

面向图像处理逆问题的扩散模型研究综述

王泽龙, 吴宇航*, 李健, 杨轩
(国防科技大学理学院, 湖南长沙 410073)

摘要: 扩散模型是一种新型生成式人工智能模型, 相比生成对抗网络、变分自编码网络、流模型等传统网络, 具有训练稳健、生成保真性与多样性高、数学可解释性强等特点, 在计算机视觉、信号处理、多模态学习等领域应用广泛。扩散模型能够充分学习挖掘训练图像的深度生成先验, 为解决图像处理逆问题提供了一类全新解决范式。为了系统性梳理扩散模型发展现状, 特别是其解决图像处理逆问题的最新进展, 对面向图像处理逆问题的扩散模型研究进行了综述, 阐述了扩散模型的基本原理及其发展现状, 重点介绍了利用扩散模型解决图像处理逆问题的主要技术路线, 以及在该方向的具体应用成果, 并展望了未来研究方向。

关键词: 扩散模型; 生成式人工智能; 图像处理; 逆问题; 深度生成先验

中图分类号: TP181 文献标志码: A 文章编号: 1001-2486(2025)04-091-20



论文
拓展

Review of diffusion models for inverse problems in image processing

WANG Zelong, WU Yuhang*, LI Jian, YANG Xuan

(College of Science, National University of Defense Technology, Changsha 410073, China)

Abstract: Diffusion models represent a novel type of generative artificial intelligence models. Compared to traditional networks such as generative adversarial networks, variational autoencoders, and flow models, diffusion models are characterized by their robust training, high fidelity and diversity in generation, and strong mathematical interpretability, and so they are widely used in fields of computer vision, signal processing, multi-modal learning and so on. Diffusion models are capable of sufficiently learning and exploring the deep generative priors from the training images, providing a novel paradigm for solving inverse problems in image processing. In order to systematically sort out the development status of diffusion model, especially the latest progress in solving the inverse problem of image processing, the research of diffusion model for the inverse problem of image processing was reviewed. The basic principle and development status of diffusion model was expounded, the main technical route of using diffusion model to solve the inverse problem of image processing and some specific application results in this direction were emphatically introduced, and the future research directions were envisioned.

Keywords: diffusion model; generative artificial intelligence; image processing; inverse problem; deep generative prior

生成式人工智能模型(简称生成模型)是一种利用人工智能技术生成文本、图像、视频等内容的深度学习网络, 其通过对训练数据的统计分布进行建模, 能够生成符合该统计分布的新样本。早期的生成模型主要有玻耳兹曼机^[1]、深度信念网络^[2]以及深度玻耳兹曼机^[3], 泛化能力较差; 之后的生成模型如生成对抗网络^[4](generative adversarial networks, GAN)、变分自编码器^[5](variational auto-encoder, VAE)、流模型^[6-7](flow models)、基于能量的模型^[8](energy based model, EBM)、自回归模型^[9]等, 在生成高质量样本方面

取得了显著进步。然而这些传统的生成模型存在着训练过程不稳定、生成样本的保真性和多样性不足等问题。在此背景下, 扩散模型应运而生, 特别是随着 OpenAI 发布的基于扩散模型的文生图模型 GLIDE^[10]以及 Stability AI 发布的开源文生图扩散模型 Stable Diffusion^[11], 扩散模型持续引领生成模型的研究热潮。从数学上讲, 扩散模型包括两个马尔可夫链, 对应着正向加噪过程与反向去噪过程, 前者通过网络训练学习训练图像的统计分布信息, 后者通过迭代推理生成新的图像

收稿日期: 2024-12-13

基金项目: 国防科技大学自主创新基金资助项目(23-ZZCX-JDZ-01)

第一作者: 王泽龙(1985—), 男, 河北博野人, 副教授, 博士, 硕士生导师, E-mail: zelong_wang@163.com

*通信作者: 吴宇航(2001—), 男, 甘肃张掖人, 硕士研究生, E-mail: yuhangwu2023@163.com

引用格式: 王泽龙, 吴宇航, 李健, 等. 面向图像处理逆问题的扩散模型研究综述[J]. 国防科技大学学报, 2025, 47(4): 91-110.

Citation: WANG Z L, WU Y H, LI J, et al. Review of diffusion models for inverse problems in image processing[J]. Journal of National University of Defense Technology, 2025, 47(4): 91-110.

数据,因此扩散模型具有很强的数学可解释性。此外,相比传统生成模型,扩散模型还具有训练稳健、生成内容可控、保真度与多样性高等优点。因此,扩散模型一经提出便在计算机视觉^[12-17]、自然语言处理^[18-21]、时间序列分析^[22-24]、多模态转换^[25-27]等领域以及诸多实际应用^[28-31]中展示出了显著优势。

图像处理逆问题主要源于航天侦察、遥感监测、医学成像、安全监控、智能驾驶等应用领域,包括图像去噪、图像超分辨、图像修复、图像解模糊、图像去雨、图像去雾、图像去马赛克等具体方向^[32-37]。事实上,相位恢复、计算成像、光谱解混、压缩感知等问题也可视为广义图像处理逆问题^[38-41],其都是从观测到的(图像)数据反演推断原始场景信息,都需要有效克服逆问题的病态性才有望获得稳定唯一解。解决图像处理逆问题的传统方法主要有以小波分析为代表的计算调和分析法^[42]、偏微分方程方法^[43]、基于统计的方法^[44]和变分推断方法^[45]。这些传统方法主要依赖模型驱动的先验信息,表征能力有限,深度学习为提取数据驱动的先验信息提供了方法途径。扩散模型作为深度学习中新型的生成模型,在解决图像处理逆问题时极具应用潜力^[46]。扩散模型能够充分学习挖掘训练图像中的深度生成先验,可为解决图像处理逆问题提供更加有效的先验约束,进而克服图像处理逆问题的病态性、显著提升图像处理质量。

目前,国内外只有较少的综述^[47-51]总结了扩散模型的起源进展,尚未检索到专门面向图像处理逆问题的扩散模型综述。因此,本文从扩散模型的基本原理出发,系统梳理总结了扩散模型的发展历程和最新进展,详细探讨了利用扩散模型解决图像处理逆问题的主要技术途径,并介绍了本团队基于扩散模型解决几类典型图像处理逆问题的基本思路和具体成果,最后探讨了扩散模型在图像处理逆问题领域中亟须解决的问题和未来研究方向。

1 扩散模型

1.1 扩散模型的基本原理

尽管早在 2015 年已经提出了扩散模型的理论构想,但目前学术界普遍将 2020 年 Ho 等提出的去噪扩散概率模型^[52](denoising diffusion probabilistic models, DDPM)视为扩散模型的起点,其灵感来自非平衡热力学^[7],核心思想是模拟热力学中的扩散过程,将结构化的数据逐步转

换成无结构的噪声数据,从而学习数据分布特性,然后通过逆向推理将噪声数据再转化生成服从原数据分布的新数据。扩散模型最早是以离散形式提出来的,后由 Song 等^[53]推广到了连续形式,这些都是无条件生成扩散模型,难以对生成内容进行控制,学者们随即提出了用于可控生成的条件扩散模型。下面针对上述几类典型形式的扩散模型分别进行阐述。

1.1.1 离散形式扩散模型

扩散模型的离散形式同样包括正向扩散与反向去噪两个过程。正向扩散过程首先产生一系列服从高斯分布的噪声数据,并将其逐步地添加到图像数据中,进而得到一个马尔可夫链:

$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_t \quad t = 1, 2, \dots, T \quad (1)$$

式中: $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 为服从标准高斯分布的噪声数据; $\alpha_t \in (0, 1)$ 为添加噪声强度相关的超参数,满足 $0 < \alpha_T < \alpha_{T-1} < \dots < \alpha_1 < 1$; \mathbf{x}_0 为初始图像数据; T 为正向过程总步数。据此可知,

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t} \mathbf{x}_{t-1}, (1 - \alpha_t) \mathbf{I}) \quad (2)$$

表明任意 t 时刻的图像数据 \mathbf{x}_t 只依赖于 $t - 1$ 时刻的图像数据 \mathbf{x}_{t-1} 。进一步利用重参数化技巧^[54]可以将式(1)与式(2)分别重写为:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon} \quad t = 1, 2, \dots, T \quad (3)$$

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (4)$$

其中, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 。由上式可知,当 $t \rightarrow T$ 时, $q(\mathbf{x}_t | \mathbf{x}_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I})$,即正向加噪过程将图像数据逐步退化成噪声数据。

反向去噪过程从服从高斯分布的噪声数据 $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 出发,通过

$$p(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}(\mathbf{x}_t, t), \boldsymbol{\Sigma}(\mathbf{x}_t, t)) \quad (5)$$

逐步实现对图像数据的去噪。其中,对 $p(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 的估计采用深度神经网络

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \quad (6)$$

实现, θ 为网络参数,其训练过程采用的损失函数是 Sohl-Dickstein 等^[55]基于变分推断得到的证据下界(evidence lower bound, ELBO),等价于优化

$$\text{loss} = \sum_{i>1} KL(q(\mathbf{x}_{i-1} | \mathbf{x}_i, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{i-1} | \mathbf{x}_i)) \quad (7)$$

其中, $KL(\cdot)$ 表示 Kullback-Leibler 散度(KL 散度)。损失函数的直观解释是在每个 t 时刻,要使 $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ 尽可能地接近正向过程中的真实后验概率 $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)$ 。考虑到 $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)$ 的方

差是一个仅与正向扩散过程参数有关的常数,因此 $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ 的协方差 $\Sigma_\theta(\mathbf{x}_t, t)$ 也可设置为与正向扩散过程参数有关的常数。此时损失函数可转化为关于均值 $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ 的函数,而由贝叶斯公式可知,均值可进一步表示成

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t) \right) \quad (8)$$

因此,训练损失函数进一步转化为关于噪声网络 $\boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t)$ 的函数,即

$$\mathbb{E}_{t \sim \mathcal{U}(1, T), \mathbf{x}_0 \sim p(\mathbf{x}_0), \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} (\|\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t)\|_2^2) \quad (9)$$

式中, \mathbb{E} 为期望算子, $\mathcal{U}(0, 1)$ 为均匀分布,噪声网络经常选择 U-Net^[56] 网络。网络参数经上述损失函数训练好后,可由式(6)迭代生成新数据。

1.1.2 连续形式扩散模型

DDPM 的正向扩散与反向去噪过程都设定成具有固定步数的离散过程,事实上其可以视为连续过程的离散采样,连续过程可用随机微分方程(stochastic differential equation, SDE)来描述^[53],其正向扩散过程可以表示为

$$d\mathbf{x} = f(\mathbf{x}, t) dt + g(t) d\mathbf{w} \quad (10)$$

式中,随机微分方程的漂移系数 $f(\mathbf{x}, t)$ 和扩散系数 $g(t)$ 分别决定了确定性变化过程与不确定性变化过程, \mathbf{w} 表示标准的布朗运动, $d\mathbf{w}$ 可以视为无穷小的高斯白噪声。与式(10)相对应的反向去噪过程为

$$d\mathbf{x} = [f(\mathbf{x}, t) - g^2(t) \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t) d\bar{\mathbf{w}} \quad (11)$$

式中: dt 为负无穷小的时间步长, $d\bar{\mathbf{w}}$ 表示时间反向的布朗运动; $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 为评分函数,可由一个神经网络 $s_\theta(\mathbf{x}_t, t)$ 进行估计,训练的损失函数为

$$\mathbb{E}_{t, p(\mathbf{x}_0), p_t(\mathbf{x}_t | \mathbf{x}_0)} (\lambda(t) \|s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t | \mathbf{x}_0)\|_2^2) \quad (12)$$

$\lambda(t)$ 为加权函数, t 服从 0 到 T 的均匀分布。

类似地,在正向扩散过程中通过优化式(12)中的损失函数,可以训练 $s_\theta(\mathbf{x}_t, t)$ 的网络参数;在反向去噪过程中,通过 $s_\theta(\mathbf{x}_t, t)$ 估计评分函数 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$,进而利用式(10)生成新数据。连续形式扩散模型的意义在于为扩散模型提供了一个数学上的统一视角,可以借助 SDE 相关的数学工具分析扩散模型性质,据此可进一步对扩散模型进行改进和优化。

