*Generative Models: Fundamentals and Applications*

# Lecture 10: Diffusion Model


**Shuigeng Zhou, Yuxi Mi**
College of CSAI

December 1, 2025

# 目录

- 扩散模型的基本概念
  - Jascha Sohl-Dickstein et al. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics, JMLR 2015
- DDPM
  - Jonathan Ho, et al. Denoising Diffusion Probabilistic Models, NIPS 2020
- Score-based SDE
  - Yang Song, et al. Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021

# 扩散模型的基本概念

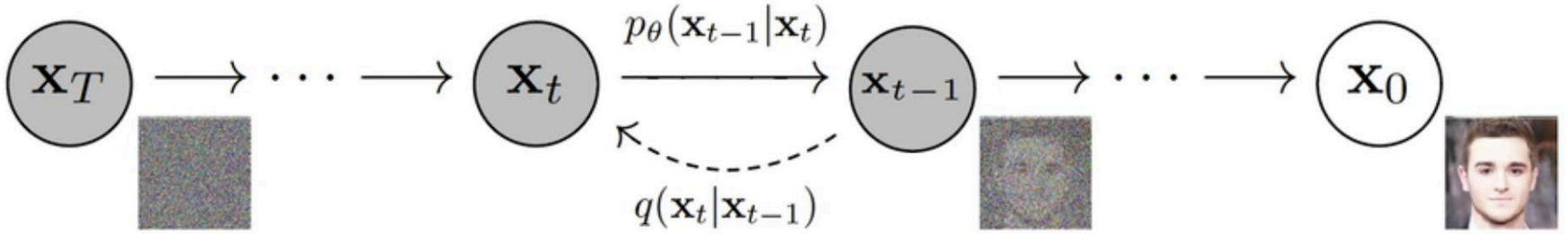
- 概率模型的目标
  - 可处理性（tractability）
    - 可以对模型进行分析评估，容易拟合数据
    - 无法恰当地描述数据集的结构
    - Eg: 高斯分布、拉普拉斯分布
  - 灵活性（flexibility）
    - 适合任意数据中的结构
    - Eg: $p(x) = \frac{\phi(x)}{Z}$

# 扩散模型的基本概念

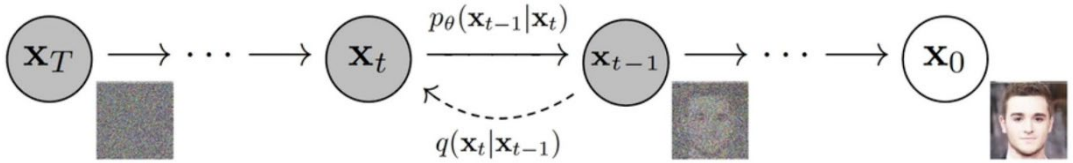
- 改善平衡的方法
  - 平均场论及其展开式
  - 变分贝叶斯
  - 对比散度
  - 最小概率流
  - 最小KL收缩
  - 正确的评分规则
  - 分数匹配
  - 伪似然…

# 扩散模型的基本概念

- 扩散模型：同时实现灵活性和可处理性
  - 基本思路
    - 使用马尔科夫链将一个分布逐步转化成另一个分布
    - 学习一个反向扩散过程，恢复数据中的结构，生成一个高度灵活且易于处理的数据生成模型



# 扩散模型的基本概念




- 正向扩散过程
  - 将原始复杂分布$q(x^{(0)})$逐步转变成简单分布$q(x^{(T)})$
  - 条件概率

$$q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) = T_\pi\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)};\beta_t\right)$$

  - $T_\pi$：马尔可夫扩散核（Markov diffusion kernel）
    - 高斯分布

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

    - 二项式分布

$$q\left(x^{(t)}|x^{(t-1)}\right) = \mathrm{Bin}\left(x^{(t)};x^{(t-1)}(1-\beta_t)+0.5\beta_t\right)$$

    - $\beta_t$：扩散速率

# 扩散模型的基本概念

- 正向扩散过程
  - 计算每次扩散后的概率分布

$$\pi(\mathbf{y}) = \int d\mathbf{y}' T_\pi(\mathbf{y}|\mathbf{y}';\beta)\,\pi(\mathbf{y}')$$

    - 高斯分布

$$\pi\left(x^{(T)}\right) = \mathcal{N}\left(x^{(T)};\mathbf{0},\boldsymbol{I}\right)$$

    - 二项分布

$$\pi\left(x^{(T)}\right) = \mathrm{Bin}\left(x^{(T)};\,0.5\right)$$

# 扩散模型的基本概念

- 正向扩散过程
  - 高斯分布：证明

回顾一下，当我们合并两个具有不同方差的高斯，$\mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I})$和$\mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{I})$，新的分布是$\mathcal{N}(\mathbf{0}, (\sigma_1^2 + \sigma_2^2)\mathbf{I})$。这里的合并标准差是$\sqrt{(1-\alpha_t) + \alpha_t(1-\alpha_{t-1})} = \sqrt{1-\alpha_t\alpha_{t-1}}$

$$q\left(x^{(t)}\middle|x^{(t-1)}\right) = \mathcal{N}\left(x^{(t)}; \sqrt{1-\beta_t}x^{(t-1)}, \beta_t I\right)$$

- 令$\alpha_t = 1 - \beta_t$和$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$
- 重参数化技巧

$$\begin{aligned} x^{(t)} &= \sqrt{\alpha_t}x^{(t-1)} + \sqrt{1-\alpha_t}\epsilon^{(t-1)}, \qquad \epsilon^{(t-1)} \sim \mathcal{N}(0,1) \\ &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x^{(t-2)} + \sqrt{1-\alpha_{t-1}}\epsilon^{(t-2)}\right) + \sqrt{1-\alpha_t}\epsilon^{(t-1)} \\ &= \sqrt{\alpha_t\alpha_{t-1}}x^{(t-2)} + \boxed{\sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}}\epsilon^{(t-2)} + \sqrt{1-\alpha_t}\epsilon^{(t-1)}} \\ &= \sqrt{\alpha_t\alpha_{t-1}}x^{(t-2)} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\epsilon}^{(t-2)}, \qquad \bar{\epsilon}^{(t-2)} \text{混合两个高斯分布} \end{aligned}$$

# 扩散模型的基本概念

- 正向扩散过程
  - 高斯分布证明

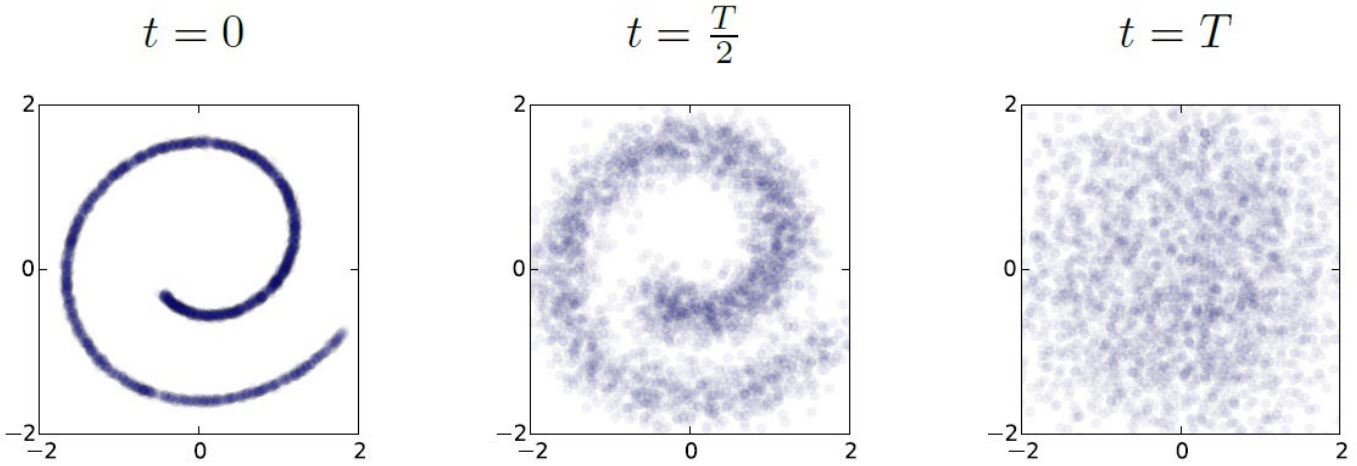
$$
\begin{aligned}
x^{(t)} &= \sqrt{\alpha_t \alpha_{t-1}} x^{(t-2)} + \sqrt{1-\alpha_t \alpha_{t-1}} \bar{\epsilon}^{(t-2)} \\
&= \sqrt{\alpha_t \alpha_{t-1} \dots \alpha_1} x^{(0)} + \sqrt{1-\alpha_t \alpha_{t-1} \dots \alpha_1} \bar{\epsilon}^{(0)} \\
&= \sqrt{\bar{\alpha}_t} x^{(0)} + \sqrt{1-\bar{\alpha}_t} \epsilon,
\end{aligned}
$$

$$
q\left(x^{(t)} \middle| x^{(0)}\right) = \mathcal{N}\left(x^{(0)}; \sqrt{\bar{\alpha}_t} x^{(0)}, (1-\bar{\alpha}_t) I\right) \xrightarrow{t \to \infty} \mathcal{N}(0, \mathrm{I})
$$

