Diffusion

参考：扩散模型之DDPM - 知乎

Step 1 前向过程(加噪)
Step 2 逆向过程(去噪)
Step 3 最大似然
Step 4 Sampling(推理)
扩展

> 苏剑林老师的 **博客** 里面对扩散模型的前向过程和反向过程有个很好的比喻——“拆楼 + 建楼”：自然图像相当于是一个“盖好的楼”，噪声相当于“水泥沙子”；前向模型相当于“拆楼”的过程，即不断地把一个楼拆成剩余的楼体和水泥沙子,直到全部变成水泥沙子（接近纯高斯噪声图）；而反向模型则相反，相当于从一堆的水泥沙子（高斯噪声），不断地去构建出楼体，直到最终建出一栋大楼出来。

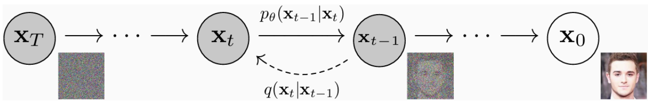

Figure 2: The directed graphical model considered in this work.

Step 1 前向过程(加噪)

$$q(x_t|x_{t-1}) \sim \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}), q(x_{1:T}|x_0) = \prod_{t=1}^{t=T} q(x_t|x_{t-1})$$

这里假定前向加噪过程满足Markov性质

利用重参数化有:

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon_{t-1} = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

由此我们可以得到任意时刻x_t的表示:

$$\begin{aligned}
x_t &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\epsilon}_{t-2} \\
&= \cdots \\
&= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon; \\
\bar{\alpha}_t &= \prod_{i=1}^{i=t}\alpha_i, \epsilon \sim \mathcal{N}(0, \mathbf{I})
\end{aligned}$$

Step 2 逆向过程(去噪)

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

这里$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$应当拟合真实的逆向分布$q(x_{t-1}|x_t, x_0)$，推导如下

$$
\begin{aligned}
q(x_{t-1}|x_t,x_0) &= \frac{q(x_{t-1},x_t,x_0)}{q(x_t,x_0)} \\
&= \frac{q(x_0)q(x_{t-1}|x_0)q(x_t|x_{t-1},x_0)}{q(x_0)q(x_t|x_0)} \\
&= q(x_t|x_{t-1},x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \\
&= \mathcal{N}(x_{t-1};\tilde{\mu}_t(x_t,x_0),\tilde{\Sigma}_t(x_t,x_0))
\end{aligned}
$$

其中

$$
\begin{aligned}
\tilde{\mu}_t(x_t,x_0) &= \frac{\sqrt{\alpha}_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 \\
&= \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t) \\
\tilde{\Sigma}_t(x_t,x_0) &= \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t\mathbf{I}
\end{aligned}
$$

方差与x_t无关，因此不用NN来拟合，直接用就可以；均值用参数为θ的NN来拟合

Step 3 最大似然

我们的目标是最大化似然函数$p_\theta(x_0)$，但有隐变量$x_{1:T}$，因此使用变分推断

$$
\begin{aligned}
\log p_\theta(x_0) &\geq \mathbb{E}_{q(x_{1:T}|x_0)} \log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \log \frac{p_\theta(x_T)p_\theta(x_0|x_1)\prod_{t=2}^T p_\theta(x_{t-1}|x_t)}{q(x_1|x_0)\prod_{t=2}^T q(x_t|x_{t-1},x_0)} \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_T)p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \prod_{t=2}^T \frac{p_\theta(x_{t-1}|x_t)}{\frac{q(x_{t-1}|x_t,x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}}\right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_T)p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \prod_{t=2}^T \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} + \log \prod_{t=2}^T \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t,x_0)}\right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_T)p_\theta(x_0|x_1)}{q(x_T|x_0)} + \log \prod_{t=2}^T \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t,x_0)}\right] \\
&= \mathbb{E}_{q(x_1|x_0)} \log p_\theta(x_0|x_1) + \mathbb{E}_{q(x_T|x_0)} \log \frac{p_\theta(x_T)}{q(x_T|x_0)} + \sum_{t=2}^T \mathbb{E}_{q(x_t,x_{t-1}|x_0)} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t,x_0)} \\
&= \mathbb{E}_{q(x_1|x_0)} \log p_\theta(x_0|x_1) + \mathbb{E}_{q(x_T|x_0)} \log \frac{p_\theta(x_T)}{q(x_T|x_0)} + \sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}\mathbb{E}_{q(x_{t-1}|x_t,x_0)} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t,x_0)} \\
&= \underbrace{\mathbb{E}_{q(x_1|x_0)} \log p_\theta(x_0|x_1)}_{\text{reconstruction term}} - \underbrace{\mathrm{KL}[q(x_T|x_0)\|p_\theta(x_T)]}_{\text{prior matching term}} - \sum_{t=2}^T \underbrace{\mathbb{E}_{q(x_t|x_0)}\mathrm{KL}[p_\theta(x_{t-1}|x_t)\|q(x_{t-1}|x_t,x_0)]}_{\text{denoising matching term}}
\end{aligned}
$$

