

ML Lecture 16:

Unsupervised Learning - Generation

臺灣大學人工智慧中心 科技部人工智慧技術暨全幅健康照護聯合研究中心 <http://aintu.tw>

Generative Models

- Component-by-component (PixelRNN)
- Variational Auto-encoder (VAE)
- Generative Adversarial Network (GAN)

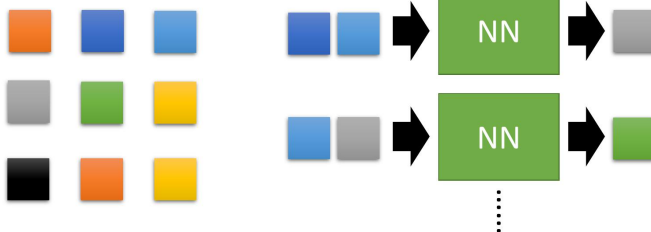
上述方法都很新，皆是都是近幾年提出的。

Component-by-component

Component-by-component

- Image generation

E.g. 3 x 3 images



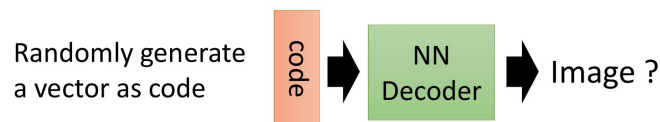
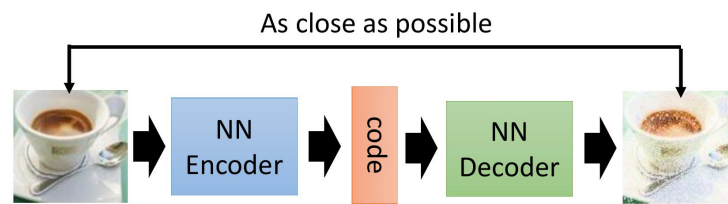
Can be trained just with a large collection of images
without any annotation

將圖片攤平，用 RNN 以之前的 pixel (RGB三圍向量)去 predict 下一個 pixel，把整張圖畫出來 (unsupervised)

不只用在圖片，還可以用在語音，像是WaveNet。也可以用在影片：給定一段 video，讓他predict 接下來會發生甚麼事。

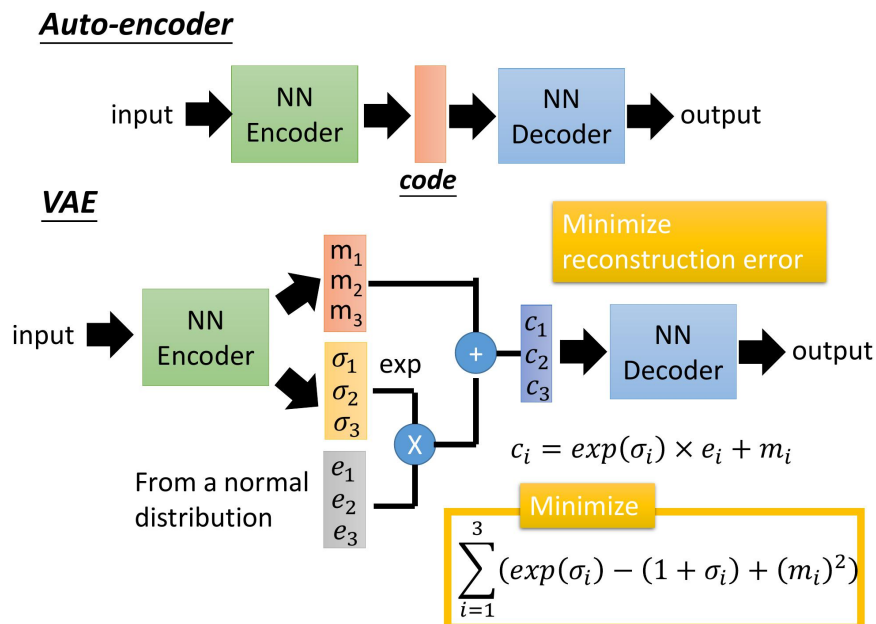
Auto-encoder

Auto-encoder



- Input: image => Encode: low dimension code => Decode: Image
 - 讓 Input & Output 兩張圖越接近越好
- Given random code => Decode: Image?
 - 效果差

