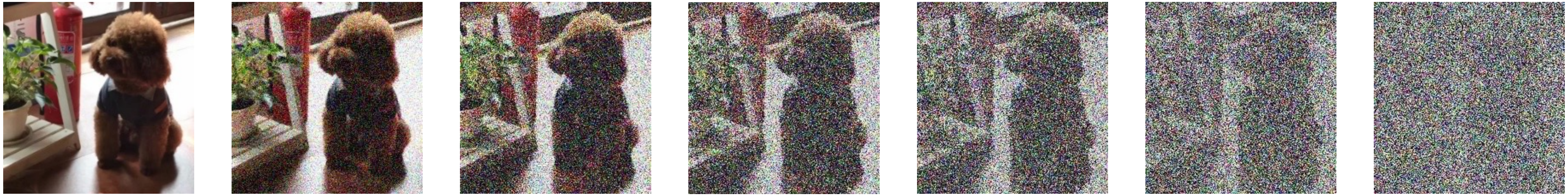
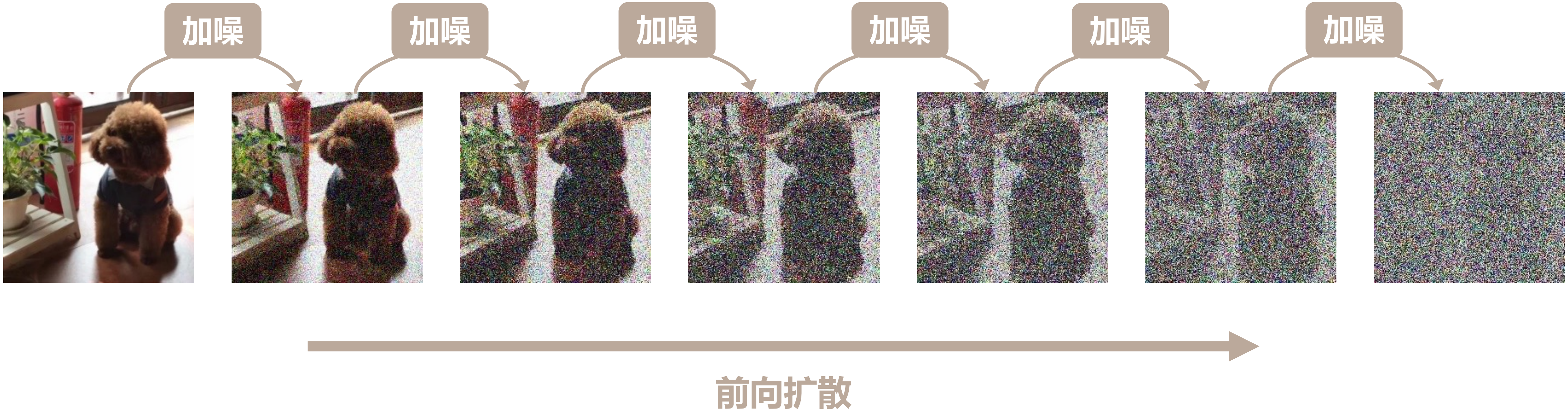
扩散模型

