

基于粗糙集的去噪扩散概率方法

余志用¹, 郭晓新², 冯月萍², 张东坡¹

(1. 新疆政法学院 信息网络安全学院, 新疆 图木舒克 844000;

2. 吉林大学 计算机科学与技术学院, 长春 130012)

摘要: 基于非 Markov 链去噪扩散隐式模型(DDIM), 提出一种粗糙集的去噪扩散概率方法, 用粗糙集理论对采样的原序列等价划分, 在原序列上构建子序列的上下近似集和粗糙度, 当粗糙度最小时获取非 Markov 链去噪扩散隐式模型的有效子序列. 利用去噪扩散概率模型(DDPM)和 DDIM 进行对比实验, 实验结果表明, 该方法获取的序列是有效子序列, 且在该序列上的采样效率优于 DDPM.

关键词: 粗糙集; 去噪扩散概率模型; 非 Markov 链去噪扩散概率模型; Markov 链

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1671-5489(2024)02-0339-08

Probability Method of Denoising Diffusion Based on Rough Sets

SHE Zhiyong¹, GUO Xiaoxin², FENG Yueping², ZHANG Dongpo¹

(1. School of Information Network Security, Xinjiang University of Political Science and Law,
Tumxuk 844000, Xinjiang Uygur Autonomous Region, China;

2. College of Computer Science and Technology, Jilin University, Changchun 130012, China)

Abstract: Based on non Markov chain denoising diffusion implicit model (DDIM), we proposed probability method of denoising diffusion based on rough sets. The rough set theory was used to equivalently partition the sampled original sequence, construct the upper and lower approximation sets and roughness of the subsequences on the original sequence, and obtain the effective subsequences of the non Markov chain DDIM when the roughness was the lowest. The comparative experiments were conducted by the denoising diffusion probability model (DDPM) and DDIM, and the experimental results show that the sequence obtained by proposed method is an effective subsequence, and the sampling efficiency on this sequence is better than that of the DDPM.

Keywords: rough set; denoising diffusion probability model; non Markov chain denoising diffusion probability model; Markov chain

扩散模型^[1]在计算机视觉、自然语言处理、时间数据建模等领域应用广泛. 从高水平的细节到生成例子的多样性, 其展现了强大的生成能力. 目前, 扩散模型已被广泛应用于各种生成建模任务, 如图像生成^[2]、图像超分辨率^[3-4]、图像插入绘画^[5]、图像编辑^[6]和图像到图像转换^[7]等. 此外, 通过扩

收稿日期: 2023-04-18.

第一作者简介: 余志用(1990—), 男, 汉族, 硕士, 讲师, 从事图像处理 and 智能决策的研究, E-mail: szy@xjzfu.edu.cn. **通信作者简介:** 冯月萍(1958—), 女, 汉族, 博士, 教授, 从事计算机图形学和图像处理的研究, E-mail: fengyp@jlu.edu.cn.

基金项目: 国家自然科学基金(批准号: 82071995)、吉林省科技发展计划重点研发项目(批准号: 20220201141GX)和新疆政法学院校长基金(批准号: XZZK2021002; XZZK2022008).

散模型学习到的潜在表示也被应用在鉴别任务中, 如图像分割^[8-9]、分类^[10]和异常检测^[11], 从而证实了扩散模型的广泛适用性.

目前, 扩散模型中的去噪扩散概率模型^[12](DDPM)在生成领域备受关注. DDPM 具有不需要变分自编码器(VAE)^[13]对准后验分布、能量模型(EBM)^[14]难以解决的分区函数、对抗神经网络(GAN)^[15]训练额外的判别器以及归一化流^[16]施加网络约束等优点. 该模型提供了可操作的描述模型的概率参数, 有足够理论支持稳定训练程序. 统一的损失函数设计具有高度的简单性. 模型旨在将先验数据分布转化为随机噪声, 进而逐步加噪扩散, 先得到全高斯噪声的分布, 然后再逐步去噪扩散, 从而重建具有相同分布的全新样本. 但 DDPM 与生成对抗网络^[17](GAN)和变异自动编码器^[18](VAE)等模型相比有一个缺点, 即采样步骤多, 导致采样时间长.

