中图法分类号:TP391.9　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-24

论文引用格式:Wan Yiqian, Zhang Yonghong, Yu Xin. Diffusion Models for Sparse-View 3D Reconstruction: A Survey[J/OL]. Journal of Image and Graphics, XXXX:1-24. DOI: 10.11834/jig.260539. (万贻前, 张永宏, 余鑫. 面向稀疏视图三维重建的扩散模型方法综述[J/OL]. 中国图象图形学报,XXXX:1-24. DOI: 10.11834/jig.260539.)[DOI:10.11834/jig.260539]

面向稀疏视图三维重建的扩散模型方法综述

万贻前,张永宏,余鑫

重庆安全技术职业学院,重庆市　404100

摘　要:稀疏视图三维重建旨在从少量图像中恢复场景的三维结构与外观。由于视角覆盖不充分、几何约束不足,重建结果容易出现结构缺失、纹理模糊和跨视角不一致。扩散模型从大规模图像数据中学习的生成先验可为未观测区域的恢复提供补充信息,但生成内容的视觉合理性并不等同于几何准确性。本文聚焦少量真实图像驱动的静态三维重建,涵盖视频扩散辅助的观测补全与反馈优化方法,单图三维生成与纯新视角生成仅作为相关技术基础介绍。依据扩散信息的使用方式,围绕扩散先验约束与生成观测驱动两种机制,系统梳理代表性方法的建模思路与技术特点,分析生成与重建之间的反馈交互及不同机制在混合方法中的作用环节。同时,归纳常用数据集与评价指标,并结合实验设置对代表性方法的实验结果进行对比分析。结合先验偏差、生成观测误差及反馈过程中的误差累积,讨论先验与真实观测的平衡、生成观测的可靠性及反馈稳定性,展望相机位姿鲁棒性、复杂场景重建、跨场景泛化与计算效率等方面的研究方向。

关键词:稀疏视图三维重建;扩散模型;生成先验;生成观测;扩散反馈

Diffusion Models for Sparse-View 3D Reconstruction: A Survey

Wan Yiqian, Zhang Yonghong, Yu Xin

Chongqing Vocational Institute of Safety Technology, Chongqing 404100, China

Abstract: Sparse-view 3D reconstruction aims to recover the structure and appearance of a scene or object from a limited number of observed images. Insufficient viewpoint coverage, limited image overlap, and occlusion make the problem underconstrained, often leading to incomplete geometry, blurred textures, floating artifacts, and inconsistent renderings. Although neural radiance fields (NeRFs) and 3D Gaussian Splatting (3DGS) improve scene representation and rendering, they do not eliminate the ambiguity caused by missing observations. Diffusion models learn regularities of appearance and structure from large-scale image datasets, providing generative priors for recovering poorly observed regions. However, plausible images may still contain structures or textures that conflict with the actual scene. The central challenge is to incorporate generative information while maintaining fidelity to observed content and consistency across viewpoints. This review focuses on static 3D reconstruction from a small number of observed images, considering both known and unknown camera poses. Single-image 3D generation and conditional novel-view generation are discussed as supporting techniques for viewpoint control, prior construction, and observation completion. Video diffusion is included where generated viewpoint sequences assist the recovery of a static scene; dynamic-scene reconstruction is outside the scope. The review introduces conditional diffusion, score functions, Score Distillation Sampling (SDS), NeRFs, 3DGS, and differentiable rendering to explain how information learned in image space can influence three-dimensional recovery. The methodological discussion is organized into three sections with distinct purposes. The first two examine diffusion-prior constraints and generated-

收稿日期:2026-08-28;**修回日期**:2026-10-25

© 中国图象图形学报版权所有

observation-driven reconstruction as mechanisms for using diffusion-derived information. The third investigates feedback between generation and reconstruction as a separate interaction dimension. These sections therefore do not define three mutually exclusive method categories. Individual methods are described through their optimization objectives, intermediate information, and data flow at each stage. When a framework combines mechanisms or supports alternative reconstruction modules, its different configurations are identified. Diffusion-prior constraints convert score estimates, denoising residuals, or distribution-matching objectives into optimization signals for a three-dimensional representation. These signals are propagated through differentiable rendering to update radiance fields, Gaussian primitives, or implicit surfaces, while reconstruction losses preserve agreement with observed images. Some methods obtain target estimates through multiple denoising steps and use the differences between these estimates and current renderings to guide optimization. Such intermediate images must be interpreted according to their role in the objective and update procedure. The effectiveness of this mechanism depends on prior accuracy, conditioning information, camera poses, and the balance between observation losses and generative guidance. Excessive prior influence can introduce unsupported content, whereas insufficient influence leaves poorly observed regions inadequately constrained. Generated-observation-driven reconstruction uses synthesized or restored RGB images, depth maps, surface normals, viewpoint sequences, or their features as auxiliary supervision or reconstruction inputs. Generated information can support scene-specific optimization or be processed by pretrained reconstruction modules. The recovery procedure may consequently involve iterative fitting, direct parameter prediction, or conditional three-dimensional generation. Feedforward prediction reduces the need for per-scene parameter optimization, although preceding view generation may still require multiple diffusion steps. Nevertheless, errors in shape, texture, visibility, depth, or scale can propagate into the final representation through supervision losses or network inputs. Increasing the number of generated views does not by itself ensure stronger or more reliable geometric constraints. Generation–reconstruction feedback concerns the interaction between generative processing and the current three-dimensional state. Rendered images, depth estimates, visibility information, or auxiliary geometry can guide restoration, and the resulting information can subsequently update the scene representation. A single restoration stage followed by reconstruction from fixed observations is distinguished from repeated feedback, in which the updated state influences later generation. Feedback may operate through prior distillation, generated-image supervision, or learned parameter updates. Image-level interaction focuses on selected viewpoints, whereas sequence-level interaction organizes neighboring views along camera trajectories. These distinctions clarify where feedback occurs and which representation is actually updated. Feedback allows generation to adapt to the reconstruction state, but it also creates additional paths for error propagation. Incorrect geometry can produce unreliable depth, visibility, or spatial correspondences that bias restoration. Subsequent reconstruction may absorb these errors and reinforce them in later rounds. Agreement between restored images and current renderings is therefore insufficient evidence of accuracy. Reliable feedback requires support from real observations and cross-view geometry, together with mechanisms for rejecting inconsistent information, adjusting its influence, and controlling update frequency. Representative methods are compared in terms of supervision, three-dimensional representation, recovery procedure, and interaction pattern. The review also summarizes commonly used object-level and scene-level datasets and discusses image quality, perceptual similarity, and geometric accuracy. Evaluation measures include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), Chamfer Distance, and F-score. Directly generated novel views and views rendered from a reconstructed representation are treated as different evaluation outputs. Differences in input-view selection, pose assumptions, scene coverage, test splits, and access to geometric ground truth limit comparisons across studies. Improved rendering quality should therefore be examined alongside geometric accuracy and cross-view consistency, rather than interpreted as sufficient evidence of successful reconstruction. Future research should strengthen the coordination between diffusion priors, observed images, and geometric constraints. Prior strength and generated supervision should account for observation coverage, uncertainty, and reconstruction progress. Feedback should be evaluated against observational and geometric evidence that is independent of the current restoration result. Robustness to pose errors, complex topology, reflective or transparent materials, and distribution shifts remains essential for practical applications. More informative evaluation settings should distinguish scene fitting from generalization and report the costs of diffusion sampling, three-dimensional recovery, and repeated feedback separately. Effi-

© 中国图象图形学报版权所有

cient sampling, diffusion distillation, selective generation, and feedforward reconstruction offer complementary approaches to reducing runtime and memory consumption. Progress along these directions can improve geometric reliability while balancing reconstruction quality, generalization, and computational efficiency.

Key words: sparse-view 3D reconstruction; diffusion models; generative priors; generated observations; diffusion feedback

论文引用格式:DOI:10. 11834/jig. 260539

0 引　言

从二维图像中恢复三维场景是计算机视觉与计算机图形学中的基础问题。随着机器人感知、增强现实、自动驾驶和数字内容生成等应用领域不断拓展,三维重建逐渐由受控环境走向复杂的非受控场景(Han等,2026)。受场景遮挡、拍摄位置、设备条件和采集成本等因素影响,实际采集通常难以获得视角覆盖充分的图像数据。当输入视角较少时,图像间的重叠范围缩小,几何约束减弱,重建结果容易出现结构缺失、细节模糊和跨视角不一致等问题。因此,如何利用少量图像恢复结构完整且跨视角一致的三维表示,已成为三维视觉领域的重要研究问题。

现有稀疏视图三维重建主要建立在传统几何方法、神经辐射场和三维高斯泼溅等技术基础上。运动恢复结构(Structure from Motion,SfM)通过图像特征匹配估计相机位姿并恢复稀疏结构(Schönberger等,2016),多视图立体视觉(Multi-View Stereo,MVS;Yuan等,2026)则在相机参数已知的基础上利用图像间的对应关系恢复稠密几何(Furukawa等,2015)。这类方法依赖充分的视角覆盖和可靠的图像对应关系,在输入视角较少或图像重叠不足时容易出现重建退化。神经辐射场(Neural Radiance Fields,NeRF)(Mildenhall等,2020)以隐式神经网络表示场景的密度和辐射信息,三维高斯泼溅(3D Gaussian Splatting,3DGS)(Kerbl等,2023)则采用显式高斯基元表示场景,提高了优化与渲染效率。尽管NeRF和3DGS改进了场景表示与渲染方式,但并未增加可用观测。视角覆盖不足时未观测区域的结构和外观仍缺少直接约束。

扩散模型能够从大规模图像数据中学习外观、语义及结构的统计规律,其生成的新视角图像、深度图或法线可以作为补充观测与真实图像共同约束三维表示。学习到的图像分布先验还可以通过得分蒸馏或分布匹配转化为优化信号引导三维表示更新。这些信息补充了稀疏输入中缺失的监督,能够缓解视角覆盖不足造成的重建欠约束问题。但扩散模型学习的仍是二维图像分布,生成内容在视觉上合理并不意味着其符合真实场景的三维几何,在相机投影、遮挡关系和场景尺度等方面仍可能存在偏差。因此,如何将二维生成先验转化为受真实观测约束的三维优化信号,兼顾视觉合理性与几何准确性,是扩散模型用于稀疏视图三维重建需要解决的核心问题。

近年来,针对稀疏视图三维重建及扩散模型在三维视觉领域应用的综述研究逐渐增多。已有工作分别从稀疏视图重建方法以及扩散模型驱动的三维生成等角度,对相关研究进行了系统梳理(Younis等,2025;Wang等,2025a)。然而,现有综述多围绕重建方法、三维表示或生成任务组织内容,对于扩散模型如何以不同形式参与稀疏视图三维重建,以及不同扩散驱动机制之间的联系与差异尚缺乏系统总结。本文通过Web of Science、IEEE Xplore、ACM Digital Library、SpringerLink、OpenReview和arXiv等平台,检索截至2026年8月4日公开的相关文献,并对CVPR、ICCV、ECCV、NeurIPS和ICLR等主要会议论文集进行补充核查。重点纳入以少量观测图像为条件,利用扩散先验恢复或优化静态三维表示的研究。一般性的单图三维生成及仅输出二维新视角图像的工作仅作为技术基础,不纳入核心方法与实验比较。对于视频扩散辅助方法,重点讨论将连续视角序列作为中间信息、最终用于恢复静态三维表示的研究,不涉及动态场景重建。同一工作同时存在预印本与正式发表版本时,优先引用正式版本;尚未正式发表的预印本,仅在正文中用于补充相关技术与发展趋势,不纳入表1及表3—表5的核心比较。整体框架如图1所示,主要工作如下:

