速覽圖像生成模型

(powered by Midjourney)

圖像生成

一張圖勝過千言萬語

一隻在奔跑的狗

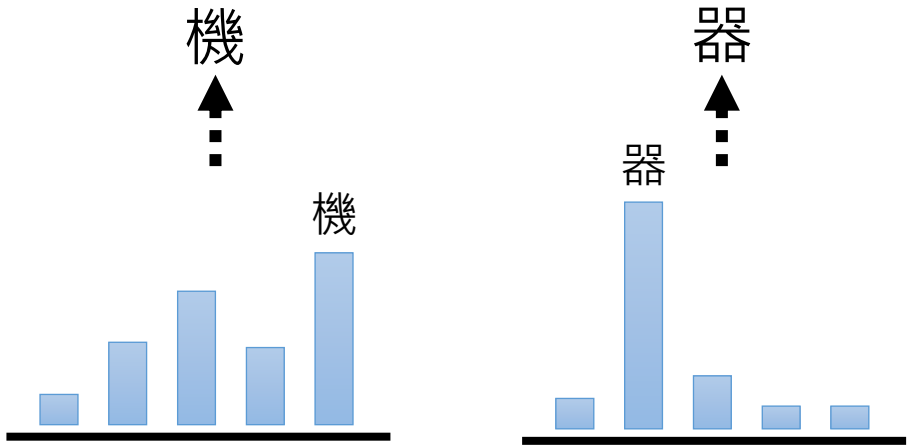
千言萬語的一部分

... 一隻在街道上奔跑的柯基狗 ...

機器需要大量腦補

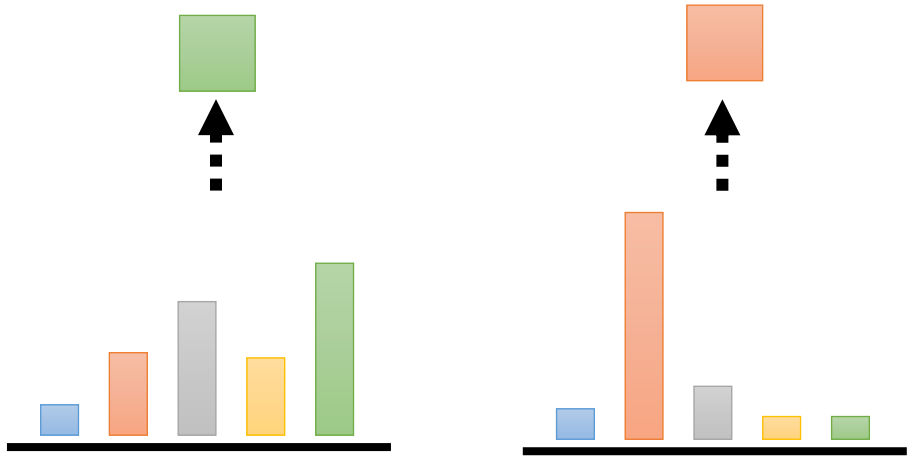
文字生成多採取 Autoregressive (各個擊破)