针对上述问题, Nichol 等^[19]提出了非 Markov 链的去噪扩散隐式模型(DDIM), 该模型主要证明了扩散过程是非 Markov 链的过程. DDIM 在原序列上随机选取子序列进行采样, 虽提高了采样效率, 但因该子序列是随机选取导致无法有效保证每次采样效果都比 DDPM 优. 针对该问题, 本文在 DDIM 的基础上提出了基于粗糙集的去噪概率扩散方法, 解决选取较佳采样子列的问题.

1 预备知识

1.1 粗糙度

给定知识库 $K=(U,R)$, 对于每个子集 $X \subseteq U$ 和一个等价关系 R , 定义两个子集:

$$RX = \cup \{Y \in U/R | Y \subseteq X\}, \tag{1}$$

$$\bar{R}X = \cup \{Y \in U/R | Y \cap X \neq \emptyset\}, \tag{2}$$

分别称为 X 的 R 下近似集和 R 上近似集. 由等价关系 R 定义的集合 X 近似粗糙度为

$$\rho_R(X) = 1 - |\underline{R}X| / |\bar{R}X|. \tag{3}$$

1.2 粒子群

粒子群优化算法(PSO)是通过个体最优解搜索全局最优解, 粒子群优化算法带有权重的速度和位置更新公式如下:

$$v_i^{k+1} = \omega v_i^k + c_1 \text{rand}_1(p\text{best}_i^k - x_i^k) + c_2 \text{rand}_2(g\text{best}^k - x_i^k), \tag{4}$$

$$x_i^{k+1} = x_i^k + v_i^k. \tag{5}$$

1.3 去噪概率扩散模型

去噪扩散概率模型主要包括前扩散和逆扩散过程, 如图 1 所示.

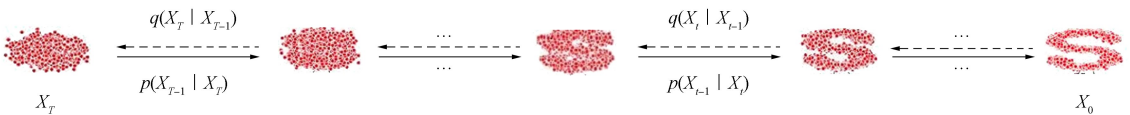


图 1 DDPM 扩散过程

Fig. 1 DDPM diffusion process

扩散过程中给定初始数据分布 $X_0 \sim q(X_0)$, 从 $t=0$ 时刻到 $t=T$ 时刻不断向对应的 t 添加高斯噪声, 该噪声的标准差是以固定值 β_t 且 $\{\beta_t \in (0, 1)\}_{t=1}^T$ 确定的, 均值是以固定值 β_t 和当前 t 时刻的数据 X_t 确定的. 且当 $T \rightarrow \infty$ 时 $\beta_1 < \beta_2 < \dots < \beta_t < \dots < \beta_{T-1} < \beta_T$, 整个扩散模型是一个 Markov 链过程.

文献[12]给出了某一时刻前扩散的条件概率分布:

$$q(X_t | X_{t-1}) = N(X_t; \sqrt{1 - \beta_t} X_{t-1}, \beta_t I), \quad q(X_{1:T} | X_0) = \prod_{t=1}^T q(X_t | X_{t-1}), \tag{6}$$

其中 I 是标准正态分布. 文献[12]令 $\alpha_t = 1 - \beta_t$ 且 $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, 给出了通过 X_0 和 α_t 表示任意时刻的 $q(X_t)$, 而不需要迭代, 推导得出如下结果:

$$X_t = \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1 - \bar{\alpha}_t} Z. \tag{7}$$

根据参数重整化得 $q(X_t | X_0) = N(X_t; \sqrt{\bar{\alpha}_t} X_0, (1 - \bar{\alpha}_t) I)$.

逆扩散过程是从高斯噪声中回复原始数据,该过程假设也是一个高斯分布,但无法逐步拟合分布,所以需要构建一个参数为 θ 的分布做估计.该过程仍然是一个 Markov 链过程.文献[12]给出了逆扩散过程如下:

$$p_\theta(X_{0:T}) = p(X_T) \prod_{t=1}^T p_\theta(X_{t-1} | X_t),$$

$$p_\theta(X_{t-1} | X_t) = N(X_{t-1}; \mu_\theta(X_t, t), \sum_{\theta} (X_t, t)).$$

文献[12]假设 $p_\theta(X_{t-1} | X_t)$ 的均值和方差分别为 $\mu_\theta(X_t, t)$, $\sum_{\theta} (X_t, t)$, 方差设置为与 β 相同的常数.