1)从扩散信息的使用方式出发,梳理扩散先验约束与生成观测驱动的重建方法,并从生成与重建

© 中国图象图形学报版权所有

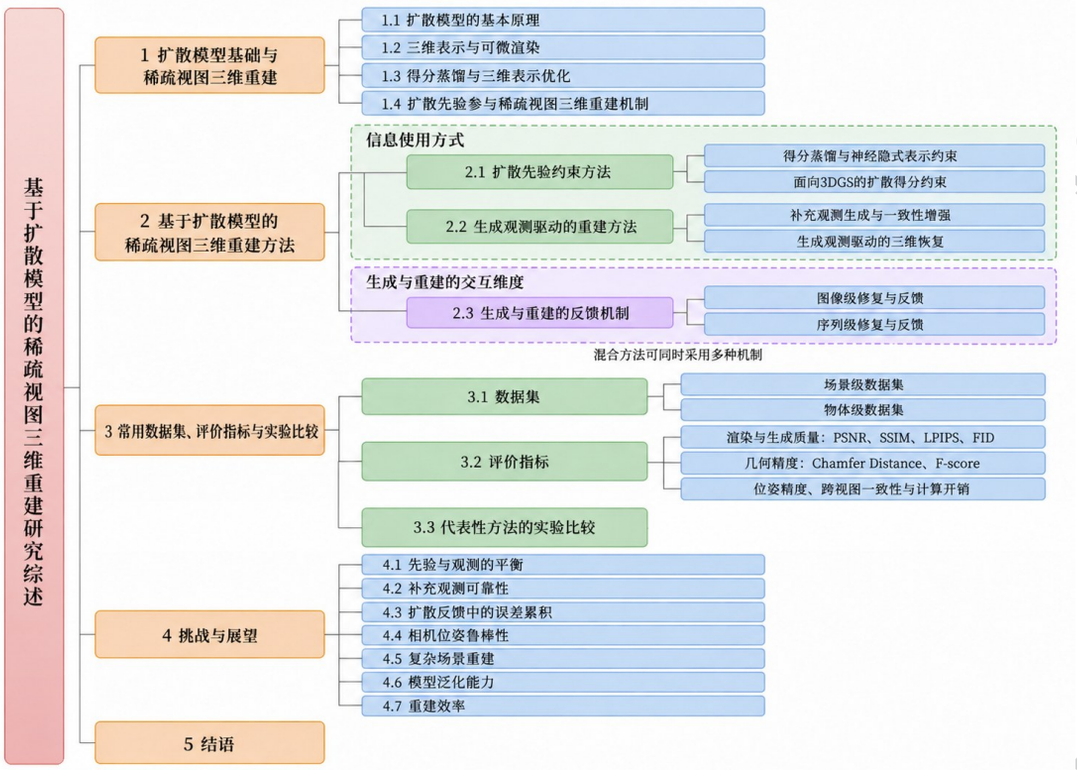


图1　本文结构框架图

Fig. 1　Organization of This Paper

的交互关系出发分析反馈机制，明确不同机制在混合方法中的作用环节及其联系。

2）从先验使用方式、监督形式、三维表示和重建过程等方面比较代表性方法，并归纳该任务常用的数据集与评价指标。

3）结合不同机制的误差传播与失效情形，讨论先验与观测的平衡、补充观测可靠性和反馈误差累积等问题，并分析相机位姿鲁棒性、复杂场景重建、模型泛化能力与计算效率等方面的研究方向。

1　扩散模型基础与稀疏视图三维重建

1.1　扩散模型的基本原理

扩散模型是一类通过正向扩散和反向去噪学习数据分布的生成模型。Sohl-Dickstein等人（2015）受非平衡热力学启发，提出了扩散概率模型。该模型在正向过程中逐步向样本中加入噪声，使数据分布逐渐接近预设的噪声分布；在反向过程中学习逐步去噪，从噪声中生成数据样本。Ho等人（2020）在此基础上提出去噪扩散概率模型（Denoising Diffusion Probabilistic Model，DDPM），利用神经网络预测各时间步样本中的噪声，并通过迭代去噪完成样本生成。其训练与生成过程如图2所示。

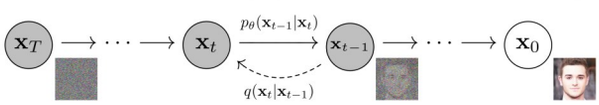


图2　扩散模型的正向加噪与反向去噪过程图（Ho等，2020）

Fig. 2　Forward diffusion and reverse denoising processes of diffusion models（Ho et al.，2020）

正向扩散按照预设的噪声调度，逐步向真实图像中加入高斯噪声。设真实图像为$\boldsymbol{x}_0$，第t个时间步的加噪结果为$\boldsymbol{x}_t$，则由$\boldsymbol{x}_{t-1}$到$\boldsymbol{x}_t$的正向转移概率可表示为：

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})=\mathcal{N}\left(\boldsymbol{x}_t;\sqrt{\alpha_t}\,\boldsymbol{x}_{t-1},(1-\alpha_t)\boldsymbol{I}\right) \tag{1}$$

式中，$\alpha_t=1-\beta_t$，β_t表示第t个时间步的噪声强度。进一步定义α_t的累积乘积为：

$$\bar{\alpha}_t=\prod_{s=1}^{t}\alpha_s \tag{2}$$

利用高斯分布的重参数化性质，可以由真实图像$\boldsymbol{x}_0$直接采样得到任意时间步t的加噪样本$\boldsymbol{x}_t$：

© 中国图象图形学报版权所有

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon,\quad \epsilon\sim\mathcal{N}(0,\boldsymbol{I}) \tag{3}$$

式中，ϵ表示从标准高斯分布中采样的随机噪声。随着t增大，$\boldsymbol{x}_t$中保留的原始图像信息逐渐减少，其分布最终接近标准高斯分布。

反向扩散以噪声样本为起点，通过逐步估计并去除噪声生成数据。在条件信息c的引导下，由$\boldsymbol{x}_t$到$\boldsymbol{x}_{t-1}$的反向转移概率可表示为：

$$p_\phi\left(\boldsymbol{x}_{t-1}\middle|\boldsymbol{x}_t,c\right) = \mathcal{N}\left(\boldsymbol{x}_{t-1};\mu_\phi\left(\boldsymbol{x}_t,t,c\right),\sigma_t^2\boldsymbol{I}\right) \tag{4}$$

式中，ϕ表示扩散模型参数，μ_ϕ和σ_t^2分别表示反向过程的均值和方差。实际训练中，扩散模型通常采用噪声预测网络ϵ_ϕ估计加噪样本$\boldsymbol{x}_t$中包含的噪声。其简化训练目标为：

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{x_0,c,\epsilon,t}\left[\left\|\epsilon-\epsilon_\phi(\boldsymbol{x}_t,t,c)\right\|_2^2\right] \tag{5}$$

训练完成后，噪声预测结果可进一步转换为条件分布得分函数的估计：

$$\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t\middle|c) \approx -\frac{\epsilon_\phi(\boldsymbol{x}_t,t,c)}{\sqrt{1-\bar{\alpha}_t}} \tag{6}$$

得分函数表示加噪样本分布的对数概率密度关于$\boldsymbol{x}_t$的梯度，反映样本向条件分布高概率区域变化的趋势。预训练扩散模型通常用于迭代去噪生成图像，其噪声预测结果还可以转换为图像分布的得分信息，作为生成先验参与其他可微模型的优化。

在条件生成中，观测图像、文本、相机位姿、深度或法线等信息均可作为条件c。为增强生成结果对条件信息的响应，部分方法采用无分类器引导（Classifier-Free Guidance，CFG）（Ho等，2022），其噪声预测结果为：

$$\hat{\epsilon}_\phi(\boldsymbol{x}_t,t,c) = \epsilon_\phi(\boldsymbol{x}_t,t,\varnothing) + s(\epsilon_\phi(\boldsymbol{x}_t,t,c) - \epsilon_\phi(\boldsymbol{x}_t,t,\varnothing)) \tag{7}$$

式中，$\varnothing$表示无条件输入，s表示引导尺度，$\hat{\epsilon}_\phi$表示经过无分类器引导后的噪声预测结果。当$s=1$时，式（7）退化为普通的条件噪声预测。增大s可以强化生成结果对条件信息的响应，但引导尺度过大可能降低生成多样性，并造成局部失真。

1.2 三维表示与可微渲染

NeRF通过神经网络表示连续的三维场景，将空间位置与观察方向映射为体密度和颜色。渲染时，沿相机射线采样空间点，根据各采样点的密度计算透射率及颜色贡献，再通过体渲染累积得到像素颜色。由于该过程可微，渲染图像与真实观测之间的误差能够反向传播至网络参数，实现对场景几何和外观的联合优化。在稀疏视图条件下，已有观测只能约束部分射线，未观测区域的密度与颜色仍具有较大的不确定性。

3DGS采用分布于三维空间的各向异性高斯基元显式表示场景，每个基元包含位置、协方差、不透明度和球谐系数等参数，分别描述其空间分布及随观察方向变化的颜色。渲染时，将高斯基元投影至图像平面，按照深度顺序进行α混合，得到像素颜色。可微光栅化使图像损失能够反向传播至高斯参数；优化过程中，还通过克隆、分裂和裁剪调整基元的数量与分布。与NeRF相比，3DGS采用显式基元和光栅化渲染，但在观测不足时，同样容易出现几何缺失、基元分布不合理和漂浮伪影。

NeRF与3DGS虽然采用不同的场景表示和渲染方式，但均可通过可微渲染接收图像空间的监督。扩散先验直接参与优化时，其提供的梯度信号可经渲染过程传递至NeRF网络参数或高斯参数；扩散模型用于补全或修复观测时，生成结果则与渲染图像构成监督损失，引导三维表示更新。对于前馈重建方法，上述监督也可用于训练预测三维表示的网络，而不必在测试阶段逐场景优化。由此，可微渲染建立了扩散生成信息与三维表示之间的联系。

1.3 得分蒸馏与三维表示优化

扩散模型的得分函数定义在二维图像空间，而三维重建的优化对象是场景的几何与外观参数，两类量之间不存在直接的梯度通路。针对这一问题，DreamFusion（Poole 等，2023）提出得分蒸馏采样（Score Distillation Sampling，SDS），将预训练的文本到图像扩散模型用作三维生成的先验。该方法通过可微渲染将二维扩散模型提供的得分信息传递至三维表示，使随机视角下的渲染结果逐步满足文本条件，实现了二维扩散先验对三维表示的优化。

设三维场景表示参数为θ，相机参数为π，对应的渲染图像$\boldsymbol{x}_\pi$可表示为：

$$\boldsymbol{x}_\pi = \mathrm{R}(\theta,\pi) \tag{8}$$

按照式（3）向渲染图像$\boldsymbol{x}_\pi$中加入高斯噪声，可得：

$$x_t = \sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_\pi + \sqrt{1-\bar{\alpha}_t}\,\epsilon,\quad \epsilon\sim\mathcal{N}(0,\boldsymbol{I}) \tag{9}$$

随后，将$\boldsymbol{x}_t$输入预训练扩散模型，并根据条件c预测其中的噪声。在计算SDS梯度时，不对预训练噪声预测网络进行反向传播，而是将噪声预测残差