由于

$$
\begin{aligned}
&\mathrm{KL}[\mathcal{N}(\mu_1,\Sigma_1)\|\mathcal{N}(\mu_2,\Sigma_2)] = \\
&\frac{1}{2}\left[\log\frac{|\Sigma_2|}{|\Sigma_1|} - d + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^{\mathrm{T}}\Sigma_2^{-1}(\mu_2-\mu_1)\right]
\end{aligned}
$$

因此有

$$
\begin{aligned}
\mathrm{KL}[p_\theta(x_{t-1}|x_t)\|q(x_{t-1}|x_t,x_0)] &= \frac{1}{2}(\tilde{\mu}_t-\mu_\theta)^{\mathrm{T}}\tilde{\Sigma}_t^{-1}(\tilde{\mu}_t-\mu_\theta) \\
&= \frac{1}{2\tilde{\sigma}_t^2}\|\tilde{\mu}_t-\mu_\theta\|_2^2
\end{aligned}
$$

注意到$\tilde{\mu}_t$的表达式中x_0,x_t均为已知，因此为了简化计算，我们让NN拟合图像x_0

$$
\mu_\theta(x_t,x_0)=\frac{\sqrt{\alpha}_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}f_\theta(x_t,t)
$$

于是

$$
\mathrm{KL}[p_\theta(x_{t-1}|x_t)\|q(x_{t-1}|x_t,x_0)]=\frac{1}{2\tilde{\sigma}_t^2}\frac{\bar{\alpha}_{t-1}\beta_t^2}{(1-\bar{\alpha}_t)^2}\|f_\theta(x_t,t)-x_0\|_2^2
$$

与原文不同，原文让NN拟合噪声

$$
\mu_\theta(x_t,x_0)=\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t))
$$

则有

$$
\mathrm{KL}[p_\theta(x_{t-1}|x_t)\|q(x_{t-1}|x_t,x_0)]=\frac{1}{2\tilde{\sigma}_t^2}\frac{(1-\alpha_t)^2}{\alpha_t(1-\bar{\alpha}_t)}\|\epsilon_\theta(x_t,t)-\epsilon_t\|_2^2
$$

最终，我们得到NN的Loss为

$$
\sum_{t=2}^{T}\mathbb{E}_{x_0,\epsilon}\left[\frac{1}{2\tilde{\sigma}_t^2}\frac{(1-\alpha_t)^2}{\alpha_t(1-\bar{\alpha}_t)}\|\epsilon_\theta(x_t,t)-\epsilon_t\|_2^2\right]
$$

原文将系数舍去并将求和改为对时间t采样

$$
\begin{aligned}
\mathcal{L}_{\text{simple}} &= \mathbb{E}_{x_0,\epsilon,t}\left[\|\epsilon_\theta(x_t,t)-\epsilon\|_2^2\right] \\
&= \mathbb{E}_{x_0,\epsilon,t}\left[\|\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon,t)-\epsilon\|_2^2\right]
\end{aligned}
$$

注意，这里ϵ_t,ϵ均表示$x_0\to x_t$过程中加的高斯噪声，也就是

$$
\begin{aligned}
x_t &= \sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon \\
&= \sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_t
\end{aligned}
$$

Step 4 Sampling(推理)

1. $x_T\sim\mathcal{N}(0,\mathbf{I})$
2. for $t=T,\ldots,1$ do
   a. $z\sim\mathcal{N}(0,\mathbf{I})$
   b. $x_{t-1}=\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t))+\tilde{\sigma}_t z$
3. return x_0

扩展

🚍 Score-Matching,SDE,Langevin Dynamics Sampling

💯 DPM-Solver

🧯 DiT

⏳ Flow Matching