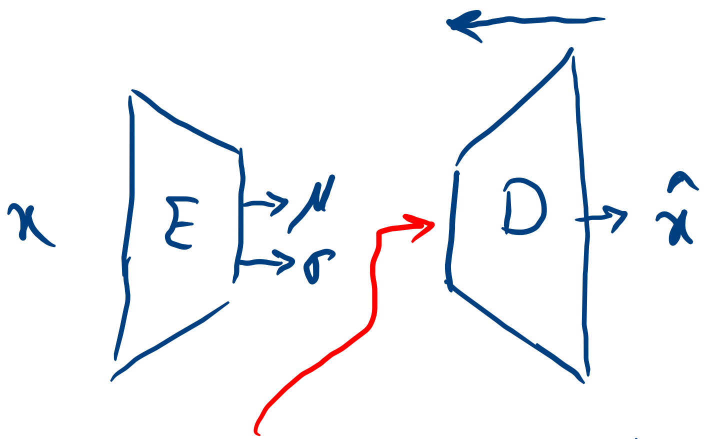
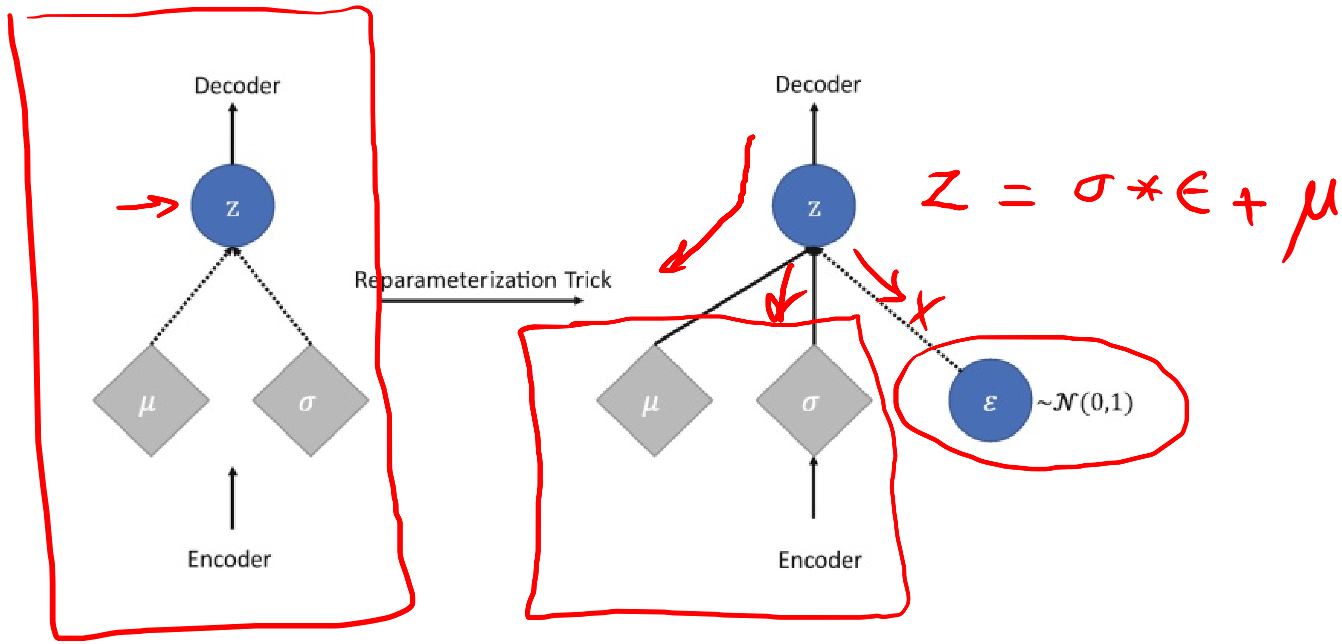
Reparametrization trick

VAE :



$$Z \sim N(\mu, \sigma^2 I)$$

$$\|x - \hat{x}\|_2^2 + KL$$



Decoder
z
μ
σ
Encoder
Reparameterization Trick
Decoder
z
$z = \sigma * \epsilon + \mu$
μ
σ
ε
$\sim\mathcal{N}(0,1)$
Encoder

DDPM

Denoising Diffusion Probabilistic Models

Jonathan Ho
UC Berkeley
jonathanho@berkeley.edu

Ajay Jain
UC Berkeley
ajayj@berkeley.edu

Pieter Abbeel
UC Berkeley
pabbeel@cs.berkeley.edu

$$x_t = a\, x_{t-1} + b\epsilon \qquad \epsilon \sim \mathcal{N}(0, I)$$

Denoising diffusion models

- **Forward / noising process**
  - Sample data $p(\mathbf{x}_0)$ → turn to noise