© 中国图象图形学报版权所有

通过可微渲染传递至三维表示。将时间步相关系数并入权重$w(t)$后,其梯度可表示为:

$$\nabla_\theta \mathcal{L}_{\mathrm{SDS}} = \mathbb{E}_{t,\epsilon,\pi}\left[w(t)\left(\hat{\epsilon}_\phi(\boldsymbol{x}_t,t,c) - \epsilon\right)\frac{\partial \boldsymbol{x}_\pi}{\partial\theta}\right] \tag{10}$$

式中,$w(t)$表示与时间步相关的权重,ϵ表示式(9)中实际加入的随机噪声,$\hat{\epsilon}_\phi(\boldsymbol{x}_t,t,c)$表示经过条件引导后的噪声预测结果。优化过程中,扩散模型参数ϕ保持不变,噪声预测残差经可微渲染传递至三维表示参数θ。

SDS根据预训练扩散模型的噪声预测结果与实际加入噪声之间的差异构造优化梯度,但较大的引导尺度可能导致颜色过饱和、纹理过度平滑和生成多样性不足。ProlificDreamer(Wang 等,2023)提出变分得分蒸馏(Variational Score Distillation,VSD),将三维表示参数建模为概率分布,并根据预训练扩散分布与当前渲染分布之间的得分差异更新三维表示,以减弱优化结果向单一高概率模式集中的倾向。

1.4 扩散先验参与稀疏视图三维重建机制

稀疏视图三维重建旨在根据少量观测图像恢复对应的三维场景。设第i幅观测图像为$\boldsymbol{I}_i$,对应相机参数为$\boldsymbol{\pi}_i$,则由N幅真实图像构成的观测集合可表示为:

$$\boldsymbol{D}_{\mathrm{obs}} = \left\{\left(\boldsymbol{I}_i,\boldsymbol{\pi}_i\right)\right\}_{i=1}^{N} \tag{11}$$

基于真实观测的重建损失可表示为:

$$\mathcal{L}_{\mathrm{obs}} = \sum_{i=1}^{N} \mathrm{d}\left(\mathrm{R}\left(\theta,\boldsymbol{\pi}_i\right),\boldsymbol{I}_i\right) \tag{12}$$

式中,$\mathrm{d}(\cdot)$表示渲染图像与观测图像之间的差异,具体可采用像素损失、结构相似性损失或感知损失。该损失能够约束已有视角下的渲染结果,但在输入视角较少时,遮挡区域和视角覆盖范围之外的场景内容仍缺少有效监督。

扩散模型从大规模图像数据中学习图像分布,可以为未观测区域的恢复提供生成先验。对于利用扩散先验约束三维表示的逐场景优化方法,其优化目标可概括为:

$$\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{obs}}(\theta) + \lambda\mathcal{L}_{\mathrm{prior}}(\theta) \tag{13}$$

式中,$\mathcal{L}_{\mathrm{prior}}$表示扩散先验约束,$\lambda$用于调节其与真实观测约束之间的相对权重。该约束可表现为显式正则项,也可如SDS一样,根据噪声预测残差构造梯度,并通过可微渲染更新三维表示。因此,式(13)中的先验项用于概括扩散模型的约束作用,不一定对应实际计算的标量损失。

2 基于扩散模型的稀疏视图三维重建方法

扩散模型为稀疏视图三维重建提供生成先验,其作用可从信息使用方式和生成与重建的交互关系两个维度分析,如图3所示。第2.1节和第2.2节分别讨论扩散先验约束与生成观测驱动,第2.3节进一步分析渲染修复及反馈机制。

本文依据各阶段的优化目标和数据流说明方法所采用的机制,对混合方法分别标明其作用环节。在交互分析中,区分一次修复后固定使用辅助观测的流程,以及更新后的三维状态再次参与生成的多轮反馈。因此,三个小节承担不同层面的机制分析,不构成三个互斥的方法类别。

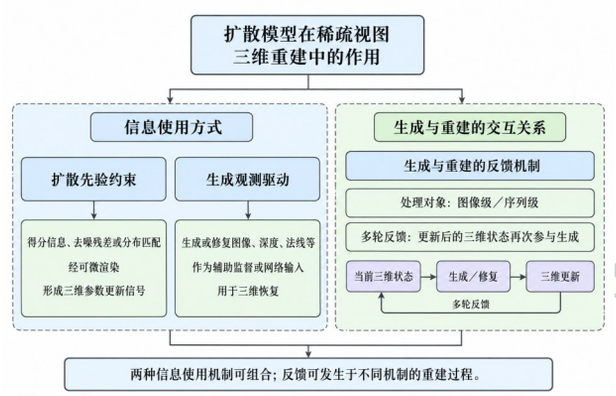


图3 扩散模型参与稀疏视图三维重建的作用方式

Fig. 3 Roles of Diffusion Models in Sparse-View 3D Reconstruction

2.1 扩散先验约束方法

扩散先验约束方法通过可微渲染,将得分信息、去噪残差或分布匹配目标转化为三维参数的更新信号,并结合真实图像的重建损失约束场景。部分方法在优化中生成目标视图估计,用于构造当前迭代的先验约束。本节分别讨论面向神经隐式表示和3DGS的相关方法。

2.1.1 得分蒸馏与神经隐式表示约束

扩散模型提供的得分信息可通过可微渲染形成对神经隐式表示的约束。DiffusioNeRF(Wynn 等,2023)是较早将此类先验用于稀疏视图NeRF重建的方法。该方法利用Hypersim数据集(Roberts 等,

© 中国图象图形学报版权所有

2021)中的RGB-D图像块训练去噪扩散模型，学习颜色与深度的局部联合分布。在NeRF优化过程中，首先从随机视点渲染RGB-D图像块，再由扩散模型估计其对数概率梯度，并通过可微渲染更新辐射场的颜色和密度参数。与依赖文本条件的SDS不同，DiffusioNeRF从外部RGB-D数据中提取几何与外观先验，能够在颜色约束之外引入局部深度信息。不过，由于训练和约束均以局部图像块为单位，该方法难以描述跨区域的遮挡关系，对场景整体结构的约束仍较弱。

为使扩散先验与目标场景建立更直接的联系，SparseFusion(Zhou等，2023)沿极线聚合多幅输入图像特征，结合相对相机位姿训练视角条件扩散模型，如图4所示。逐实例重建时，以Instant-NGP(Müller等，2022)为场景表示，在随机目标视点渲染当前图像，经编码加噪后由扩散模型逐步去噪得到目标视图估计，再根据该估计与当前渲染的差异更新场景参数。这一过程不将单次生成结果固定为伪监督，而是在优化中反复从视角条件分布中寻找与当前场景一致的高概率结果，作者称为神经模式寻优(Neural Mode Seeking)。SparseFusion主要在CO3D上训练并要求相对位姿已知，其泛化范围仍受到训练类别分布和位姿条件的限制。

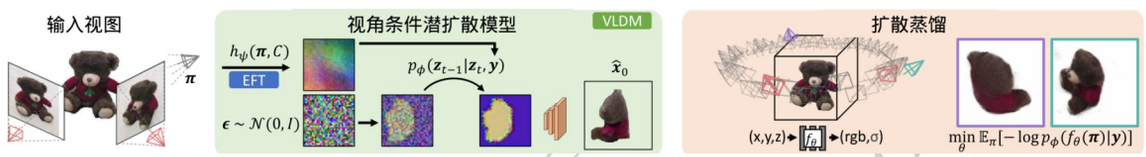


图4　SparseFusion流程框架图(Zhou等，2023)

Fig. 4　Framework of SparseFusion(Zhou et al. ,2023)

在SparseFusion基础上，Sparse3D(Zou等，2024)同样利用极线特征建立输入图像与目标视图之间的几何对应，但不再重新训练完整的图像扩散模型。该方法冻结Stable Diffusion(Rombach等，2022)的参数，仅训练特征渲染器和极线控制模块，将输入图像中的场景特征注入扩散过程。为协调场景信息与类别先验，Sparse3D提出类别得分蒸馏采样，根据多视图一致扩散模型与类别文本条件下Stable Diffusion的噪声预测差异构造优化信号，使重建结果受到输入图像约束的同时，能够利用类别先验补充未观测区域。相比主要从特定类别数据中学习目标视图分布的SparseFusion，Sparse3D保留了预训练Stable Diffusion的图像生成先验，降低了对训练类别分布的依赖。

针对输入相机位姿未知的物体重建，UpFusion(Nagoor Kani等，2024)以首幅图像建立参考坐标系，利用UpSRT将多幅输入图像聚合为集合潜表示，并结合目标视点的视图对齐特征作为扩散条件。三维恢复阶段，该方法以Instant-NGP为对象表示，通过得分蒸馏使不同视点的渲染结果与条件图像分布相匹配。扩散模型在此过程中提供分布层面的优化信号，单次生成图像不会被固定为辅助观测；该方法无需显式估计相机位姿，但仍需逐对象迭代优化。

在物体级重建之外，ReconFusion(Wu等，2024)将视角条件扩散先验用于前向拍摄和360°环绕拍摄的真实场景。该方法利用pixelNeRF(Yu等，2021)式网络提取目标视点对齐特征，并结合CLIP(Contrastive Language - Image Pre-training)(Radford等，2021)图像特征构成扩散条件。重建时，当前NeRF的目标视点渲染经编码加噪和条件去噪得到目标视图估计，据此构造采样损失与真实输入的重建损失共同优化NeRF。该方法把条件扩散模型转化为随机目标视点下的优化信号，目标视图估计随渲染状态动态更新，但由于各目标视图仍分别生成，跨视点不一致也可能传递至三维表示。

针对复杂室内场景中的对象遮挡与观测不足，DP-Recon(Ni等，2025)将室内场景分解为多个对象和背景并分别优化相应的神经隐式表示。重建过程分为几何优化和外观优化两个阶段：前者对各对象的法线图和掩膜图施加SDS约束，后者利用SDS约束颜色渲染结果。该方法还通过可见性网格估计各区域在输入视图中的可见程度，动态调整SDS损失与重建损失的相对权重。使低可见或被遮挡区域更多依赖扩散先验，观测充分的区域则主要由真实图像约束。

ZeroNVS(Sargent等，2024)将单图条件新视角

© 中国图象图形学报版权所有

生成扩展至包含物体、室内和室外场景的混合数据分布，采用尺度感知的相机条件描述视点变化。该方法主要面向单图场景生成，其相机条件设计为稀疏场景重建提供了技术基础。

2.1.2　面向3DGS的扩散得分约束

在神经隐式表示之外，SDS也被用于3DGS优化，相关研究随后从未观测区域约束、得分匹配校正以及三维表示与相机位姿联合优化等方面展开改进。

DreamGaussian（Tang等，2024a）虽主要面向文本或单幅图像驱动的三维生成，但将SDS由NeRF扩展至3DGS，为后续扩散先验驱动的稀疏视图3DGS重建提供了基础。在此基础上，SparseGS（Xiong等，2025）将扩散得分用于少量真实图像条件下的3DGS优化。在输入视角下通过图像重建损失及渲染深度与单目估计深度的Pearson相关性约束外观和几何；对邻近输入相机的虚拟视点利用单目深度重投影提供监督；对较远视点则引入SDS补充未观测区域的约束。SparseGS还通过比较α混合深度与模式选择深度，检测并删除漂浮高斯。为减小扩散先验带来的偏差，该方法将SDS权重设置得较低，因而对远距离未观测区域的约束较有限。