前向扩散

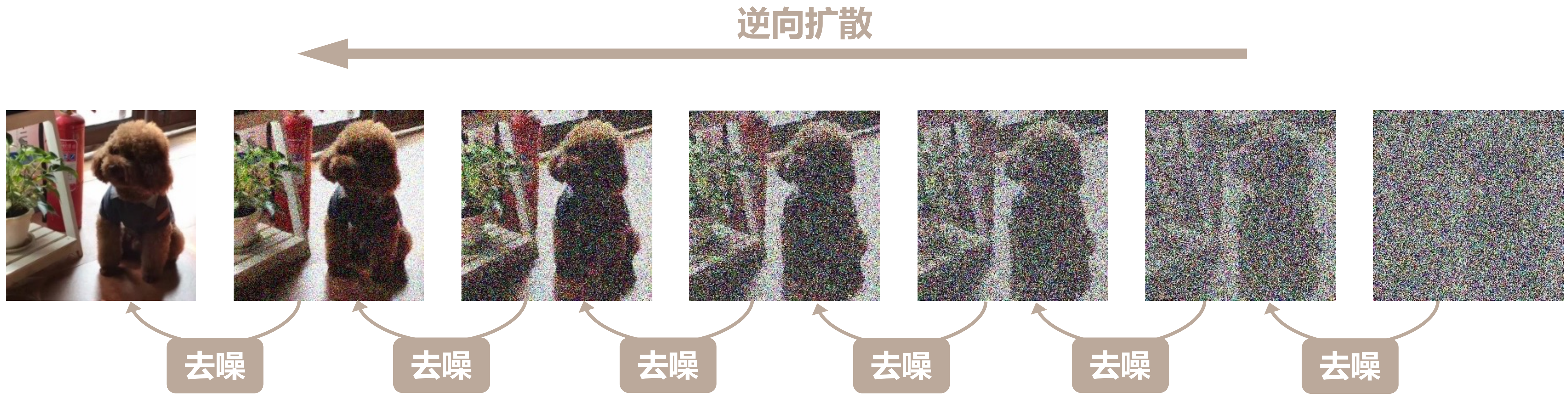


逆向扩散

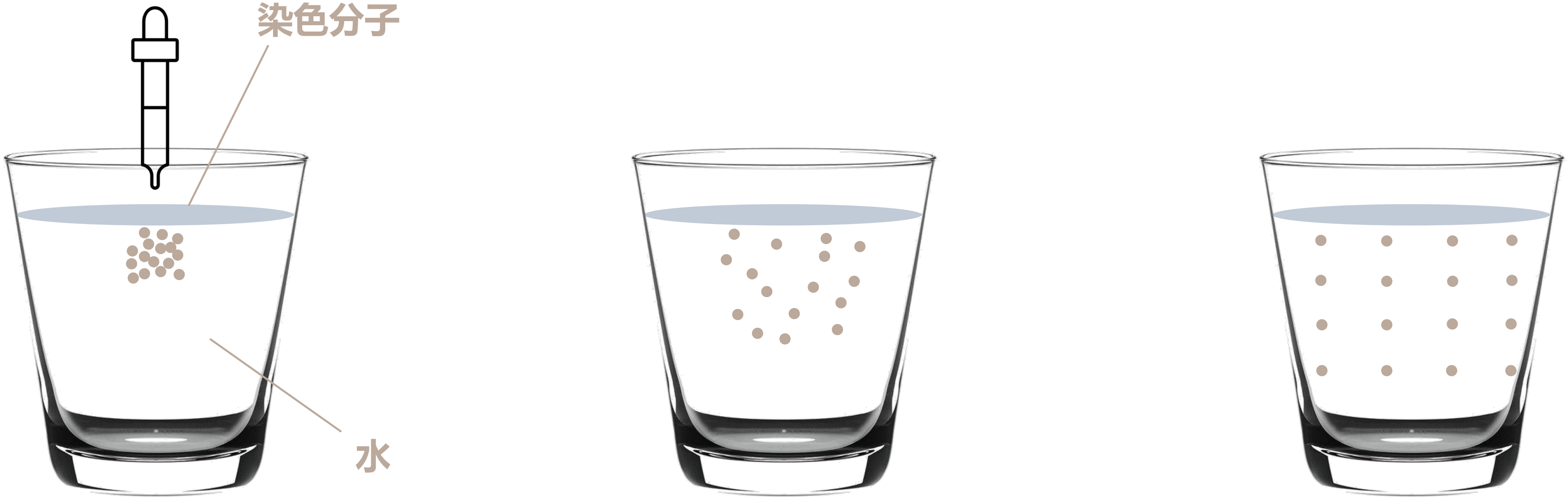
扩散模型



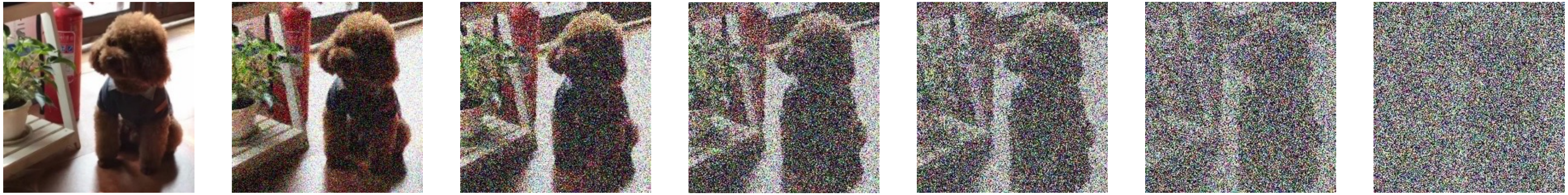
扩散模型



扩散现象



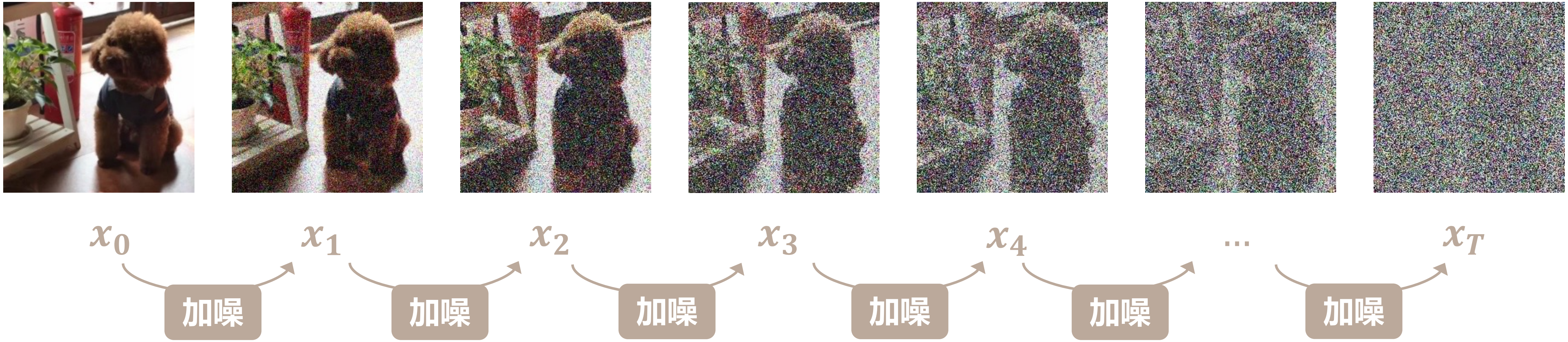
前向扩散



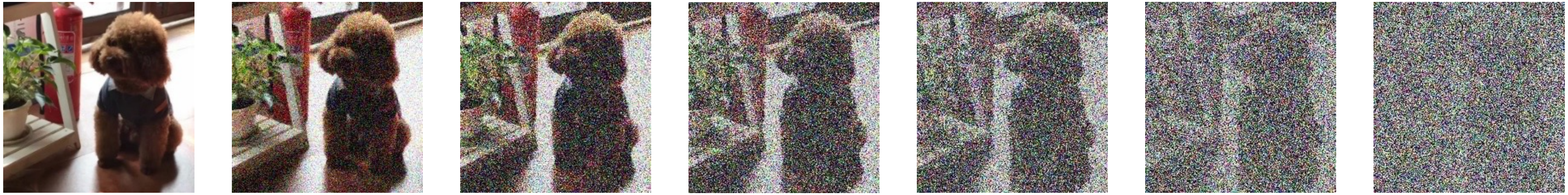
x_t

x :图片　　t:当前迭代步数

前向扩散



前向扩散



x_0 $\xrightarrow{q(x_1|x_0)}$ x_1 $\xrightarrow{q(x_2|x_1)}$ x_2 $\xrightarrow{q(x_3|x_2)}$ x_3 $\xrightarrow{q(x_4|x_3)}$ x_4 $\xrightarrow{q(x_t|x_{t-1})}$ … $\xrightarrow{q(x_T|x_{T-1})}$ x_T

