

扩散模型：方法和应用的综合调查应用

杨凌, 北京大学, 中国 张志龙*, 北京大学, 中国 宋阳, OpenAI, 美国 洪申达, 北京大学, 中国

徐润生, 加州大学洛杉矶分校, 美国

赵悦, 美国卡内基梅隆大学

张文涛, 北京大学, 中国

BIN CUI, 北京大学, 中国 MING-HSUAN YANG[†], 加州大学默塞德分校, 美国

扩散模型已成为一种强大的新型深度生成模型, 在图像合成、视频生成和分子设计等许多应用中都具有破纪录的性能。在这篇综述中, 我们概述了迅速扩展的扩散模型研究, 将研究分为三个关键领域: 高效采样、改进的似然估计和处理具有特殊结构的数据。我们还讨论了将扩散模型与其他生成模型相结合以获得增强结果的潜力。我们进一步回顾了扩散模型在计算机视觉、自然语言处理、时间数据建模以及其他科学学科的跨学科应用等领域的广泛应用。这篇综述旨在对扩散模型的状态进行背景化、深入的观察, 确定重点关注领域并指出进一步探索的潜在领域。Github: <https://github.com/YangLing0818/Diffusion-Models-Papers-Survey-Taxonomy>。

CCS 概念: • 计算方法 → 计算机视觉任务; 自然语言生成; 机器学习方法。

其他关键词和短语: 生成模型、扩散模型、基于分数的生成模型、随机微分方程

ACM 参考格式:

杨凌、张志龙、宋阳、洪申达、徐润生、赵悦、张文涛、崔斌和杨明轩。2023.扩散模型: 方法和应用的全面调查。1, 1 (2023 年 12 月), 58 页。 <https://doi.org/10.1145/3626235>

*Contributed equally.

[†]Wentao Zhang, Bin Cui, and Ming-Hsuan Yang are corresponding authors.

Authors' addresses: Ling Yang, Peking University, China, yangling0818@163.com; Zhilong Zhang, Peking University, China, zhilong.zhang@bjmu.edu.cn; Yang Song, OpenAI, USA, songyang@openai.com; Shenda Hong, Peking University, China, hongshenda@pku.edu.cn; Runsheng Xu, University of California, Los Angeles, USA, rx33386@ucla.edu; Yue Zhao, Carnegie Mellon University, USA, zhaoy@cmu.edu; Wentao Zhang, Peking University, China, wentao.zhang@pku.edu.cn; Bin Cui, Peking University, China, bin.cui@pku.edu.cn; Ming-Hsuan Yang, University of California at Merced, USA, mhyang@ucmerced.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

内容

摘要	1
目录	2
1 简介	3
2 扩散模型基础	5
2.1 去噪扩散概率模型 (DDPM)	5
2.2 基于分数的生成模型 (SGM)	8
2.3 随机微分方程 (分数 SDE)	9
3 具有高效采样的扩散模型	10
3.1 无学习采样	10
3.1.1 SDE 求解器	10
3.1.2 ODE 求解器	12
3.2 基于学习的采样	13
3.2.1 优化离散化	13
3.2.2 截断扩散	14
3.2.3 知识蒸馏	14
4 具有改进似然的扩散模型	14
4.1 噪声调度优化	14
4.2 反向方差学习	15
4.3 精确似然计算	16
5 特殊结构数据的扩散模型	17
5.1 离散数据	17
5.2 结构不变的数据	17
5.3 流形结构的数据	18
5.3.1 已知流形	18
5.3.2 学习到的流形	18
6 与其他生成模型的联系	19
6.1 大型语言模型与扩散模型的联系	19
6.2 变分自编码器与扩散模型的联系	20
6.3 生成对抗网络与扩散模型的联系	21
6.4 正则化流与扩散模型的联系	22
6.5 自回归模型与扩散模型的联系	23
6.6 基于能量的模型与扩散模型的联系	24
7 扩散的应用模型	24
7.1 无条件条件和条件扩散模型	24
7.1.1 扩散模型中的条件机制	25
7.1.2 使用 DPO/RLHF 进行扩散	25
7.1.3 标签和分类器上的条件扩散	27
7.1.4 文本、图像和语义图上的条件扩散	27
7.1.5 图上的条件扩散	27

Diffusion Models: A Comprehensive Survey of Methods and Applications	3
7.2 Computer Vision	27
7.2.1 Image Super Resolution, Inpainting, Restoration, Translation, and Editing	27
7.2.2 Semantic Segmentation	29
7.2.3 Video Generation	29
7.2.4 Generating Data from Diffusion Models	29
7.2.5 Point Cloud Completion and Generation	29
7.2.6 Anomaly Detection	31
7.3 Natural Language Generation	31
7.4 Multi-Modal Generation	32
7.4.1 Text-to-Image Generation	32
7.4.2 Scene Graph-to-Image Generation	34
7.4.3 Text-to-3D Generation	34
7.4.4 Text-to-Motion Generation	37
7.4.5 Text-to-Video Generation	38
7.4.6 Text-to-Audio Generation	38
7.5 Temporal Data Modeling	40
7.5.1 Time Series Imputation	40
7.5.2 Time Series Forecasting	40
7.5.3 Waveform Signal Processing	41
7.6 Robust Learning	41
7.7 Interdisciplinary Applications	41
7.7.1 Drug Design and Life Science	41
7.7.2 Material Design	42
7.7.3 Medical Image Reconstruction	43
8 Future Directions	43
Revisiting Assumptions	43
Theoretical Understanding	44
Latent Representations	44
AIGC and Diffusion Foundation Models	44
9 Conclusion	44
References	44

1 引言

扩散模型[111, 275, 280, 285]已成为深度生成模型中最先进的新型模型。它们打破了生成对抗网络 (GAN) [88] 在图像合成这一具有挑战性的任务中的长期主导地位 [60, 111, 280, 285]，并在多个领域展现出潜力，包括计算机视觉[4, 15, 25, 29, 112, 114, 145, 149, 171, 194, 206, 225, 257, 259, 315, 354, 355, 379, 389]、自然语言处理[9, 117, 175, 264, 361]、时间数据建模[3, 39, 159, 249, 291, 335]、多模态建模[10, 243, 255, 258, 386]，

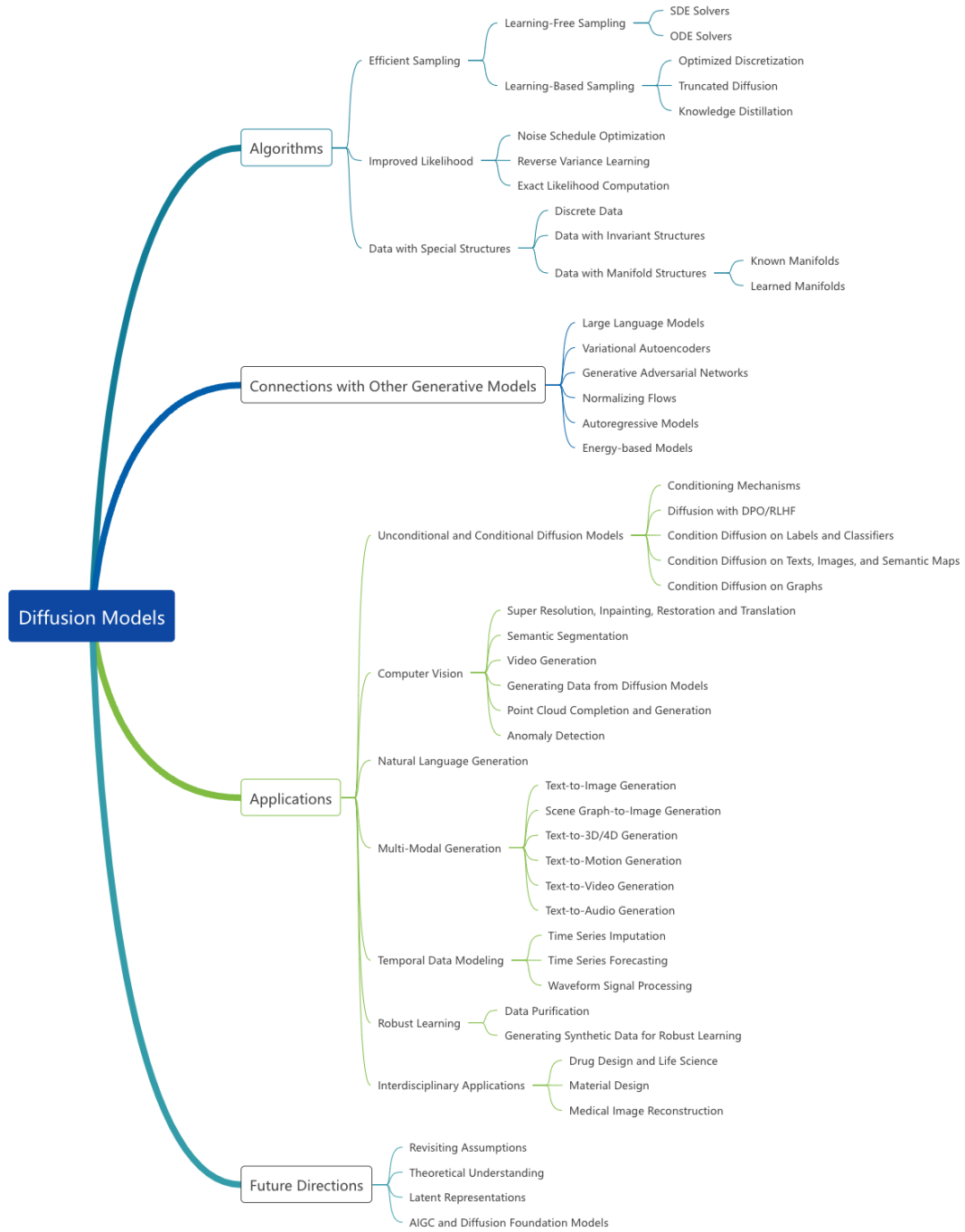


图 1. 传播模型变体的分类（第 3 至 5 节）、与其他生成模型的联系（第 6 节）、传播模型的应用（第 7 节）以及未来方向（第 8 节）。

鲁棒机器学习[23, 33, 144, 308, 357], 计算化学[5, 115, 133, 166, 169, 196, 327]和医学图像重建[31, 47 – 49, 53, 202, 230, 284, 328]等领域的跨学科应用。

人们已经开发出许多方法来改进扩散模型, 要么通过提高实证表现[214, 277, 281], 要么从理论角度扩展模型的容量[187, 188, 279, 285, 371]。在过去两年中, 扩散模型的研究主体显著增长, 这使得新研究人员越来越难以跟上该领域的最新发展。此外, 大量的工作可能会掩盖主要趋势并阻碍进一步的研究进展。本综述旨在通过全面概述扩散模型研究的现状、对各种方法进行分类以及强调关键进展来解决这些问题。我们希望本综述能为该领域的新研究人员提供一个有用的切入点, 同时为经验丰富的研究人员提供更广阔的视角。

在本文中, 我们首先解释扩散模型的基础(第2节), 简要但独立地介绍三种主要公式: 去噪扩散概率模型(DDPM)[111, 275]、基于分数的生成模型(SGM)[280, 281]和随机微分方程(Score SDE)[141, 279, 285]。所有这些方法的关键是用不断增强的随机噪声逐步扰动数据(称为“扩散”过程), 然后依次消除噪声以生成新的数据样本。我们阐明它们如何在相同的扩散原理下工作, 并解释这三个模型是如何相互联系并可以相互简化的。

接下来, 我们介绍了近期研究的分类法, 该分类法勾勒出了扩散模型领域, 将其分为三个关键领域: 有效采样(第3节)、改进的似然估计(第4节)和处理具有特殊结构的数据的方法(第5节), 例如关系数据、具有置换/旋转不变性的数据以及驻留在流形上的数据。我们通过将每个类别细分为更详细的子类别来进一步研究模型, 如图1所示。此外, 我们讨论了扩散模型与其他深度生成模型的联系(第6节), 包括变分自编码器(VAE)[156, 252]、生成对抗网络(GAN)[88]、正则化流[62, 64, 226, 254]、自回归模型[302]和基于能量的模型(EBM)[165, 283]。通过这些模型与扩散模型相结合, 研究人员有可能实现更强大的性能。

随后, 我们的调查回顾了现有研究中扩散模型的六大应用类别(第7节): 计算机视觉、自然语言处理、时间数据建模、多模态学习、稳健学习和跨学科应用。对于每项任务, 我们都提供了一个定义, 描述了如何使用扩散模型来解决它, 并总结了相关的先前工作。我们在论文的最后(第8和9节)展望了这一令人兴奋的新研究领域的未来可能发展方向。

2 扩散模型的基础

扩散模型是一类概率生成模型, 它们通过注入噪声逐渐破坏数据, 然后学习逆转此过程以生成样本。我们在图2中展示了扩散模型的直观感受。当前对扩散模型的研究主要基于三种主要公式: 去噪扩散概率模型(DDPM)[111, 214, 275]、基于分数的生成模型(SGM)[280, 281]和随机微分方程(Score SDE)[279, 285]。我们在本节中对这三种公式进行了完整的介绍, 同时讨论了它们之间的联系。

2.1 去噪扩散概率模型(DDPM)

去噪扩散概率模型(DDPM)[111, 275]使用两个马尔可夫链: 一个将数据扰动为噪声的前向链, 以及一个将噪声转换回数据的反向链。前者通常是用 Manuscript submitted to ACM 手工设计的

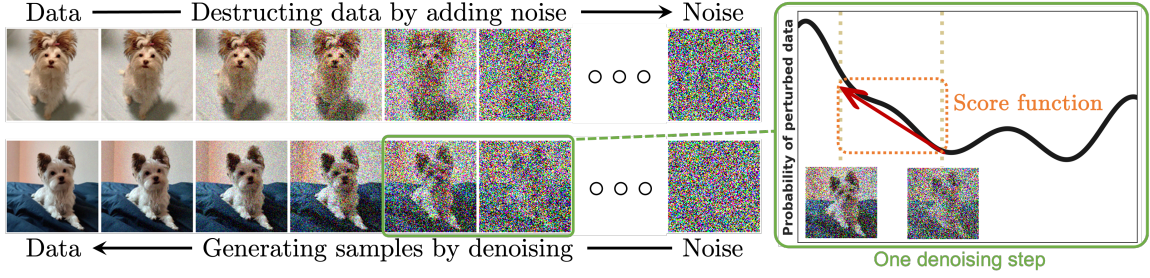


图 2. 扩散模型通过添加噪声平滑地扰乱数据，然后逆向此过程以从噪声中生成新数据。逆向过程中的每个去噪步骤通常都需要估计得分函数（参见右侧的说明图），该函数是指向可能性更高、噪声更少的数据方向的梯度。

目标是将任何数据分布转换为简单的先验分布（例如标准高斯分布），而后的马尔可夫链通过学习由深度神经网络参数化的过渡核来反转前者。随后，通过首先从先验分布中采样一个随机向量，然后通过反向马尔可夫链进行祖先采样来生成新的数据点 [158]。

正式地，给定一个数据分布 $x_0 \sim q(x_0)$ ，前向马尔可夫过程会生成一系列随机变量 $x_1, x_2 \dots x_T$ ，其过渡核为 $q(x_t | x_{t-1})$ 。利用概率链式法则和马尔可夫性质，我们可以将 $x_1, x_2 \dots x_T$ 的联合分布分解为 x_0 ，记为 $q(x_1, \dots, x_T | x_0)$ ，分解为

$$q(x_1, \dots, x_T | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}). \quad (1)$$

在 DDPM 中，我们手工制作过渡核 $q(x_t | x_{t-1})$ ，以逐步将数据分布 $q(x_0)$ 转换为可处理的先验分布。过渡核的一个典型设计是高斯扰动，最常见的过渡核选择是

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbf{I}), \quad (2)$$

其中 $\beta_t \in (0, 1)$ 是在模型训练之前选择的超参数。我们使用这个核来简化我们的讨论，尽管其他类型的核也同样适用。正如 Sohl-Dickstein 等人 (2015) [275] 所观察到的，这个高斯过渡核使我们能够边缘化公式 (1) 中的联合分布，以获得对所有 $t \in \{0, 1, \dots, T\}$ 的 $q(x_t | x_0)$ 的解析形式。具体来说，对于 $\alpha_t := 1 - \beta_t$ 和 $\bar{\alpha}_t := \prod_{s=0}^t \alpha_s$ ，我们有

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t) \mathbf{I}). \quad (3)$$

给定 x_0 ，我们可以通过对高斯向量 $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ 进行采样并应用变换，轻松获得 x_t 的样本

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon. \quad (4)$$

当 $\bar{\alpha}_T \approx 0$ 时， x_T 几乎服从高斯分布，因此有 $q(x_T) := \int q(x_T | x_0) q(x_0) dx_0 \approx \mathcal{N}(x_T; 0, \mathbf{I})$ 。

直观地说，这个正向过程会慢慢地向数据注入噪声，直到所有结构都丢失。为了生成新的数据样本，DDPM 首先从先验分布（通常很容易获得）生成非结构化噪声向量，然后通过反向时间方向上运行可学习的马尔可夫链逐渐消除其中的噪声。具体来说，反向马尔可夫链由先验分布 $p(x_T) = \mathcal{N}(x_T; 0, \mathbf{I})$ 和

Manuscript submitted to ACM 参数化

可学习转换核 $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 。我们选择先验分布 $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ ，因为前向过程的构造方式是 $q(\mathbf{x}_T) \approx \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ 。可学习转换核 $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 的形式为

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)) \quad (5)$$

其中 θ 表示模型参数，均值 $\mu_\theta(\mathbf{x}_t, t)$ 和方差 $\Sigma_\theta(\mathbf{x}_t, t)$ 由深度神经网络参数化。有了这个逆马尔可夫链，我们可以通过首先采样噪声向量 $\mathbf{x}_T \sim p(\mathbf{x}_T)$ 来生成数据样本 $\mathbf{x}_0$ ，然后从可学习的转换核 $\mathbf{x}_{t-1} \sim p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 迭代采样直到 $t = 1$ 。

此采样过程成功的关键是训练逆马尔可夫链以匹配正向马尔可夫链的实际时间反转。也就是说，我们必须调整参数 θ ，以便逆马尔可夫链 $p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T) := p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 的联合分布与正向过程 $q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T) := q(\mathbf{x}_0) \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1})$ (Eq. (1)) 的联合分布非常接近。这可以通过最小化这两者之间的 Kullback-Leibler (KL) 散度来实现：

$$\text{KL}(q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T) \parallel p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)) \quad (6)$$

$$\stackrel{(i)}{=} -\mathbb{E}_{q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)} [\log p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)] + \text{const} \quad (7)$$

$$\stackrel{(ii)}{=} \underbrace{\mathbb{E}_{q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)} \left[-\log p(\mathbf{x}_T) - \sum_{t=1}^T \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q(\mathbf{x}_t | \mathbf{x}_{t-1})} \right]}_{:= -L_{\text{VLB}}(\mathbf{x}_0)} + \text{const} \quad (8)$$

$$\stackrel{(iii)}{\geq} \mathbb{E}[-\text{对数 } p_\theta(\mathbf{x}_0)] + \text{const}, \quad (9)$$

其中，(i) 来自 KL 散度的定义，(ii) 来自 $q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)$ 和 $p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)$ 都是分布的乘积，(iii) 来自 Jensen 不等式。等式 (8) 中的第一项是数据 $\mathbf{x}_0$ 的对数似然的变分下界 (VLB)，这是训练概率生成模型的常见目标。我们使用 “const” 表示不依赖于模型参数 θ 的常数，因此它不影响优化。DDPM 训练的目标是最大化 VLB（或者等效地，最小化负 VLB），它特别容易优化，因为它是独立项的总和，因此可以通过蒙特卡洛抽样 [212] 有效地估计，并通过随机优化 [286] 有效地优化。

Ho 等人 (2020 年) [111] 建议对 L_{VLB} 中的各项重新加权，以提高样本质量，并注意到所得损失函数与噪声条件得分网络（一种基于得分的生成模型，见 Song 和 Ermon [280]）的训练目标之间存在重要的等价性。[111] 中的损失函数形式为

$$\mathbb{E}_{t \sim \mathcal{U}[1, T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} [\lambda(t) \|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|^2] \quad (10)$$

其中 $\lambda(t)$ 是正加权函数， $\mathbf{x}_t$ 根据 $\mathbf{x}_0$ 和 ϵ 通过公式 (4) 计算得出， $\mathcal{U}[1, T]$ 是集合 $\{1, 2, \dots, T\}$ 上的均匀分布， ϵ_θ 是具有参数 θ 的深度神经网络，可根据 $\mathbf{x}_t$ 和 t 预测噪声向量 ϵ 。对于特定的加权函数 $\lambda(t)$ ，此目标函数简化为公式 (8)，其形式与基于得分的生成模型在多个噪声尺度上的去噪得分匹配损失相同 [280]，这是下一节将要讨论的另一种扩散模型的公式。

2.2 基于分数的生成模型 (SGM)

基于分数的生成模型 [280, 281] 的核心是 (Stein) 分数 (又称分数或分数函数) 的概念 [126]。给定概率密度函数 $p(\mathbf{x})$ ，其分数函数定义为对数概率密度 $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ 的梯度。与统计学中常用的 Fisher 分数 $\nabla_{\theta} \log p_{\theta}(\mathbf{x})$ 不同，这里考虑的 Stein 分数是数据 $\mathbf{x}$ 的函数，而不是模型参数 θ 。它是一个矢量场，指向概率密度函数具有最大增长率的方向。

基于分数的生成模型 (SGM) [280] 的关键思想是用一系列增强的高斯噪声扰动数据，并通过训练以噪声水平为条件的深度神经网络模型（在 [280] 中称为噪声条件分数网络，NCSN），联合估计所有噪声数据分布的分数函数。样本是通过使用基于分数的采样方法在降低的噪声水平下链接分数函数生成的，这些方法包括朗之万蒙特卡洛 [96, 137, 227, 280, 285]、随机微分方程 [136, 285]、常微分方程 [141, 188, 279, 285, 371] 及其各种组合 [285]。在基于分数的生成模型的制定中，训练和采样是完全解耦的，因此在估计分数函数后可以使用多种采样技术。

与第 2.1 节中的符号类似，我们让 $q(\mathbf{x}_0)$ 为数据分布， $0 < \sigma_1 < \sigma_2 < \dots < \sigma_t < \dots < \sigma_T$ 为噪声水平序列。SGM 的一个典型示例是使用高斯噪声分布 $q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0, \sigma_t^2 \mathbf{I})$ 将数据点 $\mathbf{x}_0$ 扰动至 $\mathbf{x}_t$ 。这会产生一系列噪声数据密度 $q(\mathbf{x}_1), q(\mathbf{x}_2), \dots, q(\mathbf{x}_T)$ ，其中 $q(\mathbf{x}_t) := \int q(\mathbf{x}_t) q(\mathbf{x}_0) d\mathbf{x}_0$ 。噪声条件得分网络是一个经过训练的深度神经网络 $s_{\theta}(\mathbf{x}, t)$ ，用于估计得分函数 $\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t)$ 。从数据中学习得分函数（又称得分估计）已经建立了得分匹配 [126]、去噪得分匹配 [245, 246, 304] 和切片得分匹配 [282] 等技术，因此我们可以直接使用其中一种技术从扰动数据点训练我们的噪声条件得分网络。例如，使用去噪得分匹配和等式 (10) 中的类似符号，训练目标如下

$$\mathbb{E}_{t \sim \mathcal{U}[1, T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \mathbf{x}_t \sim q(\mathbf{x}_t | \mathbf{x}_0)} \left[\lambda(t) \sigma_t^2 \left\| \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t) - s_{\theta}(\mathbf{x}_t, t) \right\|^2 \right] \quad (11)$$

$$\stackrel{(i)}{=} \mathbb{E}_{t \sim \mathcal{U}[1, T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \mathbf{x}_t \sim q(\mathbf{x}_t | \mathbf{x}_0)} \left[\lambda(t) \sigma_t^2 \left\| \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t | \mathbf{x}_0) - s_{\theta}(\mathbf{x}_t, t) \right\|^2 \right] + \text{const} \quad (12)$$

$$\stackrel{(ii)}{=} \mathbb{E}_{t \sim \mathcal{U}[1, T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \mathbf{x}_t \sim q(\mathbf{x}_t | \mathbf{x}_0)} \left[\lambda(t) \left\| -\frac{\mathbf{x}_t - \mathbf{x}_0}{\sigma_t} - s_{\theta}(\mathbf{x}_t, t) \right\|^2 \right] + \text{const} \quad (13)$$

$$\stackrel{(iii)}{=} \mathbb{E}_{t \sim \mathcal{U}[1, T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\lambda(t) \left\| \epsilon + s_{\theta}(\mathbf{x}_t, t) \right\|^2 \right] + \text{const}, \quad (14)$$

其中 (i) 由 [304] 推导而来，(ii) 来自假设 $q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0, \sigma_t^2 \mathbf{I})$ ，以及 (iii) 来自以下事实： $\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon$ 。同样，我们用 $\lambda(t)$ 表示正权重函数，“const”表示不依赖于可训练参数 θ 的常数。将等式 (14) 与等式 (10) 进行比较，很明显，一旦我们设置 $\epsilon_{\theta}(\mathbf{x}, t) = -s_{\theta}(\mathbf{x}, t)$ ，DDPM 和 SGM 的训练目标是等价的。此外，我们可以将得分匹配推广到更高阶。数据密度的高阶导数提供了关于数据分布的额外局部信息。Meng 等人。[209] 提出了一种广义去噪得分匹配方法来有效地估计高阶得分函数。所提出的模型可以提高朗之万动力学的混合速度，从而提高扩散模型的采样效率。

对于样本生成，SGM 利用迭代方法连续从 $s_{\theta}(\mathbf{x}, T)$ 、 $s_{\theta}(\mathbf{x}, T-1)$ 、 $\dots$ 、 $s_{\theta}(\mathbf{x}, 0)$ 中生成样本。由于 SGM 中训练和推理的分离，存在许多采样方法，其中一些将在下一节中讨论。在这里，我们介绍 SGM 的第一个采样方法，称为退火朗之万

动态模型 (ALD) [280]。令 N 为每时间步的迭代次数, $s_t > 0$ 为步长。我们首先用 $\mathbf{x}_T^{(N)} \sim \mathcal{N}(0, \mathbf{I})$ 初始化 ALD, 然后依次对 $t = T, T-1, \dots, 1$ 应用朗之万蒙特卡洛方法。在每个时间步 $0 \leq t < T$, 我们从 $\mathbf{x}_t^{(0)} = \mathbf{x}_{t+}^{(N)}$ 开始, 然后根据以下更新规则对 $i = 0, 1, \dots, N-1$ 进行迭代:

$$\begin{aligned}\boldsymbol{\epsilon}^{(i)} &\leftarrow \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ \mathbf{x}_t^{(i+1)} &\leftarrow \mathbf{x}_t^{(i)} + \frac{1}{2}s_t \mathbf{s}_\theta(\mathbf{x}_t^{(i)}, t) + \sqrt{s_t} \boldsymbol{\epsilon}^{(i)}.\end{aligned}$$