機器學習是甚麼？ → 文字生成 →



影像生成好像也可以
用同樣的道理

一隻在奔跑的狗 → 影像生成 →

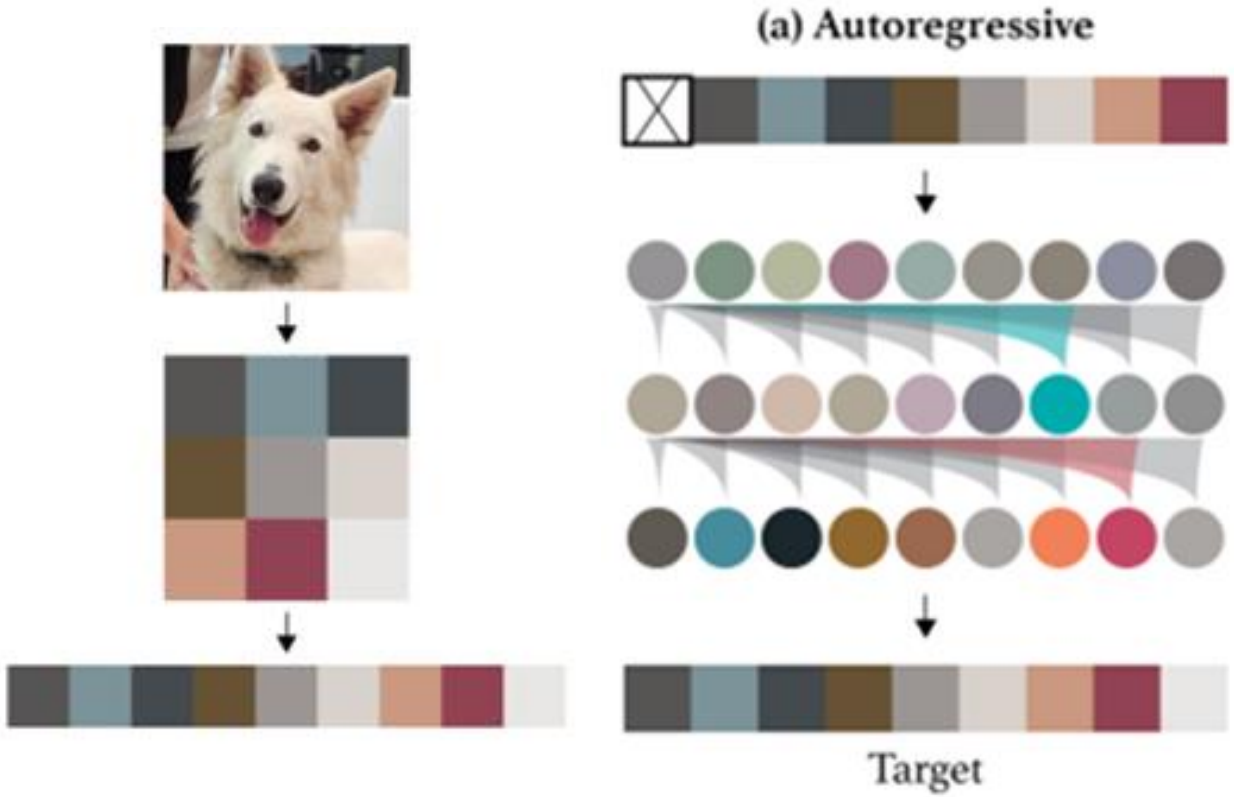


256 x 256 images → 65536 pixels

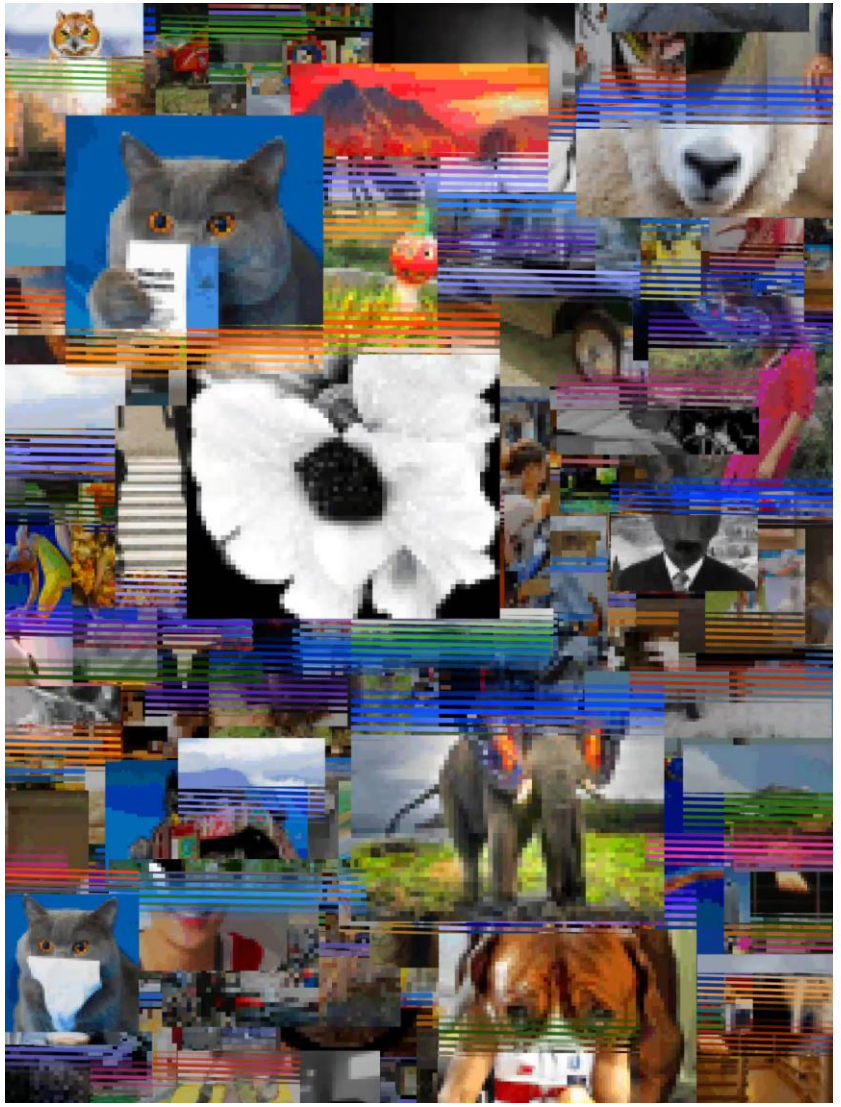
影像也可以採取 Autoregressive (各個擊破)

https://openai.com/blog/image-gpt/

影像版 GPT



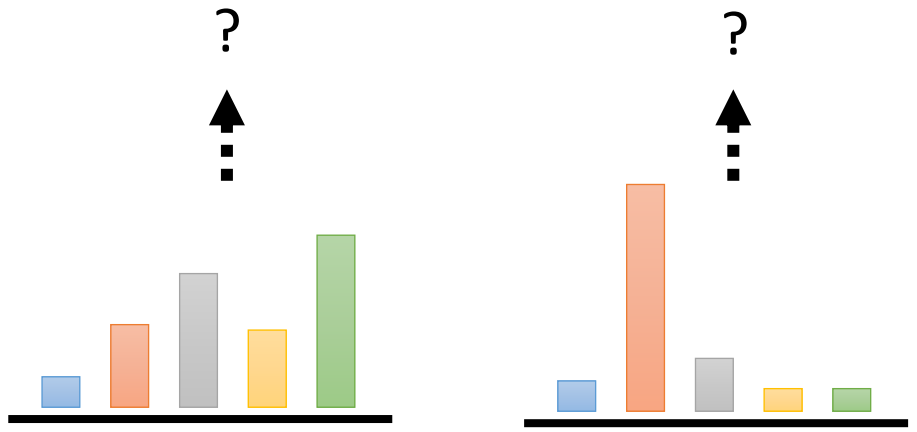
raster order



(powered by Midjourney)