输入：x_{t-1}（加噪前的图像） → $q(x_t|x_{t-1})$ → 输出：x_t（加噪后的图像）

前向扩散

$$\{\beta_t \in (0,1)\}_{t=1}^{T}$$

DDPM中$\beta_1 = 10^{-4}, \beta_T = 0.02$

方差

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

均值

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

前向扩散

重参数化:

$$z \sim \mathcal{N}\left(z, \mu, \sigma^2 \mathrm{I}\right) \longrightarrow z = \mu + \sigma \cdot \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, I)$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned} x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\ &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon \end{aligned}$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon$$

$$= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon$$

$$\boxed{x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon}$$

代入

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned} x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\ &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon \end{aligned}$$

$$\boxed{x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon}$$

代入

$$\begin{aligned} &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon\right) + \sqrt{1-\alpha_t}\epsilon \\ &= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon + \sqrt{1-\alpha_t}\epsilon \end{aligned}$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned} x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\ &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon \end{aligned}$$

$$\boxed{x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon}$$

代入

$$\begin{aligned} &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon\right) + \sqrt{1-\alpha_t}\epsilon \\ &= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \boxed{\sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon + \sqrt{1-\alpha_t}\epsilon} \end{aligned}$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon$$

$$= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon$$

$$\boxed{x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon}$$

代入

$$= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon\right) + \sqrt{1-\alpha_t}\epsilon$$

$$= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon + \sqrt{1-\alpha_t}\epsilon$$

$$\boxed{\begin{gathered}\mathcal{N}(0, \sigma_1^2 I) + \mathcal{N}(0, \sigma_2^2 I) \\ \sim\mathcal{N}\left(0, (\sigma_1^2+\sigma_2^2)I\right)\end{gathered}}$$

$$= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})) + \mathcal{N}(0, 1-\alpha_t)$$

$$= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$$

$$= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\epsilon$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned} x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\ &= \sqrt{\alpha_t \alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\epsilon \\ &= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}x_{t-3} + \sqrt{1-\alpha_t \alpha_{t-1} \alpha_{t-2}}\epsilon \end{aligned}$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned}
x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\
&= \sqrt{\alpha_t \alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\epsilon \\
&= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}x_{t-3} + \sqrt{1-\alpha_t \alpha_{t-1} \alpha_{t-2}}\epsilon \\
&= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2} \ldots \alpha_0}x_0 + \sqrt{1-\alpha_t \alpha_{t-1} \alpha_{t-2} \ldots \alpha_0}\epsilon \\
&= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon
\end{aligned}$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned}
x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}\dots\alpha_1}x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}\dots\alpha_1}\epsilon \\
&= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon
\end{aligned}$$

