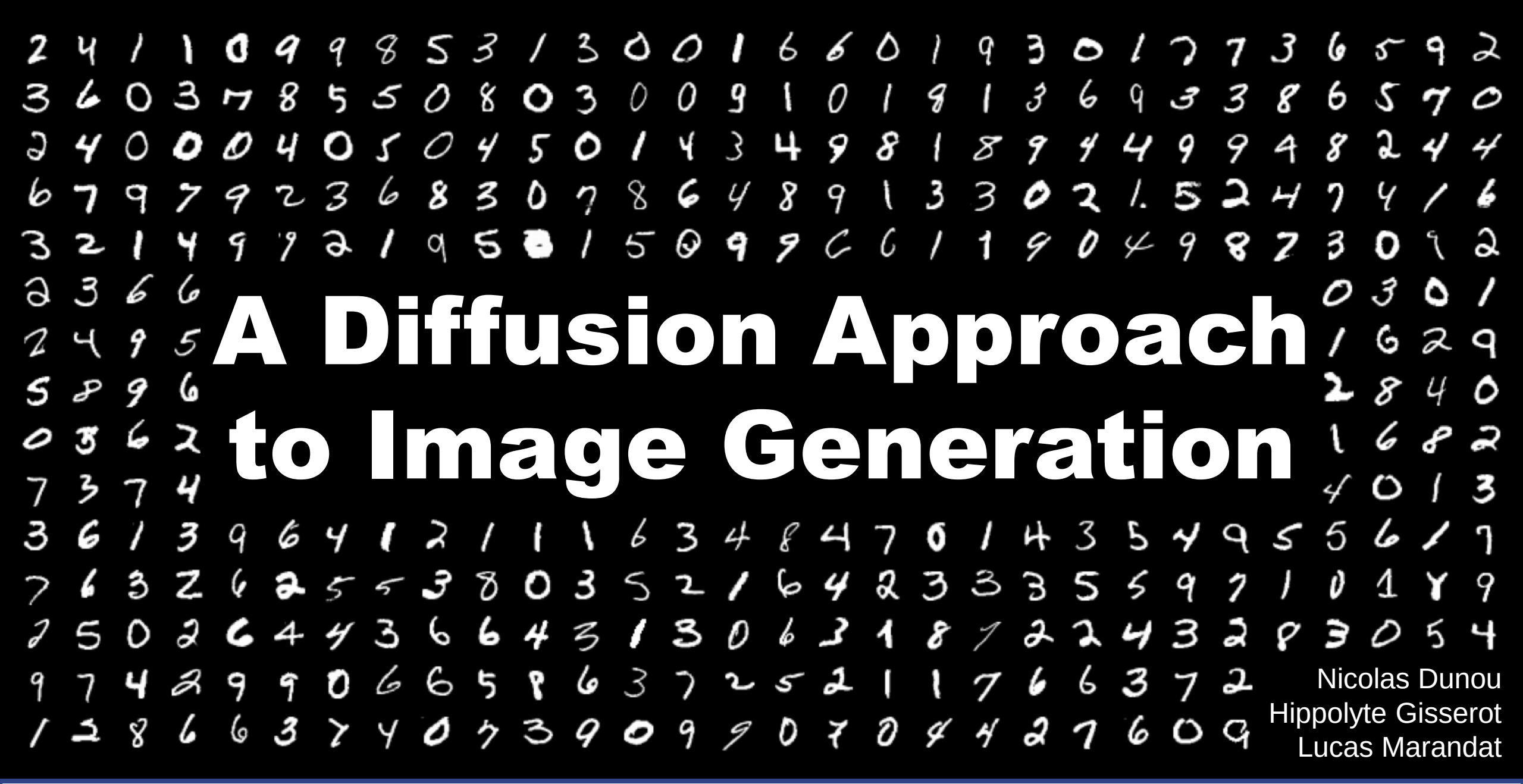


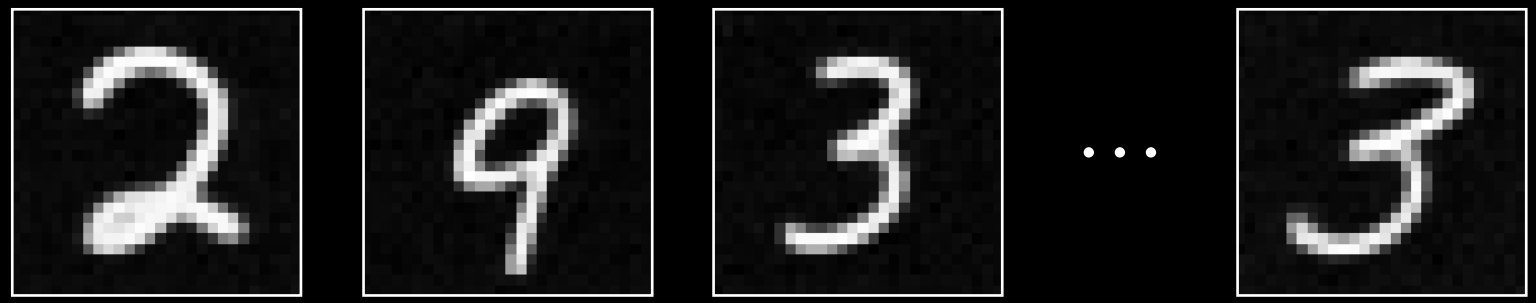
A Diffusion Approach to Image Generation

Nicolas Dunou
Hippolyte Gisserot
Lucas Marandat

Mathematical intuition

We want to sample handwritten digit images, but…

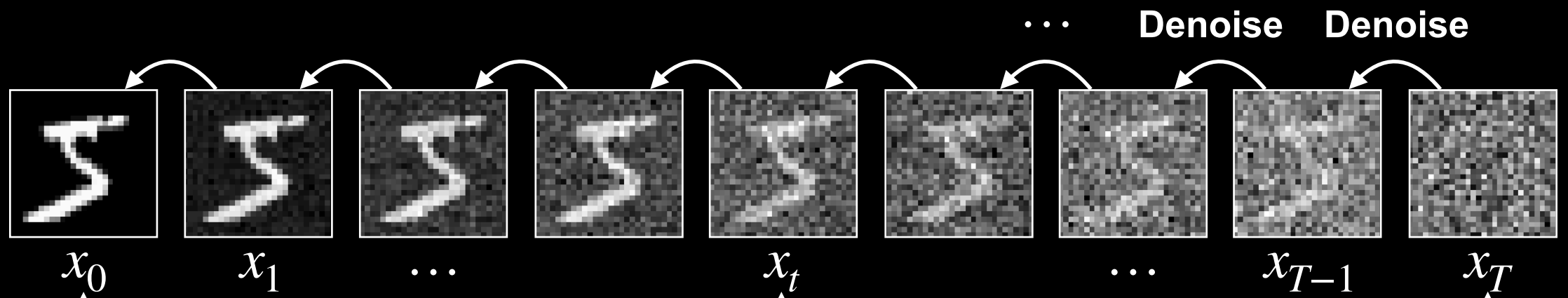
$$x \sim P(X)$$



Mathematical intuition

What if we try to learn a stepwise denoiser from a known probability?

