AI Generate Everything!

Célian Raimbault - Nuit de l'Info - Décembre 2023

A tiny glimpse at generative AIs

Some interesting tasks

- Generate images from scratch / text / image / …
- Generate audio from scratch / text / image / …
- Generate text from scratch / text / image / …
- Generate … from (you see the pattern ?)

# Image generation

Style-GAN

Cycle-GAN

Stable Diffusion
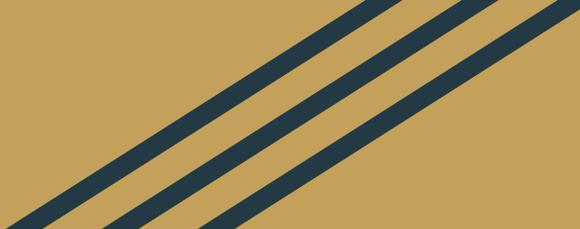
AI 101

ML 101

- A network is a function taking **inputs** and **parameters**
- We update **parameters** using a **loss** function to minimize
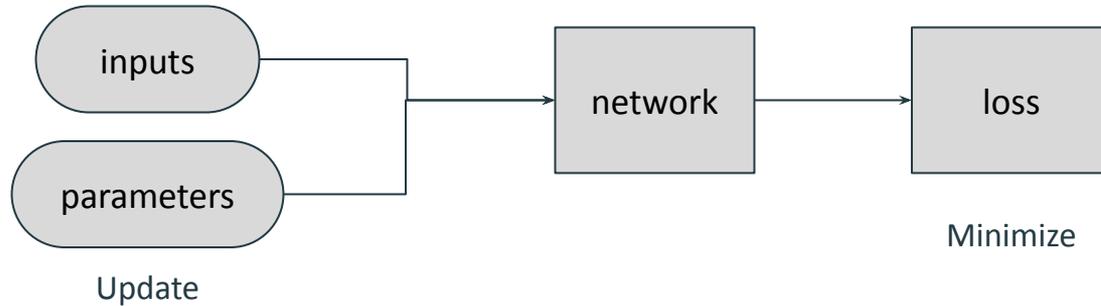
DL 101

- A **layer** is a function (sometimes with **parameters**)
- A network is made of many layers
- **Linear** layer : matrix multiplication of the inputs and the parameters
- **Convolution** layer : apply convolution filters (parameters) to the inputs
- (many other layers)

Layers: the more the better
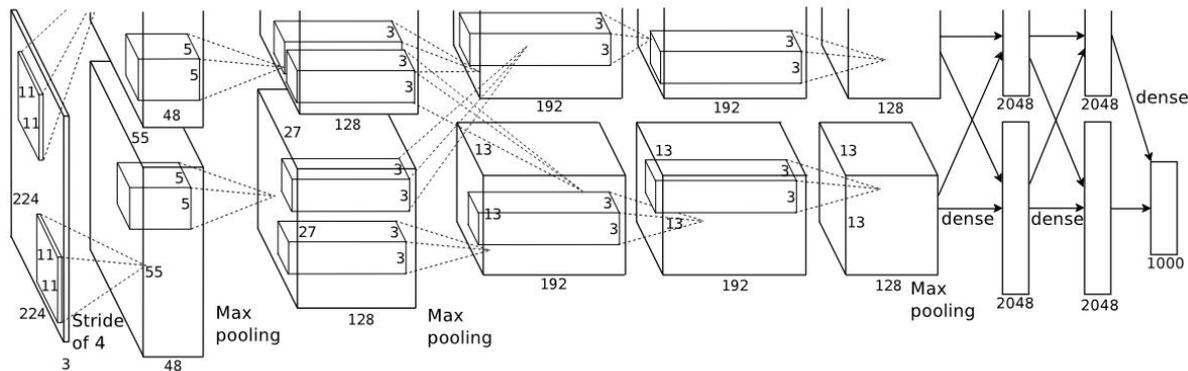

Alex-Net
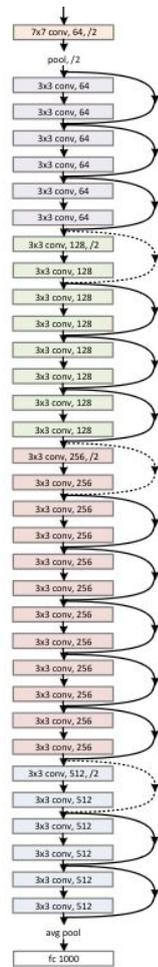
ResNet
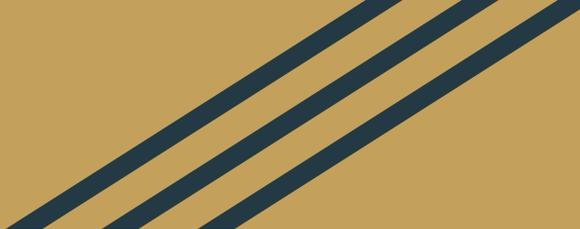
Let's go deep