针对直接使用SDS可能引起的模式偏移，IPSM-Gaussian（Wang等，2024a）从分布匹配角度重新设计扩散约束。该方法利用相机位姿和渲染深度，将已观测图像重投影至邻近虚拟视点，结合有效区域掩膜构造场景内先验作为当前渲染分布与扩散目标分布之间的中间校正分布，将原有SDS目标分解为两个匹配过程：先使渲染分布接近校正分布，再使校正分布接近扩散先验中的目标模式；深度正则和几何一致性约束分别用于提高场景内先验的可靠性并限制渲染结果与校正结果的偏差。相比SparseGS降低SDS权重的做法，IPSM-Gaussian直接调整了得分匹配的优化路径。

SparseGS与IPSM-Gaussian通常以已知输入相机位姿为前提，而有限的图像重叠增加了准确位姿估计的难度。针对这一问题，SparseAGS（Zhao等，2024）面向位姿未知的稀疏输入，聚合不同输入图像条件下的扩散噪声预测，并以六自由度相对外参和横、纵向焦距比例表示相机参数，从而同时描述相机位姿及图像裁剪引起的尺度变化。优化过程中，多视角SDS用于更新高斯表示，图像重建误差则反向传播至相机参数，实现3DGS与相机位姿的联合优化。对于初始误差较大的视图，SparseAGS先将其识别并暂时排除，再通过离散相机搜索和连续优化重新估计相机位姿。

2.2　生成观测驱动的重建方法

生成观测驱动的方法利用扩散模型生成或修复RGB图像、深度图、法线图及连续视角序列，并将这些信息或其特征用于三维恢复。生成结果可以作为逐场景优化的辅助监督，也可以作为预训练重建网络的输入。该机制关注生成信息如何进入三维恢复过程，而不限定其必须在重建开始之前一次性生成。

2.2.1　补充观测生成与一致性增强

利用扩散模型补充输入图像未覆盖的视角，首先需要根据目标相机参数生成对应图像。Zero-1-to-3（Liu等，2023）在预训练扩散模型中引入参考图像和相对相机参数，可根据方位角、俯仰角和相机距离生成指定视角图像，建立了利用条件扩散模型补充目标视角观测的基本路径。但由于各目标视图独立采样，在视角变化较大时仍容易出现结构和纹理不一致。为增强生成视图之间的一致性，Zero123++（Shi等，2023）将六个固定目标视图拼接为统一的图像网格，并在同一扩散过程中联合去噪，使不同视图共享结构和外观信息；SyncDreamer（Liu等，2024c）则在每个去噪时间步同步各视图的中间状态，并通过共享三维特征体建立跨视图空间联系。针对RGB图像难以充分约束表面几何的问题，Wonder3D（Long等，2024）联合生成颜色图和视空间法线图，SV3D（Voleti等，2024）则将图像到视频扩散模型用于环绕视角生成。上述方法主要面向单幅物体图像条件下的新视图生成或三维重建，但其视角控制、多视图联合去噪和几何属性生成机制为稀疏观测补全提供了技术基础。

与由单幅图像生成新视图的方法不同，DreamSparse（Yoo等，2023）以少量输入图像及其相机位姿作为场景条件。其几何模块沿目标视点的相机射线聚合多幅输入图像的特征，生成包含三维信息的特征图和粗略颜色图，再通过空间引导模块将这些特征注入冻结的Stable Diffusion。推理时，该方法对粗略颜色图加噪并以此作为反向扩散的初始状态，从而保留输入图像中的颜色和内容信息。该方法无需针对测试场景进行单独优化，但依赖准确的输入相机位姿。

© 中国图象图形学报版权所有

在多幅稀疏图像条件生成的基础上，MVGenMaster(Cao 等，2025)将度量几何信息显式引入多视图生成。该方法根据深度和相机位姿，将参考图像的 RGB 像素及规范坐标图(Canonical Coordinate Map，CCM)重投影至目标视点以建立像素级空间对应，并采用普吕克坐标(Plücker coordinates)表示相机射线。模型通过跨视角全局注意力处理数量可变的参考视图和目标视图，可在一次联合扩散采样过程中生成多个目标视图。DreamSparse 通过沿目标射线聚合输入特征构建三维条件，MVGenMaster 则利用显式几何重投影确定跨视图对应。

DreamSparse 和 MVGenMaster 均以已知相机位姿为生成条件。针对位姿未知的稀疏输入，NVComposer(Li 等，2025)采用图像—位姿双流扩散框架，在生成目标视图的同时估计输入图像的相机位姿。训练阶段，其几何感知特征对齐模块将扩散模型的内部特征与预训练稠密立体视觉模型生成的三维点图对齐，将多视图几何先验蒸馏至生成模型，使推理过程无需预先进行外部位姿估计或场景重建。

2.2.2　生成观测驱动的三维恢复

扩散模型生成的图像、深度、法线或连续视角序列，还需进一步转化为统一的三维表示。一类方法将其作为辅助观测，通过逐场景优化恢复 NeRF、3DGS 或神经隐式表面；另一类方法则利用预训练重建网络或三维生成模型，从生成视图及其中间特征直接预测网格、隐式表面或高斯基元。部分框架还可切换为得分蒸馏方式，因而同时涉及第 2.1 节所述的扩散先验约束。

iFusion(Wu 等，2025a)采用“分析—合成”策略反演预训练的 Zero-1-to-3，通过最小化目标图像特征与扩散模型去噪特征之间的差异，迭代估计输入图像的相对位姿。完成图像配准后，利用真实输入图像及其估计位姿对 Zero-1-to-3 进行低秩适配(Low-Rank Adaptation，LoRA)(Hu 等，2022)，增强扩散先验与目标对象的匹配程度。三维恢复阶段，适配后的模型既可生成补充视图供后续重建模块恢复三维表示，也可提供得分蒸馏约束与真实观测的重建损失共同优化三维表示，具体取决于所接入的重建模块。该框架适配灵活性，但增加了位姿求解和对象级模型适配的开销。

MVDiffusion++(Tang 等，2024b)面向单幅或少量无位姿物体图像，通过不同视图潜特征之间的自注意力隐式建模空间关系，生成预设视点下的多幅高分辨率图像。训练时随机丢弃部分目标视图以降低显存开销，推理时通过联合扩散采样获得补充视图，无需针对当前对象适配扩散模型。生成图像随后作为 NeuS(Wang 等，2021)的输入，通过逐对象优化恢复物体表面。

SpaRP(Xu 等，2024)面向无位姿稀疏物体图像，同时估计输入相机位姿并恢复纹理网格。该方法将输入图像排列为图像网格，利用二维扩散模型预测归一化对象坐标空间图(Normalized Object Coordinate Space，NOCS)和固定相机布局下的目标视图。NOCS 图建立二维像素与三维点之间的对应关系，在相机内参已知时，可通过求解透视 n 点(Perspective-n-Point，PnP)问题估计相机位姿。生成的目标视图输入预训练的两阶段三维扩散模型，由低分辨率占据场逐步恢复带颜色的符号距离场并提取纹理网格。所得网格还可通过可微渲染校正相机位姿。该方法减少了测试阶段的逐对象三维优化时间，但主要适用于前景已分离的物体图像。

CAT3D(Gao 等，2024)将补充观测驱动的重建扩展至不同类型的物体和真实场景。该方法以输入图像及其相机位姿为条件，通过跨图像注意力联合生成多个指定视点的图像，并将位置邻近的目标视点分组生成以获得较为密集的补充观测。真实图像与生成图像随后共同用于 Zip-NeRF(Barron 等，2023)优化。为降低生成视图中局部不一致对重建的影响，CAT3D 在光度损失之外加入感知损失，并根据生成视点与最近真实视点之间的距离动态调整监督权重，使邻近真实观测的生成图像在训练后期发挥更大作用。ReconFusion 在 NeRF 优化过程中反复利用扩散模型生成目标视图估计，CAT3D 则先集中生成密集的辅助视图再交由已有重建模型恢复场景。

S2D(Ji 等，2026)针对极少视图下 3DGS 远离输入视点时伪影严重的问题，引入视觉基础模型预测的点云作为结构引导。方法先用稀疏输入初始化 3DGS；优化至预设中间阶段时，在全部目标相机下渲染点云与当前 3DGS，由单步扩散修复模型结合邻近输入图像纹理、点云渲染结构和退化 3DGS 渲染，生成修复目标视图，再与真实输入共同优化。随机样本丢弃平衡两类视图采样比例，基于点云可见性置信掩膜的梯度加权降低不可靠区域影响。与反复

© 中国图象图形学报版权所有

调用扩散模型的反馈方法不同，S2D仅在中间阶段执行一轮修复，后续将修复结果作为固定辅助观测。

CAT3D和S2D通过生成若干目标视点图像补充离散观测，ReconX（Liu等，2026）则利用视频扩散模型沿相机轨迹生成连续观测序列。对于无位姿输入，该方法先借助DUSt3R（Wang等，2024b）恢复全局对齐点云，并将其编码至三维上下文空间作为结构条件，模型沿相机运动轨迹生成静态场景的连续视频帧并用于优化3DGS，整体流程如图5所示。ReconX还根据DUSt3R输出的置信图调整不同区域的监督强度，以减小不可靠生成内容的影响。相较于离散目标视图，视频序列能够提供更连续的视点覆盖。

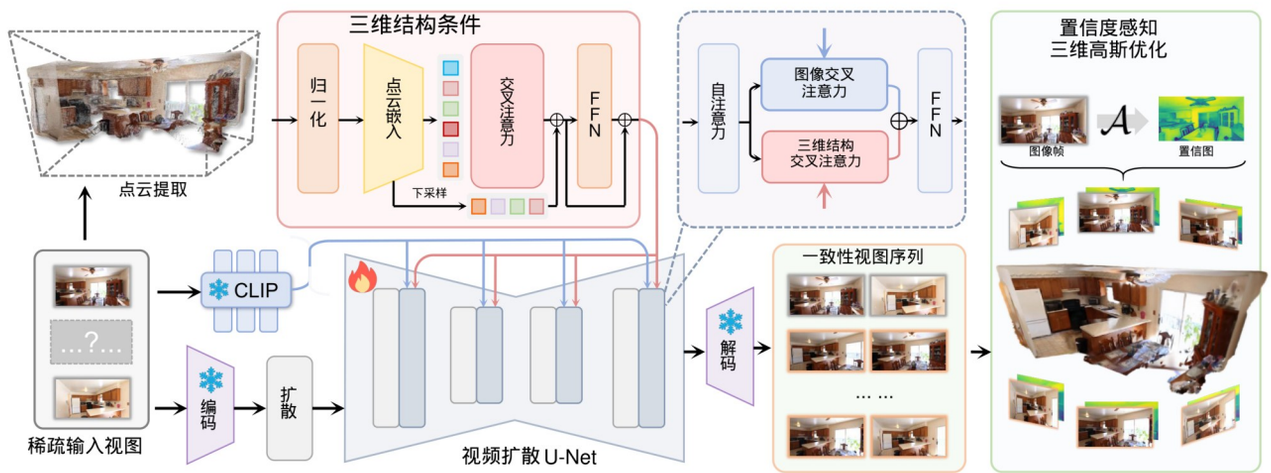


图5　ReconX总体框架图（Liu等，2026）

Fig. 5　Framework of ReconX（Liu et al.，2026）

SpatialCrafter（Zhang等，2025）将连续视角生成与前馈式3DGS预测整合在同一框架中。在视频生成阶段，该方法利用可训练相机编码器注入目标相机信息，控制视频扩散模型沿给定轨迹生成图像，并通过极线注意力将跨帧特征交互限制在符合极线几何的位置。统一尺度估计用于校准不同数据集中的相机运动尺度，使模型能够联合利用多源数据进行训练。在三维恢复阶段，SpatialCrafter融合视频潜空间中的语义特征、相机射线编码和单目深度，再利用结合选择性状态空间模型Mamba（Gu和Dao，2024）与Transformer的混合网络处理长序列特征，直接预测高斯基元。ReconX将生成序列作为辅助观测进行逐场景3DGS优化，SpatialCrafter则从生成序列的特征和深度信息中前馈预测3DGS，减少了独立优化过程。

