星、鳥）	細節：花紋和裝飾
圖 1					
圖 2					

你會發現：「大局」格大多是 ✓，「細節」格常常是 ✕。10 個字裝不下全部，編碼器會先寫大局。這就是有損壓縮。

9 動手：摘要卡寫什麼

玩法（4 個人，約 15 分鐘）

1. 一人當「編碼器」看雪人圖，其他 3 人當「畫家」。
2. 編碼器寫 3 張卡，每張**不超過 12 個字**：卡 1 只寫細節，卡 2 只寫大局，卡 3 兩樣都寫。
3. 每位畫家只照一張卡畫。3 張畫並排，和原圖比。



卡片	寫了什麼	畫得像嗎（1—5 分）
卡 1 細節		
卡 2 大局		
卡 3 都有		

你會發現：只寫細節的卡片最沒用，畫家不知道畫的是雪人。只寫大局的卡片，畫得整體最像。帽子和鈕扣，畫家不看卡也能補。VLAЕ 的摘要卡學會的就是這個。

10 詞語小卡片

自動編碼器

先把圖壓成一小張摘要卡，再照卡片畫回來的電腦程序。寫卡的叫編碼器，畫畫的叫解碼器。

變分自動編碼器 (VAE)

卡片都寫得「像樣」的自動編碼器。隨手編一張像樣的卡片，就能畫出新圖。

自迴歸

一個點一個點（或一個字一個字）接著往下，每一步都看前面已經有的。

有損壓縮

壓縮時允許丟掉一些東西。VLAЕ 的卡片留下大局，丟掉細節。

11 想一想

1. 畫家能看整張圖時，卡片會變空。為什麼畫家看得越少，卡片反而寫得越多？

2. 如果把畫家的小窗口一點點放大，你猜卡片上會發生什麼？

3. 你寫讀書筆記時，會寫故事的大局還是每個細節？什麼時候細節也值得寫？

一句話記住：畫家自己能補的，卡片就不寫。只給畫家一個小窗口，卡片就學會了只記大局。

給家長和老師：本課講的是 Xi Chen、Diederik P. Kingma、Tim Salimans 等人 2016 年的論文 [Variational Lossy Autoencoder](#) (ICLR 2017)。論文用 Bits-Back 編碼的視角解釋：自迴歸解碼器（如 PixelCNN）足夠強時，VAE 會忽略隱變量（後驗坍塌）。作者把解碼器的感受野限制在局部，迫使隱變量只編碼全局結構，成為丟棄局部紋理的有損表示；再用自迴歸流作為先驗。VLAЕ 在 MNIST、OMNIGLOT、Caltech-101 Silhouettes 上取得當時最好的對數似然，在 CIFAR-10 上達到 2.95 bits/dim。

講義中的「摘要卡」「畫家」「組長收作業」是比喻；「卡片 + 畫家的總長度」是對訓練目標（ELBO 即平均編碼長度）的簡化說法，未講 KL 項和 Bits-Back 細節。熱身圖、活動圖、「同一張卡片畫 3 次」的數字 3 都是示意，不是論文原圖。想深入可看本站的 [第 23 天中文導讀](#)，代碼見開源項目 [30u30 第 23 天目錄](#)（英文）。