

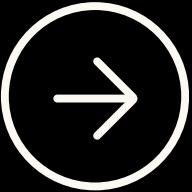


Télécom SudParis

DDIM DENOISING DIFFUSION IMPLICIT MODELS

Mayeul Ricard

SOMMAIRE



1 PAPIER

2 DATA

3 MATHS

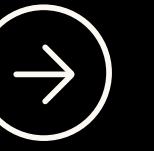
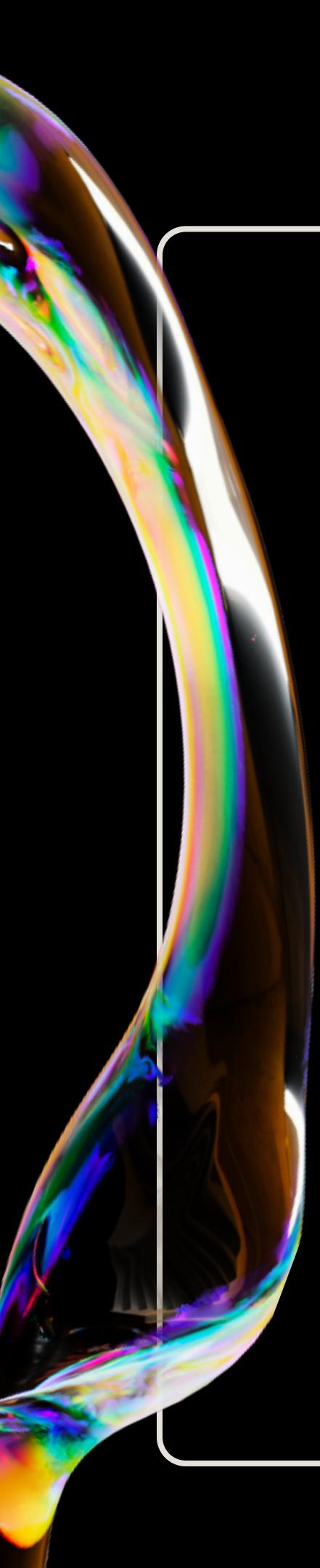
4 MODEL U-Net

