计算机科学
COMPUTER SCIENCE

生成扩散模型研究综述

闫志浩, 周长兵, 李小翠

引用本文

闫志浩, 周长兵, 李小翠. 生成扩散模型研究综述[J]. 计算机科学, 2024, 51(1): 273-283.

YAN Zhihao, ZHOU Zhangbing, LI Xiaocui. Survey on Generative Diffusion Model[J]. Computer Science, 2024, 51(1): 273-283.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230300057

生成扩散模型研究综述

闫志浩　周长兵　李小翠
中国地质大学(北京)信息工程学院　北京 100083
(2204220028@email.cugb.edu.cn)

摘　要　扩散模型在生成模型领域具有高质量的样本生成能力,一经推出就不断地刷新图像生成评价指标 FID 分数的记录,成为了该领域的研究热点,而此类相关综述在国内还鲜有介绍。因此,文中对相关扩散生成模型的研究进行汇总与分析。首先,对去噪扩散概率模型、基于分数的扩散生成模型和随机微分方程的扩散生成模型这 3 类通用模型的特点和原理进行了论述,就每一类基本扩散模型中以优化模型内部算法、高效采样为改进目标的相关衍生模型进行分析。其次,对当下扩散模型在计算机视觉、自然语言处理、时间序列、多模态和跨学科领域等方面的应用进行总结。最后,基于上述论述,分别就目前扩散生成模型存在的采样步骤多、采样时间长等局限性提出了相关建议,并结合前述研究对未来扩散生成模型的发展方向进行了研判。

关键词:深度学习;生成模型;去噪扩散概率模型;基于分数的扩散模型;随机微分方程;图像生成
中图分类号　TP183

Survey on Generative Diffusion Model

YAN Zhihao,ZHOU Zhangbing and LI Xiaocui
School of Information Engineering,China University of Geosciences(Beijing),Beijing 100083,China

Abstract　Diffusion models have shown high-quality sample generation ability in the field of generative models,and constantly set new records for image generation evaluation indicator FID scores since their introduction,and has become a research hotspot in this field. However,related reviews of this kind are scarce in China. Therefore,this paper aims to summarize and analyze the research on related diffusion generative models. Firstly,it analyzes the related derivative models in each basic diffusion model,which focus on optimizing internal algorithms and efficient sampling,by discussing the characteristics and principles of three common models:denoising diffusion probabilistic model,score-based diffusion generative model,and diffusion generative model based on random differential equations. Secondly,it summarizes the current applications of diffusion models in computer vision,natural language processing,time series,multimodal,and interdisciplinary fields. Finally,based on the above discussion,relevant suggestions for the existing limitations of diffusion generative models are proposed,such as long sampling times and multiple sampling steps,and a research direction for the future development of diffusion generative models is provided based on previous studies.

Keywords　Deep learning,Generative models,Denoising diffusion probabilistic models,Score-based diffusion models,Stochastic differential equations,Image generation

1　引言

近年来,行业间的交流愈加便利,信息量也随之激增,这对计算机硬件方面的发展提出了更高的要求。随着计算机运算能力的提升,各个领域与人工智能的结合也愈加紧密。其中机器学习作为人工智能的核心,学术界给予了其充分关注。机器学习分为有监督学习和无监督学习,其中的有监督学习需要繁琐的数据收集和标注,耗时且费力。而在无监督学习方面,为了提高模型的泛化能力、降低泛化误差,会在模型训练之前进行简单的变换处理操作来对数据进行扩充,从而得出更优秀的特征提取模型。现有的最有效的数据扩充方式是通过生成模型生成目标样本。

根据分布中提供的样本,生成模型可以对其真实的数据分布进行建模,进而训练出模型来生成新的样本。生成模型通常基于马尔可夫链、最大似然估计以及近似推理。早期的生成模型受限于玻尔兹曼机[1]、深度信念网络[2]以及深度玻尔兹曼机[3],泛化能力较差,后期的改进生成模型如生成对抗网络(Generative Adversarial Networks,GAN)[4]、变分自编码器(Variational Autoencoder,VAE)[5]、基于流的模型(Flow-based Models)[6-7]、基于能量的模型(Energy Based Model,EBM)[8]、自回归模型[9]等在生成高质量样本方面取得了巨大的成功。基于以上工作,Ho 等[10]引入了生成模型

到稿日期:2023-03-06　返修日期:2023-07-01
基金项目:国家自然科学基金(42050103)
This work was supported by the National Natural Science Foundation of China(42050103).
通信作者:周长兵(zbzhou@cugb.edu.cn)

领域的一个新的概念，即扩散概率模型（Denoising Diffusion Probabilistic Models，DDPM），该模型在前向阶段对数据逐步施加噪声，直至数据被破坏变成完全的高斯噪声，然后在逆向阶段学习，从高斯噪声将其还原为原始数据。该模型的出现成为了近年来生成模型的热门话题之一（见图 1），打破了 GAN 在具有挑战性的图像合成任务中的长期主导地位，并且在计算机视觉[11-21]、自然语言处理[22-27]、时间序列[28-30]、多模态[31-37]以及与传统科目[38-46]的结合等领域都有着不俗的表现。

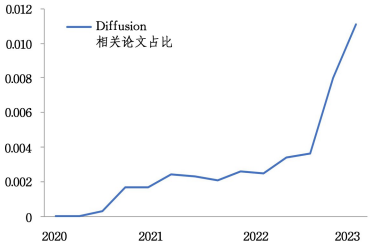

图 1 Papers With Code 网站上近年来 Diffusion 相关论文占比的变化

Fig. 1 Changes in the percentage of diffusion-related papers on the Papers With Code website in recent years

目前，系统地总结扩散模型最新进展的英文综述较少[47-49]，中文综述更是极度缺乏。因此，学术界迫切需要对扩散模型的最新文献及进展进行系统的总结、归纳，并分析当下所存在的问题，以及对未来发展趋势的预测。本文系统梳理了扩散模型从被提出至今的技术演进，并总结了该方向代表性的算法和技术。

2 生成扩散模型的介绍

扩散模型是一类概率生成模型，此模型主要分为正向扩散阶段与反向扩散阶段。如图 2 所示，在正向扩散阶段输入数据，通过逐渐添加高斯噪声的方式来对原始数据进行破坏。在反向扩散阶段，生成模型的任务是通过学习逆转扩散过程，进而从噪声数据中恢复原始输入数据。

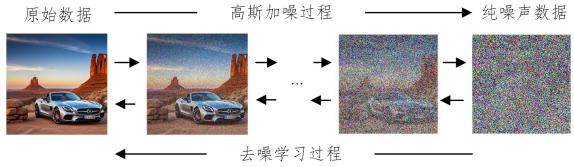

图 2 扩散模型的处理流程

Fig. 2 Process of diffusion model

目前较为主流的生成模型主要有生成对抗网络（GAN）[4]、变分自编码器（VAE）[5]、基于流的模型（Flow-based Models）[6-7]和基于能量的模型（EBM）[8]，这 4 种生成模型的基本框架如图 3 所示。GAN 在训练过程中容易出现模式崩溃，导致训练失败。相比之下，扩散模型是一种基于似然的模型，它的训练过程较为稳定且具有多样性，但其在推理过程中需要进行多次网络评估，效率较低。VAE 的潜在表征空间包含原始图像的压缩信息，但其维度会相对减少。扩散模型虽然在前向过程的最后一步完全破坏了数据，但其具有与原始数据相同的维度。Flow-based Models 与扩散模型都是将数据分布映射到高斯噪声，然而 Flow-based Models 是通过学习一个可逆和可微的函数，以一种确定的方式进行映射，导致了对网格结构的额外约束。而扩散模型由于具有可学习的正反向过程，相对而言可扩展性较好。EBM 侧重于提供密度函数非归一化形式的估计值，因此相比其他模型而言更为灵活，从而导致较难进行训练。而扩散模型基于高斯模型进行训练，结果更加稳定。