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I\right)$$

前向扩散

定义$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

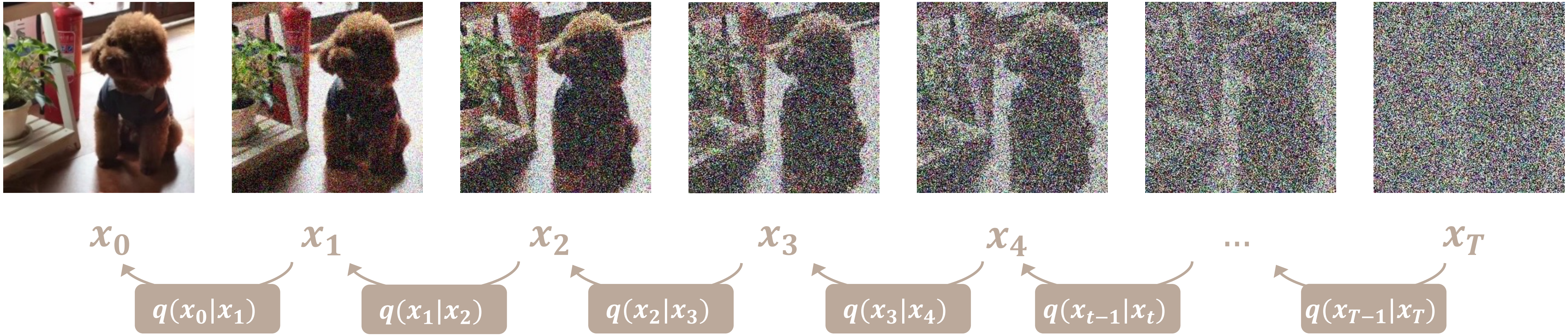
$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right)$$

$$\begin{aligned}
x_t &= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}\ldots\alpha_1}x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}\ldots\alpha_1}\epsilon \\
&= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon
\end{aligned}$$

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I\right)$$

$$\lim_{t\to\infty} q(x_t|x_0) = \mathcal{N}(0, I) \qquad \lim_{t\to\infty} q(x_t) = \mathcal{N}(0, I)$$

逆向扩散

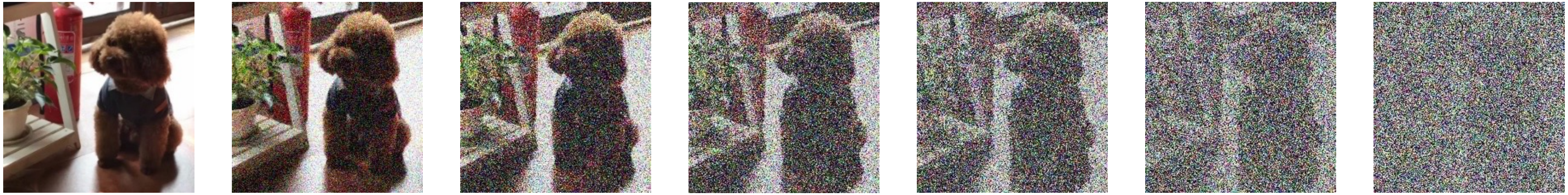


输入：x_t
（带有噪声的图像）

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

输出：x_{t-1}
（去噪后的图像）

逆向扩散



x_0 x_1 x_2 x_3 x_4 … x_T

$p_\theta(x_0|x_1)$ $p_\theta(x_1|x_2)$ $p_\theta(x_2|x_3)$ $p_\theta(x_3|x_4)$ $p_\theta(x_{t-1}|x_t)$ $p_\theta(x_{T-1}|x_T)$

输入：x_t
（带有噪声的图像）

$p_\theta(x_{t-1}|x_t)$
（神经网络）

输出：x_{t-1}
（去噪后的图像）

逆向扩散

可训练的神经网络
（主流为UNet+Attention）

固定值

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$p_\theta(x_0) = \int p_\theta(x_{0:T}) dx_{1:T}$$

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

逆向扩散

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}, \widetilde{\mu}(x_t, x_0), \widetilde{\beta}_t I\right)$$

逆向扩散

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

逆向扩散

正态分布的概率密度函数：$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2})$

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

$q(x_t|x_{t-1}, x_0) = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon \sim \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, 1-\alpha_t) \longrightarrow \frac{1}{\sqrt{2\pi(1-\alpha_t)}} exp(-\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{2(1-\alpha_t)})$

$q(x_{t-1}|x_0) = \sqrt{\bar{\alpha}_{t-1}}\,x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\epsilon \sim \mathcal{N}(\sqrt{\bar{\alpha}_{t-1}}\,x_0, 1-\bar{\alpha}_{t-1}) \longrightarrow \frac{1}{\sqrt{2\pi(1-\bar{\alpha}_{t-1})}} exp(-\frac{(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}x_0)^2}{2(1-\bar{\alpha}_{t-1})})$