  - 注意：
    - 正向扩散等价于在原始数据分布中不断加入高斯噪声
    - 扩散速率$\beta_t$对训练模型的性能非常重要

# 扩散模型的基本概念

- 正向扩散过程

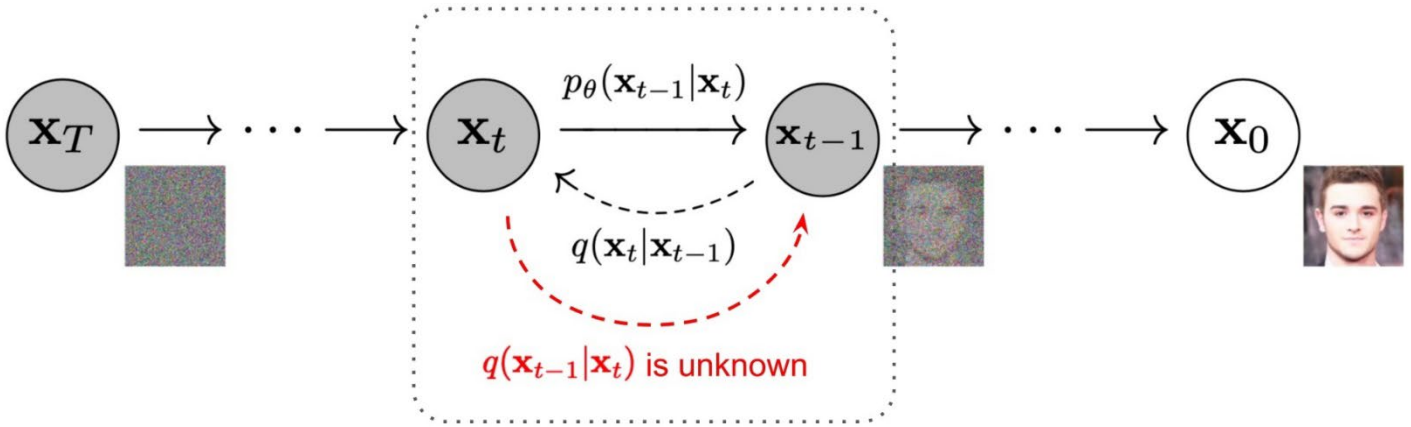



  - 联合概率

$$q\left(\mathbf{x}^{(0\cdots T)}\right) = q\left(\mathbf{x}^{(0)}\right)\prod_{t=1}^{T} q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)$$

# 扩散模型的基本概念

- 反向扩散过程




  - 将原始简单分布$p(x^{(T)})$逐步转变成复杂分布$p(x^{(0)})$
    - 原始简单分布

$$p\left(\mathbf{x}^{(T)}\right) = \pi\left(\mathbf{x}^{(T)}\right) = q\left(x^{(T)}\right)$$

# 扩散模型的基本概念

- 反向扩散过程
  - 如果$\beta_t$足够小，反向条件概率分布$p\left(x^{(t-1)}\middle|x^{(t)}\right)$的函数形式（分布族）与正向相同 （Feller,1949）
    - 高斯分布

$$p\left(\mathbf{x}^{(t-1)} \mid \mathbf{x}^{(t)}\right) = \mathcal{N}\left(\mathbf{x}^{(t-1)}; \mathbf{f}_\mu\left(\mathbf{x}^{(t)}, t\right), \mathbf{f}_\Sigma\left(\mathbf{x}^{(t)}, t\right)\right)$$

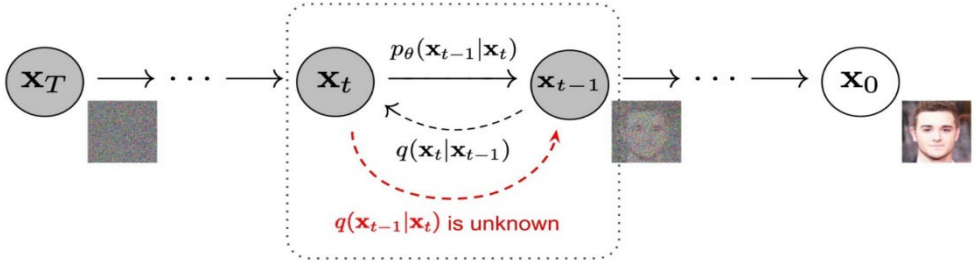
    - 二项分布

$$p\left(\mathbf{x}^{(t-1)} \mid \mathbf{x}^{(t)}\right) = \mathcal{B}\left(\mathbf{x}^{(t-1)}; \mathbf{f}_b\left(\mathbf{x}^{(t)}, t\right)\right)$$

  - 联合概率

$$p\left(\mathbf{x}^{(0\cdots T)}\right) = p\left(\mathbf{x}^{(T)}\right) \prod_{t=1}^{T} p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)$$

# 扩散模型的基本概念




- 反向扩散过程
  - 以$x_0$为条件时，反向条件概率$q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)$可控

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)=q(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{x}_0)\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)}$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{(\mathbf{x}_t-\sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{\beta_t}+\frac{(\mathbf{x}_{t-1}-\sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0)^2}{1-\bar{\alpha}_{t-1}}-\frac{(\mathbf{x}_t-\sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\right)\right)$$

$$=\exp\left(-\frac{1}{2}\left(\frac{\mathbf{x}_t^2-2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1}+\alpha_t\mathbf{x}_{t-1}^2}{\beta_t}+\frac{\mathbf{x}_{t-1}^2-2\sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0\mathbf{x}_{t-1}+\bar{\alpha}_{t-1}\mathbf{x}_0^2}{1-\bar{\alpha}_{t-1}}-\frac{(\mathbf{x}_t-\sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\right)\right)$$

$$=\exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)\mathbf{x}_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\mathbf{x}_{t-1}+C(\mathbf{x}_t,\mathbf{x}_0)\right)\right)$$

与$x_{t-1}$无关的全部项

  - 结论：$q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)\sim\mathcal{N}\left(x^{(t-1)};\tilde{\mu}_t,\tilde{\beta}_t I\right)$

# 扩散模型的基本概念

- 反向扩散过程
  - 以$x_0$为条件时，反向条件概率$q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)$可控

$$\tilde{\beta}_t = 1/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) = 1/(\frac{\alpha_t - \bar{\alpha}_t + \beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) &= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) \\ &= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0\end{aligned}$$

# 扩散模型的基本概念

- 反向扩散过程
  - 以$x_0$为条件时，反向条件概率$q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)$可控

$$x^{(t)}=\sqrt{\bar{\alpha}_t}x^{(0)}+\sqrt{1-\bar{\alpha}_t}\epsilon_t$$

$$\mathbf{x}_0=\frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t-\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t\right)$$

$$\tilde{\boldsymbol{\mu}}_t=\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t-\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t\right)$$

$$=\frac{1}{\sqrt{\alpha_t}}\left(x_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)$$

# 扩散模型的基本概念

- 模型概率

$$
\begin{aligned}
p\left(\mathbf{x}^{(0)}\right) &= \int d\mathbf{x}^{(1\cdots T)} p\left(\mathbf{x}^{(0\cdots T)}\right) \\
&= \int d\mathbf{x}^{(1\cdots T)} p\left(\mathbf{x}^{(0\cdots T)}\right) \frac{q\left(\mathbf{x}^{(1\cdots T)}|\mathbf{x}^{(0)}\right)}{q\left(\mathbf{x}^{(1\cdots T)}|\mathbf{x}^{(0)}\right)} \\
&= \int d\mathbf{x}^{(1\cdots T)} q\left(\mathbf{x}^{(1\cdots T)}|\mathbf{x}^{(0)}\right) p\left(\mathbf{x}^{(T)}\right) \prod_{t=1}^{T} \frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}
\end{aligned}
$$

  - 通过对正向轨迹的样本求平均来快速评估
  - 如果正向和反向的轨迹相同，那么只需要从正向条件概率分布中采一个样本就能精确计算

# 扩散模型的基本概念

- 模型训练：最大化对数似然函数$L$

$$
\begin{aligned}
L &= \int d\mathbf{x}^{(0)} q\left(\mathbf{x}^{(0)}\right) \log p\left(\mathbf{x}^{(0)}\right) \\
&= \int d\mathbf{x}^{(0)} q\left(\mathbf{x}^{(0)}\right) \log \left[ \int d\mathbf{x}^{(1\cdots T)} q\left(\mathbf{x}^{(1\cdots T)} \mid \mathbf{x}^{(0)}\right) p\left(\mathbf{x}^{(T)}\right) \prod_{t=1}^{T} \frac{p\left(\mathbf{x}^{(t-1)} \mid \mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)} \mid \mathbf{x}^{(t-1)}\right)} \right] \\
&\geq \int dx^{(0)} q\left(x^{(0)}\right) \left[ \int dx^{(1\ldots T)} q\left(x^{(1\ldots T)} \middle| x^{(0)}\right) \log p\left(x^{(T)}\right) \prod_{t=1}^{T} \frac{p\left(x^{(t-1)} \middle| x^{(t)}\right)}{q\left(x^{(t)} \middle| x^{(t-1)}\right)} \right] \\
&= \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log \left[ p\left(\mathbf{x}^{(T)}\right) \prod_{t=1}^{T} \frac{p\left(\mathbf{x}^{(t-1)} \mid \mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)} \mid \mathbf{x}^{(t-1)}\right)} \right] \\
&= E_q \left[ \log \frac{p_\theta\left(x^{(0\ldots T)}\right)}{q\left(x^{(1\ldots T)} \middle| x^{(0)}\right)} \right]
\end{aligned}
$$

  - 正向和反向的轨迹相同时，等号成立

# 扩散模型的基本概念

- 模型训练：最大化紧致下界$K$

$$
\begin{aligned}
K &= \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log\left[p\left(\mathbf{x}^{(T)}\right) \prod_{t=1}^{T} \frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}\right] \\
&= \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \sum_{t=1}^{T} \log\left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}\right] + \int d\mathbf{x}^{(T)} q\left(\mathbf{x}^{(T)}\right) \log p\left(\mathbf{x}^{(T)}\right) \\
&= \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \sum_{t=1}^{T} \log\left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}\right] + \int d\mathbf{x}^{(T)} q\left(\mathbf{x}^{(T)}\right) \log \pi\left(\mathbf{x}^{T}\right) \\
&= \sum_{t=1}^{T} \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log\left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}\right] - H_p\left(\mathbf{X}^{(T)}\right)
\end{aligned}
$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界
  - 去除$t=0$的影响

$$p\left(\mathbf{x}^{(0)}|\mathbf{x}^{(1)}\right)=q\left(\mathbf{x}^{(1)}|\mathbf{x}^{(0)}\right)\frac{\pi\left(\mathbf{x}^{(0)}\right)}{\pi\left(\mathbf{x}^{(1)}\right)}=T_{\pi}\left(\mathbf{x}^{(0)}|\mathbf{x}^{(1)};\beta_1\right)$$