$p_0(\mathbf{x}_0)$

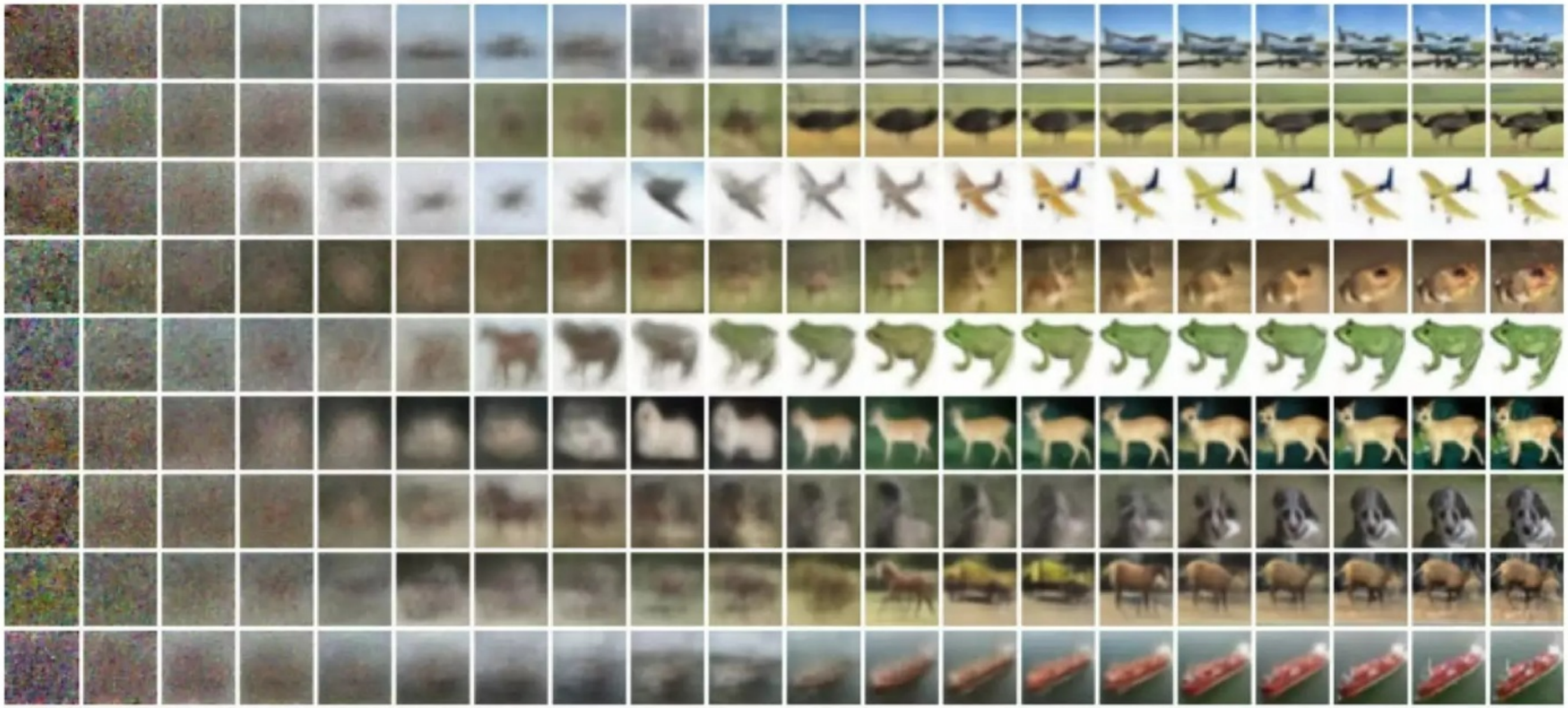
$p_T(\mathbf{x}_T) \sim \mathcal{N}(0, I)$

Clean sample $\mathbf{x}_0$ $\mathbf{x}_1$ … $\mathbf{x}_{T-1}$ $\mathbf{x}_T$ ***Pure noise***

- **Reverse / denoising process**
  - Sample noise $p_T(\mathbf{x}_T)$ → turn into data