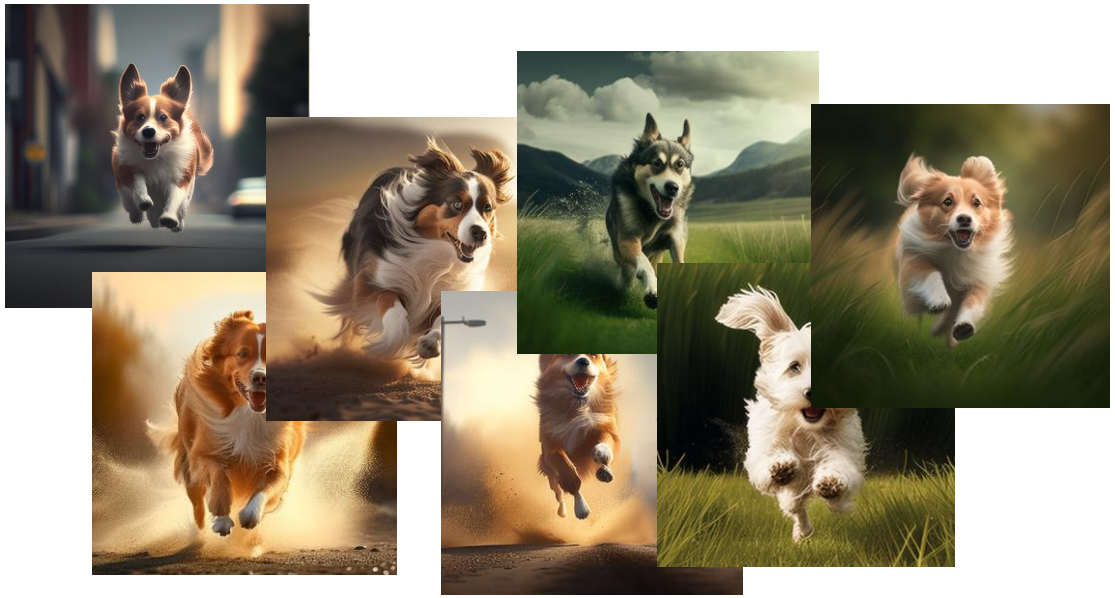
一次到位有甚麼問題

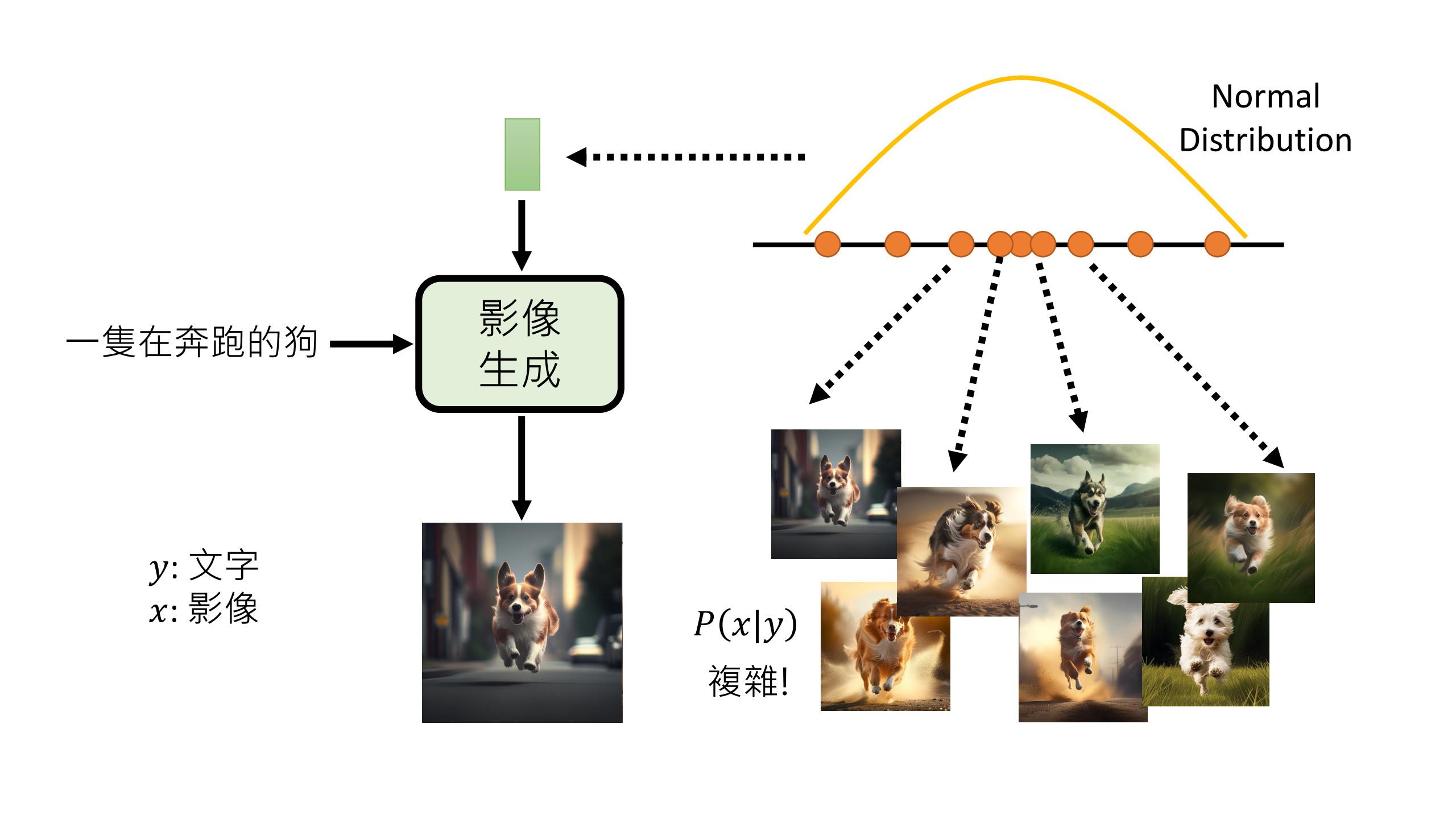
一隻在奔跑的狗 → 影像生成 →



256 x 256 images → 65536 pixels

每一個像素獨立繪製




Normal Distribution
一隻在奔跑的狗
影像生成
y: 文字
x: 影像
$P(x|y)$
複雜!

影像常用生成模型速覽

以下說明都非常簡略
詳細內容請見參考資料

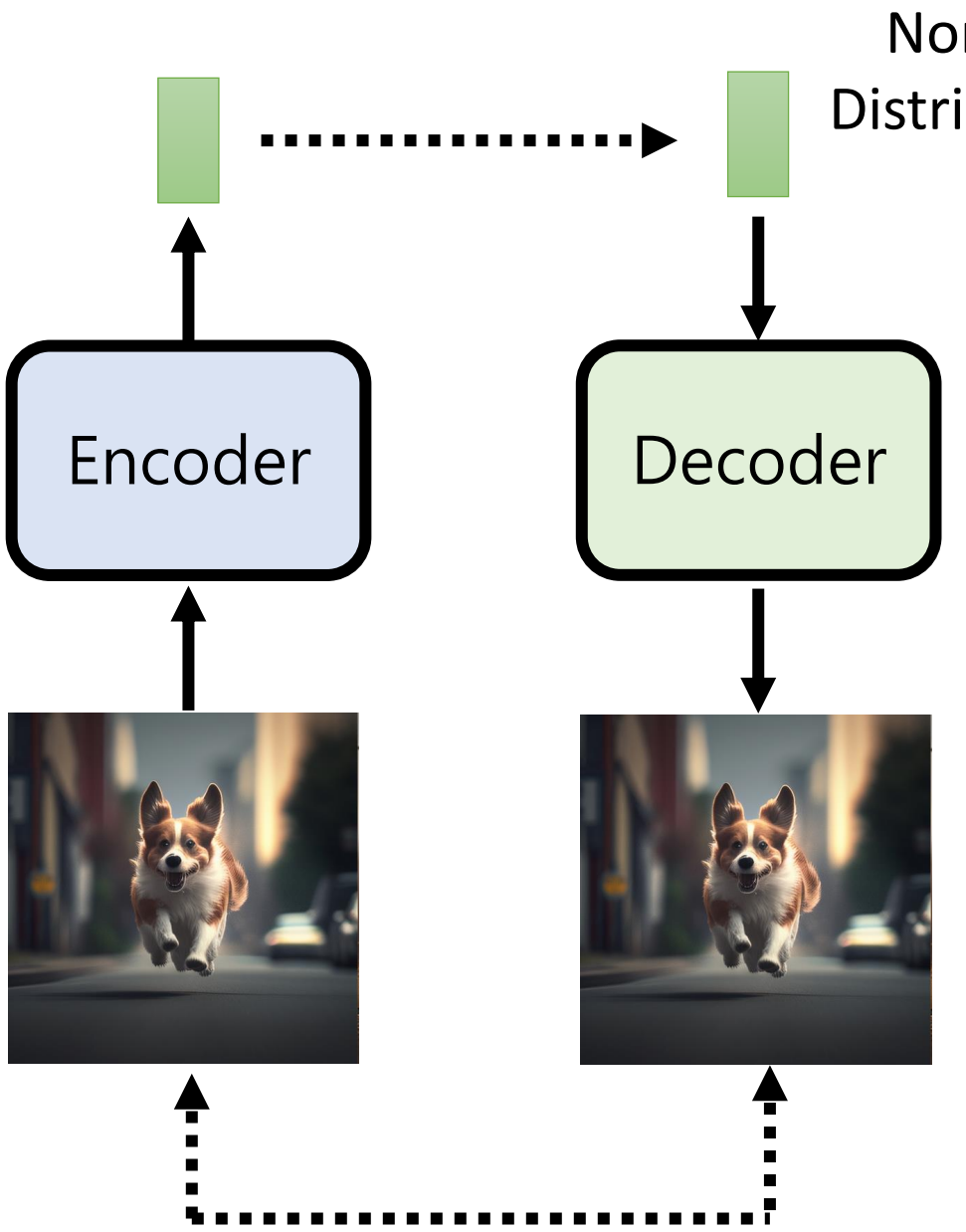
Variational Auto-encoder (VAE)