  - 紧致下界

$$K=\sum_{t=2}^{T}\int d\mathbf{x}^{(0\cdots T)}q\left(\mathbf{x}^{(0\cdots T)}\right)\log\left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right)}\right]$$

$$+\int d\mathbf{x}^{(0)}d\mathbf{x}^{(1)}q\left(\mathbf{x}^{(0)},\mathbf{x}^{(1)}\right)\log\left[\frac{q\left(\mathbf{x}^{(1)}|\mathbf{x}^{(0)}\right)\pi\left(\mathbf{x}^{(0)}\right)}{q\left(\mathbf{x}^{(1)}|\mathbf{x}^{(0)}\right)\pi\left(\mathbf{x}^{(1)}\right)}\right]-H_p\left(\mathbf{X}^{(T)}\right)$$

$$=\int dx^{(0)}q\left(x^{(0)}\right)\log\pi\left(x^{(0)}\right)-\int dx^{(1)}q\left(x^{(1)}\right)\log\pi\left(x^{(1)}\right)=H_p\left(X^{(T)}\right)-H_p\left(X^{(T)}\right)=0$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界
  - 正向扩散过程是一个马尔科夫链

$$K = \sum_{t=2}^{T} \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log \left[ \frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}, \mathbf{x}^{(0)}\right)} \right] - H_p\left(\mathbf{X}^{(T)}\right)$$

$$= \sum_{t=2}^{T} \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log \left[ \frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}, \mathbf{x}^{(0)}\right)} \frac{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(0)}\right)}{q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(0)}\right)} \right] - H_p\left(\mathbf{X}^{(T)}\right)$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界

$$\sum_{t=2}^{T}\int dx^{(0\ldots T)}q\left(x^{(0\ldots T)}\right)\log q\left(x^{(t-1)}\middle|x^{(0)}\right)$$

$$=\sum_{t=2}^{T}\int dx^{(0)}dx^{(t-1)}q\left(x^{(0)},x^{(t-1)}\right)\log q\left(x^{(t-1)}\middle|x^{(0)}\right)$$

$$=-\sum_{t=2}^{T}H_q\left(X^{(t-1)}\middle|X^{(0)}\right)$$

$$\sum_{t=2}^{T}\int dx^{(0\ldots T)}q\left(x^{(0\ldots T)}\right)\log q\left(x^{(t)}\middle|x^{(0)}\right)=-\sum_{t=2}^{T}H_q\left(X^{(t)}\middle|X^{(0)}\right)$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界

$$K = \sum_{t=2}^{T} \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log \left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}, \mathbf{x}^{(0)}\right)}\right] + \sum_{t=2}^{T} \left[H_q\left(\mathbf{X}^{(t)}|\mathbf{X}^{(0)}\right) - H_q\left(\mathbf{X}^{(t-1)}|\mathbf{X}^{(0)}\right)\right] - H_p\left(\mathbf{X}^{(T)}\right) \tag{49}$$

$$= \sum_{t=2}^{T} \int d\mathbf{x}^{(0\cdots T)} q\left(\mathbf{x}^{(0\cdots T)}\right) \log \left[\frac{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}, \mathbf{x}^{(0)}\right)}\right] + H_q\left(\mathbf{X}^{(T)}|\mathbf{X}^{(0)}\right) - H_q\left(\mathbf{X}^{(1)}|\mathbf{X}^{(0)}\right) - H_p\left(\mathbf{X}^{(T)}\right).$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界

$$\sum_{t=2}^{T}\int dx^{(0\ldots T)}q\left(x^{(0\ldots T)}\right)\log\left[\frac{p\left(x^{(t-1)}\middle|x^{(t)}\right)}{q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)}\right]$$

$$=\sum_{t=2}^{T}\int dx^{(0)}dx^{(t-1)}dx^{(t)}q\left(x^{(0)},x^{(t-1)},x^{(t)}\right)\log\left[\frac{p\left(x^{(t-1)}\middle|x^{(t)}\right)}{q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)}\right]$$

$$=\sum_{t=2}^{T}\int dx^{(0)}dx^{(t)}q\left(x^{(0)},x^{(t)}\right)\left\{-\int dx^{(t-1)}q\left(x^{(t-1)}|x^{(t)},x^{(0)}\right)\log\left[\frac{q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)}{p\left(x^{(t-1)}\middle|x^{(t)}\right)}\right]\right\}$$

$$=-\sum_{t=2}^{T}\int dx^{(0)}dx^{(t)}q\left(x^{(0)},x^{(t)}\right)D_{KL}\left(q\left(x^{(t-1)}\middle|x^{(t)},x^{(0)}\right)\|\,p\left(x^{(t-1)}\middle|x^{(t)}\right)\right)$$

# 扩散模型的基本概念

- 模型训练：最大化紧致下界

$$K = -\sum_{t=2}^{T} \int d\mathbf{x}^{(0)} d\mathbf{x}^{(t)} q\left(\mathbf{x}^{(0)}, \mathbf{x}^{(t)}\right) D_{KL}\left(q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}, \mathbf{x}^{(0)}\right) \middle|\middle| p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)\right)$$
$$+ H_q\left(\mathbf{X}^{(T)}|\mathbf{X}^{(0)}\right) - H_q\left(\mathbf{X}^{(1)}|\mathbf{X}^{(0)}\right) - H_p\left(\mathbf{X}^{(T)}\right).$$

  - 紧致下界$K$可以解析地计算

$$\hat{p}\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right) = \underset{p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right)}{\operatorname{argmax}} K$$

    - 高斯分布：估计反向传播的均值和协方差矩阵序列
    - 二项分布：估计反向传播的状态翻转概率序列

# 扩散模型的基本概念

- 特点：
  - 模型结构的极端灵活性
    - 任何光滑的目标分布都存在扩散，因此可以捕捉任意形式的数据分布
  - 易处理，精确采样
    - 该模型估计了扩散过程的细微扰动$\hat{p}(x^{(t-1)}|x^{(t)})$，比用一个单一的不可解析的归一化函数来估计描述的数据分布更容易处理

# 目录

- 扩散模型的基本概念
  - Jascha Sohl-Dickstein et al. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics, JMLR 2015
- DDPM
  - Jonathan Ho, et al. Denoising Diffusion Probabilistic Models, NIPS 2020
- Score-based SDE
  - Yang Song, et al. Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021

# DDPM

- 目标函数分解

$$\int dx^{(0)} q\left(x^{(0)}\right) \log \pi\left(x^{(0)}\right) - \int dx^{(1)} q\left(x^{(1)}\right) \log \pi\left(x^{(1)}\right) = H_p\left(X^{(T)}\right) - H_p\left(X^{(T)}\right) = 0$$

$$H_q\left(X^{(T)} \middle| X^{(0)}\right) - H_q\left(X^{(1)} \middle| X^{(0)}\right) - H_p\left(X^{(T)}\right)$$

$$= -E_q \log q\left(x^{(T)} \middle| x^{(0)}\right) + E_q \log q\left(x^{(1)} \middle| x^{(0)}\right) + E_q \log p\left(x^{(T)}\right)$$

$$= -E_q \log \frac{q\left(x^{(T)} \middle| x^{(0)}\right)}{p\left(x^{(T)}\right)} + E_q \log \frac{p\left(x^{(0)} \middle| x^{(1)}\right)\pi\left(x^{(1)}\right)}{\pi\left(x^{(0)}\right)}$$

$$= -E_q D_{KL}\left(q\left(x^{(T)} \middle| x^{(0)}\right) \| p\left(x^{(T)}\right)\right) + E_q \log p\left(x^{(0)} \middle| x^{(1)}\right)$$

# DDPM

- 目标函数分解

$$\max \quad -\sum_{t=2}^{T} \int d\mathbf{x}^{(0)} d\mathbf{x}^{(t)} q\left(\mathbf{x}^{(0)}, \mathbf{x}^{(t)}\right) D_{KL}\left(q\left(\mathbf{x}^{(t-1)} | \mathbf{x}^{(t)}, \mathbf{x}^{(0)}\right) \middle|\middle| p\left(\mathbf{x}^{(t-1)} | \mathbf{x}^{(t)}\right)\right) + H_q\left(\mathbf{X}^{(T)} | \mathbf{X}^{(0)}\right) - H_q\left(\mathbf{X}^{(1)} | \mathbf{X}^{(0)}\right) - H_p\left(\mathbf{X}^{(T)}\right).$$

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

# DDPM

- 正向扩散过程$L_T$

$$L_T = D_{\mathrm{KL}}\left(q(\mathbf{x}_T|\mathbf{x}_0) \,\|\, p_\theta(\mathbf{x}_T)\right)$$

$$q\left(x^{(t)}\middle|x^{(0)}\right) = \mathcal{N}\left(x^{(t)}; \sqrt{\bar{\alpha}_t}x^{(0)}, \sqrt{1-\bar{\alpha}_t}I\right) \xrightarrow{t\to\infty} \mathcal{N}(0, \mathrm{I})$$

  - $L_T$由扩散速率$\beta_t$决定
  - 预先设置$\beta_t$为某个常数，则$L_T$是一个常数

# DDPM

- 反向扩散过程$L_{t-1}$

$$L_{t-1} = D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0\right) \| p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)\right)$$

$$q\left(x^{(t-1)}|x^{(t)}, x^{(0)}\right) = \mathcal{N}\left(x^{(t-1)}; \tilde{\mu}_t\left(x^{(t)}, x^{(0)}\right), \tilde{\beta}_t I\right)$$

$$\mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t\right), \boldsymbol{\Sigma}_\theta\left(\mathbf{x}_t, t\right)\right)$$