$q(x_t|x_0) = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, 1-\bar{\alpha}_t) \longrightarrow \frac{1}{\sqrt{2\pi(1-\bar{\alpha}_t)}} exp(-\frac{(x_t-\sqrt{\bar{\alpha}_t}x_0)^2}{2(1-\bar{\alpha}_t)})$

逆向扩散

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

$$\propto exp(-\frac{1}{2}(\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_t}+\frac{\left((x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}\,x_0\right)^2}{1-\bar{\alpha}_{t-1}}-\frac{(x_t-\sqrt{\bar{\alpha}_t}\,x_0)^2}{1-\bar{\alpha}_t}))$$

逆向扩散

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

$$\propto exp(-\frac{1}{2}(\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_t}+\frac{\left((x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}\,x_0\right)^2}{1-\bar{\alpha}_{t-1}}-\frac{(x_t-\sqrt{\bar{\alpha}_t}\,x_0)^2}{1-\bar{\alpha}_t}))$$

$$=exp(-\frac{1}{2}(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}x_0}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}+C(x_t,x_0)))$$

逆向扩散

改写为

$$\boxed{exp(-\frac{(x-\mu)^2}{2\sigma^2})}$$

$$exp(-\frac{1}{2}((\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)x_{t-1}+C(x_t,x_0)))\mathrm{v}$$

逆向扩散

改写为

$$\boxed{exp(-\frac{(x-\mu)^2}{2\sigma^2})}$$

$$exp(-\frac{1}{2}((\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}})x_{t-1}^2-(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0)x_{t-1}+C(x_t,x_0)))\text{v}$$

$$\frac{1}{\sigma^2}=\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}, \qquad \tilde{\beta}_t=\sigma^2=\frac{1}{\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}}=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$$\tilde{\mu}_t=\frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}x_0$$

逆向扩散

改写为

$$exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$exp(-\frac{1}{2}((\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}})x_{t-1}^2-(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0)x_{t-1}+C(x_t,x_0)))$$v

$$\frac{1}{\sigma^2}=\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}},\qquad \tilde{\beta}_t=\sigma^2=\frac{1}{\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}}=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$$\tilde{\mu}_t=\frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}x_0$$

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

代入

逆向扩散

改写为 $\exp(-\frac{(x-\mu)^2}{2\sigma^2})$

$$exp(-\frac{1}{2}((\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}})x_{t-1}^2-(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0)x_{t-1}+C(x_t,x_0)))$$

$$\frac{1}{\sigma^2}=\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}},\qquad \tilde{\beta}_t=\sigma^2=\frac{1}{\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}}=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$$\tilde{\mu}_t=\frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}x_0$$

$$=\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon)$$

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

代入

逆向扩散

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, t)$$

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$x_t \xrightarrow{\text{预测}} \epsilon_\theta(x_t, t) \rightarrow p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t_t I)) \rightarrow x_{t-1} \rightarrow \cdots \rightarrow x_0$

目标函数

$$L = E_{q(x_0)}[-log p_\theta(x_0)]$$

目标函数

$$p_\theta(x_0) = \int p_\theta(x_{0:T})dx_{1:T}$$

$$L = E_{q(x_0)}[-log p_\theta(x_0)]$$

$$= -E_{q(x_0)} log(\int p_\theta(x_{0:T})dx_{1:T})$$

$$= -E_{q(x_0)} log\left(\int q(x_{1:T}|x_0)\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}dx_{1:T}\right)$$

$$= -E_{q(x_0)} log\left(E_{q(x_{1:T}|x_0)}\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right)$$

$$E(X) = \int xf(x)\,dx$$

Jensen不等式

若$f(x)$是区间$[a,b]$上的凸函数，则对任意的$x_1,x_2,\dots,x_n\in[a,b]$，则有：

$$f\left(\sum_{i=1}^{n}\frac{x_i}{n}\right)\geq\frac{\sum_{i=1}^{n}f(x_i)}{n}$$

即：

$$f\big(E(x)\big)\geq E(f(x))$$

Jensen不等式

若$f(x)$是区间$[a,b]$上的凸函数，则对任意的$x_1, x_2, \dots, x_n \in [a,b]$，则有：

$$f\left(\sum_{i=1}^{n} \frac{x_i}{n}\right) \geq \frac{\sum_{i=1}^{n} f(x_i)}{n}$$