$$x_0 \sim P_\theta(x_0 \mid x_1) P_\theta(x_1 \mid x_2) \cdots P_\theta(x_T)$$



$P_\theta(x_0 \mid x_1) = ?$

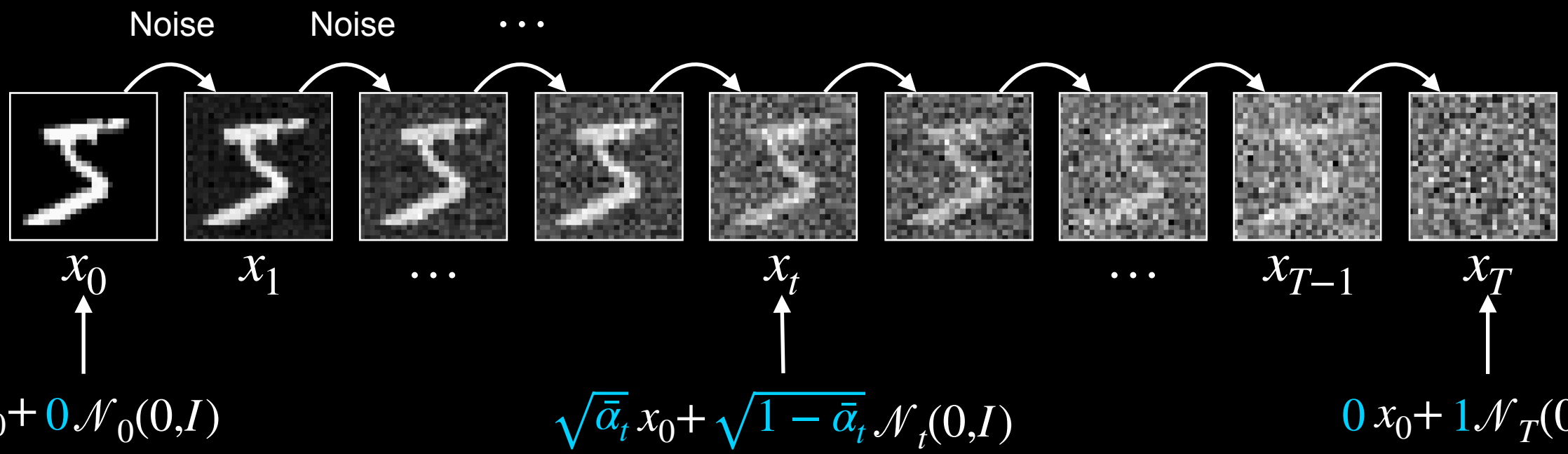
$P_\theta(x_t \mid x_{t+1}) = ?$

$P_\theta(x_T) = \mathcal{N}(0, I)$

Mathematical intuition

Noising an image is quite easy!

$$P(x_t \mid x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\, x_0,\ (1 - \bar{\alpha}_t) I)$$

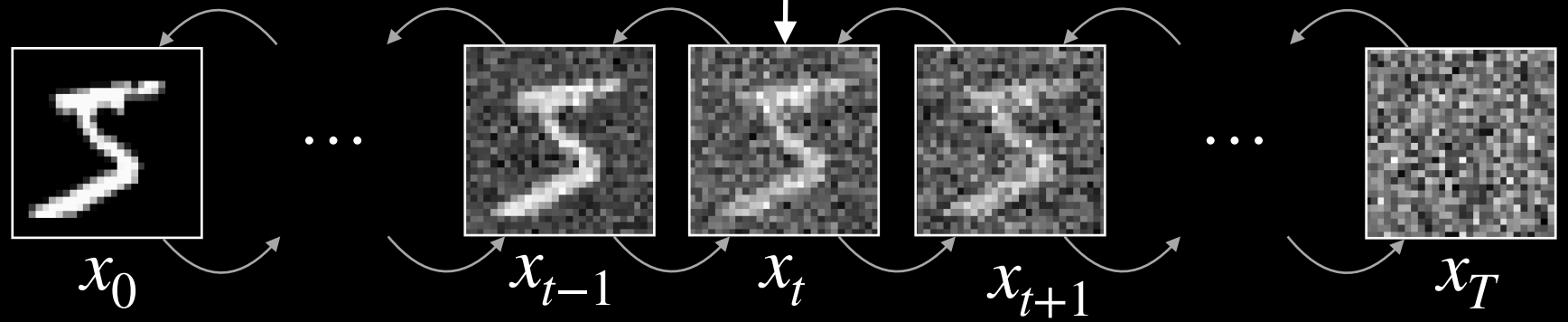


Mathematical intuition

Back to our problem

$$P(x_t \mid x_{t+1}, x_0) = \mathcal{N}\left(\tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t\right)$$

(Using Bayes' rule)

