3-12周报

本周工作:

- 1. Foundation Model 调研任务;
- 2. 点云 (和 Mesh) 数据集收集;
- 3. 寻找梯度估计的灵感: 收集传统点云降噪论文和与3D梯度有关的深度论文;
- 4. 完善公式推导,找到了一个可尝试的修正了参数的新loss。
- 5. 对模型部分代码优化,在不影响效率下大大降低了GPU显存占用;

周报内容是对工作4的阐述。

关于上周提出的 **对于** $x_a^{(t)}$ **生成策略的修改**,训练实验结果显示不合理,且基于 Nearest 生成的结果显然不通过 Jarque-Bera 检验。 (注:原方法的中间变量均通过JB检验)

1. 公式推导与分析

对部分符号进行重定义,为了降低公式编写难度,例如 $x^{(t)}$ 重定义为了 x^t 。

过去的 Loss 的依据是 Score-based,但 Diffusion的随机过程和 Langevin 过程存在区别,因此具体内容的计算上存在不同。

1.1. SBD Loss推导

定义 Diffusion Process 的分布描述:

$$q(x^{1:T}|x^0) = \prod_{t=1}^T q(x^t|x^{t-1}), \; q(x^t|x^{t-1}) = \mathcal{N}(x^t; \; \sqrt{1-eta_t}x^{t-1}, \; eta_t ext{I})$$

定义 Sampling Process 的分布描述: (带 θ 的为需要训练的

$$p_{ heta}(x^{0:T}|F_T) = p(x^T)\prod_{t=1}^T p_{ heta}(x^{t-1}|x^t, F_T)$$
 $where \ p_{ heta}(x^{t-1}|x^t, F_T) = \mathcal{N}(x^t; \ \sqrt{1-eta_t}x^{t-1}, \ eta_t ext{I}), \ p(x^T) = \mathcal{N}(x^T; x^0, L_{noise} ext{I}^3)$

其中, $F_T = EdgeFeature(x^T)$

训练方法 p_{θ} 使用它的负对数似然估计的变分上界:(**绿色部分**为相比于上行公式**修改的部分**,目的是降低阅读难度)

$$egin{aligned} \mathbb{E}[-\log p_{ heta}(x^0)] &\leq \mathbb{E}_qigg[-\log rac{p_{ heta}(x^{0:T}|F_T)}{q(x^{1:T}|x^0)}igg] \ &= E_qigg[-\log p(x^T) - \sum_{t \geq 1} \log rac{p_{ heta}(x^{t-1}|x^t,F_T)}{q(x^t|x^{t-1})}igg] \ &= E_qigg[-\log p(x^T) - \sum_{t \geq 1} \log rac{p_{ heta}(x^{t-1}|x^t,F_T)}{q(x^t|x^{t-1})} - \log rac{p_{ heta}(x^0|x^1,F_T)}{q(x^1|x^0)}igg] \end{aligned}$$

根据贝叶斯定理: $q(x^t|x^{t-1})=\dfrac{q(x^{t-1}|x^t)q(x^t)}{q(x^{t-1})}$,但是 $q(x^{t-1}|x^t)$ 是无法直接处理的。又因 $q(x^{1:T}|x^0)$ 令 $q(x^t|x^{t-1},x^0)=q(x^t|x^{t-1})$ 满足,因此引入 x^0 作为条件 $q(x^t|x^{t-1})=q(x^t|x^{t-1},x^0)=\dfrac{q(x^{t-1}|x^t,x^0)q(x^t|x^0)}{q(x^{t-1}|x^0)}.$

$$= E_q \bigg[-\log p(x^T) - \sum_{t \geq 1} \log \frac{p_{\theta}(x^{t-1}|x^t, F_T)}{q(x^{t-1}|x^t, x^0)} \cdot \frac{q(x^{t-1}|x^0)}{q(x^t|x^0)} - \log \frac{p_{\theta}(x^0|x^1, F_T)}{q(x^1|x^0)} \bigg]$$

$$\therefore \sum_{t>1} \log \left[rac{q(x^{t-1}|x^0)}{q(x^t|x^0)}
ight] = \log q(x^1|x^0) - \log q(x^T|x^0)$$

$$egin{aligned} \therefore &= E_qigg[-\lograc{p(x^T)}{q(x^T|x^0)} - \sum_{t>1}\lograc{p_ heta(x^{t-1}|x^t,F_T)}{q(x^{t-1}|x^t,x^0)} - \log p_ heta(x^0|x^1,F_T)igg] \ &= E_qigg[D_{KL}(q(x^T|x^0)\ ||\ p(x^T)) + \sum_{t>1}D_{KL}(q(x^{t-1}|x^t,x^0)\ ||\ p_ heta(x^{t-1}|x^t,F_T)) - \log p_ heta(x^0|x^1,F_T)igg] \ &= D_{KL}(q(x^T|x^0)\ ||\ p(x^T)) + E_qigg[\sum_{t>1}D_{KL}(q(x^{t-1}|x^t,x^0)\ ||\ p_ heta(x^{t-1}|x^t,F_T))igg] - \log p_ heta(x^0|x^1,F_T) \ \end{aligned}$$