Latent spaces

- Compress data, retrieve high level information (**semantics**)
- Similar representation for two similar elements
- Same representation for two different element types (image / text) -> image2text, text2speech, …
- "Arithmetic" : (word2vec)

Ground Truth

Blond_Hair

Eyeglasses

Smiling
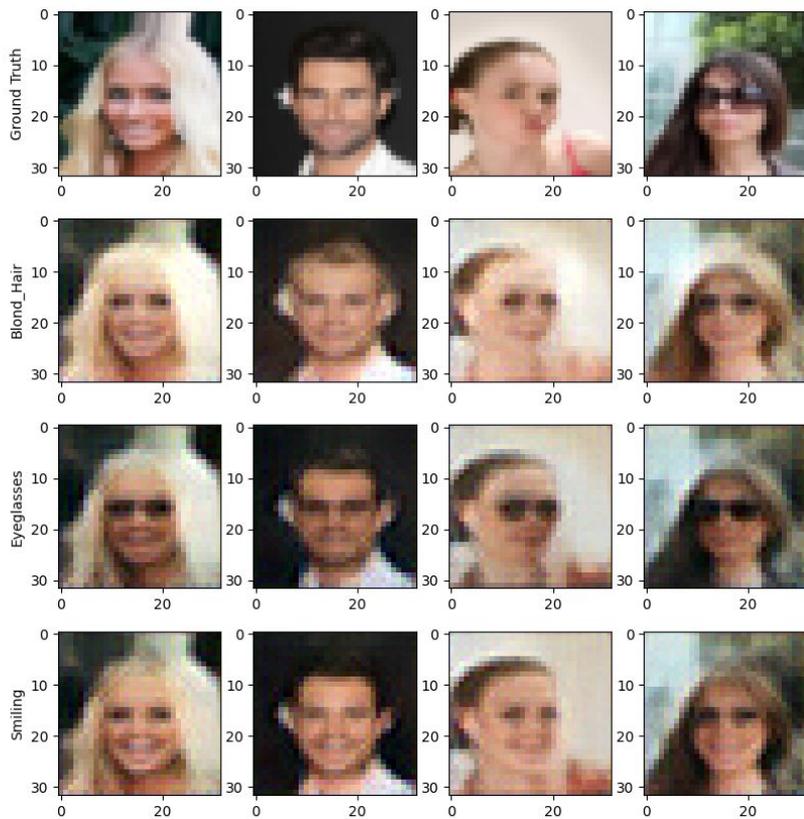
Cc618/Feature-Changer

"High resolution and colorful landscape picture of a lake. Golden hour, colorful clouds, 14mm, hdr…"

text -> vec

Latent

vec -> image

image -> vec

Latent

vec -> text
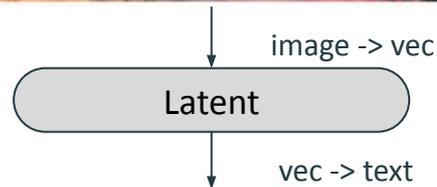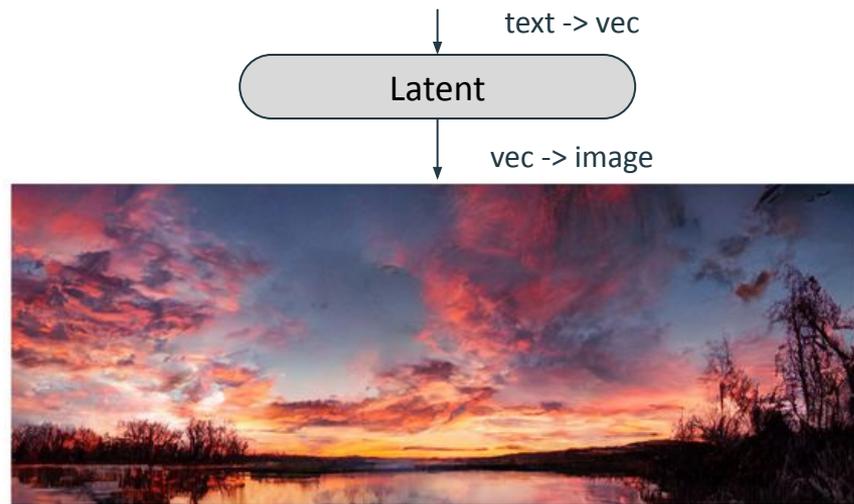
"Colorful picture of a lake took at sunset"

stable-diffusion

Auto-Encoders

- Can be used to obtain latent vectors
- Compress, obtain latent, decompress (reconstruct)