对于 $\mu_\theta(X_t, t)$ 用深度学习网络去预测,而深度学习网络最重要的是损失函数(Loss)的推导.

推导损失函数前必须已知前扩散过程的后验扩散条件概率 $q(X_{t-1} | X_t, X_0)$ 分布,设该分布为

$$q(X_{t-1} | X_t, X_0) = N(X_{t-1}; \tilde{\mu}(X_t, X_0), \tilde{\beta}_t I),$$

由 Bayes 公式、式(6)、式(7)和概率密度函数推导均值和方差:

$$q(X_{t-1} | X_t, X_0) = q(X_t | X_{t-1}, X_0) \frac{q(X_{t-1} | X_0)}{q(X_t | X_0)} \propto$$

$$\exp\left\{-\frac{1}{2}\left[\frac{(X_t - \sqrt{\alpha_t} X_{t-1})^2}{\beta_t} + \frac{(X_{t-1} - \sqrt{\bar{\alpha}_{t-1}} X_0)^2}{1 - \bar{\alpha}_{t-1}} + \frac{(X_t - \sqrt{\bar{\alpha}_t} X_0)^2}{1 - \bar{\alpha}_t}\right]\right\} =$$

$$\exp\left\{-\frac{1}{2}\left[\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1 - \bar{\alpha}_{t-1}}\right) X_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t} X_t + \frac{2\sqrt{\bar{\alpha}_t}}{1 - \bar{\alpha}_{t-1}} X_0\right) X_{t-1} + C(X_t, X_0)\right]\right\},$$

则

$$\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t, \quad (8)$$

该方差是常数:

$$\tilde{\mu}_t(X_t, X_0) = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} Z_t \right). \quad (9)$$

用最大似然函数推导深度学习网络的损失函数,主要思想是在负对数似然函数的基础上加一个 KL 散度,构建了负对数似然函数的上界,上界越小,负对数似然越小,则对数似然越大.推导过程如下:

$$-\log p_\theta(X_0) < -\log p_\theta(X_0) + D_{\text{KL}}(q(X_{1:T} | X_0) \| p_\theta(X_{1:T} | X_0)) = E_q \left[\log \frac{q(X_{1:T} | X_0)}{p_\theta(X_{0:T})} \right],$$

$$-E_{q(X_0)} \log p_\theta(X_0) \leq E_{q(X_0, T)} \left[\log \frac{q(X_{1:T} | X_0)}{p_\theta(X_{0:T})} \right].$$

则深度学习网络的损失函数为

$$\text{Loss} = E_{q(X_0, T)} \left[\log \frac{q(X_{1:T} | X_0)}{p_\theta(X_{0:T})} \right] = E_q \left[D_{\text{KL}}(q(X_T | X_0) \| p_\theta(X_T)) + \sum_{t=2}^T D_{\text{KL}}(q(X_{t-1} | X_t, X_0) \| p_\theta(X_{t-1} | X_t)) - \log p_\theta(X_0 | X_1) \right]. \quad (10)$$

根据式(8)、式(9)及两个单一变量的高斯分布 p 和 q 的 KL 散度

$$\text{KL}(p, q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2},$$

代入式(10)可得

$$\text{Loss} = E_q \left[\frac{1}{2\sigma_t^2} \|\tilde{\mu}_t(X_t, X_0) - \mu_\theta(X_t, t)\|^2 \right] + C,$$

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

$$\text{Loss} = \|\varepsilon - \varepsilon_\theta(\sqrt{\alpha_t} X_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon, t)\|^2, \quad (11)$$

其中: ϵ 为前向扩散的随机高斯噪声, 从标准正态分布中获取; $\epsilon_\theta(X_t, t)$ 为逆向扩散含参的随机高斯噪声, 需深度学习网络预测.

由于 DDPM 扩散过程依赖 Markov 链导致采样过程效率低, 因此文献[19]提出了非 Markov 链的 DDIM, 并在该模型中证明了 DDPM 的扩散过程不依赖 Markov 链, 损失函数与 DDPM 相同.