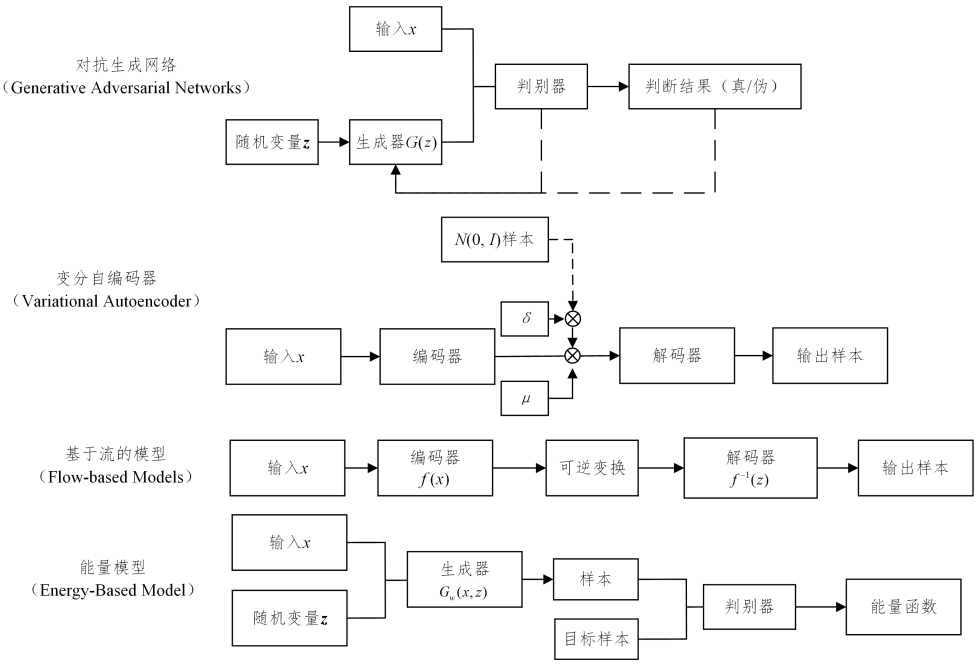

图 3 4 种主流生成模型的框架图

Fig. 3 Framework of four mainstream generative models

目前的扩散模型主要分为 3 类：去噪扩散概率模型（DDPM）[10,50-51]、基于分数的生成模型[52-54]以及基于随机微分方程的生成模型（Stochastic Differential Equations，SDEs）[55]。本文对这 3 种模型的区别进行了详细的介绍。

2.1 去噪扩散概率模型

去噪扩散概率模型的灵感来自非平衡热力学[7]，训练过程包括两个阶段：正向扩散加噪过程和逆向去噪过程。去噪扩散概率模型的处理过程如图4所示。

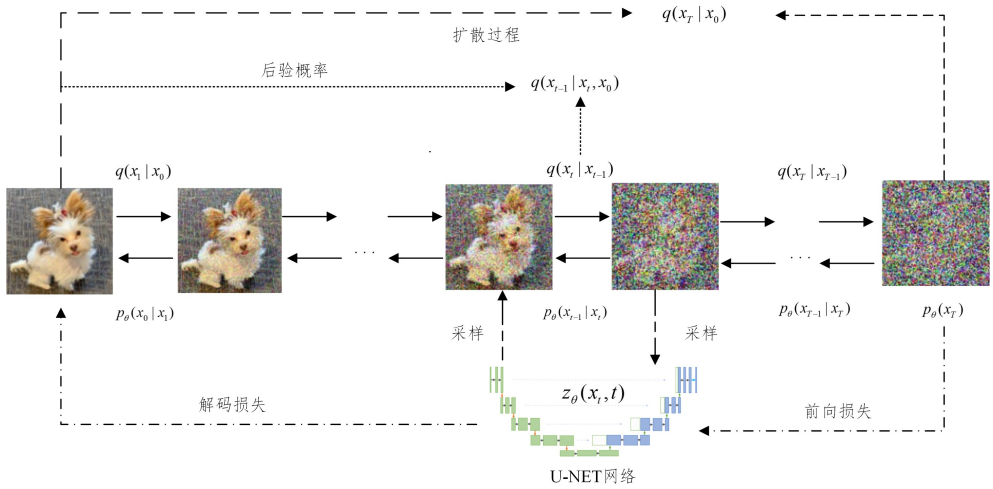

图4 去噪扩散模型的处理过程

Fig. 4 Process of denoising diffusion probabilistic models

由图4可以看出，前向扩散过程是通过给定一个从真实数据分布中采样的未损坏的数据样本$x_0 \sim q(x_0)$，来产生一系列服从高斯分布的噪声样本$x_0,\cdots,x_T$，并将其缓慢地添加到输入样本中，即可得到以下的马尔可夫过程。

$$q(x_t|x_{t-1})=N(x_t;\sqrt{1-\beta_t}\cdot x_{t-1},\beta_t\cdot \mathbf{I}),\quad \forall t\in\{1,\cdots,T\} \tag{1}$$

其中，T为总的扩散步骤，噪声的增长水平方差$\beta_t\in(0,1)$为服从高斯分布的超参数，$\mathbf{I}$代表与输入样本x_0具有相同维数的单位矩阵，$N(x;\mu,\delta)$代表生成x的均值μ和协方差δ的正态分布。且每一步的样本都只和$t-1$时刻的样本有关。因此，当t服从均匀分布时，就可以最终可以得到式(2)：

$$q(x_t|x_0)=N(x_t;\sqrt{\hat{\beta}_t}\cdot x_0,(1-\hat{\beta}_t)\cdot \mathbf{I}) \tag{2}$$

其中，$\hat{\beta}_t=\prod_{i=1}^{t}(1-\beta_t)$，因为$q(x_t|x_0)$是通过重参数化技巧[56]生成的，为了使服从正态分布的样本x标准化，需要减去平均值μ并除以标准差σ，从而得到服从标准正态分布的z，由此可通过标准逆变换推出x_t的递推公式为：

$$x_t=\sqrt{\hat{\beta}_t}\cdot x_0+\sqrt{1-\hat{\beta}_t}\cdot z_t \tag{3}$$

其中，噪声$z_t\sim N(0,\mathbf{I})$。通过上述处理原始数据就会被破坏，产生纯高斯噪声样本x_T。

根据上述情况，逆向传播过程起始于样本$x_T\sim N(0,\mathbf{I})$，并通过相反的步骤来进行去噪$p(x_{t-1}|x_t)=N(x_{t-1};\mu(x_t,t),\sum(x_t,t))$。但由于$p(x_{t-1}|x_t)$需要整条传播途中的数据，因此比较难以估计，故需要训练一个神经网络$p_\theta(x_{t-1}|x_t)=N(x_{t-1};\mu_\theta(x_t,t),\sum_\theta(x_t,t))$，其中$x_t$为每一步的噪声图像，$\theta$为模型参数，$t$为时间步长，$\mu_\theta(x_t,t)$为可学习的平均值，$\sum_\theta(x_t,t)$为可学习的协方差。

为了使反向马尔可夫链尽可能地匹配正向过程，需要不断地调整θ，使反向过程的联合分布逐渐接近正向过程。这里使用到了Sohl-Dickstein等[57]提出的最小化负对数似然的变分下界的解决方法，具体如式(4)所示：

$$\sum_{t>1}KL(q(x_{t-1}|x_t,x_0)\parallel p_\theta(x_{t-1}|x_t))+KL(q(x_T|x_0)\parallel p(x_T))-\log p_\theta(x_0|x_1) \tag{4}$$

其中，KL表示两个概率分布之间的相对熵，也称为Kullback-Leibler(KL)散度；$p(x_T)$代表反向过程的开始状态，服从$N(x_T;0,\mathbf{I})$分布。由于$KL(q(x_T|x_0)\parallel p(x_T))$不依赖于$\theta$，可将其删除，最终可得出需要进行计算的公式为$\sum_{t>1}KL(q(x_{t-1}|x_t,x_0)\parallel p_\theta(x_{t-1}|x_t))$。其目的是在每个$t$时刻，要使$p_\theta(x_{t-1}|x_t)$尽可能地接近正向过程中的真实后验概率。对于$p_\theta$中的$\sum_\theta(x_t,t)$，本文将其设置为常数，可训练的$\mu_\theta(x_t,t)$如下：

$$\mu_\theta(x_t,t)=\frac{1}{\sqrt{\alpha_t}}\left(x_t-\frac{1-\alpha_t}{\sqrt{1-\hat{\beta}_t}}\cdot z_\theta(x_t,t)\right) \tag{5}$$

其中，$z_\theta(x_t,t)$是需要预测的噪声参数，具体对应的损失函数如下：

$$\mathbb{E}_{t\sim[1,T]}\mathbb{E}_{x_0\sim p(x_0)}\mathbb{E}_{z_t\sim N(0,\mathbf{I})}\parallel z_t-z_\theta(x_t,t)\parallel^2 \tag{6}$$

其中，$\mathbb{E}$为期望值，x_t通过式(3)计算得到，x_0为从训练集随机抽取的图像。Ronneberger等[10]使用U-Net[58]卷积神经网络进行此噪声参数的训练。