除通过改变相机位置增加目标视图外，部分方法还通过扩展现有图像的视野补充场景信息。GaMO（Huang等，2025）先利用DUSt3R估计初始点云，并在此基础上优化粗略3DGS。该方法保持相机外参不变，通过减小焦距扩大视场角，再由粗略3DGS渲染颜色图和不透明度图。低不透明度区域被标记为待补全部分，粗略颜色图则为多视图扩散外绘提供已有场景的外观和空间布局参照。完成外绘后，GaMO以补全后的宽视场图像和原始输入为监督，继续优化3DGS。面向全景和弱视差采集，PanoPlane（Qureshi等，2026）利用室内布局约束补全360°全景观测，并据此细化高斯场景表示；PanoImager（Xu等，2026）则结合VGGT（Wang等，2025b）几何估计与MVGenMaster生成辅助视图，通过可靠性加权参与3DGS优化。

2.3　生成与重建的反馈机制

当扩散模型以当前三维表示的渲染结果或几何状态为条件时，生成内容会随重建状态发生变化。修复后的图像或序列可用于更新三维表示，更新结果又可提供新的渲染和几何条件。根据处理对象，这类交互主要表现为单个视点的图像级修复和沿相机轨迹的序列级修复；其中，更新后的三维状态再次参与后续生成时，便形成多轮反馈。

2.3.1　图像级修复与反馈

Deceptive-NeRF/3DGS（Liu等，2024b）较早建立

© 中国图象图形学报版权所有

了针对单个目标视点修复退化渲染，并将修复结果反馈至三维优化的基本流程。该方法利用渲染深度将邻近输入图像重投影至目标视点，根据重投影图像与当前渲染之间的颜色差异估计不确定区域，并据此引导扩散模型生成修复后的伪观测。ExtraNeRF(Shih 等，2024)则面向窄基线输入下的视角外推，通过可见性估计定位未观测区域和新显露区域，利用针对当前场景适配的扩散模型补全颜色，再由深度补全网络恢复相应几何。两种方法均根据当前三维表示确定需要修复的内容，但依据不同：Deceptive-NeRF/3DGS利用重投影误差估计渲染不确定性，ExtraNeRF则根据可见性限定颜色与深度补全的范围。

在极少视图物体重建中，扩散模型还需适应具体对象的外观及其渲染缺陷。GaussianObject(Yang 等，2024)将重建过程分为粗略结构优化、对象专属修复模型训练和反馈修复三个阶段，如图6所示。该方法首先根据输入图像掩膜和相机参数构建视觉外壳，并通过清理漂浮高斯获得粗略3DGS；随后，采用留一策略和高斯属性扰动构造“退化渲染—参考图像”对，用于微调对象专属的扩散修复模型。在反馈修复阶段，模型根据新视点与参考视点之间的距离选择待修复视点，并利用修复图像与真实输入共同优化3DGS。

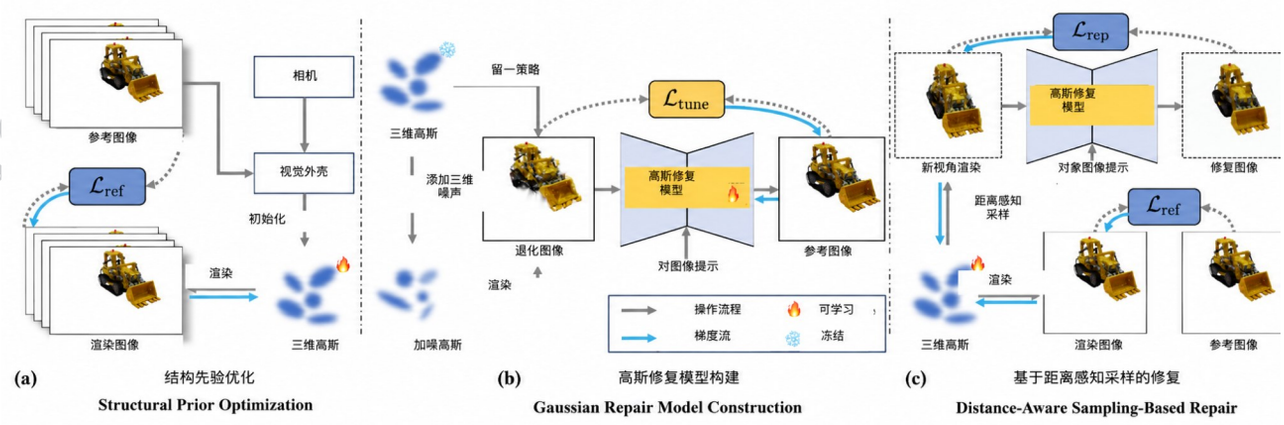


图6　GaussianObject总体框架图(Yang 等，2024)

Fig. 6　Framework of GaussianObject(Yang et al. ，2024)

RI3D(Paliwal 等，2025)进一步将扩散反馈拆分为可见区域修复和未观测区域补全两个任务，分别使用修复模型改善已有观测覆盖区域的渲染质量，并利用图像补全模型生成缺失内容。GaussianObject使用单一对象专属模型修复退化渲染，RI3D则依据区域的可见状态分配不同的生成任务，以减轻修复已有内容与生成缺失内容之间的相互干扰。为降低逐实例微调的计算开销，Difix3D+(Wu 等，2025b)训练了以参考图像为条件的通用单步扩散模型，通过插值相机逐步修复远距离视点下的NeRF或3DGS渲染，并将修复结果反馈至三维表示优化。

FixingGS(Wang 等，2025c)进一步提出连续得分蒸馏，以当前3DGS渲染结果与预训练Difix模型修复结果之间的图像残差更新高斯参数，使扩散反馈随重建状态持续变化。对于修复前后差异较大的不可靠视点，该方法沿邻近输入视点至目标视点的相机路径逐步增加监督，并利用中间视点引导远距离区域的优化。与Difix3D+阶段性生成并回写伪训练视图不同，FixingGS在3DGS优化过程中连续蒸馏当前修复结果，无需另行训练或微调扩散模型，但持续调用Difix仍会增加优化开销。

GIFSplat(Chen 等，2026a)将扩散反馈用于无位姿稀疏输入的前馈式3DGS重建。该方法首先通过高斯初始化器联合预测相机参数和初始3DGS。在每轮更新中，模型从真实输入与当前渲染的特征差异中提取观测反馈；对于新视点渲染，则利用冻结的Difix模型生成增强图像，并将增强前后的特征差异汇聚至相应高斯基元，形成生成先验反馈。两类反馈与当前高斯状态共同输入权重共享的残差更新模块，通过少量前向迭代逐步修正高斯参数。与依赖

© 中国图象图形学报版权所有

测试阶段梯度优化的FixingGS不同，GIFSplat无需反向传播，也不将增强图像持续加入训练视图集，在保留“渲染—修复—更新”反馈过程的同时避免了长时间的逐场景优化。

扩散反馈的可靠性还取决于修复模型能否适应当前场景、参考信息能否正确对齐，以及补充视点是否覆盖未观测区域。GS-GS（Kong等，2025）利用当前3DGS生成的伪视点渲染对预训练扩散模型进行几何感知的场景适配。该方法通过建立跨视角语义对应，引导扩散模型在不同视点间传递场景特征，并在生成视图用于3DGS优化时加入深度正则，以减小外观补全对几何结构的干扰。当目标视点的渲染结果退化严重时，其特征可能导致错误的跨视角检索。针对这一问题，GeoQuery（Cao等，2026）利用预测深度和相机位姿建立目标视图与参考图像之间的几何对应场，从参考图像中提取与目标位置对齐的代理查询，再在对应位置的邻域内检索跨视角特征，从而降低退化渲染特征对检索结果的影响。S2C-3D（Shen等，2026）进一步将扩散反馈与视点规划结合，根据可见性增益生成覆盖未观测区域的相机路径，并将相邻修复图像之间的一致性约束注入扩散采样过程。这样既能逐步扩展生成反馈的空间覆盖范围，也可减少不同视点独立修复造成的内容冲突。

2.3.2　序列级修复与反馈

图像级修复通常分别处理不同视点的退化渲染，难以直接约束连续视角之间的变化。序列级交互将沿相机轨迹获得的渲染组织为视频，利用视频生成模型联合修复相邻视角，再以生成结果细化三维表示。这一过程既可用于阶段性序列增强，也可进一步组织为多轮反馈：只有更新后的三维状态再次参与后续序列生成，才形成持续的生成—重建闭环。这里的视频用于组织静态场景的连续相机视角，不涉及动态场景建模。

ViewCrafter（Yu等，2025）利用少量图像估计初始点云，以沿目标相机轨迹得到的点云渲染序列为条件，利用视频扩散模型补全缺失内容并修复渲染伪影。为扩大视角覆盖，该方法结合相机轨迹规划，将生成视图用于扩展点云，再由更新后的点云引导下一轮生成，形成生成与几何更新的迭代过程。最终，扩展后的点云和生成视图可用于优化3DGS，实现可实时渲染的场景表示。

3DGS-Enhancer（Liu等，2024a）将初始3DGS沿相机轨迹生成的渲染序列输入视频扩散模型，在潜空间中恢复具有视角连续性的特征，并通过时空解码器融合首尾参考图像和原始渲染信息，所得序列再用于细化3DGS。然而，当生成过程缺少与当前场景的显式对齐时，外推和遮挡区域仍可能出现跨帧外观不一致或不存在的内容。针对这一问题，GuidedVD（Zhong等，2025）以当前3DGS的渲染序列作为场景锚点，将场景一致性约束注入视频扩散采样过程，无需重新训练视频扩散模型。GSFixer（Yin等，2026）进一步训练参考视图引导的视频修复模型，分别提取参考图像的二维语义特征和三维几何特征，以约束生成序列的外观与结构。三种方法对序列生成的约束由首尾参考图像逐步扩展至当前场景锚定和语义—几何联合引导。

在序列修复的基础上，GenFusion（Wu等，2025c）进一步将视频生成与三维表示更新组织为闭环。预训练阶段，该方法仅使用视频帧中的局部图像块建立不完整的高斯表示，再沿原相机轨迹渲染含有伪影和缺失区域的RGB-D序列，并以完整视频及其单目深度构成的RGB-D序列作为监督训练视频扩散模型。在零样本泛化阶段，预训练扩散模型保持冻结，当前高斯表示沿新相机轨迹生成RGB-D序列，经扩散模型修复和补全后加入监督集用于继续优化高斯表示。随着新轨迹不断采样，修复序列与高斯表示交替更新，从而形成持续扩展场景观测范围的闭环。在此基础上，GaussFusion（Zhu等，2026）以包含颜色、深度、法线、不透明度和协方差信息的高斯基元缓冲区为条件，通过基于流匹配的视频生成模型修复连续视点渲染，并利用修复序列重新优化3DGS。