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right)$$

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

# DDPM

- 反向扩散过程$L_{t-1}$
  - 直接设置$\Sigma_\theta(x_t,t)=\sigma_t^2 I$
    - $\sigma_t^2=\beta_t$
      - 正向扩散和反向的轨迹相同，保证了$x_0\sim\mathcal{N}(0,I)$
    - $\sigma_t^2=\tilde{\beta}_t=\frac{1-\bar{\alpha}_{t-1}}{\sqrt{1-\bar{\alpha}_t}}\beta_t$
      - $x_0$给定，则反向扩散过程可控，保证了反向扩散到一个具体的点$x_0$
  - $L_{t-1}$可重新表示为

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

# DDPM

- 反向扩散过程$L_{t-1}$
  - 训练$\mu_\theta(x_t, t)$来近似$\tilde{\mu}_t(x_t, x_0)$

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right)$$

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

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t,t) = \tilde{\boldsymbol{\mu}}_t\left(\mathbf{x}_t, \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t))\right) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right)$$

# DDPM

- 反向扩散过程$L_{t-1}$

$$L_{t-1} - C = \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\right]$$

$$= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{1}{2\|\boldsymbol{\Sigma}_\theta\|_2^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right) - \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right)\right\|^2\right]$$

$$= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\right]$$

$$= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t,t)\|^2\right]$$

与扩散速率有关，
与模型参数$\theta$无关

# DDPM

- 反向扩散过程$L_{t-1}$

$$\min L_{t-1} = D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))$$

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

$$= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\right] + C$$

$$\min L_{t-1}^{\text{simple}} = E_{x_0,\epsilon_t}[\|\epsilon_t - \epsilon_\theta(x_t, t)\|^2]$$

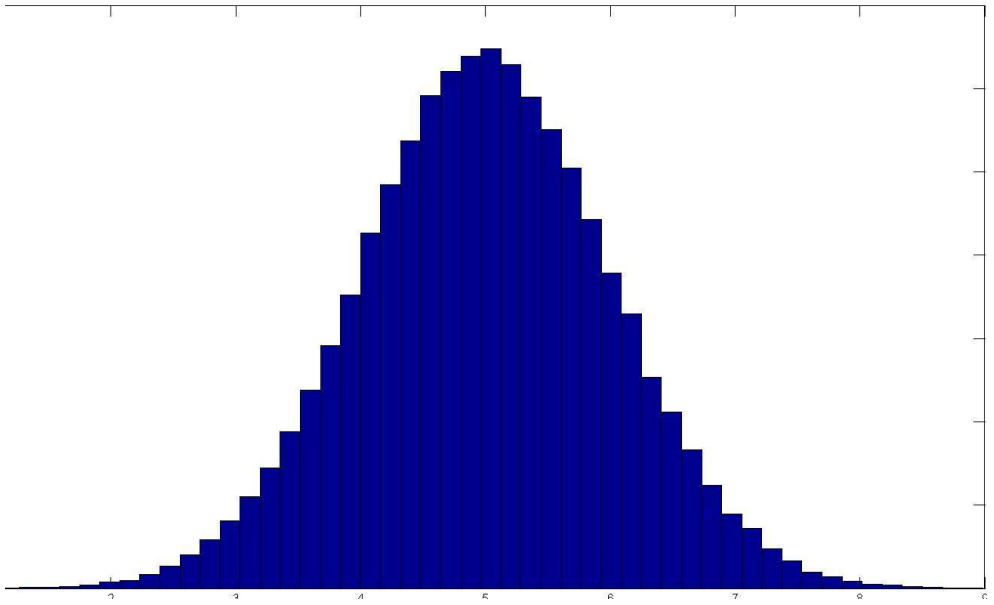
$$= E_{x_0,\epsilon_t}\left[\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right]$$

# DDPM

- 反向扩散过程$L_0$

$$L_0 = -\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)$$

  - 将图像数据$\{0,1,\ldots,255\} \rightarrow [-1,1]$
  - $x_0$是离散数据



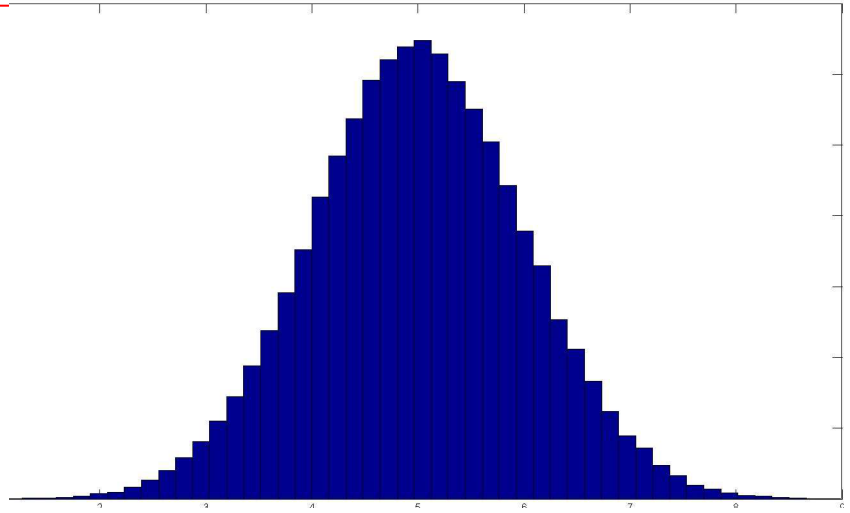
# DDPM

- 反向扩散过程$L_0$
  - 离散对数似然函数

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

$$\delta_+(x) = \begin{cases} \infty & \text{if } x = 1 \\ x + \frac{1}{255} & \text{if } x < 1 \end{cases}$$

$$\delta_-(x) = \begin{cases} -\infty & \text{if } x = -1 \\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$$



# DDPM

- 反向扩散过程$L_0$
  - 从分布$p_\theta(x_0|x_1)$中采样$x_0$，这是连续数据离散化的过程，整个过程是无损的

$$x_0 = \mu(x_1, 1)$$

$$L_0 = -\log p_\theta(x_0|x_1) = 0$$

# DDPM

- 简化后的目标函数

$$\min E_q\left[\sum_{t=2}^{T} L_{t-1}^{\text{simple}} + L_0\right]$$

$$= E_q\left[\sum_{t=2}^{T} E_{x_0,\epsilon_t}\left[\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right] - \log p_\theta(x_0|x_1)\right]$$

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

# DDPM

- 算法

**Algorithm 1** Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on

$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$

6: **until** converged

# DDPM

- 算法

**Algorithm 2** Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta (\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

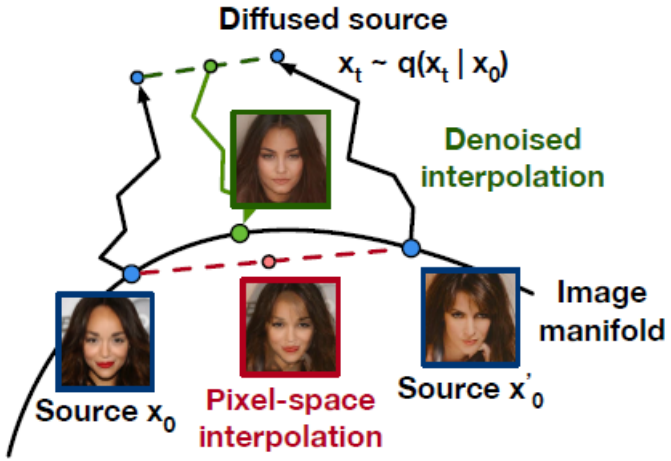
从分布$p(x_{t-1}|x_t)$ 中采样

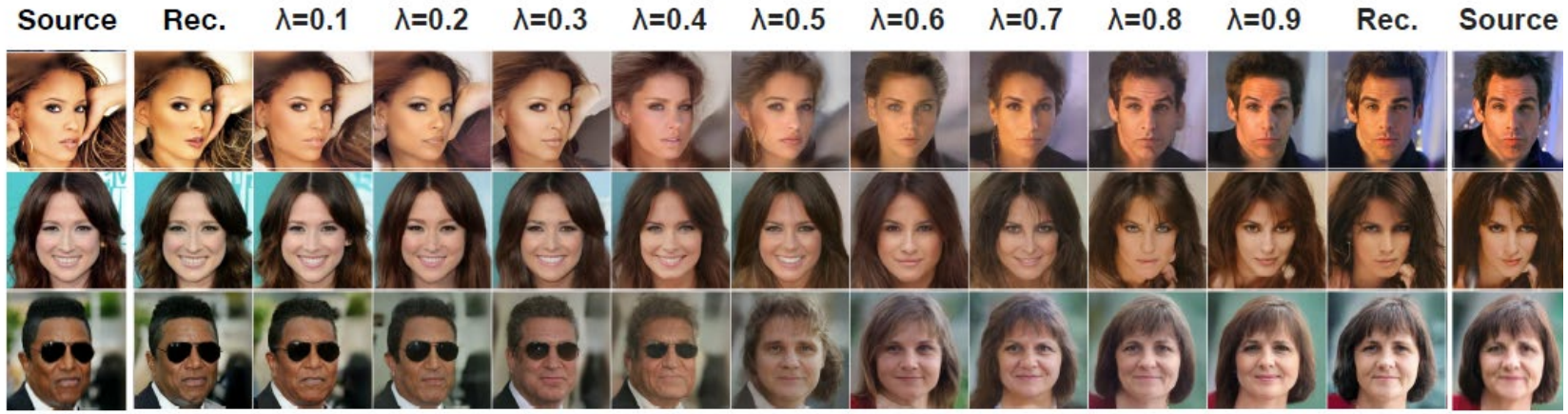
# DDPM

- 图像生成

# DDPM

- 隐空间线性插值



# 目录

- 扩散模型的基本概念
  - Jascha Sohl-Dickstein et al. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics, JMLR 2015
- DDPM
  - Jonathan Ho, et al. Denoising Diffusion Probabilistic Models, NIPS 2020
- Score-based SDE
  - Yang Song, et al. Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021