2.2 基于分数的生成扩散模型

最初的分数生成模型[54]的主要思想是用一组不同水平的高斯噪声序列对数据进行干扰，并通过训练一个以噪声为条件的深度神经网络模型，来计算出噪声数据分布的分数函数。其中的样本是通过Langevin动力学[59]生成的，每个样本$\{x_i\in R^D\}$独立同分布，分布函数为$p(x)$，概率密度函数为$\nabla_x\log p(x)$，而此密度函数是未知的，因此需要训练一个方程$s_\theta(x)$来进行估计。其中方程$s_\theta(x):R^D\to R^D$是由θ参数化的神经网络，通过最小化以下目标函数来进行训练。

$$\mathbb{E}_{p(x)}\parallel s_\theta(x)-\nabla_x\log p(x)\parallel_2^2 \tag{7}$$

但是，此方法的局限性就是在数据密度较低的区域估计不会十分准确，从而不能有效地获得高质量样本，因此在改进后的NCSN[52]算法中提出在低密度区域添加噪声的方案，这样噪声就可以填满原数据分布中概率密度较低的区域(见

图 5),并提供给数据密度小的区域如何抵达密度高区域的信息。其中噪声使用多尺度的噪声来进行数据干扰,例如当有 N 个噪声信号时,按标准差从小到大$\sigma_1<\sigma_2<\cdots<\sigma_N$进行排列,添加扰动之后的数据$p_{\sigma_i}(x)\approx p(x_0)$符合标准正态分布,之后训练一个深度神经网络$s_\theta(x,\sigma_t)$去估计分数方程$\nabla_x \log p_{\sigma_t}(x)$,$\forall t\in\{1,\cdots,T\}$,其中$\nabla_x \log\ p_{\sigma_t}(x)$通过计算后为:

$$\nabla_{x_t}\log p_{\sigma_t}(x_t \mid x)=-\frac{x_t-x}{\sigma_t^2} \tag{8}$$

并通过最小化以下目标函数来训练$s_\theta(x_t,\sigma_t)$,其中$\forall t\in\{1,\cdots,T\}$。

$$\frac{1}{T}\sum_{t=1}^{T}\lambda(\sigma_t)\mathbb{E}_{p(x)}\mathbb{E}_{x_t\sim p_{\sigma_t}(x_t\mid x)}\left\|s_\theta(x_t,\sigma_t)+\frac{x_t-x}{\sigma_t^2}\right\|_2^2 \tag{9}$$

其中,$\lambda(\sigma_t)$为权重参数,Song 等[52]将它设置为$\sigma_t{}^2$。得到$s_\theta(x_t,\sigma_t)$之后,Song 等使用退火 Langevin Dynamics[52]算法完成样本的生成工作。

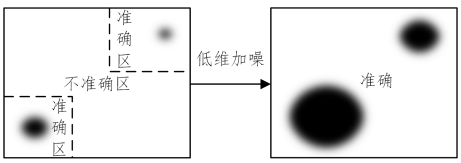

图 5 低维加噪过程图

Fig. 5 Diagram of low-dimensional noisy process

2.3 基于随机微分方程的扩散生成模型

由于噪声数量的增加会伴随样本生成质量的提高,因此 Song 等[55]提出了一个可以无穷随机生成噪声的方法,用于提高样本的生成质量。其随机方程为:

$$\mathrm{d}x=f(x,t)\mathrm{d}t+g(t)\mathrm{d}w \tag{10}$$

其中,$f(x,t)$和$g(t)$为随机微分方程的飘移系数和扩散系数,漂移系数主要是需要逐渐将原始样本数据进行无效化处理,而扩散系数主要是对噪声量增加量的控制;w 表示标准的布朗运动;$\mathrm{d}w$ 可以视为无穷小的白噪声。对于式(10)的正向加噪过程,其对应的生成样本的逆过程为:

$$\mathrm{d}x=[f(x,t)-g^2(t)\nabla_x\log\ p_t(x)]\mathrm{d}t+g(t)\mathrm{d}w \tag{11}$$

其中,$\mathrm{d}t$ 为负无穷小的时间步长,$\mathrm{d}w$ 表示反向过程中的布朗运动,当$g^2(t)\nabla_x\log\ p_t(x)$无限趋向于 $f(x,t)$时,就可实现逆向求解,而其中的$\nabla_x\log\ p_t(x)$如同上一节基于分数的生成模型一样,需要训练一个$s_\theta(x_t,t)$来进行估计,其训练的目标函数为:

$$\mathbb{E}_t[\lambda(t)\mathbb{E}_{p(x_0)}\mathbb{E}_{p_t(x_t\mid x_0)}\|s_\theta(x_t,t)-\nabla_{x_t}\log\ p_t(x_t\mid x_0)\|_2^2] \tag{12}$$

其中,λ 为一个加权函数,t 为 0 到 T 的均匀分布,通过使此目标函数最小化,估算出$\nabla_x\log\ p_t(x)$的值来进行逆向去噪,且当漂移系数为仿射生成时可以使用 denoising score matching[60],其他情况下可以用 sliced score matching[61]方法进行样本生成工作。

3 扩散模型的发展及其衍生模型

由前述内容可知,由于当前扩散模型本身对马尔可夫链生成的样本依赖性较强,在计算方面会花费大量的时间,并且随着扩散步长的增多,估算得分函数也会逐渐变得困难,因此当下研究者们也采用了多种方法去改进或解决这些问题。表 1 列出了基于前述扩散模型所改进的其他模型。由表 1 可知,目前的主要研究还是根据第 2 章所述的 3 种扩散模型基础理念来开展的。因此,本章将结合其衍生模型及其相关应用领域来展开对当前研究现状的论述工作。

表 1 扩散模型与其相关衍生模型分类

Table 1 Classification of diffusion models and their related derivative models

模式	改进目标	论文	改进点	架构名称	数据集
概率去噪扩散模型	噪声优化	Nichol 等[62]	正向过程添加余弦噪声	—	LSUN,ImageNet
		Kingma 等[63]	网络添加傅里叶特征	—	CIFAR-10,ImageNet
		San-Roman 等[64]	动态调整噪声参数	—	CelebA,LSUN
		Wang 等[65]	去噪过程中加入净化机制	GDMP	CIFAR-10,ImageNet
	改进马尔可夫链	Song 等[66]	使用非马尔可夫正向过程	DDIM	CIFAR10
		Zhang 等[67]	使用分数近似值进行快速抽样	gDDIM	CIFAR10
	多模型结合	Sinha 等[68]	加入对比表示方法	D2C	CIFAR-10,fMoW,FFHQ
		Peebles 等[69]	结合 Transformers 模型	—	ImageNet
	针对特殊数据	Giannone 等[51]	解决少样本生成问题	FSDM	CIFAR100,MinImageNet,CelebA
		Sehwag 等[70]	在低密度数据中采样	ADM	CIFAR-10,ImageNet
		Kim 等[71]	提出深度感知模型	DAG	LSUN
		Austin 等[72]	使用离散状态空间生成模型	D3PM	CIFAR-10
	优化采样效率	Watson 等[73]	重参数化和重复梯度计算	—	LSUN
		Watson 等[74]	反向过程加入动态规划	—	ImageNet
		Xiao 等[75]	最小化 KL 散度	—	CIFAR-10
		Lam 等[50]	使用双边去噪扩散模型	BDDM	CIFAR-10,CelebA
		Bao 等[76]	优化协方差设计	—	CIFAR10,CelebA,ImageNet
		Chung 等[77]	减少采样步骤	CCDF	FFHQ,AFHQ
基于分数的生成模型	改进采样算法	Song 等[52]	引入退火采样法	—	CelebA,FFHQ,LSUN
		Song 等[53]	使用近似最大似然训练	—	CIFAR-10,ImageNet
		Jolicoeur-Martineau 等[12]	改进一致性退火采样方案	—	CIFAR-10
	基础框架改进	Vahdat 等[78]	提出可变自动编码器框架	LSGM	CIFAR-10,CelebA-HQ-256,OMNIGLOT
		Zhang 等[79]	加入矩阵、进行预处理	PDS	MNIST,LSUN
		Du 等[80]	提出参数化扩散模型通用框架	FP-Diffusion	MINIST,CIFAR10