即：

$$f\big(E(x)\big) \geq E(f(x))$$

$$-E_{q(x_0)} log\left(E_{q(x_{1:T}|x_0)} \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right) \leq -E_{q(x_{0:T})} log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}$$

$$= E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = L_{VLB}$$

优化上界

$$L_{VLB} = E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

优化上界

$$L_{VLB} = E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$= E_{q(x_{0:T})}[-log p_\theta(x_T) + \sum_{t=1}^{T} log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}]$$

优化上界

$$L_{VLB} = E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$= E_{q(x_{0:T})}\left[-log p_\theta(x_T) + \sum_{t=1}^{T} log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_{q(x_{0:T})}[-log p_\theta(x_T) + \sum_{t=2}^{T} log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)} + log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}]$$

优化上界

$$L_{VLB} = E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$= E_{q(x_{0:T})}\left[-log p_\theta(x_T) + \sum_{t=1}^{T} log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_{q(x_{0:T})}\left[-log p_\theta(x_T) + \sum_{t=2}^{T} log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)} + log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right]$$

$$= E_{q(x_{0:T})}[-log p_\theta(x_T) + \sum_{t=2}^{T} log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} \cdot \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} + log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}]$$

$$= E_{q(x_{0:T})}[-log p_\theta(x_T) + \sum_{t=2}^{T} log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} + \sum_{t=2}^{T} log \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} + log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

优化上界

$$\sum_{i=2}^{T} log\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} = log\frac{q(x_2|x_0)}{q(x_1|x_0)}\cdot\frac{q(x_3|x_0)}{q(x_2|x_0)}\cdots\frac{q(x_T|x_0)}{q(x_{T-1}|x_0)} = log\frac{q(x_T|x_0)}{q(x_1|x_0)}$$

$$L_{VLB} = E_{q(x_{0:T})} log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$= E_{q(x_{0:T})}\left[-logp_\theta(x_T) + \sum_{t=1}^{T} log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_{q(x_{0:T})}\left[-logp_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right]$$

$$= E_{q(x_{0:T})}[-logp_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)}\cdot\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}]$$

$$= E_{q(x_{0:T})}[-logp_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} + \sum_{t=2}^{T} log\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

优化上界

$$L_{VLB} = E_{q(x_{0:T})} log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} = E_{q(x_{0:T})} log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$= E_{q(x_{0:T})}[-logp_\theta(x_T) + \sum_{t=2}^{T} log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} + log \frac{q(x_T|x_0)}{q(x_1|x_0)} + log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}]$$

$$= E_{q(x_{0:T})}[\sum_{t=2}^{T} log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} + log \frac{q(x_T|x_0)}{p_\theta(x_T)} - logp_\theta(x_0|x_1)]$$

KL散度：$KL(p||q)=\sum p(x)log\frac{p(x)}{q(x)}$

优化上界

$$L_{VLB}=E_{q(x_{0:T})}log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}=E_{q(x_{0:T})}log\frac{\prod_{t=1}^{T}q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T}p_\theta(x_{t-1}|x_t)}$$

$$=\ E_{q(x_{0:T})}[-logp_\theta(x_T)+\sum_{t=2}^{T}log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)}+log\frac{q(x_T|x_0)}{q(x_1|x_0)}+log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}]$$

$$=\ E_{q(x_{0:T})}[\sum_{t=2}^{T}log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)}+log\frac{q(x_T|x_0)}{p_\theta(x_T)}-logp_\theta(x_0|x_1)]$$

$$=\ E_{q(x_{0:T})}[\underbrace{-logp_\theta(x_0|x_1)]}_{L_0}+\sum_{t=2}^{T}\underbrace{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}_{L_t}+\underbrace{D_{KL}(q(x_T|x_0)||p_\theta(x_T))}_{L_T}$$