# Score-based generative modeling

- 定义概率分布

$$p_\theta(\mathbf{x}) = \frac{e^{-f_\theta(\mathbf{x})}}{Z_\theta}$$

  - $Z_\theta$是归一化常数（normalizing constant）
  - 注意
    - $Z_\theta$是一个难以处理的量
    - 最大化对数似然函数时，必须计算归一化常数$Z_\theta$
  - 解决方法
    - 限制模型结构，迫使$Z_\theta = 1$
    - 近似规则化常数，如变分推断、对比散度等

# Score-based generative modeling

- 定义得分函数

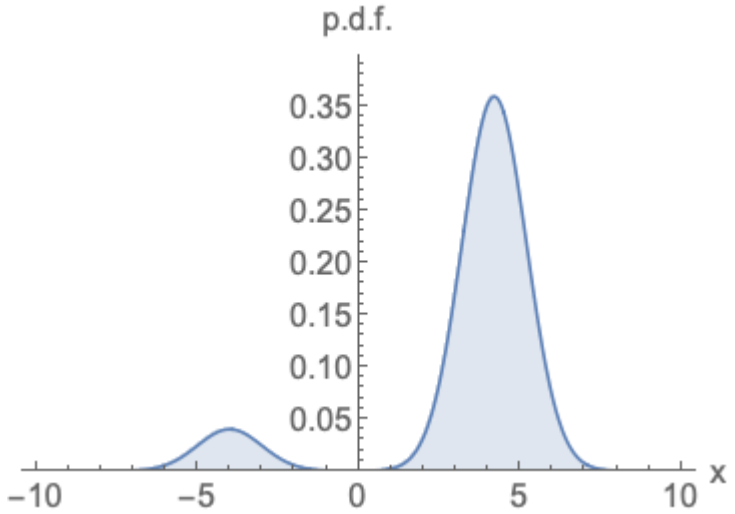
$$s_\theta(x) = \nabla_x \log p_\theta(x)$$

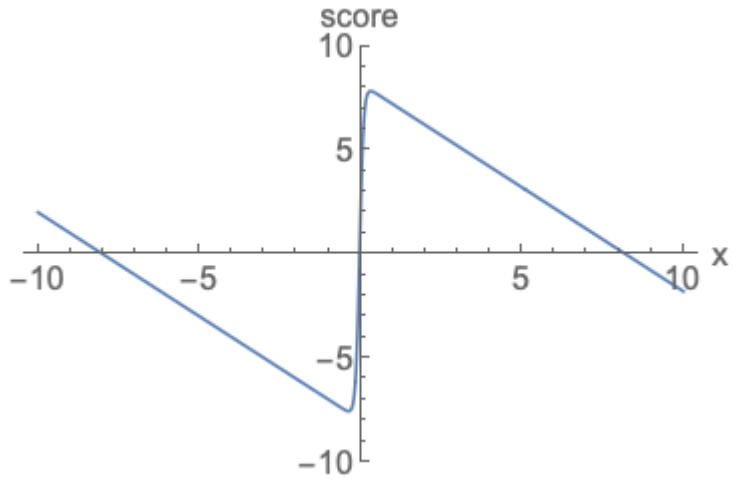
  - 估计对数似然函数的梯度
- 使用得分函数的模型统称为score-based model
  - $s_\theta(x)$和$Z_\theta$相互独立，不需要设计复杂的结构来使得$Z_\theta$易于处理

$$\mathbf{s}_\theta(\mathbf{x}) = \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}) = -\nabla_{\mathbf{x}} f_\theta(\mathbf{x}) + \underbrace{\nabla_{\mathbf{x}} \log Z_\theta}_{=0} = -\nabla_{\mathbf{x}} f_\theta(\mathbf{x}).$$

# Score-based generative modeling

- 参数化$p_\theta(x)$需要保证积分为1
- 参数化$s_\theta(x)$不需要保证积分为1



# Score-based generative modeling

- 最小化模型和数据真实得分之间的Fisher散度

$$\frac{1}{2}\mathbb{E}_{p_{\text{data}}}\left[\left\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}) - \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})\right\|_2^2\right]$$

  - 注意：真实的梯度未知
  - 得分匹配 (score matching) 方法
    - 在不知道真实数据得分的情况下，最小化Fisher散度

$s_\theta(x)$的雅克比矩阵

$$\mathbb{E}_{p_{\text{data}}(\mathbf{x})}\left[\operatorname{tr}\left(\nabla_{\mathbf{x}}\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\right) + \frac{1}{2}\left\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\right\|_2^2\right]$$

    - 无需对$s_\theta(x)$有一个较强的假设
    - 使用一个向量函数，输入和输出具有相同的维度

# Score-based generative modeling

- 降噪得分匹配 (Denoising score matching)
  - 从添加了噪声的数据分布中进行采样

$$\frac{1}{2}\mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})p_{\text{data}}(\mathbf{x})}[\|\mathbf{s}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}}\log q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x})\|_2^2]$$

  - $q_\sigma(\tilde{x}|x)$：预定义的高斯噪声分布
  - $s_{\theta^*}(x)$：最小化目标函数

$$\mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}) = \nabla_{\mathbf{x}}\log q_\sigma(\mathbf{x})$$

  - 当噪声足够小时，$q_\sigma(x) \approx p_{data}(x)$

$$\mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}) = \nabla_{\mathbf{x}}\log q_\sigma(\mathbf{x}) \approx \nabla_{\mathbf{x}}\log p_{\text{data}}(\mathbf{x})$$

# Score-based generative modeling

- 郎之万动力学 (Langevin dynamics)
  - 使用得分函数对真实数据分布$p(x)$进行MCMC采样
    - 首先从先验分布$\pi(x)$中采样$\tilde{x}_0$
    - 然后构造Markov链，递归计算

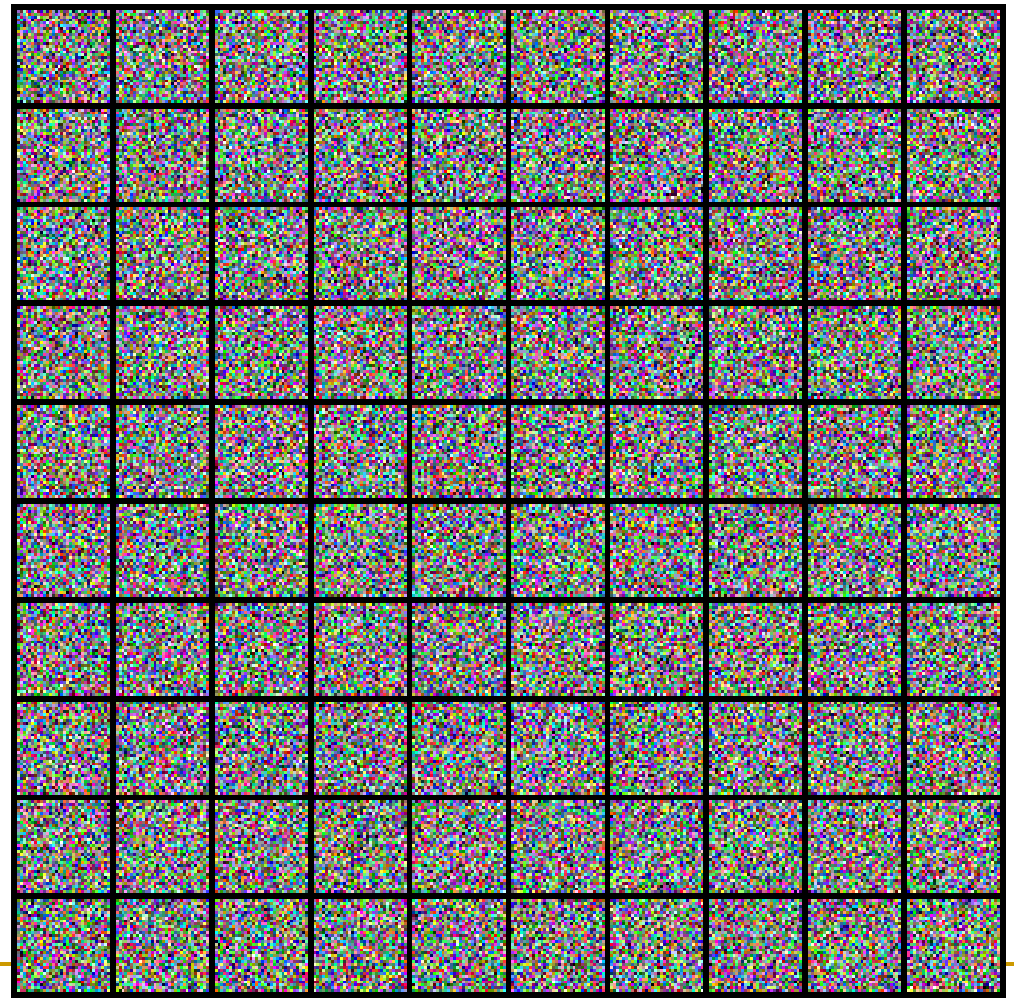
$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_{t-1} + \frac{\epsilon}{2}\nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}_{t-1}) + \sqrt{\epsilon}\,\mathbf{z}_t$$

