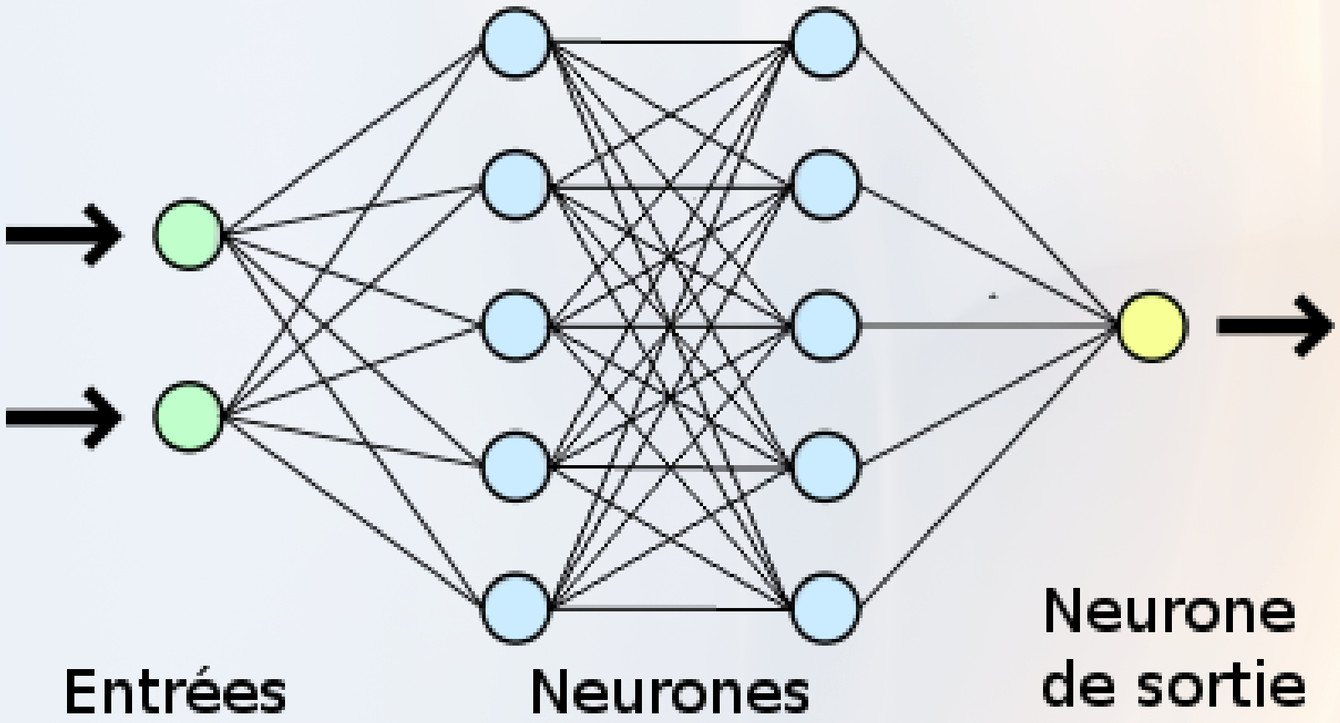
Modèle d'IA générative pour la création d'images simples