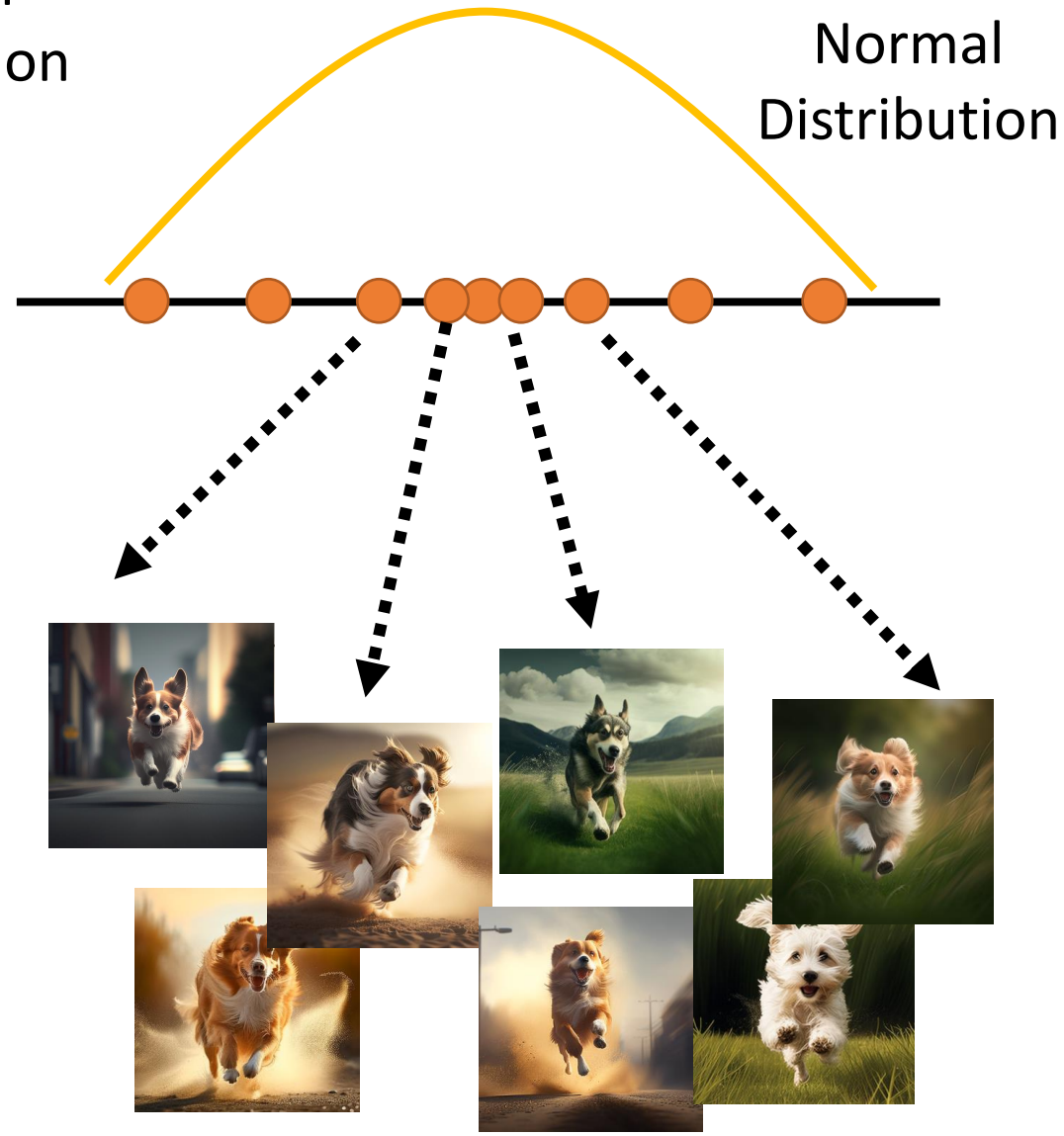
Flow-based Generative Model

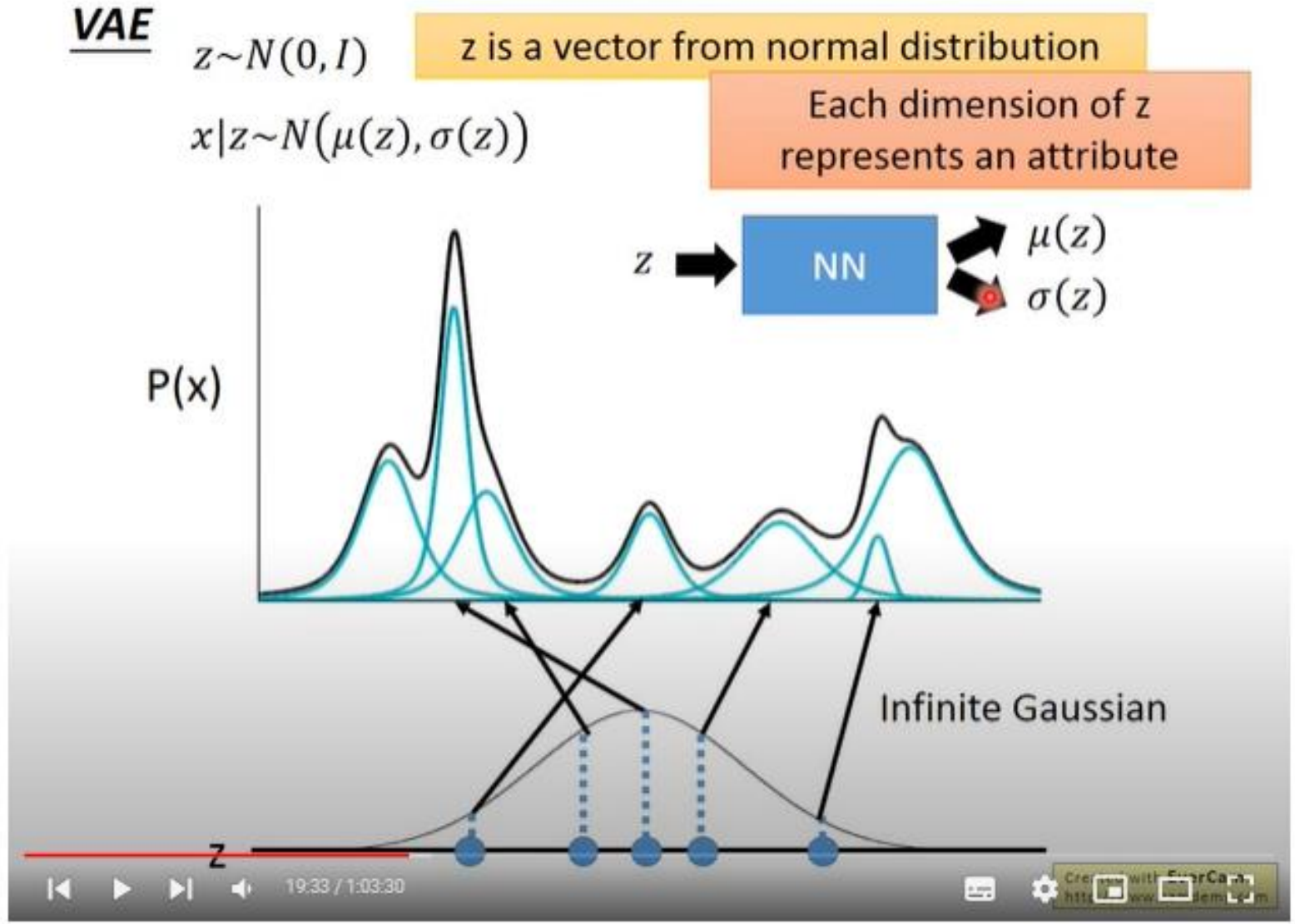
Diffusion Model

Generative Adversarial Network (GAN)

VAE


Normal
Distribution
Encoder
Decoder


Normal
Distribution



ML Lecture 18: Unsupervised Learning - Deep Generative Model (Part II)