Variance Auto-encoder

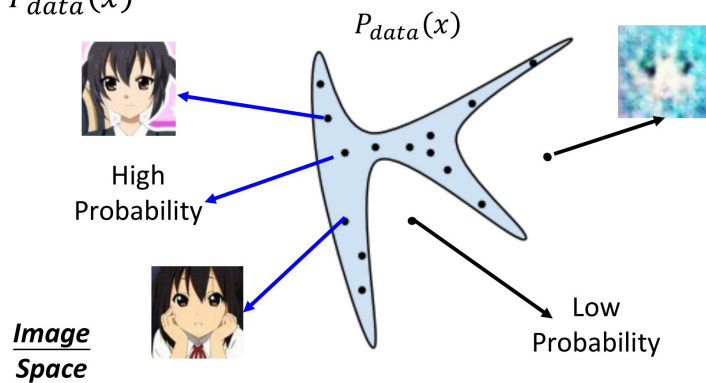


- 比起 Auto-encoder，加了小 trick：不直接 output code，而是先 output 兩個 vector，再與 random 出來的 Vector 做如圖的運算，當作 code。
- 目的是：minimize reconstruction error
- 雖然結果沒有 PixelRNN 清晰，但 code 的每一個 dimension 代表特定意思
- 也可以用來寫詩，將 IO 從 Image 換成 Sentence

Generative Adversarial Network (GAN)

Basic Idea of GAN

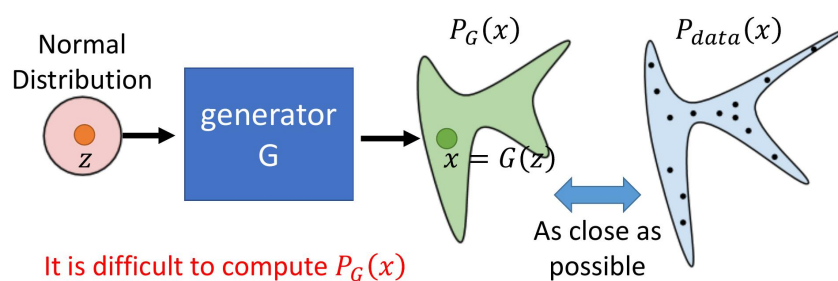
- The data we want to generate has a distribution $P_{data}(x)$



目的：從 Random 到特定的 Distribution。

Basic Idea of GAN

- A generator G is a network. The network defines a probability distribution.



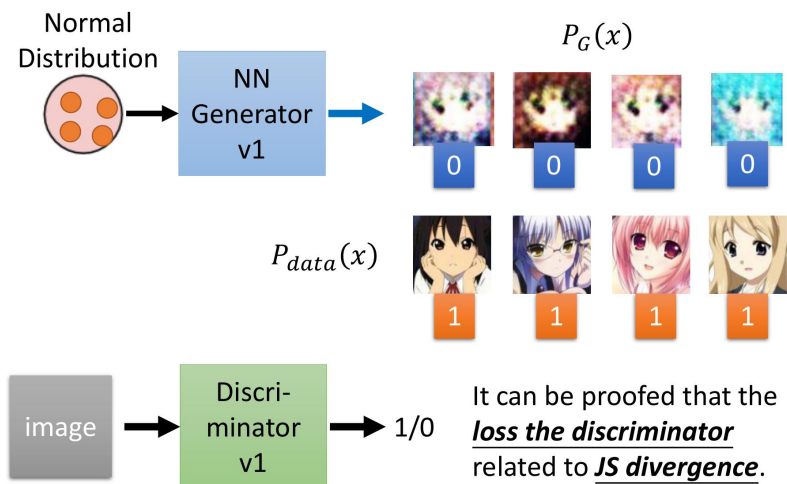
It is difficult to compute $P_G(x)$
We do not know what the distribution looks like.