SYN3R（Xu等，2025）将稀疏输入视角之间的连续视图生成转化为测试阶段的视频补全任务，直接使用预训练视频扩散模型，无需额外训练或微调。该方法首先在输入视角之间插值相机位姿，再根据当前3DGS渲染的深度，将邻近真实图像重投影至目标视图，得到扩散引导图像。在此基础上，SYN3R结合跨视图光度一致性和几何一致性估计像素级不确定性。高不确定性区域主要由生成先验补全，低不确定性区域则尽可能保留引导图像中的可靠内容。生成的伪视图用于高斯基元致密化和3DGS优化，更新后的三维表示继续参与下一轮引导图像构建，由此形成迭代反馈。与GenFusion需要预先训练

© 中国图象图形学报版权所有

重建驱动的视频扩散模型不同，SYN3R采用通用预训练模型，并根据不确定性自适应调节生成先验与三维引导的作用强度。

VidSplat(Tang 等，2026)将视频扩散先验同时用于二维高斯泼溅(2D Gaussian Splatting，2DGS)(Huang等，2024)初始化和迭代优化。初始化阶段，该方法先利用DUSt3R从稀疏输入估计点云，再沿基于可见性采样的相机轨迹，以点云渲染的RGB图像和掩膜引导视频扩散模型生成连续视图。生成序列中的关键帧与真实输入共同送入DUSt3R，得到覆盖更完整的点云用于初始化2DGS。优化阶段，2DGS渲染的RGB图像和由网格光线追踪得到的可见性掩膜共同引导分阶段去噪：在高噪声阶段强化生成内容与现有几何的一致性，随着噪声降低逐步放松约束，以修复渲染伪影并补全未观测区域。生成视图经置信度加权后加入训练集，用于继续优化和致密化2DGS。

在利用视频扩散模型补充连续观测时，虚拟相机轨迹直接影响未观测区域的覆盖范围。ExploreGS(Kim 等，2025)记录各高斯基元在不同观察方向下的可见情况，并根据候选视点相对于已有观测增加的场景覆盖计算信息增益。该方法优先选择信息增益较高的视点，逐步构建面向低覆盖区域的虚拟相机轨迹，引导视频扩散模型修复相应的渲染序列。修复后的伪观测并非以相同权重参与优化，而是依据其与已有视图的共视程度和像素级感知差异确定监督权重，再用于3DGS细化，以降低不可靠生成内容对重建结果的影响。

与直接统计高斯可见性的ExploreGS不同，G4Splat(Ni 等，2026)从全局平面结构出发规划新视点。该方法利用跨视图对应关系关联并合并各输入视图中的局部平面，得到统一的全局三维平面表示，再结合平面深度与尺度校正后的单目深度构建可见性网格。在此基础上，G4Splat以全局平面为参照评估候选视点的场景覆盖情况，选择能够观察更多未覆盖区域的新视点，并利用可见性掩膜标记遮挡和缺失区域，引导视频扩散模型进行针对性补全。

AnyRecon(Chen 等，2026b)面向输入视图数量和空间分布不固定的稀疏场景，将较长的虚拟相机轨迹划分为多个片段依次处理。每完成一个轨迹片段，生成视图便被整合到全局三维几何记忆中；在处理后续片段时，模型根据空间重叠程度和几何信息增量，从已有视图中选择参考图像，使不同片段恢复的场景信息能够持续传递。针对长轨迹中视角间隔较大、序列连续性较弱的问题，该方法取消视频潜表示的时间压缩，并通过扩散蒸馏与局部窗口稀疏注意力控制长序列的计算开销。与ExploreGS和G4Splat侧重从候选视点中规划高覆盖轨迹不同，AnyRecon进一步协调多个轨迹片段，使已恢复的几何信息能够随分段生成过程逐步扩展。

上述方法在扩散信息进入三维重建流程的位置以及生成与重建的交互方式上存在差异。为避免将信息使用机制与反馈交互混为互斥类别，表1从输入条件、三维表示、扩散信息进入方式和生成-重建交互等方面对代表性方法进行统一比较。

表1　代表性扩散驱动稀疏视图三维重建方法对比

Table 1　Comparison of the Mechanisms of Representative Diffusion-Driven Sparse-View 3D Reconstruction Methods

代表方法	位姿	三维表示	扩散信息进入方式	生成-重建交互方式	关键设计
DiffusioNeRF (Wynn 等，2023)	已知位姿	NeRF	扩散先验约束	无显式反馈	RGB-D联合图像块先验
SparseFusion (Zhou等，2023)	已知位姿	NeRF	扩散先验约束	迭代反馈	极线特征与神经模式寻优
UpFusion (Nagoor Kani 等，2024)	位姿未知	NeRF	扩散先验约束	无显式反馈	集合潜表示与得分蒸馏
ReconFusion (Wu 等，2024)	已知位姿	NeRF	扩散先验约束	迭代反馈	动态目标视图估计
DP-Recon (Ni 等，2025)	已知位姿	神经隐式表示	扩散先验约束	无显式反馈	对象分解与可见性加权

© 中国图象图形学报版权所有

表1续表

代表方法	位姿	三维表示	扩散信息进入方式	生成–重建交互方式	关键设计
SparseGS（Xiong 等，2025）	已知位姿	3DGS	扩散先验约束	无显式反馈	深度约束与漂浮高斯清理
IPSM-Gaussian（Wang 等，2024a）	已知位姿	3DGS	扩散先验约束	无显式反馈	场景内先验与分布校正
SparseAGS（Zhao 等，2024）	位姿未知	3DGS	扩散先验约束	阶段性反馈	三维表示与相机联合优化
iFusion（Wu 等，2025a）	位姿未知	可接入多种表示	两者兼有	无显式反馈	扩散反演、位姿估计与LoRA
MVDiffusion++（Tang 等，2024b）	位姿未知	NeuS表面	生成观测驱动	单次观测生成	稠密多视图联合生成
SpaRP（Xu 等，2024）	位姿未知	纹理网格	生成观测驱动	单次观测生成	NOCS位姿估计与三维扩散恢复
CAT3D（Gao 等，2024）	已知位姿	Zip-NeRF	生成观测驱动	单次观测生成	批量视图生成与距离加权
S2D（Ji 等，2026）	位姿未知	3DGS	生成观测驱动	无显式反馈	点云结构引导与置信加权
ReconX（Liu 等，2026）	位姿未知	3DGS	生成观测驱动	单次观测生成	点云条件视频扩散
SpatialCrafter（Zhang 等，2025）	已知位姿	3DGS	生成观测驱动（潜特征输入）	无显式反馈	视频潜特征直接预测高斯
GaussianObject（Yang 等，2024）	已知位姿	3DGS	生成观测驱动	阶段性反馈	对象专属修复模型
Difix3D+（Wu 等，2025b）	已知位姿	NeRF/3DGS	生成观测驱动	阶段性反馈	通用单步扩散修复
GIFSplat（Chen 等，2026a）	位姿未知	3DGS	生成观测驱动（特征输入）	迭代反馈	观测反馈与生成反馈融合
GenFusion（Wu 等，2025c）	已知位姿	3DGS	生成观测驱动	迭代反馈	RGB-D视频修复闭环
SYN3R（Xu 等，2025）	已知位姿	3DGS	生成观测驱动	迭代反馈	不确定性调节视频补全
VidSplat（Tang 等，2026）	位姿未知	2DGS	生成观测驱动	迭代反馈	视频扩散参与初始化与优化
GSFixer（Yin 等，2026）	已知位姿	3DGS	生成观测驱动	迭代反馈	二维语义与三维几何引导
G4Splat（Ni 等，2026）	位姿未知	2DGS	生成观测驱动	迭代反馈	平面几何与覆盖导向视点规划

注：“扩散信息进入方式”描述扩散先验参与三维恢复的主要形式，同一方法可兼有多种机制。生成观测驱动既包括使用生成图像、深度或法线作为辅助监督，也包括利用其潜特征进行三维预测。生成—重建交互方式依据扩散模型生成的信息与三维重建过程之间是否存在反馈及其更新方式进行判定。“无显式反馈”表示扩散信息参与重建，但重建结果不再反馈至后续生成过程；“单次观测生成”表示先生成观测，再将其用于后续重建，生成过程不根据重建结果更新；“阶段性反馈”表示在特定重建阶段利用生成结果或重建结果进行修正；“迭代反馈”表示生成与重建之间存在多轮信息反馈或更新；各标签以方法的实际处理流程为准。

© 中国图象图形学报版权所有

3 常用数据集、评价指标与实验比较

3.1 数据集

目前，扩散模型驱动的稀疏视图三维重建尚未形成专门且统一的基准数据集，相关研究多沿用新视角合成、多视图重建和三维生成任务的数据，并从已有多视角观测中选取少量图像作为输入。由于不同研究采用的输入视角数量、选取方式和空间分布并不一致，同一数据集上的实验结果也未必能够直接比较。本文按主要重建对象及常见使用方式，将相关数据集分为场景级和物体级两类，具体规模、特点及用途见表2。

表2 稀疏视图三维重建常用数据集

Table 2 Common Datasets for Sparse-View 3D Reconstruction

数据集	数据对象	数据规模与特点	主要用途
LLFF（Mildenhall等，2019）	场景级	常用LLFF-NeRF子集包含8个前向拍摄的真实场景，提供相机位姿及场景边界信息	场景重建与新视角合成
RealEstate10K（Zhou等，2018）	场景级	约8万段视频片段，包含对应约1000万帧的相机位姿，相机轨迹由SLAM和光束平差估计	新视角生成与场景重建模型训练
Replica（Straub 等，2019）	场景级	包含18个高保真室内场景，并提供稠密网格、HDR纹理及语义标注。	室内场景重建与表面恢复
Mip-NeRF 360（Barron等，2022）	场景级	包含5个室外场景和4个室内场景，具有较大的空间范围和复杂背景	无界场景重建与新视角合成
Tanks and Temples（Knapitsch等，2017）	场景级	包含21个室内外场景，以工业级激光扫描结果作为几何评测依据	室内外场景重建与几何精度评价
DL3DV（Ling等，2024）	场景级	包含10510段真实场景视频和约5120万帧图像，覆盖65类室内外拍摄地点	真实场景模型训练与跨场景测试
DTU（Jensen等，2014）	物体级	包含124个受控采集场景，提供标定相机参数和结构光扫描几何真值	物体表面重建与几何精度评价
CO3D（Reizenstein等，2021）	物体级	包含约1.9万个物体视频和150万帧图像，覆盖50个物体类别	物体重建与未见对象测试
Objaverse（Deitke等，2023）	物体级	包含80余万个三维物体模型，覆盖多种物体类别与模型形式	三维生成与重建模型预训练
GSO（Downs等，2022）	物体级	包含1000余件真实扫描物体，具有较完整的几何和纹理信息	物体重建与新视角合成
OmniObject3D（Wu等，2023）	物体级	包含6000个真实扫描物体，覆盖190个类别，提供图像、视频、点云和纹理网格	物体重建、新视角合成与三维生成