朗之万蒙特卡罗理论 [227] 保证当 $s_t \rightarrow 0$ 和 $N \rightarrow \infty$ 时, $\mathbf{x}_t^{(N)}$ 成为来自数据分布 $q(\mathbf{x}_0)$ 的有效样本。

2.3 随机微分方程 (Score SDE)

DDPM 和 SGM 可以进一步推广到无限时间步长或噪声水平的情况, 其中扰动和去噪过程是随机微分方程 (SDE) 的解。我们将此公式称为 Score SDE [285], 因为它利用 SDE 进行噪声扰动和样本生成, 而去噪过程需要估计噪声数据分布的得分函数。

评分随机微分方程 (SDE) 通过以下扩散过程将数据扰动为噪声, 该扩散过程由以下随机微分方程 (SDE) 控制 [285]:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} \quad (15)$$

其中 $\mathbf{f}(\mathbf{x}, t)$ 和 $g(t)$ 是 SDE 的扩散和漂移函数, $\mathbf{w}$ 是标准维纳过程 (又称布朗运动)。DDPM 和 SGM 中的前向过程都是此 SDE 的离散化。如 Song 等人 (2020 年) [285] 所示, 对于 DDPM, 相应的 SDE 为:

$$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}dt + \sqrt{\beta(t)}d\mathbf{w} \quad (16)$$

其中 $\beta(\frac{t}{T}) = T\beta_t$ 当 T 趋于无穷大; 对于 SGM, 相应的 SDE 由下式给出:

$$d\mathbf{x} = \sqrt{\frac{d[\sigma(t)^2]}{dt}}d\mathbf{w}, \quad (17)$$

其中 $\sigma(\frac{t}{T}) = \sigma_t$ 随着 T 趋于无穷大。这里我们使用 $q_t(\mathbf{x})$ 来表示前向过程中 $\mathbf{x}_t$ 的分布。

至关重要的是, 对于任何形式为方程 (15) 的扩散过程, Anderson [6] 证明它可以通过求解以下逆时间 SDE 来逆转:

$$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log q_t(\mathbf{x})] dt + g(t)d\bar{\mathbf{w}} \quad (18)$$

其中 $\bar{\mathbf{w}}$ 是时间倒流时的标准维纳过程, dt 表示无穷小的负时间步长。此逆 SDE 的解轨迹与正向 SDE 的解轨迹具有相同的边际密度, 只是它们沿相反的时间方向演变 [285]。直观地看, 逆时间 SDE 的解是将噪声逐渐转换为数据的扩散过程。此外, Song 等人 (2020 年) [285] 证明了存在一个常微分方程 (ODE), 即概率流 ODE, 其轨迹具有与逆时间 SDE 相同的边际密度。概率流 ODE 由以下公式给出:

$$d\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g(t)^2 \nabla_{\mathbf{x}} \log q_t(\mathbf{x}) \right] dt. \quad (19)$$

逆时间 SDE 和概率流 ODE 都允许从相同的数据分布中进行采样，因为它们的轨迹具有相同的边际。

一旦知道了每个时间步 t 的得分函数 $\nabla_x \log q_t(x)$ ，我们就可以解锁逆时间 SDE（等式 (18)）和概率流 ODE（等式 (19)），然后可以使用各种数值技术对它们进行求解来生成样本，例如退火朗之万动力学 [280]（参见第 2.2 节）、数值 SDE 求解器 [136, 285]、数值 ODE 求解器 [141, 188, 277, 285, 371] 和预测校正方法（MCMC 和数值 ODE/SDE 求解器的组合）[285]。与在 SGM 中一样，我们参数化一个时间相关的得分模型 $s_\theta(x_t, t)$ ，以通过将等式 (14) 中的得分匹配目标推广到连续时间来估计得分函数，从而得到以下目标：

$$\mathbb{E}_{t \sim \mathcal{U}[0, T], x_0 \sim q(x_0), x_t \sim q(x_t | x_0)} \left[\lambda(t) \left\| s_\theta(x_t, t) - \nabla_{x_t} \log q_{0t}(x_t | x_0) \right\|^2 \right], \quad (20)$$

其中 $\mathcal{U}[0, T]$ 表示 $[0, T]$ 上的均匀分布，其余符号遵循公式 (14)。

后续的扩散模型研究主要从三个主要方向改进这些经典方法（DDPM、SGM 和 Score SDE）：更快更高效的采样、更准确的似然和密度估计，以及处理具有特殊结构的数据（如置换不变性、流形结构和离散数据）。我们将在接下来的三节（第 3 至 5 节）中广泛概述每个方向。在表 1 中，我们列出了三种类型的扩散模型，并对其进行了更详细的分类、对应的文章和年份，以及连续和离散时间设置。

3 高效采样的扩散模型

从扩散模型生成样本通常需要涉及大量评估步骤的迭代方法。最近的大量工作都集中在加快采样过程的同时提高所得样本的质量。我们将这些有效的采样方法分为两大类：一类不涉及学习（无学习采样），另一类在扩散模型训练后需要额外的学习过程（基于学习的采样）。

3.1 无学习采样

许多扩散模型的采样器依赖于离散化方程 (18) 中的逆时间 SDE 或方程 (19) 中的概率流 ODE。由于采样成本与离散化时间步长的数量成比例增加，许多研究人员专注于开发离散化方案，以减少时间步长数量，同时最大限度地减少离散化误差。

3.1.1 SDE 求解器。 DDPM [111, 275] 的生成过程可以看作是逆时间 SDE 的一种特殊离散化。如第 2.3 节所述，DDPM 的正向过程将公式 (16) 中的 SDE 离散化，其对应的逆 SDE 形式为

$$dx = -\frac{1}{2}\beta(t)(x_t - \nabla_{x_t} \log q_t(x_t))dt + \sqrt{\beta(t)}dw \quad (21)$$

Song 等人（2020）[285] 证明，由公式 (5) 定义的逆马尔可夫链相当于公式 (21) 的数值 SDE 求解器。

噪声条件得分网络 (NCSN) [280] 和临界阻尼朗之万扩散 (CLD) [66] 都从朗之万动力学中得到启发，解决了逆时间 SDE。具体来说，NCSN 利用退火朗之万动力学（ALD，参见第 2.2 节）迭代生成数据，同时平滑地降低噪声水平，直到生成的

表 1. 列出了连续和离散设置下的三种传播模型以及相应的文章和年份。

Primary	Secondary	Tertiary	Article	Year	Setting
Efficient Sampling	Learning-Free Sampling	SDE Solvers	Song et al. [285]	2020	Continuous
			Dockhorn et al. [66]	2021	Continuous
			Jolicœur et al. [137]	2021	Continuous
			Jolicœur et al. [136]	2021	Continuous
			Chuang et al. [48]	2022	Continuous
			Song et al. [280]	2019	Continuous
		ODE Solvers	Karras et al. [141]	2022	Continuous
			Liu et al. [181]	2021	Continuous
			Song et al. [277]	2020	Continuous
			Zhang et al. [372]	2022	Continuous
	Learning-Based Sampling	Optimized Discretization	Karras et al. [141]	2022	Continuous
			Lu et al. [188]	2022	Continuous
			Zhang et al. [371]	2022	Continuous
			Watson et al. [313]	2021	Discrete
			Watson et al. [312]	2021	Discrete
			Dockhorn et al. [67]	2021	Continuous
		Knowledge Distillation	Salimans et al. [260]	2021	Discrete
			Luhman et al. [190]	2021	Discrete
		Truncated Diffusion	Meng et al. [205]	2022	Discrete
			Lyu et al. [199]	2022	Discrete
Improved Likelihood	Noise Schedule Optimization	Noise Schedule Optimization	Zheng et al. [381]	2022	Discrete
			Nichol et al. [214]	2021	Discrete
			Kingma et al. [154]	2021	Discrete
			Huang et al. [124]	2024	Discrete
			Yang et al. [351]	2024	Discrete
	Reverse Variance Learning	Reverse Variance Learning	Bao et al. [13]	2021	Discrete
			Nichol et al. [214]	2021	Discrete
	Exact Likelihood Computation	Exact Likelihood Computation	Song et al. [279]	2021	Continuous
			Huang et al. [119]	2021	Continuous
			Song et al. [285]	2020	Continuous
Data with Special Structures	Manifold Structures	Learned Manifolds	Lu et al. [187]	2022	Continuous
			Vahdat et al. [299]	2021	Continuous
			Yang et al. [346]	2024	Discrete
			Ramesh et al. [243]	2022	Discrete
			Rombach et al. [255]	2022	Discrete
		Known Manifolds	Bortoli et al. [56]	2022	Continuous
			Huang et al. [118]	2022	Continuous
			Niu et al. [219]	2020	Discrete
	Data with Invariant Structures	Data with Invariant Structures	Jo et al. [134]	2022	Continuous
			Shi et al. [267]	2022	Continuous
			Xu et al. [333]	2021	Discrete
			Meng et al. [204]	2022	Discrete
			liu et al. [184]	2023	Continuous
	Discrete Data	Discrete Data	Sohl et al. [275]	2015	Discrete
			Austin et al. [9]	2021	Discrete
			Xie et al. [326]	2022	Discrete
			Gu et al. [98]	2022	Discrete
			Campbell et al. [30]	2022	Continuous

数据分布收敛到原始数据分布。虽然 ALD 的采样轨迹不是逆时间 SDE 的精确解，但它们具有正确的边际，因此在假设朗之万动力学在每个噪声水平下都收敛到其平衡点的情况下产生正确的样本。ALD 方法通过一致退火采样 (CAS) [137] 得到进一步改进，这是一种基于分数的 MCMC 方法，可以更好地缩放时间步长和添加的噪声。受统计力学的启发，CLD 提出了一种增强 SDE，其辅助速度项类似于欠阻尼朗之万扩散。为了获得扩展 SDE 的时间反转，CLD 只需要

学习给定数据的速度条件分布的得分函数，可以说比直接学习数据得分更容易。据报道，添加的速度项可以提高采样速度和质量。

[285] 中提出的反向扩散方法以与正向随机微分方程相同的方式离散逆时间随机微分方程。对于正向随机微分方程的任何一步离散化，可以写出以下一般形式：

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{f}_i(\mathbf{x}_i) + \mathbf{g}_i \mathbf{z}_i, \quad i = 0, 1, \dots, N-1 \quad (22)$$

其中 $\mathbf{z}_i \sim \mathcal{N}(0, \mathbf{I})$ 、 $\mathbf{f}_i$ 和 $\mathbf{g}_i$ 由 SDE 的漂移/扩散系数和离散化方案决定。反向扩散建议以类似于正向 SDE 的方式离散逆时 SDE，即

$$\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{f}_{i+1}(\mathbf{x}_{i+1}) + \mathbf{g}_{i+1} \mathbf{g}_{i+1}^T s_{\theta^*}(\mathbf{x}_{i+1}, t_{i+1}) + \mathbf{g}_{i+1} \mathbf{z}_i \quad i = 0, 1, \dots, N-1 \quad (23)$$

其中 $s_{\theta^*}(\mathbf{x}_i, t_i)$ 是经过训练的噪声条件得分模型。Song 等人 (2020) [285] 证明，反向扩散方法是公式 (18) 中逆时间 SDE 的数值 SDE 求解器。此过程可应用于任何类型的正向 SDE，实证结果表明，对于一种称为 VP-SDE 的特定类型的 SDE，此采样器的性能略优于 DDPM [285]。

Jolicœur-Martineau 等人 (2021) [136] 开发了一种具有自适应步长的 SDE 求解器，以便更快地生成。通过比较高阶 SDE 求解器的输出与低阶 SDE 求解器的输出来控制步长。在每个时间步长，高阶和低阶求解器分别从前一个样本 $\mathbf{x}'_{prev}$ 生成新的样本 $\mathbf{x}'_{high}$ 和 $\mathbf{x}'_{low}$ 。然后通过比较两个样本之间的差异来调整步长。如果 $\mathbf{x}'_{high}$ 和 $\mathbf{x}'_{low}$ 相似，算法将返回 $\mathbf{x}'_{high}$ ，然后增加步长。 $\mathbf{x}'_{high}$ 和 $\mathbf{x}'_{low}$ 之间的相似性通过以下方式衡量：

$$E_q = \left\| \frac{\mathbf{x}'_{low} - \mathbf{x}'_{high}}{\delta(\mathbf{x}', \mathbf{x}'_{prev})} \right\|^2 \quad (24)$$

其中 $\delta(\mathbf{x}'_{low}, \mathbf{x}'_{prev}) := \max(\epsilon_{abs}, \epsilon_{rel} \max(|\mathbf{x}'_{low}|, |\mathbf{x}'_{prev}|))$ ，以及 ϵ_{abs} 和 ϵ_{rel} 是绝对公差和相对公差。

[285] 中提出的预测-校正方法通过结合数值 SDE 求解器（“预测器”）和迭代马尔可夫链蒙特卡罗 (MCMC) 方法（“校正器”）来解决逆 SDE。在每个时间步骤中，预测-校正方法首先使用数值 SDE 求解器生成粗样本，然后使用“校正器”通过基于分数的 MCMC 校正样本的边际分布。得到的样本具有与逆时间 SDE 的解轨迹相同的时间边际，即它们在所有时间步骤中的分布相同。经验结果表明，添加基于朗之万蒙特卡罗的校正器比使用没有校正器的额外预测器更有效 [285]。Karras 等人 (2022 年) [141] 通过提出类似朗之万的添加和去除噪声的“搅动”步骤进一步改进了 [285] 中的朗之万动力学校正器，实现了新的最先进的样本在 CIFAR-10 [16] 和 ImageNet-64 [58] 等数据集上实现了较高的质量。

3.1.2 ODE 求解器。大量关于更快扩散采样器的研究都是基于求解第 2.3 节中介绍的概率流 ODE（等式 (19)）。与 SDE 求解器相比，ODE 求解器的轨迹是确定性的，因此不受随机波动的影响。这些确定性 ODE 求解器的收敛速度通常比随机求解器快得多，但样本质量略差。

去噪扩散隐式模型 (DDIM) [277] 是加速扩散模型采样的最早研究之一。最初的动机是将原始 DDPM 扩展到非马尔可夫情况，并采用以下马尔可夫

链

$$q(\mathbf{x}_1, \dots, \mathbf{x}_T | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) \quad (25)$$

$$q_\sigma(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1} | \tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \sigma_t^2 \mathbf{I}) \quad (26)$$

$$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) := \sqrt{\bar{\alpha}_{t-1}} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0}{\sqrt{1 - \bar{\alpha}_t}} \quad (27)$$

该公式将 DDPM 和 DDIM 封装为特殊情况，其中 DDPM 对应于设置 $\sigma_t^2 = \frac{\hat{\beta}_{t-1}}{\hat{\beta}_t} \beta_t$ ，DDIM 对应于设置 $\sigma_t^2 = 0$ 。DDIM 学习马尔可夫链来逆转这个非马尔可夫扰动过程，当 $\sigma_t^2 = 0$ 时，该过程是完全确定性的。在 [141, 188, 260, 277] 中观察到，DDIM 采样过程相当于概率流 ODE 的特殊离散化方案。受对单例数据集上的 DDIM 分析的启发，广义去噪扩散隐式模型 (gDDIM) [372] 提出了一种改进的得分网络参数化方法，使得能够对更一般的扩散过程进行确定性采样，例如临界阻尼朗之万扩散 (CLD) [66] 中的扩散过程。PNDM [181] 提出了一种伪数值方法来沿 $\mathcal{R}^N$ 中的特定流形生成样本。它使用带有非线性传递部分的数值求解器来求解流形上的微分方程，然后生成样本，其中将 DDIM 封装为特例。

通过大量的实验调查，Karras 等人 (2022) [141] 表明 Heun 的 2^{nd} 阶方法 [8] 在样本质量和采样速度之间实现了极好的平衡。高阶求解器可以减小离散化误差，但代价是每个时间步对学习到的得分函数进行一次额外的评估。Heun 的方法以更少的采样步骤生成与欧拉方法质量相当甚至更好的样本。

扩散指数积分器采样器 [371] 和 DPM 求解器 [188] 利用概率流 ODE 的半线性结构来开发比通用 Runge-Kutta 方法更高效的定制 ODE 求解器。具体而言，概率流 ODE 的线性部分可以通过分析计算得到，而非线性部分可以用类似于 ODE 求解器领域中的指数积分器的技术来解决。这些方法包含 DDIM 作为一阶近似。然而，它们也允许使用更高阶的积分器，只需 10 到 20 次迭代就可以产生高质量的样本——远远少于没有加速采样的扩散模型通常需要的数百次迭代。

3.2 基于学习的抽样

基于学习的采样是另一种有效的扩散模型方法。通过使用部分步骤或训练采样器进行逆向过程，此方法可以实现更快的采样速度，但样本质量会略有下降。与使用手动步骤的无学习方法不同，基于学习的采样通常涉及通过优化某些学习目标来选择步骤。

3.2.1 优化离散化。 给定一个预先训练的扩散模型，Watson 等人 (2021) [313] 提出了一种策略，通过选择最佳 K 时间步长来最大化 DDPM 的训练目标，从而找到最佳离散化方案。这种方法的关键是观察到 DDPM 目标可以分解为各个项的总和，这使其非常适合动态规划。然而，众所周知，用于 DDPM 训练的变分下限与样本质量没有直接关系 [294]。随后的一项名为可微分扩散采样器搜索 [312] 的研究通过直接优化一个称为核初始距离 (KID) [22] 的样本质量通用指标来解决这个问题。这种优化在重新参数化的帮助下是可行的 [156, 252]

和梯度重新实现。基于截断泰勒方法, Dockhorn 等人 (2022 年) [67] 通过在一阶得分网络之上训练一个附加头, 推导出一个用于加速合成的二阶求解器。

3.2.2 截断扩散。可以通过截断正向和反向扩散过程来提高采样速度 [199, 381]。关键思想是在仅几步之后尽早停止正向扩散过程, 并以非高斯分布开始反向去噪过程。可以通过扩散来自预训练生成模型 (例如变分自动编码器 [156, 252] 或生成对抗网络 [88]) 的样本来有效地获得来自该分布的样本。

3.2.3 知识蒸馏。使用知识蒸馏的方法 [190, 205, 260] 可以显著提高扩散模型的采样速度。具体来说, 在渐进式蒸馏 [260] 中, 作者提出将整个采样过程蒸馏为一个更快的采样器, 只需要一半的步骤。通过将新的采样器参数化为深度神经网络, 作者能够训练采样器以匹配 DDIM 采样过程的输入和输出。重复此过程可以进一步减少采样步骤, 尽管步骤减少会导致样本质量下降。为了解决这个问题, 作者建议对扩散模型进行新的参数化, 并为目标函数采用新的加权方案。

4 改进似然法的扩散模型

如第 2.1 节所述, 扩散模型的训练目标是对数似然的 (负) 变分下界 (VLB)。然而, 在许多情况下, 这个界限可能不够严格 [154], 导致扩散模型的对数似然可能不是最优的。在本节中, 我们回顾了最近关于扩散模型似然最大化的研究。我们重点关注三类方法: 噪声调度优化、反向方差学习和精确对数似然评估。

4.1 噪声调度优化

在经典的扩散模型公式中, 前向过程中的噪声调度是手工制作的, 没有可训练的参数。通过联合优化前向噪声调度和扩散模型的其他参数, 可以进一步最大化 VLB, 从而实现更高的对数似然值 [154, 214]。

iDDPM [214] 的工作表明, 某种余弦噪声调度可以提高对数似然率。具体来说, 他们工作中的余弦噪声调度采用以下形式

$$\bar{\alpha}_t = \frac{h(t)}{h(0)}, \quad h(t) = \cos \left(\frac{t/T + m}{1 + m} \cdot \frac{\pi}{2} \right)^2 \quad (28)$$

其中 $\bar{\alpha}_t$ 和 β_t 在公式 (2) 和 (3) 中定义, m 是控制 $t = 0$ 处噪声尺度的超参数。他们还提出了一种在对数域中在 β_t 和 $1 - \bar{\alpha}_t$ 之间进行插值来参数化逆方差。

在变分扩散模型 (VDM) [154] 中, 作者提出通过联合训练噪声计划和其他扩散模型参数来最大化 VLB, 从而提高连续时间扩散模型的似然性。他们使用单调神经网络 $y_\eta(t)$ 对噪声计划进行参数化, 并根据 $\sigma_t^2 = \text{sigmoid}(y_\eta(t))$ 、 $q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\bar{\alpha}_t \mathbf{x}_0, \sigma_t^2 \mathbf{I})$ 和 $\bar{\alpha}_t = \sqrt{(1 - \sigma_t^2)}$ 构建前向扰动过程。此外, 作者证明了数据点 $\mathbf{x}$ 的 VLB 可以简化为仅依赖于信噪比 $R(t) := \frac{\bar{\alpha}_t^2}{\sigma_t^2}$ 的形式。具体而言, L_{VLB} 可以分解为

$$L_{VLB} = -\mathbb{E}_{\mathbf{x}_0} \text{KL}(q(\mathbf{x}_T | \mathbf{x}_0) || p(\mathbf{x}_T)) + \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_1} \log p(\mathbf{x}_0 | \mathbf{x}_1) - L_D, \quad (29)$$

其中第一项和第二项可以类似于训练变分自动编码器直接进行优化。第三项可以进一步简化为以下内容：

$$L_D = \frac{1}{2} \mathbb{E}_{\mathbf{x}_0, \epsilon \sim \mathcal{N}(0, I)} \int_{R_{\min}}^{R_{\max}} \|\mathbf{x}_0 - \tilde{\mathbf{x}}_\theta(\mathbf{x}_0, v)\|_2^2 dv, \quad (30)$$

其中 $R_{\max} = R(1)$, $R_{\min} = R(T)$, $\mathbf{x}_v = \bar{\alpha}_v \mathbf{x}_0 + \sigma_v \epsilon$ 表示通过前向扰动过程扩散 $\mathbf{x}_0$ 直到 $t = R^{-1}(v)$ 得到的噪声数据点, $\tilde{\mathbf{x}}_\theta$ 表示扩散模型预测的无噪声数据点。因此, 只要噪声时间表在 $R_{\min}$ 和 $R_{\max}$ 处共享相同的值, 它们就不会影响 VLB, 并且只会影响 VLB 的蒙特卡洛估计量的方差。

另一项研究 [123, 351] 提出通过整合跨模态信息来修改扩散轨迹。具体来说, 从任何条件输入 y 和原始样本 $\mathbf{x}_0$ 与关系网络 $r_\phi(\cdot)$ 中提取跨模态信息, 表示为 $r_\phi(y, \mathbf{x}_0)$ 。然后可以将其作为附加偏差注入前向过程以调整扩散轨迹：

$$q_t(\mathbf{x}_t | \mathbf{x}_0, y) = \mathcal{N}(\mathbf{x}_t, \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + k_t r_\phi(\mathbf{x}_0, y), (1 - \bar{\alpha}_t) I) \quad (31)$$

其中 k_t 是控制偏差项幅度的非负标量。值得注意的是, 经过这种修改, 前向过程不再是马尔可夫链。Context Diff [351] 引入了一个通用框架来联合学习跨模态关系网络 r_ϕ 和扩散模型, 并推导出这种修改后的扩散过程的 VLB 和采样程序。

4.2 反向方差学习

扩散模型的经典公式假设逆马尔可夫链中的高斯转移核具有固定的方差参数。回想一下, 我们在公式 (5) 中将逆核公式化为 $q_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$, 但通常将逆方差 $\Sigma_\theta(\mathbf{x}_t, t)$ 固定为 $\beta_t I$ 。许多方法还建议训练逆方差, 以进一步最大化 VLB 和对数似然值。

在 iDDPM [214] 中, Nichol 和 Dhariwal 提出通过用线性插值形式参数化反向方差并使用混合目标对其进行训练来学习反向方差。这可以提高对数似然率并加快采样速度, 同时不会损失样本质量。具体而言, 他们将等式 (5) 中的反向方差参数化为：

$$\Sigma_\theta(\mathbf{x}_t, t) = \exp(\theta \cdot \log \beta_t + (1 - \theta) \cdot \log \tilde{\beta}_t), \quad (32)$$

其中 $\tilde{\beta}_t := \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \cdot \beta_t$ 和 θ 联合训练以最大化 VLB。这种简单的参数化避免了估计更复杂形式的 $\Sigma_\theta(\mathbf{x}_t, t)$ 的不稳定性, 并且据报道可以提高似然值。

Analytic-DPM [13] 展示了一个显著的结果, 即可以从预先训练的得分函数中获得最佳逆方差, 其解析形式如下：

$$\Sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2 + \left(\sqrt{\frac{\tilde{\beta}_t}{\alpha_t}} - \sqrt{\beta_{t-1}} - \sigma_t^2 \right)^2 \cdot \left(1 - \bar{\beta}_t \mathbb{E}_{q_t(\mathbf{x}_t)} \frac{\|\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t)\|^2}{d} \right) \quad (33)$$

因此, 给定一个预先训练好的评分模型, 我们可以估计其一阶和二阶矩以获得最佳的反向方差。将它们插入 VLB 可以得到更紧密的 VLB 和更高的似然值。

4.3 精确似然计算

在 Score SDE [285] 公式中, 样本是通过求解以下逆 SDE 生成的, 其中, 等式 (18) 中的 $\nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t, t)$ 被学习到的噪声条件得分模型 $\mathbf{s}_\theta(\mathbf{x}_t, t)$ 取代:

$$d\mathbf{x} = f(\mathbf{x}_t, t) - g(t)^2 \mathbf{s}_\theta(\mathbf{x}_t, t) dt + g(t) d\mathbf{w}. \quad (34)$$

这里我们使用 p_θ^{sde} 来表示通过求解上述 SDE 生成的样本分布。也可以通过将得分模型代入公式 (19) 中的概率流 ODE 来生成数据, 从而得到:

$$\frac{d\mathbf{x}_t}{dt} = \underbrace{f(\mathbf{x}_t, t) - \frac{1}{2} g^2(t) \mathbf{s}_\theta(\mathbf{x}_t, t)}_{:= \tilde{f}_\theta(\mathbf{x}_t, t)} \quad (35)$$

类似地, 我们使用 p_θ^{ode} 表示通过求解此 ODE 生成的样本的分布。神经 ODE 理论 [40] 和连续正则化流 [92] 表明, 尽管计算成本较高, 但可以准确计算 p_θ^{ode} 对于 p_θ^{sde} 几项同时进行的工作 [119, 187, 279] 表明存在一个可有效计算的变分下界, 我们可以使用修改后的扩散损失直接训练我们的扩散模型以最大化 p_θ^{sde}

具体而言, Song 等人 (2021 年) [279] 证明, 使用特殊的加权函数 (似然加权), 用于训练得分 SDE 的目标隐式地最大化了数据上 p_θ^{sde} 的预期值。结果表明

$$\mathbf{D}_{KL}(q_0 \parallel p_\theta^{\text{sde}}) \leq \mathcal{L}(\theta; g(\cdot)^2) + \mathbf{D}_{KL}(q_T \parallel \pi), \quad (36)$$