2 方法描述

DDIM 是在原采样序列上随机选取一子序列进行采样, 虽提高了采样效率, 但因该子序列是随机选取, 导致无法有效保证每次采样效果都比 DDPM 优. 针对该问题, 本文在 DDIM 的基础上提出了粗糙集的去噪扩散概率方法, 解决选取子序列的问题.

本文方法在逆扩散过程中的损失函数与 DDPM 和 DDIM 相同, 只是扩散过程不依赖 Markov 链, 通过粗糙集理论获取较佳的采样子序列. 在扩散过程中每一时刻 t 对应一个 X_t , 因此 t 从 T 到 1 时刻对应的 X_t , 在粗糙集上构成一个论域 $U = \{X_T, X_{T-1}, X_{T-2}, \dots, X_t, \dots, X_2, X_1\}$, U 中元素下标表示某一时刻 t 的分布. 扩散过程是非 Markov 链, 每一时刻 t 对应的 X_t 的分布都是独立分布, 可以把一个时间段 ∇t 作为等价关系 R . 最终确定知识库^[20] $K = (U, \nabla t)$, ∇t 与 T 的规模相关. 利用等价关系 ∇t 对论域 U 进行等价划分得到的等价类记作 $U/\nabla t$.

根据 1.1 节中的定义确定任意采样子集 $Q (Q \subseteq U)$ 在论域 U 上的上下近似集及粗糙度:

$$\begin{aligned} \underline{\nabla t}Q &= \bigcup \{Y \in U/\nabla t \mid Y \subseteq Q\}, \\ \overline{\nabla t}Q &= \bigcup \{Y \in U/\nabla t \mid Y \cap Q \neq \emptyset\}, \\ \rho_{\nabla t}(Q) &= 1 - |\underline{\nabla t}Q| / |\overline{\nabla t}Q|. \end{aligned} \quad (12)$$

根据粗糙集理论知识, 集合不精确性是由边界域的存在导致, 边界域越小, 其粗糙度越小, 该知识越确定. 在论域 U 上当子集 Q 的粗糙度最小时, 此时子集 Q 中每个元素下标构成的序列是较佳的采样子序列. 因论域 U 规模大, 其子集较多, 为解决在论域 U 上获取较佳采样子序列时效率低的问题, 采用粗糙度为目标函数的粒子群优化算法解决. 本文方法流程如下.

用 Loss 函数训练 U-net 深度学习网络:

步骤 1) X_0 的概率分布 $q(X_0)$;

步骤 2) $t \in \{1, 2, \dots, T-1, T\}$;

步骤 3) ϵ 是从标准正态分布 $N(0, I)$ 中随机获取;

步骤 4) 深度学习网络训练, $\text{Loss} = \|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}X_0 + \sqrt{1-\alpha_t}\epsilon, t)\|^2$;

步骤 5) 训练完毕.

采样:

步骤 1) $X_T \sim N(0, I)$;

步骤 2) 初始化式(4)中粒子位置, 且 $c_1 = c_2 = 2$, rand_1 和 rand_2 取 $[0, 1]$ 中随机数, 惯性权值 $\omega = 0.6$;

步骤 3) for $i = 1, 2, \dots, 2^{|U|}$

步骤 4) $Q_i \subseteq U, \rho_{\nabla t}(Q_i)$;

步骤 5) if $(\rho_{\nabla t}(Q_i) < p\text{best}_i)$

步骤 6) $p\text{best}_i = \rho_{\nabla t}(Q_i)$;

步骤 7) if $(p\text{best}_i < g\text{best})$

步骤 8) $g\text{best} = p\text{best}_i$;

步骤 9) 进化式(4), (5)中的粒子位置和速度;

步骤 10) break

步骤 11) end for

步骤 12) 获取 $\rho_{\nabla t}(Q_i)$ 最小时 Q_i 集合元素的下标序列 $\{t_i, t_{i+1}, \dots, t_{i+j}\}$ 即为采样的最佳子序

列 $j = |Q_i|$;

步骤 13) for $t = t_i, t_{i+1}, \dots, t_{i+j}$

步骤 14) $Z \sim (0, I)$;

步骤 15) if ($t > t_{i+j}$);

步骤 16) $X_{t-1} = \frac{1}{\sqrt{\alpha_t}} (X_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(X_t, t)) + \sigma_t Z$;

步骤 17) end for

步骤 18) return X_0 .