4 LOSS FUNCTION

5 ENTRAINEMENT

6 RESULTATS

6 CONCLUSION



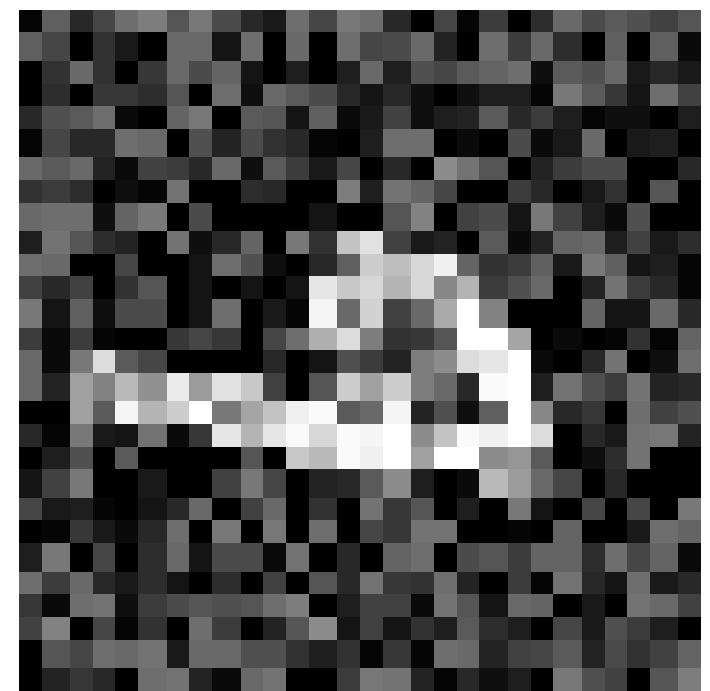
**DDIM QU'EST CE QUE
C'EST ?**

AMELIORATION DES DDPM

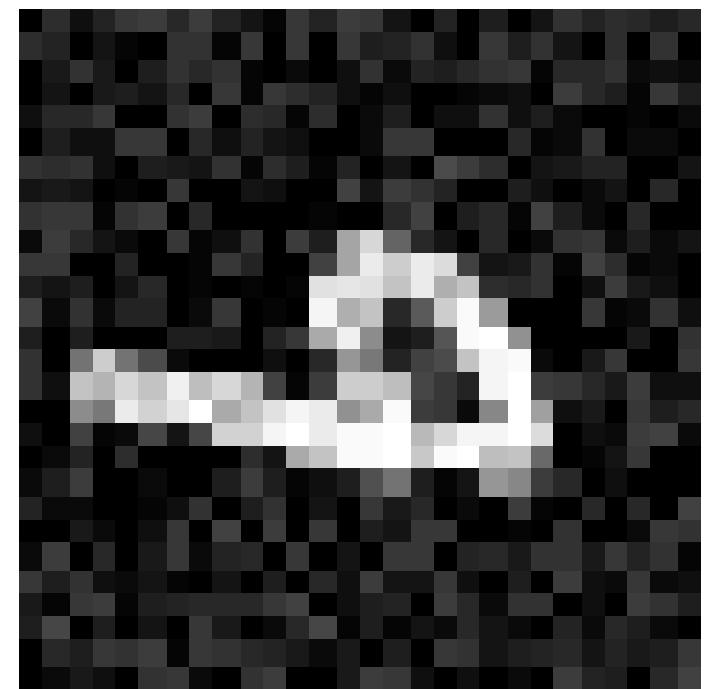
VITESSE D'INFERENCE

CHAINE NON MARKOVIENNE

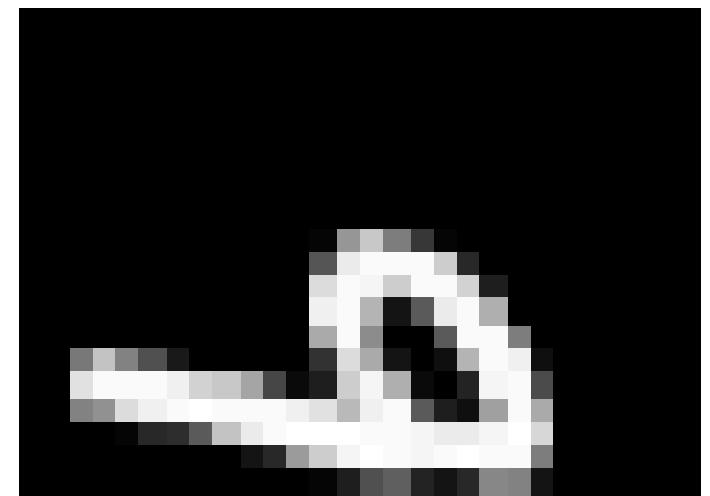
DETERMINISTE



t=224



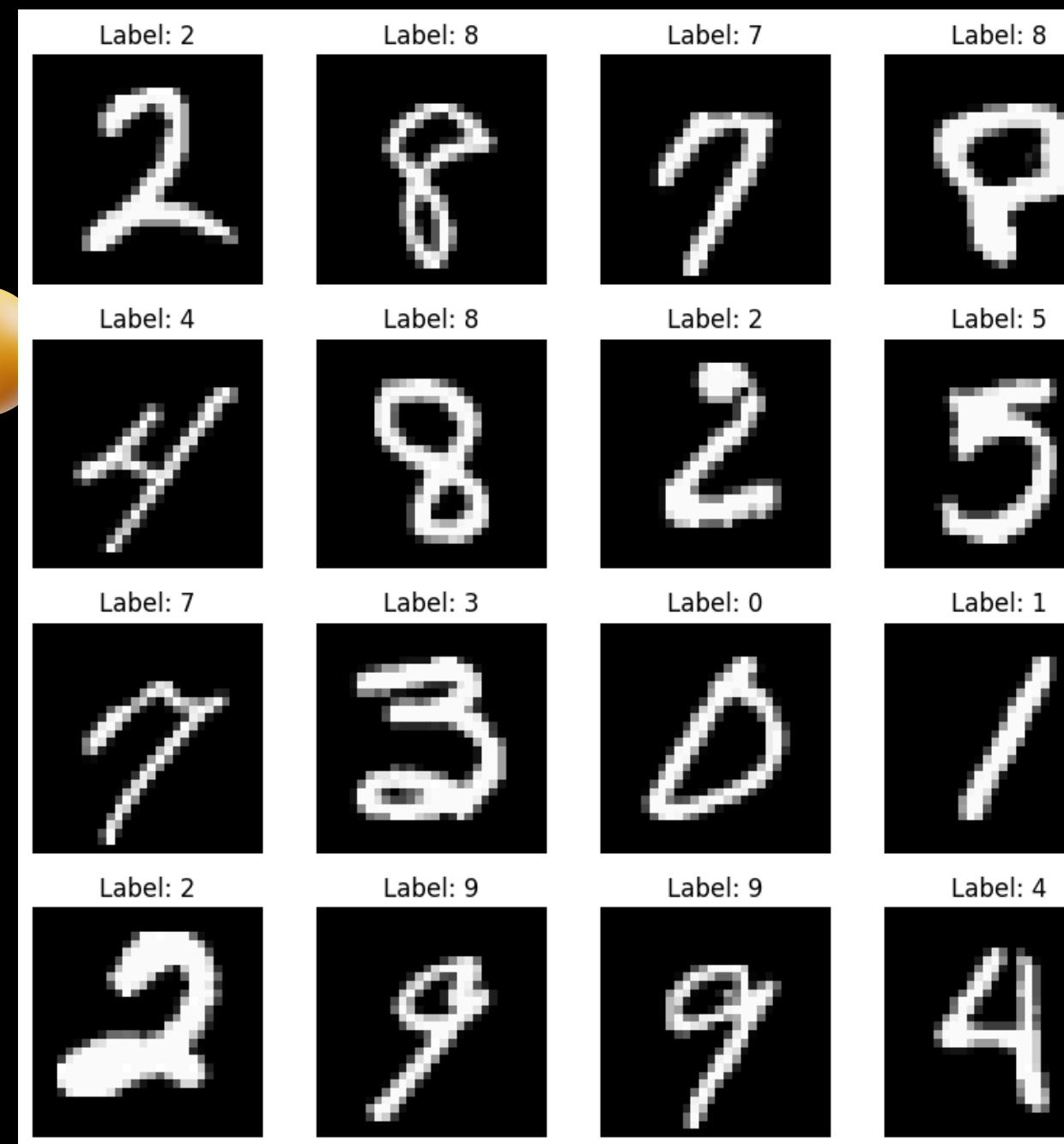
t=122



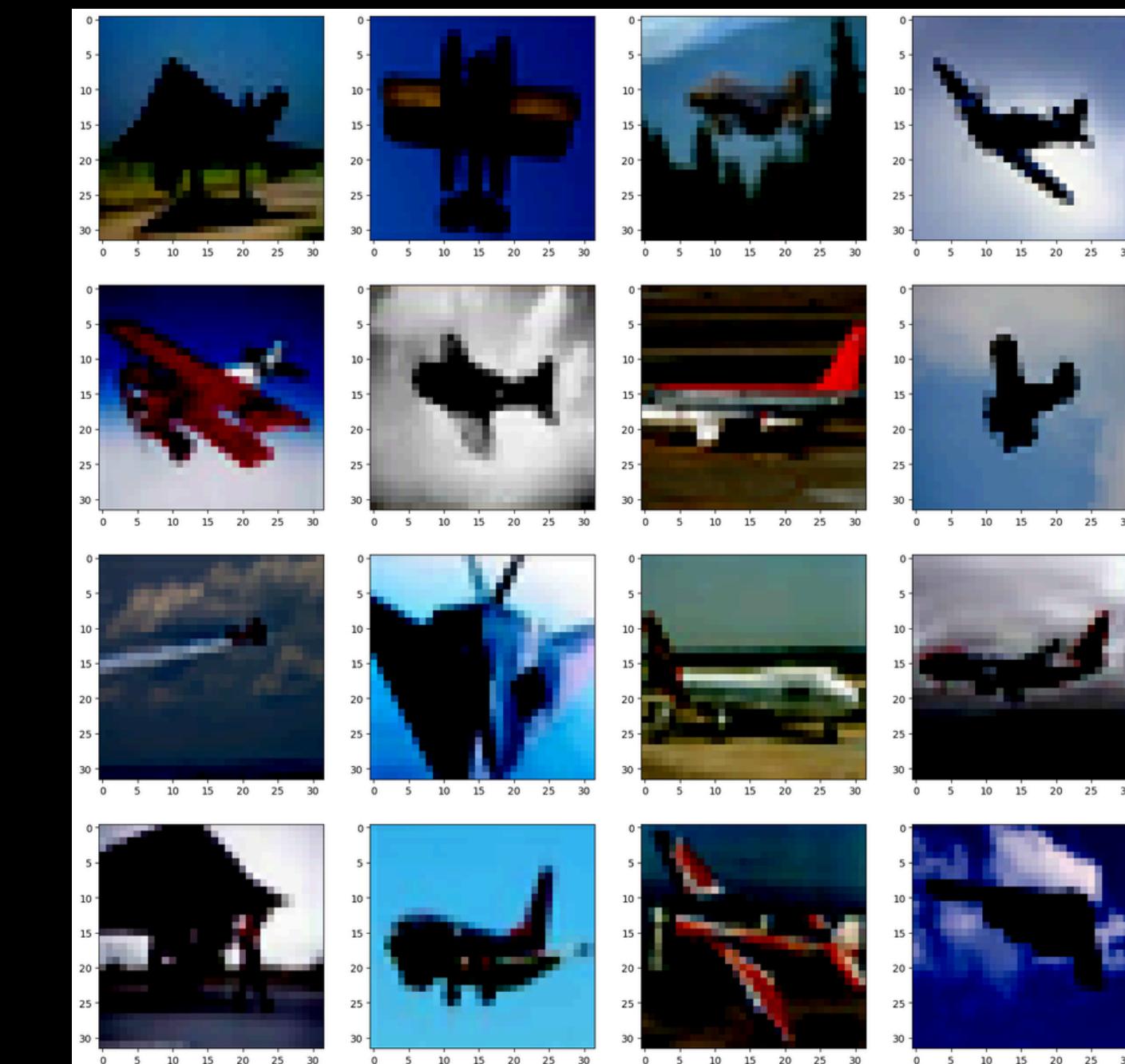
t=0

DATA

MNIST: 60 000 IMAGES



CIFAR10 AVION : 5000 IMAGES



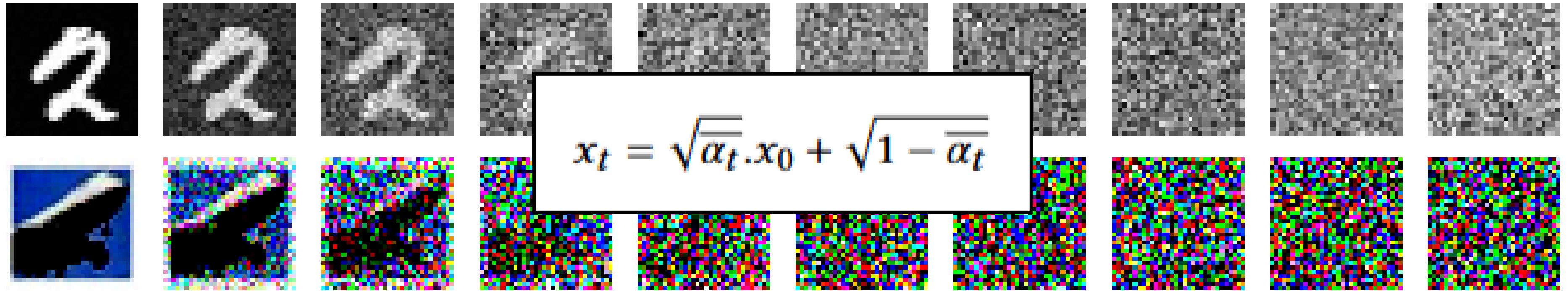