      - 其中$z_t \sim \mathcal{N}(0, I),\ \epsilon > 0$
    - 当$\epsilon \to 0,\ T \to \infty$时

$$p(\tilde{x}_T) \to p_{data}(x)$$

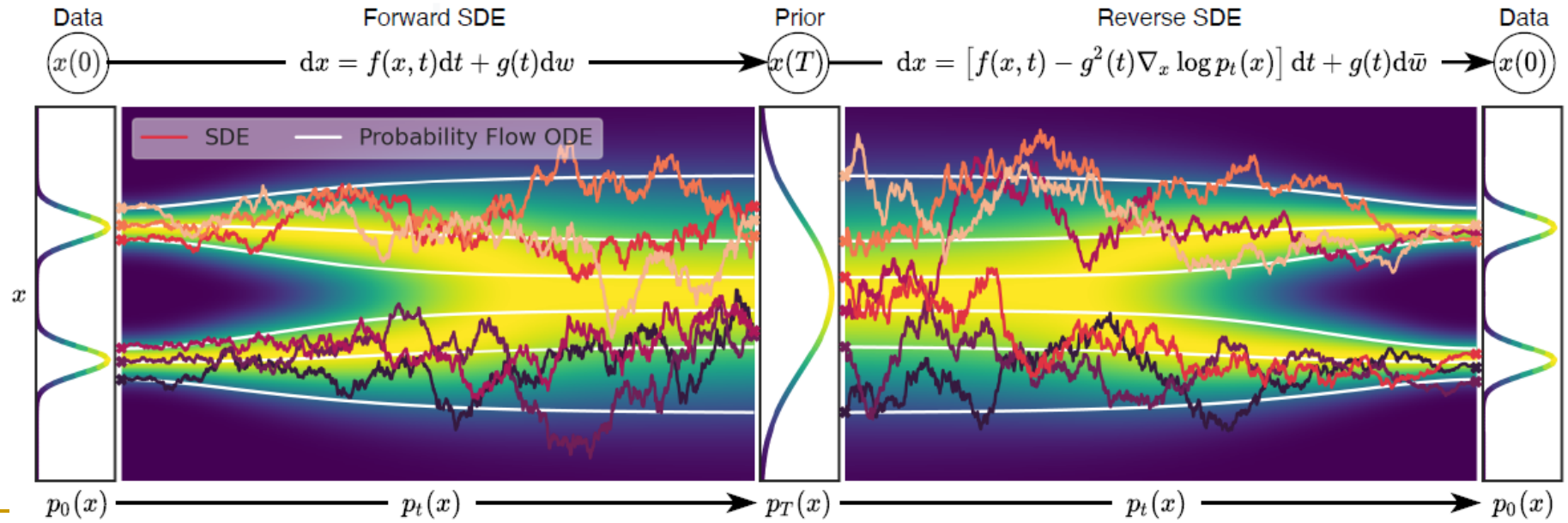
# Score-based generative modeling

- 生成效果



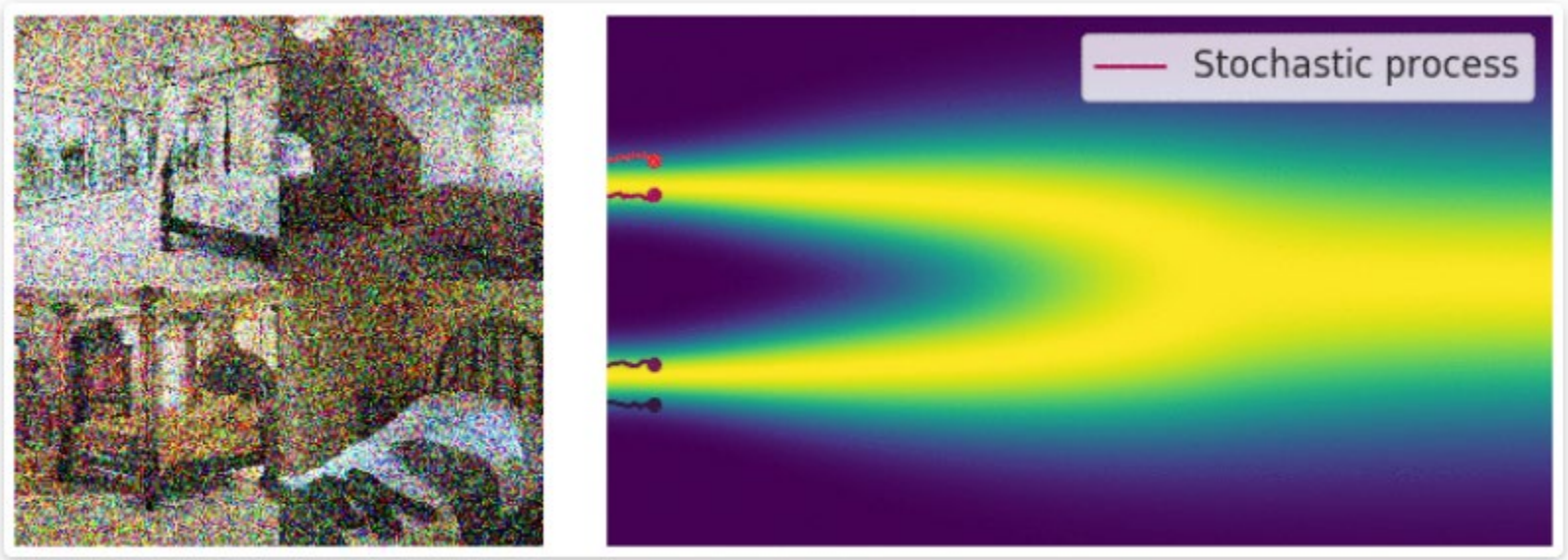
# Score-based SDE

- 提出在原始数据中加入无限数量的噪声尺度
- 噪声干扰过程是一个随时间连续的随机过程
- 定义连续时间变量$t \in [0, T]$



# Score-based SDE

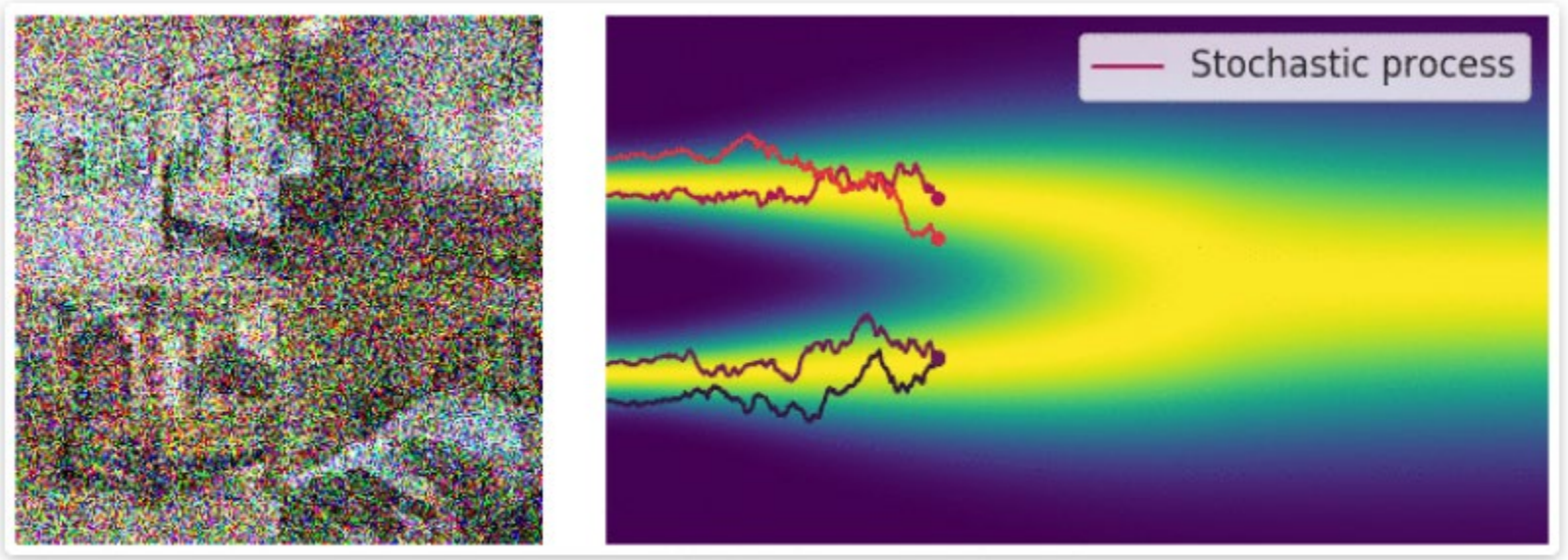
- 正向扩散：$t_1$时刻




Perturbing data to noise with a continuous-time stochastic process.

# Score-based SDE

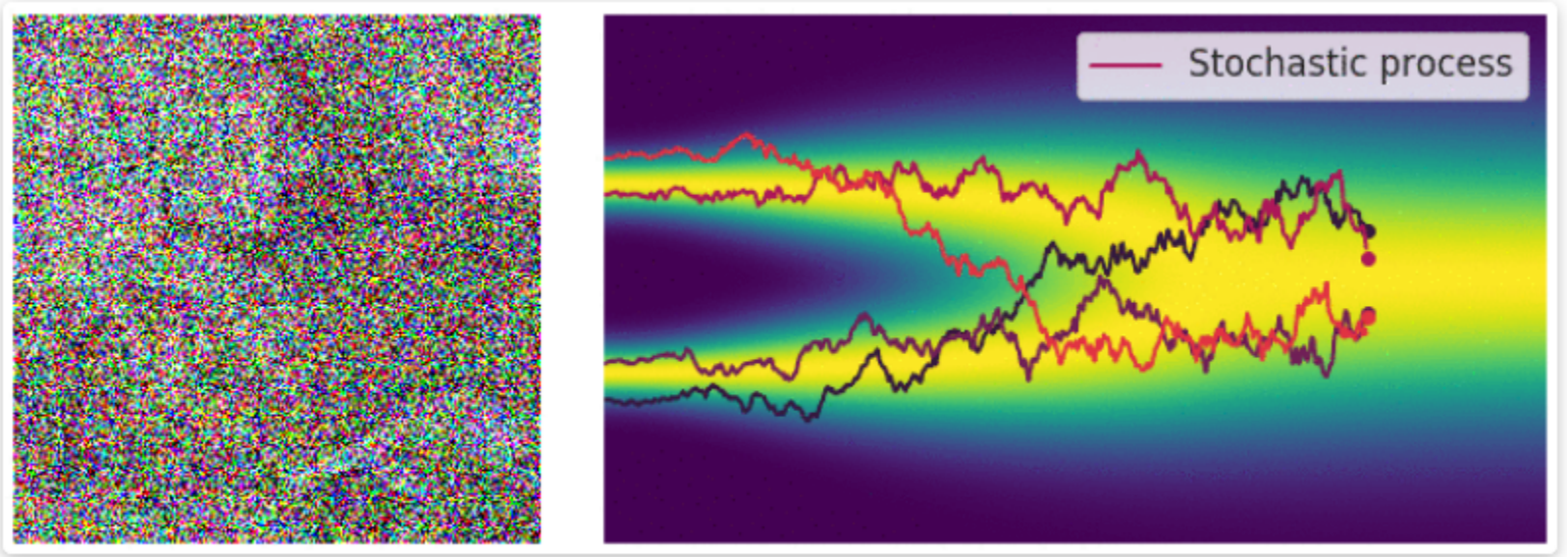
- 正向扩散： $t_2 > t_1$时刻




Perturbing data to noise with a continuous-time stochastic process.