=: loss

其中,红色项显然是常数项,对Loss的下降并不会起到任何作用,因此带 F_T 的 Diffusion 最优化问题可描述为:

$$egin{aligned} Simplify &\Rightarrow loss = E_qigg[\sum_{t>1} D_{KL}(q(x^{t-1}|x^t,x^0) \mid\mid p_{ heta}(x^{t-1}|x^t,F_T))igg] \ &\Leftrightarrow rg\min_{ heta} E_qigg[\sum_{t>1} D_{KL}(q(x^{t-1}|x^t,x^0) \mid\mid p_{ heta}(x^{t-1}|x^t,F_T))igg] \end{aligned}$$

结论:

• 引入 F_T 只影响 $q(x^{t-1}|x^t,x^0)$ 和 $p_{ heta}(x^{t-1}|x^t,F_T)$ 的相对熵;

1.1.1. 引入 Score-based 得到目标 Loss

对于 Diffusion process 来说:

$$q(x^{t-1}|x^t,x^0) = \mathcal{N}(x^{t-1}; ilde{\mu}(x^t,x^0), ilde{eta}_t)$$

分解得到: ($note:areta_t=1-arlpha_t,arlpha_t=\prod_{i=1}^Tlpha_i,lpha_t=1-eta_t$)

$$egin{aligned} q(x^{t-1}|x^t,x^0) &= q(x^t|x^{t-1},x^0) rac{q(x^{t-1}|x^0)}{q(x^t|x^0)} \ &= \mathcal{N}(x^t;\sqrt{lpha_t}x^{t-1},eta_t \mathrm{I}) rac{\mathcal{N}(x^{t-1};\sqrt{ar{lpha}_{t-1}}x^0,ar{eta}_{t-1} \mathrm{I})}{\mathcal{N}(x^t;\sqrt{ar{lpha}_t}x^0,ar{eta}_t \mathrm{I})} \ &\propto \exp\left(-rac{1}{2}(rac{(x^t-\sqrt{lpha_t}x^{t-1})^2}{eta_t} + rac{(x^{t-1}-\sqrt{ar{lpha}_{t-1}}x^0)^2}{ar{eta}_{t-1}} + rac{(x^t-\sqrt{ar{lpha}_t}x^0)^2}{ar{eta}_t})
ight) \ &= \exp\left(-rac{1}{2}((rac{lpha_t}{eta_t} + rac{1}{ar{eta}_{t-1}})(x^{t-1})^2 - 2(rac{\sqrt{lpha_t}}{eta_t}x^t + rac{\sqrt{ar{lpha}_{t-1}}}{ar{eta}_{t-1}}x^0)x^{t-1} + C(x^t,x^0))
ight) \end{aligned}$$

由此可以解得:

$$egin{aligned} ilde{eta}_t &= rac{1}{rac{lpha_t}{eta_t} + rac{1}{ar{eta}_{t-1}}} = rac{ar{eta}_{t-1}}{ar{eta}_t} eta_t \ & ilde{eta}_t &= rac{rac{\sqrt{lpha_t}}{eta_t} x^t + rac{\sqrt{ar{lpha}_{t-1}}}{ar{eta}_{t-1}} x^0}{rac{lpha_t}{eta_t} + rac{1}{ar{eta}_{t-1}}}{rac{eta_t}{eta_t}} &= rac{\sqrt{lpha_t} ar{eta}_{t-1}}{ar{eta}_t} x^t + rac{\sqrt{ar{lpha}_{t-1}} eta_t}{ar{eta}_t} x^0 \ & Reparameterizing \Rightarrow rac{1}{\sqrt{lpha_t}} (x^t (x^0, z) - rac{eta_t}{\sqrt{ar{eta}_t}} z), z \sim \mathcal{N}(0, I) \end{aligned}$$