Sarah DAHER, Camelia REKIOUA, Charles HERR, Mahir ASSANE-MOUSSABAY

Sommaire

Réseau de neurones



- entrée, poids, biais
- activation
- descente de gradient
- back-propagation

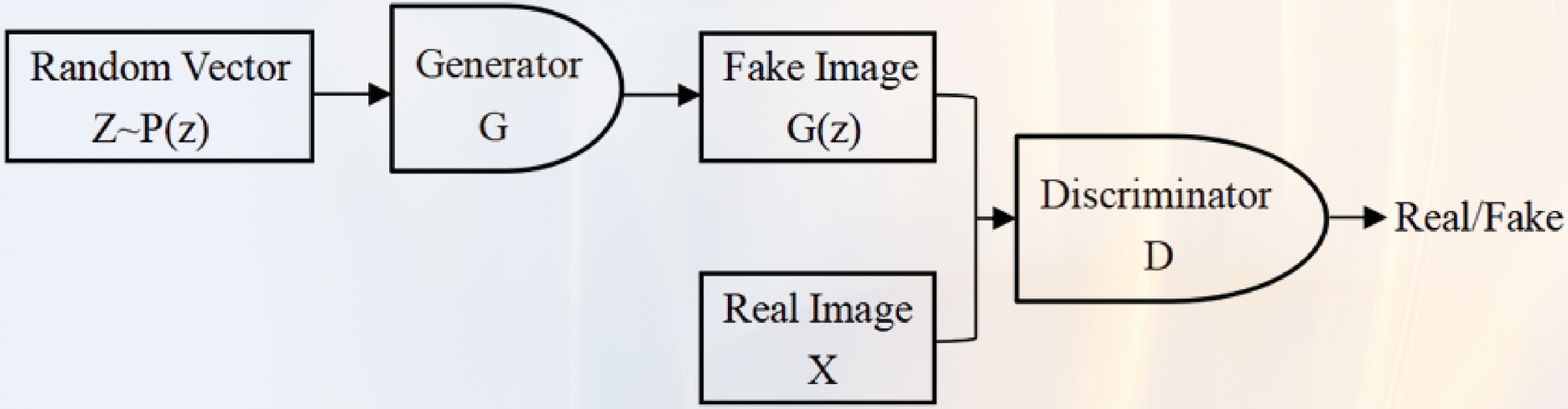
Pytorch

Opérations de base en PyTorch :

- Tenseur (torch.tensor) : structure de données principale (comme un tableau NumPy).
- Opérations classiques :
  - +, *, @ : addition, multiplication, produit matriciel.
  - .mean(), .sum() : moyennes, sommes.
- Calcul automatique de gradients (autograd) : Permet de dériver facilement des fonctions pour l'entraînement.

```
x = torch.tensor([2.0], requires_grad=True)
y = x**2 + 3*x + 1
y.backward()
print(x.grad)  # Affiche le gradient de y par rapport à x
```

