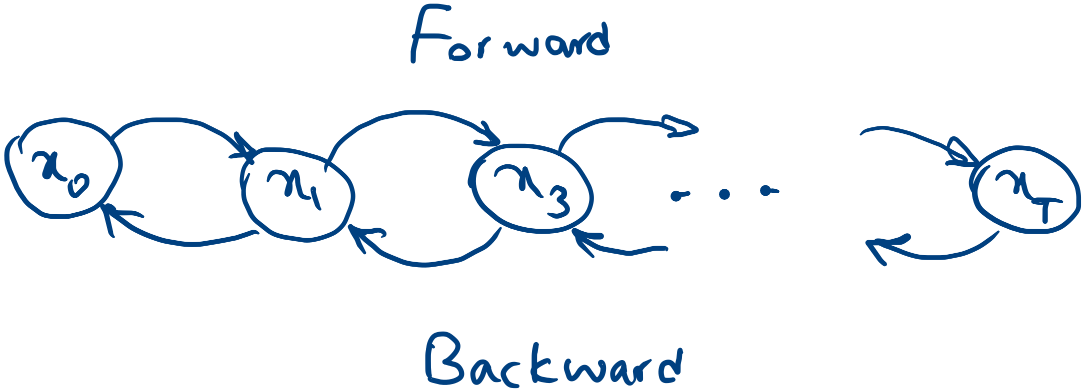
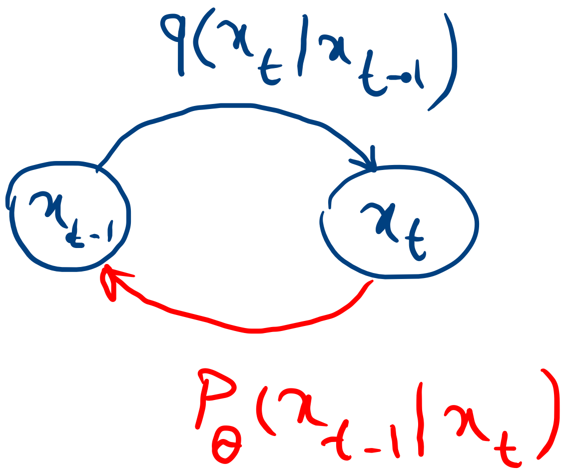


$T \simeq O(1000)$

$x_T \sim N(0, I)$



$q(x_t | x_{t-1}) = N(\mu, \Sigma)$

$$\max_\theta \underbrace{\log P(x_0)}_{\text{Evidence}} \geq \max_\theta \text{ELBO}$$

$$z : x_1, \dots, x_T$$

$$x : x_0$$

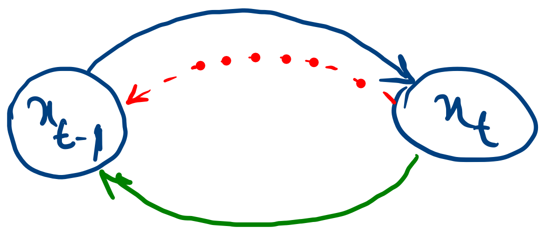
$$\text{ELBO} = -KL\Big(q(z|x) \,\|\, P(x,z)\Big)$$

$$\max_\theta -KL\Big(q(x_1, \dots, x_T | x_0) \,\|\, P(x_0, x_1, \dots, x_T)\Big)$$

$$\max_\theta -E\left[\log P(x_T) + \sum_{t=1}^{T} \log \frac{q(x_t | x_{t-1})}{P_\theta(x_{t-1} | x_t)}\right]$$

$$\cdots + \sum_{t=1}^{T} \boxed{KL\Big(\overbrace{q(x_{t-1}|x_t,x_0)}^{N(\tilde{\mu},\tilde{\Sigma})} \,\Big\|\, \overbrace{P_\theta(x_{t-1}|x_t)}^{N(\mu_\theta,I)}\Big)} + \cdots \longrightarrow N(\mu_\theta, \underset{I}{\Sigma_\theta})$$