3 实验分析

本文实验在数据集 FFHQ 上进行对比分析, 采用 Pytorch (GPU 版本) 深度学习框架, RTX A6000 显卡. 根据选取的数据集和超参数 $\beta_t (\{\beta_t \in (0, 1)\}_{t=1}^T)$ 确定 $T=100$, 其中, $c_1 = c_2 = 2$, $\text{rand}_1 = \text{rand}_2 = 0.5$, $\omega = 0.6$. 利用 U-net 网络对 $\text{Loss} = \|\epsilon - \epsilon_\theta(\sqrt{\alpha_t} X_0 + \sqrt{1-\alpha_t} \epsilon, t)\|^2$ 的模型进行训练, 训练该模型有 800 个 epoch, 每个 epoch 对应的损失函数值列于表 1, 对应关系如图 2 所示. 在表 1 中每 10 个 epoch 取一个损失值. 由图 2 可见, 当 epoch 逐渐增大时, 损失值逐渐减小. 一直达到指定的误差范围, 该模型训练完毕. U-net 网络在数据集 FFHQ 上训练本文模型的部分过程如图 3 所示.

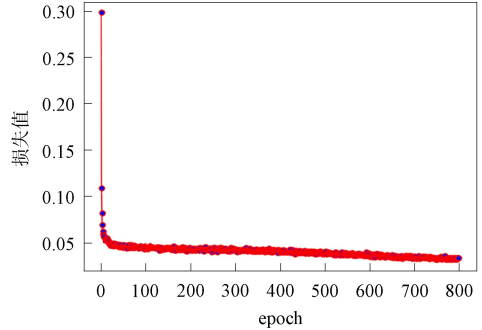


图 2 epoch 与损失函数值的对应关系

Fig. 2 Correspondence between epoch and loss function values

表 1 epoch 对应的损失函数值

Table 1 Loss function values corresponding to epoch

epoch	10	20	30	40	50	...	400	...	760	770	780	790	800
损失函数值	0.054	0.049	0.047	0.045	0.045	...	0.04	...	0.034	0.032	0.031	0.030	0.030

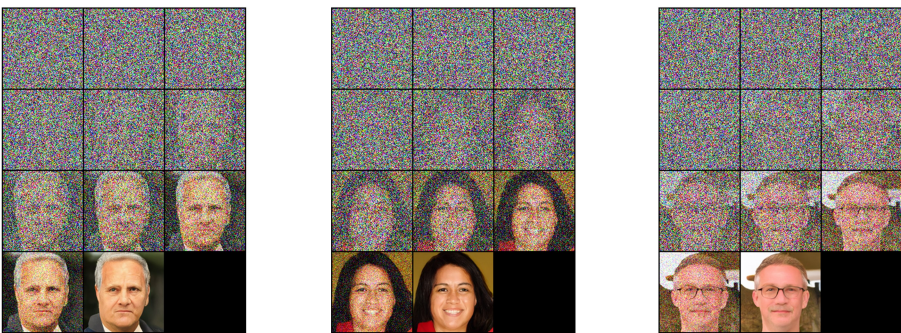


图 3 训练过程

Fig. 3 Training process

当 $T=100$ 时, 论域 $U = \{X_{100}, X_{99}, X_{98}, \dots, X_t, \dots, X_2, X_1\}$, 该论域有 2^{100} 个子集, 以粗糙度 $\rho_{\nabla_t}(Q) = 1 - \frac{|\nabla_t Q|}{|\overline{\nabla_t Q}|}$ 为目标函数, 利用粒子群优化算法求解最小粗糙度时对应的论域子集下标序列为较佳采样子序列. 迭代次数对应的粗糙度值列于表 2, 对应关系如图 4 所示.

如果不用粒子群优化算法优化本文方法需要迭代 2^{100} 次, 迭代规模庞大. 由图 4 可见, 利用粒子群优化本文方法时, 当迭代到 15 000 次时粗糙度达最小值, 效率明显提高. 当粗糙度最小值为 0.071 时对应的子集 $Q_{43} = \{X_{75}, X_{74}, X_{73}, \dots, X_{35}, X_{34}, X_{33}\}$, 该子集下标对应序列 $\{t_{75}, t_{74}, t_{73}, \dots, t_{35}, t_{34}, t_{33}\}$ 为获取的采样子列, 且 $t_{75} > t_{74} > t_{73} > \dots > t_{35} > t_{34} > t_{33}$.