其中 $\mathcal{L}(\theta; g(\cdot)^2)$ 是等式 (20) 中带有 $\lambda(t) = g(t)^2$ 的 Score SDE 目标。由于 $\mathbf{D}_{KL}(q_0 \parallel p_\theta^{\text{sde}}) = -\mathbb{E}_{q_0} \log(p_\theta^{\text{sde}}) + \text{const}$, 而 $\mathbf{D}_{KL}(q_T \parallel \pi)$ 为常数, 因此使用 $\mathcal{L}(\theta; g(\cdot)^2)$ 进行训练相当于最小化 $-\mathbb{E}_{q_0} \log(p_\theta^{\text{sde}})$, 即数据的预期负对数似然。此外, Song 等人 (2021) 和 Huang 等人 (2021) [119, 279] 为 $p_\theta^{\text{sde}}(\mathbf{x})$ 提供了以下界限:

$$-\log p_\theta^{\text{sde}}(\mathbf{x}) \leq \mathcal{L}'(\mathbf{x}), \quad (37)$$

其中 $\mathcal{L}'(\mathbf{x})$ 定义为

$$\mathcal{L}'(\mathbf{x}) := \int_0^T \mathbb{E} \left[\frac{1}{2} \|g(t) \mathbf{s}_\theta(\mathbf{x}_t, t)\|^2 + \nabla \cdot (g(t)^2 \mathbf{s}_\theta(\mathbf{x}_t, t) - f(\mathbf{x}_t, t)) \middle| \mathbf{x}_0 = \mathbf{x} \right] dt - \mathbb{E}_{\mathbf{x}_T} [\log p_\theta^{\text{sde}}(\mathbf{x}_T) | \mathbf{x}_0 = \mathbf{x}] \quad (38)$$

等式 (38) 的第一部分让人联想到隐式得分匹配 [126], 并且可以用蒙特卡洛方法有效地估计整个界限。

由于概率流 ODE 是神经 ODE 或连续正则化流的特例, 我们可以使用这些领域中成熟的方法来准确计算 $\log p_\theta^{\text{ode}}$ 具体来说, 我们有

$$\log p_\theta^{\text{ode}}(\mathbf{x}_0) = \log p_T(\mathbf{x}_T) + \int_{t=0}^T \nabla \cdot \tilde{f}_\theta(\mathbf{x}_t, t) dt. \quad (39)$$

可以使用数值 ODE 求解器和 Skilling-Hutchinson 迹估计器 [125, 274] 计算上述一维积分。不幸的是, 这个公式不能直接优化以最大化数据的 p_θ^{ode} 因为它需要为每个数据点 $\mathbf{x}_0$ 调用昂贵的 ODE 求解器。为了降低使用上述公式直接最大化 p_θ^{ode} 的成本, Song 等人 (2021) [279] 提出最大化 p_θ^{sde} 的变分下界作为最大化 p_θ^{ode} 的代理, 从而产生了一类称为 ScoreFlows 的扩散模型。

Lu 等人 (2022) [187] 进一步改进了 ScoreFlows, 他们提出不仅要最小化 vanilla 分数匹配损失函数, 还要最小化其高阶泛化。他们证明 $\log p_{\text{ode}}^{\theta}$ 可以与一阶、二阶和三阶分数匹配误差相界。基于这一理论结果, 作者进一步提出了有效的训练算法来最小化高阶分数匹配损失, 并报告了数据上改进的 p_{ode}^{θ}

5 特殊结构数据的扩散模型

虽然扩散模型在图像和音频等数据领域取得了巨大成功, 但它们不一定能无缝地转换到其他模态。许多重要的数据域具有特殊结构, 必须考虑到这些结构才能使扩散模型有效发挥作用。例如, 当模型依赖于仅在连续数据域上定义的得分函数时, 或者当数据驻留在低维流形上时, 可能会出现困难。为了应对这些挑战, 必须以各种方式调整扩散模型。

5.1 离散数据

大多数扩散模型都面向连续数据域, 因为 DDPM 中使用的高斯噪声扰动并不自然地适合离散数据, 并且 SGM 和 Score SDE 所需的得分函数仅在连续数据域上定义。为了克服这一困难, 一些研究 [9, 98, 117, 326] 在 Sohl-Dickstein 等人 (2015 年) [275] 的基础上构建, 以生成高维离散数据。具体来说, VQ-Diffusion [98] 用离散数据空间上的随机游走或随机掩码操作取代高斯噪声。前向过程的结果转换核采用以下形式

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathbf{v}^T(\mathbf{x}_t) \mathbf{Q}_t \mathbf{v}(\mathbf{x}_{t-1}) \quad (40)$$

其中 $\mathbf{v}(\mathbf{x})$ 是独热列向量, $\mathbf{Q}_t$ 是惰性随机游走的过渡核。D3PM [9] 通过使用吸收状态核或离散化高斯核构建前向噪声过程来适应扩散模型中的离散数据。Campbell 等人 (2022) [30] 提出了第一个离散扩散模型的连续时间框架。利用连续时间马尔可夫链, 他们能够推导出优于离散对应者的高效采样器, 同时对样本分布与真实数据分布之间的误差进行理论分析。

具体得分匹配 (CSM) [204] 提出了离散随机变量得分函数的推广。具体得分定义为概率相对于输入方向变化的变化率, 可视为连续 (Stein) 得分的有限差分近似。具体得分可以高效训练并应用于 MCMC。

基于随机微积分理论, 刘等人 (2023) [184] 提出了一个扩散模型框架, 用于在受限和结构化域上生成数据, 包括离散数据作为特殊情况。利用随机微积分的基本定理 Doob 的 h 变换, 可以通过在反向扩散过程中加入一个特殊的力项来限制特定区域上的数据分布。他们使用基于 EM 的优化算法对力项进行参数化。此外, 可以使用 Girsanov 定理将损失函数转换为 L_2 损失。

5.2 具有不变结构的数据

许多重要领域的数据具有不变结构。例如, 图是置换不变的, 点云是平移和旋转不变的。在扩散模型中, 这些不变性经常被忽略, 这可能

导致性能不理想。为了解决这个问题，一些研究 [56, 219] 提出赋予扩散模型考虑数据不变性的能力。

Niu 等人 (2020 年) [219] 首次利用扩散模型解决了置换不变图生成问题。他们通过使用置换等变图神经网络 [89, 265, 322] (称为 EDP-GNN) 来参数化噪声条件得分模型，从而实现这一目标。GDSS [134] 通过提出连续时间图扩散过程进一步发展了这一想法。该过程通过一组随机微分方程 (SDE) 对节点和边的联合分布进行建模，其中使用消息传递操作来保证置换不变性。

类似地，Shi 等人 (2021) [267] 和 Xu 等人 (2022) [333] 使扩散模型能够生成对平移和旋转均不变的分子构象。例如，Xu 等人 (2022) [333] 表明，以不变先验开始并以等变马尔可夫核演化的马尔可夫链可以诱导不变的边际分布，这可用于在分子构象生成中强制适当的数据不变性。正式地，让 $\mathcal{T}$ 成为旋转或平移操作。给定 $p(x_T) = p(\mathcal{T}(x_T))$, $p_\theta(x_{t-1} | x_t) = p_\theta(\mathcal{T}(x_{t-1}) | \mathcal{T}(x_t))$, Xu 等人 (2022) [333] 证明样本分布保证对 $\mathcal{T}$ 不变，即 $p_0(x) = p_0(\mathcal{T}(x))$ 。因此，只要先验核和过渡核具有相同的不变性，就可以建立一个扩散模型，生成旋转和平移不变的分子构象。

5.3 具有流形结构的数据

具有流形结构的数据在机器学习中无处不在。正如流形假设 [76] 所假设的那样，自然数据通常位于固有维数较低的流形上。此外，许多数据域具有众所周知的流形结构。例如，气候和地球数据自然位于球体上，因为这是我们星球的形状。许多研究都专注于开发流形上数据的扩散模型。我们根据流形是已知还是学习到的对它们进行分类，并在下面介绍一些代表性研究。

5.3.1 已知流形。最近的研究已将 Score SDE 公式扩展到各种已知流形。这种改编与神经 ODE [40] 和连续正则化流 [92] 到黎曼流形的推广相似 [186, 201]。为了训练这些模型，研究人员还将分数匹配和分数函数改编为黎曼流形。

基于黎曼分数的生成模型 (RSGM) [56] 适用于各种流形，包括球面和环面，只要它们满足温和的条件。RSGM 表明可以将扩散模型扩展到紧致黎曼流形。该模型还提供了在流形上反转扩散的公式。从内在角度来看，RSGM 使用测地线随机游走来近似黎曼流形上的采样过程。它使用广义去噪分数匹配目标进行训练。

相比之下，黎曼扩散模型 (RDM) [118] 采用变分框架将连续时间扩散模型推广到黎曼流形。RDM 使用对数似然的变分下界 (VLB) 作为其损失函数。RDM 模型的作者已经证明，最大化这个 VLB 相当于最小化黎曼得分匹配损失。与 RSGM 不同，RDM 采用外在观点，假设相关的黎曼流形嵌入在更高维的欧几里得空间中。

5.3.2 学习流形。根据流形假设 [76]，大多数自然数据位于具有显著降低的固有维数的流形上。因此，识别这些流形并直接在其上训练扩散模型可能具有优势，因为数据维数较低。许多最近的研究都建立在这一想法之上，首先使用自动编码器将数据压缩为低维流形，然后训练

在这个潜在空间中，我们可以使用流形模型来构建扩散模型。在这些情况下，流形由自动编码器隐式定义，并通过重构损失进行学习。为了取得成功，设计一个允许自动编码器和扩散模型联合训练的损失函数至关重要。

基于潜在分数的生成模型 (LSGM) [299] 试图通过将分数 SDE 扩散模型与变分自动编码器 (VAE) [156, 252] 配对来解决联合训练的问题。在这种配置中，扩散模型负责学习先验分布。LSGM 的作者提出了一个联合训练目标，将 VAE 的证据下限与扩散模型的分数的匹配目标合并。这为数据对数似然提供了一个新的下限。通过将扩散模型置于潜在空间内，LSGM 实现了比传统扩散模型更快的样本生成。此外，LSGM 可以通过将离散数据转换为连续潜在代码来管理离散数据。

潜在扩散模型 (LDM) [255] 不是联合训练自动编码器和扩散模型，而是分别处理每个组件。首先，训练自动编码器以产生低维潜在空间。然后，训练扩散模型以生成潜在代码。DALI-2 [243] 采用类似的策略，在 CLIP 图像嵌入空间上训练扩散模型，然后训练单独的解码器以基于 CLIP 图像嵌入创建图像。

结构引导的扩散模型对抗训练 (SADM) [346] 首次提出利用样本批次中的结构信息。具体而言，SADM 结合了对抗训练的结构鉴别器，以强制保留每个训练批次中样本之间的流形结构。这种方法利用内在数据流形来促进真实样本的生成，从而显著提升了以前的扩散模型在图像合成和跨域微调等任务中的能力。

6 与其他生成模型的联系

在本节中，我们首先介绍其他五类重要的生成模型，并分析它们的优点和局限性。然后介绍扩散模型如何与它们相联系，并说明如何通过整合扩散模型来促进这些生成模型的发展。表 2 总结了将扩散模型与其他生成模型相结合的算法，我们还在图 3 中提供了示意图。

6.1 大型语言模型及其与传播模型的联系

大型语言模型 (LLM) [1, 7, 27, 129, 348] 对人工智能界产生了深远的影响，并展示了先进的语言理解和推理能力。最近的研究开始将其令人印象深刻的推理能力扩展到视觉生成任务，以进行整体生成规划。LLM [35, 221, 348] 和扩散模型 [20, 243, 345, 351] 之间的协作可以显著改善文本-图像对齐以及生成图像的质量 [177, 295, 375]。例如，RealCompo [375] 利用 LLM 生成基于 LLM 边界框布局的图像，从而增强扩散模型的组合生成。Edit World [349] 组合了一组 LLM 和预训练的扩散模型，以生成包含大量具有世界知识的指令的图像编辑数据集 [101]。VideoTetris [295] 使用 LLM 沿时间轴分解文本提示，以更流畅、更合理的过渡指导视频生成。SemanticS DS [350] 和 Trans4D [366] 扩展了 LLM 的规划能力，以促进更复杂的 3D 和 4D 扩散生成。值得注意的是，RPG [347] 利用多模态 LLM 的视觉语言先验从文本提示中推理出互补的空间布局，并在文本引导的图像生成和编辑过程中操纵扩散模型的对象组合，在组合合成场景中实现 SOTA 性能，并为后续研究提供指导。

表 2. 扩散模型被纳入不同的生成模型。

Model	Article	Year
Large Language Model	Zhang et al. [375]	2024
	Yang et al. [349]	2024
	Yang et al. [347]	2024
	Tian et al. [295]	2024
	Yang et al. [350]	2024
	Zeng et al. [366]	2024
Variational Auto-Encoder	Luo et al. [191]	2022
	Hunag et al. [119]	2021
	Vadhat et al. [299]	2021
Generative Adversarial Network	Wang et al. [311]	2022
	Yang et al. [346]	2021
Normalizing Flow	Zhang et al.[370]	2021
	Gong et al. [87]	2021
	kim et al. [150]	2022
	Wang et al. [307]	2024
	Yang et al. [353]	2024
Autoregressive Model	Meng et al.[210]	2020
	Meng et al.[208]	2021
	Hoogeboom et al.[116]	2021
	Rasul et al. [247]	2021
Energy-based Model	Gao et al. [83]	2021
	Yu et al. [361]	2022

6.2 变分自编码器及与扩散模型的联系

变分自编码器 [68, 157, 252] 旨在学习编码器和解码器，以将输入数据映射到连续潜在空间中的值。在这些模型中，嵌入可以解释为概率生成模型中的潜在变量，概率解码器可以通过参数化的似然函数来表示。此外，假设数据 x 由某个未观测的潜在变量 z 使用条件分布 $p_\theta(x|z)$ 生成，并且 $q_\phi(z|x)$ 用于近似推断 z 。为了保证有效的推断，使用变分贝叶斯方法来最大化证据下限：

$$\mathcal{L}(\phi, \theta; \mathbf{x}) = \mathbb{E}_{q_\phi(z|\mathbf{x})} [\log p_\theta(\mathbf{x}, z) - \log q_\phi(z | \mathbf{x})] \quad (41)$$

带有 $\mathcal{L}(\phi, \theta; \mathbf{x}) \leq \log p_\theta(\mathbf{x})$ 。假设参数化似然函数 $p_\theta(x|z)$ 和参数化后验近似 $q_\phi(z|x)$ 可以逐点计算，并且可与它们的参数微分，则可以使用梯度下降最大化 ELBO。此公式允许灵活选择编码器和解码器模型。通常，这些模型由指数族分布表示，其参数由多层神经网络生成。

DDPM 可以概念化为具有固定编码器的分层马尔可夫 VAE。具体来说，DDPM 的前向过程充当编码器，该过程的结构为线性高斯模型（如公式 (2) 所述）。另一方面，DDPM 的反向过程对应于解码器，它在多个解码步骤中共享。解码器中的潜在变量的大小都与样本数据相同。

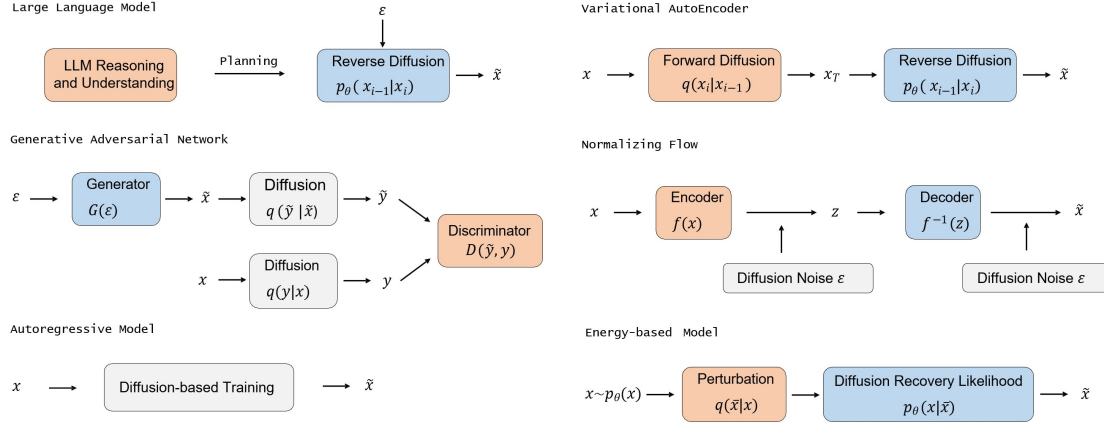


图 3. 将扩散模型与其他生成模型结合起来的示意图，例如：LLM [347]，其中扩散模型由 LLM 规划指导，VAE [255]，其中扩散模型应用于潜在空间，GAN [311]，其中噪声被注入到鉴别器输入，正则化流 [370]，其中噪声注入到流的前向和后向过程，自回归模型 [116]，其中训练目标类似于扩散模型，以及 EBM [83]，其中通过扩散恢复似然来学习一系列 EBM。

在连续时间设置中，Song 等人（2021 年）[285]、Huang 等人（2021 年）[119] 和 Kingma 等人（2021 年）[154] 证明，得分匹配目标可以通过深度分层 VAE 的证据下限 (ELBO) 来近似。因此，优化扩散模型可以看作是训练无限深的分层 VAE——这一发现支持了普遍的看法，即得分 SDE 扩散模型可以解释为分层 VAE 的连续极限。

基于潜在分数的生成模型 (LSGM) [299] 进一步推进了这一研究方向，它表明 ELBO 可以被视为潜在空间扩散背景下的专门分数匹配目标。虽然 ELBO 中的交叉熵项难以处理，但可以通过将基于分数的生成模型视为无限深的 VAE，将其转化为可处理的分数匹配目标。

6.3 基因 具有扩散性的对抗网络和连接

模型

生成对抗网络 (GAN) [51, 88, 100] 主要由两个模型组成：生成器 G 和鉴别器 D 。这两个模型通常由神经网络构建，但可以以任何形式的可微分系统实现，该系统将输入数据从一个空间映射到另一个空间。GAN 的优化可以看作是具有值函数 $V(G, D)$ 的极小-极大优化问题：

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (42)$$

生成器 G 旨在生成新示例并隐式建模数据分布。鉴别器 D 通常是一个二元分类器，用于以尽可能高的准确度从真实示例中识别出生成的示例。优化过程结束于一个鞍点，该鞍点产生生成器的最小值和鉴别器的最大值。也就是说，GAN 优化的目标是实现纳什均衡 [250]。此时，可以认为生成器已经捕获了真实示例的准确分布。

GAN 的问题之一是训练过程不稳定,这主要是由于输入数据的分布与生成数据的分布不重叠造成的。一种解决方案是在判别器输入中注入噪声,以扩大生成器和判别器分布的支持范围。利用灵活的扩散模型,Wang 等人 (2022) [311] 使用由扩散模型确定的自适应噪声计划向判别器注入噪声。另一方面,GAN 可以提高扩散模型的采样速度。Xiao 等人 (2021) [324] 表明,采样速度慢是由去噪步骤中的高斯假设引起的,这适用于较小的步长。因此,每个去噪步骤都由条件 GAN 建模,从而允许更大的步长。为了确保扩散模型能够捕捉到数据分布中真实的流形结构,SADM [346] 主张对扩散生成器进行对抗训练,以对抗极大极小博弈中的新型结构鉴别器,从而将真实的流形结构与生成的流形结构区分开来。

6.4 使用扩散模型规范流动和连接

正则化流 [63, 251] 是一种生成模型,可生成可处理的分布来对高维数据进行建模 [65, 155]。正则化流可以将简单的概率分布转换为极其复杂的概率分布,可用于生成模型、强化学习、变分推理等领域。现有的正则化流是基于变量变换公式构建的 [63, 251]。正则化流中的轨迹由微分方程表示。在离散时间设置中,正则化流中从数据 x 到潜在 z 的映射是双射序列的组合,形式为 $F = F_N \circ F_{N-1} \circ \dots \circ F_1$ 。正则化流中的轨迹 $\{x_1, x_2, \dots, x_N\}$ 满足:

$$x_i = F_i(x_{i-1}, \theta), x_{i-1} = F_i^{-1}(x_i, \theta) \quad (43)$$

对于所有 $i \leq N$ 。

与连续设置类似,归一化流允许通过改变变量公式来检索精确的对数似然。然而,双射要求限制了实际和理论背景下复杂数据的建模 [50, 317]。一些工作试图放宽这种双射要求 [65, 317]。例如,DiffFlow [370] 引入了一种生成建模算法,该算法结合了基于流的模型和扩散模型的优点。因此,与扩散概率模型相比,DiffFlow 可以产生比归一化流更清晰的边界,并且以更少的离散化步骤学习更一般的分布。隐式非线性扩散模型 (INDM) [150] 优化了潜在扩散的预编码过程,该模型首先使用归一化流将原始数据编码到潜在空间中,然后在潜在空间中进行扩散。使用非线性扩散过程,INDM 可以有效地提高似然度和采样速度。

为了扩大 CNF 的训练规模,最近的研究提出了有效的无模拟方法 [2, 179, 183],即通过参数化从噪声样本流到数据样本的矢量场。Lipman 等人 (2022) [179] 提出了流匹配 (FM) 来训练 CNF,该方法基于在噪声分布和每个数据样本之间构建显式条件概率路径。Wang 等人 (2024) [307] 对整流流 [183] 中整流的本质进行了深入分析,并将其扩展到整流扩散。此外,他们通过理论推导确定了整流扩散的基本训练目标不是直线性,而是一阶性质。Yang 等人 (2024) 进一步提出了一致性流匹配 [353],这是一种新的 FM 方法,可明确执行速度场的自治性。一致性流匹配 [353] 直接定义了从不同时间开始到同一端点的直线流,

对其速度值施加限制：

$$\begin{aligned}\mathcal{L}_\theta &= E_{t \sim \mathcal{U}} E_{x_t, x_{t+\Delta t}} \|f_\theta(t, x_t) - f_{\theta^-}(t + \Delta t, x_{t+\Delta t})\|_2^2 + \alpha \|v_\theta(t, x_t) - v_{\theta^-}(t + \Delta t, x_{t+\Delta t})\|_2^2, \\ f_\theta(t, x_t) &= x_t + (1 - t) * v_\theta(t, x_t),\end{aligned}\tag{44}$$

其中 $\mathcal{U}$ 是 $[0, 1 - \Delta t]$ 上的均匀分布， α 是正标量， Δt 表示时间间隔，它是一个小的正标量。 θ^- 表示使用指数移动平均线 (EMA) 对 θ 的过去值进行的移动平均值， x_t 和 $x_{t+\Delta t}$ 遵循预定义的分布，可以有效采样，例如 VP-SDE [111] 或 OT 路径 [179]。这样，一致性流匹配 [353] 通过以速度一致性为特征的直线流动的新概念，创新地将一致性模型和流匹配模型连接起来。

6.5 自回归模型以及与扩散模型的联系

艾尔斯

自回归模型 (ARM) 通过使用概率链式法则将数据的联合分布分解为条件分布的乘积来工作：

$$\log p(\mathbf{x}_{1:T}) = \sum_{t=1}^T \log p(x_t | \mathbf{x}_{<t})\tag{45}$$

其中 $\mathbf{x}_{<t}$ 是 $x_1, x_2, \dots, x_{t-1}$ 的简写 [17, 163]。深度学习的最新进展促进了各种数据模态 [34, 207, 264] 的重大进展，例如图像 [45, 302]、音频 [140, 301] 和文本 [18, 27, 95, 203, 211]。自回归模型 (ARM) 通过使用单个神经网络提供生成能力。从这些模型中采样需要与数据维数相同数量的网络调用。虽然 ARM 是有效的密度估计器，但采样是一个连续、耗时的过程——尤其是对于高维数据。

另一方面，自回归扩散模型 (ARDM) [116] 能够生成任意顺序的数据，包括顺序不可知的自回归模型和离散扩散模型作为特殊情况 [9, 117, 276]。ARDM 不会像 ARM 那样对表示使用因果掩蔽，而是使用与扩散概率模型类似的有效目标进行训练。在测试阶段，ARDM 能够并行生成数据，从而能够应用于一系列任意生成任务。

Ment 等人 (2021 年) [208] 将随机平滑纳入自回归生成模型，以提高样本质量。通过将原始数据分布与平滑分布（例如高斯或拉普拉斯核）进行卷积，可以使其平滑。通过自回归模型学习平滑的数据分布，然后通过应用基于梯度的去噪方法或引入另一个条件自回归模型对学习到的分布进行去噪。通过适当选择平滑度，所提出的方法可以提高现有自回归模型的样本质量，同时保留合理的似然率。

另一方面，自回归条件得分模型 (AR-CSM) [210] 提出了一种得分匹配方法来对自回归模型的条件分布进行建模。条件分布的得分函数，即 $\nabla_{x_t} \log p(x_t | \mathbf{x}_{<t})$ ，不需要进行归一化，因此可以在模型中使用更灵活、更先进的神经网络。此外，即使原始数据的维度可能非常高，也可以有效地估计单变量条件得分函数。对于推理，AR-CSM 使用朗之万动力学，只需要得分函数从密度中采样。

6.6 基于能量的模型及其与扩散模型的联系

基于能量的模型 (EBM) [36, 59, 70, 77, 81, 82, 90, 93, 94, 153, 162, 165, 213, 218, 238, 253, 325, 378] 可视为判别器的一个生成版本 [94, 130, 164, 168], 同时可以从未标记的输入数据中学习。令 $\mathbf{x} \sim p_{\text{data}}(\mathbf{x})$ 表示训练示例, $p_{\theta}(\mathbf{x})$ 表示旨在近似 $p_{\text{data}}(\mathbf{x})$ 的概率密度函数。基于能量的模型定义为:

$$p_{\theta}(\mathbf{x}) = \frac{1}{Z_{\theta}} \exp(f_{\theta}(\mathbf{x})), \quad (46)$$

其中 $Z_{\theta} = \int \exp(f_{\theta}(\mathbf{x})) d\mathbf{x}$ 是配分函数, 对于高维 $\mathbf{x}$, 该函数在分析上是难以处理的。对于图像, $f_{\theta}(\mathbf{x})$ 由具有标量输出的卷积神经网络参数化。Salimans 等人 (2021) [261] 比较了约束分数模型和基于能量的模型对数据分布分数的建模, 发现约束分数模型 (即基于能量的模型) 在使用可比模型结构时可以表现得与无约束模型一样好。

尽管 EBM 具有许多理想的特性, 但建模高维数据仍面临两个挑战。首先, 通过最大化似然来学习 EBM 需要 MCMC 方法从模型中生成样本, 这在计算上非常耗费。其次, 如 [217] 所示, 使用非收敛 MCMC 学习到的能量势是不稳定的, 因为来自长期马尔可夫链的样本可能与观察到的样本存在显著差异, 因此很难评估学习到的能量势。在最近的一项研究中, Gao 等人 (2021) [83] 提出了一种扩散恢复似然方法, 可以在扩散模型的逆过程中从一系列 EBM 中轻松学习样本。每个 EBM 都用恢复似然进行训练, 其目的是在给定更高噪声水平的噪声版本的情况下, 最大化数据在某一噪声水平下的条件概率。EBM 最大化恢复似然, 因为它比边际似然更易于处理, 因为从条件分布中采样比从边际分布中采样容易得多。