$q(x_t|x_{t-1})$



$P_\theta(x_{t-1}|x_t)$

$$\underbrace{q(x_{t-1}|x_t)}_{\text{نرمال نیست}} = \frac{\overbrace{q(x_t|x_{t-1})}^{\text{نرمال}}\, q(x_{t-1})}{q(x_t)}$$

$$\rightarrow \boxed{q(x_{t-1}|x_0,x_t)} = N(\tilde{\mu}, \tilde{\Sigma})$$

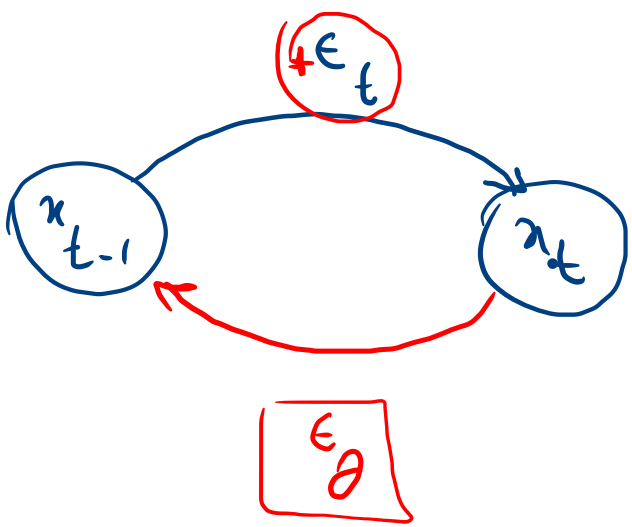
$$\min_{\theta} \sum_{t=1}^{T} E\left[\| \mu_\theta(x_t) - \tilde{\mu} \|_2^2 \right]$$

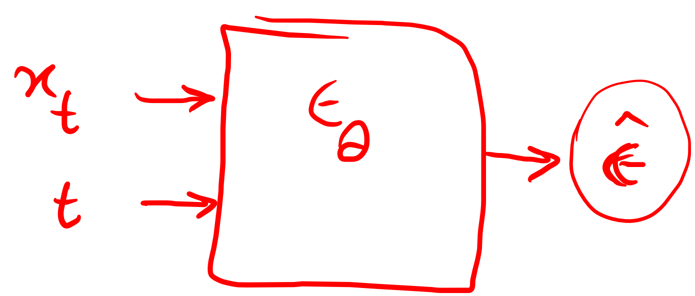
$$\mathcal{N}(x \mid \mu, \sigma^2 I)$$

$$\boxed{\min_{\theta} \sum_{t=1}^{T} E\left[\| \epsilon_\theta^{t}(x_t) - \epsilon_t \|_2^2 \right]}$$

$$\boxed{x = \mu + \sigma \epsilon}$$

$$\epsilon \sim \mathcal{N}(0, I)$$





x_0 $\qquad\qquad\qquad\qquad\qquad 0 \le \alpha_1 \le 1$

$$x_1 = \sqrt{\alpha_1}\, x_0 + \sqrt{1-\alpha_1}\, \epsilon_1 \qquad\qquad \epsilon_1 \sim N(0, I)$$

$$x_2 = \sqrt{\alpha_2}\, x_1 + \sqrt{1-\alpha_2}\, \epsilon_2$$

$$\vdots$$

$$x_T$$

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(X_0) = 1$$

$$Var(X_1) = \alpha_1 \underbrace{Var(X_0)}_{1} + (1-\alpha_1) = 1$$

$$x_0 \to x_1 \to x_2 \cdots \to x_T$$

x_0

$$x_1 = \sqrt{\alpha_1}\, x_0 + \sqrt{1-\alpha_1}\, \epsilon_1$$

$$x_2 = \sqrt{\alpha_2}\, \underline{x_1} + \sqrt{1-\alpha_2}\, \epsilon_2 = \sqrt{\alpha_2}\left(\sqrt{\alpha_1}\, x_0 + \sqrt{1-\alpha_1}\, \epsilon_1\right) + \sqrt{1-\alpha_2}\, \epsilon_2$$

$$= \sqrt{\alpha_1\alpha_2}\, x_0 + \boxed{\sqrt{\alpha_2(1-\alpha_1)}\, \epsilon_1 + \sqrt{1-\alpha_2}\, \epsilon_2}$$