表 2 迭代次数对应的粗糙度

Table 2 Roughness corresponding to number of iterations

迭代次数	50	100	150	200	250	300	...	7 500	...	14 850	149 00	149 50	15 000
粗糙度	0.764	0.729	0.696	0.702	0.624	0.599	...	0.103	...	0.072	0.071	0.071	0.071

在相同损失值下将 DDPM, DDIM 和本文方法利用 U-net 网络训练模型. DDPM 在 $\{t_{100}, t_{99}, t_{98}, \dots, t_2, t_1\}$ 序列上的采样结果与 DDIM 和本文方法在 $\{t_{75}, t_{74}, t_{73}, \dots, t_{35}, t_{34}, t_{33}\}$ 序列上的采样结果对比如图 5 所示. DDPM, DDIM 和本文方法的 FID (Fréchet inception distance) 值列于表 3.

图 5 为 DDPM 在原序列上采样得到的 3 张人脸图像, DDIM 和本文算法在获取的子序列上采样得到 3 张人脸图像. 由图 5 可见, 通过直观的对 3 种算法采样得到图像对比, DDIM 和本文算法在获取子序列上采样效果与 DDPM 在原序列上的采样效果相当. 说明本文算法获取的子序列是有效子序列, 且采样效果相似. 表 3 为通过 FID 量化评估 3 种算法采样生成图像的质量, FID 越小生成的图像质量越好. 由表 3 可见, 3 种算法 FID 值都相对较小, 说明 3 种算法采样生成的图像质量较好. 本文算法生成图像的 FID 值与 DDIM 和 DDPM 的 FID 值接近, 进一步说明本文算法获取的子列是有效子列, 且采样效果相似.

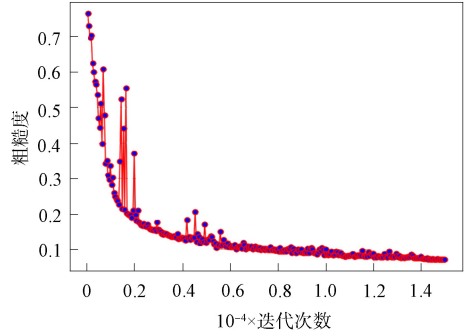


图 4 迭代次数与粗糙度的对应关系

Fig. 4 Correspondence between number of iterations and roughness

表 3 3 种算法的 FID 值

Table 3 FID values of three algorithms

算法	FID			算法	FID		
DDPM	22.71	21.36	20.24	本文	22.06	22.67	21.39
DDIM	21.98	22.52	21.02				

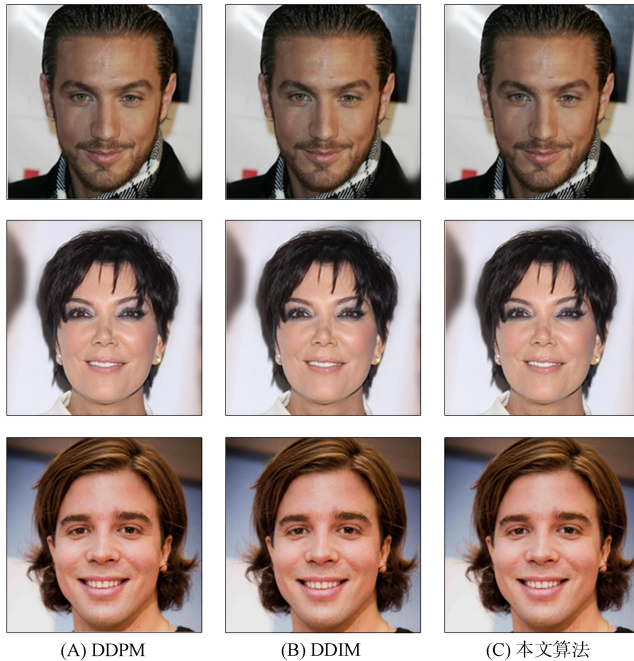


图 5 3 种算法采样结果对比

Fig. 5 Comparison of sampling results of three algorithms

验证本文算法获取子列是较佳的采样子序列. DDIM 分别在随机选取 3 个子序列和本文获取的子序列上采样结果与本文算法采样结果对比如图 6 所示. 图 6 采样结果的 FID 值列于表 4.