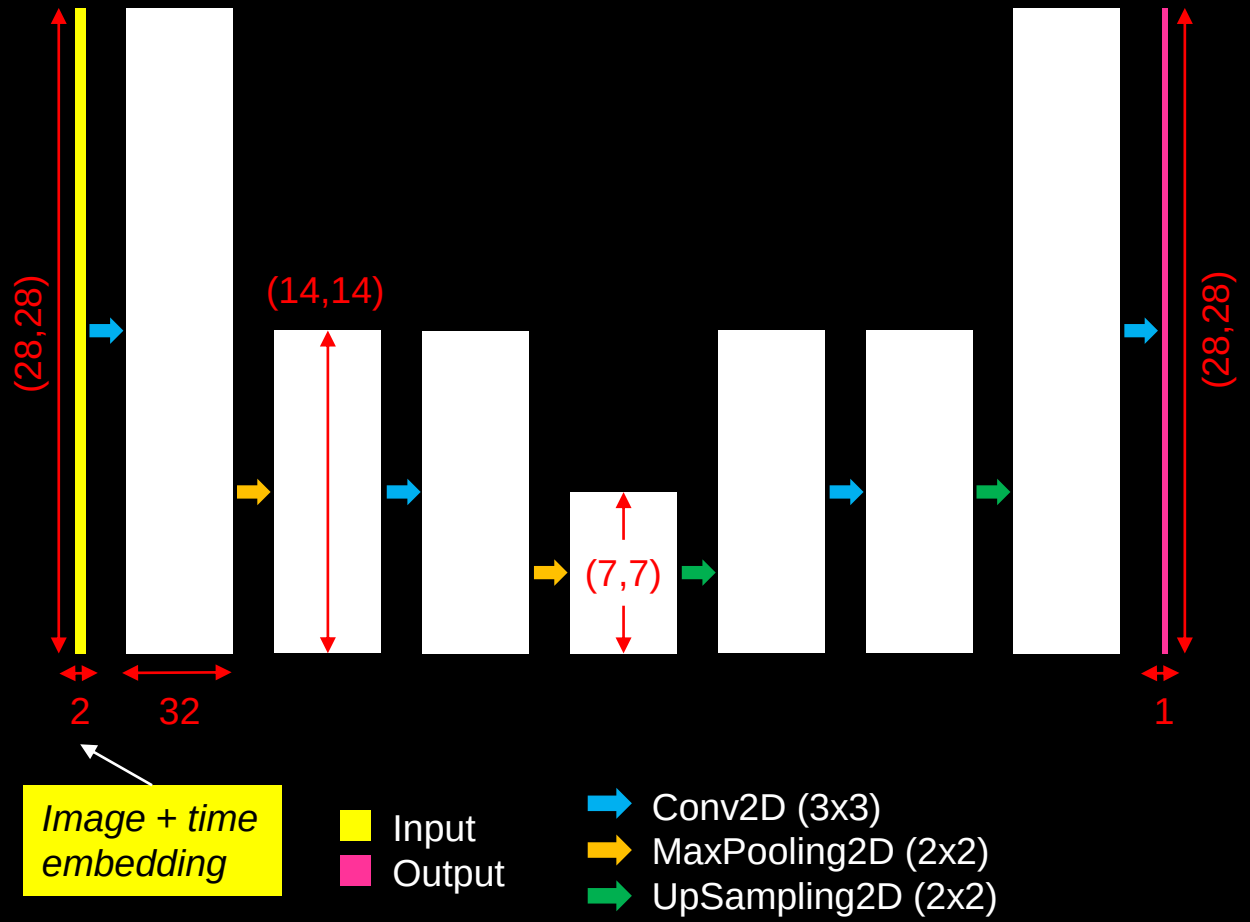


$$P_\theta(x_t \mid x_{t+1}) = \mathcal{N}\left(\mu_\theta(x_t, t), \beta_t\right) \implies \boxed{\mu_\theta(x_t, t) \approx \tilde{\mu}_t(x_t, x_0)}$$

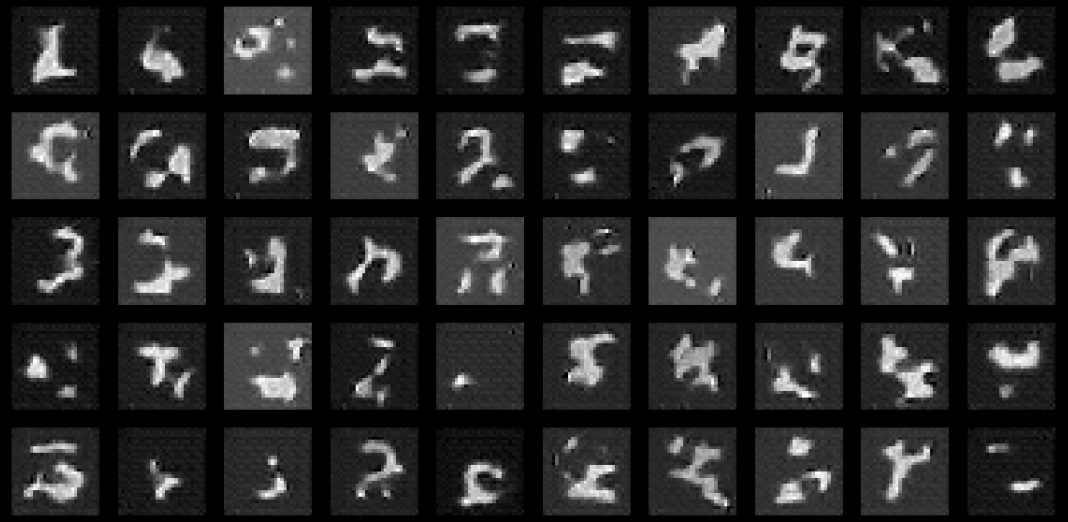
(Loss using Mean Squared Error)

Model 1

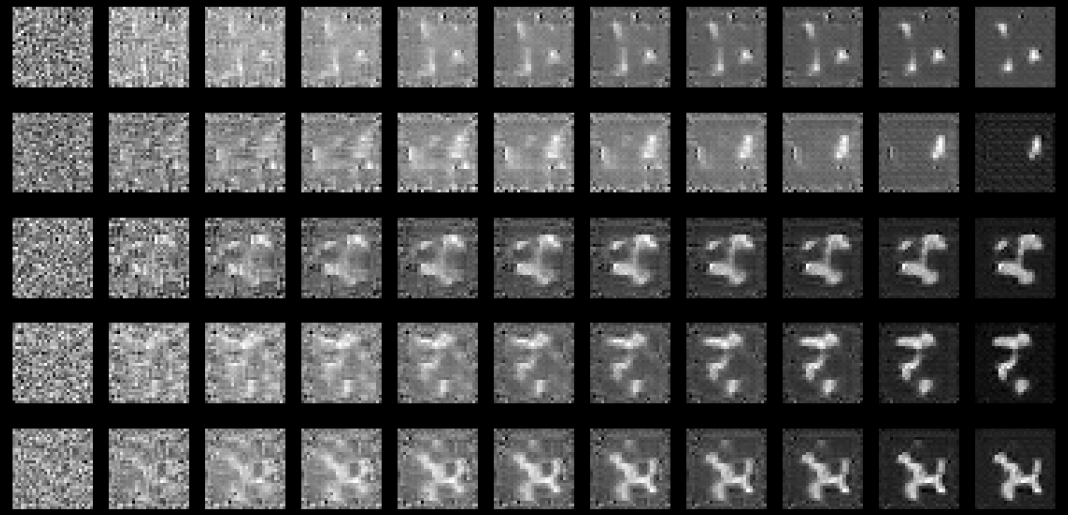
Network architecture:



Generation of 50 random images:

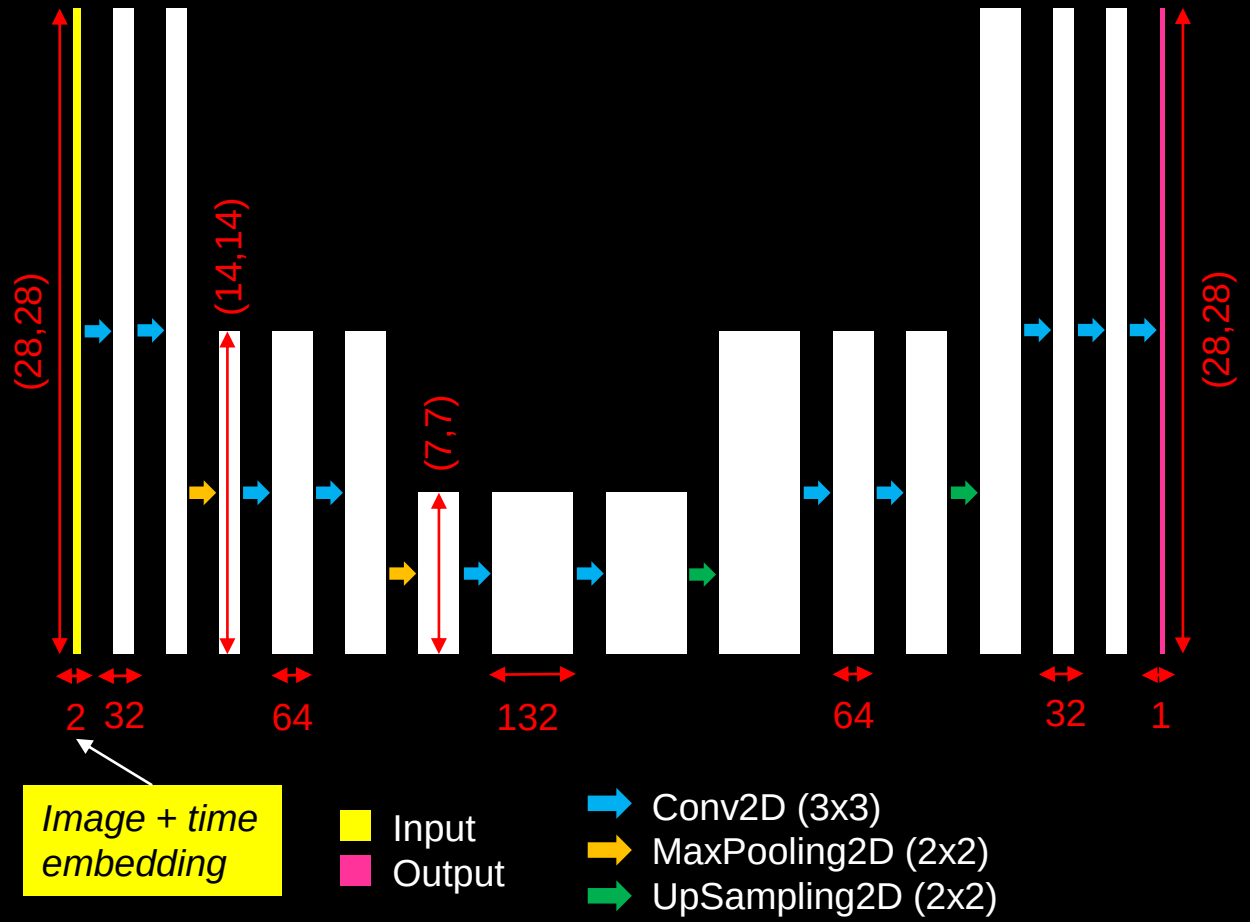


Generation of 5 random denoising paths:



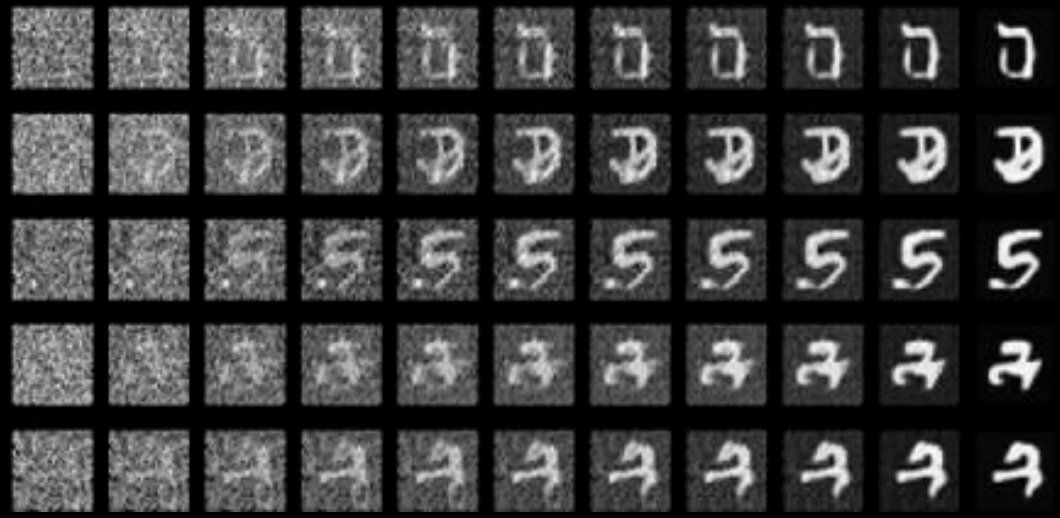
Model 2

Network architecture:



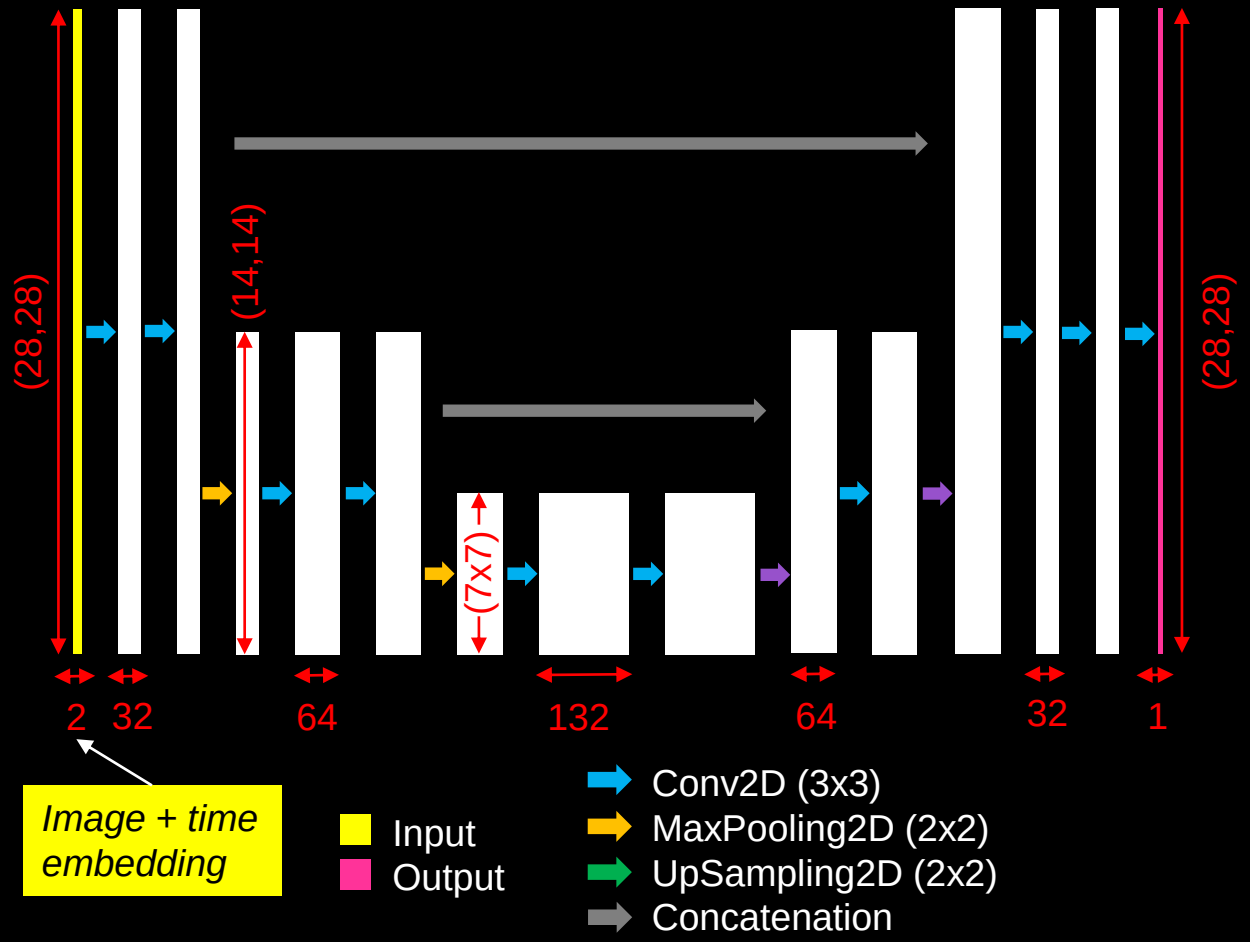
Generation of 50 random images:

Generation of 5 random denoising paths:



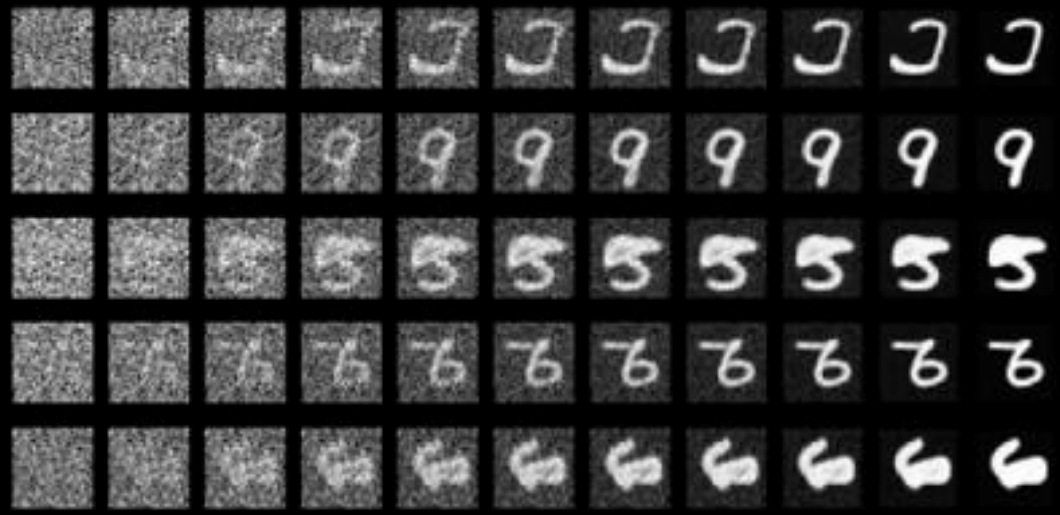
Model 3

Network architecture:



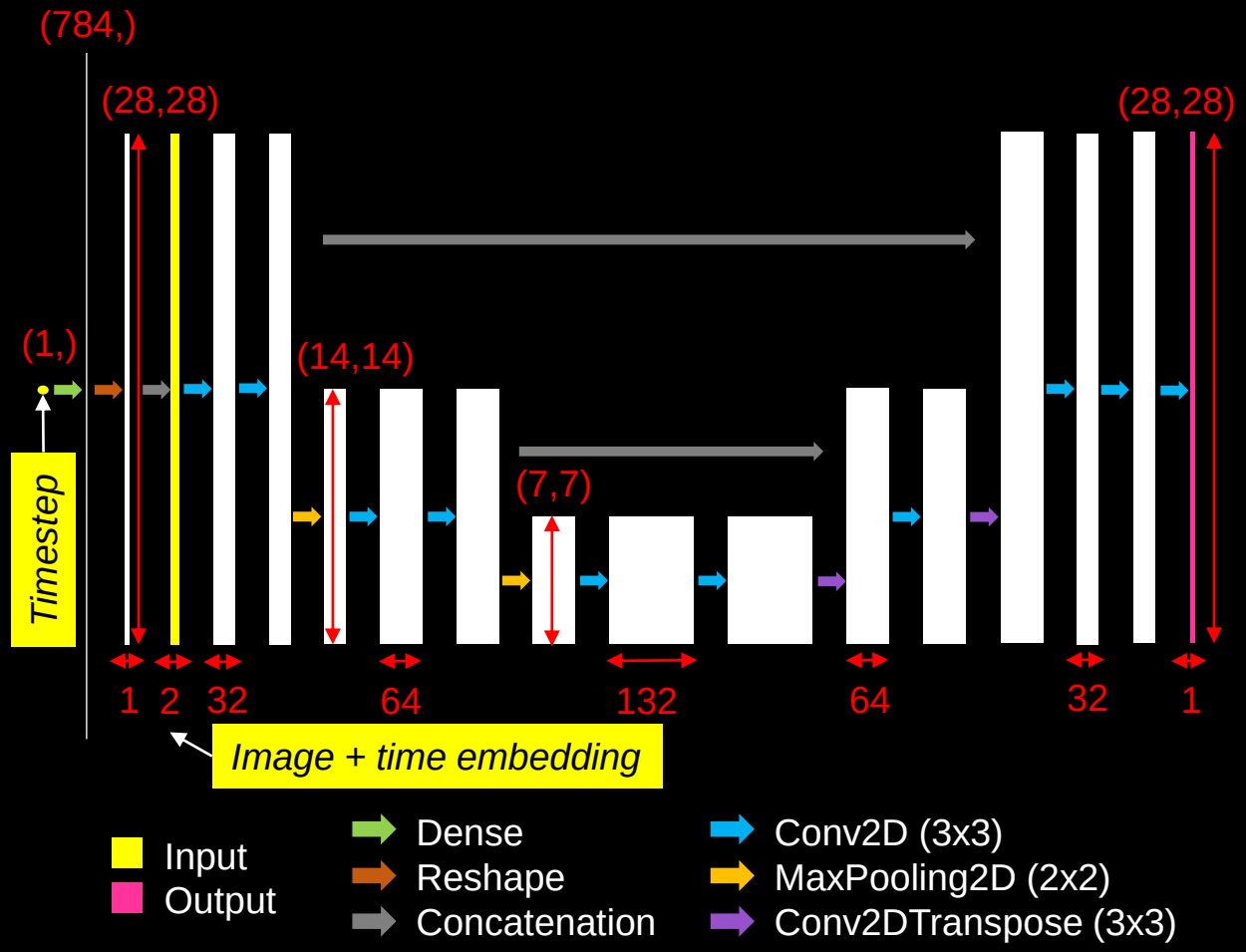
Generation of 50 random images:

Generation of 5 random denoising paths:



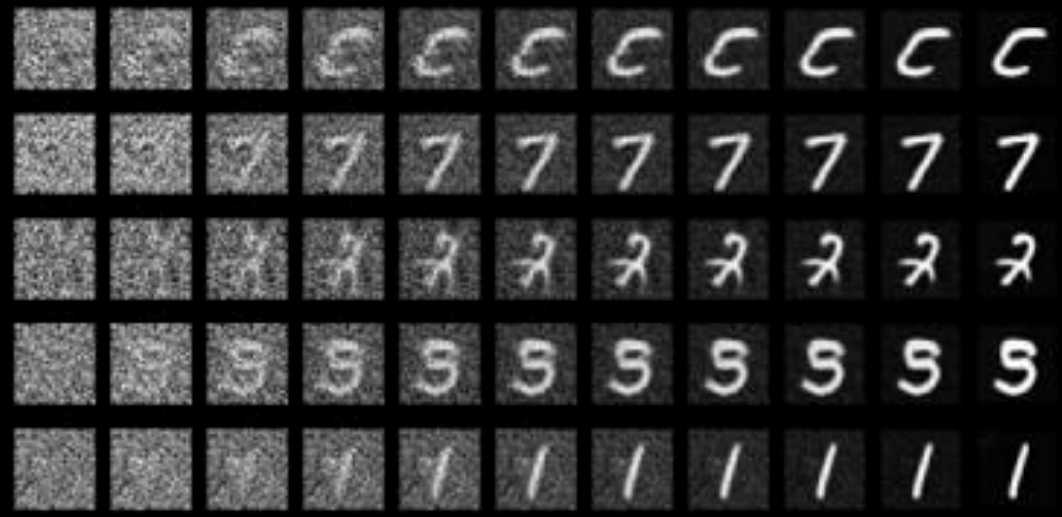
Model 4

Network architecture:



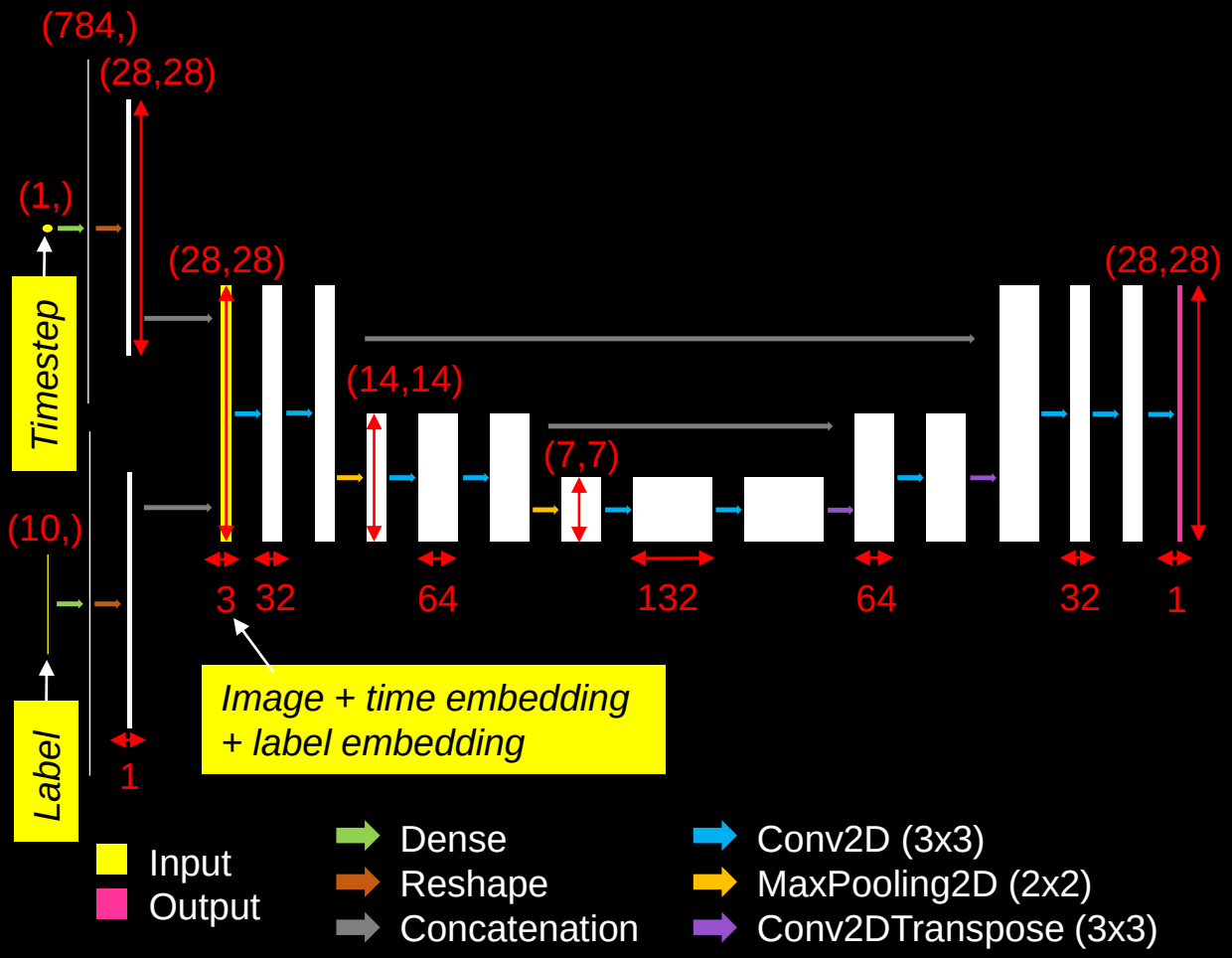
Generation of 50 random images:

Generation of 5 random denoising paths:



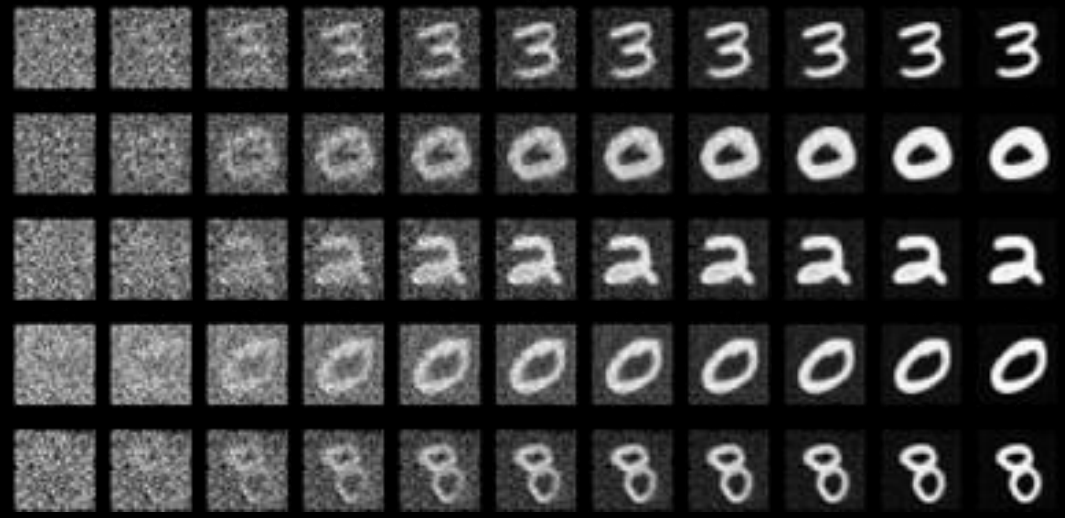
Model 5

Network architecture:



Generation of 50 random images:

Generation of 5 random denoising paths:



Evaluation metric

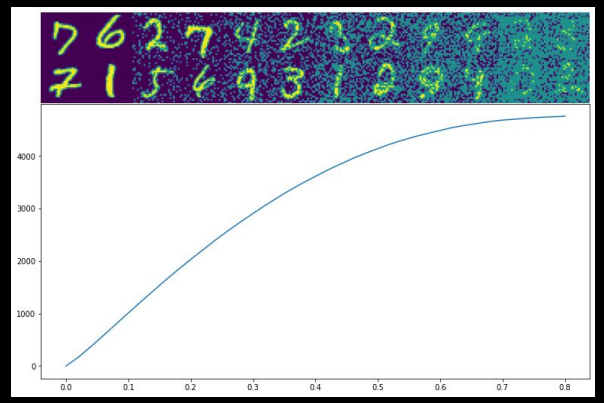
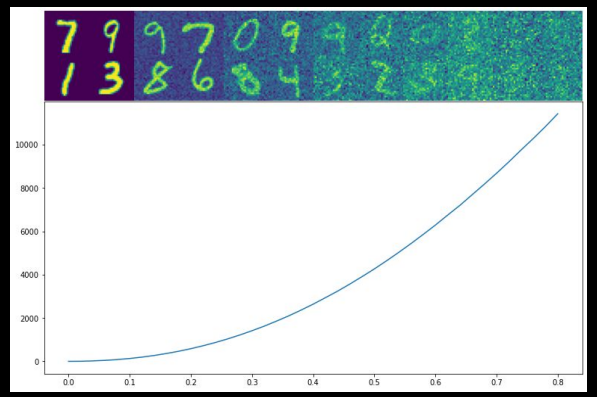
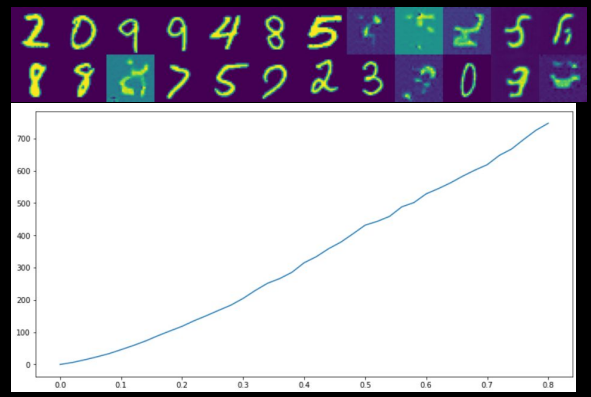
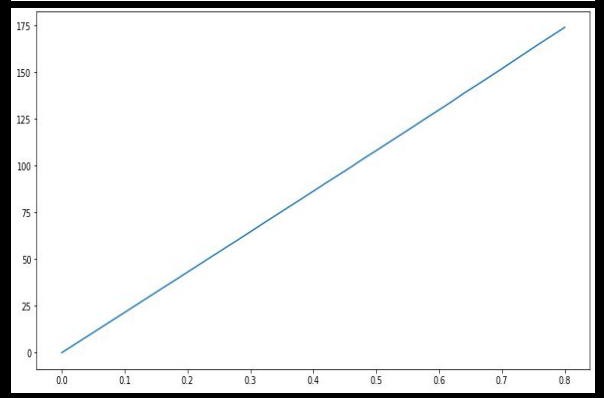
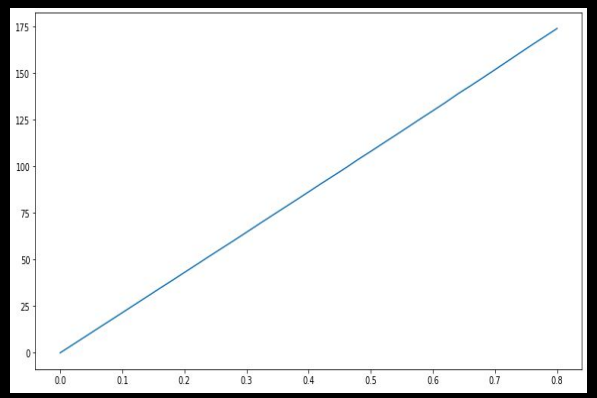
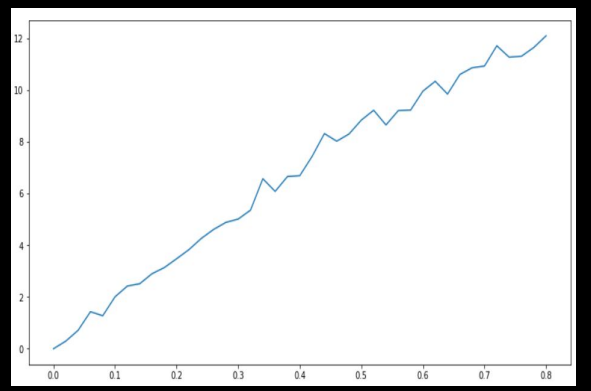
We used the **Fréchet Distance** to evaluate our models.