7 扩散模型的应用

扩散模型最近因其灵活性和强度而被应用于解决各种具有挑战性的现实任务。我们根据任务将这些应用分为六个不同的类别: 计算机视觉、自然语言处理、时间数据建模、多模态学习、稳健学习和跨学科应用。对于每个类别, 我们都对该任务进行了简要介绍, 然后详细解释了如何应用扩散模型来提高性能。表 3 总结了使用扩散模型的各种应用。

7.1 无条件条件和条件扩散模型

在介绍扩散模型的应用之前, 我们先来介绍扩散模型的两个基本应用范式, 即无条件扩散模型和条件扩散模型。作为生成模型, 扩散模型的发展历史与 VAE、GAN、流模型等生成模型非常相似, 都是先发展出无条件生成, 然后紧随其后的是条件生成。无条件生成通常用于探索生成模型性能的上限, 而条件生成则更多是应用层面的内容, 因为它可以让我们根据自己的意图来控制生成结果。扩散模型除了保证生成质量和样本多样性外, 尤其在可控性方面更胜一筹。非条件扩散模型的主要算法在第 2 至第 5 节已经讲得很清楚了, 下一篇我们将

表 3. 利用扩散模型的所有应用的总结。

Primary	Secondary	Article
Computer Vision	Super Resolution, Inpainting, Restoration, Translation, and Editing	[171], [259], [255], [189], [257], [236], [112], [16], [225], [49], [284], [47], [206], [143], [345], [347]
	Semantic Segmentation	[15], [25], [91], [329]
	Video Generation	[107], [114], [355], [369], [273], [110], [318], [237], [295]
	Point Cloud Completion and Generation	[384], [193], [198], [185], [367]
	Generating Data from Diffusion Models	[349], [26], [385]
Natural Language Generation	Natural Language Generation	[9], [175], [42], [86], [106], [61]
Temporal Data Modeling	Time Series Imputation	[291], [3], [228], [180]
	Time Series Forecasting	[248], [3], [180]
	Waveform Signal Processing	[39], [159]
Multi-Modal Learning	Text-to-Image Generation	[10], [243], [258], [215], [98], [256], [146], [334], [368], [345], [351], [347], [376]
	Scene Graph-to-Image Generation	[342]
	Text-to-3D/4D Generation	[331], [178], [234], [365], [350], [366]
	Text-to-Motion Generation	[292], [369], [151]
	Text-to-Video Generation	[273], [110], [318], [237], [108], [351], [295]
	Text-to-Audio Generation	[235], [336], [320], [170], [288], [120], [152]
Robust Learning	Robust Learning	[216], [357], [23], [308], [319], [287]
Interdisciplinary Applications	Molecular Graph Modeling	[133], [115], [343], [333], [298], [121], [124], [122]
	Material Design	[327], [196]
	Medical Image Reconstruction	[284], [47], [48], [49], [230], [328]

主要讨论条件扩散模型如何应用于不同应用场合和不同的条件形式，并选择一些典型场景进行演示。

7.1.1 扩散模型中的条件机制。利用不同形式的条件来指导扩散模型的生成方向被广泛使用，例如标签、分类器、文本、图像、语义图、图形等。然而，有些条件是结构性的和复杂的，因此对它们的条件化方法值得讨论。条件机制主要有四种，包括连接、基于梯度、交叉注意和自适应层规范化（adaLN）。连接意味着扩散模型将信息指导与扩散过程中的中间去噪目标连接起来，例如标签嵌入和语义特征图。基于梯度的机制将与任务相关的梯度纳入扩散采样过程以实现可控生成。例如，在图像生成中，可以在有噪声的图像上训练辅助分类器，然后使用梯度引导扩散采样过程朝向任意类标签。交叉注意在指导和扩散目标之间执行注意消息传递，这通常在去噪网络中以分层方式进行。adaLN 机制遵循了 GAN [142] 中自适应归一化层 [231] 的广泛使用，可扩展扩散模型 [229] 探索用自适应层归一化取代基于 Transformer 的扩散主干中的标准层规范层。它不是直接学习维度上的尺度和移位参数，而是从时间嵌入和条件的总和中回归它们。

7.1.2 使用 DPO/RLHF 进行扩散。基于大型语言模型 (LLM) 中强化学习人类反馈 (RLHF) 的成功 [11, 224]，扩散模型中的许多方法都尝试使用类似的方法进行模型对齐 [75, 167]。一些方法使用预先训练的奖励模型或训练新的奖励模型来指导生成过程。例如，ImageReward [330] 手动注释了一个人类偏好图像的大型数据集，并训练了一个奖励模型来评估图像与人类偏好之间的一致性。一些方法绕过奖励模型的训练，直接在人类偏好数据集上微调扩散模型 [339]。扩散-DPO [306] 重新制定了直接偏好优化 (DPO)，以解释扩散模型的似然概念，利用证据下限来推导可微分的目标。最近，Zhang 等人 [330] 重新制定了直接偏好优化 (DPO)，以解释扩散模型的似然概念。（2024）提出 IterComp [376]，以迭代方式将基础扩散模型与模型库中的组合感知模型偏好对齐，由六个强大的开源扩散模型组成，有效地提高了基础模型在条件

扩散生成。如图 4 所示，IterComp [376] 的表现优于其他三类条件扩散方法，同时实现了最佳推理效率。

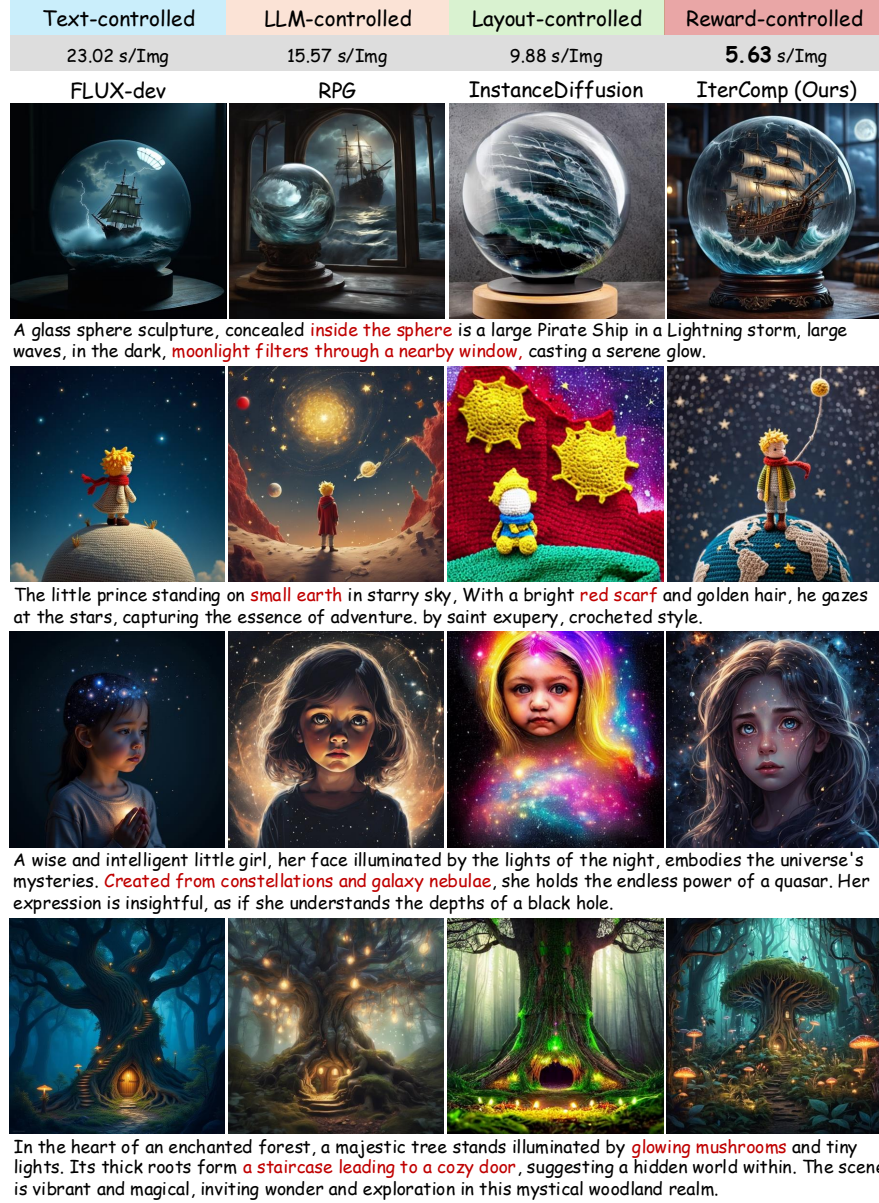


图 4. IterComp [376] 与三类组合生成方法（文本控制、LLM 控制和布局控制方法）的定性比较。彩色文本表示 IterComp [376] 在生成图像方面的优势。

7.1.3 基于标签和分类器的条件扩散。在标签的指导下对扩散过程进行条件化是将所需属性添加到生成样本中的直接方法。然而，当标签有限时，很难使扩散模型充分捕获数据的整个分布。SGGM [352] 提出了一种基于自生成的分层标签集的自引导扩散过程，而 You 等人 (2023) [359] 通过对偶伪训练证明了大规模扩散模型和半监督学习器可以通过少量标签互惠互利。Dhariwal 和 Nichol [60] 提出了分类器指导，通过使用额外训练的分类器来提高扩散模型的样本质量。Ho 和 Salimans [113] 联合训练了一个条件和非条件扩散模型，发现可以将得到的条件和非条件分数结合起来，以获得与使用分类器指导获得的样本质量和多样性之间的权衡。

7.1.4 文本、图像和语义图上的条件扩散。最近的研究开始在更多语义（如文本、图像和语义图）的指导下对扩散过程进行条件化，以更好地表达样本中的丰富语义。DiffuSeq [86] 以文本为条件，并提出了一个 seq-to-seq 扩散框架，可帮助完成四项 NLP 任务。SDEdit [206] 以样式化图像为条件进行图像到图像的转换，而 LDM [255] 将这些语义条件与灵活的潜在扩散统一起来。请注意，如果条件和扩散目标属于不同的模态，预对齐 [243, 342] 是加强引导扩散的一种实用方法。unCLIP [243] 和 ConPreDiff [345] 在文本到图像的生成中利用 CLIP 潜在特征，从而对齐图像和文本之间的语义。RPG [347] 以互补的矩形和轮廓区域为条件，实现文本到图像的组合生成和复杂的文本引导图像编辑。ContextDiff [351] 提出了一种通用的前向-后向一致性扩散模型，可以更好地适应各种输入模态。

7.1.5 图上的条件扩散。图结构数据通常表现出节点之间的复杂关系，因此对于扩散模型来说，对图进行条件化极其困难。SGDiff [342] 提出了第一个专门为场景图到图像生成而设计的扩散模型，该模型采用了一种新颖的掩蔽对比预训练。这种掩蔽预训练范式具有通用性，可以扩展到任何跨模态扩散架构，用于粗粒度和细粒度指导。其他图条件扩散模型主要用于图生成研究。Graphusion [343] 对图数据集的潜在聚类进行条件化，以生成与数据分布高度一致的新二维图。BindDM [121]、IPDiff [124] 和 IRDiff [122] 提出以三维蛋白质图作为条件，生成具有等变扩散的三维分子。

7.2 计算机视觉

7.2.1 图像超分辨率、修复、恢复、翻译和编辑。生成模型已用于解决各种图像恢复任务，包括超分辨率、修复和翻译 [16, 58, 74, 128, 171, 225, 244, 379]。图像超分辨率旨在从低分辨率输入中恢复高分辨率图像，而图像修复则围绕重建图像中缺失或损坏的区域。

有几种方法利用扩散模型来完成这些任务。例如，通过重复细化实现的超分辨率 (SR3) [259] 使用 DDPM 来实现条件图像生成。SR3 通过随机迭代去噪过程进行超分辨率。级联扩散模型 (CDM) [112] 由多个连续的扩散模型组成，每个模型都会生成分辨率不断提高的图像。SR3 和 CDM 都将扩散过程直接应用于输入图像，这会导致更大的评估步骤。为了能够在有限的计算资源下训练扩散模型，一些方法 [255, 299] 使用



图 5.LDM 产生的图像超分辨率结果[255]。

预训练的自动编码器。潜在扩散模型 (LDM) [255] 简化了去噪扩散模型的训练和采样过程，同时不牺牲质量。

对于修复任务，RePaint [189] 采用了增强的去噪策略，使用重采样迭代来更好地调节图像。ConPreDiff [345] 提出了一种基于上下文预测的通用扩散模型，以持续改进无条件/条件图像生成和图像修复（见图 6）。同时，Palette [257] 采用条件扩散模型为四个图像生成任务创建了一个统一的框架：着色、修复、反裁剪和 JPEG 恢复。图像转换侧重于合成具有特定期望的图像



图 6.ConPreDiff 产生的图像修复结果[345]。

样式 [128]。SDEdit [206] 使用随机微分方程 (SDE) 来提高保真度。具体来说，它以 Manuscript submitted to ACM 开头

通过向输入图像添加噪声，然后通过 SDE 对图像进行去噪。去噪扩散恢复模型 (DDRM) [143] 利用预先训练的去噪扩散生成模型来解决线性逆问题，并在多个图像数据集上展示了 DDRM 在不同测量噪声量下实现超分辨率、去模糊、修复和着色的多功能性。有关更多文本到图像的扩散模型，请参阅第 7.4.1 节。

7.2.2 语义分割。语义分割旨在根据既定的对象类别标记每个图像像素。生成式预训练可以提高语义分割模型的标签利用率，最近的研究表明，通过 DDPM 学习到的表示包含对分割任务有用的高级语义信息 [15, 91]。利用这些学习到的表示的少样本方法已经胜过 VDVAE [44] 和 ALAE [232] 等替代方案。类似地，解码器去噪预训练 (DDeP) [25] 将扩散模型与去噪自动编码器 [305] 相结合，并在标签高效的语义分割方面取得了有希望的结果。ODISE [329] 探索了开放词汇分割任务的扩散模型，并提出了一种新的隐式字幕器来为图像生成字幕，以更好地利用预先训练的大规模文本到图像扩散模型。

7.2.3 视频生成。由于视频帧的复杂性和时空连续性，在深度学习时代生成高质量视频仍然是一项挑战 [341, 362]。最近的研究转向扩散模型来提高生成视频的质量 [114]。例如，灵活扩散模型 (FDM) [107] 使用生成模型来允许对任意子集的视频帧进行采样（给定任何其他子集）。FDM 还包括为此目的设计的专门架构。此外，残差视频扩散 (RVD) 模型 [355] 采用自回归、端到端优化的视频扩散模型。它通过修改确定性的下一帧预测来生成未来帧，使用通过逆扩散过程产生的随机残差。有关更多文本到视频的扩散模型，请参阅第 7.4.5 节。

7.2.4 从扩散模型生成数据。从生成模型合成数据集可以有效地推进各种任务，如分类[12, 309, 385]。最近的研究已经开始利用扩散模型来实现视觉任务的这一目标。例如，Trabucco 等人。[297] 采用扩散模型对少量图像分类进行有效的数据增强。DistDiff [385] 提出了一种无需训练的数据扩展框架，该框架具有分布感知扩散模型。它构建分层原型来近似真实数据分布，并通过分层能量指导优化生成过程中的潜在数据点。InstructPix2Pix [26] 利用两个大型预训练模型（即 GPT-3 和稳定扩散）生成大量输入目标指令三元组示例数据集，并在数据集上训练指令跟踪图像编辑模型。为了使图像编辑能够反映来自现实世界和虚拟媒体的具有挑战性的世界知识和动态，EditWorld [349] 引入了一项名为世界指导图像编辑的新任务，如图 7 中所示的数据示例。EditWorld 提出了一个创新的组合框架，该框架包含一组预先训练的 LLM 和扩散模型（如图 8 所示），用于合成一个世界指导的训练数据集，用于遵循指令的图像编辑。

7.2.5 点云补全与生成。点云是捕捉现实世界物体的重要 3D 表示形式。然而，由于部分观察或自我遮挡，扫描通常会生成不完整的点云。最近的研究应用了扩散模型来解决这一挑战，利用它们推断缺失部分以重建完整的形状。这项工作对许多下游任务（如 3D 重建、增强现实和场景理解）具有重要意义 [194, 198, 367]。

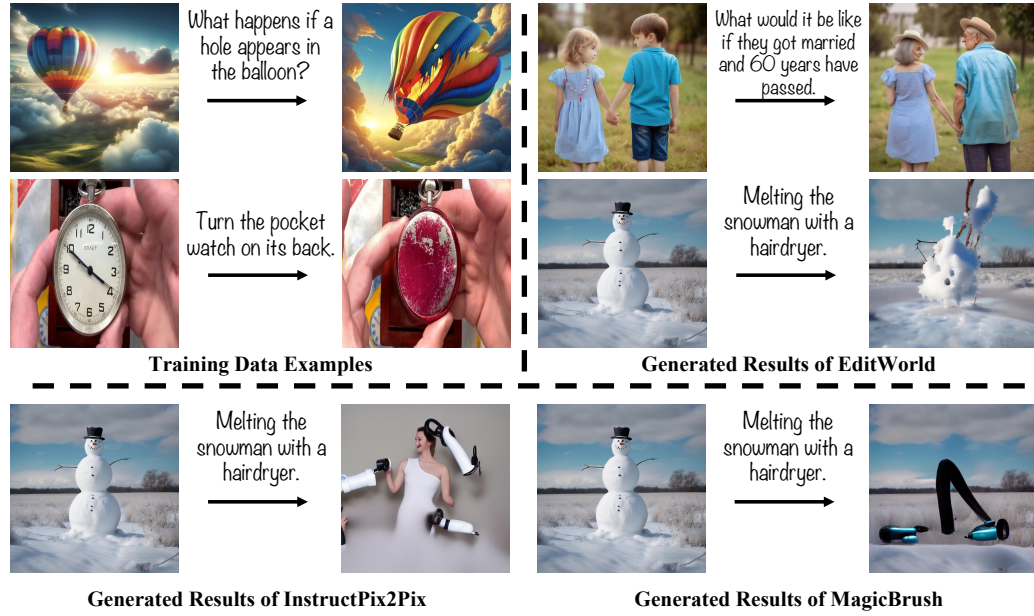


图 7.将 EditWorld [349] 与 InstructPix2Pix 和 MagicBrush 进行比较。

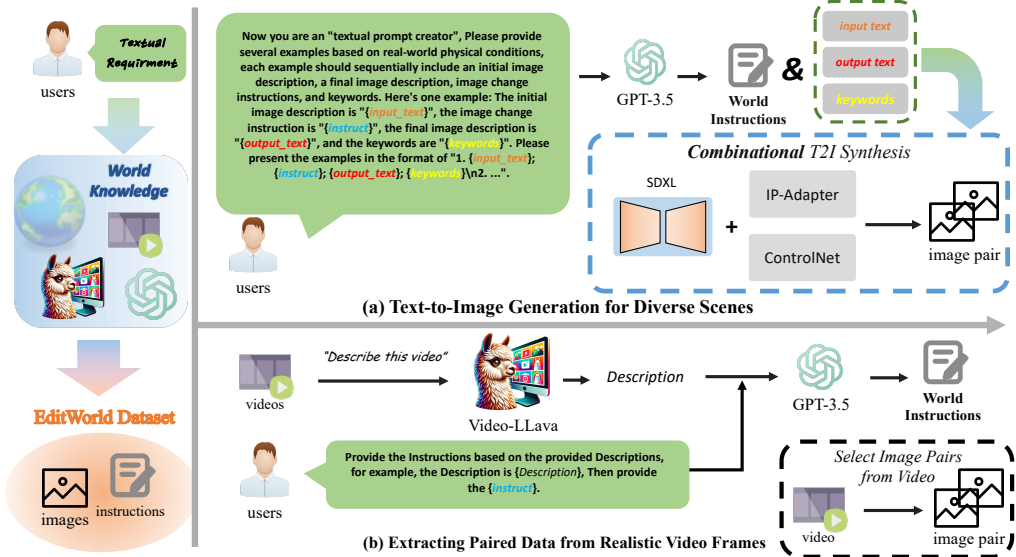


图 8.EditWorld [349] 从两个不同的分支生成世界指示图像编辑的训练数据集。

Luo 等人在 2021 年 [193] 中提出了将点云视为热力学系统中的粒子的方法, 使用热浴促进从原始分布到噪声分布的扩散。同时, 点体素扩散 (PVD) 模型 [384] 将去噪扩散模型与 3D 形状的点体素表示结合起来。

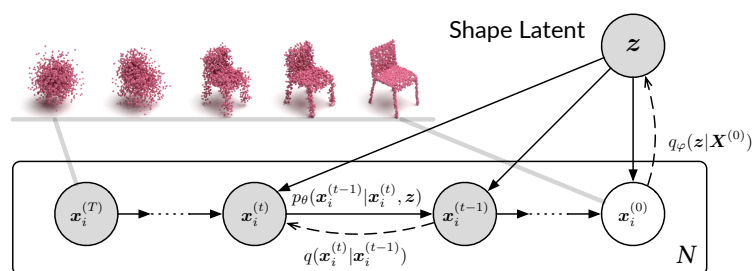


图9.点云扩散过程的有向图模型[193]。

点扩散细化 (PDR) 模型 [198] 使用条件 DDPM 从部分观测中生成粗略完成；它还在生成的点云和地面实况之间建立逐点映射。

7.2.6 异常检测。异常检测是机器学习 [266, 380] 和计算机视觉 [337] 中的一个关键且具有挑战性的问题。生成模型已被证明拥有强大的异常检测机制 [84, 105, 323]，可以对正常或健康的参考数据进行建模。AnoDDPM [323] 利用 DDPM 破坏输入图像并重建图像的健康近似值。这些方法可能比基于对抗训练的替代方法表现更好，因为它们可以通过有效的采样和稳定的训练方案更好地对较小的数据集进行建模。DDPM-CD [84] 通过 DDPM 将大量无监督遥感图像纳入训练过程。利用预先训练的 DDPM 并应用来自扩散模型解码器的多尺度表示来检测遥感图像的变化。

7.3 自然语言生成

自然语言处理旨在理解、建模和管理来自文本或音频等不同来源的人类语言。文本生成已成为自然语言处理中最关键和最具挑战性的任务之一 [127, 172, 173]。它旨在根据输入数据（例如序列和关键字）或随机噪声，用人类语言编写可信且可读的文本。已经开发了许多基于扩散模型的文本生成方法。离散去噪扩散概率模型 (D3PM) [9] 引入了类似扩散的生成模型，用于字符级文本生成 [38]。它通过超越具有均匀转移概率的腐败过程来推广多项式扩散模型 [117]。大型自回归语言模型 (LM) 能够生成高质量文本 [27, 46, 242, 373]。为了在实际应用中可靠地部署这些 LM，通常希望文本生成过程是可控的。这意味着我们需要生成能够满足所要求（例如主题、句法结构）的文本。在不重新训练的情况下控制语言模型的行为是文本生成中一个主要且重要的问题 [54, 147]。模拟位 [42] 生成模拟位来表示离散变量，并通过自调节和不对称时间间隔进一步提高样本质量。

尽管最近的方法在控制简单句子属性（例如情绪）方面取得了重大成功 [160, 338]，但在复杂、细粒度的控制（例如句法结构）方面进展甚微。为了解决更复杂的控制问题，Diffusion-LM [175] 提出了一种基于连续扩散的新语言模型。Diffusion-LM 从一系列高斯噪声向量开始，逐步将其去噪为与单词相对应的向量。逐步去噪步骤有助于产生分层的连续潜在表示。这种分层和连续的潜在变量可以使简单的基于梯度的方法实现复杂的控制。

类似地, DiffuSeq [86] 也在潜在空间中进行扩散过程, 并提出了一种新的条件扩散模型来完成更具挑战性的文本到文本生成任务。Ssd-LM [106] 在自然词汇空间而不是学习到的潜在空间上进行扩散, 使模型能够结合分类器指导和模块化控制, 而无需对现成的分类器进行任何调整。CDCD [61] 提出使用在时间和输入空间上都是连续的扩散模型来对分类数据 (包括文本) 进行建模, 并设计了一种分数插值技术来进行优化。

7.4 多模态生成

7.4.1 文本到图像生成。视觉语言模型最近因其众多潜在应用而备受关注 [240]。文本到图像生成是从描述性文本生成相应图像的任务 [69, 146, 300]。混合扩散 [10] 利用预训练的 DDPM [60] 和 CLIP [240] 模型, 并提出了一种用于通用目的的基于区域的图像编辑解决方案, 该解决方案使用自然语言指导, 适用于真实多样的图像。另一方面, unCLIP (DALLE-2) [243] 提出了一种两阶段方法, 一个可以生成以文本标题为条件的基于 CLIP 的图像嵌入的先验模型, 以及一个可以生成以图像嵌入为条件的图像的基于扩散的解码器。最近, Imagen [258] 提出了一种文本到图像的扩散模型和一个用于性能评估的综合基准。它表明 Imagen 与 VQ-GAN+CLIP [52]、潜在扩散模型 [188] 和 DALL-E 2 [243] 等最先进的方法相比表现良好。