$$= \sqrt{\alpha_1\alpha_2}\, x_0 + \sqrt{1-\alpha_1\alpha_2}\, \tilde{\epsilon}$$

$$\tilde{\epsilon} \sim N(\underbrace{\mu = ?}_{0},\ \underbrace{\sigma^2 = ?}_{1-\alpha_1\alpha_2})$$

$$\tilde{\epsilon} \sim N(0, I)$$

$$\alpha_2(1-\alpha_1) + 1 - \alpha_2 = 1 - \alpha_1\alpha_2$$

$$\bar{\alpha}_t = \alpha_1 \alpha_2 \cdots \alpha_t$$

$$x_2 = \sqrt{\alpha_1 \alpha_2}\, x_0 + \sqrt{1 - \alpha_1 \alpha_2}\, \epsilon$$

$$x_3 = \sqrt{\alpha_1 \alpha_2 \alpha_3}\, x_0 + \sqrt{1 - \alpha_1 \alpha_2 \alpha_3}\, \epsilon$$

$$\boxed{x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon} \longrightarrow x_T \approx \epsilon$$

$$0 \leq \alpha_i \leq 1 \qquad \bar{\alpha}_t \to 0$$

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$$

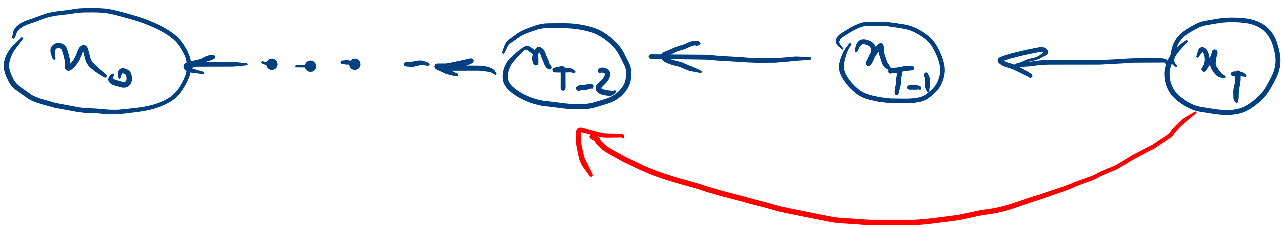
$$\boxed{x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta}{\sqrt{\bar{\alpha}_t}}}$$

$$\boxed{x_t = \sqrt{\alpha_{t-1}}\, x_{t-1} + \sqrt{1-\alpha_{t-1}}\, \epsilon}$$

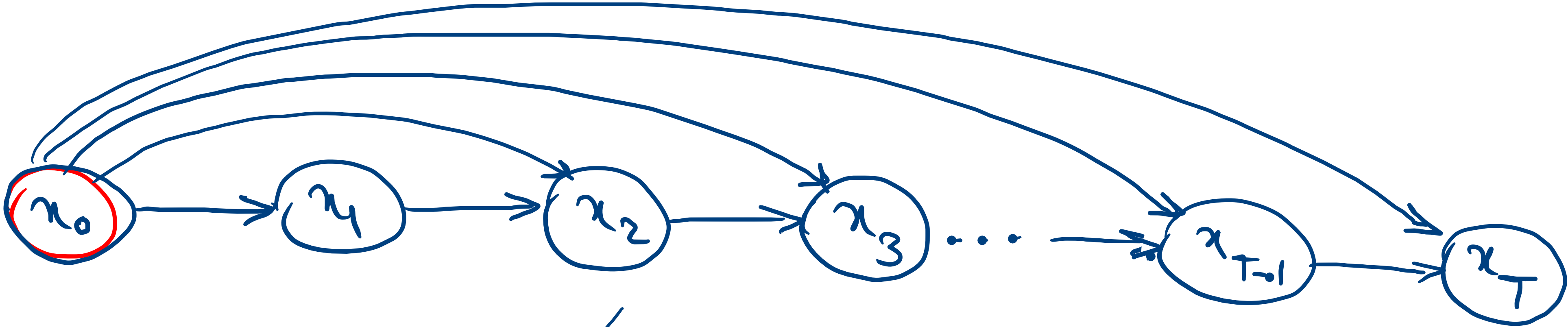
DDIM : Denoising Diffusion Implicit Model