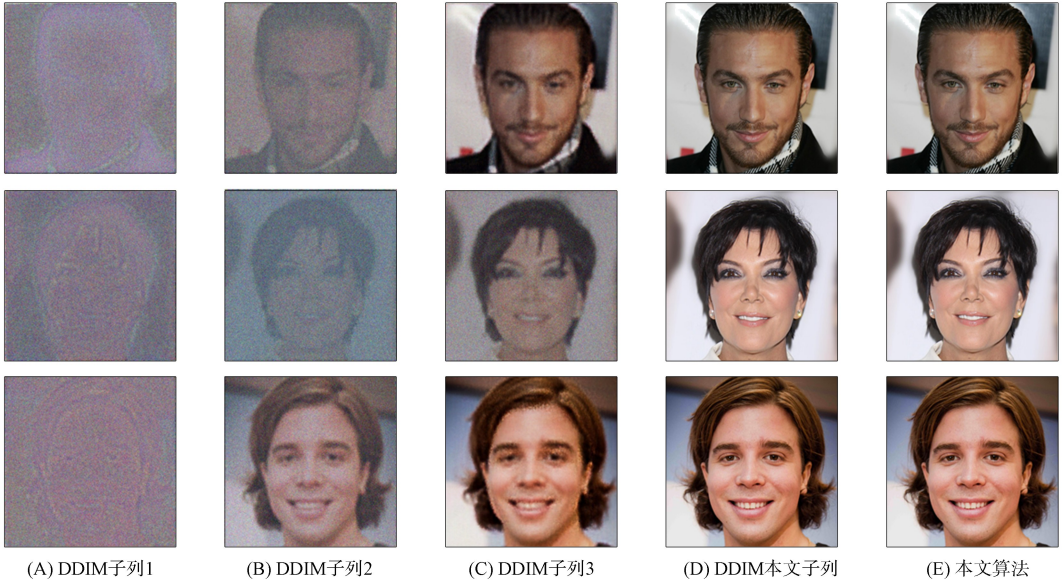


图 6 不同子列采样结果对比

Fig. 6 Comparison of sampling results of different sub columns

表 4 图 6 采样结果的 FID 值

Table 4 FID values of sampling results in Figure 6

子列	FID			子列	FID		
DDIM 子列 1	329.99	338.51	309.16	DDIM 本文子列	21.98	22.52	21.02
DDIM 子列 2	227.23	221.69	219.91	本文算法	22.06	22.67	21.39
DDIM 子列 3	105.79	107.25	90.67				

为说明本文算法获取子序列采样的优越性, 图 6 中的 3 个随机子序列都是等长的. 从 DDIM 在 3 个随机子序列采样效果与 DDIM 和本文算法在本文获取的子序列采样结果对比, 在本文获取的子序列上采样效果明显优于 3 个随机子序列上的采样效果. 说明本文获取的采样子序列不但采样有效, 且采样效果更佳. 对表 4 中不同子序列生成图像的 FID 对比分析, DDIM 和本文方法在获取子序列上生成图像的 FID 明显低于 DDIM 在其他随机子序列上的 FID. 进一步说明本文方法获取的子序列是较佳采样子序列.

由图 5 和表 3 可见, 虽然 DDPM 在原序列上采样效果与本文方法在获取子序列上的采样结果相似, 但在理论上 DDPM 比本文方法的采样效率低. 原因是 DDPM 在原序列上采样, 而本文方法在子序列上采样. DDPM 的采样时间分别为 103, 104, 102 s, 本文方法的采样时间分别为 21, 22, 20 s. 实验结果表明, 本文方法在获取的子序列上采样效率是 DDPM 在原序列上采样效率的 4 倍多, 说明本文方法采样效率明显高于 DDPM 的采样效率.