Experiments

◆ Progressive Generation



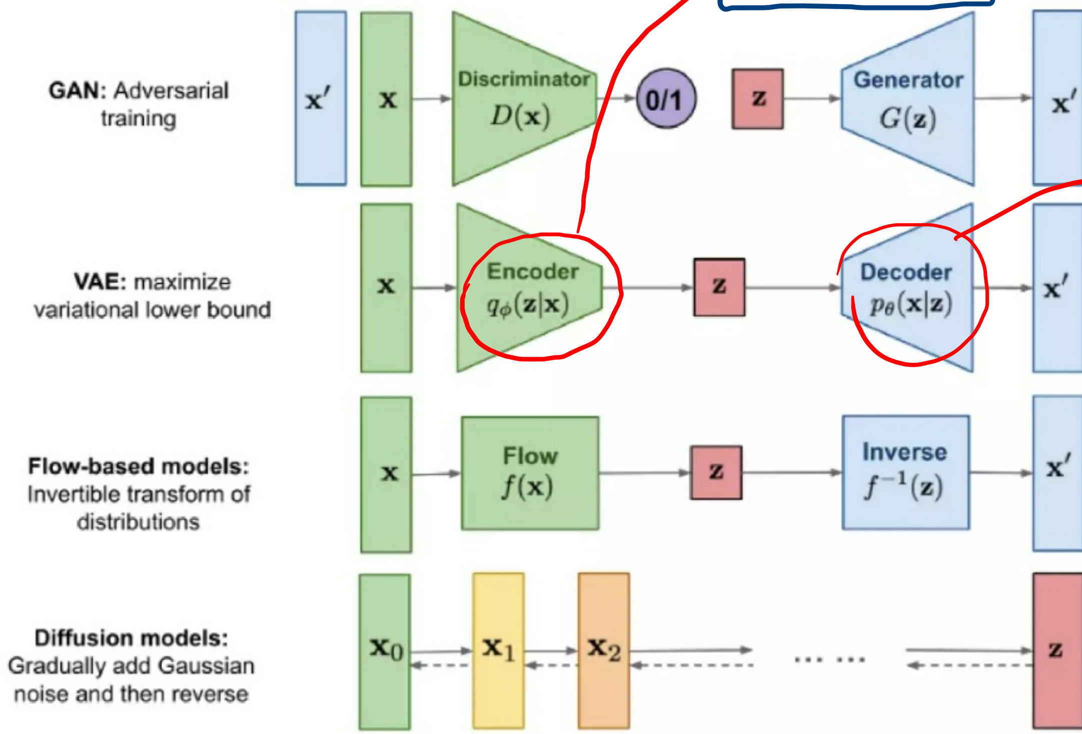
x_T

x_0

Generative Models

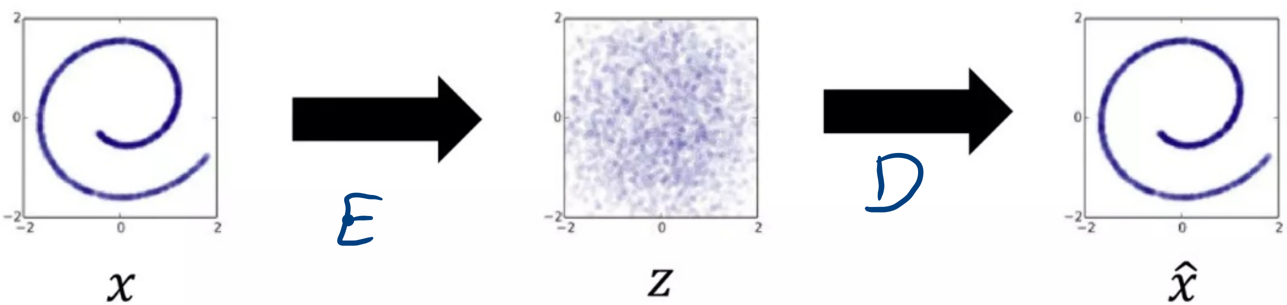
◆ Generative Models

$\boxed{q_\phi(z|x)}$ $\boxed{p_\theta(x|z)}$



Generative Models

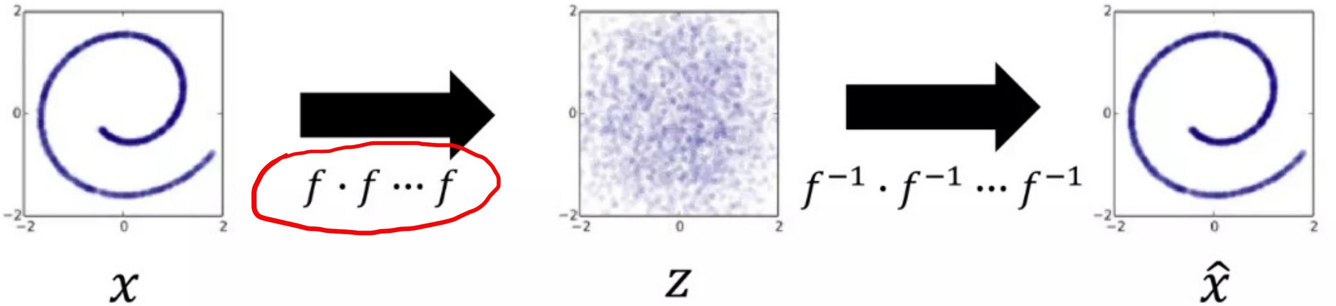
- Variational Auto Encoders

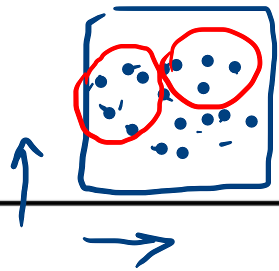


$\mathcal{N}(0, I)$

Generative Models

◆ Normalizing Flow Models

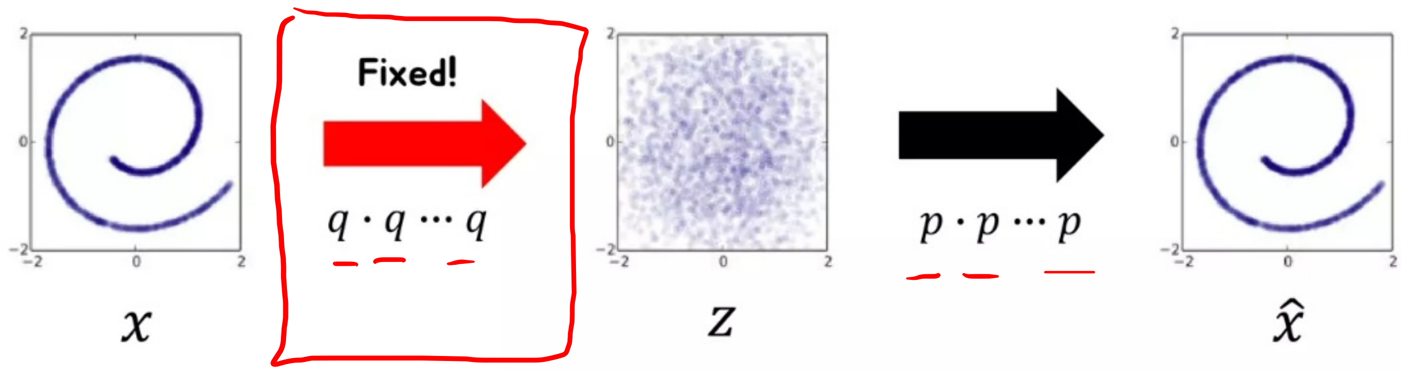




Generative Models

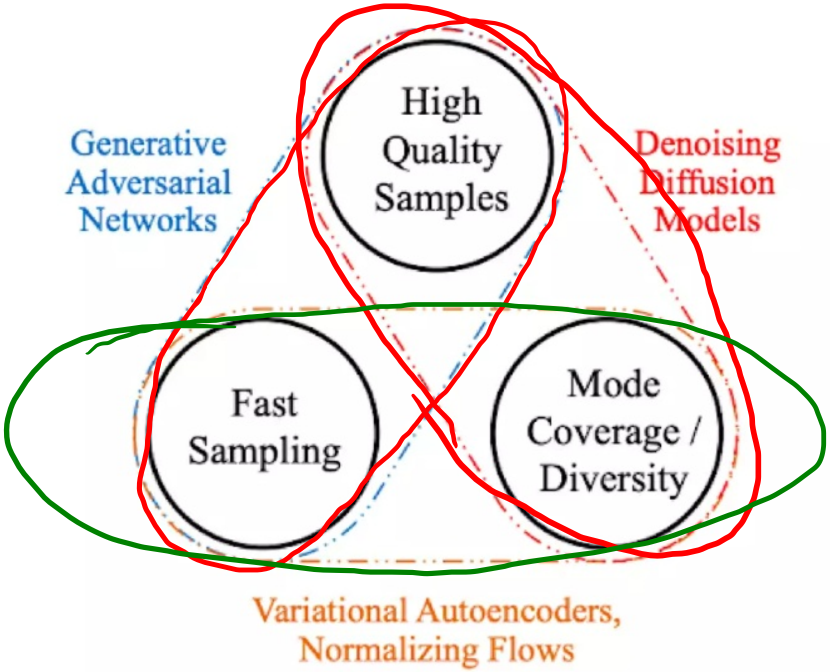
- Diffusion Models

DDPM



Generative Models

◆ Generative Models



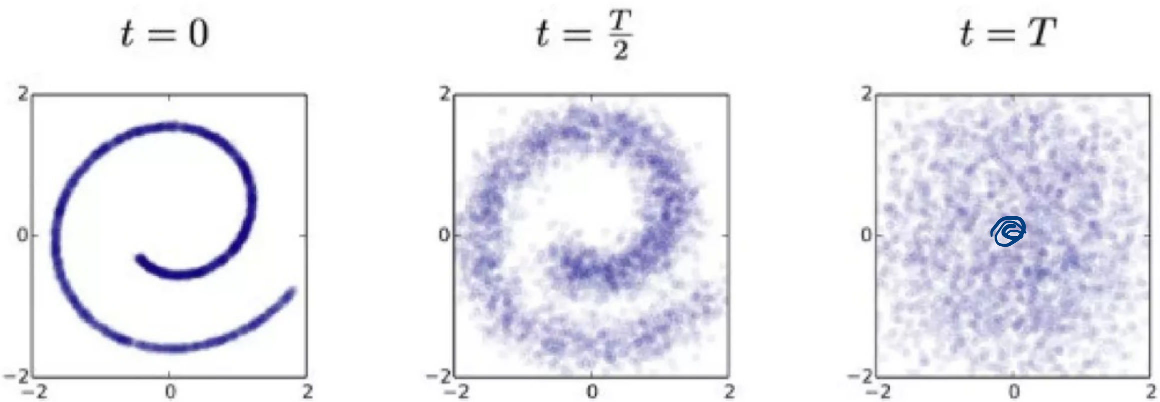
Mode collapse

Diffusion Probabilistic Models

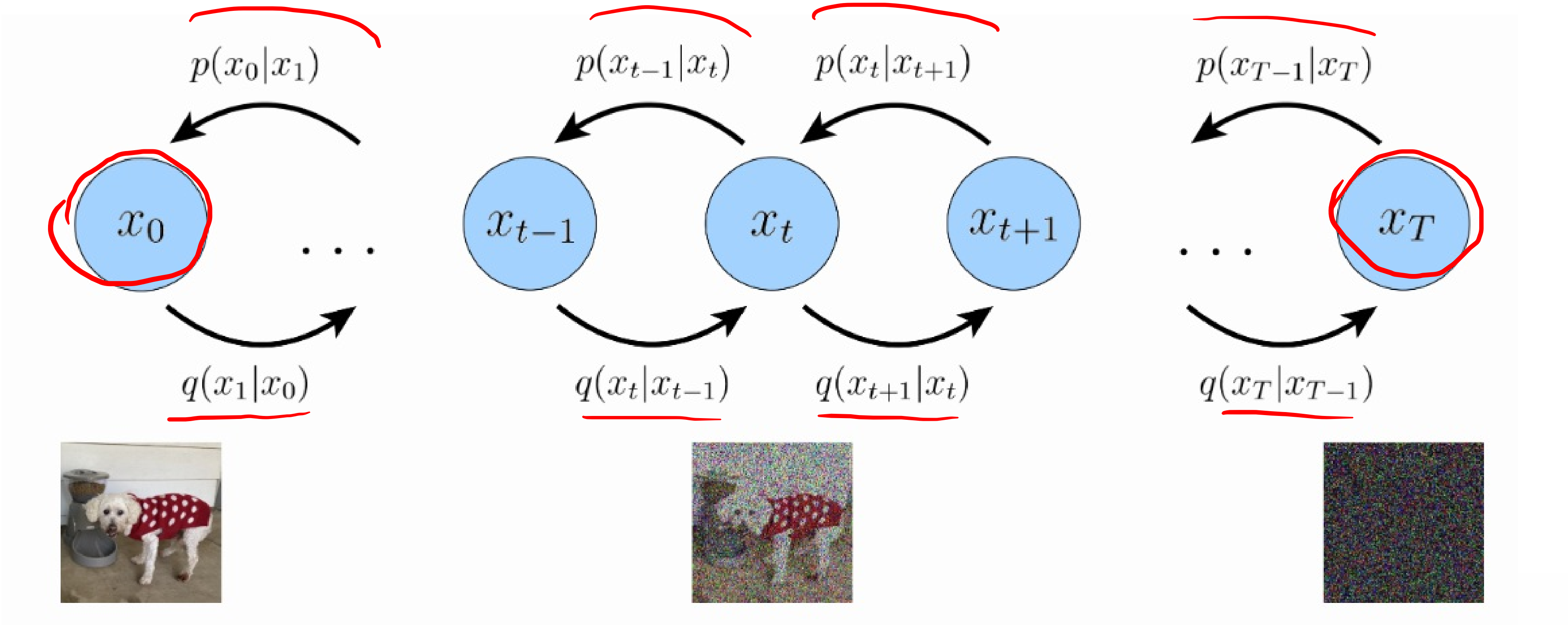
◆ **Diffusion?**

Backward

$T = 1000$



$x_0 \longrightarrow x_{500} \longrightarrow x_{1000}$