KL散度：$KL(p||q) = \sum p(x) log \frac{p(x)}{q(x)}$

优化上界

$$L_t = E_{q(x_{0:T})} D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}, \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I\right)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}, \mu_\theta(x_t, t), \tilde{\beta}_t I\right)$$

正态分布的KL散度：$D_{KL}(\mathcal{N}(\mu_1, \sigma_1^2)||\mathcal{N}(\mu_2, \sigma_2^2)) = log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$

KL散度：$KL(p||q) = \sum p(x) log \frac{p(x)}{q(x)}$

优化上界

$$L_t = E_{q(x_{0:T})} D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}, \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

正态分布的KL散度：$D_{KL}(\mathcal{N}(\mu_1, \sigma_1^2)||\mathcal{N}(\mu_2, \sigma_2^2)) = log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$

$$L_t = E_{q(x_{0:T})} \left[\frac{1}{2||\Sigma_\theta(x_t, t)||^2} ||\tilde{\mu}(x_t, x_0) - \mu_\theta(x_t, t)||^2 \right] + C$$

KL散度：$KL(p||q) = \sum p(x) log \frac{p(x)}{q(x)}$

优化上界

$$L_t = E_{q(x_{0:T})} D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}, \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

正态分布的KL散度：$D_{KL}(\mathcal{N}(\mu_1, \sigma_1^2)||\mathcal{N}(\mu_2, \sigma_2^2)) = log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$

$$L_t = E_{q(x_{0:T})} \left[\frac{1}{2||\Sigma_\theta(x_t, t)||^2} ||\tilde{\mu}(x_t, x_0) - \mu_\theta(x_t, t)||^2 \right] + C$$

$$\tilde{\mu}(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t)$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}})\epsilon_\theta(x_t, t)$$

KL散度：$KL(p||q)=\sum p(x)log\frac{p(x)}{q(x)}$

优化上界

$$L_t=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2\|\Sigma_\theta(x_t,t)\|^2}\|\tilde{\mu}(x_t,x_0)-\mu_\theta(x_t,t)\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2\|\Sigma_\theta(x_t,t)\|^2}\left\|\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t)-\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}})\epsilon_\theta(x_t,t)\right\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{\beta_t^2}{2\alpha_t(1-\bar{\alpha}_t)\|\Sigma_\theta(x_t,t)\|^2}\|\epsilon_t-\epsilon_\theta(x_t,t)\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{\beta_t^2}{2\alpha_t(1-\bar{\alpha}_t)\|\Sigma_\theta(x_t,t)\|^2}\left\|\epsilon_t-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon\,,t)\right\|^2\right]+C$$

KL散度：$KL(p||q)=\sum p(x)log\frac{p(x)}{q(x)}$

优化上界

$$L_t=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2\|\Sigma_\theta(x_t,t)\|^2}\|\tilde{\mu}(x_t,x_0)-\mu_\theta(x_t,t)\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2\|\Sigma_\theta(x_t,t)\|^2}\left\|\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon)-\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}})\epsilon_\theta(x_t,t)\right\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{\beta_t^2}{2\alpha_t(1-\bar{\alpha}_t)\|\Sigma_\theta(x_t,t)\|^2}\|\epsilon-\epsilon_\theta(x_t,t)\|^2\right]+C$$

$$=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{\beta_t^2}{2\alpha_t(1-\bar{\alpha}_t)\|\Sigma_\theta(x_t,t)\|^2}\left\|\epsilon-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon,t)\right\|^2\right]+C$$

$$L_{simple}=E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}\left[\left\|\epsilon-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon,t)\right\|^2\right]$$