https://youtu.be/8zomhgKrsmQ

影像常用生成模型速覽

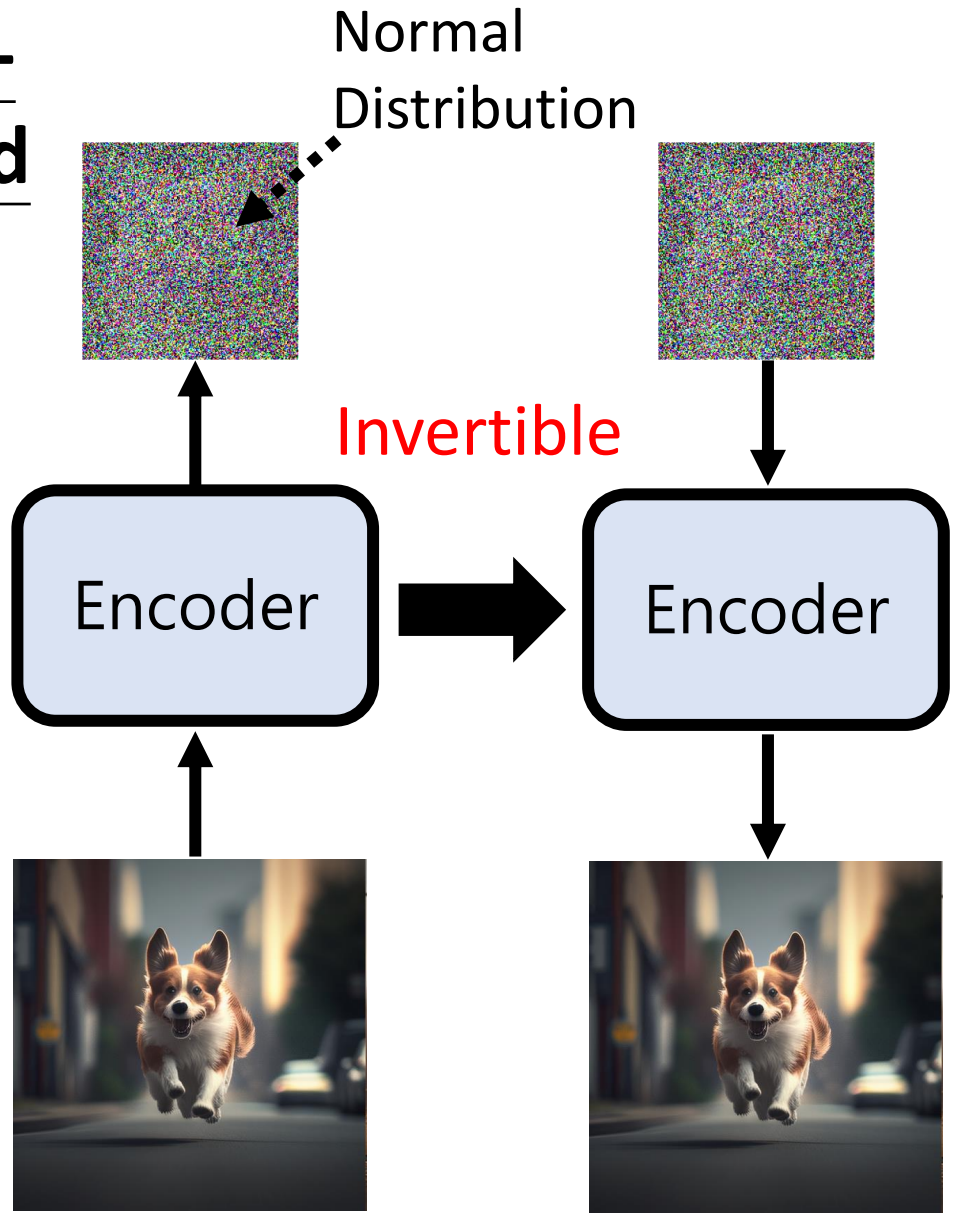
Variational Auto-encoder (VAE)