1.1.3 条件扩散模型

扩散模型无论是离散形式还是连续形式,在生成图片时都是从高斯噪声数据出发,利用反向

去噪过程逐步生成与训练图像数据统计分布一致的新图片数据。由于初始噪声数据具有随机性,且反向去噪过程中每步都施加随机噪声,因此扩散模型每次生成的图像内容大概率是不同的,即生成的图像具有高多样性。该特性适用于数据增强等实际任务,但解决图像处理逆问题需要生成的图像逼近原始图像,显然扩散模型的不可控生成难以胜任该任务。

针对扩散模型的可控生成,专家学者提出了条件扩散模型。假设引入条件 y 来控制扩散模型的生成过程,则反向去噪过程的迭代估计依赖条件概率分布 $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, y)$ 。Dhariwal 等^[57]证明了增加条件 y 对正向扩散过程没有影响,即条件扩散模型的正向扩散过程与(无条件)扩散模型的正向去噪过程完全一样,引入条件只影响反向去噪过程。据此, Dhariwal 等提出了一种利用图片类别标签引导图像生成的方法,即 Classifier-Guidance,这大幅度提升了扩散模型生成相应类别图像的质量,并在评价生成图像质量的启发式分数^[58](inception score, IS)和弗雷歇距离^[59](Fréchet inception distance, FID)两个指标上优于传统生成模型。具体而言,上述方法从连续形式扩散模型出发,引入条件 y 后,反向去噪过程(11)中评分函数 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 变为条件评分函数 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x} | y)$,而后者依据贝叶斯公式可写为

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x} | y) = \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \nabla_{\mathbf{x}} \log p_t(y | \mathbf{x}) \quad (13)$$

式中,第一项 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 称为先验评分函数,就是无条件扩散模型本身的评分函数;第二项 $\nabla_{\mathbf{x}} \log p_t(y | \mathbf{x})$ 称为似然评分函数,其中 $p_t(y | \mathbf{x})$ 可代表一个分类器, \mathbf{x} 与 y 分别是分类器的输入与类别标签。分类器 $p_t(y | \mathbf{x})$ 经独立训练后,在可控生成时对 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x} | y)$ 的估计可表示为

$$s_\theta(\mathbf{x}_t, t, y) = s_\theta(\mathbf{x}_t, t) + \nabla_{\mathbf{x}} \log p_t(y | \mathbf{x}) \quad (14)$$

即条件扩散模型可视为由无条件扩散模型经过分类器引导修正后得到。

相比 Classifier-Guidance 方法需要独立训练分类器实现条件引导,谷歌团队在 2022 年提出了基于 Classifier-Free^[60]的条件扩散模型,其特点是避免使用分类器,而是直接训练 $s_\theta(\mathbf{x}_t, t, y)$ 来估计条件评分函数 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x} | y)$ 。Classifier-Free 方法对于生成结果的控制更精确,生成质量更好,但对于任给的不同条件都需要重新训练神经网络,故训练成本太高。

上述条件扩散模型是探索条件引导下的可控

扩散模型先驱,为后续应用扩散模型解决图像处理逆问题提供了参考借鉴。

1.2 扩散模型的优化提升

尽管扩散模型一经提出便引起了广泛关注,但在实际应用中仍面临着瓶颈问题,例如经典扩散模型 DDPM 是基于马尔可夫链的,正反过程都需要大量迭代,训练成本和推理复杂度高,同时生成质量以及适用条件距离实际应用还有一定差距,因此需要在生成质量与生成效率两个方面探索扩散模型的优化提升方法。目前图像质量评价主流指标有:IS^[58]、FID^[59]、Recall(召回率)^[61]、负对数似然(negative log likelihood, NLL)、核心初始距离^[62](kernel inception distance, KID),以及用于表征正确分类的对抗样本数量与对抗样本总数的 Robust 指标等。

1.2.1 噪声优化

高斯噪声是扩散模型涉及的主要噪声类型,优化噪声可以提升扩散模型效能,使得生成结果质量提升。针对 DDPM, Nichol 等^[63]在正向扩散过程中添加余弦噪声,获得了更好的对数似然性,并且通过为反向去噪过程添加可学习的方差实现了对生成步骤的优化。而 Kingma 等^[64]在预测噪声时使用了傅里叶特征,并通过分析扩散模型的变分下限发现扩散损失只受信噪比函数极值的影响。San-Roman 等^[65]使用 VGG-11 卷积神经网络实现对噪声参数的动态调整,生成最优的噪声参数,可生成具有更低 FID 的图像样本。此外, Wang 等提出的提纯导向扩散模型^[66](guided diffusion model for purification, GDMP)将提纯过程嵌入反向去噪过程,通过选择合适的扩散时间步长,使高斯噪声在抵消对抗性扰动的同时能够保留图像的主要内容。

上述噪声优化方法的主要性能评估如表 1 所示,噪声优化方法主要目的在于降低生成图像的 FID 指标,同时有助于提升后续目标检测分类等任务,例如 GDMP 方法重点在于保证模型在面对对抗样本时仍然能够做出正确的预测。

1.2.2 马尔可夫链优化

去噪扩散隐式模型^[67](denoising diffusion implicit models, DDIM)是一种基于非马尔可夫过程的扩散模型,能够在保持训练损失函数不变的前提下实现生成推理的加速。DDIM 可以使用具有任意步骤的正向扩散过程来训练神经网络,但在反向去噪过程中只从其中某些步骤进行迭代推理(采样),这便通过减少采样步数加速了扩散模

表 1 噪声优化性能评估

Tab. 1 Performance evaluation of noise optimization

方法	数据集	指标
Nichol 等 ^[63]	ImageNet 64 × 64	FID: 2.92 Recall: 0.71
Kingma 等 ^[64]	ImageNet 64 × 64	FID: 3.40
San-Roman 等 ^[65]	LSUN Church 256 × 256	FID: 54.09
Wang 等 ^[66]	ImageNet 256 × 256	Robust: 70.94

型的反向去噪过程,且通过对采样步骤和方差的控制优化显著降低了对生成图像质量的影响。基于 DDIM, Zhang 等提出了广义去噪扩散隐式模型^[68](generalized denoising diffusion implicit models, gDDIM),其从数值计算的角度出发,在求解随机微分方程时使用评分函数的特定近似值来获得 DDIM,并证明了使用确定性的采样方案相比随机采样方案会更快地实现生成推理。

上述马尔可夫链优化方法的主要性能评估如表 2 所示,其中 Steps 指标表示采样步数。马尔可夫链优化主要目的是在有效缩减采样步数的同时,保持甚至提高图像生成质量。

表 2 马尔可夫链优化性能评估

Tab. 2 Performance evaluation of Markov chain optimization

方法	数据集	指标
Song 等 ^[67]	CelebA 64 × 64	Steps: 10 FID: 17.33
Zhang 等 ^[68]	CIFAR-10 64 × 64	Steps: 10 FID: 2.25

1.2.3 采样算法优化

采样算法优化可以提升扩散模型的生成质量以及效率。在提升生成质量方面, Song 等^[69]通过使用依赖噪声生成尺度的采样策略,提高了扩散模型在高分辨率图像生成的稳定性; Jolicoeur-Martineau 等^[13]提出了一致性采样方案,结合由评分函数和对抗式目标函数组成的混合目标函数,提升了扩散模型的生成图像质量; Liu 等^[70]从求解流形中的微分方程视角出发,通过计算发现使用伪数值方法生成的样本质量高于常规数值方法。在提升生成效率方面, Jolicoeur-Martineau 等提出了一种具有自适应步长的高效 SDE 求解器^[71],其仅需要进行两次评分函数的评估,不依赖于负样本,比传统 Euler-Maruyama 求解器快 2 ~ 10 倍; Bortoli 等^[72]通过解决路径空间上的熵正则

化最优传输问题,提高扩散模型的生成效率。Dockhorn 等将扩散模型与统计力学相关联,提出了一种临界阻尼 Langevin 扩散模型^[73] (critically-damped Langevin diffusion, CLD),其在模型中添加一个需要学习的速度变量,进而从训练数据中学习以该速度为条件的条件评分函数,相较于直接学习训练数据的评分函数,在降低学习成本的同时提升了扩散模型生成高分辨率图像的能力。

上述采样算法优化方法的主要性能评估如表3所示。采样算法优化使得扩散模型的求解效率有所提高,同时较为显著地提升了图像的生成质量。

表3 采样算法优化性能评估

Tab.3 Performance evaluation of sampling algorithm optimization

方法	数据集	指标
Song 等 ^[69]	CelebA 64 × 64	FID: 10.23
Liu 等 ^[70]	CelebA 64 × 64	Steps: 10 FID: 17.3
Jolicoeur-Martineau 等 ^[71]	FFHQ 256 × 256	FID: 15.68
Bortoli 等 ^[72]	CelebA 32 × 32	Steps: 40 FID: 5.0
Dockhorn 等 ^[73]	CIFAR-10 64 × 64	FID: 2.25

1.2.4 多模型结合

在扩散模型与传统生成模型结合方面,Sinha 等^[74]提出了具备对比表征的扩散解码模型,通过学习扩散先验分布来改进生成质量,再使用对比自监督学习来提高表征质量,最终实现少快拍条件图像生成;Peebles 等^[75]结合 Transformers 模型^[76]将反向去噪过程中的神经网络由传统的 U-Net 网络替换为 Transformers 网络,并发现当提高网络深度广度或者增加字符时,可以获得更好的训练效果;Wang 等将 GAN 模型与扩散模型相结合^[77],提高了生成样本的稳定性。此外,Zhang 等^[78]在微分方程的基础上,将标准化流^[6]与 SDE 相结合,通过联合训练正向与反向神经网络,最小化关于二者差异的共同成本函数,生成效果得到显著提升;Kim 等^[79]也采用了类似的思路,在 SDE 模型基础上提出了一种非线性扩散模型,融合了可训练的标准化流与扩散过程,通过流网络在隐空间中进行线性扩散来学习噪声分布,再回

到原空间实现非线性扩散。

上述多模型结合方法的主要性能评估如表4所示,可以看出,多模型结合方法通过扩散模型与传统生成模型的结合,实现了在多种数据集上的生成图像视觉质量的提升。

表4 多模型结合性能评估

Tab.4 Performance evaluation of multi-model combination

方法	数据集	指标
Sinha 等 ^[74]	FFHQ 256 × 256	FID: 13.04
Peebles 等 ^[75]	ImageNet 256 × 256	FID: 2.27 IS: 278.24
Wang 等 ^[77]	FFHQ 1 024 × 1 024	FID: 2.83 Recall: 0.49
Zhang 等 ^[78]	CIFAR-10 64 × 64	NLL: 3.04 FID: 14.14
Kim 等 ^[79]	CIFAR-10 64 × 64	NLL: 3.01 FID: 3.25

1.2.5 其他方面优化

Watson 等^[80]在反向去噪过程中融入了动态规划算法,将 ELBO 分解为单独的 KL 散度项,通过最小化 KL 散度来最大化 ELBO,优化传统的推理路径,提高了推理效率。Xiao 等^[81]在反向去噪过程中利用 GAN 区分真实样本和生成样本,从而提高了最小化 KL 散度的推理效率。Bao 等^[82]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^[83]认为从高斯噪声开始反向去噪完全没有必要,其使用随机差分方程的收缩理论优化了反向去噪过程中的初始化图像,显著减少了反向去噪过程中的步数,提高了生成效率。针对高维数据,Deasy 等^[84]提出将广义高斯噪声引入扩散模型,以提升评分函数估计精度、可控生成的收敛性以及在不平衡数据集上的生成性能,从而改进传统扩散模型在高维数据生成上的表现。Karras 等^[85]认为基于扩散模型的理论和实践目前过于复杂,提出了一个清晰分离具体设计选择的设计空间来解决这个问题,据此改进了训练过程和采样过程中网络架构、训练求解器、参数选择等,实现了生成性能的显著提升。