50k 32x32 GAN 1 minute

DDPM 20 hours



x10 speed up



DDPM: $q(x_t \mid x_{t-1})$ $\quad$ $q(x_{t-1} \mid x_t)$

DDIM: $q_\sigma(x_t \mid x_0, x_{t-1})$

$$\sum_{t=1}^{T} KL\left(\underbrace{q(x_{t-1} \mid x_0, x_t)}_{\mathcal{N}(\tilde{\mu}, \tilde{\Sigma})} \,\|\, p_\theta(x_{t-1} \mid x_t) \right)$$

$$q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0) = \mathcal{N}\left(\underbrace{\sqrt{\alpha_{t-1}}\boldsymbol{x}_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\frac{\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}}}_{\tilde{\mu}}, \underbrace{\sigma_t^2\boldsymbol{I}}_{\tilde{\Sigma}}\right)$$

$\bar{\alpha}_t$

α_t

Bayes' rule

$$q_\sigma(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{x}_0) = \frac{q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\,q_\sigma(\boldsymbol{x}_t|\boldsymbol{x}_0)}{q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)},$$

DDPM:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1};\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0),\tilde{\beta}_t\mathbf{I}), \tag{6}$$

where $\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t$ and $\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$ (7)

$$\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}).$$

prediction of $\boldsymbol{x}_0$ given $\boldsymbol{x}_t$:

$$x_0 = f_\theta^{(t)}(\boldsymbol{x}_t) := (\boldsymbol{x}_t - \sqrt{1-\overline{\alpha}_t} \cdot \epsilon_\theta^{(t)}(\boldsymbol{x}_t))/\sqrt{\overline{\alpha}_t}$$

$\overline{\alpha}_t \to 0$

We can then define the generative process with a fixed prior $p_\theta(\boldsymbol{x}_T) = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and

$$p_\theta^{(t)}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \begin{cases} \mathcal{N}(f_\theta^{(1)}(\boldsymbol{x}_1), \sigma_1^2\boldsymbol{I}) & \text{if } t = 1 \\ q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, f_\theta^{(t)}(\boldsymbol{x}_t)) & \text{otherwise,} \end{cases}$$

DDPM:

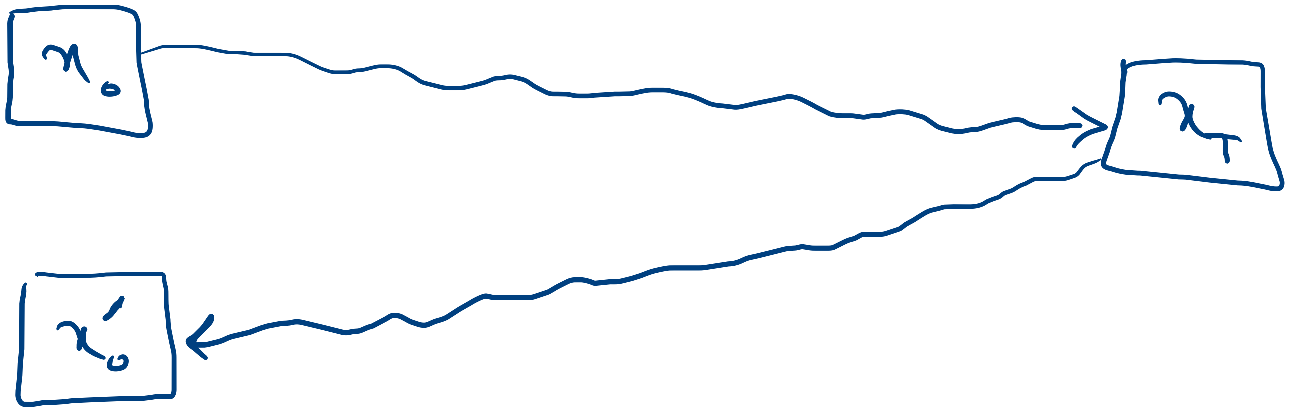
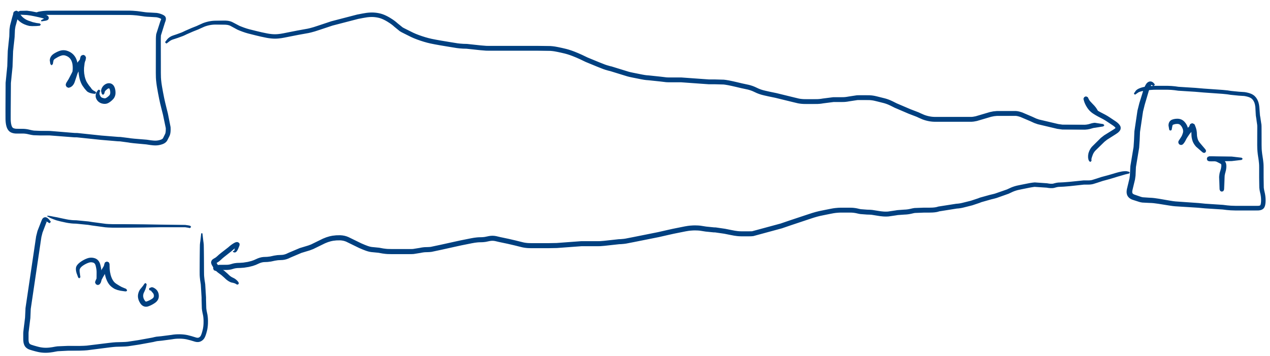