$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) \coloneqq \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) \coloneqq \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

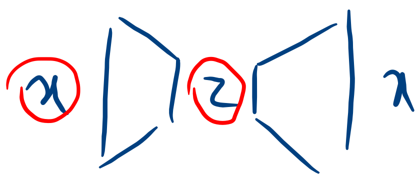
$$p_\theta(\mathbf{x}_{0:T}) \coloneqq p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \coloneqq \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$\underline{\underline{x}}$

observed

$\underline{\underline{z}}$ $\to$ latent/hidden variables

latent space



VAE		Diff
x	$\to$	x_0
z	$\to$	$x_1, x_2, \ldots, x_T$

$$\underbrace{x_0} \to \underbrace{x_1 \to x_2 \to \cdots \to x_T}_{\substack{\text{hidden} \\ \text{Latent}}}$$

$$\log p_\theta(x) \geq -KL\left(q_\phi(z|x) \,\|\, p_\theta(x,z)\right)$$

$$z \longrightarrow x_1, \ldots, x_T$$

$$x \longrightarrow x_0$$

$$\max_{\theta,\phi} \; -KL\left(q_\phi(z|x) \,\|\, p_\theta(x,z)\right)$$

$$\boxed{\max_{\theta} \; -KL\left(q(x_1,\ldots,x_T|x_0) \,\|\, p_\theta(x_0,x_1,\ldots,x_T)\right)}$$

$$x_t \longrightarrow \boxed{\begin{matrix} NN \\ \theta \end{matrix}} \longrightarrow x_{t-1}$$

$$q(x_1, \dots, x_T \mid x_0) = q(x_1 \mid x_0)\, q(x_2 \mid x_1) \cdots q(x_T \mid x_{T-1}) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

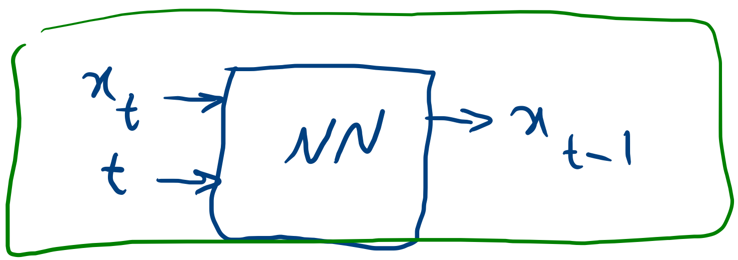
$$x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_T$$

$$p_\theta(x_0, x_1, \dots, x_T) = \underline{p(x_T)}\, \underline{p_\theta(x_{T-1} \mid x_T)}\, p_\theta(x_{T-2} \mid x_{T-1}) \cdots p_\theta(x_0 \mid x_1)$$

$$x_T \rightarrow x_{T-1} \rightarrow x_{T-2} \rightarrow \cdots \rightarrow x_1 \rightarrow x_0$$