DATA

CIFAR10 : 50 000 IMAGES



MATHS

FORWARD PROCESS



$$q(x_{1:T} | x_0) := \prod_{t=1}^T q(x_t | x_{t-1})$$

$$q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)\mathbf{I})$$

$$q(x_t | x_0) = \mathcal{N}(x_t | \sqrt{\alpha_t}x_0, (1 - \alpha_t)\mathbf{I})$$

MATHS

$$q(x_{1:T} \mid x_0) = q(x_T \mid x_0) \prod_{t=2}^T q(x_{t-1} \mid x_t, x_0),$$

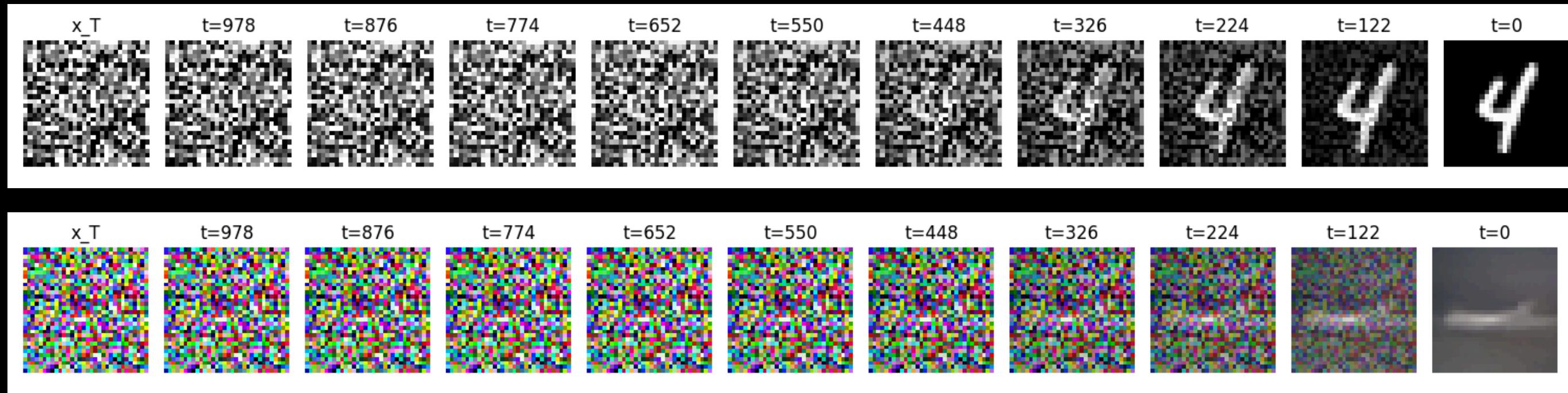
$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}\left(\sqrt{\alpha_{t-1}}x_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \cdot \frac{x_t - \sqrt{\alpha_t}x_0}{\sqrt{1 - \bar{\alpha}_t}}, \sigma_t^2 \mathbf{I}\right)$$

$$f_\theta^{(t)}(x_t) := \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon_\theta^{(t)}(x_t)}{\sqrt{\alpha_t}}.$$

$$p_\theta^{(t)}(x_{t-1} \mid x_t) = \begin{cases} \mathcal{N}\left(f_\theta^{(1)}(x_1), \sigma_1^2 \mathbf{I}\right), & \text{si } t = 1, \\ q(x_{t-1} \mid x_t, f_\theta^{(t)}(x_t)), & \text{sinon.} \end{cases}$$