image colorization



DDIM: $\boxed{\sigma = 0}$

$$
\begin{tikzcd}
\boxed{x_0} \arrow[rr, squiggly, shift left] & & \boxed{x_T} \arrow[ll, squiggly, shift left] \arrow[dll] \\
\boxed{x_0'} & &
\end{tikzcd}
$$

$x_T \sim N(0, I)$

$\boxed{x_0' \simeq x_0}$

$x_{t-1} \leftarrow x_t$

$\epsilon_t \qquad x_0$

$$\boxed{q_\sigma(x_{t-1} \mid x_0, x_t)} = N(\mu, \overset{0}{\sigma_t^2 I})$$

$$x_{t-1} = \mu$$

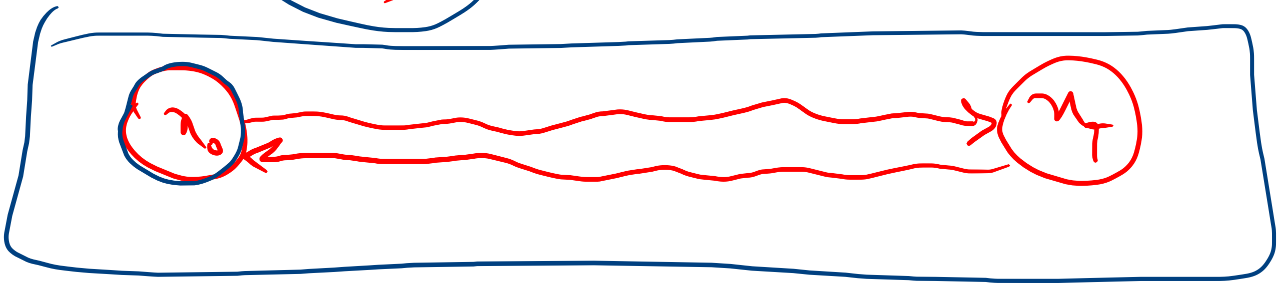
$$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{\boldsymbol{x}_t - \sqrt{1-\alpha_t}\,\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}}\right)}_{\text{“ predicted } \boldsymbol{x}_0\text{”}} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}_{\text{“direction pointing to } \boldsymbol{x}_t\text{”}} + \cancel{\underbrace{\sigma_t\epsilon_t}_{\text{random noise}}}$$