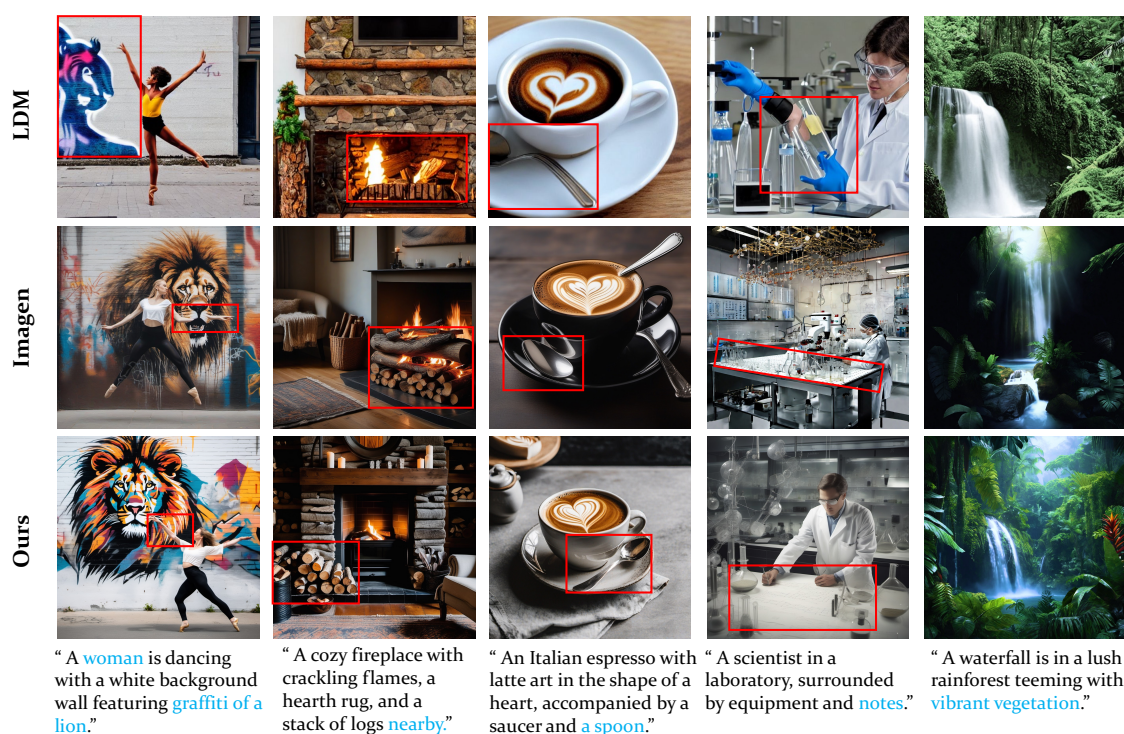


图 10. 使用 LDM、Imagen 和 ContextDiff 演示各种文本提示的文本转图像功能的合成示例 [351]。手稿已提交给 ACM

借助引导扩散模型 [60, 113] 生成照片级真实感样本的能力和文本转图像模型处理自由形式提示的能力, GLIDE [215] 将引导扩散应用于文本条件图像合成应用。VQ-Diffusion [98] 提出了一种用于文本转图像生成的矢量量化扩散模型, 它消除了单向偏差并避免了累积预测误差。Versatile Diffusion [334] 提出了第一个统一的多流多模态扩散框架, 支持图像转文本、图像变异、文本转图像和文本变异, 并可进一步扩展到其他应用, 如语义风格解缠、图像-文本双引导生成、潜在图像转文本转图像编辑等。继 Versatile Diffusion 之后, UniDiffuser [14] 提出了一种基于 Transformer 的统一扩散模型框架, 该框架可以拟合多模态数据分布, 同时处理文本到图像、图像到文本以及联合图像-文本生成任务。ConPreDiff [345] 首次将上下文预测纳入文本到图像的扩散模型, 在不增加推理成本的情况下显著提高了生成性能。ContextDiff [351] 通过将包含交互和对齐的跨模态上下文纳入正向和反向过程, 提出了通用的上下文化扩散模型。图 10 给出了这些模型的定性比较。

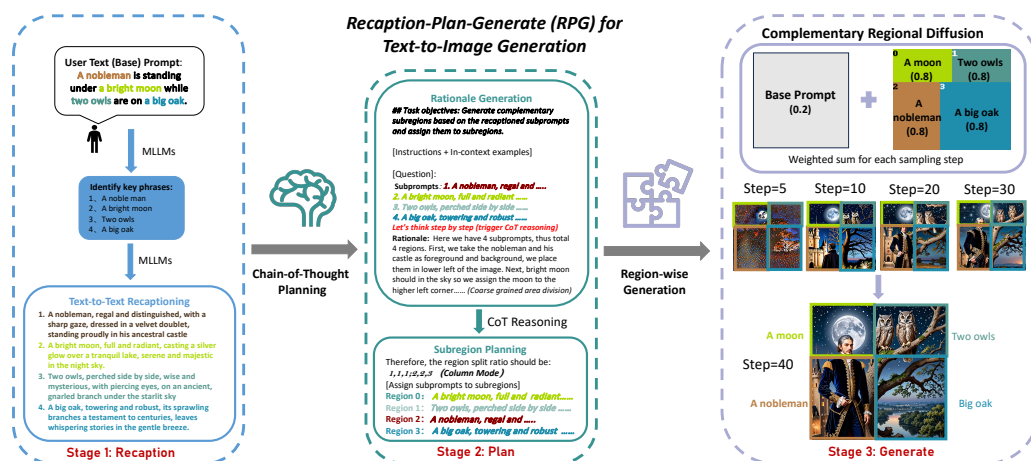


图 11. RPG [347] 文本到图像生成的框架概览。

扩散模型研究的一个有趣的新方向是利用预训练的文本到图像扩散模型对合成结果进行更复杂或更细粒度的控制。DreamBooth [256] 提出了第一种解决主题驱动生成这一新挑战性问题的技术, 它允许用户仅从几张随意拍摄的主题图像中, 重新语境化主题、修改其属性、原始艺术再现等。与那些以文本提示为条件的图像扩散模型不同, ControlNet [368] 试图控制预先训练的大型扩散模型, 以支持额外的语义图, 如边缘图、分割图、关键点、形状法线、深度等。然而, 大多数方法在处理涉及具有多种属性和关系的多个对象的复杂文本提示时往往会面临挑战。为此, RPG [347] 提出了一个全新的无需训练的文本到图像生成/编辑框架, 利用多模态 LLM [377] 强大的思路链推理能力来增强文本到图像扩散模型的组合性。这个新的 RPG 框架以闭环方式统一了文本引导的图像生成 (图 11) 和图像编辑 (图 12) 任务。值得注意的是, 如图 13 所示, RPG 的表现优于所有 SOTA 方法, 例如 SDXL [233] 和 DALL-E 3 [20], 证明了其 Manuscript submitted to ACM

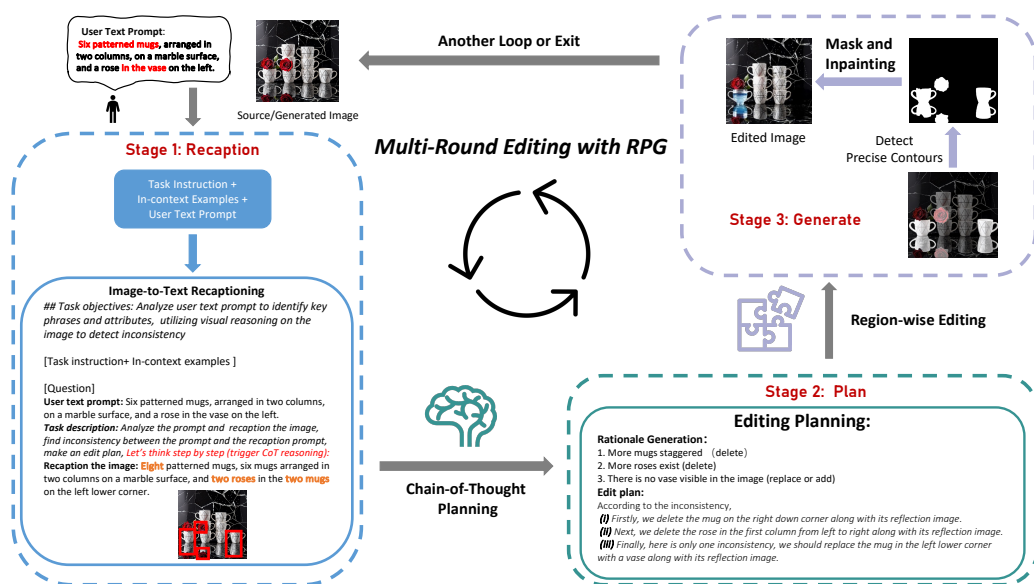


图 12.RPG[347] 可以在闭环系统中统一文本引导的图像生成和编辑

op 方法。

优越性。此外，RPG 框架易于使用，并且可以推广到不同的 MLLM 架构和扩散主干（例如 ControlNet）。

7.4.2 场景图到图像生成。尽管文本到图像生成模型在自然语言描述方面取得了令人兴奋的进展，但它们仍难以忠实地再现具有许多对象和关系的复杂句子。从场景图 (SG) 生成图像是生成模型的一项重要且具有挑战性的任务 [135]。传统方法 [109, 135, 176] 主要从 SG 预测类似图像的布局，然后根据布局生成图像。然而，这样的中间表示会丢失 SG 中的一些语义，而最近的扩散模型 [255] 也无法解决这一限制。SGDiff [342] 提出了第一个专门用于从场景图生成图像的扩散模型（图 14），并学习连续的 SG 嵌入来调节潜在在扩散模型，该模型已通过设计的掩蔽对比预训练在 SG 和图像之间进行了全局和局部语义对齐。与非扩散和扩散方法相比，SGDiff 可以生成更好地表达 SG 中密集和复杂关系的图像。然而，高质量的成对 SG 图像数据集稀缺且规模较小，如何利用大规模文本图像数据集来增强训练或提供语义扩散先验以实现更好的初始化仍然是一个悬而未决的问题。

7.4.3 文本到 3D 生成。3D 内容生成 [139, 178, 234, 331] 在游戏、娱乐和机器人模拟等各种应用中都需求旺盛。使用自然语言增强 3D 内容生成可以极大地帮助新手和经验丰富的艺术家。DreamFusion [234] 采用预先训练的 2D 文本到图像扩散模型来执行文本到 3D 的合成。它使用概率密度蒸馏损失优化了随机初始化的 3D 模型（神经辐射场，NeRF），该模型利用 2D 扩散模型作为参数图像生成器优化的先验。为了快速、高分辨率地优化 NeRF，Magic3D [178] 提出了一个基于级联低分辨率图像扩散先验和高分辨率

Manuscript submitted to ACM



Prompt: A beautiful landscape with a river in the middle, the left of the river is **in the evening** and **in the winter** with a big iceberg and a small village while some people are skiing on the river and some people are skating, the right of the river is **in the summer** with a **volcano** **in the morning** and a small village while some people are playing.

Left Prompt: A Chinese general wearing a crown, with whiskers and **golden Chinese** style armor, standing with a **majestic dragon head on his chest**, symbolizing his strength, wearing **black and gold boots**. His appearance exudes a sense of authority, wisdom, and an unyielding spirit, embodying the ideal ancient Chinese hero.

Right Prompt: This painting is a quintessential example of ancient Chinese ink art, At the top of the painting, towering mountains shrouded in **mist rise majestically**. The mountains' craggy peaks are sketched with fine, precise lines, typical of traditional Chinese ink art. **A slender swirling mists**, meandering waterfall begins its descent here, its water appearing almost ethereal amidst the soft. In the middle section, **the waterfall cascades energetically**, creating a dynamic contrast with the serene mountains above. **Lush pine trees**, rendered with graceful, flowing brush strokes, flank the waterfall. These trees appear to dance with the rhythm of the water, adding a vibrant life to the scene. At the bottom, the waterfall concludes its journey in a tranquil pool. The water's surface is calm, reflecting the surrounding nature and the sky above. Here, **delicate flowers** and **small shrubs** are **depicted along the water's edge**, symbolizing peace and harmony with nature.

图 13. 与之前的 SOTA 模型相比, RPG [347] 表现出在生成的图像中表达复杂和组合文本提示的卓越能力 (彩色文本表示关键部分)。

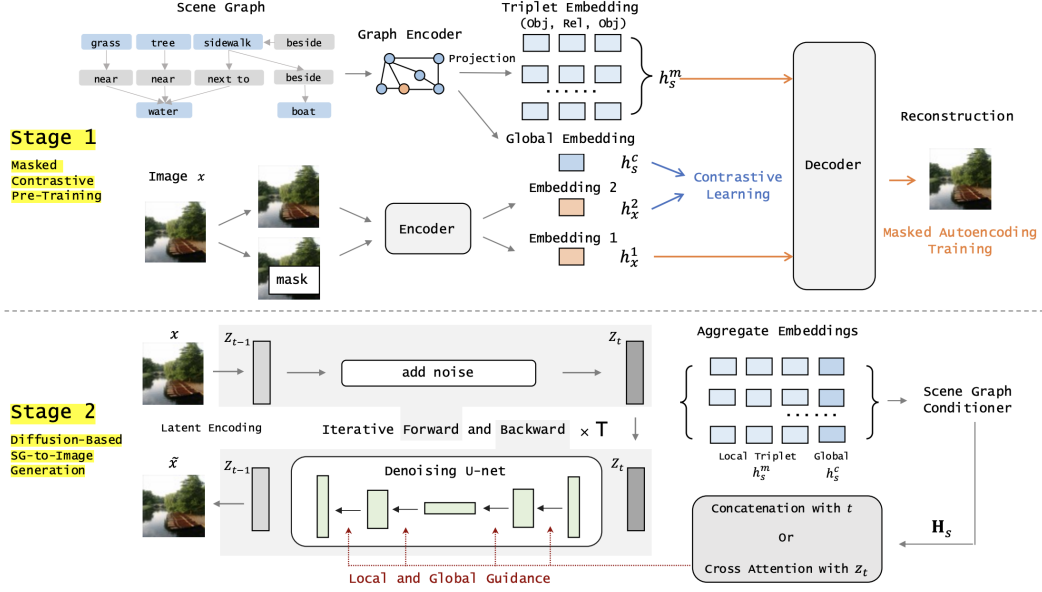
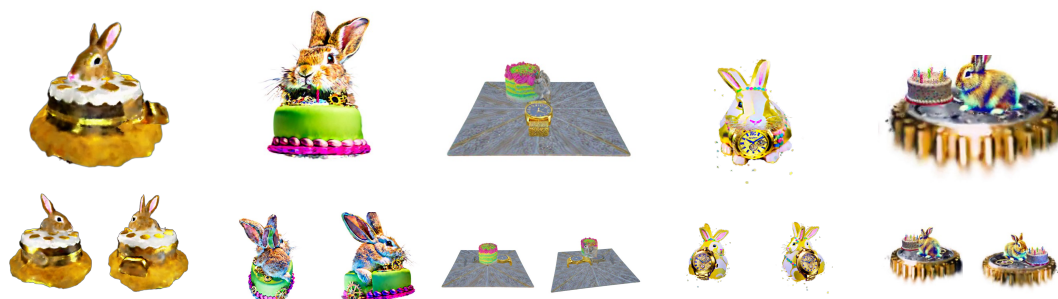


图 14.SGDiff [342] 利用掩蔽对比预训练进行基于场景图的图像扩散生成。

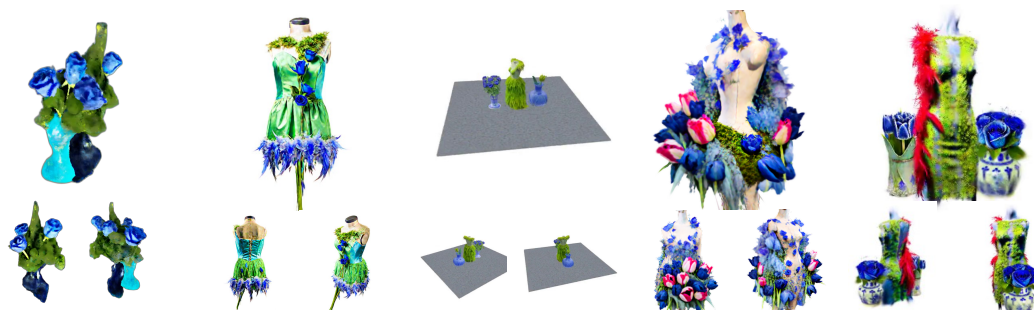
潜在扩散先验。为了实现高保真 3D 创作，Make-It-3D [290] 通过在正面视图中结合参考图的约束和在新视图中结合扩散先验来优化神经辐射场，将粗模型增强为纹理点云，并使用扩散先验和参考图中的高质量纹理提高真实感。ProlificDreamer [310] 提出了变分分数蒸馏 (VSD)，基于文本提示作为随机变量优化 3D 场景的分布，使所有视角的渲染图像分布与预训练的 2D 扩散模型紧密对齐，使用 KL 散度作为度量。IPDreamer [365] 进一步提出了一种新颖的 3D 对象合成框架，使用户能够毫不费力地创建可控的高质量 3D 对象。它擅长合成高质量的 3D 对象，可以与提供的复杂图像提示高度一致。

对组合式 3D 数据分布进行建模是生成模型的一项基本且关键的任务。当前的前馈方法 [270, 271] 主要能够生成单个物体，由于训练数据有限，在创建包含多个物体的更复杂场景时面临挑战。最近，提出了一系列可学习的布局组合方法 [43, 73, 80, 104, 303]。这些方法组合多个物体临时辐射场，然后根据外部反馈优化辐射场的位置。例如，Epstein 等 [73] 提出仅基于来自大型预训练文本到图像模型的知识来学习合理布局的分布。Vil esov 等 [303] 介绍了一种基于蒙特卡罗采样和物理约束的优化方法。然而，这些形式的布局指导相对较粗，对于细粒度控制来说表达力不够。Yang 等 [303] 提出了一种基于蒙特卡罗采样和物理约束的优化方法。然而，这些形式的布局指导相对较粗，对于细粒度控制来说表达力不够。（2024）[350] 通过将确保视图一致性和明确区分对象的语义嵌入纳入 SDS 流程（即 SemanticSDS [350]）来解决此问题，这些流程灵活且富有表现力，可用于优化 3D 场景。如图 15 所示，SemanticSDS [350] 在合成文本到 3D 生成中可以实现优于现有方法的精度和质量。

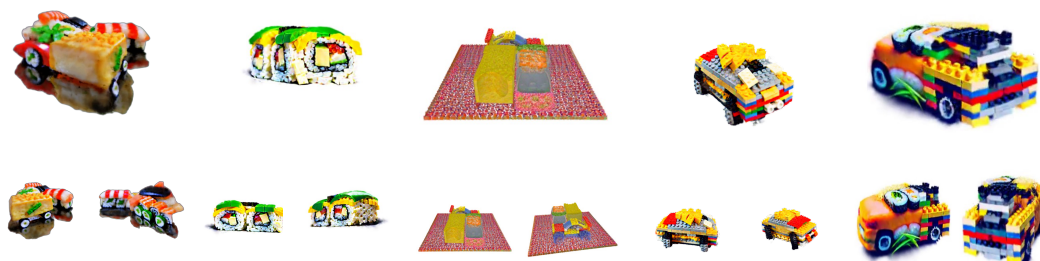
GraphDreamer LucidDreamer GALA3D GSGEN SemanticSDS



A rabbit sits atop a large, expensive watch with many shiny gears, made half of iron and half of gold, eating a birthday cake that is in front of the rabbit



A mannequin adorned with a dress made of feathers and moss stands at the center, flanked by a vase with a single blue tulip and another with blue roses.



A car with the front right side made of cheese, the front left side made of sushi, and the back made of LEGO.

图 15. SemanticSDS [350] 实现了优于最先进基线的组合文本到 3d 生成结果，特别是在生成具有不同属性的多个对象时。

7.4.4 文本到动作生成。人体动作生成是计算机动画的一项基本任务，应用范围从游戏到机器人 [369]。生成的动作通常是一系列人体姿势

由关节旋转和位置表示。运动扩散模型 (MDM) [292] 采用无分类器扩散的生成模型来生成人体运动, 该模型基于变换器, 结合了运动生成文献中的见解, 并使用运动位置和速度的几何损失对模型进行正则化。FLAME [151] 涉及一种基于变换器的扩散, 以更好地处理运动数据, 它可以管理可变长度的运动并很好地处理自由格式的文本。值得注意的是, 它可以逐帧和逐关节编辑运动的各个部分, 而无需任何微调。

7.4.5 文本到视频生成。基于扩散的文本到图像生成方面的最新进展推动了文本到视频生成的发展 [110, 273, 318]。Make-A-Video [273] 提出通过时空分解扩散模型将基于扩散的文本到图像模型扩展到文本到视频。它利用联合文本图像先验来绕过对成对文本视频数据的需求, 并进一步提出了用于高清、高帧率文本到视频生成的超分辨率策略。Imagen Video [110] 通过设计级联视频扩散模型来生成高清视频, 并将一些在文本到图像设置中效果良好的发现转移到视频生成, 包括冻结的 T5 文本编码器和无分类器指导。Tune-A-Video [318] 为文本到视频生成引入了一次性视频调整, 从而消除了使用大规模视频数据集进行训练的负担。它采用高效的注意力调节和结构反转来显著增强时间一致性。Text2Video-Zero [148] 使用预训练的文本到图像扩散模型实现零样本文本到视频合成, 通过潜在代码中的运动动态和跨帧注意力确保时间一致性。其目标是实现价格合理的文本引导视频生成和编辑, 而无需额外的微调。FateZero [237] 是第一个使用预训练的文本到图像扩散模型进行时间一致的零样本文本到视频编辑的框架。它融合了 DDIM 反转和生成过程中的注意力图, 以最大限度地保持编辑过程中运动和结构的一致性。ContextDiff [351] 将关于文本条件和视频样本之间相互作用的跨模态上下文信息合并到正向和反向过程中, 形成用于文本到视频生成的前向-后向一致的视频扩散模型。

大多数文本到视频的扩散模型都是在固定大小的视频数据集上训练的, 因此通常只能生成相对较少的帧, 这在生成较长的视频时会导致质量显著下降。一些进展[108,295,388]试图通过各种策略来克服这一限制。Vlogger[388]采用掩蔽扩散模型进行条件帧输入, 从而促进较长的视频生成, StreamingT2V[108]利用类似ControlNet的条件机制来实现自回归视频生成。最近的VideoTetris[295]引入了一种时空组合扩散方法, 用于处理具有多个对象的场景并遵循渐进的复杂提示(即组合文本到视频生成)。此外, VideoTetris还开发了一种新的视频数据预处理方法和一种称为参考帧注意的一致性正则化方法, 通过增强的运动动态和提示语义来改进自回归长视频生成。图 16 中的定性比较表明, VideoTetris 不仅可以生成优质的构图视频, 还可以制作符合构图提示同时保持最佳一致性的高质量长视频。

7.4.6 文本转音频生成。文本转音频生成是将普通语言文本转换为语音输出的任务 [170, 320]。Grad-TTS [235] 提出了一种新的文本转语音模型, 该模型具有基于分数的解码器和扩散模型。它逐渐转换编码器预测的噪声, 并通过单调对齐搜索的方法与文本输入进一步对齐 [239]。Grad-TTS2 [152] 以自适应方式改进了 Grad-TTS。Diffsound [336] 提出了一种基于离散扩散模型的非自回归解码器 [9, 275], 它可以预测所有梅尔频谱图

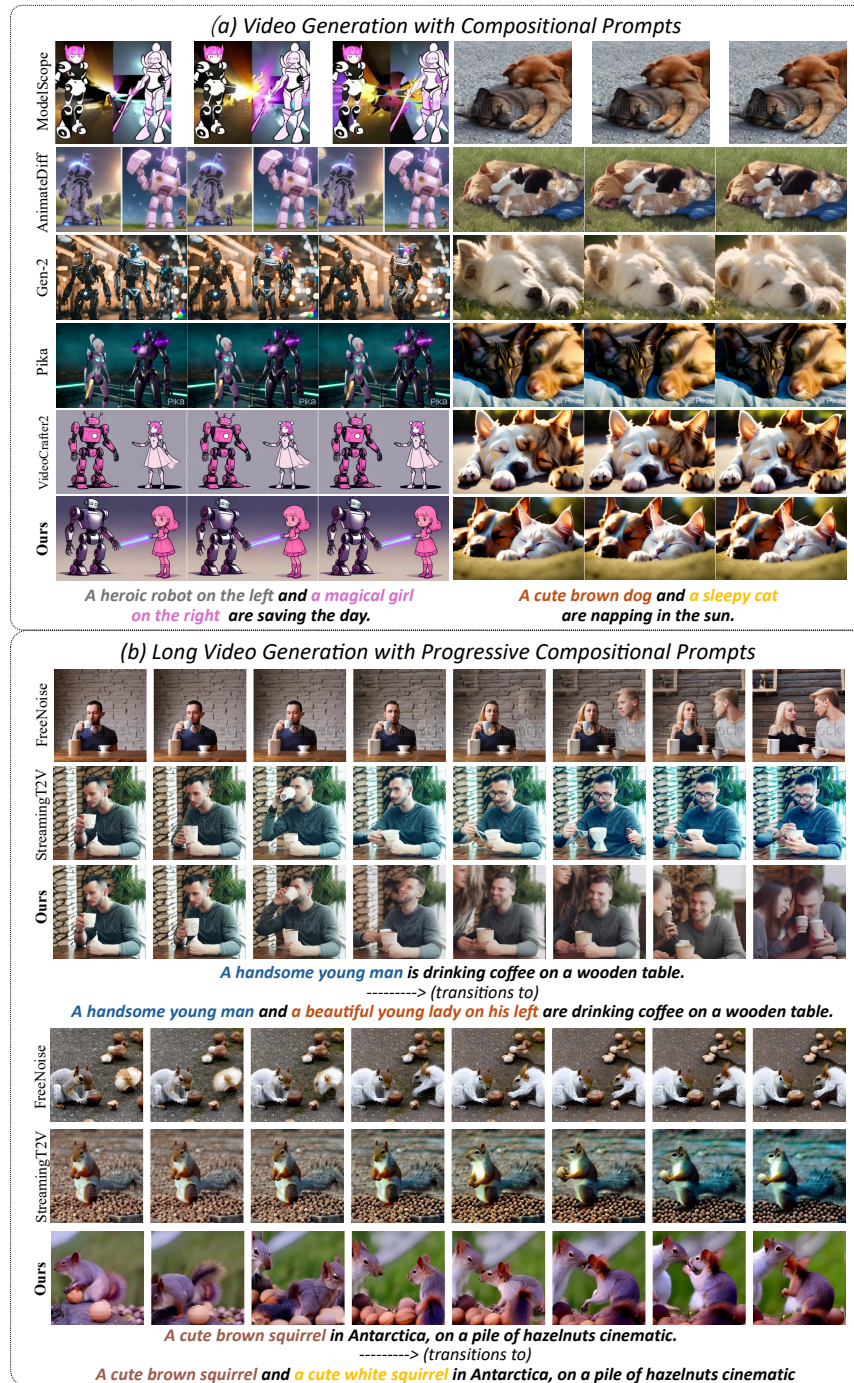


图 16. 将 VideoTetris [295] 与开源或商业 T2V 模型在短视频和长视频生成中进行比较。

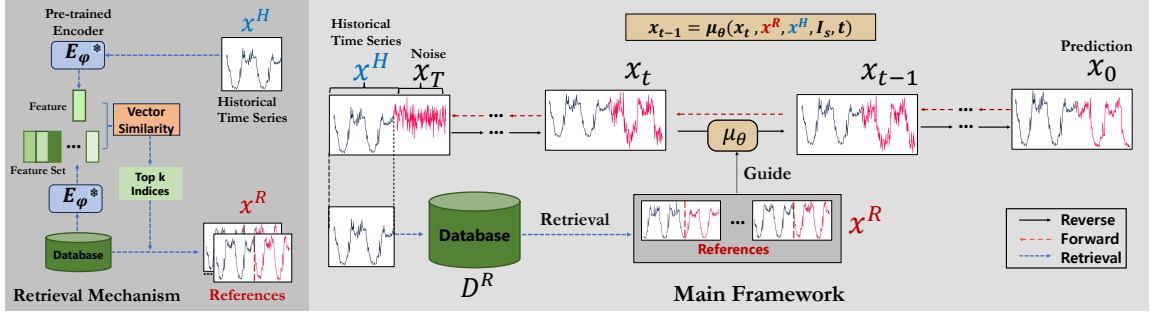


图 17. 用于时间序列预测的检索增强扩散模型 (RATD [180]) 概述。历史时间序列 x^H 输入检索模块，以获得相应的参考 x^R 。之后， x^H 与噪声连接，作为模型 μ_θ 的主要输入。 x^R 将用作去噪过程的指导。