<https://blog.openai.com/generative-models/>

Generator負責產出與 Target 相同 Distribution 的 Vector。

Discriminator負責分辨真偽 (真正的 Target & Generator 生的 Target)

Basic Idea of GAN



將兩個 Model 一起訓練，然後去 Minimize JS Divergence。

Basic Idea of GAN

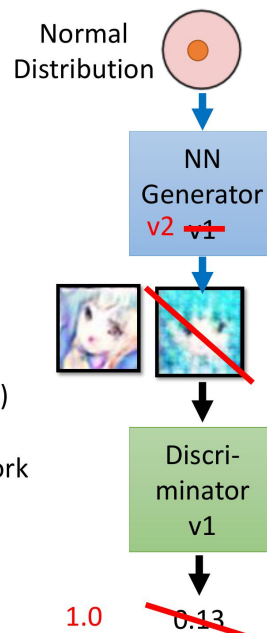
• Next step:

- Updating the parameters of generator
- To minimize the JS divergence

➡ The output be classified as "real" (as close to 1 as possible)

Generator + Discriminator = a network

Using gradient descent to update the parameters in the generator, but fix the discriminator

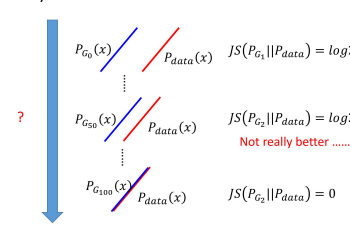
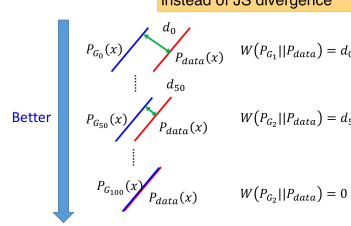


以 Gradient Descent 去 Update。

WGAN

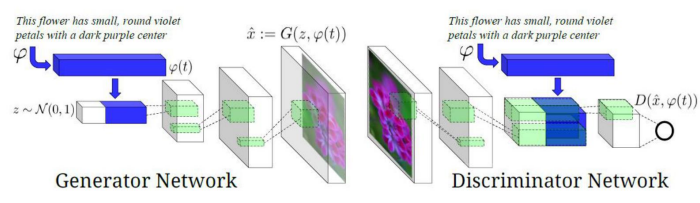
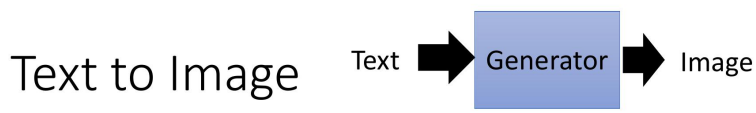
Using Wasserstein distance instead of JS divergence

GAN的進化版，原始的GAN在變好的過程中，JS divergence 不會逐漸著這變小，比較難train。

GAN	WGAN
<p>Why GAN is hard to train?</p>  <p>$p_{G_0}(x)$ $p_{data}(x)$ $JS(p_{G_1} p_{data}) = \log 2$</p> <p>?</p> <p>$p_{G_{50}}(x)$ $p_{data}(x)$ $JS(p_{G_2} p_{data}) = \log 2$ Not really better</p> <p>$p_{G_{100}}(x)$ $p_{data}(x)$ $JS(p_{G_2} p_{data}) = 0$</p>	<p>WGAN</p> <p>Using Wasserstein distance instead of JS divergence</p>  <p>$p_{G_0}(x)$ $p_{data}(x)$ d_0 $W(p_{G_1} p_{data}) = d_0$</p> <p>$p_{G_{50}}(x)$ $p_{data}(x)$ d_{50} $W(p_{G_2} p_{data}) = d_{50}$</p> <p>$p_{G_{100}}(x)$ $p_{data}(x)$ $W(p_{G_2} p_{data}) = 0$</p>