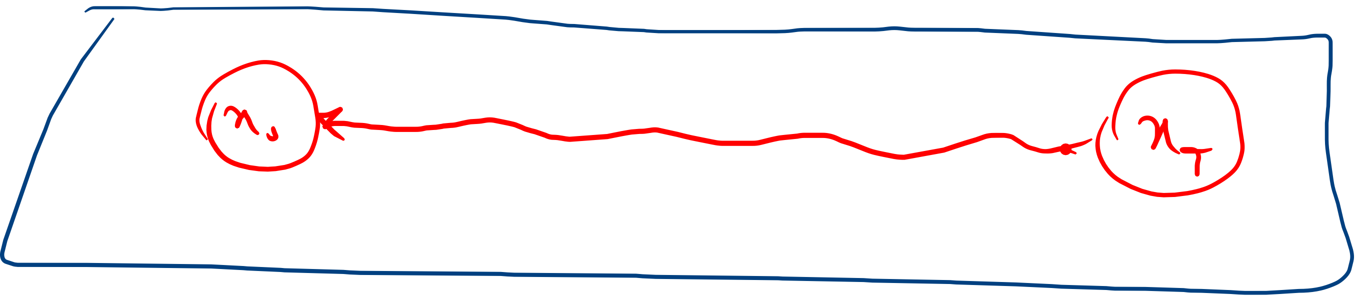
Sigma modulates the stochasticity of the process:

If $\sigma_t = \sqrt{(1-\alpha_{t-1})/(1-\alpha_t)}\sqrt{1-\alpha_t/\alpha_{t-1}}$: **Definition of the original DDPM**

If $\sigma_t = 0$: **The forward process becomes deterministic**

$\sigma = 0$





$\sigma =$

x_0

$T = 1000$

Table 1: CIFAR10 and CelebA image generation measured in FID. $\eta = 1.0$ and $\hat{\sigma}$ are cases of DDPM (although Ho et al. (2020) only considered $T = 1000$ steps, and $S < T$ can be seen as simulating DDPMs trained with S steps), and $\eta = 0.0$ indicates DDIM.

	S	CIFAR10 (32 × 32) 10	20	50	100	1000	CelebA (64 × 64) 10	20	50	100	1000
η	0.0	**13.36**	**6.84**	**4.67**	**4.16**	4.04	**17.33**	**13.73**	**9.17**	**6.53**	3.51
	0.2	14.04	7.11	4.77	4.25	4.09	17.66	14.11	9.51	6.79	3.64
	0.5	16.66	8.35	5.25	4.46	4.29	19.86	16.06	11.01	8.09	4.28
	1.0	41.07	18.36	8.01	5.78	4.73	33.12	26.03	18.48	13.93	5.98
	$\hat{\sigma}$	367.43	133.37	32.72	9.99	**3.17**	299.71	183.83	71.71	45.20	**3.26**