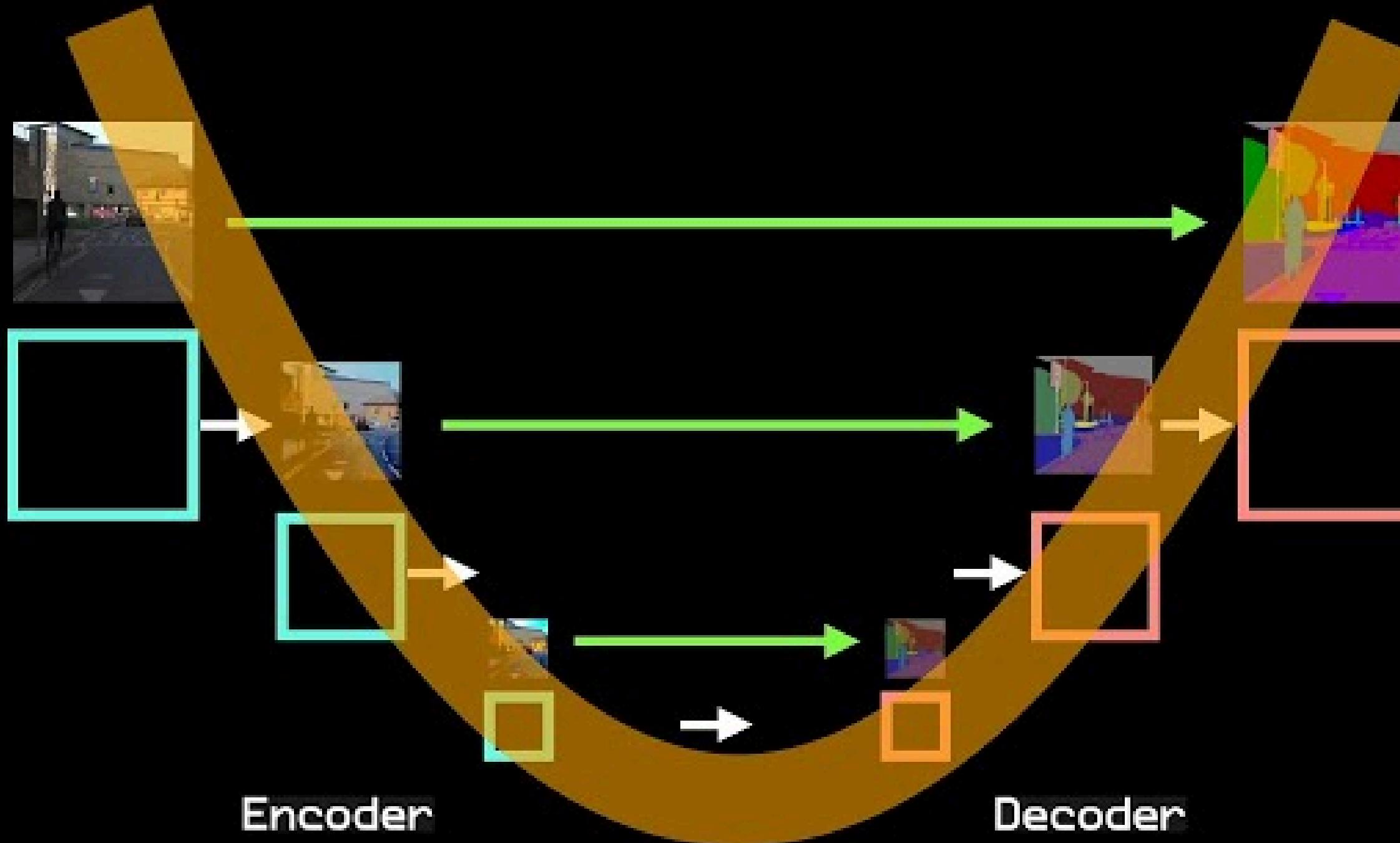
MATHS

BACKWARD PROCESS



$$x_{T-1} = \sqrt{\alpha_{T-1}} \underbrace{\left(\frac{x_T - \sqrt{1 - \bar{\alpha}_T} \cdot \epsilon_{\theta}^{(\tau)}(x_T)}{\sqrt{\bar{\alpha}_T}} \right)}_{\text{"predicted } x_0\text{"}} + \underbrace{\sqrt{1 - \bar{\alpha}_{T-1} - \sigma_{\tau}^2} \cdot \epsilon_{\theta}^{(\tau)}(x_T) + \underbrace{\sigma_{\tau} \epsilon_{\tau}}_{\text{random noise}}}_{\text{"direction pointing to } x_t\text{"}}$$