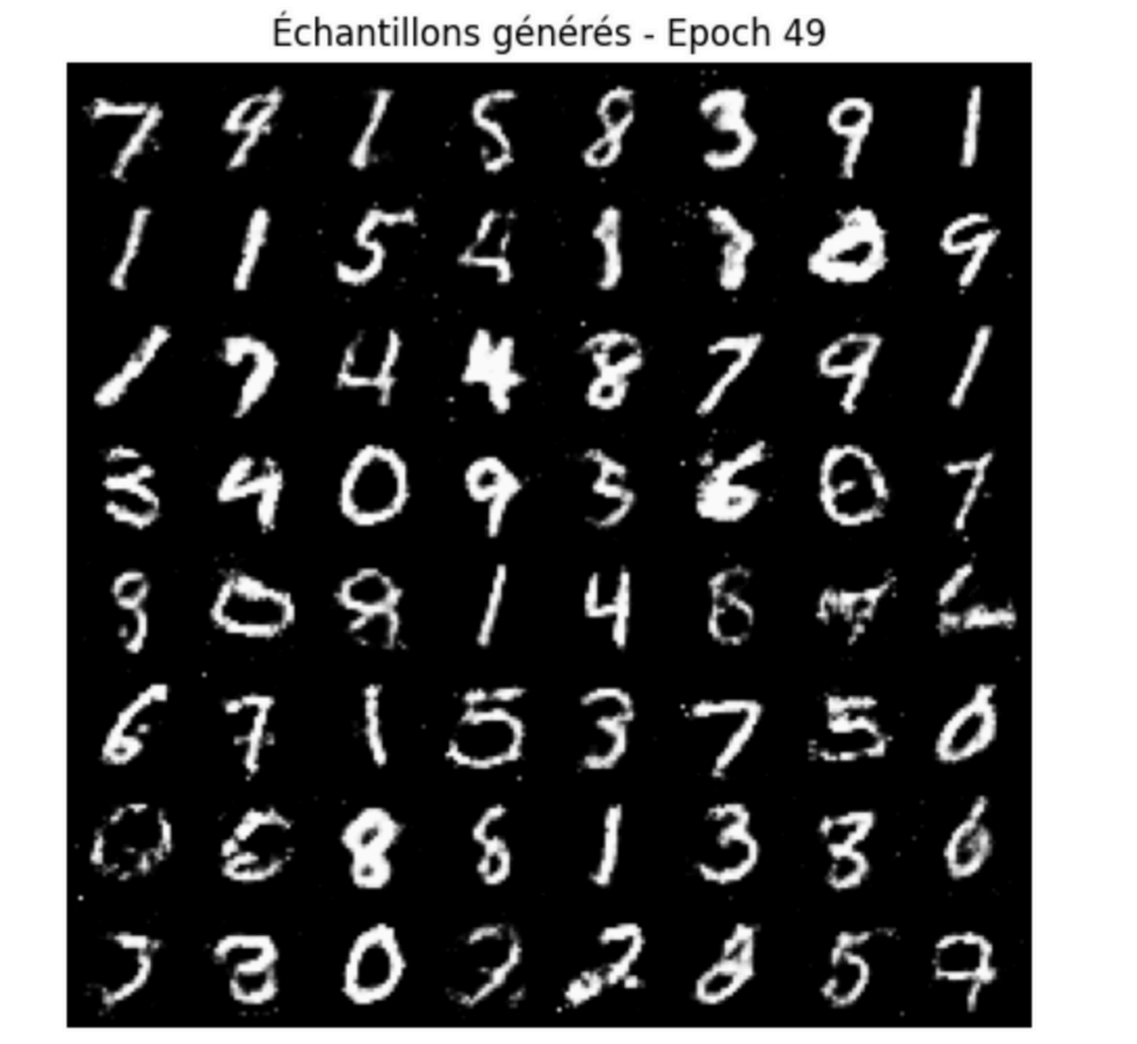
GANs



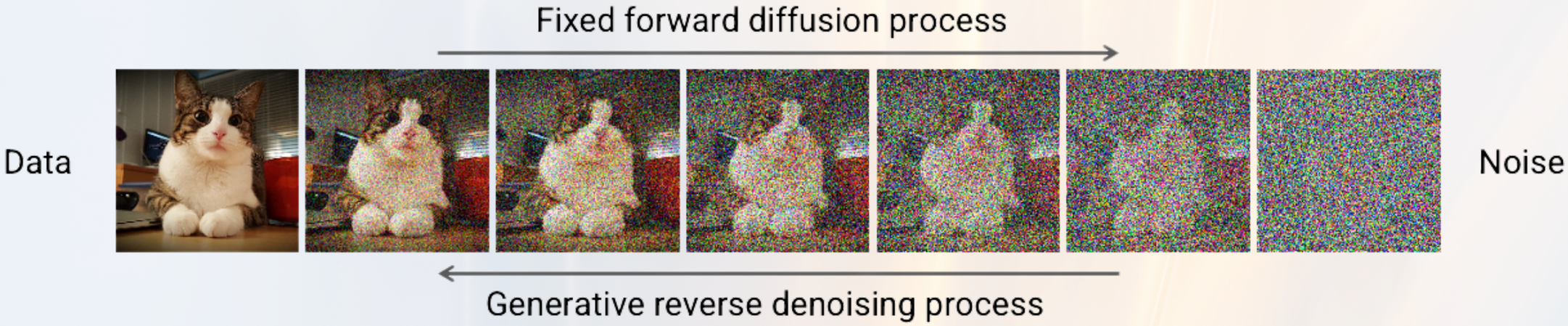
$$\mathcal{L}_G(\theta) = \mathbb{E}_{z \sim p_z} \left[\log D_{\theta'}(G_\theta(z)) \right]$$

$$\mathcal{L}_D(\theta') = \mathbb{E}_{x \sim p_{\text{data}}} \left[\log D_{\theta'}(x) \right] + \mathbb{E}_{z \sim p_z} \left[\log(1 - D_{\theta'}(G_\theta(z))) \right]$$