（续表）

模式	改进目标	论文	改进点	架构名称	数据集
基于随机微分方程的生成模型	多模型结合	Ho 等[81]	使用无分类器引导生成	—	ImageNet
		Zhang 等[82]	提出扩散归一化流方法	DiffFlow	MNIST，CIFAR-10
		Kim 等[83]	提出非线性扩散模型	INDM	CelebA
	改进采样算法	Dockhorn 等[84]	加入临界阻尼 Langevin	CLD	MINIST，CIFAR10
		Bortoli 等[85]	优化迭代比例拟合过程	DSB	MINIST，CelebA
		Liu 等[86]	使用伪数值方法进行去噪	PNDM	CIFAR-10，CelebA，LSUN
		Jolicoeur-Martineau 等[87]	提出自适应步长解决方案	—	CIFAR-10

3.1 基于概率去噪扩散模型的优化

目前概率去噪扩散模型内部有丰富的参数，导致模型的性能严重依赖于模型参数的选择，因此以优化各项参数为目标的相关衍生模型也逐渐被提出。

3.1.1 噪声优化

在噪声优化方面，基于概率扩散模型[10]的研究，Nichol 等[62]发现通过在正向加噪过程中添加一定的余弦噪声可以获得更好的对数似然性，并且在反向去噪过程中添加了可学习的方差，以减少采样步骤。Kingma 等[63]将傅里叶特征添加到网络的输入中来预测噪声，并分析扩散模型的变分下限（VLB），通过分析发现扩散损失只会受到信噪比函数极值的影响。还有一种动态调整噪声参数[64]的新方法，其使用 VGG-11 卷积神经网络来训练出最合适的噪声参数，生成的图像样本也具有更高的 FID 值。在样本生成的过程中，不可避免地会伴有对抗性攻击来扰乱样本的生成，提纯噪音框架 GDMP[65]，通过将净化过程加入到去噪概率模型的去噪过程中，其可以选择合适的扩散时间步长让高斯噪声淹没对抗性扰动，并同时保留输入图像中的主要内容，从而提高分类的正确性。

3.1.2 改进马尔可夫链

对于传统的正向马尔可夫过程，去噪扩散隐式模型（DDIM）[66]证明了使用非马尔可夫过程也能够获得不错的生成结果。在去噪过程中首先预测正常样本，然后将它用到下一步的估计中，这种改变使得模型的采样速度变得更快，并且对生成样本质量的影响也较小。在 DDIM 模型被提出之后，Zhang 等[67]提出了 gDDIM 模型，将注意力转移到数值角度后发现，在求解相应的随机微分方程时，可以使用分数的特定近似值来获得 DDIM，并解释了使用确定性的抽样方案相比随机的方案会更加快速地进行采样。

3.1.3 多模型结合

在扩散模型与经典模型结合方面，Sinha 等[68]提出了具有对比表示学习思想的扩散解码模型，他们通过学习扩散先验分布来改进生成，再使用对比自监督学习来提高表示的质量。在生成任务上优于当时的 VAE[5]模型。Peebles 等[69]结合 Transformers[88]模型将反向生成图像中常用的 U-Net 网络替换为了 Transformers 网络，并发现当提高网络的深度广度或者增加 token 时，可以获得更好的训练效果。GAN[4]模型与扩散模型相结合[89]弥补了 GAN 在生成样本时不稳定的缺陷，它通过扩散模型将噪声注入到自适应噪声生成计划鉴别器中，来弥补由输入数据和生成数据的分布不重叠造成的生成样本不稳定问题。

3.1.4 针对特殊数据

对于样本量较少且主要特征集中在低密度区域的非常规数据，Sehwag 等[70]改进了扩散模型的采样过程。他们在每一个时间步骤中使用两个额外的分类器来优化采样，一个是将扩散关注度从高密度区域转向低密度区域，另一个是能保证生成关注度停留在低样本的数据流形上，使其可以在数据密度较低的区域生成高质量样本。

对于少样本的数据，FSDM[51]就是一个利用条件 DDPM 进行少样本生成的框架，其使用 VIT[90]框架来聚合图像块信息，从而训练 FSDM 去学习已有类别的小规模图像的生成过程，并将已学习到的各种条件分布生成为更加丰富而复杂的样本，用于弥补样本少的缺点。DAG[71]将关注点聚焦在具有几何性质的图像上，其提出了一种有效利用内部表示来对扩散模型生成图像并进行深度感知的方法。

对于离散的数据处理，Austin 等[72]提出了用离散扩散模型来进行样本生成，在正向过程中加入多个过渡矩阵，并提出了将变分下限与辅助交叉熵损失集合起来的新的损失函数，其解决方案在图像生成的对数似然性方面超过了连续扩散模型的性能。

3.1.5 超参数优化

由于扩散模型对马尔可夫链的依赖较强，在正反向的过程中模型处理效率会相对较低。针对这个问题，Watson 等[73]使用重参数化和重复梯度计算的方法来优化扩散模型的快速采样器，并将 KID 差异指标作为损失函数，使用随机梯度下降来对其进行优化，并且通过一种特别的抽样参数化组来减少抽样步骤，进而获得更好的效果。Lam 等[50]提出了双边去噪扩散模型，该模型使用调度网络和评分网络对正向和反向过程进行参数化训练处理，此模型的样本生成步骤相比之前明显变少。

3.1.6 降低 KL 散度

Watson 等[74]在反向去噪过程中将动态规划算法融入到模型中，该算法利用了证据下界（ELBO）可以被分为单独的相对熵（KL 散度）项的情况，通过最小化 KL 散度来最大化 ELBO，从而找到最佳的推理路径来提高推理效率。Xiao 等[75]在反向去噪过程中整合 GAN 来区分真实样本和去噪后的样本的区别，从而最小化 KL 散度来提高推理效率。

3.1.7 减少采样步骤

由于扩散模型在整个时间步长的迭代生成过程中的效率较低，Bao 等[76]使用对角和完全协方差来优化时间步长，提高了 DDPM 的生成效率。Chung 等[77]使用随机差分方程的收缩理论发现，对正向过程中初始化的图像进行优化，能够明显减少反向去噪过程中的步骤，从而提高生成效率。

3.2 基于分数的生成扩散模型优化

3.2.1 改进采样算法

在前述分数模型[54]的基础上，为了改进当下训练模型仅

限于低分辨率图像且生成图像不稳定的情况，Song 等[52]使用了新的策略来决定噪声生成尺度。在采样方面，他们建议将指数移动平均应用于参数，并为 Langevin 动力学选择超参数。而且他们还将分数与损失匹配的加权组合最小化，用于分数扩散模型的近似最大似然训练[53]。对于采样过程中所用到的退火采样法，Jolicoeur-Martineau 等[12]使用了一种更加稳定的一致性退火采样方案，并提出了一个由去噪分数和对抗目标组成的混合训练公式。

3.2.2 训练梯度优化

由于分数的生成模型需要多次迭代的顺序计算，这使得它们的推断变得非常缓慢。针对这个问题，LSGM[78]提出了一个可变自动编码器框架在潜在空间中训练分数生成模型的方法，其主要是将分数生成模型应用于非连续数据，并在更小的空间中学习更平滑的模型，从而减少网络评估并进行更快的采样。预条件扩散采样(PDS)[79]模型在保持其目标分布的同时，通过矩阵预处理重新表述扩散过程，从而避免原来存在的病态曲率，提升模型效率。

3.2.3 其他改进方面

当前分数的生成模型，在正向过程中大部分还是通过人工来进行设计，Du 等[80]通过结合黎曼几何和蒙特卡罗方法的理念，来分析其模型之间的深层联系，据此提出了基于正向过程的参数化扩散模型的通用框架，并通过在标准数据集上的测试，证明了此方法的有效性。

3.3 基于随机微分方程的生成扩散模型的优化

3.3.1 多模型结合

在基础模型的改进方面，Zhang 等[82]在微分方程的基础上，引入了一种将标准化流[6]与随机微分方程(SDE)[55]相结合的建模方法，主要是通过联合训练正反 SDE 神经网络，使两者之间差异的共同成本函数最小化，后向 SDE 扩散过程以高斯分布开始，以期望的数据分布结束。同样使用此方法的还有 Kim 等[83]，他们提出的模型主要基于 SDE 模型原有的线性扩散模式，提出了一种非线性扩散模型，其主要使用可训练的标准化流与扩散过程相结合的模型，通过流网络在潜在空间中进行线性扩散来学习噪声分布，再将其用在数据空间上进行非线性扩散。对于当前模型使用分类器引导导致采样结果严重受限于该分布的局部领域，Ho 等[81]介绍了一种无分类器的引导方法，其思想基于从贝叶斯规则衍生出来的隐式分类器，它只需要一个条件扩散模型和一个无条件扩散模型，就能生成极高保真度的样本。