$$= p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)$$

$T = 1000$



$x_T \sim N(0, I)$

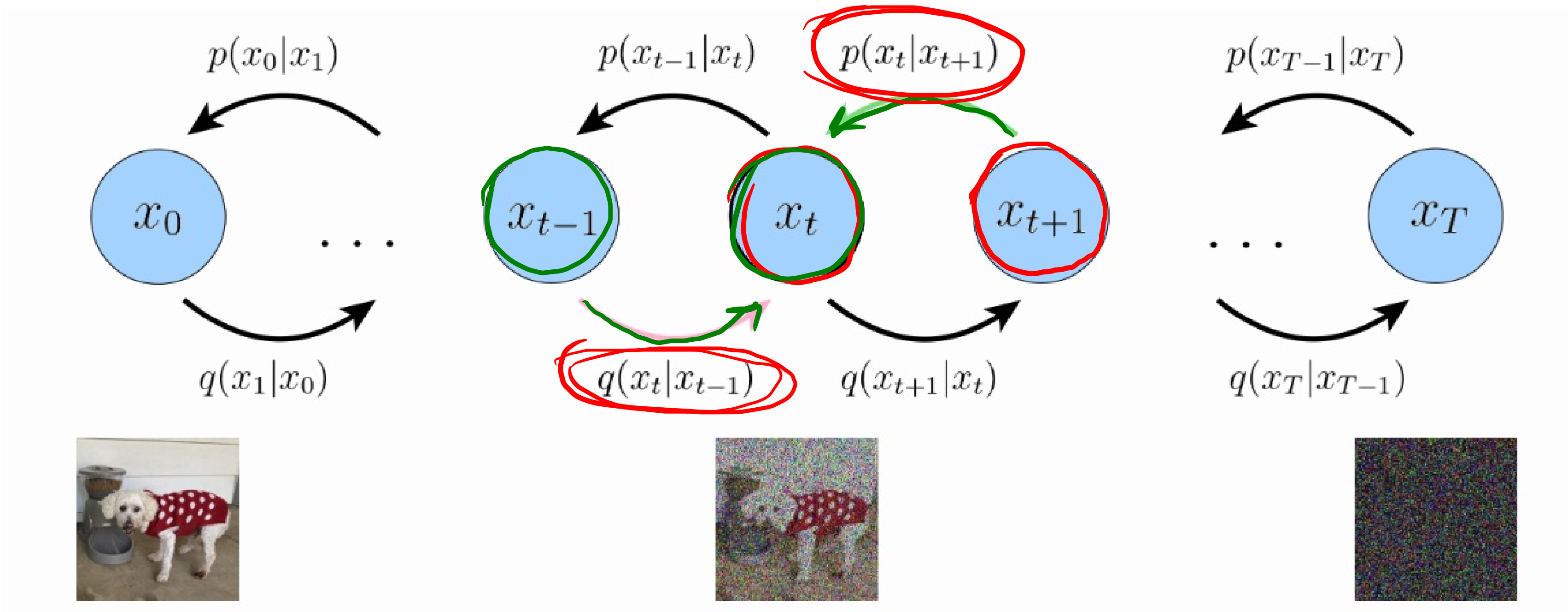
$$-KL\left(q(x_1,\dots,x_T|x_0) \,\|\, P_\theta(x_0,x_1,\dots,x_T)\right)$$

$$= -E_q\left[\log \frac{q(x_1,\dots,x_T|x_0)}{P_\theta(x_0,x_1,\dots,x_T)}\right] \doteq -E\left[\log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{P(x_T)\prod_{t=1}^{T} P_\theta(x_{t-1}|x_t)}\right]$$

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

$$\max_{\theta} L(\theta)$$

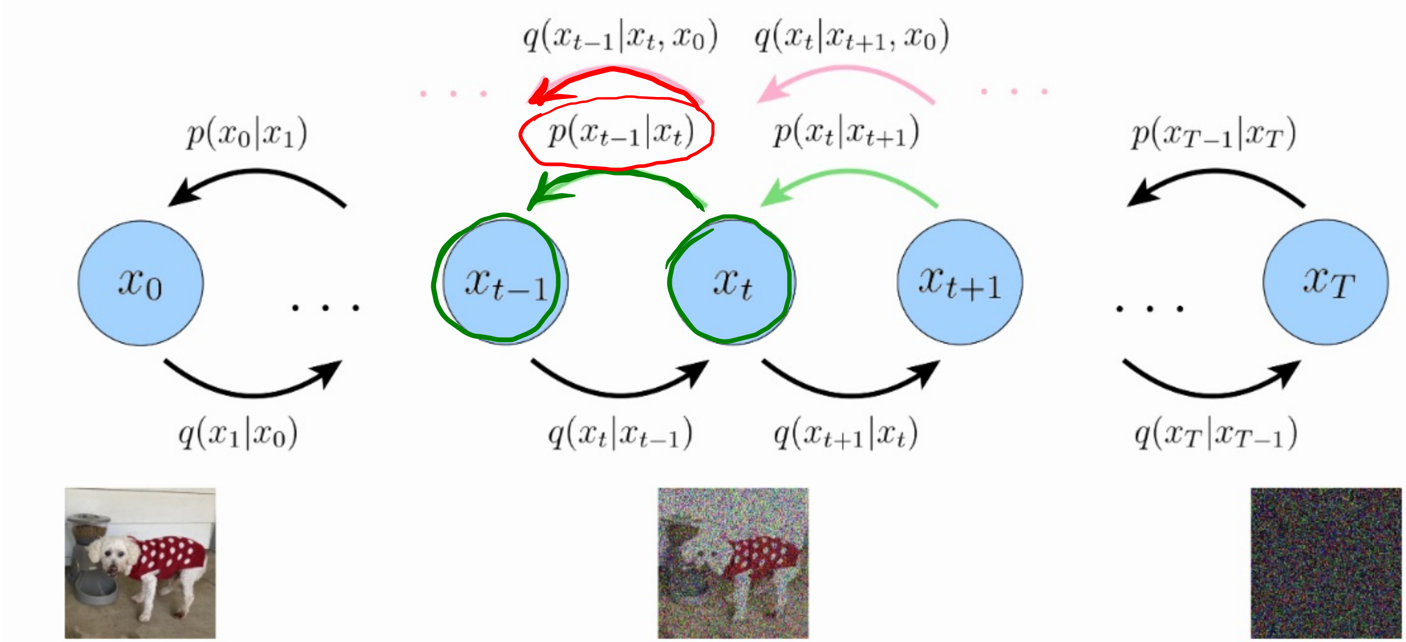
$$
\begin{aligned}
\log p(\boldsymbol{x}) &\geq \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)}\right] \\
&= \underbrace{\mathbb{E}_{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\right]}_{\text{reconstruction term}} - \underbrace{\mathbb{E}_{q(\boldsymbol{x}_{T-1}|\boldsymbol{x}_0)}\left[\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1}) \,||\, p(\boldsymbol{x}_T))\right]}_{\text{prior matching term}} \\
&\quad - \sum_{t=1}^{T-1} \underbrace{\mathbb{E}_{q(\boldsymbol{x}_{t-1},\boldsymbol{x}_{t+1}|\boldsymbol{x}_0)}\left[\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \,||\, p_\theta(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1}))\right]}_{\text{consistency term}}
\end{aligned}
$$