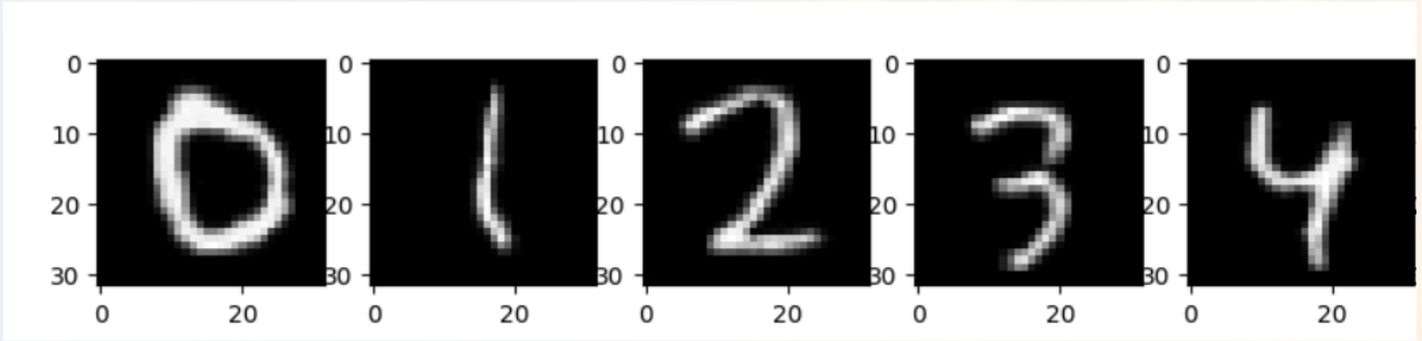
Résultats sur MNIST

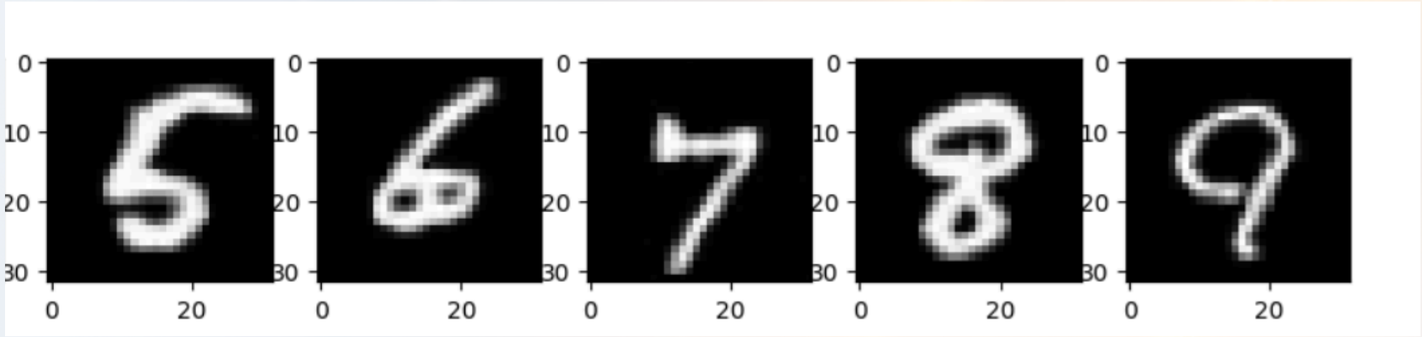


Modèle de diffusion

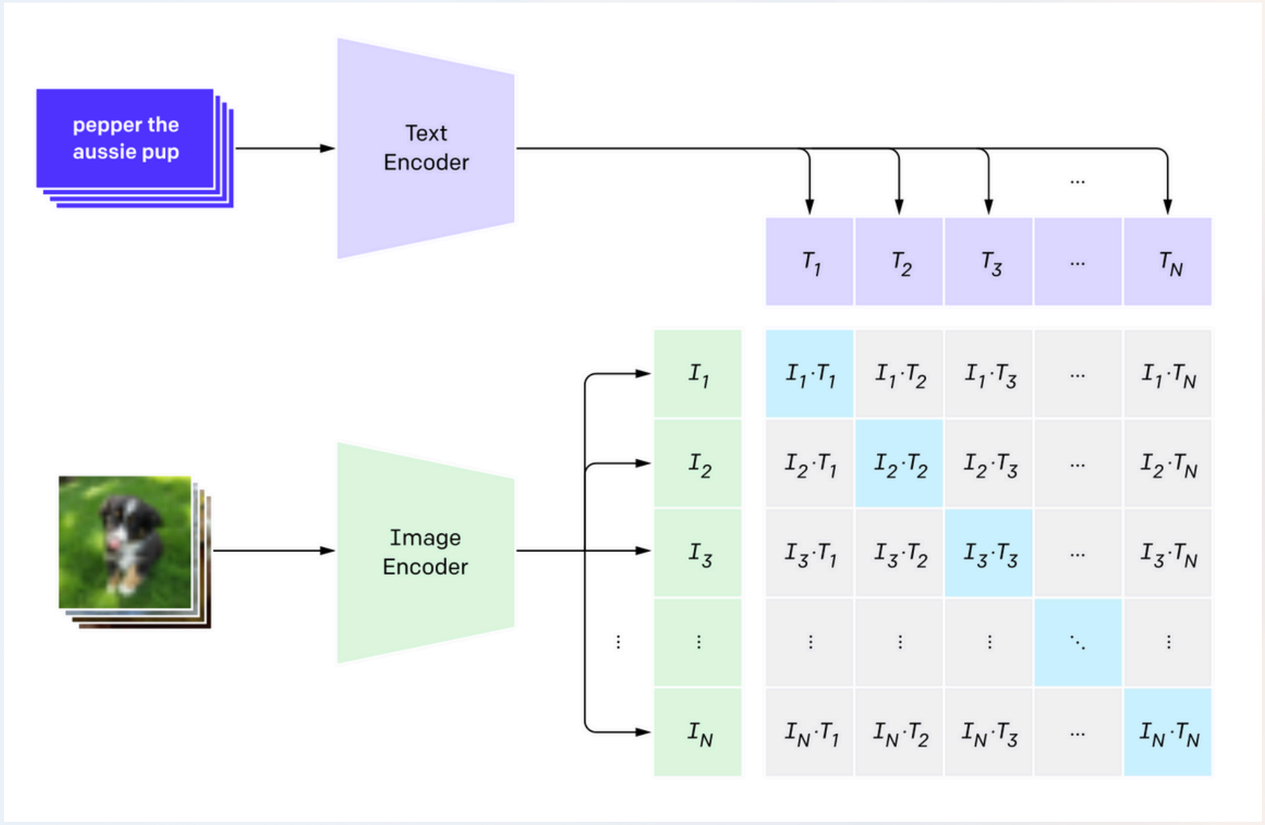


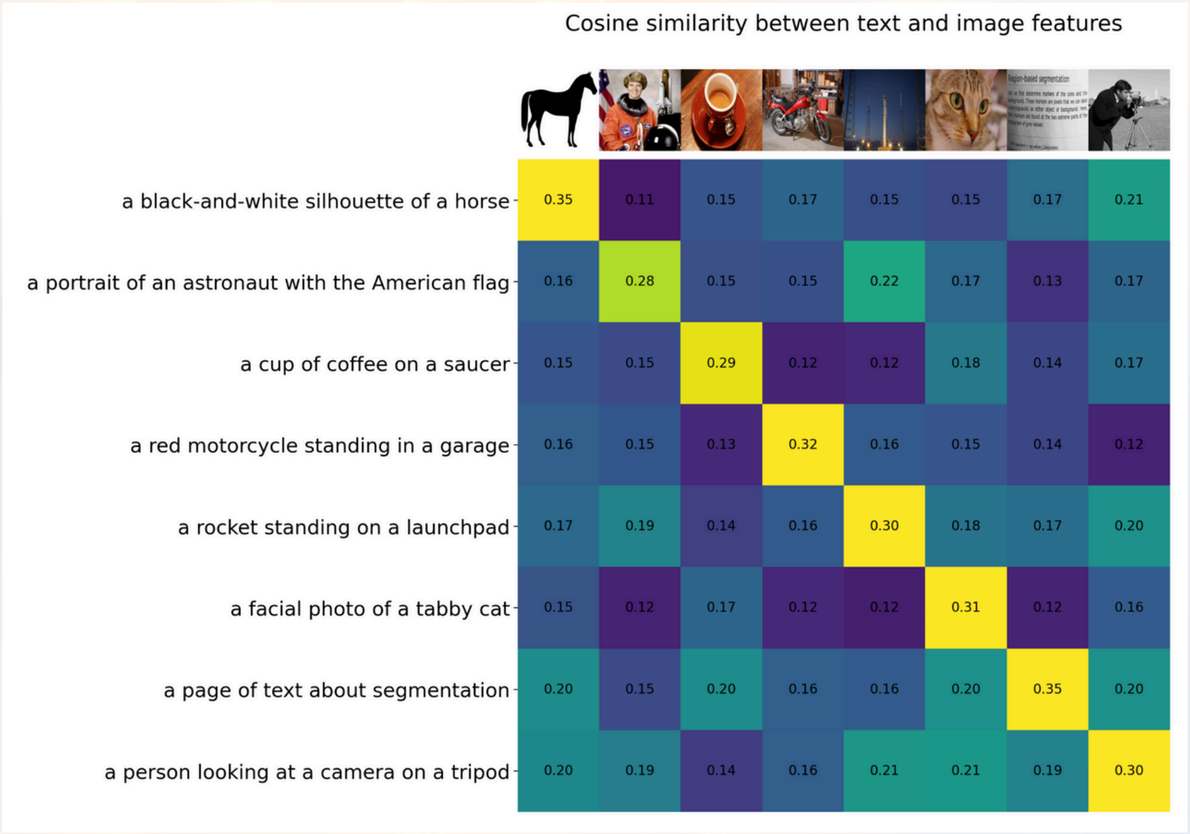