上述其他方面优化方法的主要性能评估如表5所示,可以看出其他方面的优化方法同样实现了扩散模型在各类数据集上的生成质量提升,但是大部分方法在高维数据生成上并未展示出实际效果。

表 5 其他方面优化性能评估

Tab. 5 Performance evaluation of other optimization

方法	数据集	指标
Watson 等 ^[80]	ImageNet 64 × 64	NLL: 3.55
		IS: 9.63
Xiao 等 ^[81]	CIFAR - 10 64 × 64	FID: 3.75
		Recall: 0.57
Bao 等 ^[82]	CelebA 64 × 64	FID: 5.96
Chung 等 ^[83]	AFHQ - 256	FID: 15.53
Deasy 等 ^[84]	CelebA 64 × 64	IS: 2.111
		KID: 0.129
Karras 等 ^[85]	AFHQv2 64 × 64	FID: 1.96

最后梳理总结扩散模型的优化提升方法。首先,上述优化提升方法不论是去噪过程、采样算法,还是改变神经网络构型、优化损失函数,都是在扩散模型基本框架下实施的,都未突破扩散模型的基本原理,这便于在保持扩散模型固有优势前提下提升模型性能。其次,上述优化改进确实能够提升模型性能,主要表现于生成质量与生成效率两个方面。生成质量的提升主要表现在 IS、FID 等指标及视觉效果上,生成效率的提升表现在采样过程的步数更精简、用时更少。目前,改进的扩散模型已在多种实际应用中取得实效,例如 DDIM 已经被广泛应用于加速模型生成过程。上述优化提升方法的主要创新点、优势及不足在表 6 中列出,可供读者参考。

表 6 扩散模型优化提升典型方法综述

Tab. 6 Review on typical methods for optimization and improvement of diffusion model

优化方向	方法	创新点及优势	不足
噪声优化	Nichol 等 ^[63]	正向过程添加余弦噪声,获得更优对数似然	训练数据集分辨率低,高维数据生成能力未经验证
	Kingma 等 ^[64]	网络添加傅里叶特征,实现方差最小化	训练数据集分辨率低,高维数据生成能力未经验证
	San-Roman 等 ^[65]	自适应动态调整噪声参数,无须手动调整即可获得最优结果	调整步骤耗时较长
	Wang 等 ^[66]	去噪过程中加入净化机制,同时去除扰动噪声和高斯噪声	额外的净化过程导致耗时较长
马尔可夫链优化	Song 等 ^[67]	使用非马尔可夫正向过程,减少了抽样步骤,实现采样加速	生成质量有损失
	Zhang 等 ^[68]	使用分数近似值进行快速抽样,扩展了 DDIM 的应用范围	生成质量有损失
采样算法优化	Song 等 ^[69]	使用近似最大似然,增强了生成高维数据的能力	训练集有偏差
	Jolicoeur-Martineau 等 ^[13]	提出一致性采样方案和混合训练方案,提高生成样本的质量	训练数据集分辨率低,高维数据生成能力未经验证
	Dockhorn 等 ^[73]	引入速度变量,实现更高质量的采样,且可拓展至高维数据	训练成本高
多模型结合	Sinha 等 ^[74]	引入对比自监督方法,适用于小样本条件	计算效率较低
	Peebles 等 ^[75]	引入 Transformers 学习噪声,具有良好的可扩展性	训练成本高,GFLOPS 越高效果越好
	Wang 等 ^[77]	基于 GAN 的架构结合扩散模型,实现更逼真的生成结果	本质上是一类新型 GAN 框架,对扩散探索不足
	Zhang 等 ^[78]	结合标准化流模型,减少离散化步骤,采样效率更高	需额外学习正向过程,训练效率低
	Kim 等 ^[79]	提出非线性扩散模型,生成结果更接近最大似然估计	需要额外训练流网络构建非线性扩散模型,训练评估耗时更长

续表

优化方向	方法	创新点及优势	不足
其他方面	Watson 等 ^[80]	引入动态规划算法,自适应地为不同扩散模型提供最佳采样步数	提供的采样步数在 ELBO 上最佳,但不一定能获得最优的生成质量
	Xiao 等 ^[81]	使用 GAN 辅助最小化 KL 散度,减少去噪步数,提高生成效率	由于 GAN 模型的固有问题,鲁棒性不足
	Bao 等 ^[82]	优化协方差以提高计算效率,给出了协方差的解析形式	训练数据集分辨率低,高维数据生成能力未经验证
	Chung 等 ^[83]	提出更优的正向过程初始化,显著减少采样步数	初始化选择缺乏明确标准,鲁棒性不足
	Deasy 等 ^[84]	引入广义高斯噪声,有效提升了评分函数估计精度、可控生成的收敛性等	训练数据集分辨率低,高维数据生成能力未经验证
	Karras 等 ^[85]	提供一个明确区分具体设计选择的设计空间,改进训练与采样过程,降低模型复杂度	训练数据集分辨率低,高维数据生成能力未经验证,同时训练成本较高

2 基于扩散模型求解图像处理逆问题的主要技术路线

图像处理逆问题一般都对应明确的正过程模型,用于描述原始图像与观测图像之间的关系,该正过程模型可表示为

$$y = f(x) + n \quad (15)$$

其中 y 为观测图像, $f(\cdot)$ 为图像退化函数, x 为原始图像, n 为观测噪声。当退化函数为线性函数 $f(x) = Ax$ 时,式(14)可重写为

$$y = Ax + n \quad (16)$$

后续若无特殊说明,主要考虑线性模型(16)。求解图像处理逆问题即从观测图像中恢复重构原始图像,目前应用扩散模型求解该逆问题的主流技术路线有空间分解、后验采样、隐式扩散模型和知识蒸馏等,其中空间分解与后验采样主要考虑如何以观测图像为条件实现对扩散模型的可控生成,隐式扩散模型和知识蒸馏主要考虑如何加快可控生成的速度。

除 1.2 节中所涉及的评价图像生成质量的指标外,针对图像处理逆问题的主流评价指标还有:峰值信噪比^[86] (peak signal-to-noise ratio, PSNR)、结构相似度^[87] (structural similarity, SSIM)、学习感知图像块相似度^[88] (learned perceptual image patch similarity, LPIPS)、融合视觉信息保真度^[89] (visual information fidelity for fusion, VIFF)、互信息量^[90] (mutual information, MI) 和交叉熵^[91] (cross entropy, CE)^[91] 等。

2.1 空间分解

空间分解(又称空间投影)是第一个利用扩散模型解决图像处理线性逆问题的重要思路^[92-96],其从几何的视角审查了扩散模型与图像处理逆问题的关系,其中较为典型的模型有去噪扩散零空间模型^[93] (denoising diffusion null-space model, DDNM)、去噪扩散复原模型^[94] (denoising diffusion restoration models, DDRM) 与流形约束梯度^[95] (manifold constrained gradient, MCG) 等。

2.1.1 DDNM

DDNM 的基本原理是利用线性代数中的零域值域分解(range-null space decomposition, RNSD)将图像空间分解为零域与值域两部分,其中值域部分可以通过退化算子和观测图像直接得到,而零域部分则是通过训练好的扩散模型来重建,即

$$\hat{x}_{0|t} = A^\dagger y + (I - A^\dagger A)x_{0|t} \quad (17)$$

式中, A 为图像退化算子, A^\dagger 为其伪逆, $x_{0|t}$ 为扩散模型对原始图像的估计。DDNM 基本原理如图 1 所示。DDNM 确保了重构图像与原始图像的数据一致性,同时通过扩散模型的去噪能力,有效地处理了图像中的噪声问题。

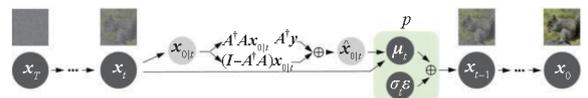


图1 DDNM 结构图^[93]

Fig.1 DDNM structure illustration^[93]

此外该方法还具有以下两个优势:一是该方

法只需要一个训练好的扩散模型,通过数学上的零域值域分解,就能够实现多种图像处理任务,因此在内存和时间消耗上与原扩散模型基本相同;二是该方法无须针对特定任务进行训练或优化,极大地简化解解决图像处理逆问题的过程,例如其能够处理从 2 倍到 256 倍的超分辨率任务、采样率从 100% 到 0.1% 的压缩感知问题,以及任意形状的图像修复任务。

2.1.2 DDRM

DDRM 利用扩散模型和流形约束,以无监督的方式高效地解决多种线性逆问题,基本原理如图 2 所示。DDRM 的基本思想是将逆问题建模为从观测图像的后验分布中进行采样的问题,继而采用了一种不同于变分推断的方法来学习该后验分布。此外,DDRM 引入了由数据逼近项的梯度构成的校正项,通过投影到数据流形的切空间实现流形约束。这样可以有效平衡逆问题的数据逼近项与先验约束项,从而提高图像重构的准确性和稳定性。

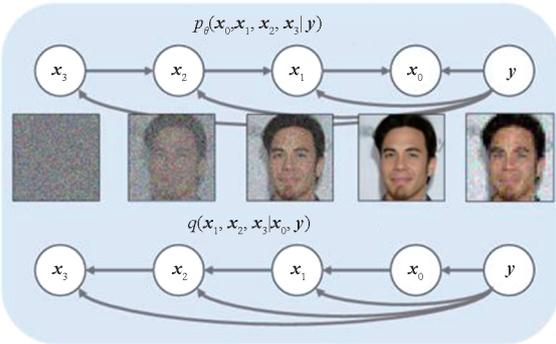


图 2 DDRM 结构图^[94]

Fig. 2 DDRM structure illustration^[94]

DDRM 的优势在于以下几点:①可以用于解决多种线性逆问题,而不需要针对特定问题进行额外训练;②可以无监督地学习后验分布,无须使用额外的训练数据;③流形约束可以提高重建的准确性和稳定性,减少噪声的影响。DDRM 的主要不足在于迭代过程需要进行奇异值分解(singular value decomposition, SVD),对算力要求较高。

2.1.3 MCG

传统扩散模型在解决图像处理逆问题时,通常采用递归迭代的方式进行反向去噪和投影,容易导致样本路径偏离数据流形,从而累积重构误差,影响重构质量。为解决该问题,Chung 等^[95]从几何视角解释了扩散模型和数据流形之间的关系,提出了基于 MCG 的图像处理逆问题求解方法,显著提高了图像重构的准确性和稳定性。

MCG 的核心思想是通过引入流形约束来引导扩散过程,使其更接近数据流形,从而提高图像重构性能。MCG 用于图像修复的流程如图 3 所示,实验结果表明,MCG 在图像修复、图像着色和计算机断层扫描重构等任务中取得了优异性能。

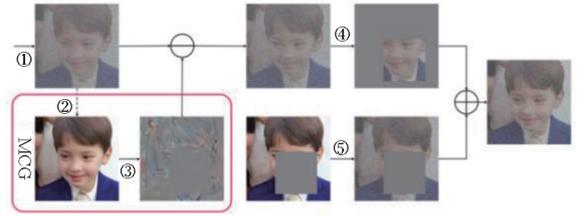


图 3 MCG 用于图像修复示意图^[95]

Fig. 3 Structure illustration of MCG used for image inpainting^[95]

2.2 后验采样

后验采样是利用扩散模型解决图像处理逆问题的另一重要思路,其从统计的视角出发,通过估计条件评分函数实现条件扩散模型的反向去噪生成。依据不同的估计方法,后验采样又包括多种具体实现形式,例如扩散后验采样^[97](diffusion posterior sampling, DPS)、变分扩散采样^[98](variational diffusion sampling, VDS)及其多种改进^[97-101]。

2.2.1 DPS

DPS 由 Chung 等于 2023 年提出,该工作从条件扩散模型出发,通过观测图像的控制,引导反向去噪过程朝着符合数据一致性的方向演化,其基本原理如图 4 所示。具体而言,条件扩散模型的条件引导主要依赖于式(13),其中先验评分函数可以由无条件扩散模型获得,而似然评分函数依赖对 \$p_t(y | x)\$ 的估计。利用 Tweedie 公式,由反向去噪过程中的噪声图像可以估计原始图像的后验均值,即

$$\hat{x}_0 \doteq \mathbb{E}[x_0 | x_t] = \frac{1}{\sqrt{\alpha_t}} [x_t + (1 - \bar{\alpha}_t) \nabla_{x_t} \log p_t(x_t)] \quad (18)$$

