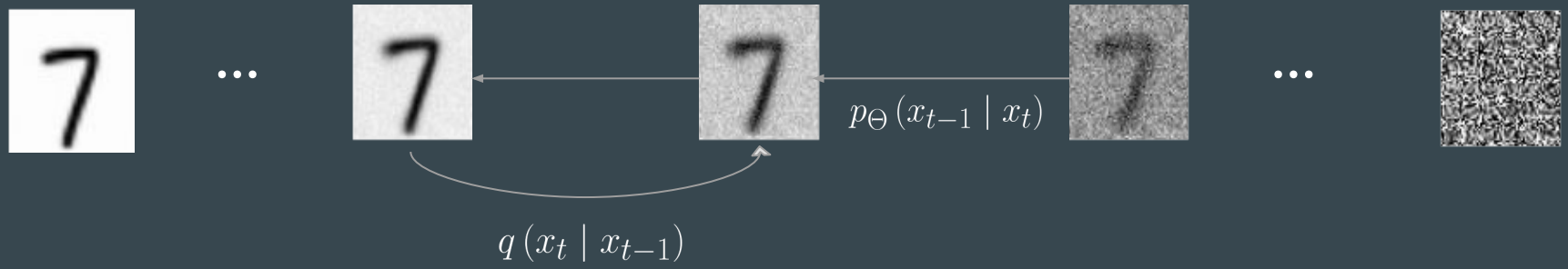
Diffusion Model

Kayo Christian, Vallaeys Théophane, Wemaere Maximilien

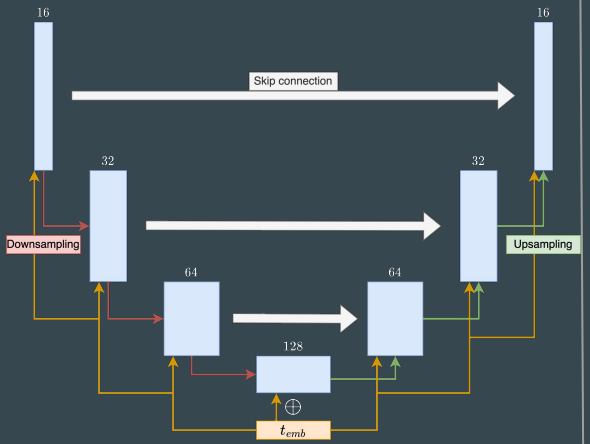
Working principle of the diffusion model

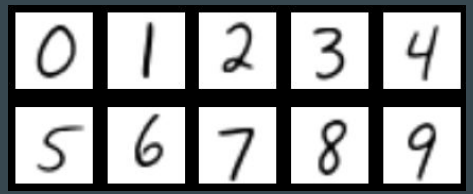
- Inspired by thermodynamics
- Based on learning how to noise to be able to denoise
- The concept is also borrowed from Markov Chains
- Originally proposed by Solh-Dickstein in 2015