MODELE U-NET

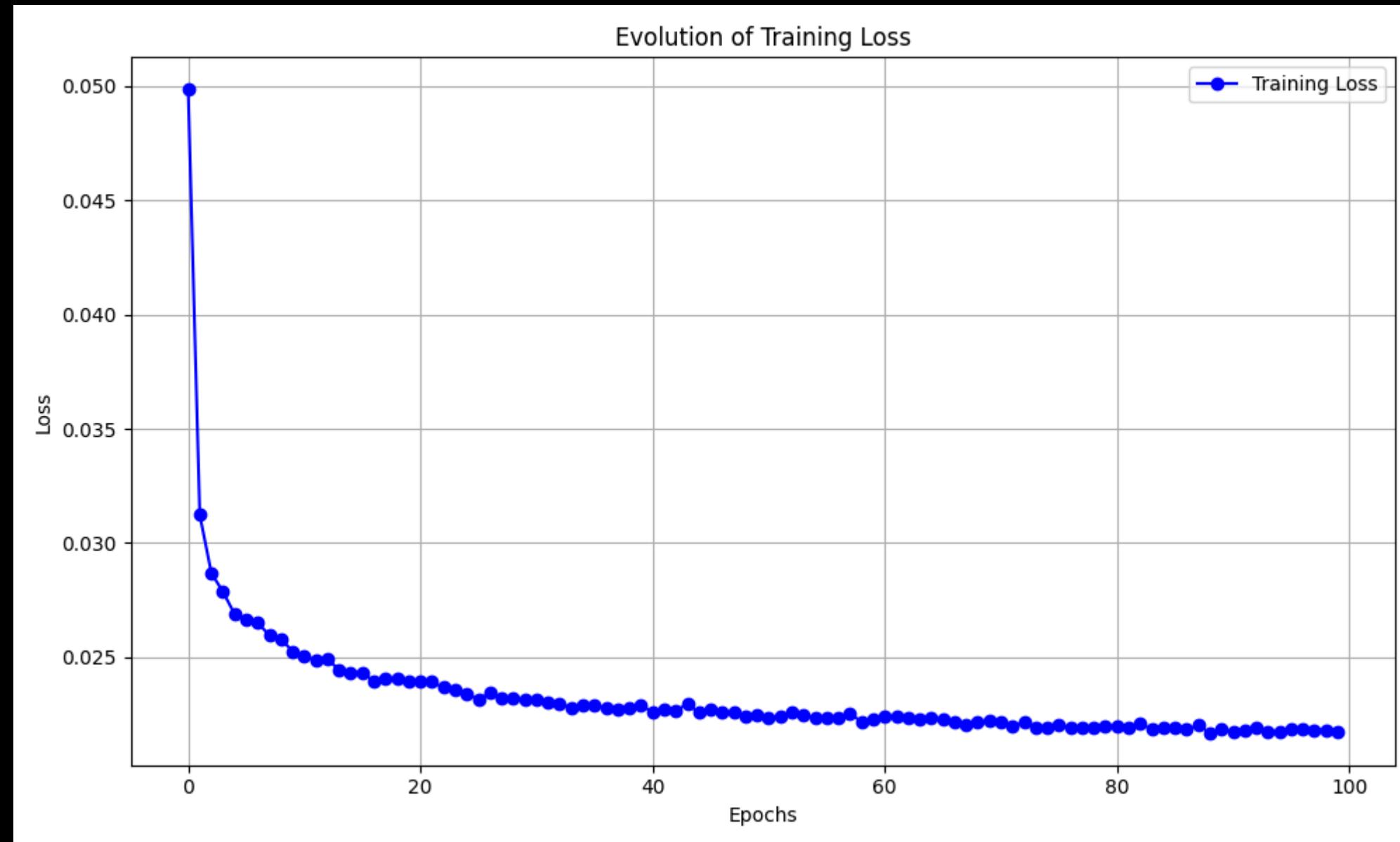


LOSS FUNCTION

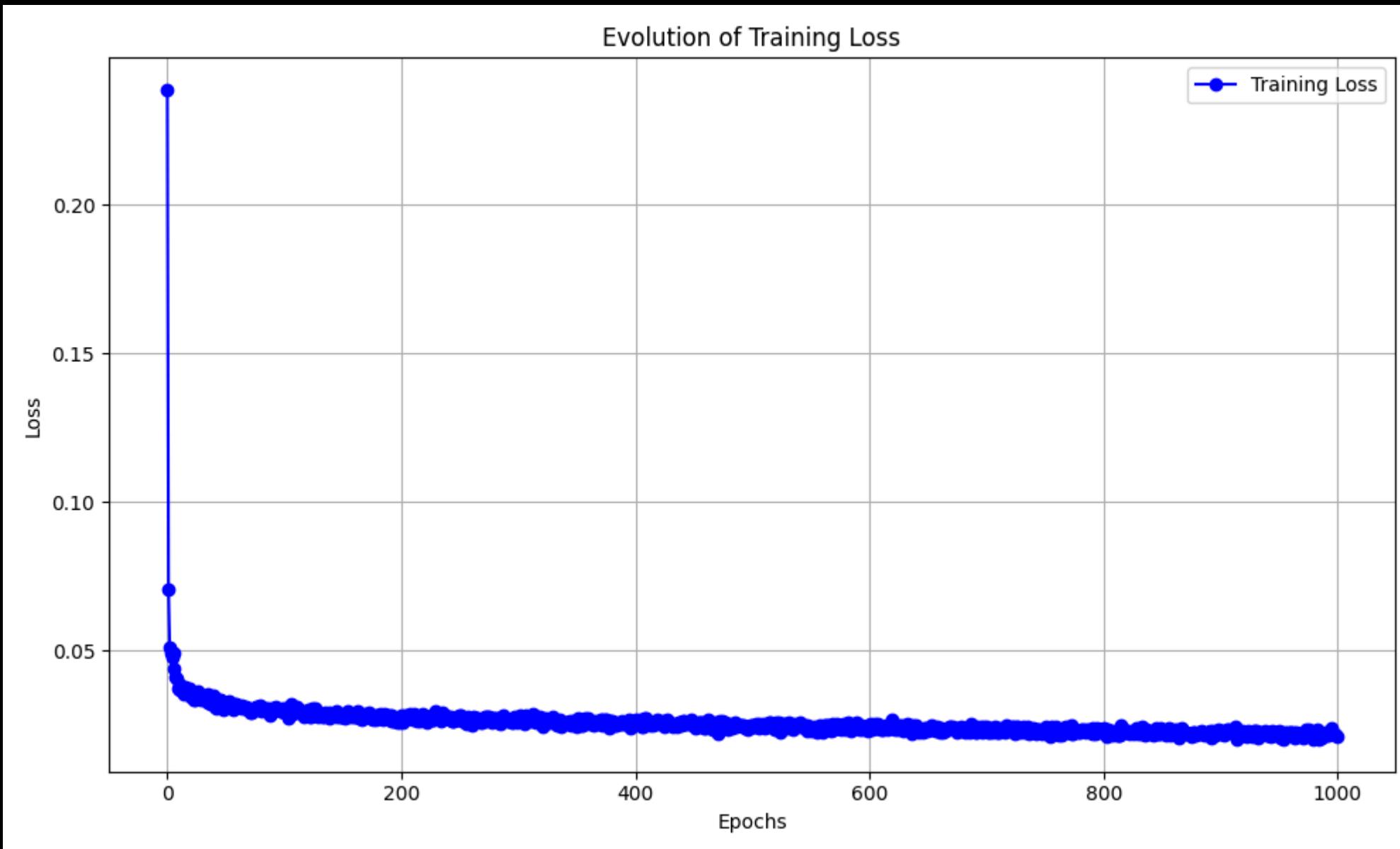
$$\frac{\gamma_t}{2\sigma^2} \|\epsilon_t - \epsilon_{\theta}^{(t)}(x_t)\|^2$$