因此近似成立 \$p_t(y | x) = p(y | \hat{x}_0)\$;再结合正

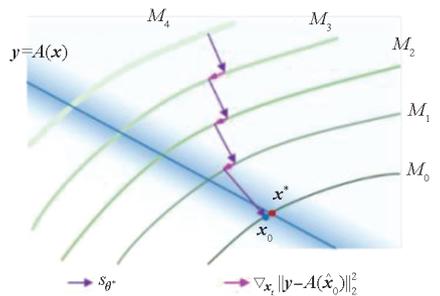


图 4 DPS 几何概念图^[97]

Fig. 4 Conceptual illustration of the geometries of DPS^[97]

过程模型(15)便可求得 $p_t(\mathbf{y}|\mathbf{x})$ 的解析表达式。可以发现,似然评分函数融合了扩散模型与逆问题的正过程模型,实现了先验信息与观测数据的有机结合。

DPS 针对逆问题的正过程模型中的高斯噪声与泊松噪声,都给出了具体的似然评分函数形式,拓展了扩散模型在不同噪声类型下的线性图像处理逆问题中的应用;同时 DPS 还可以处理非线性逆问题,只要其退化函数能够通过自动微分获得梯度,DPS 就可以在同一框架下有效求解非线性图像处理逆问题。此外, DPS 避免了 DDRM 涉及的奇异值分解,计算复杂度相对较低,一定程度上提高了计算效率。

2.2.2 VDS

DPS 在似然评分函数计算中需要利用后验均值实现对原始图像的估计,这导致在反向去噪过

程中容易累计估计误差。为了解决该问题, Mardani 等提出了 VDS,其基本原理如图 5 所示。该方法由扩散模型引出原始图像的先验分布,并将观测图像表示为原始图像的似然函数,进而使用变分推理来估计给定观测图像后原始图像的后验分布。具体实现上, VDS 通过极小化 KL 散度实现对原始图像后验分布的估计,进而在变分分布的方差接近于零的条件下,将 KL 散度极小化转化为优化问题:

$$\min_{\boldsymbol{\mu}} \|\mathbf{y} - f(\boldsymbol{\mu})\|_2^2 + \mathbb{E}_{t, \varepsilon} (\lambda(t) \|\varepsilon_{\theta}(\mathbf{x}_t; t) - \varepsilon\|_2^2) \quad (19)$$

式中, $\boldsymbol{\mu}$ 为待估原始图像, $\lambda(t)$ 为关于时间的系数函数, ε_{θ} 为噪声估计网络。可以看出,相比 DPS 从后验近似角度的优化步骤, VDS 使用变分优化,不再受扩散模型需要按时间倒序反向去噪的约束,能够加速推理生成。

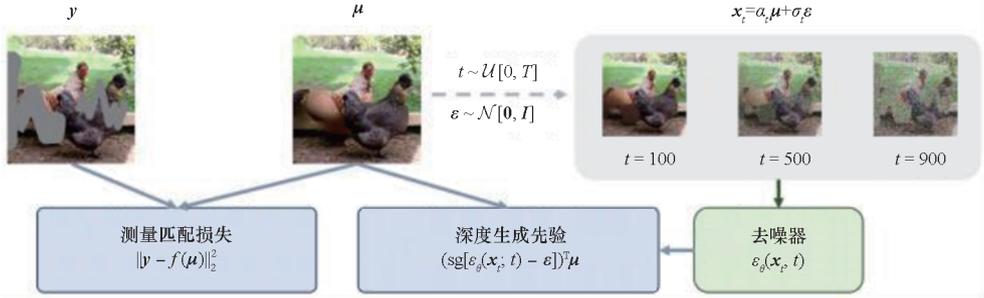


图5 VDS 结构图^[98]

Fig. 5 VDS structure illustration^[98]

VDS 在扩散模型的框架下为解决图像处理逆问题提供了一个新的变分视角,将迭代采样过程转化为随机优化模型,便于应用轻量级迭代实现逆问题解算,在多种线性和非线性图像处理逆问题上的实验结果验证了 VDS 的优越性。

空间分解及后验采样方法性能评估如表 7 所示,该两类技术路线着重于提升扩散模型求解图像处理逆问题的求解质量,在多种评估指标上都有较为显著的提升,但是在求解速度方面难以满足实际需求,部分限制了其实际应用价值。后续要介绍的两类技术路线,即隐式扩散模型和知识蒸馏,则着重于在保证求解质量的前提下,提高扩散模型的计算效率,实现更快速的求解,以提升其实际的应用价值。

2.3 隐式扩散模型

传统的扩散模型大多是直接在图像域上进行正向扩散和反向去噪的,在迭代步骤不变的前提下,若图像域维度较大,则每一步的迭代时间会显著变长,直接降低图像生成的效率。为此,2022 年

表7 空间分解及后验采样方法性能评估

Tab. 7 Performance evaluation of spatial decomposition and posterior sampling methods

方法	数据集及逆问题	指标
DDNM ^[93]	ImageNet 256 × 256 图像去噪	PSNR: 32.06
		SSIM: 0.968
		FID: 3.89
DDRM ^[94]	ImageNet 256 × 256 图像去噪	PSNR: 35.65
		SSIM: 0.96
		KID: 0.70
MCG ^[95]	ImageNet 256 × 256 图像重建	FID: 34.8
		LPIPS: 0.308
DPS ^[97]	ImageNet 256 × 256 图像动态模糊去除	FID: 56.08
		LPIPS: 0.389
VDS ^[98]	ImageNet 256 × 256 图像重建	PSNR: 23.29
		SSIM: 0.87
		KID: 0.86

4 月 Rombach 等提出了隐式扩散模型^[11] (latent diffusion model, LDM), 其通过编码解码网络建立了图像域与隐空间之间的映射, 进而在隐空间实现扩散模型的正向扩散与反向去噪过程, 模型结构如图 6 所示。由于隐空间的维度一般远小于图像域的维度, 则隐空间上的扩散模型相比在图像域具有更低的计算复杂度。

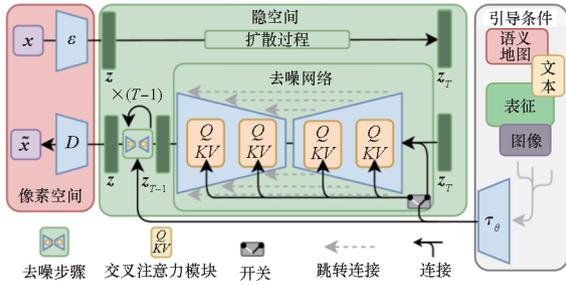


图 6 隐式扩散模型结构图^[11]

Fig. 6 LDM structure illustration^[11]

具体实现上, LDM 使用了预训练的感知压缩模型作为编码解码器, 在预训练过程中损失函数不仅使用了感知损失和基于片的对抗损失, 还为了进一步提高编码解码重构精度考虑了两种正则化, 即 KL 正则化与矢量量化 (vector quantization, VQ) 正则化。实验结果表明 LDM 在多个无条件任务 (例如图像合成、修复、随机超分辨率等) 和数据集上具有优异性能, 并为处理百万级维度图像提供了计算思路和条件。针对有条件任务, LDM 在神经网络中引入交叉注意机制来融合多模态条件, 并增强底层 U-Net 主干网络, 可实现文生图、图生图等有条件任务。LDM 是扩散模型发展历史上里程碑式节点, 在很多有影响力的软件中得到广泛使用, 如图像生成软件 Stable Diffusion、Dalle2、Imagen 等。

由于编码解码器的非线性映射, 利用扩散模型解决图像处理逆问题的传统方法并不能简单地推广到 LDM。Rout 等^[102] 第一个提出了利用预训练的 LDM 解决线性逆问题的框架, 即基于隐式扩散模型的后验采样 (posterior sampling with latent diffusion models, PSLD), 在随机遮挡修复、去噪、解模糊等领域显著加快了处理速度。He 等^[103] 针对从稀疏测量重建 CT 图像的问题, 提出了隐式扩散迭代重建 (iterative reconstruction based on latent diffusion, LDIR), 它是一种零样本学习方法, 通过预训练的 LDM 提取深度生成先验辅助迭代重建。Song 等^[104] 提出了基于预训练的隐式扩散模型的一般逆问题求解算法 ReSample, 通过解决反向采样过程中的优化问题来实现硬数

据一致性 (hard data consistency), 进而提出了新的重采样方案, 将数据一致的样本映射到噪声数据流形上, 并从理论上证明了其优势。

2.4 知识蒸馏

知识蒸馏是加速扩散模型采样的一种深度学习技术, 最早由 Hinton 等^[105] 在 2015 年提出, 其核心思想是将一个大型、复杂的模型 (通常称为教师模型) 的知识转移到一个小型、更简单的模型 (称为学生模型)。知识蒸馏的主要目的是提高学生模型的效率和可部署性, 同时保持或接近教师模型的性能。知识蒸馏引入了软标签的概念, 区别于传统的硬标签 (即类别标签), 软标签包含了对每个类别的概率分布信息, 通常由教师模型的输出构成。学生模型的训练损失函数通常包括针对硬标签的传统损失 (如交叉熵损失) 与针对软标签的损失 (如 KL 散度)。

知识蒸馏应用于扩散模型有多种方式。Salimans 等^[106] 在 2022 年提出的渐进式蒸馏扩散模型 (模型结构如图 7 所示) 以及 Song 等^[107] 在 2023 年提出的一致蒸馏, 其学生模型训练结束后仍然是无条件引导的, 生成结果不可控。

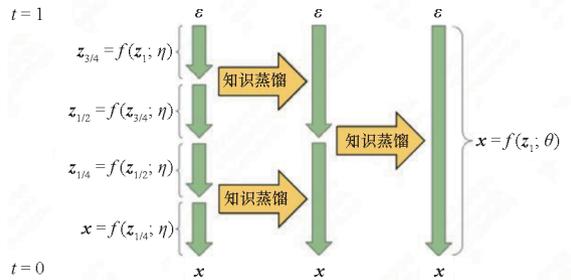


图 7 渐进式蒸馏扩散模型结构图^[106]

Fig. 7 Structure illustration of progressive distillation diffusion model^[106]

Meng 等^[108] 在 2023 年提出了一种将无分类器引导扩散模型蒸馏成快速采样模型的方法, 即给定一个预训练的无分类器引导模型, 首先学习一个单一模型来匹配组合条件和无条件模型的输出, 然后逐步将该模型提炼成需要更少采样步骤的扩散模型。Li 等^[109] 通过引入高效的网络结构和改进渐进蒸馏的方式提出了一种高效的文生图扩散模型 SnapFusion, 其通过识别原始模型的冗余和数据蒸馏减少图像解码器的计算, 并通过新的训练策略和引入无分类器引导的正则化来增强渐进蒸馏。Gu 等^[110] 提出了 BOOT 新技术, 通过无数据蒸馏算法克服了现有蒸馏方法对大量训练数据的依赖。Sauer 等^[111] 提出了一种对抗扩散蒸馏 (adversarial diffusion distillation, ADD), 用于

解决渐进式蒸馏在少步推理中产生模糊结果的问题。Lin 等^[112]也提出了一种将渐进式蒸馏和对抗式蒸馏相结合的方法 SDXL-Lightning。此外, Feng 等^[113-114]专门针对计算成像问题设计了基于评分先验的训练损失函数,同时提升了成像速度与质量。Mammadov 等^[115]与 Lee 等^[116]在分期变分推理(amortized variational inference)框架下提出了新的扩散模型知识蒸馏损失函数,用于提升知识蒸馏的稳健性,两者的不同在于前者学生模型使用的是流模型,而后者参考了一致模型,使用的是扩散模型本身的 U-Net。