场景级数据集主要包括LLFF、RealEstate10K、Mip-NeRF 360、Tanks and Temples、Replica和DL3DV。LLFF（Mildenhall等，2019）由手持相机采集的真实场景构成，并采用运动恢复结构方法估计相机位姿和场景深度边界。NeRF及后续研究通常使用其中8个前向拍摄场景评价已知相机位姿条件下的场景重建与新视角合成性能。Mip-NeRF 360（Barron等，2022）包含5个室外场景和4个室内场景，拍摄范围覆盖场景周围的多个方向，近景与远景之间存在较大的深度和尺度差异，主要用于无界场景重建与新视角合成。Replica（Straub 等，2019）是面向室内空间建模的高保真三维场景数据集，共包含18个房间或建筑尺度场景，覆盖公寓、办公室、住宅和酒店等环境。每个场景提供稠密网格、高动态范围（High Dynamic Range，HDR）纹理、语义类别与实例标注，以及玻璃和镜面等表面信息。Tanks and Temples（Knapitsch等，2017）由真实条件下采集的室内外场景视频构成，并以工业激光扫描结果作为几何评测依据，适合评价大尺度场景的结构恢复精度。RealEstate10K和DL3DV规模较大，主要用于场

© 中国图象图形学报版权所有

景级模型训练。RealEstate10K(Zhou等,2018)包含约8万个视频片段及其对应的相机轨迹,覆盖约1000万帧,相机位姿由同步定位与建图(Simultaneous Localization and Mapping,SLAM)和光束平差估计。该数据集常用于训练新视角生成和场景重建模型。DL3DV(Ling等,2024)包含10510段真实场景视频和约5120万帧图像,覆盖65类室内外拍摄地点,并对反射、透明度和光照条件等场景属性进行标注。与LLFF和Mip-NeRF 360等主要用于逐场景重建测试的数据集相比,RealEstate10K和DL3DV可为训练能够处理未见场景的重建模型提供更多数据。

物体级数据集主要包括以单个物体或受控小型场景为中心采集的DTU、CO3D、Objaverse、Google Scanned Objects(GSO)和OmniObject3D。DTU(Jensen等,2014)在受控环境中利用工业机械臂从固定位置采集图像,并设置多种光照条件,同时提供经过标定的相机参数和结构光扫描结果。其采集流程和视角设置较为统一,常用于评价稀疏输入条件下的物体表面重建精度。CO3D(Reizenstein等,2021)由真实环境中拍摄的物体视频构成,包含近1.9万个视频、150万帧图像和50个物体类别,并提供相机位姿、前景掩膜和三维点云等标注。与DTU相比,CO3D中的物体外观、背景和相机轨迹更加多样,常用于物体新视角合成和类别重建,并测试模型处理训练阶段未见物体的能力。

Objaverse、GSO和OmniObject3D均提供可渲染的三维物体模型,但其数据来源和主要用途有所不同。Objaverse(Deitke等,2023)从网络平台收集了80余万个三维模型,并附有文字描述、标签和动画等元数据。研究者可从指定相机位置渲染多视角图像,用于扩散模型和三维重建模型的预训练。GSO(Downs等,2022)包含1000余件真实日常物体的扫描模型,保留了较完整的几何结构和表面纹理,可通过渲染构造具有准确相机参数和几何信息的多视角数据。OmniObject3D(Wu等,2023)包含6000个真实扫描物体,覆盖190个类别,并提供纹理网格、点云、多视角渲染图像和真实拍摄视频,可用于物体表面重建、新视角合成和三维生成。相比主要用于受控几何评测的DTU,Objaverse更适合大规模模型预训练,Google Scanned Objects和OmniObject3D则能够为训练和测试提供真实扫描几何。

3.2 评价指标

扩散模型驱动的稀疏视图三维重建通常从图像重建质量、感知质量、生成分布质量和几何精度等方面进行评价。图像重建质量主要采用峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)(Huynh-Thu等,2008)和结构相似性指数(Structural Similarity Index Measure,SSIM)(Wang等,2004)衡量,感知质量常用学习感知图像块相似度(Learned Perceptual Image Patch Similarity,LPIPS)(Zhang等,2018)评价。生成分布质量可采用Fréchet Inception Distance(FID)(Heusel等,2017)衡量,几何精度则主要采用CD(Chamfer Distance)(Fan等,2017)和F-score(Tatarchenko等,2019)等指标。

上述指标的解释需要结合具体评价对象和实验协议。扩散模型直接生成的目标视图主要反映二维生成质量,NeRF、3DGS或网格在测试视点下的渲染结果反映三维表示的新视角合成能力,二者不能作为同一评价对象直接比较。PSNR、SSIM、LPIPS和FID也不能单独证明三维几何准确;在具有几何真值的数据集上,还需结合CD、F-score等指标,并说明坐标与尺度对齐方式、表面采样方法及距离阈值。

3.3 代表性方法的实验比较

表3—表5汇总了代表性方法在重建质量、场景几何精度及重建耗时方面的实验结果。表3按数据集和输入视图设置分别列示,表4整理VidSplat论文所报告的TNT与Replica几何重建结果,表5整理G4Splat论文中针对单个场景的耗时与质量比较。各表保留来源论文的评价口径;对于由不同文献报告的结果,仅作为相应实验条件下的性能参考,不将相同数据集名称和输入视图数视为评价协议完全一致的依据,也不据此进行跨表综合排名。

表3表明,在所列数据集与输入设置下,不同方法的渲染质量和几何精度存在差异,其相对表现需结合具体指标理解。表4中,VidSplat在来源论文所采用的TNT与Replica评价协议下取得了较好的几何结果;表5中,G4Splat获得较低的CD和较高的PSNR,但耗时高于所列GenFusion与Difix3D+配置,体现出重建质量与计算开销之间的取舍。上述比较反映的是完整方法在特定条件下的表现,不能将性能差异单独归因于扩散反馈或某一种几何约束。

© 中国图象图形学报版权所有

表3　代表性方法在不同数据集与输入设置下的重建质量

Table 3　Reconstruction Quality of Representative Methods across Different Datasets and Input Settings

/(a)Mip-NeRF 360(Barron等,2022):3、6和9输入视图

方法	3视图	6视图	9视图	均值
ReconFusion(Wu 等,2024)	15.50 / 0.358 / 0.585	16.93 / 0.401 / 0.544	18.19 / 0.432 / 0.511	16.87 / 0.397 / 0.547
CAT3D(Gao 等,2024)	16.62 / 0.377 / 0.515	17.72 / 0.425 / 0.482	18.67 / 0.460 / 0.460	17.67 / 0.421 / 0.486
ReconX(Liu 等,2026)	17.16 / 0.435 / 0.407	19.20 / 0.473 / 0.378	20.13 / 0.482 / 0.356	18.83 / 0.463 / 0.380
SpatialCrafter(Zhang 等,2025)	17.32 / 0.439 / 0.418	19.42 / 0.510 / 0.371	20.19 / 0.536 / 0.345	18.98 / 0.495 / 0.378

注:结果整理自SpatialCrafter(Zhang等,2025)原文表2,保留其中与本文讨论范围相关的方法。单元格内数值依次为PSNR↑、SSIM↑和LPIPS↓;"均值"为3、6和9幅输入视图结果的算术平均值。各项结果沿用来源论文的实验设置。**(b)GSO(Downs 等,2022):6幅输入图像,位姿未知**

方法	F-score↑(%)	PSNR↑	LPIPS↓
iFusion(Wu 等,2025a)	88.5	16.2	0.151
SpaRP(Xu 等,2024)	96.9	19.3	0.123

注:结果整理自SpaRP(Xu等,2024)原文表2,评测对象为GSO中的1030个物体,每个物体使用6幅渲染图像作为输入。预测网格与真实网格经旋转、尺度及ICP对齐,并归一化至单位包围盒,F-score阈值为0.05。PSNR与LPIPS用于评价重建网格的渲染结果。**(c)LLFF(Mildenhall等,2019):6幅输入图像**

方法	三维表示	PSNR↑	SSIM↑	LPIPS↓
DiffusioNeRF(Wynn 等,2023)	NeRF	23.60	0.775	0.235
ReconFusion(Wu 等,2024)	NeRF	24.25	0.815	0.152
IPSM-Gaussian (Wang 等,2024a)	3DGS	23.94	0.818	0.135

注:结果整理自IPSM-Gaussian(Wang等,2024a)论文附录表12,其中DiffusioNeRF与ReconFusion的结果由该文引用自ReconFusion论文;IPSM-Gaussian采用原文报告的三次实验均值。该分表比较采用不同三维表示的扩散先验约束方法,不代表所有方法均在同一实现下重新训练。**(d)Replica(Straub 等,2019):5幅输入图像**

表4　典型扩散反馈方法的场景几何重建精度

Table 4　Geometric Reconstruction Accuracy of Representative Diffusion-Feedback Methods

方法	CD↓(cm)	F-score↑	NC↑	PSNR↑	SSIM↑	LPIPS↓
GenFusion(Wu 等,2025c)	13.05	41.60	69.33	20.14	0.801	0.258
Difix3D+(Wu 等,2025b)	13.71	43.11	65.34	19.42	0.779	0.231
G4Splat(Ni 等,2026)	6.61	65.14	83.98	23.90	0.836	0.199

注:结果整理自G4Splat(Ni等,2026)原文表1,采用该文的场景划分与评价设置。CD以cm计,F-score的距离阈值为5 cm,NC表示表面法向一致性;各项数值保留原文的统计尺度。

4　挑战与展望

扩散模型通过先验约束和生成观测扩展了稀疏视图三维重建的信息来源,生成与重建之间的交互又使这些信息能够随重建状态进行调整。进一步提高重建质量,需要兼顾先验与真实观测的协调、生成信息的可靠性及交互过程的稳定性。其中,得分偏差可能影响三维参数的优化方向,生成观测中的错误可能被重建过程吸收,多轮反馈则可能进一步累积已有偏差。这些问题分别对应信息使用和交互过程中的不同环节,在混合方法中也可能同时出现。此外,相机位姿鲁棒性、复杂场景重建、跨场景泛化与计算效率,仍是实际应用需要持续研究的共性问题。

4.1　先验与观测的平衡

在扩散先验约束过程中,扩散模型提供的得分信息或分布匹配信号通过可微渲染作用于三维参数。然而,扩散先验反映的是训练数据中的统计规律,其引导方向未必符合目标场景的真实结构。当输入观测不足或先验权重过大时,三维表示可能逐渐偏向先验所倾向的外观与形状,造成几何偏移或细节失真;先验权重过低,又难以补充未观测区域的

© 中国图象图形学报版权所有

表4　典型扩散反馈方法的场景几何重建精度

Table 4　Geometric Reconstruction Accuracy of Representative Diffusion-Feedback Methods

方法	TNT CD ↓	TNT F-score ↑	Replica CD ↓	Replica NC ↑	Replica F-score ↑
Difix3D+(Wu 等,2025b)	0.82	0.21	0.70	51.11	4.77
GuidedVD(Zhong等,2025)	0.81	2.91	0.12	67.91	40.45
VidSplat(Tang 等,2026)	0.66	12.80	0.06	88.42	80.79

注:结果整理自VidSplat(Tang等,2026)原文表1。TNT采用6个测试场景,每个场景使用5幅输入图像;Replica采用8个测试场景,每个场景使用10幅输入图像。CD、F-score和NC分别表示倒角距离、几何重建F-score和法向一致性,各项数值保留来源论文的评价尺度。原论文未明确给出上述指标的完整计算参数及数值尺度,故表中保留原始报告值,不另行换算。