Reconstructed

Decoder

Latent

Encoder

Real
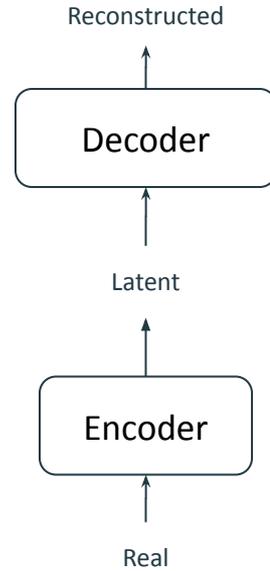
VAE: Variational Auto-Encoders

- Latent space is constrained
- Ability to sample from it (gaussian distribution)

Reconstructed

Decoder

Constrained Latent

Encoder

Real

Image latent +
Eyeglasses latent

Blurry, isn't it ?

VAE
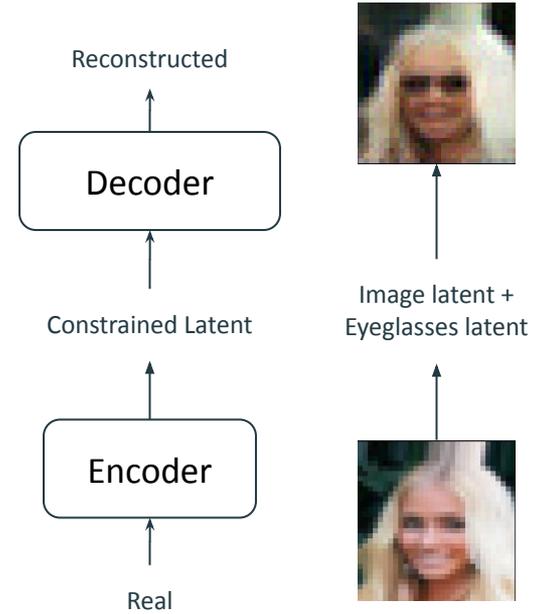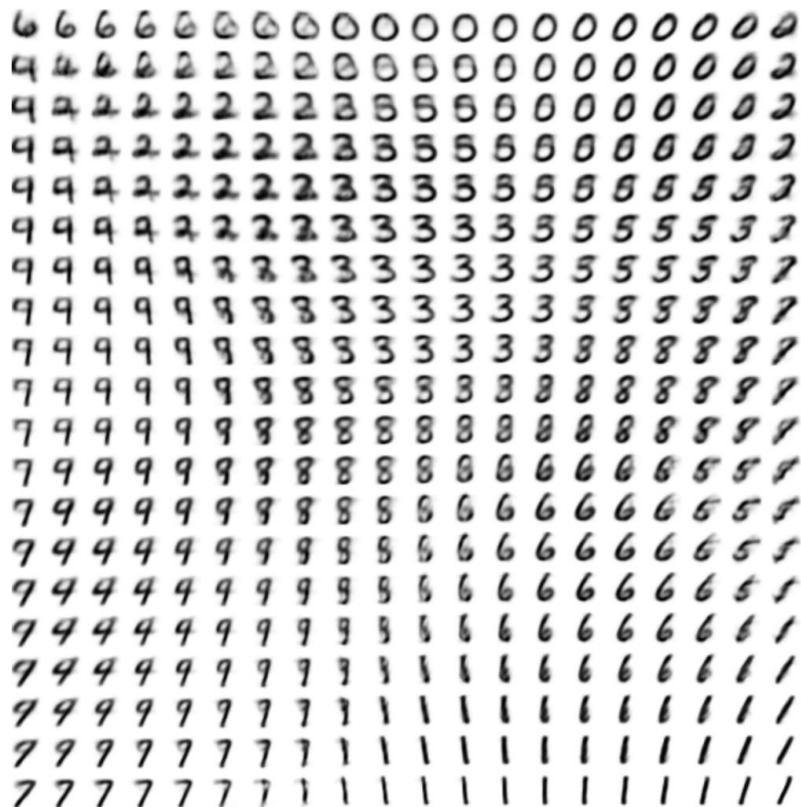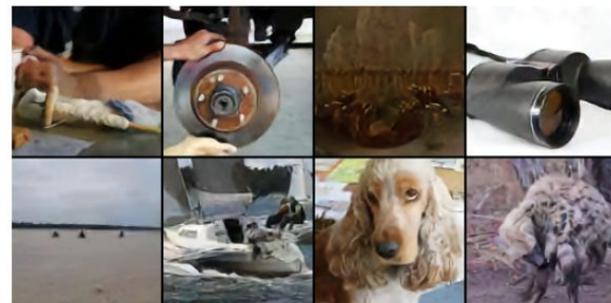
VQ-VAE