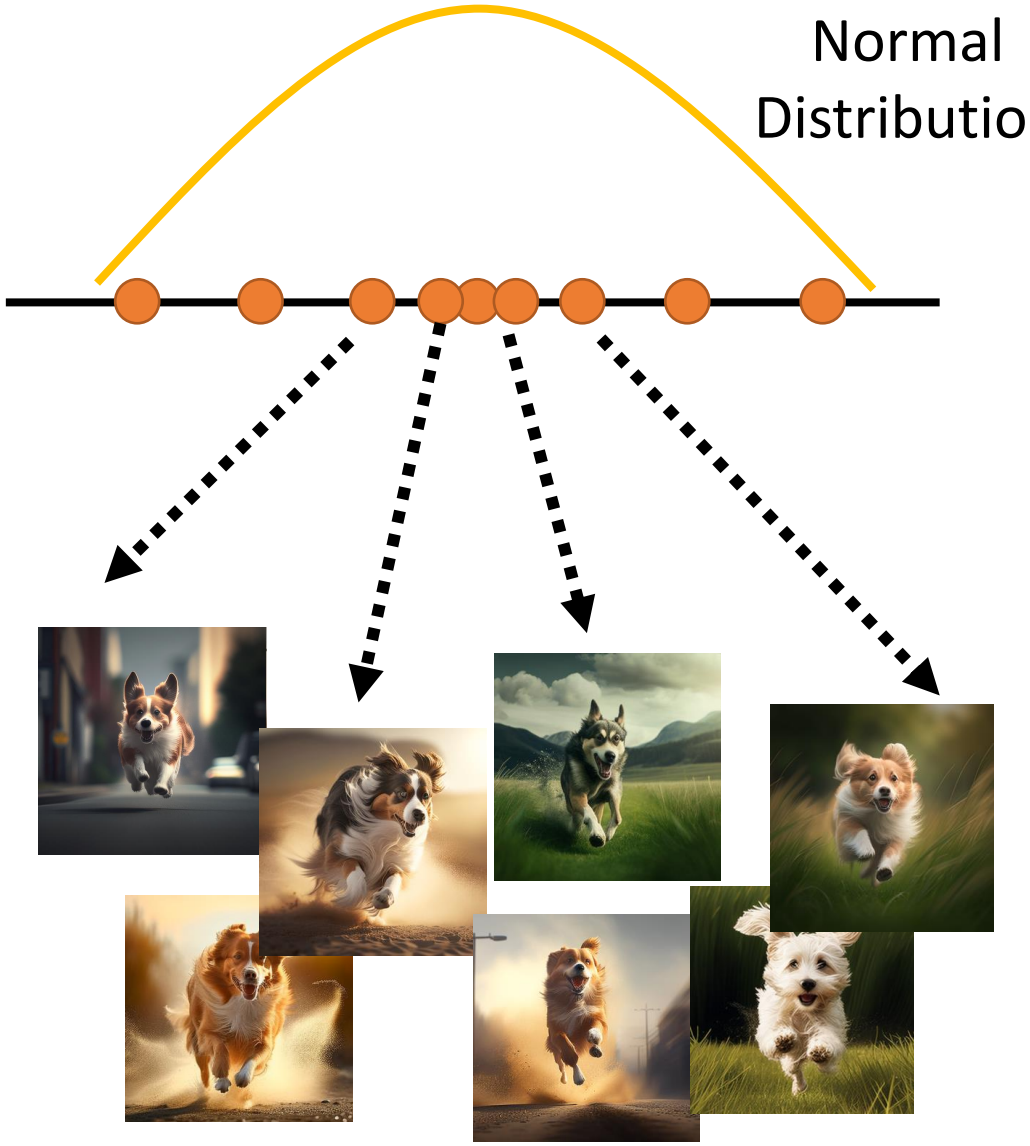
Flow-based Generative Model

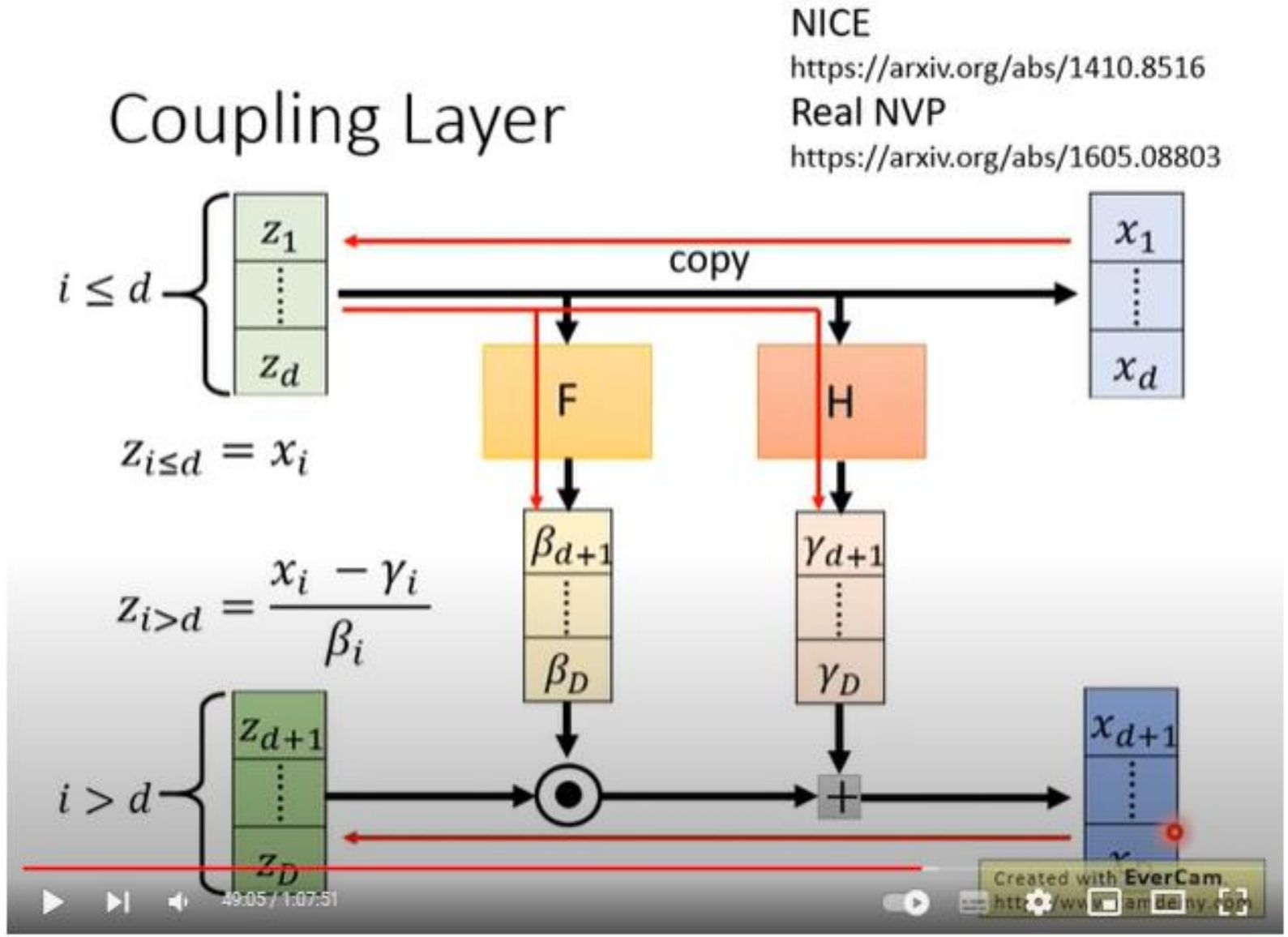
Diffusion Model

Generative Adversarial Network (GAN)

Flow-based


Normal Distribution
Invertible
Encoder
Encoder


Normal Distribution



Flow-based Generative Model

https://youtu.be/uXY18nzdSsM

影像常用生成模型速覽

Variational Auto-encoder (VAE)