$0 \leq \eta \leq 1$

$\eta = 0$ DDIM

$\eta = 1$ DDPM

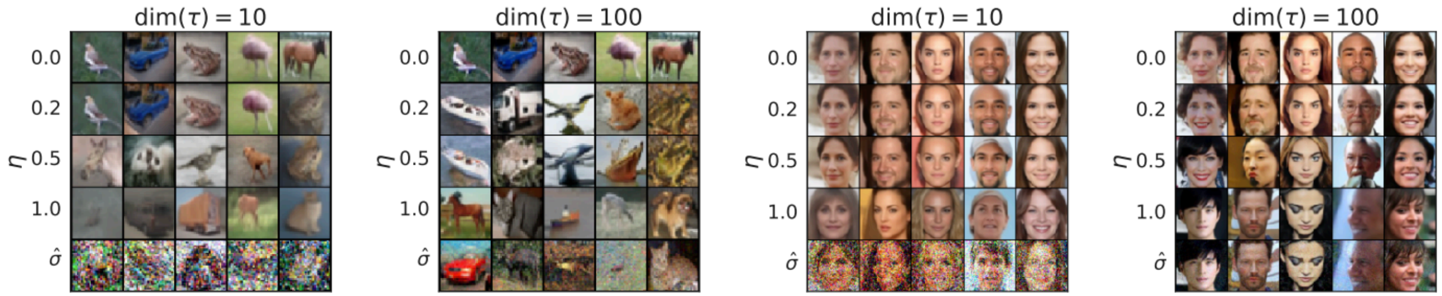
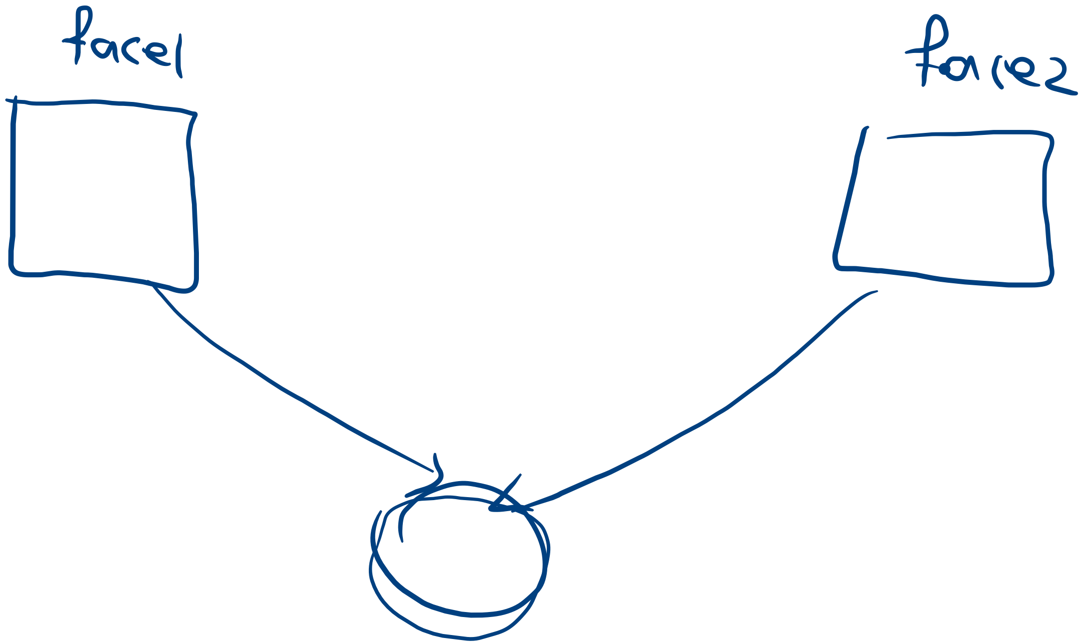


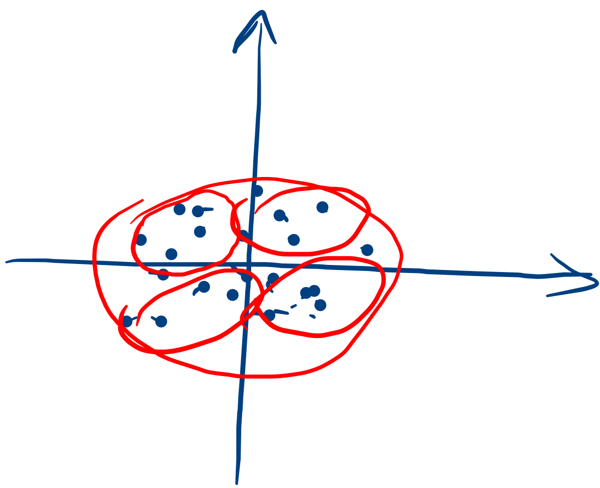
Figure 3: CIFAR10 and CelebA samples with $\dim(\tau) = 10$ and $\dim(\tau) = 100$.

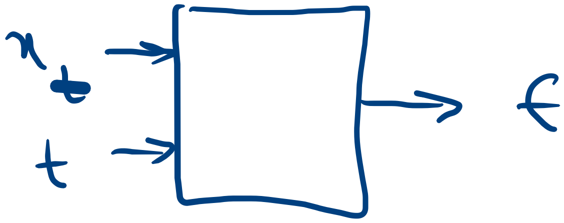
Comparison

- Sample Quality and Efficiency
- Sample Consistency
- Interpolation in deterministic generative process
- Reconstruction from latent space


x_0
x_T
x_0'


face1
face2





x
t
€