A distance between probability distribution that can be **estimated by sampling**.

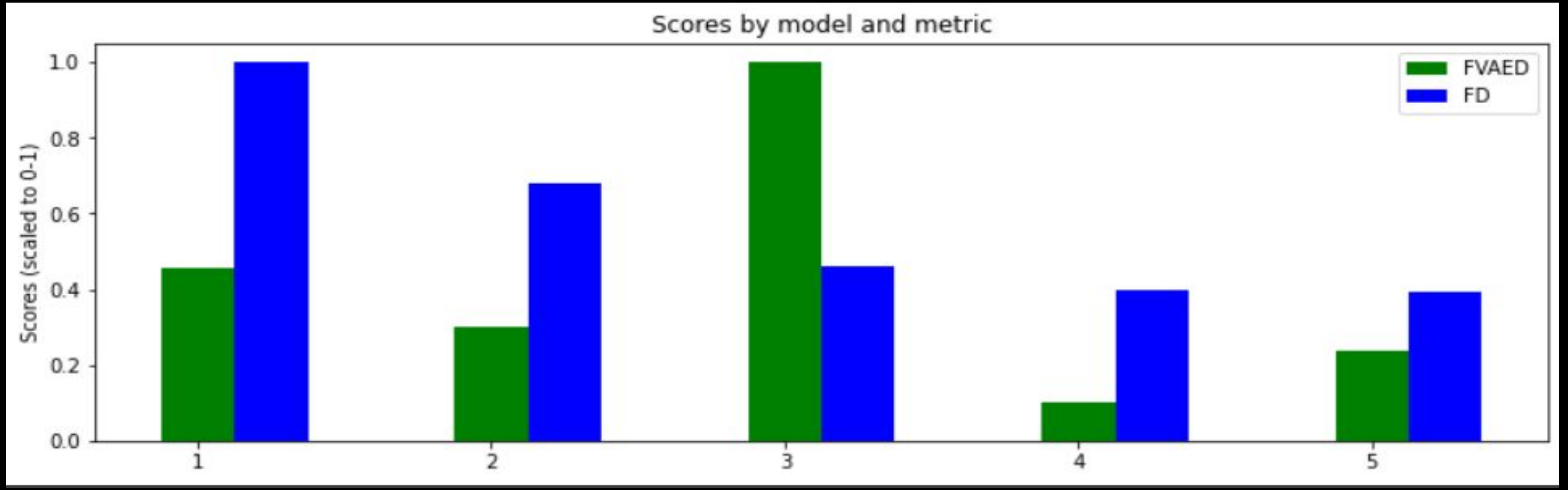
$$FD(X,Y)^2 = |\mu_X - \mu_Y|^2 + \mathrm{Tr}(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y})$$

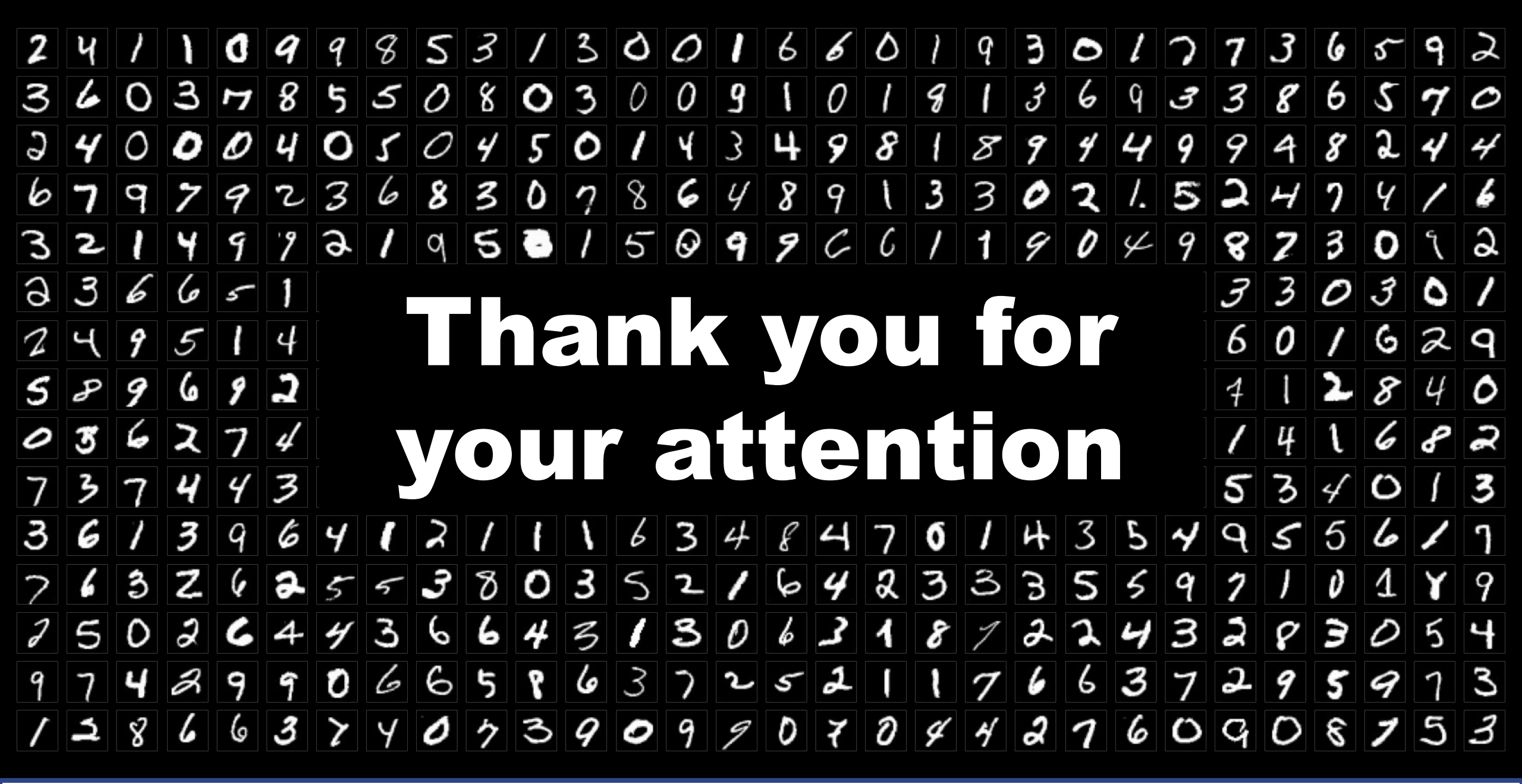
To obtain better results, we choose to apply the Fréchet Distance to an **embedding** obtained with a **Variationnal Auto Encoder**

Metric comparison

	Dead pixels	Gaussian noise	Invasion by another set
FD			
FD + VAE			

Results evaluation




Thank you for
your attention