隐式扩散模型及知识蒸馏典型方法性能评估如表 8 所示,可以发现,根据逆问题种类和数据集分辨率的不同,传统的基于扩散模型的方法求解逆问题通常需要几分钟至十几分钟不等,而隐式扩散模型与知识蒸馏两类技术路线大大提升了求解速度,已将求解速度提升到了秒级,同时保证了求解质量与传统方法相近。在扩散模型处理图像逆问题展现了较强的优势后,研究者们就致力于实现处理质量和计算效率的平衡,随着对网络结构、采样过程等步骤的不断优化调整,能够实现在更低的 FLOPS 下生

成质量相近甚至更优的处理结果。上述各类方法的主要创新点、优势及不足在表 9 中列出,可供读者参考。

表 8 隐式扩散模型及知识蒸馏典型方法性能评估
Tab. 8 Performance evaluation of typical implicit diffusion model and knowledge distillation methods

方法	数据集及逆问题	指标
Rombach 等 ^[11]	ImageNet 256 × 256 4 倍图像超分辨	FID: 2.4
		IS: 174.9
Song 等 ^[104]	CelebA-HQ 图像高斯模糊 去除	Time: 4.5 s
		LPIPS: 0.148
		PSNR: 30.69
Mammadov 等 ^[115]	CelebA 64 × 64 图像高斯模糊 去除	SSIM: 0.832
		Memory: 5 009 MB
		FID: 86.23
		PSNR: 23.81
Lee 等 ^[116]	FFHQ 256 × 256 4 倍图像超分辨	SSIM: 0.860
		Time: 0.035 s
		FID: 23.96
		Time: 0.04 s

表 9 扩散模型求解图像处理逆问题典型方法综述

Tab. 9 Review on typical methods of diffusion modeling for solving inverse problems in image processing

技术路线	方法	创新点及优势	不足
空间分解	DDNM ^[93]	分解图像空间为零域与值域,可使用预训练模型求解逆问题	仅适用于线性逆问题,且退化算子的获取在部分极端条件下较为困难
	DDRM ^[94]	利用扩散模型和流形约束,以无监督的方式高效地解决多种线性逆问题	仅适用于线性逆问题,且迭代过程需要进行 SVD,对算力要求较高
	MCG ^[95]	引入流形约束引导扩散过程,显著提高了图像重构的准确性和稳定性	在训练数据集较为复杂且多样化时表现较差
后验采样	DPS ^[97]	给出了具体的似然评分函数形式,在不同噪声类型下均有良好表现	给出的近似解较为宽松,在强干扰条件下重构准确性有待加强
	VDS ^[98]	引入变分推理方法估计后验分布,能够加快推理生成速度	对迭代步数等超参数较为敏感,需要针对不同任务手工调整
隐式扩散模型	Rombach 等 ^[11]	通过编码解码网络,在隐空间实现扩散过程,降低计算复杂度	只能在四倍缩放取得较好效果,使用更低的缩放倍数效果不佳
	PSLD ^[102]	第一个提出利用预训练 LDM 解决逆问题,提升了求解速度	训练数据集质量较差,且并未在非线性的逆问题上展现性能
	LDIR ^[103]	使用预训练的 LDM 辅助迭代重建,可实现零样本学习	仅适用于线性逆问题,在非线性的逆问题上的应用前景不明朗
	Song 等 ^[104]	解决了反向采样过程中的优化问题,实现硬数据一致性	需要对解码器进行反向传播,计算量大,且对超参数较为敏感

续表

技术路线	方法	创新点及优势	不足
知识蒸馏	Salimans 等 ^[106]	通过迭代,每次将扩散模型采样器采样步数减半,显著提高采样速度	泛化性较差,且针对随机性采样器蒸馏的效果不佳
	Song 等 ^[107]	通过学习一致性函数,能够实现单步生成,大大提高生成效率	一致性模型训练和推理成本较高,且受边界条件限制,泛化能力差
	Meng 等 ^[108]	将无分类器引导扩散模型蒸馏为单模型,显著提高了采样效率	在四步和八步采样能获得较好性能,但是两步和一步采样性能不佳
	Li 等 ^[109]	引入全新的训练策略,并使用无分类器引导的正则化增强渐进蒸馏	模型规模较大,训练成本高,且在两步和一步采样上性能不足
	BOOT ^[110]	无需数据,只需学习时间条件模型,即可预测教师模型的输出	生成质量严重受限于教师模型,且在训练数据易于获得的任务中性能与传统模型存在差距
	ADD ^[111]	结合了对抗损失和分数蒸馏损失,同在单步条件下优于现有的方法	模型规模较大,训练成本较高,且生成样本多样性较低
	SDXL-Lightning ^[112]	结合渐进式蒸馏和对抗式蒸馏的优点,既保证生成图像的质量,又保证模式的覆盖范围	需要为每个推理步数设置生成单独的检查点,不够灵活,且 U-Net 不是一步生成的最佳架构
	Feng 等 ^[113-114]	设计了全新的代理函数替代原有的训练损失函数,同时提升了成像速度与质量	代理函数精确度相较于原损失函数有所不足,在需要高精度求解的任务上可能表现不佳
	Mammadov 等 ^[115]	基于流模型提出了新的扩散模型知识蒸馏损失函数,用于提升知识蒸馏的稳健性	为保持简单性,网络架构缺乏对空间信息的归纳偏置,导致重建结果存在轻微的背景噪声
	Lee 等 ^[116]	将测量值直接映射到对应干净数据的隐式后验分布,实现高效的单一后验采样	需要额外训练隐式分布的神经网络,相比基准方法,需要更高的训练成本,同时泛化能力较弱

3 基于扩散模型求解图像处理逆问题的典型应用案例

针对图像处理逆问题,本团队对扩散模型进行了较为深入的研究与探索,在图像解模糊、图像乘性噪声(相干斑)抑制、图像融合与图像去雨等方面取得了系列成果,在此一并进行介绍。

3.1 图像解模糊

图像解模糊旨在从模糊图像中恢复清晰的原始图像。本团队针对带模糊核误差的图像解模糊问题,在扩散模型框架下提出了一种基于即插即用的图像解模糊方法,其基本原理是利用分离变量技术将优化问题分解为逼近项子问题和去噪子问题进行交替迭代,进而可以将任何学习到的去噪器作为先验插入上述迭代中。我们尝试将扩散模型作为生成去噪器集成到传统的即插即用方法

中,充分发挥扩散模型对深度生成先验的表征能力,使恢复的图像具有更好的视觉效果。

具体实现上,首先建立带核误差约束项 $\|\mathbf{h}\|_2^2$ 的优化模型:

$$\min_{\mathbf{x}, \mathbf{h}} \frac{1}{2\sigma_n^2} \|(\mathbf{K} + \mathbf{h})\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \Phi(\mathbf{x}) + \frac{\nu_h}{2} \|\mathbf{h}\|_2^2 \quad (20)$$

式中, \mathbf{K} 与 \mathbf{h} 分别为已知的模糊核与未知的核误差; $\Phi(\mathbf{x})$ 为图像 \mathbf{x} 的先验约束项,即去噪器; σ_n^2 、 λ 与 ν_h 分别为决定优化模型权重系数的相关参数。然后借助半二次分裂方法将扩散模型作为去噪器引入到即插即用方法,再结合扩散模型的采样步骤实现对模型的迭代求解,迭代步骤如式(21)所示,其中,第一个与第四个迭代更新涉及扩散模型的反向去噪过程,而第二个与三个迭代更新主要针对核误差与观测图像的引导修正,

所提方法如图8所示。

$$\begin{cases} \mathbf{x}_0^{(t)} = \arg \min_z \|\mathbf{z} - \mathbf{x}_t\|_2^2 + \lambda \Phi(\mathbf{z}) \\ \mathbf{h}^{(t)} = \arg \min_h \frac{1}{2\sigma_n^2} \|(\mathbf{K} + \mathbf{h})\mathbf{x}_0^{(t)} - \mathbf{y}\|_2^2 + \frac{\nu_h}{2} \|\mathbf{h}\|_2^2 \\ \hat{\mathbf{x}}_0^{(t)} = \arg \min_x \|\mathbf{y} - (\mathbf{K} + \mathbf{h}^{(t)})\mathbf{x}\|_2^2 + 2\gamma\sigma_n^2 \|\mathbf{x} - \mathbf{x}_0^{(t)}\|_2^2 \\ \mathbf{x}_{t-1} \leftarrow \hat{\mathbf{x}}_0^{(t)} \end{cases} \quad (21)$$

其中, γ 与 \mathbf{z} 分别为算法引入的拉格朗日乘子与增广变量。

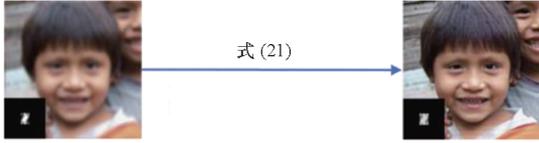


图8 图像解模糊示意图

Fig. 8 Diagram of image deblurring

实验结果如图9和表10所示。从视觉效果上看,所提方法视觉质量最好;从定量评价上看,所提方法在SSIM与LPIPS上与对比方法的最优结果相当,但在PSNR上优于对比方法。正是由于扩散模型能够挖掘图像的深度生成先验,才使得所提方法能够在重构质量上取得较大优势。

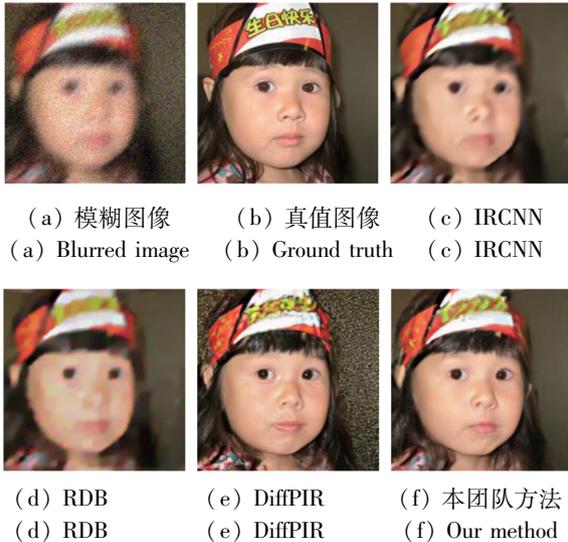


图9 图像解模糊重构结果(运动模糊)

Fig. 9 Restoration results of image deblurring(motion blur)

表10 图像解模糊量化评估

Tab. 10 Quantitative evaluation of image deblurring

运动长度	对比方法最优结果			本团队方法		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
10	23.47	0.81	0.49	23.81	0.82	0.50
20	22.81	0.83	0.53	22.97	0.87	0.52
30	20.95	0.80	0.55	21.62	0.79	0.51

3.2 相干斑抑制

合成孔径雷达(synthetic aperture radar, SAR)的相干成像机理导致SAR图像总会伴随相干斑噪声,这是一种乘性噪声,满足

$$\mathbf{y} = \mathbf{x} \odot \mathbf{n} \quad (22)$$

式中: \mathbf{x} 和 \mathbf{y} 分别为干净SAR图像和散斑SAR图像; \mathbf{n} 为相干斑噪声,通常可假设其满足伽马分布; \odot 为向量逐元素相乘。相干斑抑制可视为图像处理逆问题,对此本团队提出了一种基于扩散后验采样的相干斑抑制方法,即SAR-DPS^[117],如图10所示。具体而言,该方法在式(22)提供的正过程模型基础上,对扩散模型中的似然得分函数进行了重新估计:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t) = \frac{\partial \hat{\mathbf{x}}_0}{\partial \mathbf{x}_t} \frac{\partial R}{\partial \hat{\mathbf{x}}_0} \quad (23)$$

式中, $\hat{\mathbf{x}}_0$ 的定义同式(18), $R = \log p(\mathbf{y} | \mathbf{x}_t)$ 。式(23)中的两项导数可通过下式计算得到:

$$\begin{cases} \frac{\partial R}{\partial \hat{\mathbf{x}}_0} = \frac{L}{\hat{\mathbf{x}}_0 \odot \hat{\mathbf{x}}_0} \odot (\mathbf{y} - \hat{\mathbf{x}}_0) \\ \frac{\partial \hat{\mathbf{x}}_0}{\partial \mathbf{x}_t} = \frac{1}{\sqrt{\alpha_t}} [\mathbf{I} + (1 - \bar{\alpha}_t) \mathcal{J}(s_\theta(\mathbf{x}_t, t))] \end{cases} \quad (24)$$