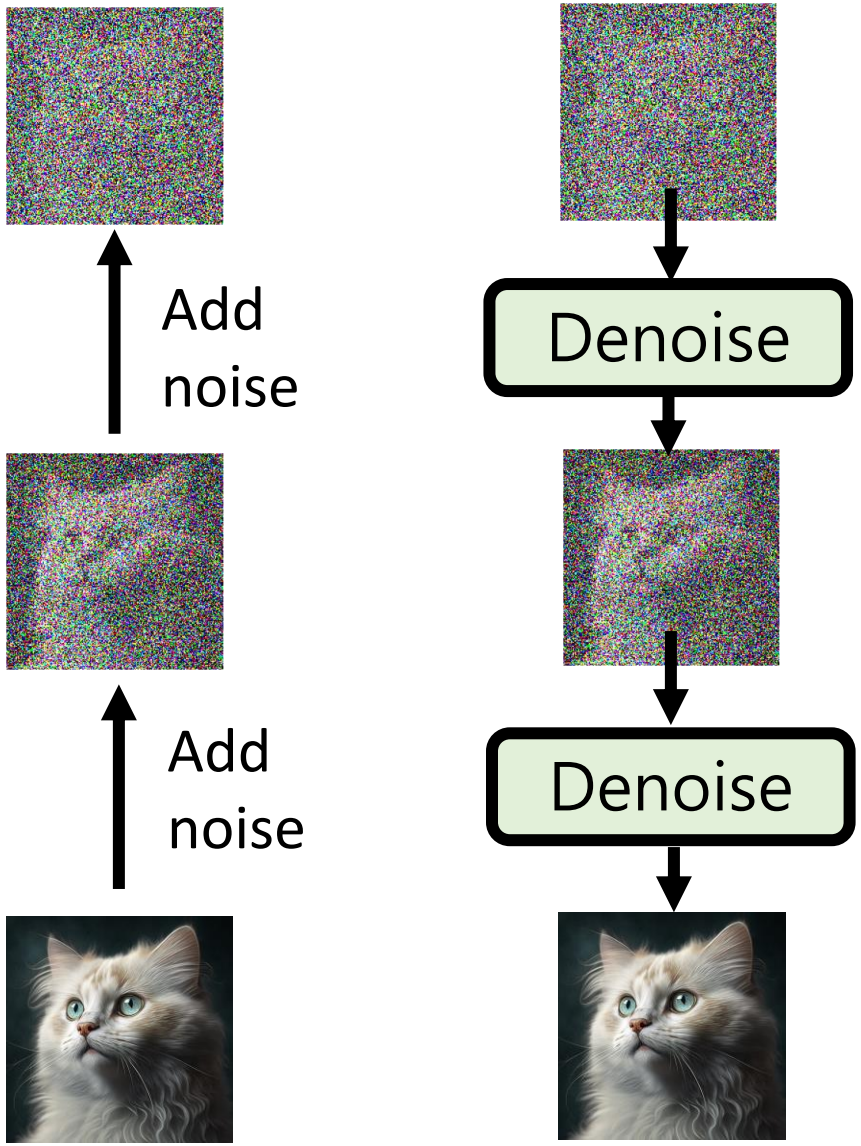
Flow-based Generative Model

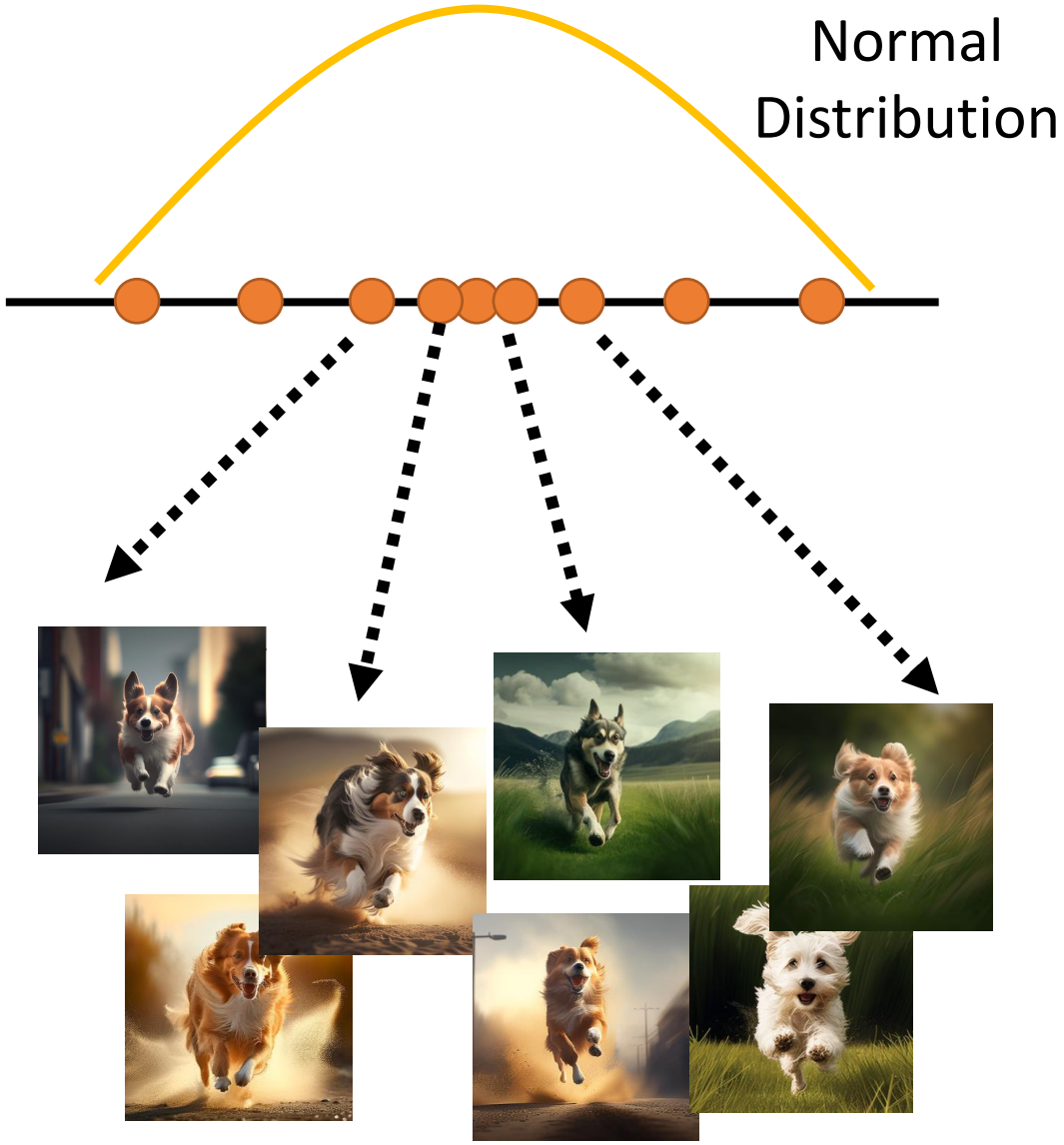
Diffusion Model

Generative Adversarial Network (GAN)

Diffusion

下一段課程會再細講





影像常用生成模型速覽

Variational Auto-encoder (VAE)

Flow-based Generative Model

Diffusion Model

Generative Adversarial Network (GAN)