3.3.2 改进采样算法

对于当前数值 SDE 求解器需要大量的分数网络来进行评估这一问题，Jolicoeur-Martineau 等[87]设计了一个优化后的 SDE 求解器，其具有自适应步长，并可以逐个为基于分数的生成模型量身定制，且只需要进行两次评分函数评估。基于随机微分方程的生成模型在前向过程中必须花费大量时间才能使最终噪声分布服从高斯分布，基于此问题，Bortoli 等[85]通过解决路径空间上的熵正则化最优传输问题，也叫薛定谔桥问题，来提高前向(后项)生成效率。

Dockhorn 等[84]将扩散模型与统计力学相关联，提出了一种新的临界阻尼 Langevin 扩散模型(CLD)，其主要通过在数据中添加另一个需要学习的速度变量，学习给定数据的速度条件分布函数，会比直接学习数据的分数更加容易，且更容易生成高分辨率图像。

Liu 等[86]把扩散模型过程看作是求解流形中的微分方程，并在计算中发现使用常规数值方法求解反向过程所返回的样本质量较差，伪数值方法的效率反而很高。他们将数值方法分为两部分，即梯度部分和传递部分，目的是最终使传递部分尽可能地接近目标流形。

3.3.3 其他改进方面

针对高维的数据，Deasy 等[91]提出将噪声引入高斯去噪分数匹配，以实现扩散强度的可控性，并通过添加重尾分布来改进分数估计、可控采样收敛以及无条件不平衡数据集的生成性能，从而改进原 SDE 模型在高维数据生成上的表现。

对于当下的扩散模型步骤单元之间的黑盒问题，Karras 等[92]提出将模型分离成相互独立的单元，并且这种拆分对单个单元的改变不会影响到其他单元的状态。Karras 等在这方面主要有两个贡献，一是使用 Heun 的方法作为常微分方程求解器的采样过程，另一个是通过对神经网络的输入及其对应的标签进行预处理，从而训练基于分数的模型。

4 扩散模型的应用

随着扩散模型的发展，可以发现扩散生成模型在表示能力、泛化能力、灵活性方面都有不错的优势，这些优势特征使得扩散模型的身影遍布诸多领域。本章将讨论扩散生成模型在计算机视觉、自然语言处理、时间序列、多模态、跨学科方面的应用。

4.1 计算机视觉

4.1.1 提高图像分辨率

对于当下单图像超分辨率(SISR)所存在的过度平滑、模式崩溃和内存占用问题，SRDiff[93]结合扩散生成模型，利用马尔可夫链将高分辨率图像(HR)转换为潜在的简单分布，然后在反向过程中生成对超高分辨率图像(SR)的预测。在此过程中使用以低分辨率图像(LR)编码器编码的 LR 信息作为条件噪声，逐步对高分辨率图像进行去噪处理。

SR3[18]模型使用迭代细化来提高图像分辨率，解决了图像分辨率提升单程化的缺陷，其主要通过结合 DDPM 模型的随机去噪过程来实现超高分辨率图像的生成。原始噪声图像从纯高斯噪声开始，使用不同的噪声水平去训练 U-Net 模型，从而实现去噪过程的迭代优化。CDM[94]模型将多个扩散模型组成一条流水线，此种级联的方式在不同的空间分辨率上采用不同的生成模型，一种是生成低分辨率数据的基础扩散模型，另一种是生成超高分辨率的 SR3 模型，将图像提高到超高分辨率。

4.1.2 图像合成领域

目前图像合成领域主要是使用 GAN 模型进行训练生成，对于 GAN 所具有的训练不稳定且数据覆盖不全等问题，UNIT-DDPM[95]模型结合 DDPM 模型基于非配对的图到图的任务，引入了元数据域与目标数据域，通过将其中一个域的去噪分数匹配最小化来形成联合分布，并将其分布作为马尔可夫链进行更新，最终通过马尔可夫链蒙特卡洛方法去噪生成最终的样本。Wang 等[96]也将 DDPM 模型应用于语义图像合成领域，他们将噪声图像提供给 U-Net 结构的编码器，

而语义布局则通过多层空间自适应归一化算子提供给解码器，并通过引入无分类器引导的采样策略，进一步提高了采样质量以及语义可解释性。

受到自然语言领域 BART[97] 模型的影响，当前图像生成任务的主要问题是在单一尺度上处理整个图像，Image-BART[19] 模型通过学习反转多项式扩散过程来解决自回归图像合成问题，通过引入情景信息来减小自回归模型的曝光误差，解决自由形式的图像修复而无需特定掩模训练。

4.1.3 多维图像领域

对于 3D 图像的生成领域的研究，Zhou 等[98] 提出了一个形状生成补全的统一框架(PVD)，它能够合成高保真形状，补全部分点云，并从真实物体的单视角深度扫描中生成多个完成结果。Luo 等[99] 提出了一个用于点云生成的概率模型，他们将点云的生成看作学习将噪声分布转换为所需形状分布的反向扩散过程，其模型可以用在点云形状补全、上采样、合成和数据增强等方面(见图 6)。

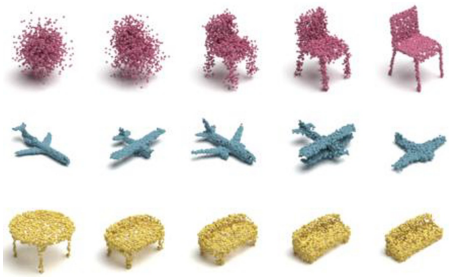

图 6 3D 模型噪声扩散过程[99]

Fig. 6 3D model noise diffusion process[99]

对于同时生成多个 3D 图像方面，Lee 等[100] 使用离散扩散模型来学习场景尺度类别分布，并使用得出的类别分布来表示场景，从而将多个对象分配到对应的语义类别中。

4.2 自然语言处理

扩散模型在计算机视觉领域的广泛应用，使自然语言处理领域的研究者们开始考虑是否可以将去噪扩散模型应用到本领域。但相比图像的连续空间，文本序列之间是离散的。因此，基于此问题，研究者们提出了两种解决思路。

4.2.1 将离散文本映射到连续的表征空间

Difformer[22] 将 Transformer 模型与扩散模型相结合，其中包括额外的锚点损失函数、嵌入层的归一化模块以及额外的高斯噪声因子，以保证将离散数据转为连续数据进行训练。Diffusion-LM[25] 提出了一种新的基于连续扩散的非自回归语言模型，它将高斯噪声向量迭代去噪为单词向量，从而使向量之间产生层次连续的潜在关系。DiffuSeq[23] 提出添加一个 embedding 层，将离散文本映射到连续的表征空间，在反向过程中通过训练模型来寻找近似的文本分布序列。

4.2.2 泛化扩散模型

相比上述将离散文本向连续空间改造的方式，Diffus-ER[24] 模型将重点放在泛化扩散模型上，它的正向加噪过程没有使用原来的高斯噪声，而是将文本的删除、添加、修改视为加噪处理过程，在反向去噪建模过程中，需要学习文本的逆变换来完成目标文本的生成(见图 7)。DiffusionBERT[26] 结合 BERT[101] 模型在训练过程提出了一种新的时间步长和根据每个 token 的信息来控制每一步需要添加噪声程度的调度方案。

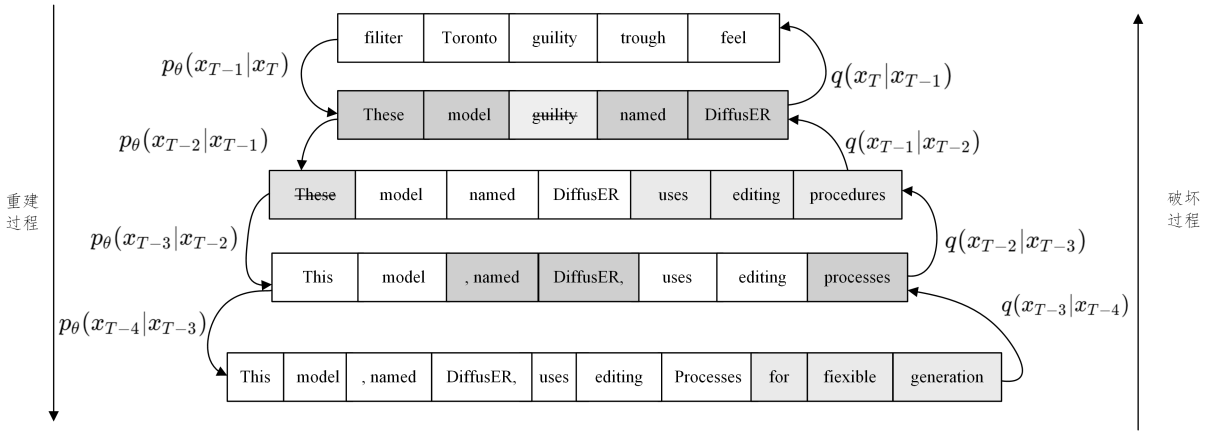

图 7 DiffusER 文本生成过程[24]