Résultats sur MNIST



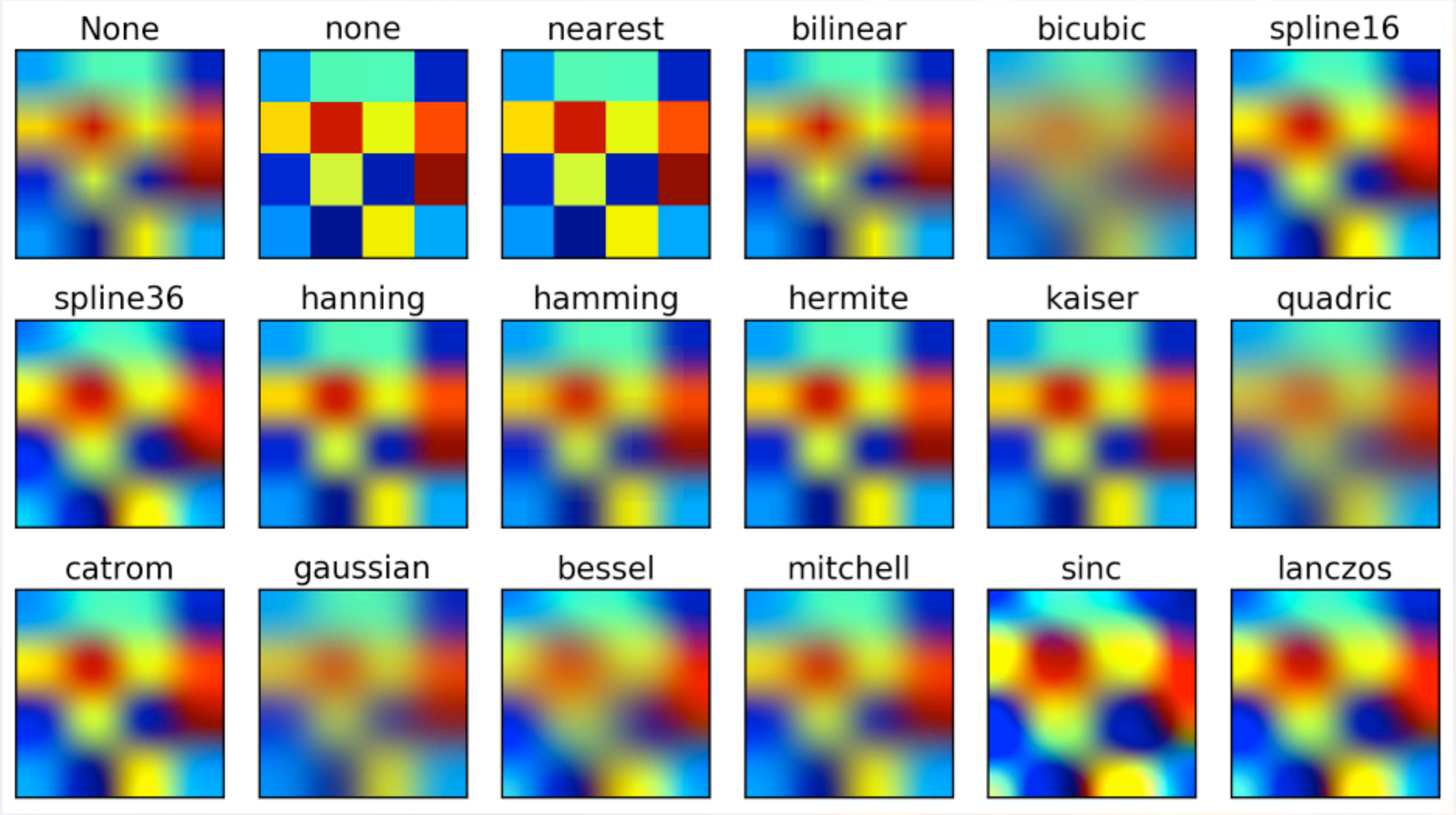


CLIP


pepper the aussie pup
Text Encoder
Image Encoder
T_1 T_2 T_3 ... T_N
I_1 I_2 I_3 ⋮ I_N
$I_1 \cdot T_1$ $I_1 \cdot T_2$ $I_1 \cdot T_3$ ... $I_1 \cdot T_N$
$I_2 \cdot T_1$ $I_2 \cdot T_2$ $I_2 \cdot T_3$ ... $I_2 \cdot T_N$
$I_3 \cdot T_1$ $I_3 \cdot T_2$ $I_3 \cdot T_3$ ... $I_3 \cdot T_N$
$I_N \cdot T_1$ $I_N \cdot T_2$ $I_N \cdot T_3$ ... $I_N \cdot T_N$