GAN

Decoder

$P(x)$ and $P'(x)$ are close

Discriminator has poor performance

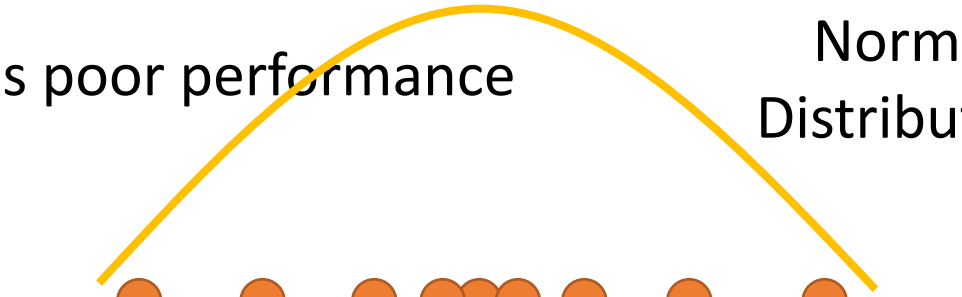
Real / Generated

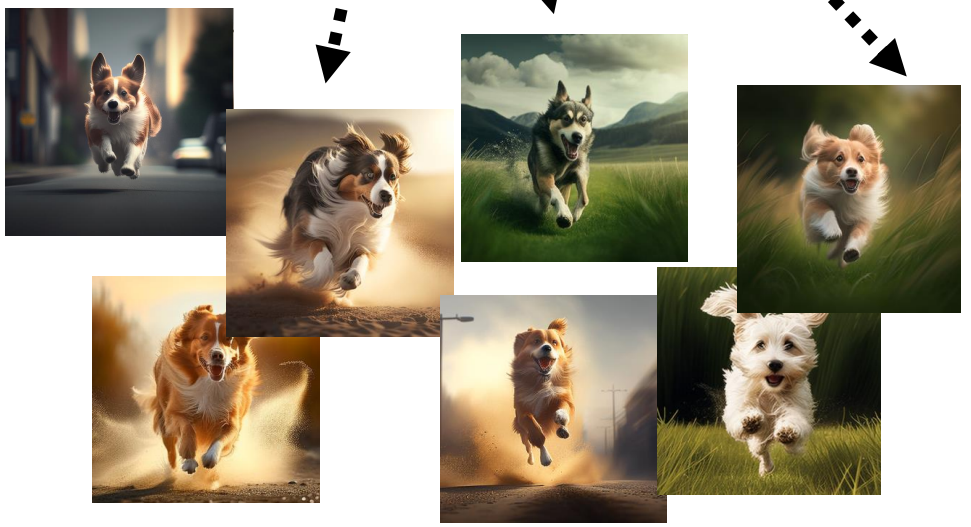
Discriminator

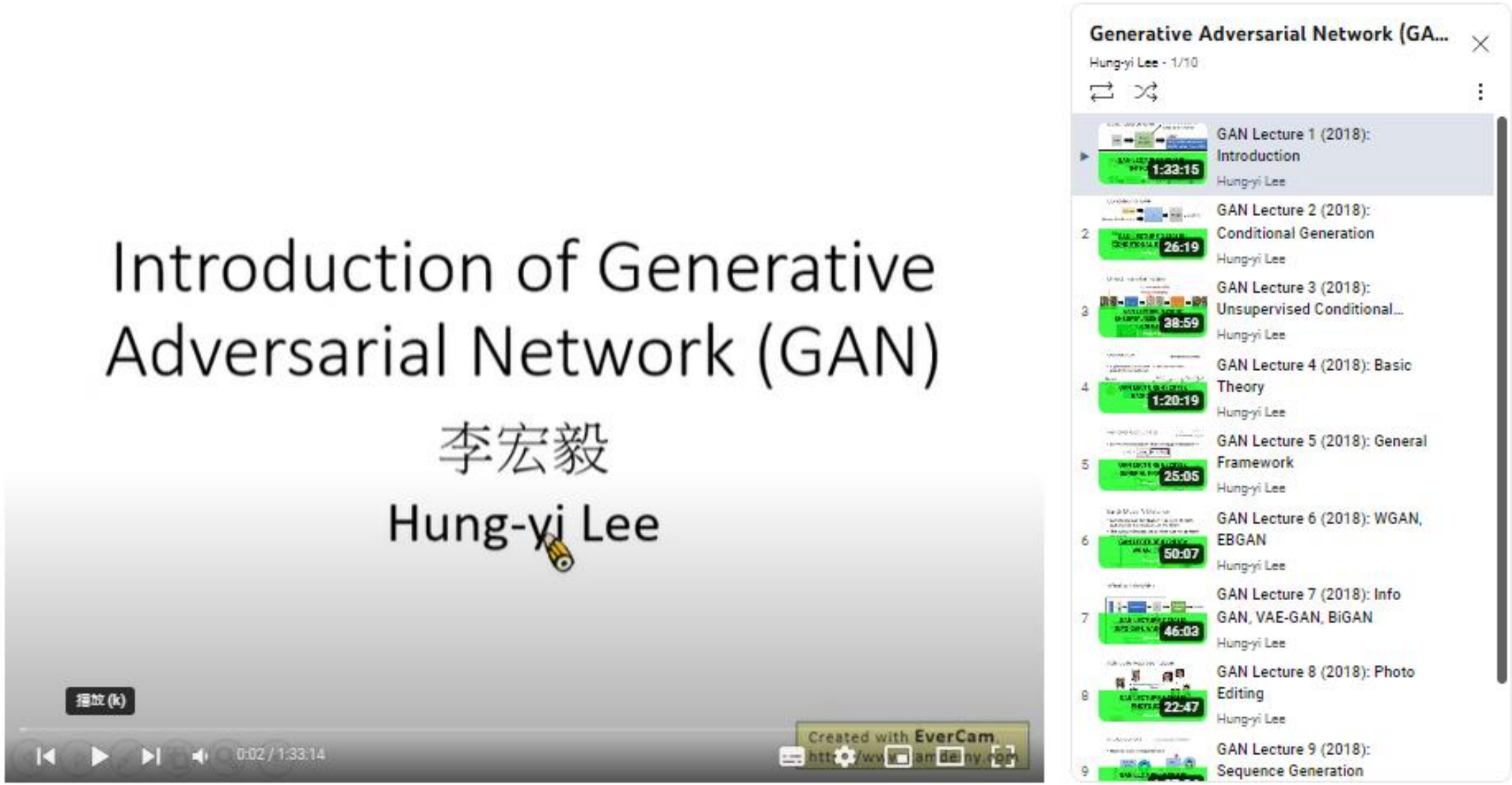
$P'(x)$

$P(x)$

Normal Distribution







https://www.youtube.com/watch?v=DQNNMiAP5lw&list=PLJV_el3uVTsMq6JEFPW35BCiOQTsoqwNw

VAE

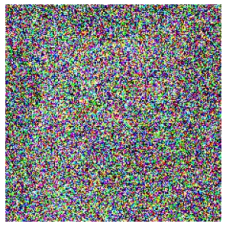
Encoder

Decoder

Flow-based

Encoder

Invertible

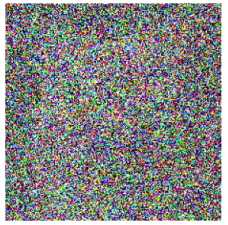


Encoder

Diffusion

Add noise

X N

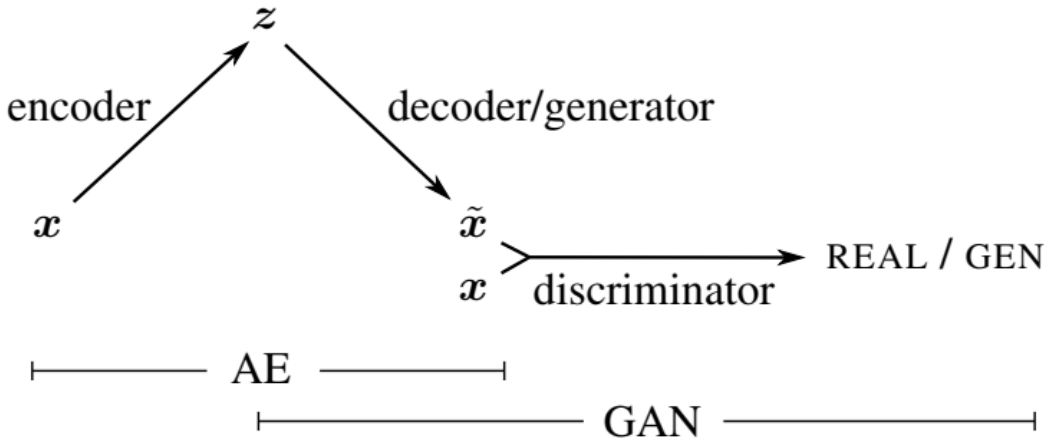


X N

Denoise

VAE + GAN

https://arxiv.org/abs/1512.09300

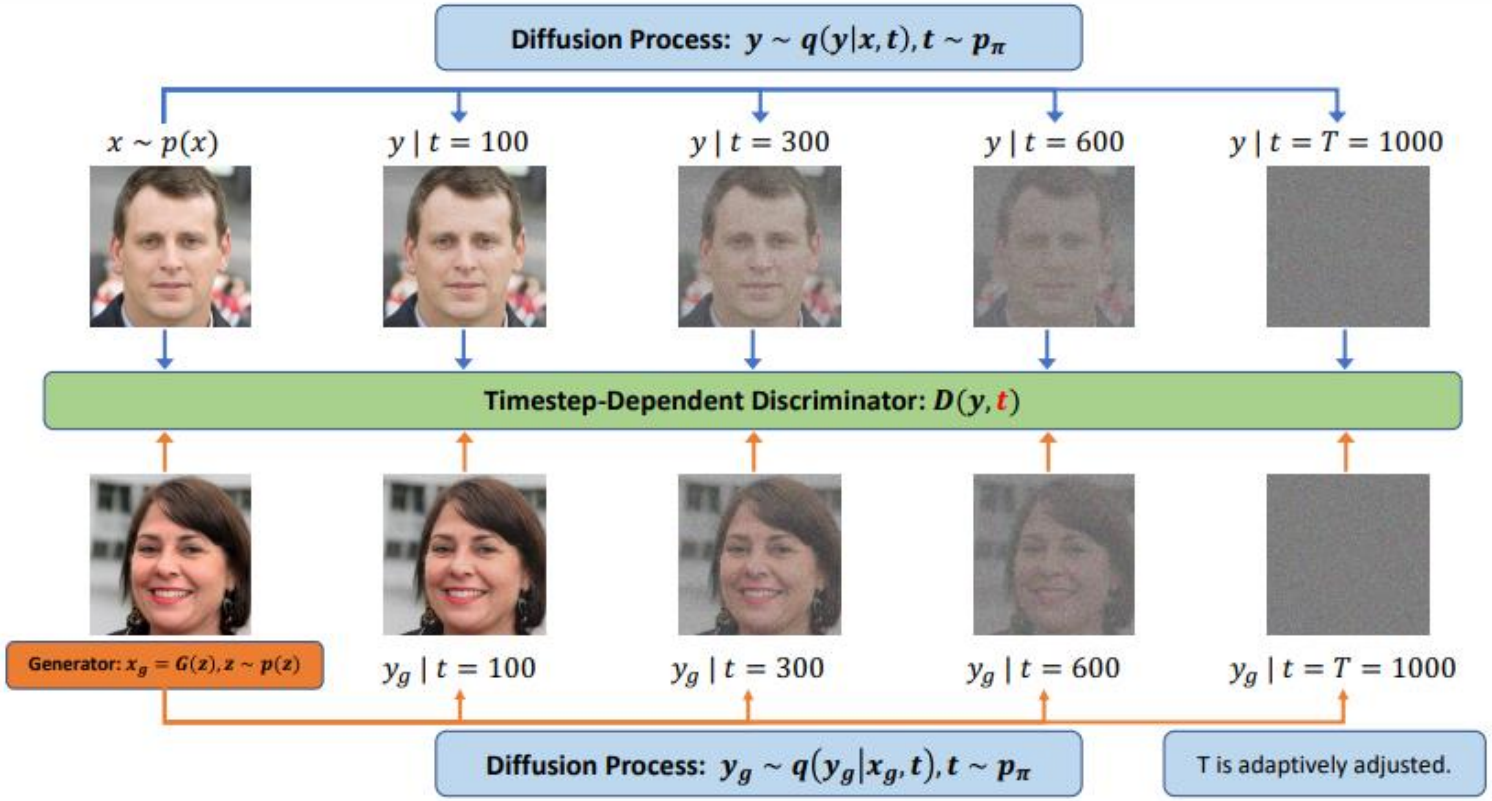


Flow + GAN

https://arxiv.org/abs/1705.08868

Diffusion + GAN

https://arxiv.org/abs/2206.02262



影像常用生成模型速覽

- Variational Auto-encoder (VAE)
- Flow-based Generative Model
- Diffusion Model
- Generative Adversarial Network (GAN)