Score-based. Score diffusion. Model.

Score-based Generative Modeling through stochastic Differential Equation.

→ 是 DDPE 更一般的形式，可以用 SDE 统一.

① 扩散过程 → (前向过程) 没有参数.

$$\boxed{x_T} \rightarrow \cdots \rightarrow \boxed{x_t} \underset{q(x_t \mid x_{t-1})}{\overset{p_\theta(x_{t-1} \mid x_t)}{\rightleftarrows}} \boxed{x_{t-1}} \rightarrow \cdots \rightarrow x_0.$$

$X \sim N(\mu, \sigma^2)$. some Image.

Add small amount of Gaussian noise to the sample in T step, the data sample x_0 gradually loses its distinguishable features and the $T \to \infty$, X_T is equivalent to an isotropic Gaussian noise.

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I).$$

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

β_t is a variance schedule. $\{\beta \in (0,1)\}_{t=1}^{t}$.

任意时刻的 $q(x_t)$ 可以基于 x_0 和 β_t 计算:

Let $\alpha_t = 1 - \beta_t$. and $\bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$

利用重参数化技巧. $z_{t-1} \sim N(0, I)$

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t} * z_{t-1}$$

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

$X \sim N(\mu_1, \sigma_1)$

$Y \sim N(\mu_2, \sigma_2)$.

$aX + bY$ 叠加后的分布:

$Z \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

$$x_{t-1} = \sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, z_{t-2}.$$

$$\text{——} = \text{——}\; x_{t-2} + \sqrt{1-\alpha_t}\, z_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\, z_{t-2}.$$

$\sqrt{\alpha}$ … $\sqrt{\alpha_t \alpha_{t-1}}$ …

可以看作 $\sim N(0, (1-\alpha_{t-1} + 1-\alpha_t\alpha_{t-1}))$

$\sim N(0,$

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

$$\vdots$$

$$= \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, z.$$

$\prod_{t=1}^{T} \alpha_t$ 的连乘.

$$\therefore \boxed{q(x_t|x_0) = N(x_t; \sqrt{\bar\alpha}\, x_0, (1-\bar\alpha)I).}$$

② 反向过程：希望从随机噪声逐步还原 x_0.

(去噪过程) $\rightarrow p(x_0|x_T)$.

Given $q(x_t|x_{t-1})$ 是高斯分布，且 β_t 足够小. $q(x_{t-1}|x_t)$ 仍满足高斯分布

构建一个参数为θ的网络，拟合 $q(x_{t-1}|x_t, x_0)$.

$$= N(x_t, x_0; \mu_\theta(x_t, x_0), \Sigma_\theta(x_t, x_0))$$

$$p_\theta(x_0) = \int p_\theta(x_{0:T})\, d x_{1:T} \quad ?$$

$p_\theta(x_{0:T})$ → joint distribution called reversed process.

$x_{1:T}$ are latent of the same distribution

为什么用 q_θ 近似了 p_θ ??

$p_\theta(x_{0:T})$. defined as Markov chain. & $p(x_T) = N(0, I)$.

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

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

$\because$ Markov chain. ignore x_0

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

$$N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= N(x_{t-1}; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \cdot \frac{N(x_0, \sqrt{\bar\alpha_{t-1}}\, x_0, \sqrt{1-\bar\alpha_{t-1}} I)}{N(x_0, \sqrt{\bar\alpha_t}\, x_0, \sqrt{1-\bar\alpha_t} I)}$$

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

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

$$\frac{x_{t-1}^2}{1-\bar\alpha_{t-1}}$$

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

$$= \exp\left(-\frac{1}{2}\left(\overbrace{\left(\frac{1}{\beta_t} - \frac{1}{1-\bar\alpha_t}\right) x_t^2}^{const} + \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar\alpha_{t-1}}\right) x_{t-1}^2 + \overbrace{\left(\frac{\bar\alpha_{t-1}}{1-\bar\alpha_{t-1}} - \frac{\bar\alpha_t}{1-\bar\alpha_t}\right) x_0^2}^{const} - \left(\frac{2\sqrt{1-\beta_t}}{\beta_t} x_t + \frac{2\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_{t-1}} x_0\right) x_{t-1} + \underbrace{\frac{2\sqrt{\bar\alpha_t}}{1-\bar\alpha_t} x_t x_0}_{const}\right)\right)$$

→ 因为 given x_t, x_0.