其中, \mathbf{I} 表示单位矩阵, $\mathcal{J}(\cdot)$ 表示对向量函数求雅可比矩阵, L 为SAR图像视数。进一步将式(24)代入式(23)可得似然得分函数的估计式。

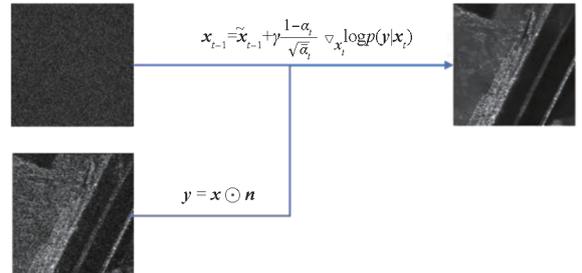


图10 SAR图像的相干斑抑制示意图

Fig. 10 Despeckling diagram of synthetic SAR images

SAR-DPS在合成SAR图像上的实验结果如图11与表11所示,表明SAR-DPS在视觉效果和定量评价方面都取得了较好性能。SAR-DPS可以作为基于扩散模型的SAR相干斑抑制的一般框架,可以探索网络结构优化、改进采样方法等多种技术,进一步提高其性能。

3.3 图像融合

与传统图像处理逆问题不同,图像融合没有原始图像真值可参考,故难以建立其正过程模型。针对这一问题,本团队针对光学与雷达图像融合问题,从传统优化模型出发,利用扩散模型挖掘的图像深度生成先验,提升图像融合质

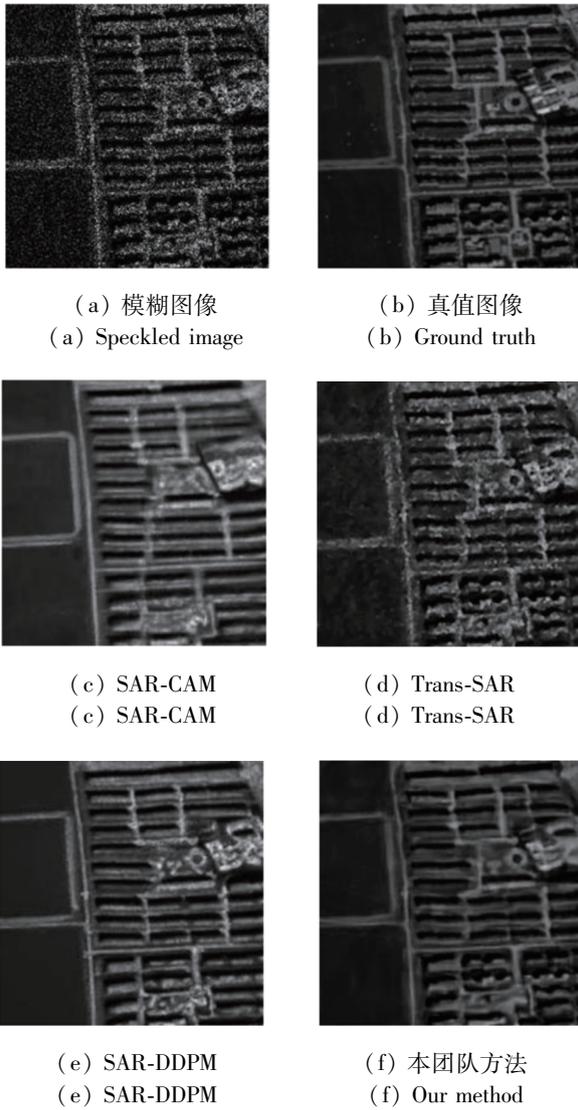


图 11 合成 SAR 图像的相干斑抑制结果

Fig. 11 Despeckling results of synthetic SAR images

表 11 合成 SAR 图像的相干斑抑制定量评价

Tab. 11 Quantitative evaluation of despeckling for synthetic SAR images

方法	SAR-CAM	Trans-SAR	SAR-DDPM	SAR-DPS
PSNR	25.71	24.60	25.99	27.67
SSIM	0.82	0.79	0.84	0.88

量。基本思路如下:首先建立结合稀疏先验和总变分先验设计图像融合和优化模型,该模型可转化为对融合图像拉普拉斯分布的极大似然估计问题,进而在扩散后验采样框架下推导出由光学图像与雷达图像引导的扩散模型反向去噪过程,并设计了基于扩散后验采样的图像融合算法^[118]。

具体而言,将光学图、雷达图像和融合图像分

别记为 $o \in \mathbf{R}^{H \times W}$ 、 $r \in \mathbf{R}^{H \times W}$ 和 $f \in \mathbf{R}^{H \times W}$,以融合图像 f 为扩散目标,扩散模型的反向去噪过程可以写作:

$$df = \left[-\frac{\beta(t)}{2}f - \beta(t)\nabla_{f_i} \log p_i(f_i | o, r) \right] dt + \sqrt{\beta(t)}dw \quad (25)$$

其中,扩散模型的漂移系数和扩散系数分别为 $-\beta(t)/2$ 与 $\sqrt{\beta(t)}$ 。

对条件评分函数的估计,可转化为对估计量 \hat{f}_0 的引导修正,在此利用结合稀疏约束与总变分约束的优化模型实现引导修正,即

$$\hat{f} = \arg \min_f [\|f - \hat{f}_0\|_1 + a\|f - o\|_1 + b\|f - r\|_1 + c\|\nabla(f - o)\|_1 + d\|\nabla(f - r)\|_1] \quad (26)$$

式中, \hat{f}_0 为依据式 (18) 利用 f_i 对融合图像的估计, a, b, c, d 为非负正则化参数。已经证明,式 (26) 中最小化 ℓ_1 范数等价于对拉普拉斯分布的极大似然估计,进而式 (26) 等价于对融合图像的后验分布估计。具体图像融合流程如图 12 所示。

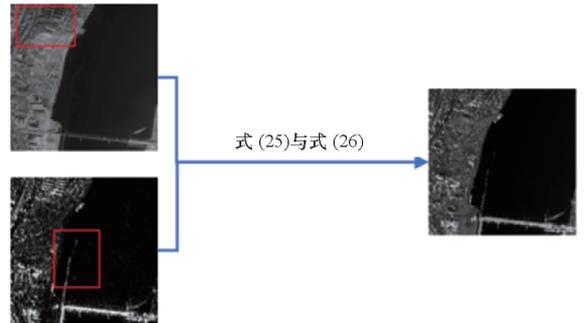


图 12 图像融合示意图

Fig. 12 Diagram of image fusion

在实测数据上对上述方法进行验证,实验结果如图 13 与表 12 所示。表 12 展示的定量评价结果表明本团队方法在 MI、CE、PSNR 和 SSIM 指标上达到了最优,在 VIFF 指标上达到了次优,总体表现较好,与图 13 所示的视觉效果一致。



(a) 光学图像 (a) Optical image (b) SAR 图像 (b) SAR image

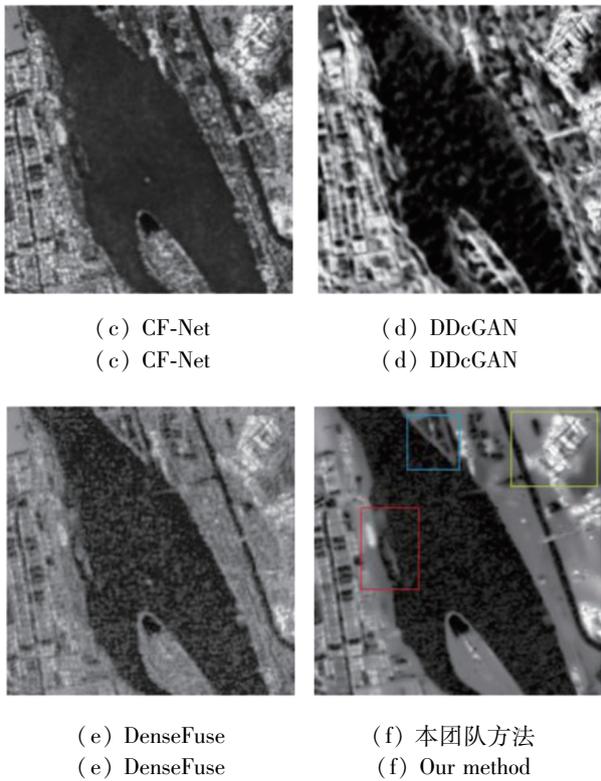


图 13 图像融合结果
Fig. 13 Results of image fusion

表 12 图像融合定量评价

Tab. 12 Quantitative evaluation of image fusion

方法	VIFF	MI	CE	PSNR	SSIM
CF-Net	0.22	1.52	0.19	14.21	0.48
DDcGAN	0.15	1.17	0.39	11.17	0.11
DenseFuse	0.36	2.01	0.51	14.59	0.51
本团队方法	0.30	2.32	0.53	14.94	0.55

3.4 图像去雨

图像去雨面向的问题是成像时因降雨环境导致的图像中成像场景与雨条纹的混合现象,图像去雨旨在解耦图像中的成像场景与雨条纹,恢复清晰图像。Özdenizci 等^[119]在 2023 年尝试利用扩散模型实现图像去雨,采用无分类器方案,即直接将清晰图像与对应的雨合成图像成对输入扩散模型进行训练,训练成本较高,且训练好的模型难以适用其他数据集,因此方法应用具有较大局限性。

本团队基于后验采样的扩散模型提出了一种针对传统去雨方法的精炼模型^[120],实现了去雨效果的增强,同时仅需使用预训练模型而无需额外的训练。该方法从传统去雨结果出发,重点解决雨条纹的残留问题,基本思路是将残留雨条纹的图像视为待修复的图像,残留雨条纹的位置即

为图像需要修复的像素,因此可将其转化为图像修复问题,具体流程如图 14 所示。该方法的关键在于检测定位残留雨条纹,用于生成图像修复问题的掩码,这里考虑使用传统去雨模型输出雨条纹 R 的估计掩码 K ,即

$$K_{ij} = \begin{cases} 0 & R_{ij} < T \\ 1 & R_{ij} \geq T \end{cases} \quad (27)$$

其中,下标表示像素位置, T 为阈值参数。上式表示在雨条纹强度较大的位置容易出现雨残留,因此需要在该位置修复图像。进一步利用掩码对传统去雨模型输出的残留雨条纹图像进行挖取(即点乘),得到图像修复的观测图像,进而在图像处理逆问题框架下利用扩散模型实现清晰图像的高质量重构。

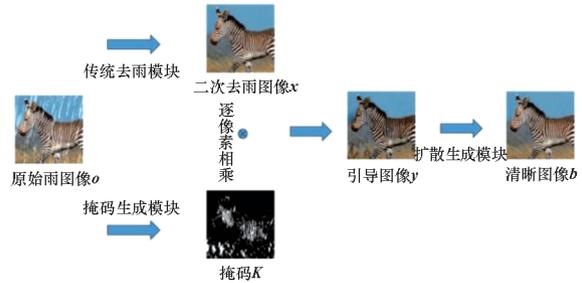


图 14 图像去雨流程

Fig. 14 Flowchart of image deraining

图像去雨实验结果如图 15 与表 13 所示,表明所提方法有效去除了对比方法无法去除的雨条纹(红框中所示),并恢复了背景图像。从指标上看,所提方法总体与对比方法相当,但 FID 指标提升明显。