# Score-based SDE

- 正向扩散： $t_3 > t_2 > t_1$时刻




Perturbing data to noise with a continuous-time stochastic process.



- 结论：随着随机过程的加深，原图的信息被大量的隐藏起来

# Score-based SDE

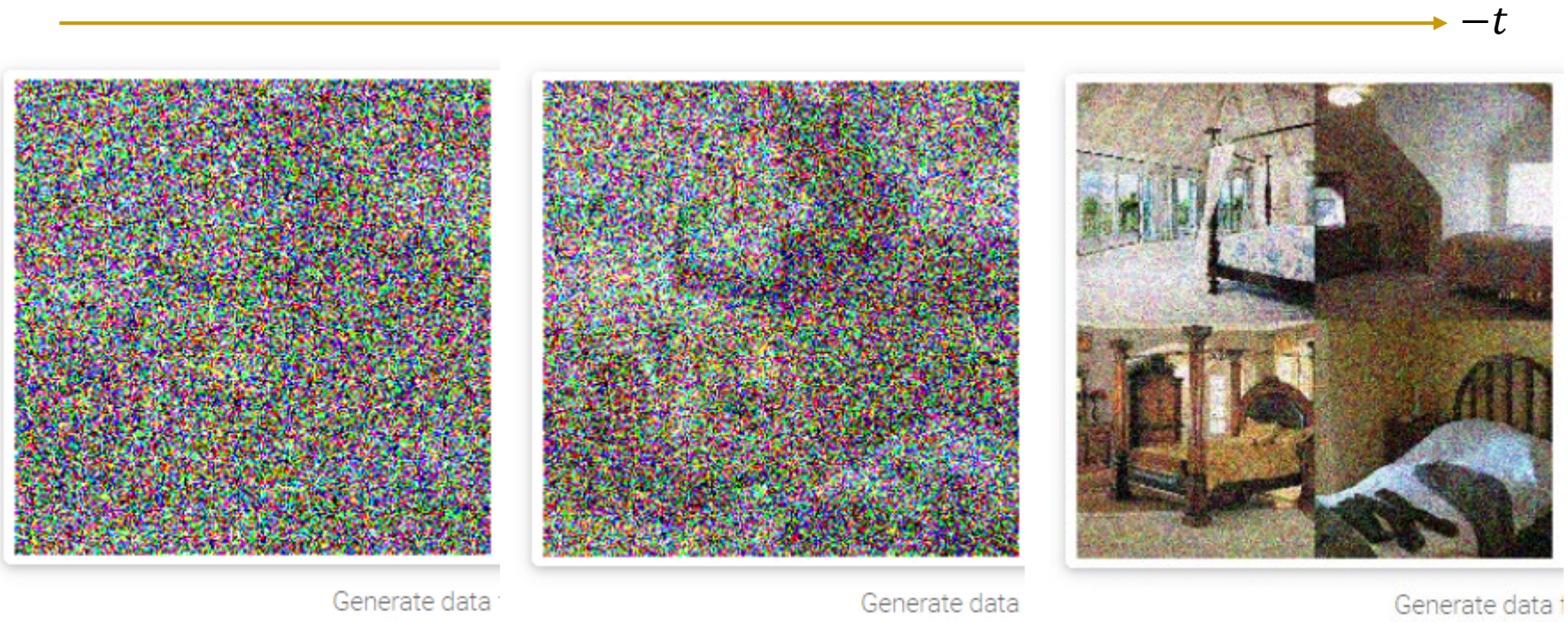
- 正向SDE

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

  - $w$则表示为标准的布朗运动
  - $dw$可以视为无穷小的白噪声
  - $g(t)$: $\mathbb{R} \mapsto \mathbb{R}$表示$x(t)$的扩散系数
  - $f(x,t)$: $\mathbb{R}^d \mapsto \mathbb{R}^d$表示$x(t)$的漂移系数
  - 该方程的解是一组连续的随机变量$\{x(t)\}_{t\in[0,T]}$
  - $p_t(x)$： $x(t)$的边缘概率密度函数
  - $p_{st}(x(t)|x(s))$： 转移核函数

# Score-based SDE

- 反向扩散：生成样本



# Score-based SDE

- 反向SDE

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}},$$

  - $\bar{w}$表示为标准的布朗运动，时间从$T$到0
  - $d\bar{w}$可以视为无穷小的白噪声
  - $dt$表示负的无穷小时间步
  - 估计$\nabla_x \log p_t(x)$?
    - 定义得分函数

# Score-based SDE

对原始数据$x(0)$加噪声得到数据$x(t)$,
这是加噪后的t时刻的得分模型

- 构造目标函数

$$\mathbb{E}_t\left\{\lambda(t)\mathbb{E}_{\mathbf{x}(0)}\mathbb{E}_{\mathbf{x}(t)|\mathbf{x}(0)}\left[\left\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}(t),t)-\nabla_{\mathbf{x}(t)}\log p_{0t}(\mathbf{x}(t)\mid\mathbf{x}(0))\right\|_2^2\right]\right\}$$

  - $\mathbb{E}_t$：对时间$t$连续泛化
  - $\lambda(t)$是一个正加权函数

$$\lambda \propto 1/\mathbb{E}\left[\left\|\nabla_{\mathbf{x}(t)}\log p_{0t}(\mathbf{x}(t)\mid\mathbf{x}(0))\right\|_2^2\right]$$

  - 转移核函数$p_{0t}(x(t)|x(0))$怎么选择？

# Score-based SDE

- 如果漂移系数$f(x,t)$是关于$x$的仿射变换
  - 转移核函数选择高斯核

    Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. 2019.
- 如果是一般的漂移系数$f(x,t)$
  - 柯尔莫哥洛夫向前方程 (Kolmogorov's forward equations)

    Bernt Øksendal. Stochastic differential equations. 2003.
  - Sliced score matching
    - 其中$\mathbb{E}[\mathrm{v}]=0$, $\mathbb{E}_t\left\{\lambda(t)\mathbb{E}_{\mathbf{x}(0)}\mathbb{E}_{\mathbf{x}(t)}\mathbb{E}_{\mathbf{v}\sim p_{\mathbf{v}}}\left[\frac{1}{2}\left\|\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}(t),t)\right\|_2^2+\mathbf{v}^{\mathrm{T}}\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}(t),t)\mathbf{v}\right]\right\}$

# Score-based SDE

- 对时间离散化后的特例
  - SMLD

    Y Song, et al. Generative modeling by estimating gradients of the data distribution. NIPS 2019
  - DDPM
    - 正向扩散的数据

$$\mathbf{x}_i = \sqrt{1-\beta_i}\mathbf{x}_{i-1} + \sqrt{\beta_i}\mathbf{z}_{i-1}, \quad i = 1, \cdots, N.$$

    - 当$N \to \infty$时，记$\bar{\beta}_i = N\beta_i$，$\beta\left(\frac{i}{N}\right) = \bar{\beta}_i$
    - 方差保持（VP）

$$\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$$

# Score-based SDE

- sub-VP SDE

$$\mathrm{d}\mathbf{x} = \boxed{-\frac{1}{2}\beta(t)\mathbf{x}}\,\mathrm{d}t + \boxed{\sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)\mathrm{d}s})}}\mathrm{d}\mathbf{w}.$$

漂移系数$f(x,t)$　　扩散系数$g(t)$

  - 仿射漂移系数$f(x,t)$
    - 转移核函数$p_{0t}(x(t)|x(0))$：高斯核

# Score-based SDE

- 求解反向SDE
  - 在训练得分模型$s_\theta(x)$之后，如何构建反向SDE，然后用数值方法对其进行模拟，从而生成样本

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}},$$

  - 时间离散化
    - Euler-Maruyama方法

      P. E. Kloeden, E. Platen. Numerical solution of stochastic differential equations, volume 23. Springer Science & Business Media, 2013.
    - DDPM：祖先采样 (Ancestral sampling)

$$\mathbf{x}_{i-1} = \frac{1}{\sqrt{1-\beta_i}}(\mathbf{x}_i + \beta_i \mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_i, i)) + \sqrt{\beta_i}\mathbf{z}_i$$

# Score-based SDE

- 反向扩散采样器（reverse diffusion samplers）
  - 时间离散化
    - 正向SDE

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{f}_i(\mathbf{x}_i) + \mathbf{G}_i\mathbf{z}_i, \quad i = 0, 1, \cdots, N-1$$

    - 反向SDE

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}(t)\mathbf{G}(t)^{\mathsf{T}}\nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathrm{d}t + \mathbf{G}(t)\mathrm{d}\bar{\mathbf{w}}$$

$$\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{f}_{i+1}(\mathbf{x}_{i+1}) + \mathbf{G}_{i+1}\mathbf{G}_{i+1}^{\mathsf{T}}\mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_{i+1}, i+1) + \mathbf{G}_{i+1}\mathbf{z}_{i+1}$$