在每个步骤中预测 token，然后在后续步骤中细化预测的 token。EdiTTS [288] 利用基于分数的文本转语音模型来细化经过粗略修改的梅尔频谱图先验。ProDiff [120] 不是估计数据密度梯度，而是通过直接预测干净数据来参数化去噪扩散模型。

7.5 时间数据建模

7.5.1 时间序列插补。时间序列数据广泛应用于许多重要的实际应用 [72, 223, 341, 374]。然而，时间序列通常会因多种原因而包含缺失值，这些原因可能是机械或人为错误造成的 [272, 289, 356]。近年来，插补方法得到了广泛的应用，包括确定性插补 [32, 37, 197] 和概率插补 [78]，包括基于扩散的方法。基于条件分数的插补扩散模型 (CSDI) [291] 提出了一种利用基于分数的扩散模型的新型时间序列插补方法。具体而言，为了利用时间数据中的相关性，它采用自监督训练的形式来优化扩散模型。它在一些现实世界数据集中的应用表明了它优于以前的方法。受控随机微分方程 (CSDE) [228] 提出了一种新的概率框架，用于使用神经控制随机微分方程对随机动力学进行建模。结构化状态空间扩散 (SSSD) [3] 集成了条件扩散模型和结构化状态空间模型 [97]，特别用于捕捉时间序列中的长期依赖关系。它在时间序列插补和预测任务中都表现良好。

7.5.2 时间序列预测。时间序列预测是预测或预报一段时间内的未来值的任务。最近，神经方法被广泛用于解决预测问题，采用单变量点预测方法 [222] 或单变量概率方法 [262]。在多变量设置中，我们还有点预测方法 [174] 以及概率方法，它们使用高斯 copula [263]、GAN [358] 或正则化流 [249] 明确地对数据分布进行建模。TimeGrad [248] 提出了一种用于预测多元概率时间序列的自回归模型，该模型通过估计其梯度在每个时间步骤中从数据分布中抽样。它利用了扩散概率模型，该模型与分数匹配和基于能量的方法密切相关。具体来说，它通过优化数据似然的变分界限来学习梯度，并在推理过程中使用朗之万采样 [280] 通过马尔可夫链将白噪声转换为感兴趣的分布样本。为了处理复杂的时间序列预测，如图 17 所示，刘等人 (2024) 针对

首次引入检索增强时间序列扩散 (RATD) [180], 可以更充分地利用数据集并在去噪过程中提供有意义的指导。

7.5.3 波形信号处理。在电子学、声学和一些相关领域中, 信号的波形用其图形形状表示, 该图形是时间函数, 与时间和幅度尺度无关。WaveGrad [39] 引入了一种用于波形生成的条件模型, 该模型估计数据密度的梯度。它接收高斯白噪声信号作为输入, 并使用基于梯度的采样器迭代细化信号。WaveGrad 通过调整细化步骤的数量, 自然地以推理速度换取样本质量, 并在音频质量方面建立了非自回归和自回归模型之间的联系。DiffWave [159] 提出了一种通用且有效的扩散概率模型, 用于条件或无条件波形生成。该模型是非自回归的, 通过优化数据似然的变分界限的变体进行有效训练。此外, 它在不同的波形生成任务中产生高保真音频, 例如类条件生成和无条件生成。

7.6 稳健学习

鲁棒学习是一类防御方法, 有助于学习对对抗性扰动或噪声具有鲁棒性的网络[23、216、232、308、319、357]。对抗性训练[200]被视为图像分类器防御对抗性攻击的标准方法, 而对抗性净化作为一种替代防御方法已表现出显著的性能[357], 它使用独立的净化模型将受攻击的图像净化为干净的图像。给定一个对抗性样本, DiffPure[216]在前向扩散过程后用少量噪声对其进行扩散, 然后通过反向生成过程恢复干净的图像。自适应去噪净化 (ADP) [357]表明, 用去噪分数匹配[304]训练的EBM可以在短短几步内有效净化受攻击的图像。它进一步提出了一种有效的随机净化方案, 在净化之前向图像中注入随机噪声。投影梯度下降 (PGD) [23] 提出了一种基于随机扩散的新型预处理鲁棒性方法, 旨在实现与模型无关的对抗防御, 并产生高质量的去噪结果。此外, 一些研究建议应用引导扩散过程进行高级对抗净化 [308, 319]。

7.7 跨学科应用

7.7.1 药物设计和生命科学。图神经网络 [102, 322, 344, 383] 和相应的表示学习 [103] 技术在许多领域取得了巨大成功 [21, 293, 321, 332, 340, 387], 包括对分子 / 蛋白质进行建模, 从属性预测 [71, 85] 到分子 / 蛋白质生成 [131, 138, 195, 268], 其中分子自然地由节点边图表示。一方面, 最近的研究提出针对具有生物学或物理洞察力的分子 / 蛋白质 [182, 364] 预训练 GNN/transformer [192, 382], 并取得了显著的成果。另一方面, 越来越多的研究开始利用基于图的扩散模型来增强分子或蛋白质的生成。扭转扩散 [133] 提出了一种新的扩散框架, 该框架在扭转角空间上进行操作, 扩散过程在超空间上, 并采用外在到内在的评分模型。GeoDiff [333] 证明使用等变马尔可夫核演化的马尔可夫链可以产生不变分布, 并进一步设计马尔可夫核的模块以保留所需的等方差性质。还有其他研究将等方差性质融入三维分子生成 [115] 和蛋白质生成 [5, 19]。受模拟分子动力学的经典力场方法的启发, ConfGF [267] 直接估计分子构象生成中原子坐标对数密度的梯度场。

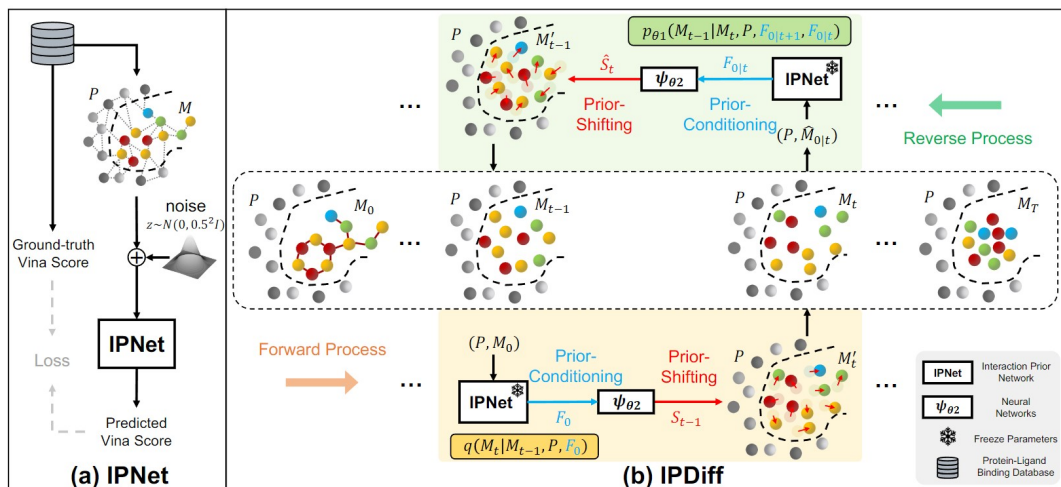


图 18. IPDiff [124] 将蛋白质-配体相互作用纳入分子扩散模型的正向和逆向过程。

最近，给定一个靶蛋白，扩散模型开始推动能够与靶标紧密结合的 3D 小药物分子的设计 [99, 122, 124]。IP Diff [124] 提出了一种用于基于结构的药物设计 (SBDD) 的新型 3D 分子扩散模型。如图 18 所示，在提出的先验条件和先验转移机制下，口袋-配体相互作用在正向和反向过程中都得到了明确考虑。值得注意的是，IPDiff 在结合相关指标和分子特性方面优于所有以前的基于扩散和自回归生成模型。BindDM [121] 为 SBDD 任务提出了一种分层的复合子复合扩散模型，该模型结合了 3D 分子扩散生成所必需的结合自适应子复合体。IRDiff [124] 为 SBDD 任务提出了一种基于相互作用的检索增强 3D 分子扩散模型，名为 IRDIFF。如图 19 所示，该模型使用信息丰富的外部目标感知参考来指导 3D 分子生成，设计了两种新颖的增强机制，即检索增强和自我增强，以结合必需的蛋白质分子结合结构来进行目标感知分子生成。

还有一些研究使用扩散模型进行蛋白质生成，例如 DiffAb。DiffAb [196] 首次提出了一种基于扩散的 3D 抗体设计框架，该框架对决定抗体互补性的互补决定区 (CDR) 的序列和结构进行建模。实验表明，DiffAb 可用于各种抗体设计任务，例如联合生成序列结构、设计具有固定框架的 CDR 以及优化抗体。SMCDiff [298] 提出首先通过 E(3) 等变图神经网络学习多样化和更长的蛋白质骨架结构的分布，然后在给定基序的情况下从该分布中有效地采样支架。生成结果表明设计的骨架与 AlphaFold2 预测的结构很好地对齐。

7.7.2 材料设计。固态材料是众多关键技术的关键基础 [28]。晶体扩散变分自编码器 (CDVAE) [327] 通过提出一个噪声条件得分网络，将稳定性作为归纳偏差，该网络同时利用了置换、平移、旋转和周期不变性

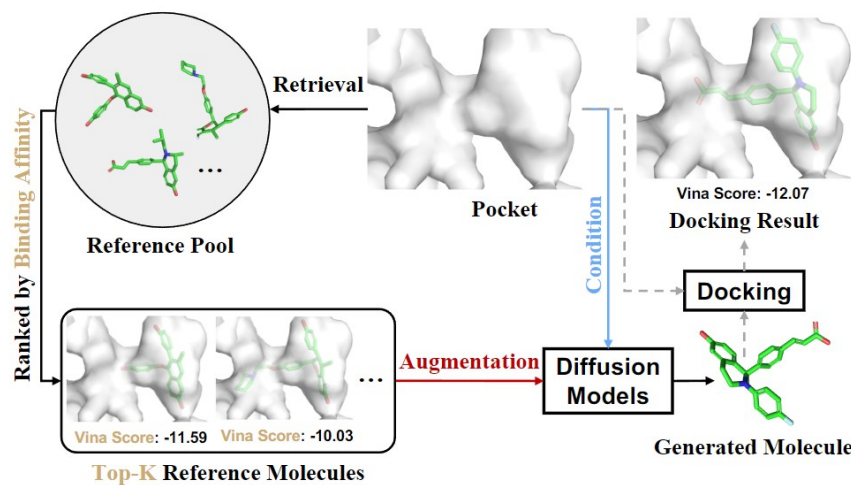


图 19. IRDiff[122] 设计了一个基于交互的检索增强生成框架

为 SBDD 工作。

属性。Luo 等人 (2022 年) [196] 利用等变扩散对互补决定区域的序列和结构进行建模, 并明确针对特定抗原结构, 以原子分辨率生成抗体。

7.7.3 医学图像重建。逆问题是从观察到的测量值中恢复未知信号, 它是计算机断层扫描 (CT) 和磁共振成像 (MRI) 医学图像重建中的一个重要问题 [47, 48, 230, 284, 328]。Song 等人 (2021) [284] 利用基于分数的生成模型重建与先验和观察到的测量值一致的图像。Chung 等人 (2022) [49] 训练具有去噪分数匹配的连续时间相关分数函数, 并在评估阶段在数值 SDE 求解器和数据一致性步骤之间进行迭代以进行重建。Peng 等人 (2022) [230] 通过给定观察到的 k 空间信号逐渐引导反向扩散过程来执行 MR 重建, 并提出了一种由粗到细的采样算法以实现高效采样。

8 未来方向

传播模型研究尚处于早期阶段, 在理论和实证方面都有很大改进空间。如前文所述, 关键研究方向包括高效采样和改进可能性, 以及探索传播模型如何处理特殊数据结构、与其他类型的生成模型交互以及如何针对各种应用进行定制。此外, 我们预计未来传播模型研究可能会扩展到以下途径。

重新审视假设。需要重新审视和分析扩散模型中的许多典型假设。例如, 扩散模型的前向过程完全消除数据中的任何信息并使其等同于先验分布的假设可能并不总是成立。实际上, 在有限的时间内完全消除信息是不可能实现的。了解何时停止前向噪声过程以在采样效率和样本质量之间取得平衡是非常有趣的[79]。薛定谔桥的最新进展

和最优传输[41, 55, 57, 269, 278]提供了有希望的替代解决方案, 为能够在有限时间内收敛到指定先验分布的扩散模型提出了新的公式。

理论理解。扩散模型已成为一个强大的框架, 尤其是它是唯一一个可以在大多数应用中与生成对抗网络 (GAN) 相媲美而无需诉诸对抗训练的框架。利用这一潜力的关键是了解为什么以及何时扩散模型比其他特定任务更有效。重要的是确定哪些基本特征将扩散模型与其他类型的生成模型区分开来, 例如变分自动编码器、基于能量的模型或自回归模型。理解这些区别将有助于阐明为什么扩散模型能够生成质量极佳的样本, 同时实现最高似然。同样重要的是需要制定理论指导, 以系统地选择和确定扩散模型的各种超参数。

潜在表示。与变分自动编码器或生成对抗网络不同, 扩散模型在潜在空间中提供良好数据表示的效果较差。因此, 它们不能轻易用于基于语义表示处理数据等任务。此外, 由于扩散模型中的潜在空间通常具有与数据空间相同的维数, 因此采样效率会受到负面影响, 并且模型可能无法很好地学习表示方案 [132]。

AIGC 和扩散基础模型。从稳定扩散到 ChatGPT, 人工智能生成内容 (AIGC) 引起了学术界和工业界的广泛关注。生成式预训练是 GPT-1/2/3/4 [220, 224, 241, 242] 和 (Visual) ChatGPT [316] 的核心技术, 它配备了大型语言模型 (LLM) [296] 和可视化基础模型 [24, 360, 363], 表现出良好的生成性能和令人惊讶的涌现能力 [314]。将生成式预训练 (仅解码器) 从 GPT 系列转移到扩散模型类、评估基于扩散的大规模生成性能以及分析扩散基础模型的涌现能力是很有趣的。此外, 将 LLM 与扩散模型相结合已被证明是一个新的有前途的方向 [347, 349]。

9 结论

我们从各个角度全面介绍了扩散模型。我们首先介绍了三种基本公式: DDPM、SGM 和 Score SDE。然后, 我们讨论了最近改进扩散模型的努力, 重点介绍了三个主要方向: 采样效率、似然最大化和针对特殊结构数据的新技术。我们还探讨了扩散模型与其他生成模型之间的联系, 并概述了将两者结合起来的潜在好处。对六个领域的应用的调查表明了扩散模型的广泛潜力。最后, 我们概述了未来研究的可能途径。

参考

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat 等人。2023 年。Gpt-4 技术报告。arXiv 预印本 arXiv:2303.08774 (2023)。[2] Michael Samuel Albergo 和 Eric Vanden-Eijnden。2022 年。使用随机插值构建正则化流。在第十一届国际学习表征会议上。[3] Juan Miguel Lopez Alcaraz 和 Nils Strodthoff。2022 年。基于扩散的时间序列插值和预测与结构化状态空间模型。arXiv 预印本 arXiv:2208.09399 (2022)。[4] Tomer Amit, Eliya Nachmani, Tal Shaharabany 和 Lior Wolf。2021 年。Segdiff: 使用扩散概率模型进行图像分割。arXiv 预印本 arXiv:2112.00390 (2021 年)。

- [5] Namrata Anand and Tudor Achim. 2022 年。使用等变去噪扩散概率模型生成蛋白质结构和序列。arXiv 预印本 arXiv:2205.15019 (2022)。[6] Brian DO Anderson. 1982 年。逆时间扩散方程模型。随机过程及其应用 12, 3 (1982), 313 – 326。[7] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen 等人。2023 年。Pa lm 2 技术报告。arXiv 预印本 arXiv:2305.10403 (2023)。[8] Uri M Ascher 和 Linda R Petzold. 1998. 常微分方程和微分代数方程的计算机方法。第 61 卷。暹罗。[9] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow 和 Rianne van den Berg. 2021. 离散状态空间中的结构化去噪扩散模型。神经信息处理系统进展。[10] Omri Avrahami, Dani Lischinski 和 Ohad Fried. 2022. 用于文本驱动的自然图像编辑的混合扩散。在 IEEE 计算机视觉和模式识别会议上。18208 – 18218。[11] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan 等人。2022. 使用从人类反馈中进行强化学习来训练一个有用且无害的助手。arXiv preprint arXiv:2204.05862 (2022)。[12] Hritik Bansal 和 Aditya Grover. 2023 年。将现实留给想象：通过生成的数据集进行鲁棒分类。在国际学习表征会议上。[13] 鲍帆、李崇轩、朱俊和张波。2021 年。Analytic-DPM：扩散概率模型中最优逆方差的解析估计。在国际学习表征会议上。[14] 鲍帆、聂申、薛凯文、李崇轩、蒲诗、王耀乐、岳刚、曹越、苏航和朱俊。2023 年。一个 Transformer 可适应大规模多模态扩散中的所有分布。arXiv preprint arXiv:2303.06555 (2023)。[15] Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov 和 Artem Babenko. 2021 年。使用扩散模型进行标签高效语义分割。在国际学习表征会议上。[16] Georgios Batzolis, Jan Stanczuk, Carola-Bibiane Schönlieb 和 Christian Etmann. 2021 年。使用基于分数的扩散模型进行条件图像生成。arXiv 预印本 arXiv:2111.13606 (2021)。[17] S amy Bengio 和 Yoshua Bengio. 2000 年。使用神经网络解决联合分布中的维数灾难。IEEE 神经网络学习系统汇刊 (2000)。[18] Yoshua Bengio, Réjean Ducharme, Pascal Vincent 和 Christian Janvin. 2003. 一种神经概率语言模型。机器学习研究杂志 3 (2003), 1137 – 1155。[19] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov 和 Philip E Bourne. 2000. 蛋白质数据库。核酸研究 28, 1 (2000), 235 – 242。[20] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo 等。2023. 通过更好的字幕改进图像生成。计算机科学。https://cdn.openai.com/papers/dall-e-3.pdf (2023)。[21] Piotr Bielak, Tomasz Kajdanowicz 和 Nitesh V Chawla. 2021 年。Graph Barlow Twins：一种用于图的自监督表示学习框架。arXiv 预印本 arXiv:2106.02466 (2021)。[22] Miko ĳ aj Bińkowski, Dougal J. Sutherland, Michael Arbel 和 Arthur Gretton. 2018 年。揭开 MMD GAN 的神秘面纱。在国际学习表示会议上。[23] Tsachi Blau, Roy Ganz, Bahjat Kavar, Alex Bronstein 和 Michael Elad. 2022 年。使用扩散模型的威胁模型不可知对抗性防御。arXiv 预印本 arXiv:2207.08089 (2022)。[24] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill 等人。2021 年。关于基础模型的机遇和风险。arXiv 预印本 arXiv:2108.07258 (2021)。[25] Emmanuel Asiedu Brempong, Simon Kornblith, Ting Chen, Niki Parmar, Matthias Minderer 和 Mohammad Norouzi. 2022 年。用于语义分割的去噪预训练。在 IEEE 计算机视觉与模式识别会议上。4175 – 4186。[26] Tim Brooks, Aleksander Holynski 和 Alexei A Efros. 2023 年。Instructpix2pix：学习遵循图像编辑指令。在 IEEE 计算机视觉和模式识别会议上。18392 – 18402。[27] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell 等人。2020 年。语言模型是小样本学习者。在神经信息处理系统进展中。[28] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev 和 Aron Walsh. 2018 年。分子和材料科学的机器学习。自然 559, 7715 (2018), 547 – 555。[29] Ruojin Cai, Guandao Ya

[34] Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu 和 William T Freeman. 2022 年。Maskgit：带掩码的生成图像转换器。在 IEEE 计算机视觉与模式识别会议上。11315 – 11325. [35] ChatGPT 简介。2022 年。ChatGPT 简介。[36] Tong Che, Ruixiang Zhang, Jascha Sohl-Dickstein, Hugo Larochelle, Liam Paull, Yuan Cao 和 Yoshua Bengio. 2020 年。你的 GAN 实际上是基于能量的模型，你应该使用鉴别器驱动的潜在采样。arXiv 预印本 arXiv:2003.06060 (2020). [37] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag 和 Yan Liu. 2018 年。具有缺失值的多元时间序列的循环神经网络。科学报告 8, 1 (2018), 1 – 12. [38] Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Phillipp Koehn 和 Tony Robinson. 2013 年。用于衡量统计语言建模进展的 10 亿字基准。arXiv 预印本 arXiv:1312.3005 (2013). [39] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi 和 William Chan. 2020 年。WaveGrad：估计用于波形生成的梯度。arXiv 预印本 arXiv:2009.00713 (2020). [40] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt 和 David Duvenaud. 2018 年。神经常微分方程。arXiv 预印本 arXiv:1806.07366 (2018). [41] Tianrong Chen, Guan-Hong Liu 和 Evangelos Theodorou. 2021. 使用前向-后向 SDE 理论进行薛定谔桥的似然训练。在国际学习表征会议上。[42] Ting Chen, Ruixiang Zhang 和 Geoffrey Hinton. 2022. 模拟位：使用具有自调节的扩散模型生成离散数据。arXiv 预印本 arXiv:2208.04202 (2022). [43] Yongwei Chen, Tengfei Wang, Tong Wu, Xingang Pan, Kui Jia 和 Ziwei Liu. 2024. Comboverse：使用空间感知扩散指导创建组合 3d 资产。arXiv 预印本 arXiv:2403.12409 (2024). [44] Rewon Child. 2020 年。非常深的 VAE 可以泛化自回归模型，并且在图像上的表现优于它们。在国际学习表征会议上。[45] Rewon Child, Scott Gray, Alec Radford 和 Ilya Sutskever. 2019 年。使用稀疏变换器生成序列。CoRR abs/1904.10509 (2019 年)。arXiv:1904.10509 <http://arxiv.org/abs/1904.10509> [46] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann 等人。2022 年。PaLM：使用路径扩展语言建模。(2022 年)。[47] Hyungjin Chung, Eun Sun Lee 和 Jong Chul Ye. 2022 年。使用正则化逆扩散对 MRI 图像进行去噪和超分辨率。arXiv 预印本 arXiv:2203.12621 (2022 年)。[48] Hyungjin Chung, Byeongsu Sim 和 Jong Chul Ye. 2022 年。更接近扩散更快：通过随机收缩加速逆问题的条件扩散模型。在 IEEE 计算机视觉和模式识别会议上。12413 – 12422. [49] Hyungjin Chung 和 Jong Chul Ye. 2022 年。基于分数的加速 MRI 扩散模型。医学图像分析 (2022), 102479. [50] Rob Cornish, Anthony Caterini, George Deligiannidis 和 Arnaud Doucet. 2020. 使用连续索引正则化流放宽双射约束。在国际机器学习会议上。2133 – 2143. [51] Antonia Criswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta 和 Anil A Bharath. 2018. 生成对抗网络：概述。IEEE 信号处理杂志 35, 1 (2018), 53 – 65. [52] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato 和 Edward Raff. 2022. Vqgan-clip：使用自然语言指导的开放域图像生成和编辑。arXiv 预印本 arXiv:2204.08583 (2022). [53] Salman UH Dar, Şaban Öztürk, Yilmaz Korkmaz, Gokberk Elmas, Muzaffer Özbey, Alper Güngör 和 Tolga Çukur. 2022 年。用于加速 MRI 重建的自适应扩散先验。arXiv 预印本 arXiv:2207.05876 (2022). [54] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski 和 Rosanne Liu. 2019 年。即插即用语言模型：一种简单的受控文本生成方法。在国际学习表征会议上。[55] Valentin De Bortoli, Arnaud Doucet, Jeremy Heng 和 James Thornton. 2021 年。使用分数匹配模拟扩散桥。arXiv 预印本 arXiv:2111.07243 (2021). [56] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh 和 Arnaud Doucet. 2022 年。基于黎曼分数的生成建模。arXiv 预印本 arXiv:2202.02763 (2022). [57] Valentin De Bortoli, James Thornton, Jeremy Heng 和 Arnaud Doucet. 2021 年。扩散薛定谔桥及其在基于分数的生成建模中的应用。《神经信息处理系统进展》，第 34 卷。17695 – 17709. [58] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li 和 Li Fei-Fei. 2009. Imagenet：一个大规模分层图像数据库。在 IEEE C 会议上