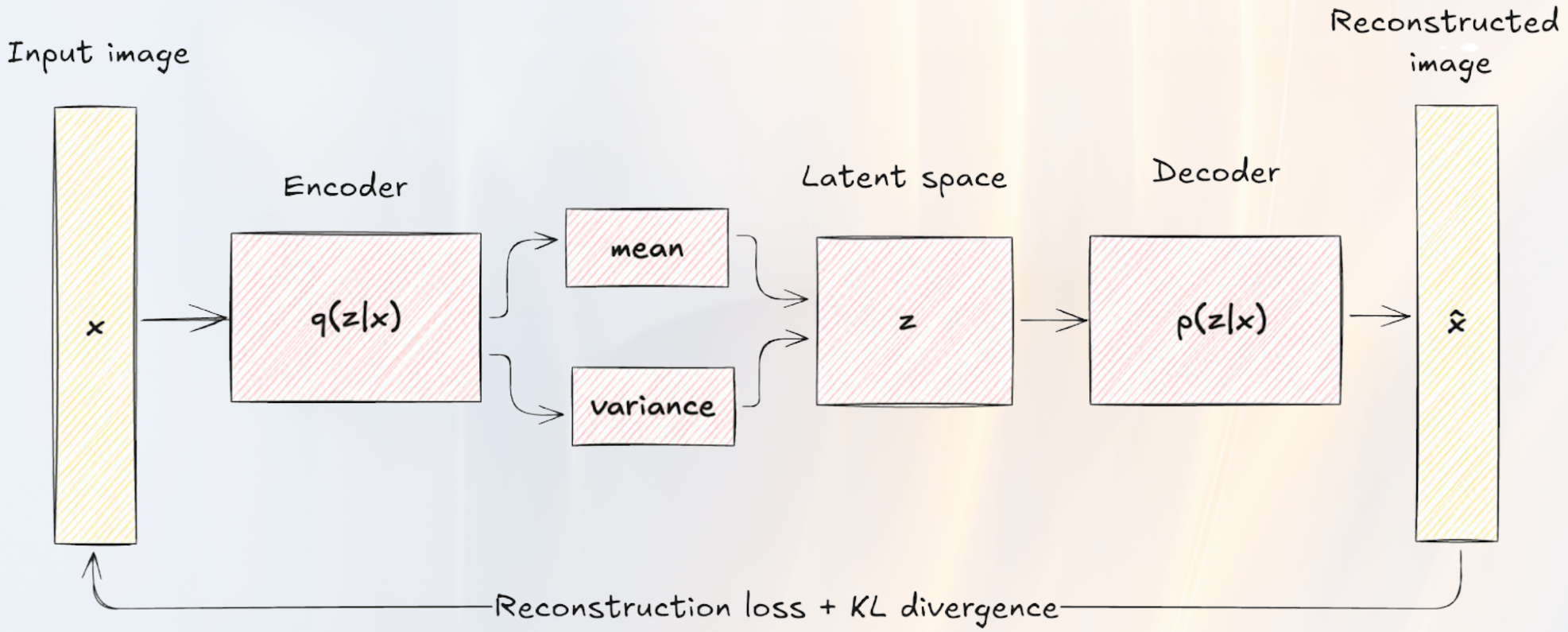

Cosine similarity between text and image features
a black-and-white silhouette of a horse 0.35 0.11 0.15 0.17 0.15 0.15 0.17 0.21
a portrait of an astronaut with the American flag 0.16 0.28 0.15 0.15 0.22 0.17 0.13 0.17
a cup of coffee on a saucer 0.15 0.15 0.29 0.12 0.12 0.18 0.14 0.17
a red motorcycle standing in a garage 0.16 0.15 0.13 0.32 0.16 0.15 0.14 0.12
a rocket standing on a launchpad 0.17 0.19 0.14 0.16 0.30 0.18 0.17 0.20
a facial photo of a tabby cat 0.15 0.12 0.17 0.12 0.12 0.31 0.12 0.16
a page of text about segmentation 0.20 0.15 0.20 0.16 0.16 0.20 0.35 0.20
a person looking at a camera on a tripod 0.20 0.19 0.14 0.16 0.21 0.21 0.19 0.30