GAN: Generative Adversarial Networks

- **Generator**: given random latent vector, **generate element**
- **Discriminator**: guesses whether **an element is generated or not**
1. Generate element
2. Train generator: try to fool the discriminator
3. Train discriminator: try to classify well generated / real element

Real or generated ?

Discriminator

Generated   Real

Generator   Dataset

Random latent

Zebras ⟷ Horses

$$D_X \qquad D_Y$$

$$G$$

$$X \qquad Y$$

$$F$$
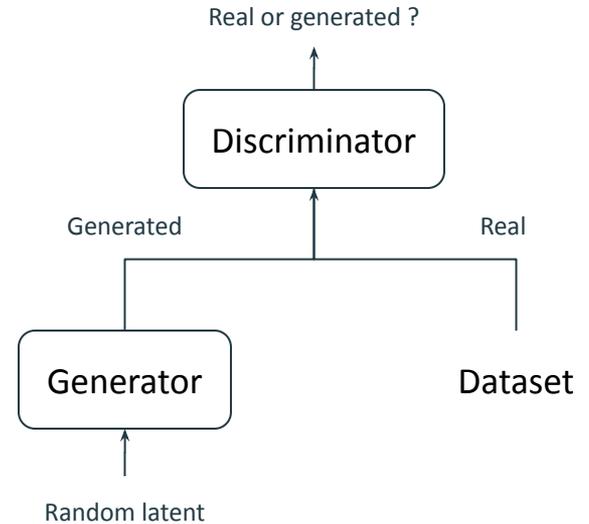
Style-GAN
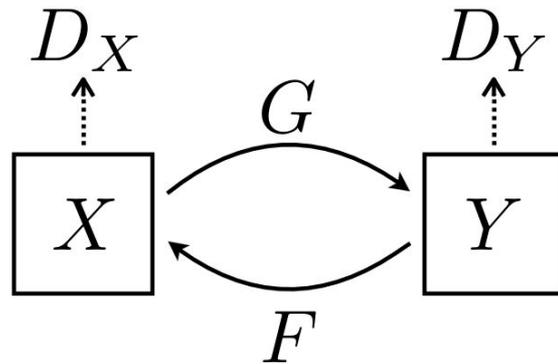
Cycle-GAN

Diffusion Models

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

$$q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

$\mathbf{x}_T$ · · · $\mathbf{x}_t$ $\mathbf{x}_{t-1}$ · · · $\mathbf{x}_0$
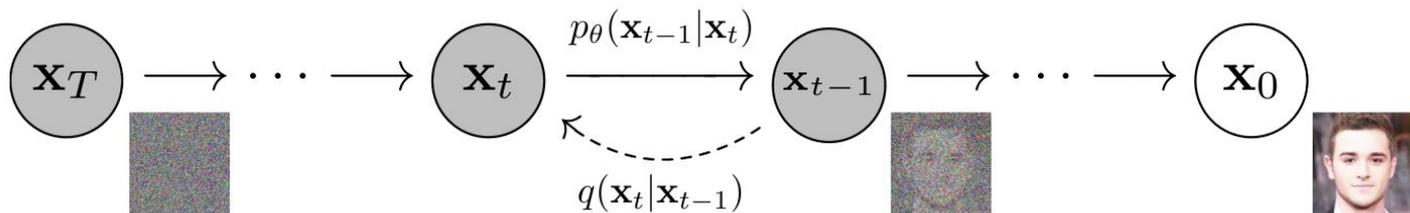
- Start from random noise
- Remove noise step by step

The fun part

Usages

Conditioning

- Generating random elements is not really useful
- Guide generation using conditioning

Examples

- Condition text generation by text (GPT) -> translation
- Condition image generation by image and text (stable-diffusion) -> inpainting
- Condition audio generation by audio and text (music-gen) -> guided audio generation

Image generation

Stable Diffusion

a very cute cat laying by a big bike.