[64] Laurent Dinh, Jascha Sohl-Dickstein 和 Samy Bengio. 2017 年。使用 Real NVP 进行密度估计。在国际学习表征会议上。https://openreview.net/forum?id=HkpbH9lx [65] Laurent Dinh, Jascha Sohl-Dickstein, Hugo Larochelle 和 Razvan Pascanu. 2019 年。深度混合模型的 RAD 方法。arXiv 预印本 arXiv:1903.07714 (2019)。[66] Tim Dockhorn, Arash Vahdat 和 Karsten Kreis. 2021 年。具有临界阻尼朗之万扩散的基于分数的生成模型。在国际学习表征会议上。[67] Tim Dockhorn, Arash Vahdat 和 Karsten Kreis. 2022. GENIE : 高阶去噪扩散求解器。神经信息处理系统进展 (2022)。[68] Carl Doersch. 2016. 变分自动编码器教程。arXiv 预印本 arXiv:1606.05908 (2016)。[69] Yifan Du, Zikang Liu, Junyi Li 和 Wayne Xin Zhao. 2022. 视觉语言预训练模型综述。arXiv 预印本 arXiv:2202.10936 (2022)。[70] Yilun Du 和 Igor Mordatch. 2019. 基于能量的模型中的隐式生成和泛化。arXiv 预印本 arXiv:1903.08689 (2019)。[71] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik 和 Ryan P Adams. 2015 年。用于学习分子指纹的图卷积网络。《神经信息处理系统进展》，第 28 卷。[72] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li 和 Cuntai Guan. 2021 年。通过时间和上下文对比进行时间序列表征学习。arXiv 预印本 arXiv:2106.14112 (2021)。[73] Dave Epstein, Ben Poole, Ben Mildenhall, Alexei A Efros 和 Aleksander Holynski. 2024. 通过布局学习解开 3D 场景生成。在国际机器学习会议上。[74] Patrick Esser, Robin Rombach 和 Bjorn Ommer. 2021. 驯服 Transformer 以实现高分辨率图像合成。在 IEEE 计算机视觉与模式识别会议上。12873 – 12883。[75] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee 和 Kimin Lee. 2024. 强化学习用于微调文本到图像扩散模型。神经信息处理系统进展 36 (2024 年)。[76] Charles Fefferman, Sanjoy Mitter 和 Hariharan Narayanan. 2016. 测试流形假设。美国数学学会杂志 29, 4 (2016), 983 – 1049。[77] Chelsea Finn, Paul Christiano, Pieter Abbeel 和 Sergey Levine. 2016 年。生成对抗网络、逆强化学习和基于能量的模型之间的联系。arXiv 预印本 arXiv:1611.03852 (2016 年)。[78] Vincent Fortuin, Dmitry Baranchuk, Gunnar Ratsch 和 Stephan Mandt. 2020 年。Gp-vae : 深度概率时间序列插补。在国际人工智能与统计会议上。PMLR, 1651 – 1661。[79] Giulio Franzese, Simone Rossi, Lixuan Yang, Alessandro Finamore, Dario Rossi, Maurizio Filippone 和 Pietro Michiardi. 2022. 多少才够？基于分数的生成模型中的扩散时间研究。arXiv 预印本 arXiv:2206.05173 (2022)。[80] Gege Gao, Weiyang Liu, Anpei Chen, Andreas Geiger 和 Bernhard Schölkopf. 2024. Graphdreamer : 从场景图进行组合 3D 场景合成。在 IEEE/CVF 计算机视觉与模式识别会议论文集上。21295 – 21304。[81] Ruiqi Gao, Yang Lu, Junpei Zhou, Song-Chun Zhu 和 Ying Nian Wu. 2018. 通过多网格建模和采样学习生成卷积网络。在 IEEE 计算机视觉与模式识别会议上。9155 – 9164。[82] Ruiqi Gao, Erik Nijkamp, Diederik P Kingma, Zhen Xu, Andrew M Dai 和 Ying Nian Wu. 2020 年。基于能量的模型的流量对比估计。在 IEEE 计算机视觉与模式识别会议上。7518 – 7528。[83] Ruiqi Gao, Yang Song, Ben Poole, Ying Nian Wu 和 Diederik P Kingma. 2020 年。通过扩散恢复可能性学习基于能量的模型。arXiv 预印本 arXiv:2012.08125 (2020)。[84] Wele Gedara Chaminda Bandara, Nithin Gopalakrishnan Nair 和 Vishal M Patel. 2022 年。使用去噪扩散概率模型进行遥感变化检测（分割）。arXiv 电子版 (2022 年), arXiv – 2206。[85] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals 和 George E Dahl. 2017 年。量子化学的神经信息传递。国际机器学习会议。1263-1272。[86] Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu 和 Lingpeng Kong. 2023 年。使用扩散模型的序列到序列文本生成。国际学习表征会议。[87] Wenbo Gong 和 Yingzhen Li. 2021 年。使用规范化流解释扩散得分匹配。arXiv 预印本 arXiv:2107.10072 (2021 年)。[88] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville 和 Yoshua Bengio. 2014. 生成对抗网络。神经信息学进展

[93] Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi 和 Kevin Swersky. 2019 年。您的分类器实际上是一种基于能量的模型，您应该像对待能量模型一样对待它。arXiv preprint arXiv:1912.03263 (2019)。

[94] Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud 和 Richard Zemel. 2020 年。消除中间人：无需采样即可训练和评估基于能量的模型。arXiv preprint arXiv:2002.05616 (2020)。

[95] Alex Graves. 2013 年。使用循环神经网络生成序列。CoRR abs/1308.0850 (2013)。arXiv:1308.0850 <http://arxiv.org/abs/1308.0850> [96] Ulf Grenander 和 Michael I Miller. 1994 年。复杂系统中的知识表示。皇家统计学会杂志：B 系列（方法论）56，4（1994 年），549–581。

[97] Albert Gu, Karan Goel 和 Christopher Re. 2021 年。使用结构化状态空间高效建模长序列。在国际学习表征会议上。

[98] Shuyang Gu, Dong Chen, Jianmin Bao, Fang Wen, Bo Zhang, Dongdong Chen, Lu Yuan 和 Baining Guo. 2022 年。用于文本到图像合成的矢量量化扩散模型。在 IEEE 计算机视觉与模式识别会议上。10696–10706。

[99] Jiaqi Guan, Wesley Wei Qian, Xinggang Peng, Yufeng Su, Jian Peng 和 Jianzhu Ma. 2023 年。用于目标感知分子生成和亲和力预测的 3D 等变扩散。国际学习表征会议。

[100] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao 和 Jieping Ye. 2021 年。生成对抗网络综述：算法、理论和应用。IEEE 知识与数据工程学报 (2021)。

[101] David Ha 和 Jürgen Schmidhuber. 2018 年。世界模型。arXiv 预印本 arXiv:1803.10122 (2018)。

[102] William L Hamilton, Rex Ying 和 Jure Leskovec. 2017 年。大型图上的归纳表示学习。在第 31 届国际神经信息处理系统会议论文集上。1025–1035。

[103] William L Hamilton, Rex Ying 和 Jure Leskovec. 2017 年。图上的表示学习：方法和应用。arXiv 预印本 arXiv:1709.05584 (2017)。

[104] Haonan Han, Rui Yang, Huan Liao, Jiankai Xing, Zunnan Xu, Xiaoming Yu, Junwei Zha, Xiu Li 和 Wanhua Li. 2024 年。REPARO：具有可区分 3D 布局对齐的组合 3D 资产生成。arXiv 预印本 arXiv:2405.18525 (2024)。

[105] Songqiao Han, Xiyang Hu, Hailiang Huang, Mingqi Jiang 和 Yue Zhao. 2022。AD Bench：异常检测基准。arXiv 预印本 arXiv:2206.09426 (2022)。

[106] Xiaochuang Han, Sachin Kumar 和 Yulia Tsvetkov. 2022。Ssd-lm：用于文本生成和模块化控制的半自回归单纯形扩散语言模型。arXiv 预印本 arXiv:2210.17432 (2022)。

[107] William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach 和 Frank Wood. 2022。长视频的灵活扩散建模。arXiv 预印本 arXiv:2205.11495 (2022)。

[108] Roberto Henschel, Levon Khachatryan, Daniil Hayrapetyan, Hayk Poghosyan, Vahram Tadevosyan, Zhangyang Wang, Shant Navasardyan 和 Humphrey Shi. 2024 年。StreamingT2V：从文本生成一致、动态且可扩展的长视频。arXiv 预印本 arXiv:2403.14773 (2024)。

[109] Roei Herzig, Amir Bar, Huijuan Xu, Gal Chechik, Trevor Darrell 和 Amir Globerson. 2020 年。学习场景图到图像生成的规范表示。欧洲计算机视觉会议。210–227。

[110] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet 等人。2022 年。Imagen 视频：使用扩散模型生成高清视频。arXiv 预印本 arXiv:2210.02303 (2022)。

[111] Jonathan Ho, Ajay Jain 和 Pieter Abbeel. 2020 年。去噪扩散概率模型。在《神经信息处理系统进展》第 33 卷中。6840–6851。

[112] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi 和 Tim Salimans. 2022 年。用于高保真图像生成的级联扩散模型。J. Mach. Learn. Res. 23 (2022), 47–1。

[113] Jonathan Ho 和 Tim Salimans. 2022 年。无分类器扩散指导。arXiv 预印本 arXiv:2207.12598 (2022)。

[114] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi 和 David J Fleet. 2022 年。视频扩散模型。arXiv 预印本 arXiv:2204.03458 (2022)。

[115] Emiel Hoogeboom, Victor Garcia Satorras, Clement Vignac 和 Max Welling. 2022 年。3D 分子生成的等变扩散。arXiv 电子版 (2022), arXiv – 2203。

[116] Emiel Hoogeboom, Alexey A Gritsenko, Jasmijn Bastings, Ben Poole, Rianne van den Berg 和 Tim Salimans. 2021。自回归扩散模型。在国际学习代表会议上。

[117] Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré 和 Max Welling. 2021。A rgmax 流和多项式扩散：学习分类分布。神经信息处理进展

[122] 黄志林, 杨凌, 周祥鑫, 秦楚军, 余毅杰, 郑夏武, 周子坤, 张文涛, 王宇, 杨文明. 2024. 用于蛋白质特异性 3D 分子生成的基于交互的检索增强扩散模型. 在国际机器学习会议上. [123] 黄志林, 杨凌, 周祥鑫, 张志龙, 张文涛, 郑夏武, 陈杰, 王宇, 崔斌, 杨文明. 2024. 结合感知 3d 分子扩散模型的蛋白质-配体相互作用先验. 在国际学习代表会议上. [124] 黄志林, 杨凌, 周祥鑫, 张志龙, 张文涛, 郑夏武, 陈杰, 王宇, 崔斌, 杨文明. 2024. 结合感知三维分子扩散模型的蛋白质-配体相互作用先验. 国际学习表征会议. [125] Michael F Hutchinson. 1989. 拉普拉斯平滑样条影响矩阵迹的随机估计. 统计通讯-Simulation and Computation 18, 3 (1989), 1059 – 1076. [126] Aapo Hyvärinen. 2005 年. 通过分数匹配估计非正则化统计模型. J. Mach. Learn. Res. 6 (2005), 695 – 709. [127] Touseef Iqbal 和 Shaima Qureshi. 2020 年. 综述: 深度学习中的文本生成模型. 沙特国王大学计算机与信息科学杂志 (2020 年). [128] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou 和 Alexei A Efros. 2017 年. 使用条件对抗网络进行图像到图像转换. IEEE 计算机视觉和模式识别会议. 1125 – 1134. [129] Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand 等人. 2024 年. 专家混合. arXiv 预印本 arXiv:2401.04088 (2024). [130] Long Jin, Justin Lazarow 和 Zhuowen Tu. 2017 年. 使用卷积网络进行内省分类. 神经信息处理系统进展, 第 30 卷 30. 823 – 833. [131] Wengong Jin, Regina Barzilay 和 Tommi Jaakkola. 2018 年. 用于分子图生成的连接树变分自动编码器. 国际机器学习会议. 2323 – 2332. [132] Bowen Jing, Gabriele Corso, Renato Berlinghieri 和 Tommi Jaakkola. 2022 年. 子空间扩散生成模型. arXiv 预印本 arXiv:2205.01490 (2022). [133] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay 和 Tommi Jaakkola. 2022 年. 扭转扩散用于分子构象生成. arXiv 预印本 arXiv:2206.01729 (2022 年). [134] Jaehyeong Jo, Seul Lee 和 Sung Ju Hwang. 2022 年. 通过随机微分方程系统实现基于分数的图生成建模. 国际机器学习会议. PMLR, 10362 – 10383. [135] Justin Johnson, Agrim Gupta 和 Li Fei-Fei. 2018 年. 从场景图生成图像. IEEE 计算机视觉和模式识别会议论文集. 1219 – 1228. [136] Alexia Jolicoeur-Martineau, Ke Li, Rémi Piché-Taillefer, Tal Kachman 和 Ioannis Mitliagkas. 2021 年. 使用基于分数的模型生成数据时必须快速行动. (2021 年). [137] Alexia Jolicoeur-Martineau, Remi Piche-Taillefer, Rémi Tachet des Combes 和 Ioannis Mitliagkas. 2021 年. 对抗性得分匹配和改进的图像生成采样. ArXiv abs/2009.05475 (2021 年). [138] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko 等人. 2021 年. 使用 AlphaFold 实现高精度蛋白质结构预测. 《自然》596, 7873 (2021), 583 – 589. [139] Heewoo Jun 和 Alex Nichol. 2023 年. Shap-e: 生成条件 3d 隐式函数. arXiv pre print arXiv:2305.02463 (2023). [140] Nal Kalchbrenner, Erich Elsen, Karen Simonyan, Seb Noury, Norman Casagrande, Edward Lockhart, Florian Stimberg, Aäron van den Oord, Sander Dieleman 和 Koray Kavukcuoglu. 2018 年. 高效神经音频综合. 在国际机器学习会议上. 2410 – 2419. [141] Tero Karras, Miika Aittala, Timo Aila 和 Samuli Laine. 2022 年. 阐明基于扩散的生成模型的设计空间. arXiv 预印本 arXiv:2206.00364 (2022). [142] 泰罗·卡拉斯, 萨穆利·莱恩, 蒂莫·艾拉. 2019. 一种基于风格的生成对抗网络生成器架构. IEEE/CVF 计算机视觉和模式识别会议论文集. 4401 – 4410. [143] Bahjat Kawar, Michael Elad, Stefano Ermon 和 Jiaming Song. 2022 年. 去噪扩散恢复模型. arXiv 预印本 arXiv:2201.11793 (2022). [144] Bahjat Kawar, Roy Ganz 和 Michael Elad. 2022 年. 通过稳健的分类器指导增强基于扩散的图像合成. arXiv 预印本 arXiv:2208.08664 (2022). [145] Bahjat Kawar, Gregory Vaksman 和 Michael Elad. 2021 年. 通过从后验分布中采样进行随机图像去噪. 在 IEEE/CVF 国际计算机视觉会议论文集. 1866-1875 年. [146] Bahjat Kawar, Shiran Za

[150] Dongjun Kim, Byeonghu Na, Se Jung Kwon, Dongsoo Lee, Wanmo Kang 和 Il-chul Moon. 2022 年. 隐式非线性扩散模型的最大似然训练. 神经信息处理系统的发展. [151] Jihoon Kim, Jiseob Kim 和 Sungjoon Choi. 2022 年. Flame: 基于自由形式语言的运动合成与编辑. arXiv 预印本 arXiv:2209.00349 (2022). [152] Sungwon Kim, Heeseung Kim 和 Sungroh Yoon. 2022 年. Guided-TTS 2: 一种用于使用未转录数据的高质量自适应文本转语音的扩散模型. arXiv 预印本 arXiv:2205.15370 (2022). [153] Taesup Kim 和 Yoshua Bengio. 2016. 具有基于能量的概率估计的深度定向生成模型. arXiv 预印本 arXiv:1606.03439 (2016). [154] Diederik Kingma, Tim Salimans, Ben Poole 和 Jonathan Ho. 2021. 变分扩散模型. 在《神经信息处理系统进展》第 34 卷中. 21696 – 21707. [155] Diederik P Kingma 和 Prafulla Dhariwal. 2018. Glow: 具有可逆 1x1 卷积的生成流. arXiv 预印本 arXiv:1807.03039 (2018). [156] Diederik P Kingma 和 Max Welling. 2013. 自动编码变分贝叶斯. arXiv 预印本 arXiv:1312.6114 (2013). [157] Diederik P Kingma, Max Welling 等人. 2019 年. 变分自动编码器简介. 机器学习基础与趋势® 12, 4 (2019 年), 307 – 392. [158] Daphne Koller 和 Nir Friedman. 2009 年. 概率图形模型: 原理和技术. 麻省理工学院出版社. [159] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao 和 Bryan Catanzaro. 2020 年. Diffwave: 一种用于音频合成的多功能扩散模型. arXiv 预印本 arXiv:2009.09761 (2020 年). [160] Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher 和 Nazneen Fatema Rajani. 2020. Gedi: 生成式鉴别器引导序列生成. arXiv 预印本 arXiv:2009.06367 (2020). [161] Alex Krizhevsky. 2009. 从微小图像中学习多层特征. (2009 年). [162] Rithesh Kumar, Anirudh Goyal, Aaron Courville 和 Yoshua Bengio. 2019. 基于能量模型的最大熵生成器. arXiv 预印本 arXiv:1901.08508 (2019). [163] Hugo Larochelle 和 Iain Murray. 2011. 神经自回归分布估计器. 第十四届国际人工智能与统计会议论文集, AISTATS. [164] Justin Lazarow, Long Jin 和 Zhuowen Tu. 2017. 用于生成模型的内省神经网络. IEEE 国际计算机视觉会议论文集. 2774 – 2783. [165] Yann LeCun, Sumit Chopra, Raia Hadsell, Marc ' Aurelio Ranzato 和 Fugie Huang. 2006. 基于能量的学习教程. 预测结构化数据 1, 0 (2006 年). [166] Jin Sub Lee 和 Philip M Kim. 2022. ProteinSGM: 基于分数的从头蛋白质设计生成模型. bioRxiv (2022 年). [167] Kimin Lee, Hao Liu, Moonkyung Ryu, Olivia Watkins, Yuqing Du, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh 和 Shixiang Shane Gu. 2023. 使用人工反馈对齐文本到图像模型. arXiv 预印本 arXiv:2302.12192 (2023). [168] Kwonjoon Lee, Weijian Xu, Fan Fan 和 Zhuowen Tu. 2018 年. Wasserstein 内省神经网络. 在 IEEE 计算机视觉与模式识别会议上. 3702 – 3711. [169] Seul Lee, Jaehyeong Jo 和 Sung Ju Hwang. 2022 年. 使用基于分数的分布外生成探索化学空间. arXiv 预印本 arXiv:2206.07632 (2022). [170] Alon Levkovitch, Eliya Nachmani 和 Lior Wolf. 2022 年. 用于去噪扩散 TTS 模型的零样本语音条件. arXiv 预印本 arXiv:2206.02246 (2022). [171] 李浩英, 杨一凡, 常孟, 冯华军, 徐志海, 李琪, 陈悦婷. 2022. SRDiff: 具有扩散概率模型的单图像超分辨率. 神经计算 479 (2022), 47 – 59. [172] 李俊毅, 唐天一, 何高乐, 蒋金浩, 胡晓轩, 谢普照, 陈志鹏, 于卓浩, 赵鑫, 温继荣. 2021. Textbox: 一个统一的、模块化的、可扩展的文本生成框架. arXiv 预印本 arXiv:2101.02046 (2021). [173] 李俊毅, 唐天一, 赵鑫, 温继荣. 2021. 用于文本生成的预训练语言模型: 一项调查. arXiv preprint arXiv:2105.10311 (2021). [174] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyu Zhou, Wenhui Chen, Yu-Xiang Wang 和 Xifeng Yan. 2019 年. 增强 Transformer 在时间序列预测中的局部性并突破内存瓶颈. 神经信息处理系统进展, 第 32 卷. [175] Xiang Lisa Li, John Thackstun, Ishaan Gulrajani, Percy Liang 和 Tatsunori B Hashimoto. 2022 年. Diffusion-LM 改进了可控文本生成. arXiv preprint arXiv:2205.14217 (2022). [176] Yikang Li, Tao Ma, Yeqi Bai, Nan Duan, Sining Wei 和 Xiaogang Wang. 2019 年. Pastegan: 一种从场景图生成图像的半参数方法. 神经信息处理系统进展 32 (2019). [177] Long Lian, Boyi Li, Adam Yala 和 Trevor Darrell. 2023 年. 基于法学的扩散: 增强快速理解

[180] 刘经纬、杨凌、李红艳、洪申达。2024 年。用于时间序列预测的检索增强扩散模型。神经信息处理系统进展。[181] 刘路平、任毅、林志杰、赵舟。2021 年。流形上扩散模型的伪数值方法。国际学习表征会议。[182] 刘胜超、郭洪宇、唐健。2023 年。使用 $se(3)$ 不变去噪距离匹配进行分子几何预训练。国际学习表征会议。[183] 刘兴超、龚成月等。2022 年。直流快速：学习使用整流流生成和传输数据。第十一届国际学习表征会议。[184] 刘星超、吴乐萌、叶茂等。2023 年。在受限域上学习扩散桥。在国际学习表征会议上。[185] 刘星超、吴乐萌、叶茂和刘强。2022 年。让我们架起桥梁：理解和扩展扩散生成模型。arXiv 预印本 arXiv:2208.14699 (2022)。[186] Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser Nam Lim 和 Christopher M De Sa。2020 年。神经流形常微分方程。神经信息处理系统进展 33 (2020), 17548 – 17558。[187] 陆程、郑凯文、包帆、陈建飞、李崇轩和朱军。2022。基于分数的扩散 ODE 的最大似然训练，通过高阶去噪分数匹配。在国际机器学习会议上。14429 – 14460。[188] 程璐、周宇浩、包帆、陈建飞、李崇轩和朱军。2022。DPM-Solver：一种用于扩散概率模型采样的快速 ODE 求解器，大约 10 个步骤。arXiv 预印本 arXiv:2206.00927 (2022)。[189] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte 和 Luc Van Gool。2022。重绘：使用去噪扩散概率模型进行修复。在 IEEE 计算机视觉与模式识别会议上。11461 – 11471。[190] Eric Luhman 和 Troy Luhman。2021 年。迭代生成模型中的知识蒸馏以提高采样速度。arXiv 预印本 arXiv:2101.02388 (2021)。[191] Calvin Luo。2022 年。理解扩散模型：一个统一的视角。arXiv 预印本 arXiv:2208.11970 (2022)。[192] Shengjie Luo, Tianlang Chen, Yixian Xu, Shuxin Zheng, Tie-Yan Liu, Liwei Wang 和 Di He。2023 年。一个 Transformer 可以理解 2d 和 3d 分子数据。在国际学习表征会议上。[193] Shitong Luo 和 Wei Hu。2021 年。用于 3d 点云生成的扩散概率模型。在 IEEE 计算机视觉与模式识别会议上。2837 – 2845。[194] 罗世通和胡伟。2021。基于分数的点云去噪。IEEE/CVF 国际计算机视觉会议论文集。4583 – 4592。[195] 罗世通、石晨、徐敏凯和唐健。2021。通过动态图分数匹配预测分子构象。神经信息处理系统进展，第 34 卷。19784 – 19795。[196] 罗世通、苏宇峰、彭新刚、王胜、彭健和马建柱。2022。基于扩散的生成模型的抗原特异性抗体设计和优化。bioRxiv (2022)。[197] 罗永红、蔡祥瑞、张颖、徐军等人 2018。基于生成对抗网络的多元时间序列插补。载于《神经信息处理系统进展》，第 31 卷。[198] 吕兆阳、孔志峰、徐旭东、潘亮和林大华。2021。一种用于 3D 点云补全的条件点扩散-细化范式。载于国际学习表征会议。[199] 吕兆阳、徐旭东、杨策远、林大华和戴波。2022。通过提前停止扩散过程加速扩散模型。arXiv 预印本 arXiv:2205.12524 (2022)。[200] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras 和 Adrian Vladu。2018。面向抵抗对抗攻击的深度神经网络。在国际学习表征会议上。[201] Emile Mathieu 和 Maximilian Nickel。2020。黎曼连续正则化流。神经信息处理系统进展 33 (2020), 2503 – 2515。[202] Siyuan Mei, Fuxin Fan 和 Andreas Maier。2022。使用基于分数的生成模型在 CBCT 投影中进行金属修复。arXiv 预印本 arXiv:2209.09733 (2022)。[203] Gábor Melis, Chris Dyer 和 Phil Blunsom。2018。论神经语言模型评估的最新进展。在国际学习表征会议上。https://openreview.net/forum?id=ByJHuTgA- [204] Chenlin Meng, Kristy Choi, Jiaming Song 和 Stefano Ermon。2022 年。具体分数匹配：离散数据的广义分数匹配。神经信息处理系统进展。[205] Chenlin Meng, Ruiqi Gao, Diederik P Kingma, Stefano Ermon, Jonathan Ho 和 Tim Salimans。2022 年。论引导扩散模型的提炼。

- [208] 孟晨林、宋嘉明、宋扬、赵胜佳和 Stefano Ermon. 2021 年。带有分布平滑的改进自回归建模。国际学习表征会议。[209] 孟晨林、宋扬、李文哲和 Stefano Ermon. 2021 年。通过去噪估计数据分布的高阶梯度。神经信息处理系统进展 34 (2021), 25359 – 25369. [210] 孟晨林、余澜涛、宋扬、宋嘉明和 Stefano Ermon. 2020 年。自回归分数匹配。神经信息处理系统进展 33 (2020), 6673 – 6683. [211] Stephen Merity、Nitish Shirish Keskar 和 Richard Socher. 2018 年。正则化和优化 LSTM 语言模型。在国际学习表征会议上。https://openreview.net/forum?id=SyGPP0TZ [212] Nicholas Metropolis 和 Stanislaw Ulam. 1949 年。蒙特卡罗方法。美国统计协会杂志 44, 247 (1949 年), 335-341. [213] Jiquan Ngiam、Zhenghao Chen、Pang W Koh 和 Andrew Y Ng. 2011 年。学习深度能量模型。在国际机器学习会议上。1105-1112. [214] Alexander Quinn Nichol 和 Prafulla Dhariwal. 2021 年。改进的去噪扩散概率模型。在国际机器学习会议上。8162-8171. [215] Alexander Quinn Nichol、Prafulla Dhariwal、Aditya Ramesh、Pranav Shyam、Pamela Mishkin、Bob McGrew、Ilya Sutskever 和 Mark Chen. 2022 年。GLIDE：使用文本引导扩散模型实现逼真的图像生成和编辑。国际机器学习会议。16784-16804. [216] Weili Nie、Brandon Guo、Yujia Huang、Chaowei Xiao、Arash Vahdat 和 Anima Anandkumar. 2022 年。对抗性净化的扩散模型。arXiv 预印本 arXiv:2205.07460 (2022)。[217] Erik Nijkamp、Mitch Hill、Tian Han、Song-Chun Zhu 和 Ying Nian Wu. 2019. 基于能量模型的 mcmc 最大似然学习剖析。arXiv 预印本 arXiv:1903.12370 (2019)。[218] Erik Nijkamp、Mitch Hill、Song-Chun Zhu 和 Ying Nian Wu. 2019. 学习面向基于能量的模型的非收敛短期 MCMC。arXiv 预印本 arXiv:1904.09770 (2019)。[219] Chenhao Niu、Yang Song、Jiaming Song、Shengjia Zhao、Aditya Grover 和 Stefano Ermon. 2020. 通过基于分数的生成模型生成置换不变图。在国际人工智能与统计会议上。PMLR, 4474 – 4484. [220] OpenAI. 2023. GPT-4 技术报告。arXiv 预印本 arXiv:2303.08774 (2023)。[221] R OpenAI. 2023 年。Gpt-4 技术报告。arxiv 2303.08774. 查看文章 2 (2023), 3. [222] Boris N Oreshkin、Dmitri Carpv、Nicolas Chapados 和 Yoshua Bengio. 2019 年。N-BEATS：用于可解释时间序列预测的神经基础扩展分析。在国际学习代表会议上。[223] Boris N. Oreshkin、Dmitri Carpv、Nicolas Chapados 和 Yoshua Bengio. 2020 年。N-BEATS：用于可解释时间序列预测的神经基础扩展分析。在国际学习代表会议上。[224] Long Ouyang、Jeffrey Wu、Xu Jiang、Diogo Almeida、Carroll Wainwright、Pamela Mishkin、Chong Zhang、Sandhini Agarwal、Katarina Slama、Alex Ray 等人。2022 年。训练语言模型以遵循带有人类反馈的指令。神经信息处理系统进展 35 (2022), 27730 – 27744. [225] Muzaffer Özbey、Salman UH Dar、Hasan A Bedel、Onat Dalmaz、Şaban Öztürk、Alper Güngör 和 Tolga Çukur. 2022 年。使用对抗扩散模型进行无监督医学图像翻译。arXiv 预印本 arXiv:2207.08208 (2022)。[226] George Papamakarios、Eric T Nalisnick、Danilo Jimenez Rezende、Shakir Mohamed 和 Balaji Lakshminarayanan. 2021 年。概率建模和推理的规范化流。J. Mach. Learn. Res. 22, 57 (2021 年), 1-64. [227] Giorgio Parisi. 1981 年。相关函数和计算机模拟。核物理 B 180, 3 (1981 年), 378-384. [228] Sung Woo Park、Kyungjae Lee 和 Junseok Kwon. 2021 年。神经马尔可夫控制的 SDE：连续时间数据的随机优化。在国际学习表征会议上。[229] William Peebles 和 Saining Xie. 2022 年。使用 Transformer 的可扩展扩散模型。arXiv preprint arXiv:2212.09748 (2022)。[230] Cheng Peng、Pengfei Guo、S Kevin Zhou、Vishal M Patel 和 Rama Chellappa. 2022 年。通过扩散模型采样实现高性能、可靠的欠采样 MR 重建。在国际医学图像计算和计算机辅助干预会议上。Springer, 623 – 633. [231] Ethan Perez、Florian Strub、Harm De Vries、Vincent Dumoulin 和 Aaron Courville. 2018 年。电影：具有一般条件层的视觉推理。在 AAAI 人工智能会议论文集第 32 卷。[232] Stanislav Pidhorskyi、Donald A Adjeroh 和 Gianfranco Doretto. 2020 年。对抗性潜在自动编码器。在 IEEE 计算机视觉与模式识别会议上。14104 – 14113. [233] 达斯汀·波德尔、锡安·英格兰什、凯尔·拉克