Fig. 7 DiffusER's text generation process[24]

4.3 时间序列

对于时间序列原本使用自回归模型进行插补的模式，CSDI[28] 提出使用基于条件分数的扩散模型替换自回归模型来学习条件分布，它将观察数据作为扩散模型的条件输入，使模型利用观察值中的信息进行去噪处理。并在之后的训练过程中使用了一种自监督的方法，使观察值分离为条件信息和插补目标来弥补真值缺少的情况。SSSD[29] 集成了条件扩散模型和结构化状态空间模型[102]，其模型善于捕捉时间序列中的长期依赖关系，在时间序列归并和预测任务中都有很好的表现。

在多元概率时间序列预测方面，TimeGrad[30] 基于能量生成模型，利用 RNN[103] 与扩散模型相结合的方式捕获时间序列特征。在此过程中通过优化数据似然的变分界来学习梯度，并在推理时使用 Langevin 采样，通过马尔可夫链将白噪声转换为感兴趣分布的样本。

4.4 多模态

4.4.1 文本转图像

对于文本转图像领域，使用扩散模型可以将输入的描述性文本生成相互之间毫无关系的物体和形状(见图 8)。VQ-Diffusion[31] 解决了之前生成模型的单项偏差，使用掩蔽机制来避免推理过程中误差的累计。DALLE-2[32] 提出了一种两阶段方法，第一阶段是用 CLIP[104] 图像和文本为条件的嵌入

先验模型，第二阶段是基于扩散模型的解码器，该解码器可以完成图像嵌入工作，进而生成最终图像。Imagen[33]模型由一个用于文本序列的编码器和一个用于生成高分辨率图像的级联扩散模型组成，并且改进了原有的 U-Net 模型来提升效率。

图 8 通过 Imagen 生成“舞者在月亮上跳舞”的图片

Fig. 8 Image of a dancer dancing on the moon generated by Imagen

对于文本到 3D 图像的生成领域，OpenAi 公司提出了 Point-E 模型[34]。该模型首先使用文本到图像的扩散模型生成单个合成视图，然后使用第二个扩散模型生成 3D 点云并对生成的图像进行调节。

4.4.2 文本转语音

在文本转语音领域中，Grad-TTS[35]提出了一种新颖的文本到语音模型，即带有分数的解码器，通过逐渐转换编码器预测的噪声并通过单调对齐搜索与文本输入对齐来生成梅尔频谱图。Diff-TTS[36]通过用噪声增量填充中间表示来解决由双射约束对模型宽度限制所导致的有效容量不足的问题。轻量级扩散模型 ResGrad[37]使用残差作为生成目标来改进原来需要从头到尾合成语音的过程，并将现有的 TTS 模型的推理过程变为即插即用的方式。

4.4.3 文本转视频

Dreamix[105]在推理时使用扩散模型，根据所提供的文本信息，将低分辨率信息与高分辨率信息相结合来进行视频编辑，并通过微调模型的初步阶段，来提高编辑视频的保真度。Tune-A-Video[106]将文本生成视频问题看作生成一些列连续图像的问题，通过提出一种稀疏因果注意力机制将原本的图像生成中的空间自注意力扩展到时空域中，进而完成视频的生成工作。

4.5 跨学科领域

4.5.1 医学图像领域

在医学领域的图像生成任务上，对于从测量数据重建图像的逆问题，有学者[42-43]利用分数的生成模型来重建与先验数据一致的图像。Kim 等[38]提出了由扩散模块和变形模块所组成的模型(DDM)，此模型可以学习源体积和目标体积之间的空间变形信息，并通过生成变化过程的图像来生成 4D(3D 图像加时间)的心脏数据。

在进行医学缺陷检测任务时，可以使用 DDPM 模型来代替原始的自编码器模型[39-41]，以训练健康的图像。并且在推理时，可通过减去原始图像中生成的健康图像样本来检测异常。

4.5.2 分子建模领域

在蛋白质分子建模方面，Anand 等[46]使用扩散生成模型来学习蛋白质旋转和平移等动态的结构信息，进而生成蛋白质的基础结构与序列。ProteinSGM[45]将蛋白质的建模过程表述为图像修复问题，并基于条件扩散的生成方法对蛋白质结构进行精确建模。DiffFolding[44]将蛋白质骨架结构看作一系列连续的角度，用来捕捉组成氨基酸残基的相对方向，结合扩散生成模型由随机未折叠的结构来生成新的稳定折叠结构。

结束语 基于以上论述可以发现，扩散模型相对于传统的生成模型具有一定的优势，并在生成领域有着巨大的潜力，但是还是存在一些问题亟待解决。

1)在正向过程中，扩散模型还是以将原始图像转为全高斯噪声图为目标，这种方式在推理过程中存在多个采样步骤和采样时间长的问题，导致了推理过程的时间成本过高，限制了其应用范围和效果。如何正确且合理地停止前向加噪过程[107]，并在预期时间内收敛到特定的先验分布[85]以及加入自适应机制[87]是当前需要解决的关键问题。

2)扩散模型的生成过程使用了很长的马尔可夫链，从而使得整个生成过程变得相当黑盒化。这种黑盒化的特性使得扩散模型存在着难以捕捉的依赖关系，进而难以直观地理解整个生成过程，从而导致扩散模型的改进和优化受到了很大的限制。针对这一问题，是否可以将整个扩散模型分解为相互独立的单元[92]，从而方便进行白盒化处理，以便更好地理解和优化模型；以及能否优化默认的马尔可夫链[66-67]，使用相对易捕捉易训练的模型加以替代，是当前亟需研究的方向。

3)基于扩散模型改进的衍生模型还是以 DDPM[10]的原始设定来开展工作？在这种情况下，是否可以将扩散模型视为一种广义的模型类型来开展研究。例如，仅基于其采样算法、扩散方案以及构建先验分布等思想进行模型改进工作，而不依赖于 DDPM 的整体思想。这样一来，扩散模型就更容易与其他现有模型相结合，其应用范围也会变得更加广泛，研究意义也将更大。

4)对于扩散模型生成样本的评估指标主要基于 FID 分数，但是此分数无法有效地评价样本使用模型之后的恢复效果，对模型生成样本的多样性也无法评判。因此，是否可以新建一种评估指标来评估上述标准也是未来值得思考的问题。

5)在训练扩散模型时，通常会将证据下界(ELBO)作为训练目标，从而最小化后验分布与先验分布之间的 KL 散度，进而最大化数据对数似然。然而，ELBO 和负对数似然(NLL)能否同时优化的理论研究尚未得到证实，这就导致真实样本和生成后的目标样本之间存在着潜在不匹配的问题。此问题会使训练出的模型在实际应用中表现不佳，从而影响其可靠性和实用性。

综上所述，扩散生成模型未来的研究方向还是优化采样算法、降低模型复杂度、提高采样效率等。对此可以考虑将原本的马尔可夫链蒙特卡洛逐步采样算法转化为其他更有效的算法，如使用哈密顿蒙特卡罗方法(HMC)；并引入预训练模型来初始化模型参数，减少训练步骤，使用更优的超参数来开展模型的训练加快训练过程。以上几个优化方面都值得引起更多的关注与研究。

参 考 文 献

[1] SMOLENSKY P. Information processing in dynamical systems: Foundations of harmony theory[R]. Colorado Univ. at Boulder Dept. of Computer Science, 1986.

[2] HINTON G E,OSINDERO S,TEH Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation,2006,18(7):1527-1554.

[3] HINTON G E,SALAKHUTDINOV R R. A Better Way to Pretrain Deep Boltzmann Machines[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012:2447-2455.

[4] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al. Generative adversarial networks[J]. Communications of the ACM,2020,63(11):139-144

[5] KINGMA D P,WELLING M. Auto-Encoding Variational Bayes[J]. arXiv:1312.6114,2022.

[6] ZHANG M,SUN Y,MCDONAGH S,et al. Flow Based Models For Manifold Data[J]. arXiv:2109.14216,2023.

[7] REZENDE D,MOHAMED S. Variational Inference with Normalizing Flows[C]// Proceedings of the 32nd International Conference on Machine Learning. 2015:1530-1538.