當變得更好時，得到的 JS divergence 越低

Text To Image



Scott Reed, Zeynep Akata, Xincheng Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee, "Generative Adversarial Text-to-Image Synthesis", ICML 2016

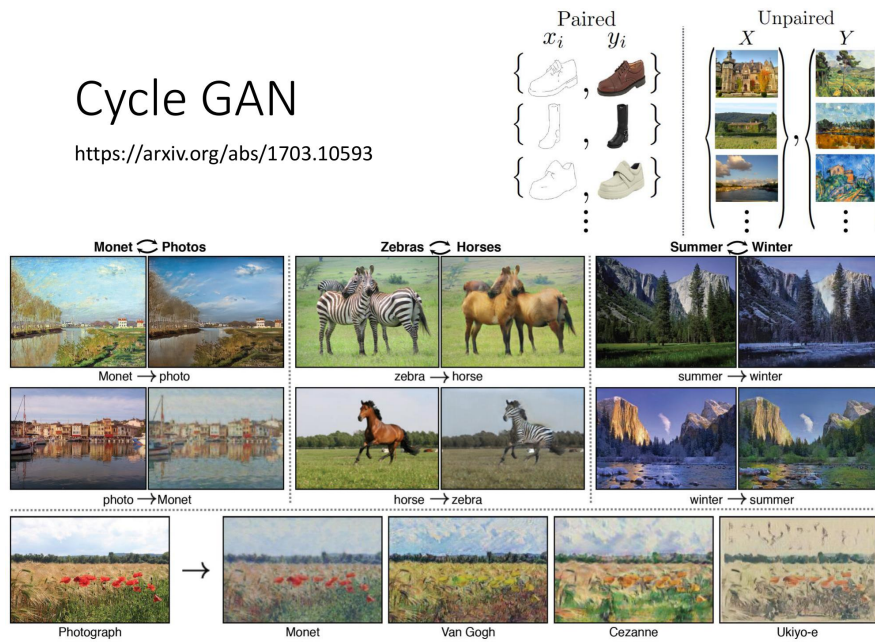
Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris Metaxas, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks", arXiv preprint, 2016

Scott Reed, Zeynep Akata, Santosh Mohan, Samuel Tenka, Bernt Schiele, Honglak Lee, "Learning What and Where to Draw", NIPS 2016

Cycle GAN

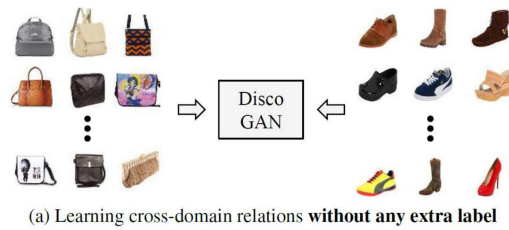
Cycle GAN

<https://arxiv.org/abs/1703.10593>



Disco GAN

Disco GAN



(a) Learning cross-domain relations **without any extra label**



(b) Handbag images (input) & **Generated** shoe images (output)



(c) Shoe images (input) & **Generated** handbag images (output)

<https://arxiv.org/abs/1703.05192>

GAN 的各種加強版

So many GANs
..... Just name a few

Modifying the Optimization of GAN	Different Structure from the Original GAN
fGAN	Conditional GAN
WGAN	Semi-supervised GAN
Least-square GAN	InfoGAN
Loss Sensitive GAN	BiGAN
Energy-based GAN	Cycle GAN
Boundary-seeking GAN	Disco GAN
Unroll GAN	VAE-GAN
.....