CLIP

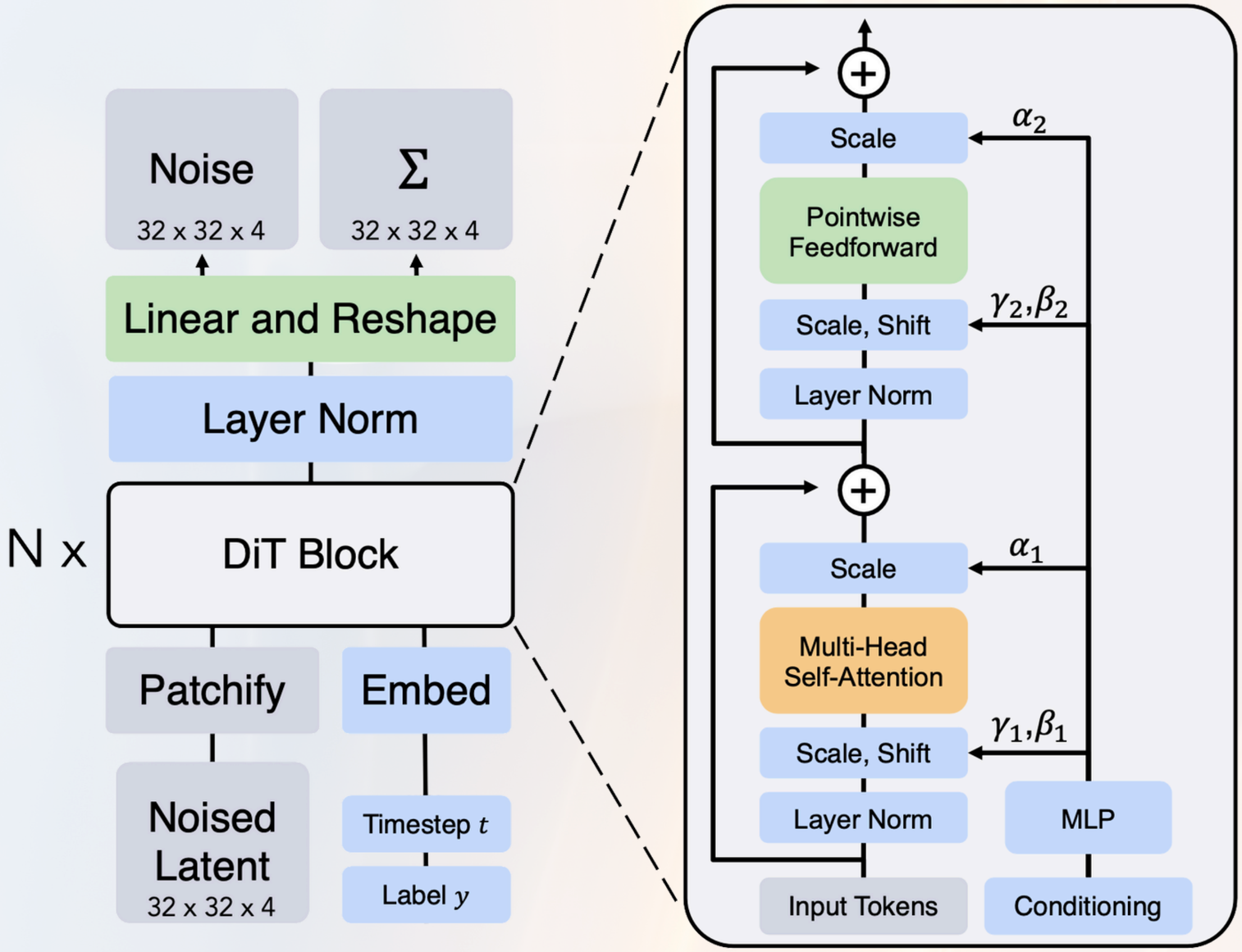


Pre-processing strategies

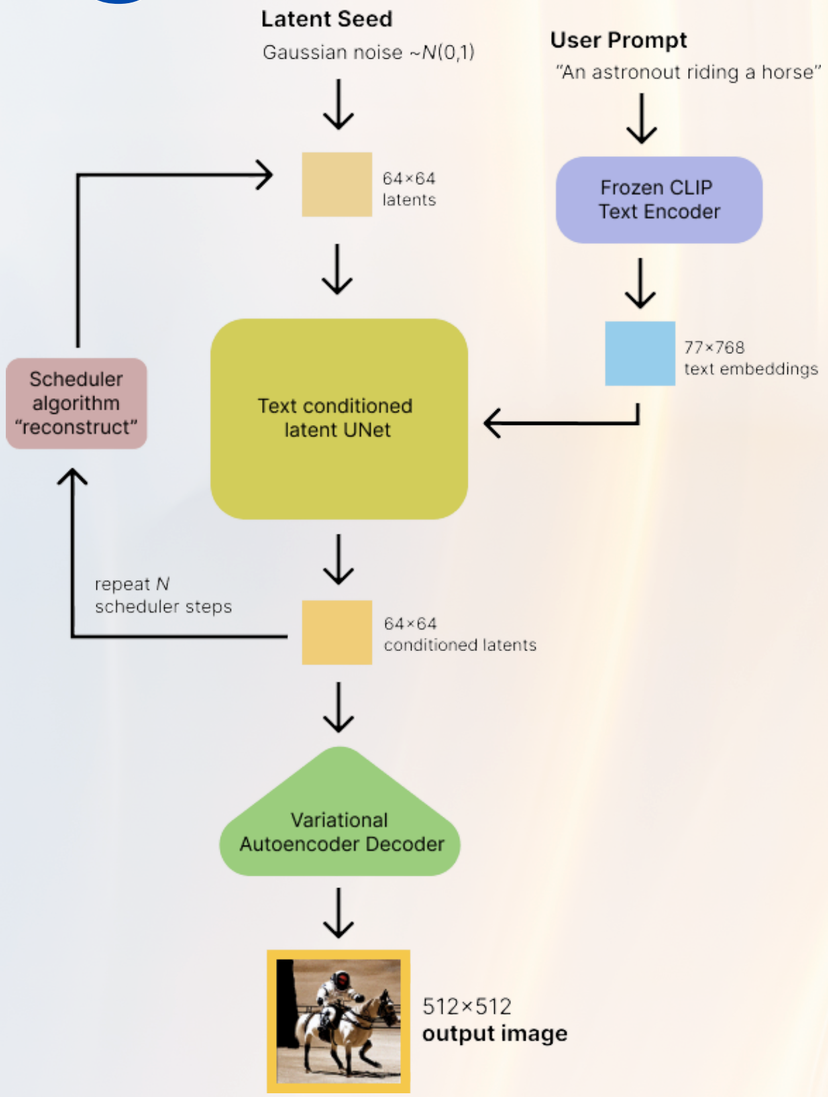
VAEs


Input image
Encoder
Latent space
Decoder
Reconstructed image
x
q(z|x)
mean
variance
z
p(z|x)
x̂
Reconstruction loss + KL divergence

Transformers



Architecture globale



Merci pour votre attention