$$
\begin{aligned}
\log p(\boldsymbol{x}) &= \log \int p(\boldsymbol{x}_{0:T}) d\boldsymbol{x}_{1:T} \\
&= \log \int \frac{p(\boldsymbol{x}_{0:T}) q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)}{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} d\boldsymbol{x}_{1:T} \\
&= \log \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\frac{p(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \right] \\
&\geq \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{\prod_{t=1}^{T} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T) p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \prod_{t=2}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1}) \prod_{t=1}^{T-1} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T) p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \prod_{t=1}^{T-1} p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1}) \prod_{t=1}^{T-1} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T) p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1})} \right] + \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \prod_{t=1}^{T-1} \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})}{q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \right] + \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T)}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1})} \right] + \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\sum_{t=1}^{T-1} \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})}{q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \right] + \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T)}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1})} \right] + \sum_{t=1}^{T-1} \mathbb{E}_{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)} \left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})}{q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \right] + \mathbb{E}_{q(\boldsymbol{x}_{T-1}, \boldsymbol{x}_T \mid \boldsymbol{x}_0)} \left[\log \frac{p(\boldsymbol{x}_T)}{q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1})} \right] + \sum_{t=1}^{T-1} \mathbb{E}_{q(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t, \boldsymbol{x}_{t+1} \mid \boldsymbol{x}_0)} \left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})}{q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right] \\
&= \underbrace{\mathbb{E}_{q(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \right]}_{\text{reconstruction term}} - \underbrace{\mathbb{E}_{q(\boldsymbol{x}_{T-1} \mid \boldsymbol{x}_0)} \left[\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_T \mid \boldsymbol{x}_{T-1}) \parallel p(\boldsymbol{x}_T)) \right]}_{\text{prior matching term}} \\
&\qquad - \sum_{t=1}^{T-1} \underbrace{\mathbb{E}_{q(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t+1} \mid \boldsymbol{x}_0)} \left[\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \parallel p_{\boldsymbol{\theta}}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t+1})) \right]}_{\text{consistency term}}
\end{aligned}
$$

$$
\begin{aligned}
\log p(\boldsymbol{x}) &\geq \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)}\right] \\
&= \underbrace{\mathbb{E}_{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\right]}_{\text{reconstruction term}} - \underbrace{\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_T \mid \boldsymbol{x}_0) \,||\, p(\boldsymbol{x}_T))}_{\text{prior matching term}} \\
&\quad - \sum_{t=2}^{T} \underbrace{\mathbb{E}_{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\mathcal{D}_{\mathrm{KL}}(q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) \,||\, p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t))\right]}_{\text{denoising matching term}}
\end{aligned}
$$

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



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

$=$

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

$q(x_{t-1} | x_0, x_t)$

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

$$q(x_{t-1}|x_t) = \frac{q(x_t|x_{t-1})\, q(x_{t-1})}{q(x_t)}$$

$$\boxed{x_t} = a x_{t-1} + b\epsilon \longrightarrow x_{t-1} = \frac{1}{a} x_t - \frac{b}{a}\epsilon$$

$$q(x_t|x_{t-1}) = N(a x_{t-1}, b^2 I)$$

نرمال نیست $\leftarrow q(x_{t-1}|x_t)$

$\rightarrow q(x_{t-1}|x_t) \rightarrow$ نرمال نیست

$$\begin{tikzcd} x_{t-1} \arrow[r, bend left, "q"] & x_t \arrow[l, bend left, "p_\theta"] \end{tikzcd}$$

$$p_\theta = \cancel{q(x_{t-1} | x_t)}$$

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

$$KL\Big(\underbrace{q(x_{t-1}\mid x_0, x_t)}_{\boxed{\mathcal{N}(\tilde{\mu}, \tilde{\Sigma})}} \,\|\, \underbrace{p_\theta(x_{t-1}\mid x_t)}_{\substack{\text{Normal} \\ \boxed{\mathcal{N}(\mu_\theta, \Sigma_\theta)}}}\Big)$$

Diffusion Probabilistic Models

- Diffusion Loss

$$\boxed{D_{KL}\left(q_\phi(z|x)||p_\theta(z)\right)} + \boxed{\mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]}$$

Regularization　　Reconstruction

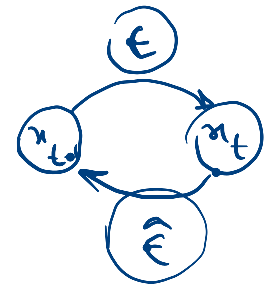
Learning β_t

$$\mathbb{E}_q\Bigg[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p(\mathbf{x}_T))}_{L_T} + \sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\Bigg]$$