$$\|\epsilon_t - \epsilon_{\theta}^{(t)}(x_t)\|^2$$

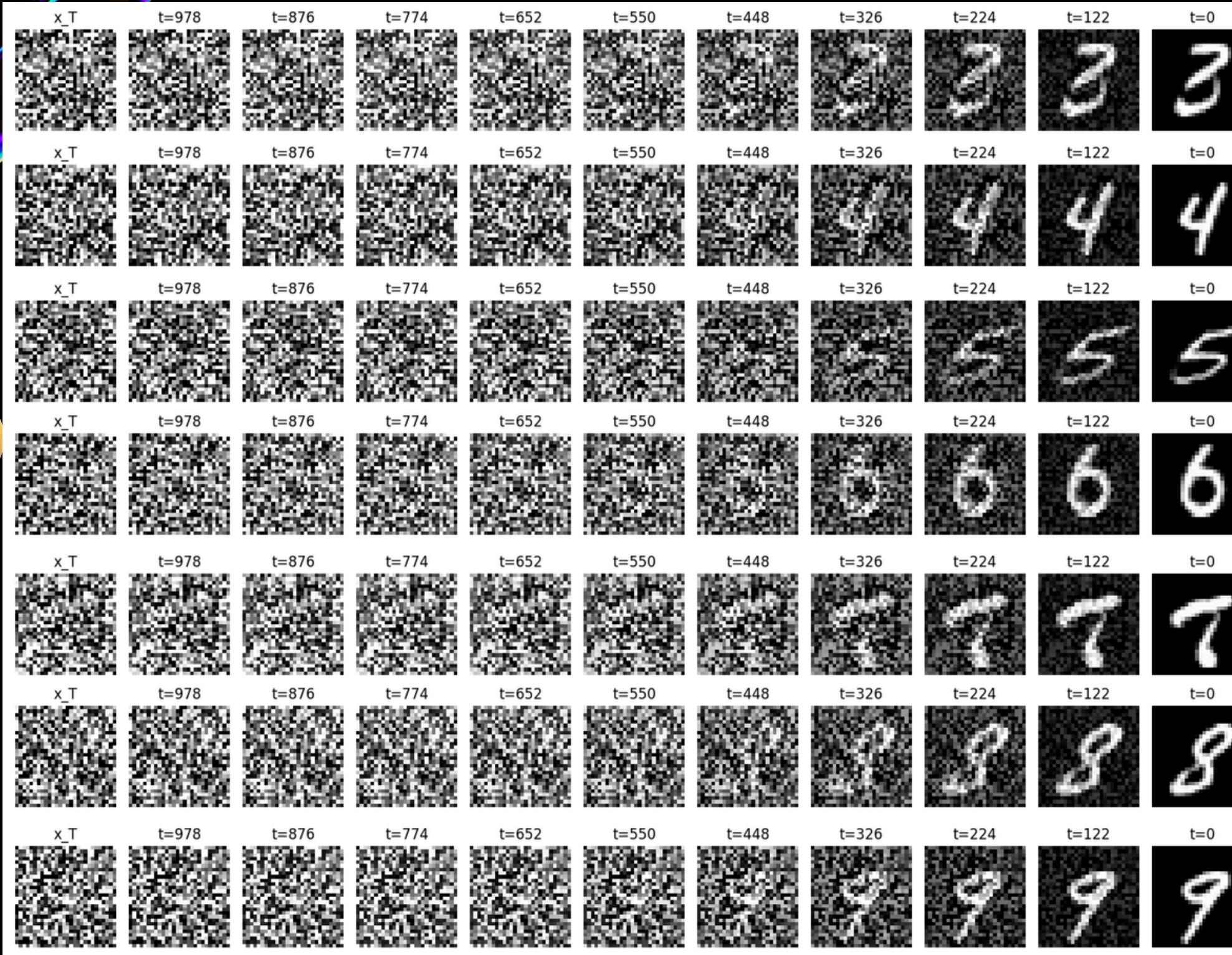
ENTRAINEMENT MNIST



ENTRAINEMENT CIFAR10 (AVION)