  - 两个采样器；PC采样器和概率流ODE采样器

# Score-based SDE

- 预测器-校正器 (Predictor-Corrector, PC) 采样器
  - 预测器
    - 在每个时间步内，数值 SDE 求解器首先给出下一个时间步的样本估计值
  - 校正器
    - 基于得分的 MCMC 方法纠正了估计样本的边际分布

$$\mathbf{x}_i \leftarrow \mathbf{x}_i + \epsilon_i \mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_i, i) + \sqrt{2\epsilon_i}\mathbf{z}$$

# Score-based SDE

- 预测器-校正器 (Predictor-Corrector, PC) 采样器

**Algorithm 1** Predictor-Corrector (PC) sampling

**Require:**

$N$: Number of discretization steps for the reverse-time SDE

$M$: Number of corrector steps

Initialize $\mathbf{x}_N \sim p_T(\mathbf{x})$

**for** $i = N - 1$ **to** $0$ **do**

  $\mathbf{x}_i \leftarrow \text{Predictor}(\mathbf{x}_{i+1})$

  **for** $j = 1$ **to** $M$ **do**

    $\mathbf{x}_i \leftarrow \text{Corrector}(\mathbf{x}_i)$

**return** $\mathbf{x}_0$

# Score-based SDE

记为$\tilde{f}(x,t)$

- 确定性采样器：概率流常微分方程(ODE)
  - 对于所有的扩散过程，存在一个对应的确定过程

$$\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x},t) - \frac{1}{2}g(t)^2\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right]\mathrm{d}t$$

  - 本质上是一个神经常微分方程 (Neural ODE)
    - 构造了连续的归一化流 (normalized flow)
  - 优点：
    - 精确计算似然函数
    - 控制隐式表示：将任何数据$x(0)$编码到隐空间$x(T)$
    - 唯一可识别编码：插值计算
    - 高效采样

# Score-based SDE

- PC采样和概率流ODE结合

$$\mathbf{x}_i = (2 - \sqrt{1-\beta_{i+1}})\mathbf{x}_{i+1} + \frac{1}{2}\beta_{i+1}\mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_{i+1}, i+1)$$

**Algorithm 3** PC sampling (VP SDE)

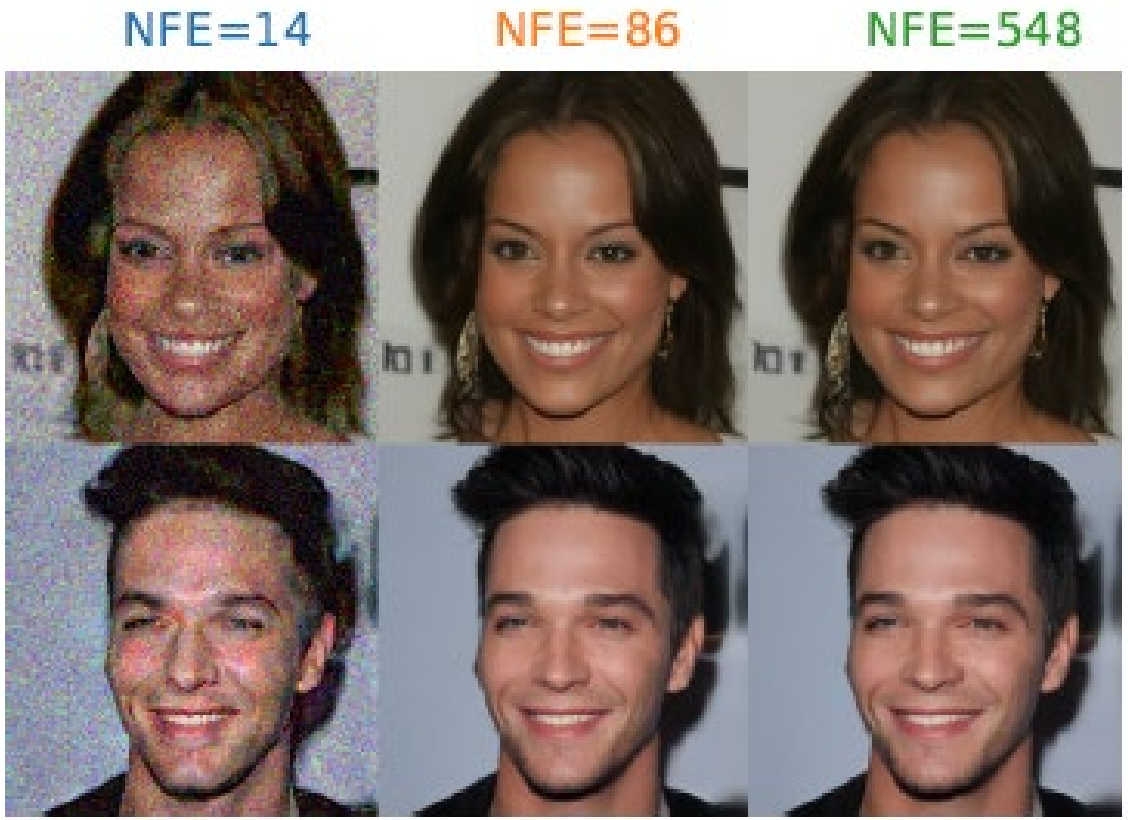
$\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
**for** $i = N-1$ **to** $0$ **do**
  $\mathbf{x}'_i \leftarrow (2 - \sqrt{1-\beta_{i+1}})\mathbf{x}_{i+1} + \beta_{i+1}\mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_{i+1}, i+1)$
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{x}_i \leftarrow \mathbf{x}'_i + \sqrt{\beta_{i+1}}\mathbf{z}$
  **for** $j = 1$ **to** $M$ **do**
    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{x}_i \leftarrow \mathbf{x}_i + \epsilon_i \mathbf{s}_{\boldsymbol{\theta}*}(\mathbf{x}_i, i) + \sqrt{2\epsilon_i}\mathbf{z}$
**return** $\mathbf{x}_0$

Predictor

Corrector

# Score-based SDE

- 得分函数评估(score function evaluations (NFE))



# Score-based SDE

- 插值计算

# Score-based SDE

- 可控生成

$$\mathrm{d}\mathbf{x} = \{\mathbf{f}(\mathbf{x}, t) - g(t)^2[\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \nabla_{\mathbf{x}} \log p_t(\mathbf{y} \mid \mathbf{x})]\}\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}.$$

前四行是车

后四行是马

# Score-based SDE

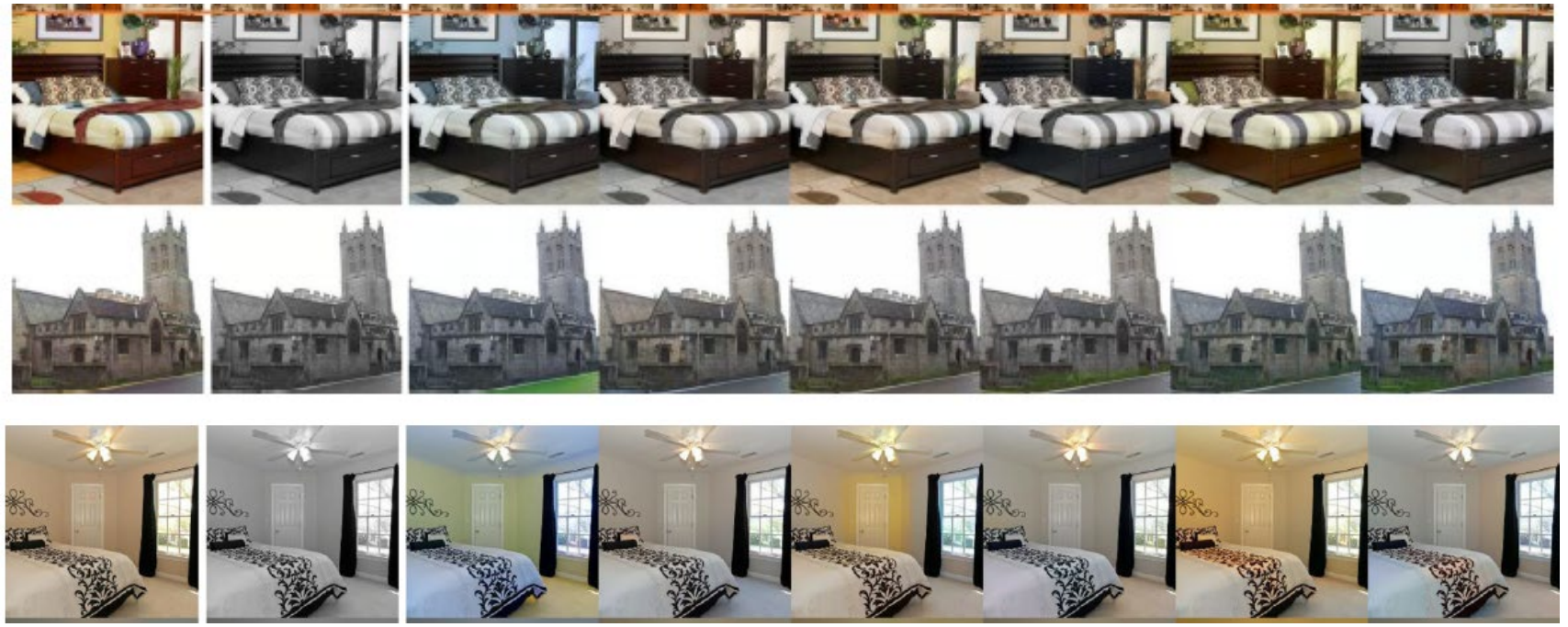
- 可控生成：图像修复

第一列是原始图像
第二列是蒙版图像
其余列是采样图像

# Score-based SDE

- 可控生成：图像着色

第一列是原始图像
第二列是灰度图像
其余列是采样图像



# Thanks!


**Questions?**