Regularization　　Reconstruction

const

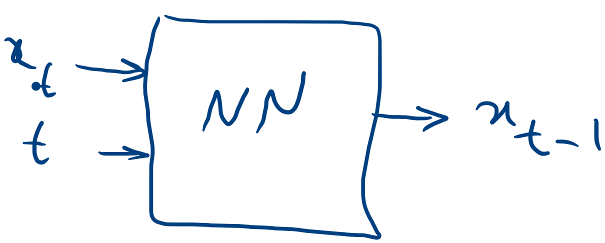
$$\mathbb{E}_q\Bigg[\overset{1}{\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\,\|\,p(\mathbf{x}_T))}_{L_T}} + \boxed{\overset{2}{\sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\,\|\,p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}}}} \overset{3}{\underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}}\Bigg]$$

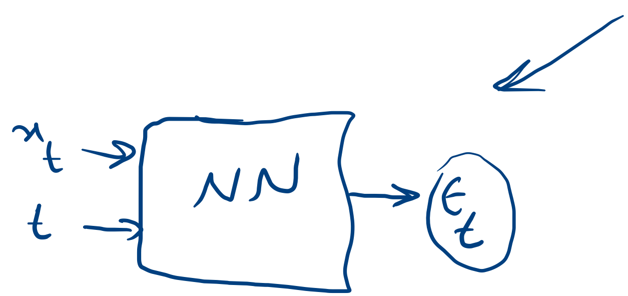


$$\boxed{L_{t-1} = \mathbb{E}_q\left[\frac{1}{2\sigma_t^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\right] + C}$$

$$\boxed{\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]}$$

$$\boxed{L_{\mathrm{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]}$$

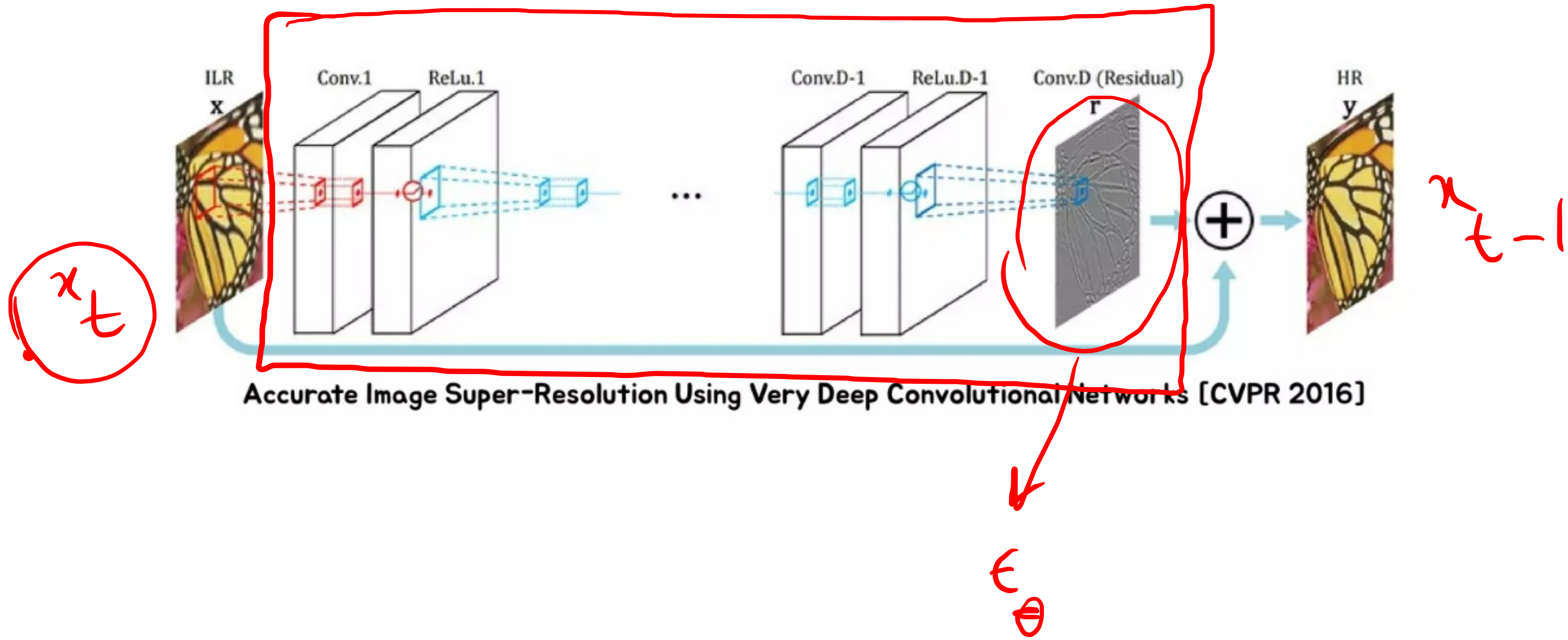




$$\boxed{x_{t-1} = x_t - \epsilon_t}$$

DDPM

◆ Residual Estimation



Accurate Image Super-Resolution Using Very Deep Convolutional Networks [CVPR 2016]

DDPM

◆ Loss Simplification

$$\mathbb{E}_q\left[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\right]$$

$$\Downarrow$$

$$L_{\mathrm{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$