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

$\boxed{ax^2+b}$ 可以看成 $a(x+\frac{b}{2a})^2+C$，均值为 $-\frac{b}{2a}$，方差 $\frac{1}{a}$.

$$\therefore q(x_{t-1}|x_t,x_0) = N(x_t, x_0; \mu_\theta(x_t,x_0), \Sigma_\theta(x_t,x_0)).$$

$$\mu = \left(\frac{\sqrt{\alpha_t}}{\beta_t} x_t + \frac{2\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_{t-1}} x_0\right) \Big/ \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar\alpha_{t-1}}\right) = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} x_t + \frac{\sqrt{\bar\alpha_{t-1}}\,\beta}{1-\bar\alpha_t} x_0$$

$$\Sigma = \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar\alpha_{t-1}}\right)^{-1} = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t$$ const.

$$\mu = \frac{\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_{t-1}}x_0}{\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar\alpha_{t-1}}} = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})x_t + \beta_t\sqrt{\bar\alpha_{t-1}}\,x_0}{\alpha_t(1-\bar\alpha_{t-1}) + \beta_t}$$

$$= \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})x_t + \beta_t\sqrt{\bar\alpha_{t-1}}\,x_0}{\alpha_t - \bar\alpha_t + (1-\alpha_t)}$$

$$= \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})x_t + \beta_t\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_t}$$

可以简化为:

$$\boxed{= \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t + \frac{\beta_t\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_t}x_0}$$

这里的 x_0 其实是要生成的样本

x_0 可以用 x_t 表示: $x_t = \sqrt{\bar\alpha_t}\,x_0 + \sqrt{1-\bar\alpha_t}\,z_t$

$$\boxed{x_0 = \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,z_t\right)}$$

所以

$$\mu = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t + \frac{\beta_t\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_t}\cdot\frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,z_t\right)$$

$$\frac{\beta_t\sqrt{\bar\alpha_{t-1}}\left(x_t - \sqrt{1-\bar\alpha_t}\,z_t\right)}{\sqrt{\bar\alpha_t}(1-\bar\alpha_t)}$$

$$= \left(\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} + \frac{\beta_t\sqrt{\bar\alpha_{t-1}}}{\sqrt{\bar\alpha_t}(1-\bar\alpha_t)}\right)x_t - \frac{\beta_t\sqrt{\bar\alpha_{t-1}(1-\bar\alpha_t)}}{\sqrt{\bar\alpha_t}(1-\bar\alpha_t)}z_t$$

$$= \frac{\sqrt{\alpha_t\bar\alpha_t}(1-\bar\alpha_{t-1}) + (1-\alpha_t)\sqrt{\bar\alpha_{t-1}}}{\sqrt{\bar\alpha_t}\,(1-\bar\alpha_t)}x_t - \frac{(1-\alpha_t)\sqrt{\bar\alpha_{t-1}-\bar\alpha_t\bar\alpha_{t-1}}}{\sqrt{\bar\alpha_t}(1-\bar\alpha_t)}z_t$$

$$\frac{\sqrt{\alpha_t\bar\alpha_t} - \bar\alpha_{t-1}\sqrt{\alpha_t\bar\alpha_t} + \sqrt{\bar\alpha_{t-1}} - \alpha_t\sqrt{\bar\alpha_{t-1}}}{\sqrt{\bar\alpha_t}(1-\bar\alpha_t)} - \frac{(1-\alpha_t)\sqrt{1-\bar\alpha_t}}{\sqrt{\alpha_t}\,(1-\bar\alpha_t)}z_t$$

$$\frac{\sqrt{\alpha_t\bar\alpha_{t-1}\alpha_t} - \bar\alpha_{t-1}\sqrt{\alpha_t\bar\alpha_{t-1}\alpha_t} + \sqrt{\bar\alpha_{t-1}} - \alpha_t\sqrt{\bar\alpha_{t-1}}}{\sqrt{\bar\alpha_{t-1}}\left(\cancel{\alpha_t} - \bar\alpha_t + 1 - \cancel{\alpha_t}\right)} - \frac{\beta_t}{\sqrt{\alpha_t(1-\bar\alpha_t)}}z_t$$

$$= \frac{\sqrt{\bar\alpha_{t-1}}\,(\cancel{1-\bar\alpha_t})}{\sqrt{\bar\alpha_t}\,(\cancel{1-\bar\alpha_t})}$$

$$= \frac{1}{\sqrt{\alpha_t}} x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} z_t$$

z_t 其实是由 $x_0 \to x_t$ 时的噪音 ε.