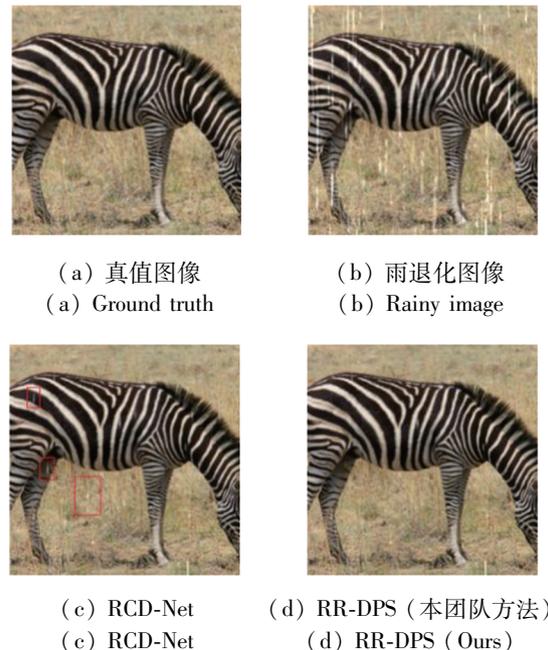


图 15 图像去雨实验结果

Fig. 15 Experiment results of image deraining

表 13 图像去雨定量评价

Tab. 13 Quantitative evaluation of image deraining

方法	PSNR	SSIM	FID	LPIPS
RCD-Net	29.73	0.968 8	7.61	0.015
RR-DPS	29.92	0.969 7	6.57	0.013

4 总结与展望

扩散模型在图像处理逆问题领域展现出了显著优势,能够为多种图像处理逆问题提供全新的解决视角和思路。然而,由于图像处理逆问题本身的复杂性(非线性、强噪声、高病态等),以及扩散模型相关理论与方法尚在发展之中,目前面向图像处理逆问题的扩散模型还面临以下几个问题:

1) 计算复杂度高。由于扩散模型的反向去噪过程需要逐步进行推理,每次推理需要执行一次深度学习网络前向传播过程,因此计算复杂度较高。尽管研究者们提出了很多加速策略,如隐式扩散模型、知识蒸馏等,相较于传统神经网络的一次前向传播,扩散模型仍然耗时较长,这限制了扩散模型在实时解决图像处理逆问题中的应用。

2) 扩散模型提供的深度生成先验与观测图像提供的测量数据间的平衡问题。利用扩散模型解决图像处理逆问题时,扩散模型提供了深度生成先验,而测量图像时提供测量数据用来实现条件引导,两者的权重如何平衡尚缺少理论依据。若先验权重过大,容易生成虚假内容,与观测图像不一致;若测量数据权重过大,则缺少先验约束从而难以克服逆问题的病态性。

3) 非线性图像处理逆问题的扩散模型稳健求解问题。图像处理逆问题包括相位恢复、非线性混合条件下的图像去雨、带空变模糊核的图像解模糊等非线性问题,利用扩散模型解决此类问题尽管可采用类似处理线性模型的方法,但考虑到非线性逆问题通常存在多个局部最优解,因此扩散模型求解的收敛性以及唯一性难以保证。

4) 扩散模型解决极端条件下图像处理逆问题的性能问题。随着扩散模型在图像处理逆问题中的应用越来越广,其面临的处理条件越来越复杂,包括训练样本数量少、观测图像质量差、噪声类型多样等,需要进一步提升扩散模型在解决极端条件下的图像处理逆问题的性能。

5) 图像处理逆问题的扩散模型多模态信息联合引导问题。智能时代产生了大量除图像外的文本、音频、视频等多模态信息,利用多模态信息

可以提升扩散模型对图像处理逆问题的求解能力,增强鲁棒性和可靠性。如何更有效地理解、融合和利用这些多模态信息、实现扩散模型的联合引导,将成为解决图像处理逆问题值得关注的研究方向。

综上所述,扩散模型是生成式人工智能的最具潜力的发展方向,已经成为解决图像处理逆问题的有效途径,未来迫切需要针对上述方向开展深入研究,并积极探索扩散模型在图像处理逆问题领域中的应用方式与应用场景,显著提升其处理效能。

参考文献 (References)

- [1] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory [R/OL]. (1986 - 01 - 01) [2024 - 12 - 11]. https://bg.sunwayinfo.com.cn/report_detail?id=e8c45a11-d22d-11ea-bd16-0cc47a2ac945.
- [2] HINTON G E, OSINDERO S, THE Y W. A fast learning algorithm for deep belief nets [J]. *Neural Computation*, 2006, 18(7): 1527 - 1554.
- [3] SALAKHUTDINOV R, HINTON G. A better way to pretrain deep Boltzmann machines [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, 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, WELING M. Auto-encoding variational Bayes [EB/OL]. (2013 - 12 - 20) [2024 - 12 - 11]. <https://arxiv.org/abs/1312.6114>.
- [6] ZHANG M T, SUN Y T, MCDONAGH S, et al. Flow based models for manifold data [EB/OL]. (2021 - 09 - 29) [2024 - 12 - 11]. <https://arxiv.org/pdf/2109.14216v1>.
- [7] REZENDE D J, MOHAMED S. Variational inference with normalizing flows [C]//Proceedings of the 32nd International Conference on Machine Learning - Volume 37, 2015: 1530 - 1538.
- [8] LECUN Y, CHOPRA S, HADSELL R, et al. A tutorial on energy-based learning [M]. Cambridge, Massachusetts: The MIT Press, 2006.
- [9] VAN DEN OORD A, KALCHBRENNER N, VINYALS O, et al. Conditional image generation with PixelCNN decoders [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 4797 - 4805.
- [10] NICHOL A, DHARIWAL P, RAMESH A, et al. GLIDE: towards photorealistic image generation and editing with text-guided diffusion models [EB/OL]. (2021 - 12 - 20) [2024 - 12 - 11]. <https://arxiv.org/abs/2112.10741>.
- [11] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models [C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 10674 - 10685.
- [12] CHENG S, CHEN Y J, CHIU W C, et al. Adaptively-

- realistic image generation from stroke and sketch with diffusion model [C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023; 4043–4051.
- [13] JOLICOEUR-MARTINEAU A, PICHÉ-TAILLEFER R, DES COMBES R T, et al. Adversarial score matching and improved sampling for image generation [EB/OL]. (2020–09–11) [2024–12–11]. <https://arxiv.org/abs/2009.05475>.
- [14] CHEN T, ZHANG R X, HINTON G. Analog bits: generating discrete data using diffusion models with self-conditioning [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [15] GU Z X, CHEN H X, XU Z E, et al. DiffusionInst: diffusion model for instance segmentation [C]//Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024.
- [16] XU J L, WANG X T, CHENG W H, et al. Zero-shot text-to-3D synthesis using 3D shape prior and text-to-image diffusion models [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023; 20908–20918.
- [17] SAHARIA C, HO J, CHAN W, et al. Image super-resolution via iterative refinement [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 4713–4726.
- [18] GAO Z J, GUO J L, TAN X, et al. Empowering diffusion models on the embedding space for text generation [C]//Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 2024; 4664–4683.
- [19] GONG S S, LI M K, FENG J T, et al. DiffuSeq: sequence to sequence text generation with diffusion models [C]//Proceedings of the International Conference on Learning Representations (ICLR), 2023.
- [20] HE Z F, SUN T X, TANG Q, et al. DiffusionBERT: improving generative masked language models with diffusion models [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023; 4521–4534.
- [21] LIN Z H, GONG Y Y, SHEN Y L, et al. Text generation with diffusion language models: a pre-training approach with continuous paragraph denoise [C]//Proceedings of the 40th International Conference on Machine Learning (ICML), 2023.
- [22] TASHIRO Y, SONG J M, SONG Y, et al. CSDI: conditional score based diffusion models for probabilistic time series imputation [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021; 24804–24816.
- [23] ALCARAZ J M L, STRODTHOFF N. Diffusion-based time series imputation and forecasting with structured state space models [EB/OL]. (2022–08–19) [2024–12–11]. <https://arxiv.org/abs/2208.09399>.
- [24] 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.
- [25] GU S Y, CHEN D, BAO J M, et al. Vector quantized diffusion model for text-to-image synthesis [C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022; 10686–10696.
- [26] SAHARIA C, CHAN W, SAXENA S, et al. Photorealistic text-to-image diffusion models with deep language understanding [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022; 36479–36494.
- [27] 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.
- [28] KIM B, YE J C. Diffusion deformable model for 4D temporal medical image generation [C]//Proceedings of the Medical Image Computing and Computer Assisted Intervention-MICCAI 2022, 2022; 539–548.
- [29] SONG Y, SHEN L Y, XING L, et al. Solving inverse problems in medical imaging with score-based generative models [EB/OL]. (2021–11–15) [2024–12–11]. <https://arxiv.org/abs/2111.08005>.
- [30] WU K E, YANG K K, VAN DEN BERG R, et al. Protein structure generation via folding diffusion [J]. *Nature Communications*, 2024, 15(1): 1059.
- [31] LEE J S, KIM J, KIM P M. Score-based generative modeling for de novo protein design [J]. *Nature Computational Science*, 2023, 3(5): 382–392.
- [32] EL HAKOUME A, AFRAITES L, LAGHRIB A. An improved coupled PDE system applied to the inverse image denoising problem [J]. *Electronic Research Archive*, 2022, 30(7): 2618–2642.
- [33] HAN J, ZHANG S L, ZHENG S Z, et al. Bias analysis and correction for ill-posed inversion problem with sparsity regularization based on L1 norm for azimuth super-resolution of radar forward-looking imaging [J]. *Remote Sensing*, 2022, 14(22): 5792.
- [34] LUO X Y, CHEN Y, CHENG J, et al. A mechanism-based method for image inpainting [J]. *Applied Mathematics Letters*, 2024, 154: 109113.
- [35] BUCCINI A, DONATELLI M, RAMLAU R. A semiblind regularization algorithm for inverse problems with application to image deblurring [J]. *SIAM Journal on Scientific Computing*, 2018, 40(1): A452–A483.
- [36] SUN G M, LENG J S, CATTANI C. A particular directional multilevel transform based method for single-image rain removal [J]. *Knowledge-Based Systems*, 2020, 200: 106000.
- [37] TOLEYEV Y, SHAO W Z. Single image haze removal: comparative studies with advanced matting approaches [J]. *Journal of Physics: Conference Series*, 2018, 1069: 012145.
- [38] LEE S Y, CHO D H, JUNG C, et al. Denoising low-intensity diffraction signals using k -space deep learning: applications to phase recovery [J]. *Physical Review Research*, 2021, 3(4): 043066.
- [39] DUN X, IKOMA H, WETZSTEIN G, et al. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging [J]. *Optica*, 2020, 7(8): 913–922.
- [40] NEZHAD Z H, KARAMI A, HEYLEN R, et al. Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(6): 2377–2389.