其中,
$$x^t(x^0,z)=\sqrt{ar{lpha}_t}x^0+\sqrt{ar{eta}_t}z$$

这里开始,我**引入 Score-based** 作为 p_{θ} 中计算梯度的模型,把上面的最优化问题进行限制。定义 Score-based 计算梯度:

$$\sqrt{ar{lpha}_t}
abla_x log[s_{ heta}(x_a^{t-1}|x_a^t, F_T)] pprox -z_{ heta}(x^t, F_T) \propto \min\{||x_i^0 - x_a^t||_2^2 \ |x_i^0 \in x^0\}, \ x_a^t = rac{x^t}{\sqrt{\overline{lpha}_t}}$$

Sampling Process 同理得到:

$$p_{ heta}(x^{t-1}|x^t, F_T) = \mathcal{N}(x^{t-1}; \mu_{ heta}(x^t, t, F_T), \Sigma_{ heta}(x^t, t)), \Sigma_{ heta}(x^t, t) = \sigma^2 \mathbf{I} = \tilde{eta}_t \mathbf{I}$$

$$\Rightarrow \mu_{ heta}(x^t, t, F_T) = \frac{1}{\sqrt{\alpha_t}} (x^t(x^0, z) - \frac{eta_t}{\sqrt{ar{eta}_t}} z_{ heta}(x^t, F_T))$$

对于两个高斯分布的 KL 散度来说(借鉴VAE的推导),我们可以得到如下推导:

$$\begin{split} D_{KL}(q \mid\mid p_{\theta}) &= D_{KL}(\mathcal{N}(x^{t-1}; \tilde{\mu}(x^{t}, x^{0}), \tilde{\beta}_{t}) \mid\mid \mathcal{N}(x^{t-1}; \mu_{\theta}(x^{t}, t, F_{T}), \Sigma_{\theta}(x^{t}, t))) \\ &= \frac{1}{2}(n + \frac{1}{\tilde{\beta}_{t}^{2}} \mid\mid \tilde{\mu}(x^{t}, x^{0}) - \mu_{\theta}(x^{t}, t, F_{T}) \mid\mid^{2} - n + \log 1) \\ &= \frac{1}{2\tilde{\beta}_{t}^{2}} \mid\mid \left(\frac{1}{\sqrt{\alpha_{t}}} (x^{t}(x^{0}, z) - \frac{\beta_{t}}{\sqrt{\tilde{\beta}_{t}}} z)\right) - \left(\frac{1}{\sqrt{\alpha_{t}}} (x^{t}(x^{0}, z) - \frac{\beta_{t}}{\sqrt{\tilde{\beta}_{t}}} z_{\theta}(x^{t}, F_{T}))\right) \mid\mid_{2}^{2} \\ &= \frac{1}{2\tilde{\beta}_{t}^{2}} \mid\mid \left|\frac{\beta_{t}}{\sqrt{\tilde{\beta}_{t}}} \left(z_{\theta}(x^{t}, F_{T}) - z\right)\right\mid\mid_{2}^{2} \\ &: : p(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^{n}|\Sigma|}} e^{-\frac{(x-\mu)^{T}\Sigma^{-1}(x-\mu)}{2}} \propto p(x; \mu, \sigma^{2}) = \frac{1}{\sqrt{(2\pi)^{n}\sigma}} e^{-\frac{(x-\mu)^{2}}{2\sigma^{2}}}, when \Sigma = \sigma^{2}I \\ &: : \Rightarrow \frac{1}{2\tilde{\beta}_{t}^{2}} \mid\mid \frac{\beta_{t}\sqrt{\tilde{\alpha}_{t}}}{\sqrt{\tilde{\beta}_{t}}} \left(\nabla_{x}log[s_{\theta}(x_{a}^{t-1}|x_{a}^{t}, F_{T})] - \nabla_{x}q(x_{a}^{t})\right) \mid\mid_{2}^{2} \end{split}$$

其中, $\nabla_x q(x_a^t)$ 为我们通过算法估计的梯度方向,目前是个待改善内容。

综上所述:

$$\mathcal{L}(x_a^{0:T},\{eta_i\}_{i=1}^T) = \sum_{t>1} rac{1}{2 ilde{eta}_t^2} \mathbb{E}_qigg[igg| rac{eta_t\sqrt{ar{lpha}_t}}{\sqrt{ar{eta}_t}} \Big(
abla_x log[s_ heta(x_a^{t-1}|x_a^t,F_T)] -
abla_x q(x_a^t) \Big) igg| igg|_2^2 igg]$$

结论:

• loss形式上和当前使用的loss一致,但细节参数不同,下周会对这个新loss进行验证。