综上所述, 本文在 DDIM 的基础上通过粗糙集和粒子群优化算法相结合获取较佳采样子序列. 实验验证结果表明, 本文方法获取的采样子序列上生成图像效果与 DDPM 采样效果相似, 说明本文获取的子列是有效子列; DDIM 和本文方法在获取子序列上的采样效果明显优于 DDIM 在多个随机子序列上的采样结果, 说明本文获取的有效子序列是较佳子序列, 进一步说明 DDIM 在随机子序列上的采样结果不稳定; 最后通过本文方法和 DDPM 分别在获取的子序列和原序列上的采样时间对比, 表明本文方法采样效率明显高于 DDPM 的采样效率.

参 考 文 献

[1] SOHL-DICKSTEIN J, WEISS E, MAHESWARANATHAN N, et al. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics [C]//International Conference on Machine Learning. [S. l.]: PMLR, 2015: 2256-2265.

- [2] LU C, ZHOU Y H, BAO F, et al. Dpm-Solver: A Fast Ode Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 5775-5787.
- [3] KAWAR B, ELAD M, ERMON S, et al. Denoising Diffusion Restoration Models [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 23593-23606.
- [4] SAHARIA C, HO J, CHAN W, et al. Image Super-resolution via Iterative Refinement [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(4): 4713-4726.
- [5] LUGMAYR A, DANELLJAN M, ROMERO A, et al. Repaint: Inpainting Using Denoising Diffusion Probabilistic Models [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2022: 11461-11471.
- [6] AVRAHAMI O, LISCHINSKI D, FRIED O. Blended Diffusion for Text-Driven Editing of Natural Images [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2022: 18208-18218.
- [7] SAHARIA C, CHAN W, CHANG H W, et al. Palette: Image-to-Image Diffusion Models [C]//*ACM SIGGRAPH 2022 Conference Proceedings*. New York: ACM, 2022: 1-10.
- [8] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-Resolution Image Synthesis with Latent Diffusion Models [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2022: 10684-10695.
- [9] LI M Y, LIN J, MENG C L, et al. Efficient Spatially Sparse Inference for Conditional Gans and Diffusion Models [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 28858-28873.
- [10] MIKUNI V, NACHMAN B. Score-Based Generative Models for Calorimeter Shower Simulation [EB/OL]. (2022-06-17)[2023-02-10]. <https://arxiv.org/abs/2206.11898>.
- [11] PINAYA W H L, GRAHAM M S, GRAY R, et al. Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models [C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer, 2022: 705-714.
- [12] HO J, JAIN A, ABBEEL P. Denoising Diffusion Probabilistic Models [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 6840-6851.
- [13] GUU K, HASHIMOTO T B, OREN Y, et al. Generating Sentences by Editing Prototypes [J]. *Transactions of the Association for Computational Linguistics*, 2018, 6: 437-450.
- [14] YU L T, SONG Y, SONG J M, et al. Training Deep Energy-Based Models with f-Divergence Minimization [C]//*International Conference on Machine Learning*. [S.l.]: PMLR, 2020: 10957-10967.
- [15] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative Adversarial Networks: An Overview [J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 53-65.
- [16] GARCÍA G G, CASAS P, FERNÁNDEZ A, et al. On the Usage of Generative Models for Network Anomaly Detection in Multivariate Time-Series [J]. *ACM SIGMETRICS Performance Evaluation Review*, 2021, 48(4): 49-52.
- [17] PEI H Z, REN K, YANG Y Q, et al. Towards Generating Real-World Time Series Data [C]//*2021 IEEE International Conference on Data Mining (ICDM)*. Piscataway, NJ: IEEE, 2021: 469-478.
- [18] LIANG D W, KRISHNAN R G, HOFFMAN M D, et al. Variational Autoencoders for Collaborative Filtering [C]//*Proceedings of the 2018 World Wide Web Conference*. New York: ACM, 2018: 689-698.
- [19] NICHOL A Q, DHARIWAL P. Improved Denoising Diffusion Probabilistic Models [C]//*International Conference on Machine Learning*. [S.l.]: PMLR, 2021: 8162-8171.
- [20] 余志用, 段超, 张雷. 变精度最小平方粗糙熵的图像分割算法 [J]. *计算机工程与科学*, 2019, 41(4): 657-664. (SHE Z Y, DUAN C, ZHANG L. An Image Segmentation Algorithm Using Variable Precision Least Square Rough Entropy [J]. *Computer Engineering & Science*, 2019, 41(4): 657-664.)