- [41] XIE Z H, LIU L J, CHEN Z H. Image compressed sensing: from deep learning to adaptive learning[J]. Knowledge-Based Systems, 2024, 293: 111659.
- [42] MO C J, TIAN J R, WU F G, et al. Remote sensing images fusion algorithm based on weighted wavelet analysis [J]. Journal of Jilin University (Science Edition), 2018, 56(3): 617–624.
- [43] WANG P, SULAIMANI H J, CHA J C. Digital model creation and image meticulous processing based on variational partial differential equation [J]. Applied Mathematics and Nonlinear Sciences, 2022, 7(2): 467–476.
- [44] BIOUSCAS-DIAS J M, PLAZA A, DOBIGEON N, et al. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 5(2): 354–379.
- [45] WAINWRIGHT M J, JORDAN M I. Graphical models, exponential families, and variational inference [J]. Foundations and Trends in Machine Learning, 2008, 1(1/2): 1–305.
- [46] CHUNG H, NAM H, YE J C. Review of diffusion models: theory and applications[J]. Journal of the Korean Society for Industrial and Applied Mathematics, 2024, 28(1): 1–21.
- [47] CROITORU F A, HONDRU V, IONESCU R T, et al. Diffusion models in vision: a survey[EB/OL]. (2022–09–10)[2024–12–11]. <https://arxiv.org/abs/2209.04747>.
- [48] YANG L, ZHANG Z L, SONG Y, et al. Diffusion models: a comprehensive survey of methods and applications[J]. 2023, 56(4): 1–39.
- [49] CAO H Q, TAN C, GAO Z Y, et al. A survey on generative diffusion model[J]. 2024, 36(7): 2814–2830.
- [50] ARELLANO S, OTERO B, TOUS R. Exploring image transformations with diffusion models: a survey of applications and implementation code[C]//Proceedings of the Machine Learning, Optimization, and Data Science: 9th International Conference, 2023: 19–33.
- [51] 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 Z H, ZHOU Z B, LI X C. Survey on generative diffusion model [J]. Computer Science, 2024, 51(1): 273–283. (in Chinese)
- [52] HO J, JAIN A, ABBEEL P, et al. Denoising diffusion probabilistic models [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020, 574: 6840–6851.
- [53] SONG Y, SOHLG-DICKSTEIN J, KINGMA D P, et al. Score-based generative modeling through stochastic differential equations [EB/OL]. (2020–11–26)[2024–12–11]. <https://arxiv.org/abs/2011.13456>.
- [54] WILSON J T, MORICONI R, HUTTER F, et al. The reparameterization trick for acquisition functions [EB/OL]. (2017–12–01)[2024–12–11]. <https://arxiv.org/abs/1712.00424>.
- [55] SOHL-DICKSTEIN J, WEISS E A, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics [C]//Proceedings of the 32nd International Conference on Machine Learning, 2015: 2256–2265.
- [56] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, 2015: 234–241.
- [57] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 8780–8794.
- [58] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 2234–2242.
- [59] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6629–6640.
- [60] HO J, SALIMMANS T. Classifier-Free diffusion guidance [EB/OL]. (2022–07–26)[2024–12–11]. <https://arxiv.org/abs/2207.12598>.
- [61] KYNKÄÄNNIEMI T, KARRAS T, LAINE S, et al. Improved precision and recall metric for assessing generative models[J]. Statistics, 2019, 353: 3927–3936.
- [62] BIŃKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying MMD GANs [EB/OL]. (2018–01–04)[2024–12–11]. <http://arxiv.org/abs/1801.01401v4>.
- [63] NICHOL A, DHARIWAL P. Improved denoising diffusion probabilistic models [C]//Proceedings of the 38th International Conference on Machine Learning, 2021: 8162–8171.
- [64] KINGMA D P, SALIMANS T, POOLE B, et al. Variational diffusion models [C]//Proceedings of the 35th Conference on Neural Information Processing Systems, 2021.
- [65] SAN-ROMAN R, NACHMANI E, WOLF L. Noise estimation for generative diffusion models [EB/OL]. (2021–04–06)[2024–12–11]. <https://arxiv.org/abs/2104.02600>.
- [66] WANG J Y, LYU Z Y, LIN D H, et al. Guided diffusion model for adversarial purification [EB/OL]. (2022–05–30)[2024–12–11]. <https://arxiv.org/abs/2205.14969>.
- [67] SONG J M, MENG C L, ERMON S. Denoising diffusion implicit models [EB/OL]. (2020–10–06)[2024–12–11]. <https://arxiv.org/abs/2010.02502>.
- [68] ZHANG Q S, TAO M L, CHEN Y X. gDDIM: generalized denoising diffusion implicit models [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [69] SONG Y, ERMON S. Improved techniques for training score-based generative models [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 12438–12448.
- [70] LIU L P, REN Y, LIN Z J, et al. Pseudo numerical methods for diffusion models on manifolds [EB/OL]. (2022–02–20)[2024–12–11]. <https://arxiv.org/abs/2202.09778>.
- [71] JOLICOEUR-MARTINEAU A, LI K, Piché-Taillefer R, et al. Gotta go fast when generating data with score-based models [EB/OL]. (2021–05–28)[2024–12–11]. <https://arxiv.org/abs/2105.14080>.
- [72] BORTOLI V D, THORNTON J, HENG J, et al. Diffusion schrödinger bridge with applications to score-based generative modeling [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 17695–17709.
- [73] DOCKHORN T, VAHDAT A, KREIS K. Score-based

- generative modeling with critically-damped Langevin diffusion [EB/OL]. (2021 - 12 - 14) [2024 - 12 - 11]. <https://arxiv.org/abs/2112.07068>.
- [74] SINHA A, SONG J M, MENG C L, et al. D2C: diffusion-denoising models for few-shot conditional generation [EB/OL]. (2021 - 06 - 12) [2024 - 12 - 11]. <https://arxiv.org/abs/2106.06819>.
- [75] PEEBLES W, XIE S N. Scalable diffusion models with transformers [C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 4172 - 4182.
- [76] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000 - 6010.
- [77] WANG Z D, ZHENG H J, HE P C, et al. Diffusion-GAN: training GANs with diffusion [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [78] ZHANG Q S, CHEN Y X. Diffusion normalizing flow [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 16280 - 16291.
- [79] KIM D J, NA B, KWON S J, et al. Maximum likelihood training of implicit nonlinear diffusion models [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 32270 - 32284.
- [80] WATSON D, HO J, NOROUZI M, et al. Learning to efficiently sample from diffusion probabilistic models [EB/OL]. (2021 - 06 - 07) [2024 - 12 - 11]. <https://arxiv.org/abs/2106.03802>.
- [81] XIAO Z S, KREIS K, VAHDAT A. Tackling the generative learning trilemma with denoising diffusion GANs [EB/OL]. (2021 - 12 - 15) [2024 - 12 - 11]. <https://arxiv.org/abs/2112.07804>.
- [82] BAO F, LI C X, SUN J C, et al. Estimating the optimal covariance with imperfect mean in diffusion probabilistic models [EB/OL]. (2022 - 06 - 15) [2024 - 12 - 11]. <https://arxiv.org/abs/2206.07309>.
- [83] CHUNG H, SIM B, YE J C. Come-closer-diffuse-faster: accelerating conditional diffusion models for inverse problems through stochastic contraction [C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 12403 - 12412.
- [84] DEASY J, SIMIDJIEVSKI N, LIÒ P. Heavy-tailed denoising score matching [EB/OL]. (2021 - 12 - 17) [2024 - 12 - 11]. <https://arxiv.org/abs/2112.09788>.
- [85] KARRAS T, AITTALA M, LAINE S, et al. Elucidating the design space of diffusion-based generative models [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 26565 - 26577.
- [86] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment [J]. Electronics Letters, 2008, 44(13): 800.
- [87] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600 - 612.
- [88] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 586 - 595.
- [89] HAN Y, CAI Y Z, CAO Y, et al. A new image fusion performance metric based on visual information fidelity [J]. Information Fusion, 2013, 14(2): 127 - 135.
- [90] CHEN Y, XUE Z Y, BLUM R S. Theoretical analysis of an information-based quality measure for image fusion [J]. Information Fusion, 2008, 9(2): 161 - 175.
- [91] DENG L Y. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning [J]. Technometrics, 2006, 48(1): 147 - 148.
- [92] KAWAR B, VAKSMAN G, ELAD M. SNIPS: solving noisy inverse problems stochastically [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 21757 - 21769.
- [93] WANG Y H, YU J W, ZHANG J. Zero-shot image restoration using denoising diffusion null-space model [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [94] KAWAR B, ELAD M, ERMON S, et al. Denoising diffusion restoration models [EB/OL]. (2022 - 01 - 27) [2024 - 12 - 11]. <https://arxiv.org/abs/2201.11793>.
- [95] CHUNG H J, SIM B, RYU D, et al. Improving diffusion models for inverse problems using manifold constraints [EB/OL]. (2022 - 06 - 02) [2024 - 12 - 11]. <https://arxiv.org/abs/2206.00941>.
- [96] ZHU Y Z, ZHANG K, LIANG J Y, et al. Denoising diffusion models for plug-and-play image restoration [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023: 1219 - 1229.
- [97] CHUNG H J, KIM J, MCCANN M T, et al. Diffusion posterior sampling for general noisy inverse problems [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [98] MARDANI M, SONG J M, KAUTZ J, et al. A variational perspective on solving inverse problems with diffusion models [C]//Proceedings of the International Conference on Learning Representations, 2024.
- [99] MENG X, KABASHIMA Y. Diffusion model based posterior sampling for noisy linear inverse problems [C]//Proceedings of the 16th Asian Conference on Machine Learning, 2025: 623 - 638.
- [100] SONG J M, VAHDAT A, MARDANI M, et al. Pseudoinverse-guided diffusion models for inverse problems [C]//Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
- [101] FEI B, LYU Z Y, PAN L, et al. Generative diffusion prior for unified image restoration and enhancement [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 9935 - 9946.
- [102] ROUT L, RAOOF N, DARAS G, et al. Solving linear inverse problems provably via posterior sampling with latent diffusion models [C]//Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023: 49960 - 49990.
- [103] HE L C, YAN H Y, LUO M T, et al. Iterative reconstruction based on latent diffusion model for sparse data reconstruction [EB/OL]. (2023 - 07 - 22) [2024 - 12 -

- 11]. <https://arxiv.org/abs/2307.12070v1>.
- [104] SONG B, KWON S M, ZHANG Z C, et al. Solving inverse problems with latent diffusion models via hard data consistency[EB/OL]. (2023-07-16)[2024-12-11]. <https://arxiv.org/abs/2307.08123>.
- [105] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09)[2024-12-11]. <https://arxiv.org/abs/1503.02531v1>.
- [106] SALIMANS T, HO J. Progressive distillation for fast sampling of diffusion models[EB/OL]. (2022-02-01)[2024-12-11]. <https://arxiv.org/abs/2202.00512>.
- [107] SONG Y, DHARIWAL P, CHEN M, et al. Consistency models[EB/OL]. (2023-03-02)[2024-12-11]. <https://arxiv.org/abs/2303.01469>.
- [108] MENG C L, ROMBACH R, GAO R Q, et al. On distillation of guided diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023; 14297-14306.
- [109] LI Y Y, WANG H, JIN Q, et al. SnapFusion: text-to-image diffusion model on mobile devices within two seconds[C]//Proceedings of the 37th Conference on Neural Information Processing Systems(NeurIPS), 2023; 20662-20678.
- [110] GU J T, ZHAI S F, ZHANG Y Z, et al. BOOT: data-free distillation of denoising diffusion models with bootstrapping[EB/OL]. (2023-06-08)[2024-12-11]. <https://arxiv.org/abs/2306.05544>.
- [111] SAUER A, LORENZ D, BLATTMANN A, et al. Adversarial diffusion distillation[C]//Proceedings of the Computer Vision-ECCV 2024; 18th European Conference, 2024.
- [112] LIN S C, WANG A, YANG X. SDXL-Lightning: progressive adversarial diffusion distillation [EB/OL]. (2024-02-21)[2024-12-11]. <https://arxiv.org/abs/2402.13929>.
- [113] FENG B, BOUMAN K. Efficient Bayesian computational imaging with a surrogate score-based prior[C]//Proceedings of the 11th International Conference on Learning Representations(ICLR), 2023.
- [114] FENG B T, SMITH J, RUBINSTEIN M, et al. Score-based diffusion models as principled priors for inverse imaging[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023; 10486-10497.
- [115] MAMMADOV A, CHUNG H, YE J C. Amortized posterior sampling with diffusion prior distillation [EB/OL]. (2024-07-25)[2024-12-11]. <https://arxiv.org/abs/2407.17907>.
- [116] LEE S, PARK D, KONG I, et al. Diffusion prior-based amortized variational inference for noisy inverse problems[C]// Proceedings of the Computer Vision-ECCV 2024; 18th European Conference, 2024; 288-304.
- [117] WANG Z L, HAN J L, ZHANG C L. Diffusion posterior sampling for SAR despeckling[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-19.
- [118] ZHANG C L, WANG Z L, HAN J L. Diffusion posterior sampling for remote sensing image fusion[C]// Proceedings of the 7th International Conference on Vision, Image and Signal Processing, 2023; 28-33.
- [119] ÖZDENIZCI O, LEGENSTEIN R. Restoring vision in adverse weather conditions with patch-based denoising diffusion models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10346-10357.
- [120] WU Y H, ZHANG C L, CHANG Y J, et al. Refinement for image rain removal based on diffusion model [C]// Proceedings of the 16th International Conference on Graphics and Image Processing (ICGIP 2024), 2025; 755-763.