[8] LECUN Y,CHOPRA S,HADSELL R,et al. A Tutorial on Energy-Based Learning[M]//Predicting Structured Data. 2006.

[9] VAN DEN OORD A,VINYALS O,KALCHBRENNER N,et al. Conditional Image Generation with PixelCNN Decoders[C]// Advances in Neural Information Processing Systems. 2016:4797-4805.

[10] HO J,JAIN A,ABBEEL P. Denoising Diffusion Probabilistic Models[C]// Advances in Neural Information Processing Systems. 2020:6840-6851.

[11] CHENG S I,CHEN Y J,CHIU W C,et al. Adaptively-Realistic Image Generation From Stroke and Sketch With Diffusion Model[C]// 2023 IEEE/CVF Winter Conference on Applications of Computer Vision(WACV). 2023:4043-4051.

[12] JOLICOEUR-MARTINEAU A,PICHÉ-TAILLEFER R,COMBES R T DES,et al. Adversarialscore matching and improved sampling for image generation[J]. arXiv:2009.05475,2020.

[13] CHEN T,ZHANG R,HINTON G. Analog Bits:Generating Discrete Data using Diffusion Models with Self-Conditioning[J]. arXiv 2208.04202,2022.

[14] GU Z,CHEN H,XU Z,et al. DiffusionInst:Diffusion Model for Instance Segmentation[J]. arXiv:2212.02773,2022.

[15] XU J,WANG X,CHENG W,et al. Dream3D:Zero-Shot Text-to-3D Synthesis Using 3D Shape Prior and Text-to-Image Diffusion Models[J]. arXiv:2212.14704,2022.

[16] YE M,WU L,LIU Q. First Hitting Diffusion Models for Generating Manifold,Graph and Categorical Data[J]. arXiv:2209.01170,2022.

[17] FURUSAWA C,KITAOKA S,LI M,et al. Generative Probabilistic Image Colorization[J]. arXiv:2109.14518,2021.

[18] SAHARIA C,HO J,CHAN W,et al. Image Super-Resolution Via Iterative Refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,45(4):4713-4726.

[19] ESSER P,ROMBACH R,BLATTMANN A,et al. ImageBART:Bidirectional Context with Multinomial Diffusion for Autoregressive Image Synthesis[C]// Advances in Neural Information Processing Systems. 2021:3518-3532.

[20] BATZOLIS G,STANCZUK J,SCHÖNLIEB C B,et al. Non-Uniform Diffusion Models[J]. arXiv:2207.09786,2022.

[21] LEE S,CHUNG H,KIM J,et al. Progressive Deblurring of Diffusion Models for Coarse-to-Fine Image Synthesis[J]. arXiv:2207.11192,2022.

[22] GAO Z,GUO J,TAN X,et al. Difformer:Empowering Diffusion Models on the Embedding Space for Text Generation[J]. arXiv:2212.09412,2023.

[23] GONG S,LI M,FENG J,et al. DiffuSeq:Sequence to Sequence Text Generation with Diffusion Models[J]. arXiv:2210.08933,2022.

[24] REID M,HELLENDOORN V J,NEUBIG G. DiffusER:Discrete Diffusion via Edit-based Reconstruction[J]. arXiv:2210.16886,2022.

[25] LI X L,THICKSTUN J,GULRAJANI I,et al. Diffusion-LM Improves Controllable Text Generation[J]. arXiv:2205.14217,2022.

[26] HE Z,SUN T,WANG K,et al. DiffusionBERT:Improving Generative Masked Language Models with Diffusion Models[J]. arXiv:2211.15029,2022.

[27] LIN Z,GONG Y,SHEN Y,et al. GENIE:Large Scale Pre-training for Text Generation with Diffusion Model[J]. arXiv:2212.11685,2022.

[28] TASHIRO Y,SONG J,SONG Y,et al. CSDI:Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation[C]// Advances in Neural Information Processing Systems. 2021:24804-24816.

[29] ALCARAZ J M L,STRODTHOFF N. Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models[J]. arXiv:2208.09399,2022.

[30] RASUL K,SEWARD C,SCHUSTER I,et al. Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting[C]// Proceedings of the 38th International Conference on Machine Learning. 2021:8857-8868.

[31] GU S,CHEN D,BAO J,et al. Vector Quantized Diffusion Model for Text-to-Image Synthesis[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2022:10686-10696.

[32] RAMESH A,DHARIWAL P,NICHOL A,et al. Hierarchical Text-Conditional Image Generation with CLIP Latents[J]. arXiv:2204.06125,2022.

[33] SAHARIA C,CHAN W,SAXENA S,et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding[J]. arXiv:2205.11487,2022.

[34] NICHOL A,JUN H,DHARIWAL P,et al. Point-E:A System for Generating 3D Point Clouds from Complex Prompts[J]. arXiv:2212.08751,2022.

[35] POPOV V,VOVK I,GOGORYAN V,et al. Grad-TTS:A Diffusion Probabilistic Model for Text-to-Speech[C]// Proceedings of the 38th International Conference on Machine Learning. 2021:8599-8608.

[36] JEONG M,KIM H,KIM H,et al. Diff-TTS:A Denoising Diffusion Model for Text-to-Speech[J]. arXiv:2014.01409,2021.

[37] CHEN Z,WU Y,LENG Y,et al. ResGrad:Residual Denoising Diffusion Probabilistic Models for Text to Speech[J]. arXiv:2212.14518,2022.

[38] KIM B,YE J C. Diffusion Deformable Model for 4D Temporal

Medical Image Generation[J]. arXiv:2206.13295,2022.
[39] WOLLEB J,BIEDER F,SANDKÜHLER R,et al. Diffusion Models for Medical Anomaly Detection[J]. arXiv:2203.04306,2022.
[40] SANCHEZ P,KASCENAS A,LIU X,el al. What is Healthy? Generative Counterfactual Diffusion for Lesion Localization [C]// Deep Generative Models. Cham:Springer Nature Switzerland. 2022:34-44.
[41] WYATT J,LEACH A,SCHMON S M,et al. AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2022:649-655.
[42] SONG Y,SHEN L,XING L,et al. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models[J]. arXiv:2111.08005,2021.
[43] CHUNG H,YE J C. Score-based diffusion models for accelerated MRI[J]. arXiv:2110.05243,2022.
[44] WU K E,YANG K K,BERG R VAN DEN,et al. Protein structure generation via folding diffusion[J]. arXiv:2209.15611,2022.
[45] LEE J S,KIM P M. ProteinSGM:Score-based generative modeling for de novo protein design[J]. Nature Computational Science,2023,3(5):382-392.
[46] ANAND N,ACHIM T. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models [J]. arXiv:2205.15019,2022.
[47] CROITORU F A,HONDRU V,IONESCU R T,et al. Diffusion Models in Vision:A Survey[J]. arXiv:2209.04747,2022.
[48] YANG L,ZHANG Z,SONG Y,et al. Diffusion Models:A Comprehensive Survey ofMethods and Applications[J]. arXiv:2209.00796,2022.
[49] CAO H,TAN C,GAO Z,et al. A Survey on Generative Diffusion Model[J]. arXiv:2209.02646,2022.
[50] LAM M W Y,LAM M W Y,WANG J,et al. Bilateral Denoising Diffusion Models. [J]. arXiv:2108.11514,2021.
[51] GIANNONE G,NIELSEN D,WINTHER O. Few-Shot Diffusion Models[J]. arXiv:2205.15463,2022.
[52] SONG Y,ERMON S. Improved Techniques for Training Score-Based Generative Models[C]// Advances in Neural Information Processing Systems. 2020:12438-12448.
[53] SONG Y,DURKAN C,MURRAY I,et al. Maximum Likelihood Training of Score-Based Diffusion Models[C]// Advances in Neural Information Processing Systems. 2021:1415-1428.
[54] SONG Y,ERMON S. Generative Modeling by Estimating Gradients ofthe Data Distribution[C]// Advances in Neural Information Processing Systems. 2019.
[55] SONG Y,SOHL-DICKSTEIN J,KINGMA D P,et al. Score-Based Generative Modeling through Stochastic Differential Equations[J]. arXiv:2011.13456,2021.
[56] WILSON J T,MORICONI R,HUTTER F,et al. The reparameterization trick for acquisition functions[J]. arXiv:1712.00424,2017.
[57] SOHL-DICKSTEIN J,WEISS E,MAHESWARANATHAN N,et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics[C]// Proceedings of the 32nd International Conference on Machine Learning. PMLR. 2015:2256-2265.
[58] RONNEBERGER O,FISCHER P,BROX T. U-Net:Convolutional Networks for Biomedical Image Segmentation[J]. arXiv:1505.04597,2015.
[59] WELLING M,TEH Y W. Bayesian Learning via Stochastic Gradient Langevin Dynamics[C]// Proceedings of the 28th International Conference on Machine Learning(ICML-11). 2011:681-688.
[60] VINCENT P. A Connection Between Score Matching and Denoising Autoencoders[J]. Neural Computation,2011,23(7):1661-1674.
[61] SONG Y,GARG S,SHI J,et al. Sliced Score Matching:A Scalable Approach to Density and Score Estimation[C]// Proceedings of The 35th Uncertainty in Artificial Intelligence Conference. 2020:574-584.
[62] NICHOL A Q,DHARIWAL P. Improved Denoising Diffusion Probabilistic Models[C]// Proceedings of the 38th International Conference on Machine Learning. 2021:8162-8171.
[63] KINGMA D,SALIMANS T,POOLE B,et al. Variational Diffusion Models[C]// Advances in Neural Information Processing Systems. 2021:21696-21707.
[64] SAN-ROMAN R,NACHMANI E,WOLF L. Noise Estimation for Generative Diffusion Models[J]. arXiv:2104.02600,2021.
[65] WANG J,LYU Z,LIN D,et al. Guided Diffusion Model for Adversarial Purification[J]. arXiv:2205.14969,2022.
[66] SONG J,MENG C,ERMON S. Denoising Diffusion Implicit Models[J]. arXiv:2010.02502,2020.
[67] ZHANG Q,TAO M,CHEN Y. gDDIM:Generalized denoising diffusion implicit models[J]. arXiv:2206.05564,2022.
[68] SINHA A,SONG J,MENG C,et al. D2C:Diffusion-Denoising Models for Few-shot Conditional Generation. [J]. arXiv:2106.06819,2021.
[69] PEEBLES W,XIE S. Scalable Diffusion Models with Transformers[J]. arXiv:2212.09748,2022.
[70] SEHWAG V,HAZIRBAS C,GORDO A,et al. Generating High Fidelity Data From Low-Density Regions Using Diffusion Models[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2022:11492-11501.
[71] KIM G,JANG W,LEE G,et al. DAG:Depth-Aware Guidance with Denoising Diffusion Probabilistic Models[J]. arXiv:2212.08861,2023.
[72] AUSTIN J,JOHNSON D D,HO J,et al. Structured Denoising Diffusion Models in Discrete State-Spaces[C]// Advances in Neural Information Processing Systems. 2021:17981-17993.
[73] WATSON D,CHAN W,HO J,et al. Learning Fast Samplers for Diffusion Models by Differentiating Through Sample Quality [J]. arXiv:2202.05830,2022.
[74] WATSON D,HO J,NOROUZI M,et al. Learning to efficiently sample from diffusion probabilistic models[J]. arXiv:2106.03802,2021.
[75] XIAO Z,KREIS K,VAHDAT A. Tackling the generative learning trilemma with denoising diffusion GANs[J]. arXiv:2112.07804,2021.