GÉNÉRATION MNIST



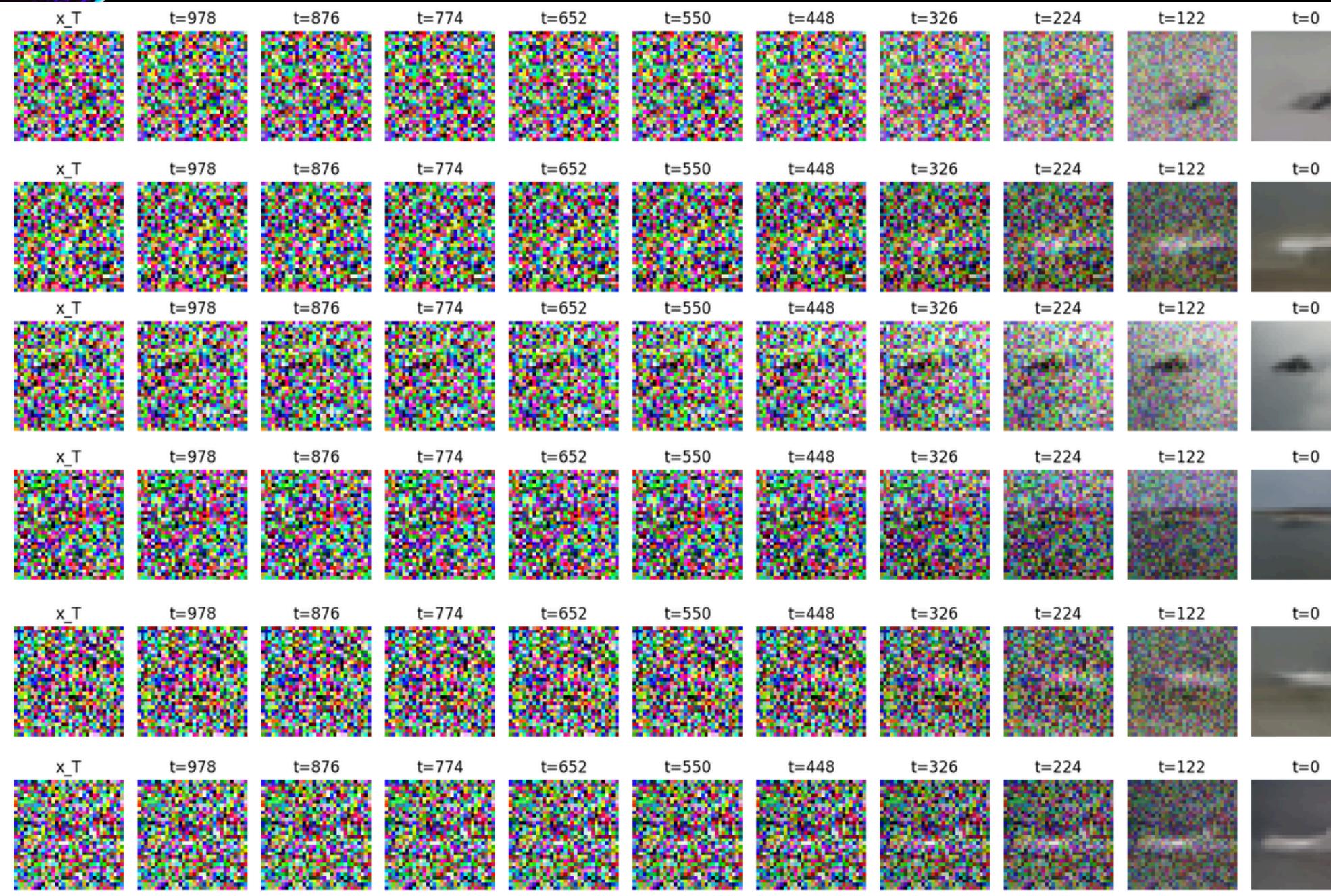
T=1000 ÉTAPES

100 EPOCHS

GPU NVIDIA H100 NVL

ENVIRON 50 MINUTES

GÉNÉRATION CIFAR10 (AVION)



T=1000 ÉTAPES

1000 EPOCHS

GPU NVIDIA H100 NVL

ENVIRON 50 MINUTES



CONCLUSION

- MODÈLE TRÈS PERFORMANT SUR MNIST
 - GÉNÉRATION FIDÈLE
 - IMAGES NETTES
- MODÈLE MOINS PERFORMANT SUR CIFAR10 (AVION)
 - PLUS D'IMAGE
 - SURAPPRENTISSAGE
 - AMÉLIORATION DU RÉSEAU U-NET