china airlines plain on the ground at an airport with baggage cars nearby.
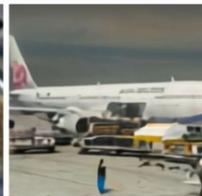
a shiba inu wearing a beret and black turtleneck

Validation

Ours

panda mad scientist mixing sparkling chemicals, artstation

DALL-E 1 (left) / 2 (right)

Audio generation
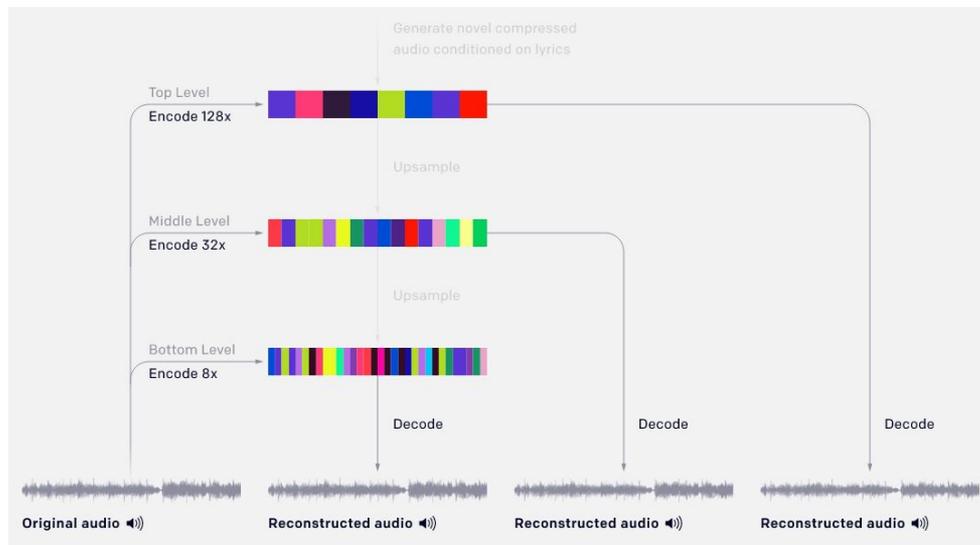
Jukebox

[examples](#)

MusicGen

[demo](#)

https://talks.celian.dev/231201.html
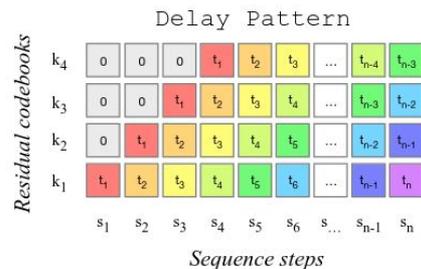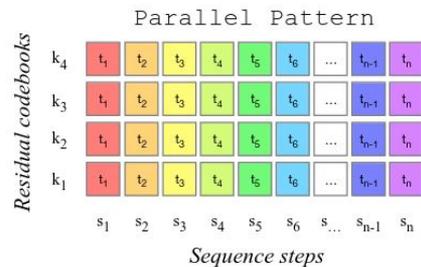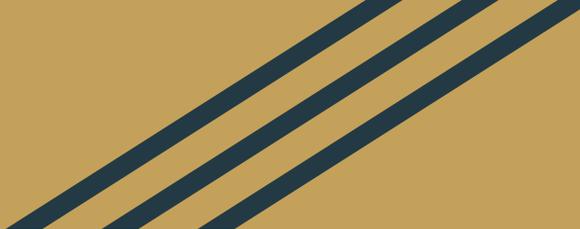
Questions ?

Sources - Examples

- StyleGAN : https://arxiv.org/abs/1812.04948
- Stable Diffusion : https://arxiv.org/abs/2112.10752
- CycleGAN : https://arxiv.org/abs/1703.10593
- Alex-Net : link
- ResNet : https://arxiv.org/abs/1512.03385
- Word2vec : https://arxiv.org/abs/1301.3781
- Feature-Changer : https://github.com/Cc618/Feature-Changer
- VAE : https://arxiv.org/abs/1312.6114
- DDPM (Diffusion) : https://arxiv.org/abs/2006.11239
- DALL-E : (1) https://arxiv.org/abs/2102.12092 (2) https://arxiv.org/abs/2204.06125
- VQ-VAE : https://arxiv.org/abs/1711.00937
- GPT : https://openai.com/research/better-language-models
- Jukebox : https://arxiv.org/abs/2005.00341
- MusicGen : https://arxiv.org/abs/2306.05284