[76] BAO F，LI C，SUN J，et al. Estimating the Optimal Covariance with Imperfect Mean in Diffusion Probabilistic Models[J]. arXiv：2212.08861，2022.

[77] CHUNG H，SIM B，YE J C. Come-Closer-Diffuse-Faster：Accelerating Conditional Diffusion Models for Inverse Problems Through Stochastic Contraction[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022：12413-12422.

[78] VAHDAT A，KREIS K，KAUTZ J. Score-based Generative Modeling in Latent Space[C]// Advances in Neural Information Processing Systems. 2021：11287-11302.

[79] ZHANG L，ZHU X，FENG J. Accelerating Score-based Generative Models with Preconditioned Diffusion Sampling[J]. arXiv：2207.02196，2022.

[80] DU W，YANG T，ZHANG H，et al. A Flexible Diffusion Model [J]. arXiv：2206.10365，2022.

[81] HO J，SALIMANS T. Classifier-Free Diffusion Guidance[J]. arXiv：2207.12598，2022.

[82] ZHANG Q，CHEN Y. Diffusion Normalizing Flow [J]. arXiv：2110.07579，2021.

[83] KIM D，NA B，KWON S J，et al. Maximum Likelihood Training of Implicit Nonlinear Diffusion Models[J]. arXiv：2205.13699，2022.

[84] DOCKHORN T，VAHDAT A，KREIS K. Score-Based Generative Modeling with Critically-Damped Langevin Diffusion[J]. arXiv：2112.07068，2022.

[85] BORTOLI V D，THORNTON J，HENG J，et al. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling[J]. arXiv：2106.01357，2021.

[86] LIU L，REN Y，LIN Z，et al. Pseudo Numerical Methods for Diffusion Models on Manifolds[J]. arXiv：2202.09778，2022.

[87] JOLICOEUR-MARTINEAU A，LI K. Gotta Go Fast When Generating Data with Score-Based Models [J]. arXiv：2105.14080，2021.

[88] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is All you Need[C]// Advances in Neural Information Processing Systems. 2017.

[89] WANG Z，ZHENG H，HE P，et al. Diffusion-GAN：Training GANs with Diffusion[J]. arXiv：2206.02262，2022.

[90] DOSOVITSKIY A，BEYER L，KOLESNIKOV A，et al. An Image is Worth 16x16 Words：Transformers for Image Recognition at Scale[J]. arXiv：2010.11929，2021.

[91] DEASY J，SIMIDJIEVSKI N，LIÒ P. Heavy-tailed denoising score matching[J]. arXiv：2112.09788，2022.

[92] KARRAS T，AITTALA M，AILA T，et al. Elucidating the Design Space of Diffusion-Based Generative Models [J]. arXiv：2206.00364，2022.

[93] LI H，YIFAN Y，CHANG M，et al. SRDiff：Single Image Super-Resolution with Diffusion Probabilistic Models[J]. arXiv：2104.14951，2021.

[94] HO J，SAHARIA C，CHAN W，et al. Cascaded Diffusion Models for High Fidelity Image Generation [J]. arXiv：2106.15282，2021.

[95] SASAKI H，WILLCOCKS C G，BRECKON T P. UNIT-DDPM：UNpaired Image Translation with Denoising Diffusion Probabilistic Models[J]. arXiv：2104.05358，2021.

[96] WANG W，BAO J，ZHOU W，et al. Semantic Image Synthesis via Diffusion Models[J]. arXiv：2207.00050，2022.

[97] LEWIS M，LIU Y，GOYAL N，et al. BART：Denoising Sequence-to-Sequence Pre-training for Natural Language Generation，Translation，and Comprehension[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020：7871-7880.

[98] ZHOU L Q，DU Y L，WU J J，et al. 3D Shape Generation and Completion through Point-Voxel Diffusion [J]. arXiv：：2104.03670，2021.

[99] LUO S T，HU W. Diffusion Probabilistic Models for 3D Point Cloud Generation[J]. arXiv：2103.01458，2021.

[100] LEE J，IM W，LEE S，et al. Diffusion Probabilistic Models for Scene-Scale 3D Categorical Data[J]. arXiv：2301.00527，2023.

[101] DEVLIN J，CHANG MW，LEE K，et al. BERT：Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. Minneapolis，Minnesota：Association for Computational Linguistics，2019：4171-4186.

[102] GU A，GOEL K，RÉ C. Efficiently Modeling Long Sequences with Structured State Spaces[J]. arXiv：2111.00396，2022.

[103] SCHMIDT R M. Recurrent Neural Networks(RNNs)：A gentle Introduction and Overview[J]. arXiv：1912.05911，2019.

[104] RADFORD A，KIM J W，HALLACY C，et al. Learning Transferable Visual Models From Natural Language Supervision[J]. arXiv：2103.00020，2021.

[105] MOLAD E，HORWITZ E，VALEVSKI D，et al. Dreamix：Video Diffusion Models are General Video Editors [J]. arXiv：2302.01329，2023.

[106] WU J Z，GE Y，WANG X，et al. Tune-A-Video：One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation [J]. arXiv：2212.11565，2022.

[107] FRANZESE G，ROSSI S，YANG L，et al. How Much is Enough? A Study onDiffusion Times in Score-based Generative Models[J]. arXiv：2206.05173，2022.

YAN Zhihao，born in 1995，postgraduate. His main research interest is deep learning.

ZHOU Zhangbing，born in 1974，Ph.D，professor，is a member of CCF (No. 28475M). His main research interests include wireless sensor networks，services computing and business process management.

（责任编辑：喻藜）