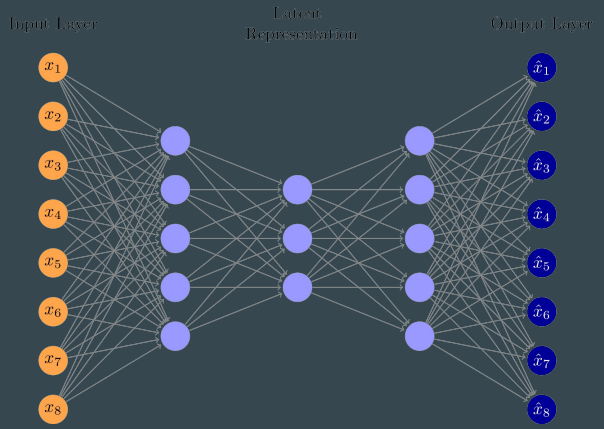
U-net vs Autoencoder vs VAE

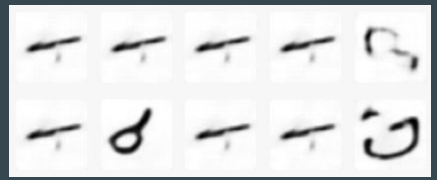
U-Net



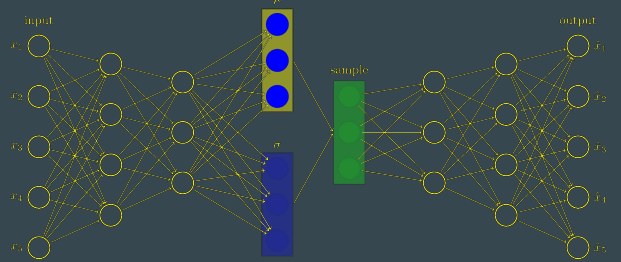


Autoencoder

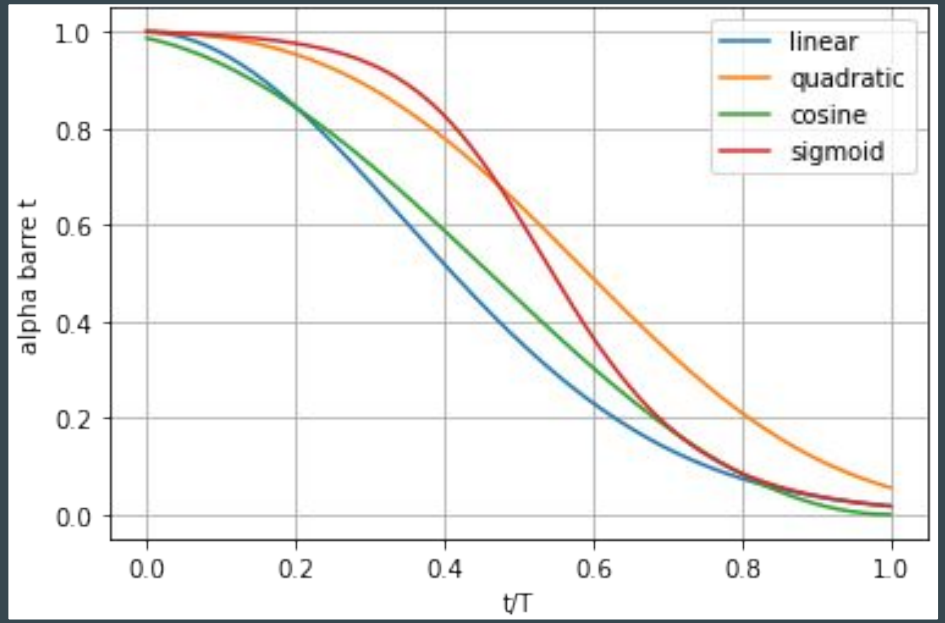




VAE



Different schedulers



FID:

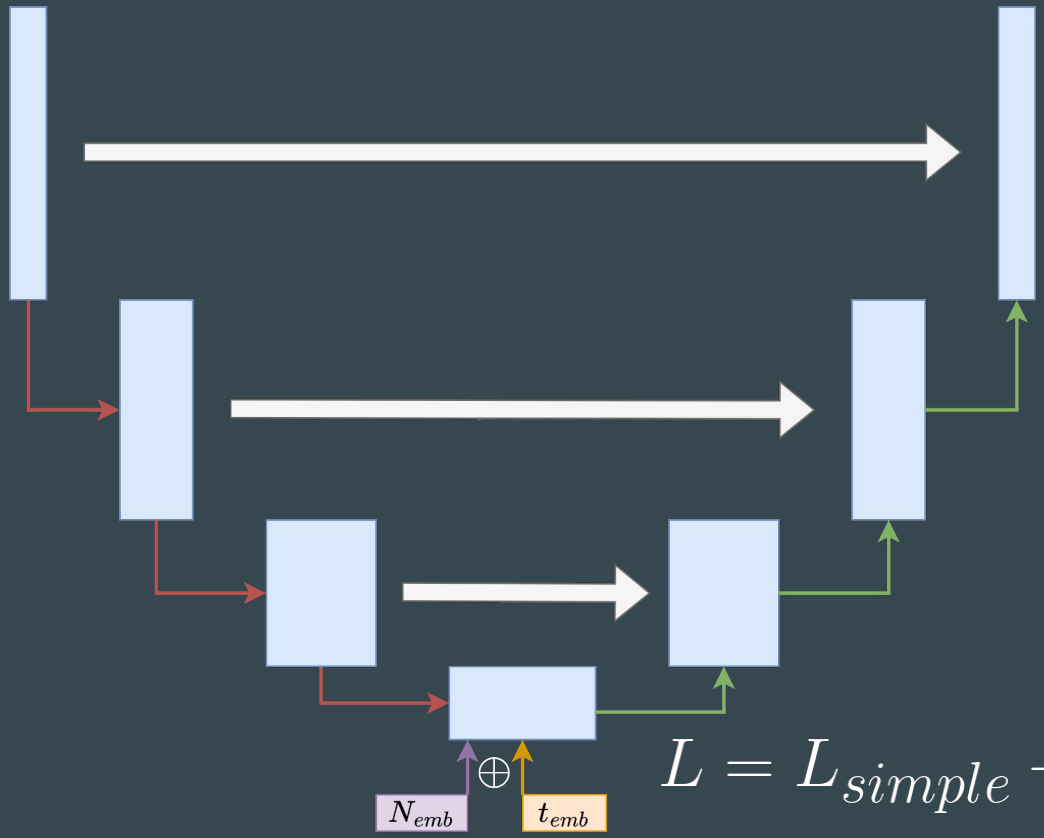
Linear: 41.43

Quadratic: 85.88

Cosine: 215.80

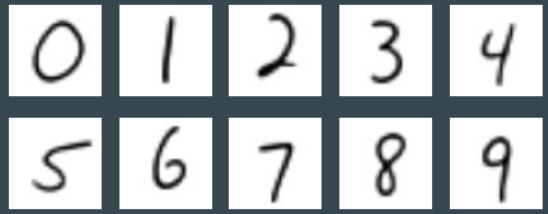
Sigmoid: 95.74

Embedding of the number to generate / Discriminator



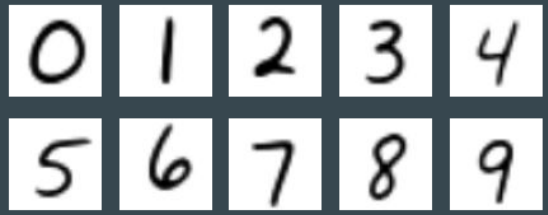
$$L = L_{simple} + \lambda_{\mathcal{D}} L_{\mathcal{D}}$$

With only N_{emb}



With N_{emb} and $\mathcal{D}$

May reduce variance (see GANs)

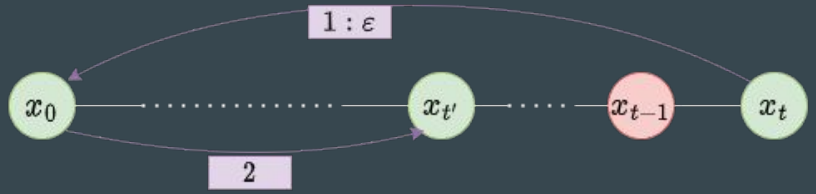


Faster image sampling

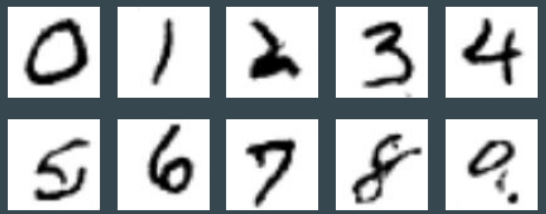
DDIM

Idea : instead of sampling using $q(x_{t-1}|x_t)$

we will predict x_0 and ε $(q(x_0|x_t))$ and then $q(x_{t'}|x_0, \varepsilon)$

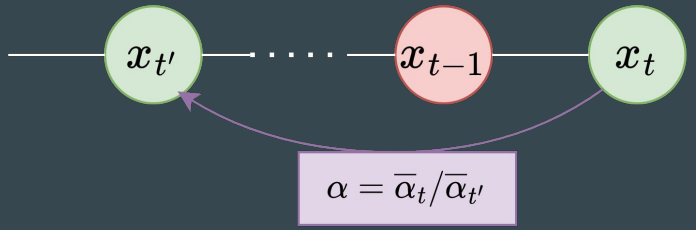


With 50 steps :

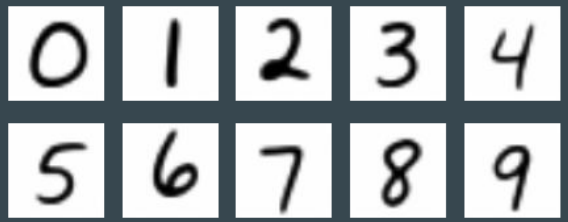


Reducing T

Training with T_{large}
Only predicting T_{small} steps



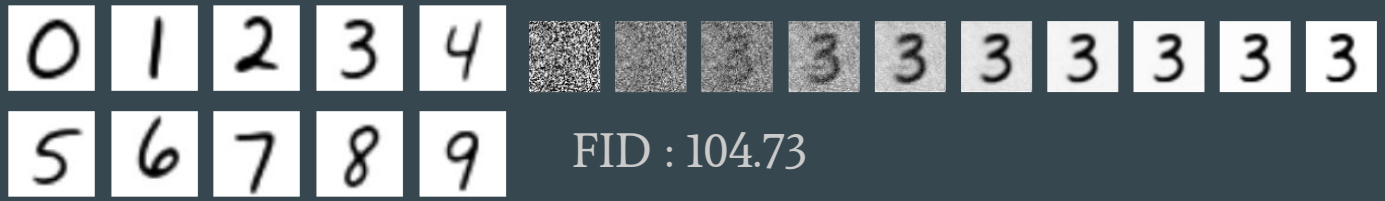
With 10 steps :



Adding variance to the sampling process

$$\mu_{t-1} = \mu_{pred}(\mu_t) + \sigma_t \varepsilon'_t$$

$\sigma_t = 0$



FID : 104.73

$\sigma_t = \sqrt{\beta_t}$

FID : 21.19

$\sigma_t = \sqrt{\tilde{\beta}_t}$

FID : 17.46

Conclusion and further research

Using the FID with a pre-trained InceptionV3 network was not reflective on sample quality on MNIST: **should be fine-tuned**.


FID of 339.05


FID of 111.02

Variance was not taken into account: **may be learned**.

Unstable results between two experiments, with large differences in the results: **should be investigated further**.

Denoising result was basically settled from the first step. **Should use a better noise scheduling, or another noising method** (gaussian blur, or noise inside a VAE latent space).