[237] 齐晨阳、村晓东、张勇、雷晨阳、王新涛、单颖、陈奇峰。2023 年。FateZero：融合注意力机制实现零样本文本视频编辑。arXiv 预印本 arXiv:2303.09535 (2023)。[238] 邱义轩、张凌松、王晓。2019 年。用于训练基于能量的隐变量模型的无偏对比散度算法。国际学习表征会议。[239] Lawrence R Rabiner。1989 年。隐马尔可夫模型教程及其在语音识别中的应用。IEEE 论文集 77, 2 (1989), 257 – 286。[240] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark 等人, 2021 年。通过自然语言监督学习可迁移视觉模型。国际机器学习会议。8748 – 8763。[241] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever 等人, 2018 年。通过生成式预训练提高语言理解能力。(2018 年)。[242] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever 等人, 2019 年。语言模型是无监督的多任务学习器。OpenAI 博客 1, 8 (2019), 9。[243] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu 和 Mark Chen。2022 年。使用剪辑潜在数据的分层文本条件图像生成。arXiv 预印本 arXiv:2204.06125 (2022)。[244] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen 和 Ilya Sutskever。2021 年。零样本文本到图像生成。在国际机器学习会议上。8821 – 8831。[245] Martin Raphan 和 Eero P Simoncelli。2007 年。学习无监督贝叶斯。在神经信息处理系统的进展中。1145 – 1152。[246] Martin Raphan 和 Eero P Simoncelli。2011。无先验或监督的最小二乘估计。神经计算 23, 2 (2011), 374 – 420。[247] Kashif Rasul, Calvin Seward, Ingmar Schuster 和 Roland Vollgraf。2021。用于多元概率时间序列预测的自回归去噪扩散模型。在国际机器学习会议上。8857 – 8868。[248] Kashif Rasul, Calvin Seward, Ingmar Schuster 和 Roland Vollgraf。2021。用于多元概率时间序列预测的自回归去噪扩散模型。在国际机器学习会议上。8857 – 8868。[249] Kashif Rasul, Abdul-Saboor Sheikh, Ingmar Schuster, Urs M Bergmann 和 Roland Vollgraf。2020。通过条件正则化流进行多元概率时间序列预测。国际学习表征会议。[250] Lillian J Ratliff, Samuel A Burden 和 S Shankar Sastry。2013。连续博弈中局部纳什均衡的表征与计算。2013 年第 51 届 Allerton 通信、控制和计算年会 (Allerton)。IEEE, 917 – 924。[251] Danilo Rezende 和 Shakir Mohamed。2015。带正则化流的变分推理。国际机器学习会议。1530 – 1538。[252] Danilo Jimenez Rezende, Shakir Mohamed 和 Daan Wierstra。2014。深度生成模型中的随机反向传播和近似推理。国际机器学习会议。1278 – 1286。[253] Benjamin Rhodes, Kai Xu 和 Michael U Gutmann。2020 年。伸缩密度比估计。神经信息处理系统进展, 第 33 卷。4905-4916。[254] Oren Rippel 和 Ryan Prescott Adams。2013 年。使用深度密度模型进行高维概率估计。arXiv 预印本 arXiv:1302.5125 (2013)。[255] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser 和 Björn Ommer。2022 年。使用潜在扩散模型进行高分辨率图像合成。IEEE 计算机视觉与模式识别会议。10684-10695。[256] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein 和 Kfir Aberman。2022 年。DreamBooth：针对主题驱动生成的文本到图像扩散模型的微调。arXiv 预印本 arXiv:2208.12242 (2022)。[257] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet 和 Mohammad Norouzi。2022 年。Palette：图像到图像扩散模型。计算机图形学和交互技术特别兴趣小组会议论文集。1-10。[258] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes 等人。2022 年。具有深度语言理解的逼真文本到图像扩散模型。arXiv 预印本 arXiv:2205.11487 (2022)。[259] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet 和 Mohammad Norouzi。2022 年。通过迭代细化实现图像超分辨率。IEEE 模式分析与机器智能交易 (2022)。[260] Tim Salimans 和 Jonathan Ho。2021 年。用于快速采样扩散模型的渐进式蒸馏。在国际学习会议上

[265] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner 和 Gabriele Monfardini. 2008 年。图神经网络模型。IEEE 神经网络交易 20, 1 (2008 年), 61-80. [266] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth 和 Georg Langs. 2017 年。使用生成对抗网络进行无监督异常检测以指导标记发现。在国际医学成像信息处理会议上。Springer, 146-157. [267] Chence Shi, Shitong Luo, Minkai Xu 和 Jian Tang. 2021 年。学习梯度场用于分子构象生成。在国际机器学习会议上。9558-9568. [268] 石晨、徐敏凯、朱兆成、张伟南、张明和唐健。2020 年。Graphaf：一种基于流的分子图生成自回归模型。arXiv 预印本 arXiv:2001.09382 (2020)。[269] 石宇阳、Valentin De Bortoli, George Deligiannidis 和 Arnaud Doucet. 2022 年。使用扩散薛定谔桥的条件模拟。arXiv 预印本 arXiv:2202.13460 (2022)。[270] 石一春、王鹏、叶江龙、麦龙、李克杰和杨晓。2024 年。MVDream：用于 3D 生成的多视图扩散。第十二届国际学习表征会议。[271] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu 和 Gordon Wetzstein. 2023 年。使用三平面扩散生成 3d 神经场。IEEE/CVF 计算机视觉与模式识别会议论文集。20875-20886. [272] Ikaro Silva, George Moody, Daniel J Scott, Leo A Celi 和 Roger G Mark. 2012 年。预测 ICU 患者住院死亡率：2012 年心脏病学物理网络/计算挑战赛。2012 年心脏病学计算。IEEE, 245-248. [273] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni 等人。2022。制作视频：无需文本视频数据即可生成文本转视频。arXiv 预印本 arXiv:2209.14792 (2022)。[274] 约翰·斯基林。1989。巨维矩阵的特征值。《最大熵与贝叶斯方法》。Springer, 455 – 466 页。[275] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan 和 Surya Ganguli. 2015。使用非平衡热力学的深度无监督学习。国际机器学习会议。2256 – 2265 页。[276] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan 和 Surya Ganguli. 2015。使用非平衡热力学的深度无监督学习。在国际机器学习会议上, Francis R. Bach 和 David M. Blei (编辑)。2256 – 2265. [277] Jiaming Song, Chenlin Meng 和 Stefano Ermon. 2020 年。去噪扩散隐式模型。在国际学习表征会议上。[278] Ki-Ung Song. 2022 年。在生成模型中应用基于正则化薛定谔桥的随机过程。arXiv 预印本 arXiv:2208.07131 (2022 年)。[279] Yang Song, Conor Durkan, Iain Murray 和 Stefano Ermon. 2021 年。基于分数的扩散模型的最大似然训练。在神经信息处理系统进展中, 第 34 卷。1415 – 1428. [280] Yang Song 和 Stefano Ermon. 2019。通过估计数据分布梯度进行生成建模。神经信息处理系统进展, 第 32 卷。[281] 宋杨和 Stefano Ermon. 2020。改进的基于分数的生成模型训练技术。神经信息处理系统进展, 第 33 卷。12438 – 12448. [282] 宋杨, Sahaj Garg, 石嘉欣和 Stefano Ermon. 2019。切片分数匹配：一种可扩展的密度和分数估计方法。在第三十五届人工智能不确定性会议论文集 (UAI 2019), 以色列特拉维夫, 2019 年 7 月 22 日至 25 日。204。http://auai.org/uai2019/proceedings/papers/204.pdf [283] Yang Song 和 Diederik P Kingma. 2021 年。如何训练基于能量的模型。arXiv 预印本 arXiv:2101.03288 (2021)。[284] Yang Song, Liyue Shen, Lei Xing 和 Stefano Ermon. 2021 年。使用基于分数的生成模型解决医学成像中的逆问题。在国际学习表征会议上。[285] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon 和 Ben Poole. 2020 年。通过随机微分方程实现基于分数的生成建模。在国际学习表征会议上。[286] James C Spall. 2012 年。随机优化。在《计算统计手册》中。Springer, 173-201. [287] Jiachen Sun, Weili Nie, Zhiding Yu, Z Morley Mao 和 Chaowei Xiao. 2022 年。PointDP：针对 3D 点云识别中对抗攻击的扩散驱动净化。arXiv 预印本 arXiv:2208.09801 (2022)。[288] Jaesung Tae, Hyeonju Kim 和 Taesu Kim. 2021。EdiTTS：基于分数的可控文本转语音编辑。arXiv 预印本 arXiv:2110.02584 (2021)。[289] 谭华春、冯广东、冯建帅、王武红、张玉金和李峰。2013。一种基于张量的缺失流量检测方法

[294] Lucas Theis, Aäron van den Oord 和 Matthias Bethge. 2015 年。关于生成模型评估的说明。arXiv 预印本 arXiv:1511.01844 (2015 年)。[295] Ye Tian, Ling Yang, Haotian Yang, Yuan Gao, Yufan Deng, Jingmin Chen, Xintao Wang, Zhaochen Yu, Xin Tao, Pengfei Wan, Di Zhang 和 Bin Cui. 2024 年。VideoTetris：面向组合式文本到视频生成。神经信息处理系统的发展 (2024 年)。[296] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar 等人。2023。Llama：开放高效的基础语言模型。arXiv 预印本 arXiv:2302.13971 (2023)。[297] Brandon Trabucco, Kyle Doherty, Max Gurinas 和 Ruslan Salakhutdinov. 2023。使用扩散模型进行有效的数据增强。在国际学习表征会议上。[298] Brian L Trippe, Jason Yim, Doug Tischer, Tamara Broderick, David Baker, Regina Barzilay 和 Tommi Jaakkola. 2023。针对基序支架问题的 3D 蛋白质骨架扩散概率建模。在国际学习表征会议上。[299] Arash Vahdat, Karsten Kreis 和 Jan Kautz. 2021。基于分数的潜在空间生成模型。在神经信息处理系统进展中, 第 3 卷 34。11287 – 11302。[300] Dani Valevski, Matan Kalman, Yossi Matias 和 Yaniv Leviathan. 2022 年。UniTune：通过在单幅图像上微调图像生成模型实现文本驱动的图像编辑。arXiv 预印本 arXiv:2210.09477 (2022)。[301] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior 和 Koray Kavukcuoglu. 2016 年。WaveNet：原始音频的生成模型。在第 9 届 ISCA 语音合成研讨会上。[302] Aäron van den Oord, Nal Kalchbrenner 和 Koray Kavukcuoglu. 2016。像素循环神经网络。在国际机器学习会议上, Maria-Florina Balcan 和 Kilian Q. Weinberger (编辑)。1747 – 1756 年。[303] Alexander Vilesov, Pradyumna Chari 和 Achuta Kadambi. 2023。Cg3d：通过高斯分层实现文本到 3D 的组合生成。arXiv 预印本 arXiv:2311.17907 (2023 年)。[304] Pascal Vincent. 2011。分数匹配与去噪自动编码器之间的联系。神经计算 23, 7 (2011 年), 1661 – 1674。[305] Pascal Vincent, Hugo Larochelle, Yoshua Bengio 和 Pierre-Antoine Manzagol. 2008。使用去噪自动编码器提取和组合稳健特征。国际机器学习会议。1096 – 1103。[306] Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty 和 Nikhil Naik. 2024。使用直接偏好优化的扩散模型对齐。IEEE/CVF 计算机视觉与模式识别会议论文集。8228 – 8238。[307] Fu-Yun Wang, Ling Yang, Zhaoyang Huang, Mengdi Wang 和 Hongsheng Li. 2024。整流扩散：整流流中不需要直线性。arXiv 预印本 arXiv:2410.07303 (2024)。[308] 王进一, 吕昭阳, 林大华, 戴波, 付鸿飞。2022。对抗性净化的引导扩散模型。arXiv 预印本 arXiv:2205.14969 (2022)。[309] 王宇飞, 郑嘉怡, 徐灿, 耿秀波, 沉涛, 陶重阳, 蒋大新。2022。KnowDA：用于少样本 NLP 中数据增强的一体化知识混合模型。arXiv 预印本 arXiv:2206.10265 (2022)。[310] 王正义, 陆成, 王一凯, 包凡, 李崇轩, 苏航, 朱军。2023。ProlificDreamer：通过变分分数蒸馏实现高保真且多样化的文本到 3D 生成。arXiv preprint arXiv:2305.16213 (2023)。[311] 王振东, 郑黄杰, 何鹏程, 陈伟柱, 周明远。2022。Diffusion-GAN：使用 Diffusion 训练 GAN。arXiv preprint arXiv:2206.02262 (2022)。[312] Daniel Watson, William Chan, Jonathan Ho 和 Mohammad Norouzi. 2021。通过样本质量区分学习扩散模型的快速采样器。在国际学习表征会议上。[313] Daniel Watson, Jonathan Ho, Mohammad Norouzi 和 William Chan. 2021。学习从扩散概率模型中高效采样。arXiv preprint arXiv:2106.03802 (2021)。[314] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler 等人。2022 年。大型语言模型的新兴能力。机器学习研究汇刊 (2022 年)。[315] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis 和 Peyman Milanfar. 2022 年。通过随机细化进行去模糊。在 IEEE 计算机视觉与模式识别会议上。16293 – 16303。[316] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang 和 Nan Duan. 2023 年。视觉聊天 GPT：使用视觉基础模型进行交谈、绘图和编辑。arXiv 预印本 arXiv:2303.04671 (2023)。[317] 吴浩, Jonas Köhler 和

[322] 吴宗翰、潘诗睿、陈峰文、龙国栋、张成琪和 S Yu Philip. 2020 年。图神经网络综合综述。IEEE 神经网络与学习系统学报 32, 1 (2020), 4–24. [323] Julian Wyatt, Adam Leach, Sebastian M Schmon 和 Chris G Willcocks. 2022 年。AnoDDPM：使用单纯形噪声的去噪扩散概率模型进行异常检测。在 IEEE 计算机视觉与模式识别会议上。650–656. [324] 肖志胜、Karsten Kreis 和 Arash Vahdat. 2021 年。使用去噪扩散 GAN 解决生成学习三难困境。arXiv 预印本 arXiv:2112.07804 (2021). [325] 谢建文、陆阳、朱松春和吴英年. 2016 年。生成式卷积网络理论。国际机器学习会议。2635–2644. [326] 谢攀、张启鹏、李泽贤、唐浩、杜耀和胡晓辉. 2022 年。使用 CodeUnet 的矢量量化扩散模型生成文本到手势姿势序列。arXiv 预印本 arXiv:2208.09141 (2022). [327] 谢天、傅翔、Octavian-Eugen Ganeva、Regina Barzilay 和 Tommi S Jaakkola. 2021 年。用于周期性材料生成的晶体扩散变分自编码器。国际学习表征会议。[328] 谢宇桐和李全政. 2022 年。用于欠采样医学图像重建的测量条件下的去噪扩散概率模型。arXiv 预印本 arXiv:2203.03623 (2022). [329] Jiarui Xu, Sifei Liu, Arash Vahdat, Wonmin Byeon, Xiaolong Wang 和 Shalini De Mello. 2023. 使用文本到图像扩散模型的开放词汇全景分割。在 IEEE 计算机视觉与模式识别会议上。[330] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang 和 Yuxiao Dong. 2024. Imagereward：学习和评估人类对文本到图像生成的偏好。神经信息处理系统进展 36 (2024). [331] 徐家乐、王新涛、程伟豪、曹彦培、单颖、郗晓虎、高胜华. 2022 年。Dream3D：使用 3D 形状先验和文本到图像扩散模型的零样本文本到 3D 合成。arXiv 预印本 arXiv:2212.14704 (2022). [332] 徐明浩、王航、倪冰冰、郭宏宇、唐健. 2021 年。具有局部和全局结构的自监督图级表征学习。国际机器学习会议。11548–11558. [333] 徐敏凯、余岚涛、宋阳、石晨、Stefano Ermon、唐健. 2021 年。GeoDiff：用于分子构象生成的几何扩散模型。国际学习表征会议。[334] 徐兴干、王张扬、张伟、王凯、史汉弗莱. 2022. 多功能扩散：文本、图像和变体合二为一的扩散模型。arXiv 预印本 arXiv:2211.08332 (2022). [335] 严体金、张宏伟、周同、詹玉峰、夏元庆. 2021. ScoreGrad：使用基于连续能量的生成模型进行多元概率时间序列预测。arXiv 预印本 arXiv:2106.10121 (2021). [336] 杨东超、于建伟、王鹤林、王文、翁超、邹月贤、于东. 2022. DiffSound：用于文本到声音生成的离散扩散模型。arXiv preprint arXiv:2207.09983 (2022). [337] 杨杰、徐瑞杰、齐志全和石勇. 2021 年。图像视觉异常检测：一项调查。arXiv preprint arXiv:2109.13157 (2021). [338] 杨凯文和丹·克莱因. 2021 年。FUDGE：具有未来鉴别器的受控文本生成。（2021 年）。[339] 杨凯、陶建、吕嘉飞、葛春江、陈嘉欣、沈伟翰、朱小龙和李秀. 2024 年。使用人类反馈微调扩散模型（无需任何奖励模型）。IEEE/CVF 计算机视觉与模式识别会议论文集。8941–8951. [340] 杨玲和洪申达. 2022 年。用于自监督图表征学习的全粒度自我语义传播。arXiv 预印本 arXiv:2205.15746 (2022). [341] 杨凌和洪申达. 2022. 使用迭代双线性时间-频谱融合的无监督时间序列表征学习。国际机器学习会议。25038–25054. [342] 杨凌、黄志林、宋扬、洪申达、李国豪、张文涛、崔斌、Bernard Ghanem 和杨明轩. 2022. 使用蒙版对比预训练的基于扩散的场景图到图像生成。arXiv 预印本 arXiv:2211.11138 (2022). [343] 杨凌、黄志林、张志龙、刘忠义、洪申达、张文涛、杨文明、崔斌、张露霞. 2024. Graphusion：图生成的潜在扩散。IEEE 知识与数据工程汇刊 (2024). [344] 杨凌、李亮亮、张子伦、周新宇、周尔金、刘宇. 2020. Dpgn：用于少样本学习的分布传播图网络。在 IEEE 计算机视觉和模式识别会议上。13390–13399. [345] 杨凌、刘经纬、洪申达、张志龙、黄志林、蔡哲明、张文涛、崔斌. 2023. 即兴表演

[351]杨凌, 张志龙, 余兆晨, 刘经纬, 徐敏凯, Stefano Ermon, 崔斌. 2024.用于文本引导视觉生成和编辑的跨模态情境化扩散模型。在国际学习代表会议上。 [352]杨凌, 张志龙, 张文涛, 洪申达. 2023.具有自引导潜在扩散的基于分数的图生成模型。(2023)。 <https://openreview.net/forum?id=AykEgQNPJEK> [353] 杨凌、张子祥、张志龙、刘兴超、徐敏凯、张文涛、孟陈林、Stefano Ermon 和崔斌. 2024.一致性流匹配：定义具有速度一致性的直线流。 arXiv preprint arXiv:2407.02398 (2024)。 [354] Ruihan Yang 和 Stephan Mandt. 2022 年。具有条件扩散模型的有损图像压缩。 arXiv preprint arXiv:2209.06950 (2022)。 [355] Ruihan Yang, Prakhar Srivastava 和 Stephan Mandt. 2022. 视频生成的扩散概率建模。 arXiv 预印本 arXiv:2203.09481 (2022)。 [356] Xiuwen Yi, Yu Zheng, Junbo Zhang 和 Tianrui Li. 2016. ST-MVL：填充地理传感时间序列数据中的缺失值。在第 25 届国际人工智能联合会议论文集上。 [357] Jongmin Yoon, Sung Ju Hwang 和 Juho Lee. 2021 年。基于分数的生成模型的对抗性净化。国际机器学习会议。12062 – 12072。 [358] Jinsung Yoon, Daniel Jarrett 和 Mihaela Van der Schaar . 2019. 时间序列生成对抗网络。神经信息处理系统进展, 第 32 卷。 [359] Zebin You, Yong Zhong, Fan Bao, Jiacheng Sun, Chongxuan Li 和 Jun Zhu. 2023. 扩散模型和半-监督学习者通过少量标签实现互惠互利。 arXiv 预印本 arXiv:2302.10586 (2023)。 [360] 于嘉辉、王子瑞、Vijay Vasudevan、Legg Yeung、Mojtaba Seyedhosseini 和吴永辉. 2022. Coca：对比字幕是图像文本基础模型。 arXiv 预印本 arXiv:2205.01917 (2022)。 [361] 于佩玉, 谢思瑞, 马晓建, 贾宝雄, 庞波, 高瑞琪, 朱一新, 朱松春, 吴英念. 2022.用于可解释文本建模的基于潜在扩散能量的模型。在国际机器学习会议上。 25702 – 25720。 [362] Sihyun Yu, Jihoon Tack, Sangwoo Mo, Hyunsu Kim, Junho Kim, Jung-Woo Ha 和 Jinwoo Shin. 2022. 使用动态感知隐式生成对抗网络生成视频。 arXiv 预印本 arXiv:2202.10571 (2022)。 [363] Lu Yuan, Dongdong Chen, Yi-Ling Chen, Noel Codella, Xiyang Dai, Jianfeng Gao, Houdong Hu, Xuedong Huang、Boxin Li、Chunyan Li 等人. 2021 年。Florence：计算机视觉的新基础模型。 arXiv 预印本 arXiv:2111.11432 (2021)。 [364] Shehryar Zaidi, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye Teh, Alvaro Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu 和 Jonathan Godwin. 2023 年。通过去噪进行预训练以预测分子特性。在国际学习表征会议上。 [365] 曾博涵, 李尚林, 冯玉堂, 杨凌, 李红, 高思成, 刘家明, 何从辉, 张文涛, 刘建庄, 张保昌, 严水成. 2023. Ipdreamer：具有图像提示的外观可控的 3D 对象生成。 arXiv 预印本 arXiv:2310.05375 (2023)。 [366] 曾博涵、杨凌、李思宇、刘家明、张子祥、田娟熙、朱凯新、郭永振、王福云、徐敏凯、Stefano Ermon 和张文涛. 2024. Trans4D：用于组合文本到 4D 合成的现实几何感知过渡。 arXiv 预印本 arXiv:2410.07155 (2024)。 [367] Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler 和 Karsten Kreis. 2022 年。LION：用于 3D 形状生成的潜在点扩散模型。神经信息处理系统进展。 [368] Lvmin Zhang 和 Maneesh Agrawala. 2023 年。在文本到图像的扩散模型中添加条件控制。 arXiv 预印本 arXiv:2302.05543 (2023)。 [369] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang 和刘子伟. 2022 年。Motiondiffuse：基于扩散模型的文本驱动人体运动生成。 arXiv 预印本 arXiv:2208.15001 (2022 年)。 [370] 张勤胜和陈永欣. 2021. 扩散归一化流。神经信息处理系统进展, 第 34 卷。16280 – 16291。 [371] 张勤胜和陈永欣. 2022. 使用指数积分器的扩散模型快速采样。 arXiv 预印本 arXiv:2204.13902 (2022)。 [372] 张勤胜、陶莫蕾、陈永欣. 2022 年。gDDIM：广义去噪扩散隐式模型。 arXiv 预印本 arXiv:2206.05564 (2022)。 [373] 张苏珊、史蒂芬·罗勒、纳曼·戈亚尔、迈克尔·阿特克斯、陈莫亚、Shuohui Chen、Christopher Dewan、Mona Diab、Xian Li、Xi Victoria Lin 等人. 2022 年。Opt：开放预训练的 Transformer 语言模型。 arXiv preprint arXiv:2205.01068 (2022)。 [374] 张文睿、杨凌、耿诗佳、洪申达. 2022 年。用于自监督时间序列重构的交叉重构 Transformer

[380] 赵越, Zain Nasrullah, 郑力。2019.PyOD：用于可扩展异常值检测的 Python 工具箱。机器学习研究杂志 20 (2019), 1 – 7。[381] 郑黄杰, 何鹏程, 陈伟柱, 周明远。2022.截断扩散概率模型。arXiv 预印本 arXiv:2202.09671 (2022)。[382]周庚模, 高志峰, 丁乾坤, 郑航, 徐宏腾, 魏哲伟, 张林峰, 柯国林。2023. Uni-mol：通用 3D 分子表示学习框架。在国际学习代表会议上。[383]周杰, 崔甘曲, 胡胜定, 张正彦, 杨程, 刘志远, 王立峰, 李长城, 孙茂松。2020. 图神经网络：方法与应用综述。AI Open 1 (2020), 57 – 81。[384] 周林琪、杜逸伦、吴嘉俊。2021. 通过点体素扩散实现 3d 形状生成与完成。在 IEEE/CVF 国际计算机视觉会议论文集上。5826 – 5835。[385] 朱浩伟、杨玲、雍俊海、张文涛和王斌。2024 年。使用扩散模型进行分布感知数据扩展。arXiv preprint arXiv:2403.06741 (2024)。[386] Ye Zhu, Yu Wu, Kyle Olszewski, Jian Ren, Sergey Tulyakov 和 Yan Yan。2022 年。用于跨模态和条件生成的离散对比扩散。arXiv 预印本 arXiv:2206.07771 (2022)。[387]朱艳桥, 徐一尘, 于峰, 刘强, 吴树, 王亮。2020.深度图对比表示学习。arXiv 预印本 arXiv:2006.04131 (2020)。[388]庄少斌, 李昆昌, 陈新元, 王耀辉, 刘紫薇, 乔宇, 王亚丽。2024.视频博主：让你的梦想成为视频博客。arXiv 预印本 arXiv:2401.09414 (2024)。[389]罗兰·S·齐默尔曼、卢卡斯·肖特、杨松、本杰明·邓恩和大卫·A·克林特。2021.基于分数的生成分类器。arXiv 预印本 arXiv:2110.00473 (2021)。