变成仅与 x_t 和 z_t 有关

$$\mu = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_t \right).$$

前向时刻的 z_t 可以作为标签.

把 Transformer 和 Unet 结合起来.

$q(x_{t-1} | x_t) \sim (\mu, \Sigma)$. $\quad x_{t-1} = \tilde{\Sigma} x_t + \mu_\theta(x_t, z_t)$.

---

③ 计算 Loss.

Usually optimize log likelihood:

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

即 $x_0 \sim q(x_0)$. 与 $p_\theta(x_0)$ 的交叉熵.

$$-\int_{x_0} q(x_0) \log p_\theta(x_0)\, dx_0$$

For equation (3):

$$-\log p_\theta(x_0) \le -\log p_\theta(x_0) + D_{KL}\left(q(x_{1:T} | x_0) \,\|\, p_\theta(x_{1:T} | x_0)\right).$$

$$= -\log p_\theta(x_0) + \int_x q(x_{1:T} | x_0) \log \frac{q(x_{1:T} | x_0)}{p_\theta(x_{1:T} | x_0)}\, dx.$$

$$= -\log p_\theta(x_0) + E_{x_{1:T} \sim q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T} | x_0)}{p_\theta(x_{0:T}) / p_\theta(x_0)}\right].$$

$p_\theta(x_{1:T} | x_0) = \frac{p_\theta(x_{0:T})}{p(x_0)}$.

$p_\theta(x_{0:T}) = p(x_{1:T} | x_0) \cdot p(x_0)$ chain rule.

$$= \cancel{-\log p_\theta(x_0)} + E_q\left[\log \frac{q(x_{1:T} | x_0)}{p_\theta(x_{0:T})} + \underbrace{\log p_\theta(x_0)}_{\text{与 } q \text{ 无关}}\right]$$

$$= E_q\left[\log \frac{q(x_{1:T} | x_0)}{p_\theta(x_{0:T})}\right].$$

取期望 $E_{q(x_0)}$，Fubini 定理. 交叉熵的上界

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

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

$$= E_q \left[-\log p(x_T) - \sum_{t=1} \log \frac{p_\theta(x_{t-1} | x_t)}{q(x_t | x_{t-1})} \right]$$

---

for equation (5):

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

Markov chain.

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

$$= E_q \left[\log \frac{\prod_{t=1}^{T} p_\theta(x_{t-1} | x_t) \cdot p_\theta(x_T)}{\prod_{t=1}^{T} q(x_t | x_{t-1})} \right]$$

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

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

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

$$= E_q \Big[\underbrace{D_{KL}(q(x_T | x_0) \| p_\theta(x_T))}_{L_T \to \text{const}} + \sum_{t=2}^{T} \underbrace{D_{KL}(q(x_{t-1} | x_t, x_0) \| p_\theta(x_{t-1} | x_t))}_{L_{t-1}} - \underbrace{\log p_\theta(x_0 | x_1)}_{L_0} \Big]$$

$L_{1:T}$ 可以看作 $q(x_{t-1} | x_t, x_0) = N(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$ 和

$p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta)$ 之间的KL散度.

$\hookrightarrow \sigma_t^2$. const.

$$\therefore L_{t-1} = E_q \left[\frac{1}{2\sigma_t^2} \| \tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t) \|^2 + C \right].$$

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

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z_t. \qquad x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\, z_t\right).$$

$$\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t) := \tilde{\mu}_t\left(x_t, \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\, z_t\right)\right) - \mu_\theta(x_t, t).$$

即 given x_t, predict z_t.

---

② 逆扩散过程 → (Reverse process).

Hypothesis: the reverse process: $q(x_{t-1} | x_t)$.

If the β_t is small enough, $q(x_{t-1} | x_t)$ can also be Gaussian.

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

③ Loss:

$$L = E_q\big(D_{KL}(q(x_T \| x_0) \,\|\, p_\theta(x_T)\big)$$ → 无参，仅与 β_t 有关.

$$+ \sum_{t=2}^{T} D_{KL}\big(q(x_{t-1} | x_t, x_0) \,\|\, p_\theta(x_{t-1} | x_t)\big) - \log p_\theta(x_0 | x_t).$$

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