表5　代表性方法的重建耗时与质量比较

Table 5　Reconstruction Time and Quality of Representative Methods

/实验设置:Replica(Straub 等,2019)room_0:5幅输入,单张NVIDIA A100

方法	时间 ↓(min)	CD ↓(cm)	PSNR ↑
GenFusion(Wu 等,2025c)	41.4	15.66	16.49
Difix3D+(Wu 等,2025b)	68.7	17.79	12.94
G4Splat(Ni 等,2026)	73.3	8.77	20.26

注:表中重建耗时及质量指标整理自 G4Splat(Ni 等,2026)原文的比较结果。实验采用 Replica 数据集的 room_0 场景,输入5幅图像,使用单张 NVIDIA A100 GPU。耗时数据沿用原论文的报告值;由于原文未明确说明各方法计时的具体起止点及所包含的处理阶段,表中时间仅作为原文实验条件下的参考,不宜据此直接判定不同方法的绝对运行效率。CD 和 PSNR 分别表示倒角距离和峰值信噪比,评价方向分别为越低越好和越高越好;CD以cm计。

约束。固定权重通常难以适应不同区域的观测条件及重建状态的变化。后续研究可根据视角覆盖、重投影误差和优化阶段动态调节先验强度,在观测充分区域优先保持输入一致性,在观测不足区域适度引入生成约束,并抑制与真实观测明显冲突的参数更新。

4.2　补充观测可靠性

生成的RGB图像、深度和法线并非真实采集数据,其可靠性受视角覆盖、遮挡程度和场景结构影响。当生成观测作为辅助监督参与重建时,其中的结构和纹理错误可能被三维模型反复拟合,造成几何偏移、纹理错位或漂浮伪影;当生成观测或其特征输入重建网络时,这些偏差也可能传递至预测的三维表示。尤其在完全未观测区域,生成内容即使视觉上合理,也缺少真实图像的验证。后续研究需要结合跨视角重投影误差、深度与法线一致性以及生成不确定性,对辅助观测进行筛选和加权。对于与真实输入明显冲突的内容,应降低其监督权重或予以剔除;对于缺少观测支持的区域,则应保留不确定性,避免将生成结果直接视为可靠真值。

4.3　扩散反馈中的误差累积

在生成与重建的多轮反馈中,当前三维状态参与观测修复,修复结果又用于更新三维表示。若初始几何存在偏差,错误的深度、可见性或空间对应关系便可能影响生成过程,而生成结果又可能通过后续优化强化原有错误。随着反馈轮数增加,渲染与修复结果可能逐渐趋于一致,却共同偏离真实场景。因此,二者之间的差异减小不能单独作为反馈有效的依据。后续研究应在每轮更新中结合真实观测和跨视角几何关系检验修复结果,对缺少可靠支持的内容降低监督权重,并根据误差变化控制反馈范围和轮数。当某一区域的偏差持续增大时,应停止其反馈更新,避免错误在循环中进一步传播。

4.4　相机位姿鲁棒性

相机内参与位姿共同决定不同图像之间的投影关系,也是扩散条件与三维表示建立空间对应的基础。手持拍摄、网络图像和非受控采集数据可能缺少可靠的相机内外参数,自动估计又容易受到纹理不足、视角间隔和遮挡的影响。错误位姿会造成图像重投影错位,并在三维表示中产生重复表面、漂浮点或结构变形。后续方法需要将位姿估计纳入重建

© 中国图象图形学报版权所有

过程，在优化三维表示的同时检测并修正异常相机。对于偏差较大的输入视图，可通过重新初始化、离散搜索或连续优化恢复其相机状态；对于位姿完全未知的输入，则需要利用跨视角特征对应和当前三维结构交替估计相机参数。

4.5 复杂场景重建

细长结构、复杂拓扑和严重遮挡区域在有限视角下通常缺少足够的几何对应，容易出现表面断裂、局部缺失和纹理模糊。透明或反光表面的观测颜色还会随视角和光照发生变化，无法直接对应表面颜色和几何位置。扩散模型虽可以生成视觉上完整的内容，但无法确定这些内容是否符合真实结构。后续研究可区分几何、材质和光照对图像形成的影响，并在三维表示中引入适用于透明、反射等材质的成像模型。对于真实观测无法确定的区域，还需保留不确定性。

4.6 模型泛化能力

扩散先验和三维重建模型都会继承训练数据中的类别、场景与成像偏差。当测试对象、拍摄设备、光照条件或场景环境与训练数据差异较大时，生成内容和几何预测的误差通常会增加。现有数据集在物体类别、特殊材质、复杂场景和非受控采集条件上的覆盖仍然有限，也缺少统一的跨类别、跨场景和跨设备测试设置。后续研究可联合利用真实多视角视频、扫描物体和合成三维模型，并通过跨域训练或测试阶段适配减少数据分布变化的影响。实验评价也应增加未见类别、未见场景和不同采集设备下的测试，以区分模型对已有数据的拟合能力与实际泛化能力。

4.7 重建效率

扩散模型通常需要多步去噪，NeRF和3DGS等三维表示也可能依赖逐场景优化。两类过程结合后，生成、渲染和参数更新需要反复执行，增加了处理时间和显存占用。部分反馈方法还会多次生成伪观测并重新优化场景，计算开销随视角数量和反馈轮数继续增加。后续研究可通过扩散蒸馏或少步采样减少去噪次数，利用前馈模型提供三维初始化，并只对低置信度区域和观测不足视角调用扩散模型。相关实验还应分别报告扩散采样时间、三维优化时间和显存占用，以便判断效率提升来自采样过程、重建过程还是生成调用次数的减少。

5 结 语

总体来看，扩散模型为稀疏视图条件下的缺失信息恢复提供了新的技术路径，推动了生成先验与三维重建的进一步融合。现有方法虽改善了未观测区域的重建效果，但生成内容与真实几何之间的偏差仍未得到根本解决，方法的稳定性、泛化能力和计算效率也有待提高。未来研究应加强生成先验与真实观测、跨视角几何约束的协调，在提高重建可靠性的同时兼顾泛化能力与计算效率，推动稀疏视图三维重建向几何可信和实际可用发展。

参考文献(References)

Barron J T, Mildenhall B, Verbin D, Srinivasan P P and Hedman P. 2022. Mip-NeRF 360: unbounded anti-aliased neural radiance fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE: 5470-5479 [DOI: 10.1109/CVPR52688.2022.00539]

Barron J T, Mildenhall B, Verbin D, Srinivasan P P and Hedman P. 2023. Zip-NeRF: anti-aliased grid-based neural radiance fields//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 19697-19705 [DOI: 10.1109/ICCV51070.2023.01804]

Cao C J, Yu C H, Liu S, Wang F, Xue X Y and Fu Y W. 2025. MVGenMaster: scaling multi-view generation from any image via 3D priors enhanced diffusion model//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 6045-6056 [DOI: 10.1109/CVPR52734.2025.00567]

Cao X, Li Y Z, Zhang Y M, Song J Y, Yan C, Li W, et al. 2026. GeoQuery: geometry-query diffusion for sparse-view reconstruction//Proceedings of the ACM SIGGRAPH 2026 Conference Papers. Los Angeles, CA, USA: ACM: 1-11 [DOI: 10.1145/3799902.3811222]

Chen T Y, Xiang W, Han K, Lu Y, Wu D, Liu G W, et al. 2026a. GIFSplat: generative prior-guided iterative feed-forward 3D Gaussian splatting from sparse views//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, CO, USA: IEEE: 26031-26040.

Chen Y T, Guo S, Jin R B, Yang T S, Cai X, Luo Y W, et al. 2026b. AnyRecon: arbitrary-view 3D reconstruction with video diffusion model[EB/OL]. [2026-08-04]. https://arxiv. org/abs/2604.19747 [DOI: 10.48550/arXiv. 2604.19747]

© 中国图象图形学报版权所有

Deitke M, Schwenk D, Salvador J, Weihs L, Michel O, VanderBilt E, et al. 2023. Objaverse: a universe of annotated 3D objects//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 13142-13153 [DOI: 10.1109/CVPR52729.2023.01263]

Downs L, Francis A, Koenig N, Kinman B, Hickman R, Reymann K, et al. 2022. Google Scanned Objects: a high-quality dataset of 3D scanned household items//Proceedings of the IEEE International Conference on Robotics and Automation. Philadelphia, PA, USA: IEEE: 2553-2560 [DOI: 10.1109/ICRA46639.2022.9811809]

Fan H Q, Su H and Guibas L J. 2017. A point set generation network for 3D object reconstruction from a single image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 2463-2471 [DOI: 10.1109/CVPR.2017.264]

Furukawa Y and Hernández C. 2015. Multi-view stereo: a tutorial. Foundations and Trends in Computer Graphics and Vision, 9(1-2): 1-148 [DOI: 10.1561/0600000052]

Gao R Q, Holynski A, Henzler P, Brussee A, Martin-Brualla R, Srinivasan P P, et al. 2024. CAT3D: create anything in 3D with multi-view diffusion models. Advances in Neural Information Processing Systems, 37: 75468-75494 [DOI: 10.52202/079017-2403]

Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces//Proceedings of the Conference on Language Modeling.

Han X G, Xiu Y L, Xu Z, Lian Z H, Peng S D, Yao Y, et al. 2026. Frontiers and prospects of 3D reconstruction and generation. Journal of Image and Graphics, 31(6): 2182-2197 (韩晓光, 修宇亮, 徐震, 连宙辉, 彭思达, 姚遥等. 2026. 三维重建与生成前沿进展与展望. 中国图象图形学报, 31(6): 2182-2197) [DOI: 10.11834/jig.260070]

Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S. 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Advances in Neural Information Processing Systems, 30.

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33: 6840-6851

Ho J and Salimans T. 2022. Classifier-free diffusion guidance [EB/OL]. [2026-08-04]. https://arxiv.org/abs/2207.12598 [DOI: 10.48550/arXiv.2207.12598]

Hu E J, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. 2022. LoRA: low-rank adaptation of large language models//Proceedings of the 10th International Conference on Learning Representations. Virtual Conference: OpenReview.net

Huang B B, Yu Z H, Chen A P, Geiger A and Gao S H. 2024. 2D Gaussian splatting for geometrically accurate radiance fields//Proceedings of the ACM SIGGRAPH 2024 Conference Papers. ACM: 1-11 [DOI: 10.1145/3641519.3657428]

Huang Y C, Chien H J, Lin C Y, Chen Y H and Liu Y L. 2025. GaMO: geometry-aware multi-view diffusion outpainting for sparse-view 3D reconstruction [EB/OL]. [2026-08-04]. https://arxiv.org/abs/2512.25073 [DOI: 10.48550/arXiv.2512.25073]

Huynh-Thu Q and Ghanbari M. 2008. Scope of validity of PSNR in image/video quality assessment. Electronics Letters, 44(13): 800-801 [DOI: 10.1049/el:20080522]

Jensen R R, Dahl A L, Vogiatzis G, Tola E and Aanæs H. 2014. Large scale multi-view stereopsis evaluation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE: 406-413 [DOI: 10.1109/CVPR.2014.59]

Ji Y Z, Tian Q J, Zhu H, Jiang X Q, Cao G Z, Ma L Z, et al. 2026. S2D: sparse to dense lifting for 3D reconstruction with minimal inputs//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, CO, USA: IEEE: 7491-7502.

Kerbl B, Kopanas G, Leimkühler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4): 139:1-139:14 [DOI: 10.1145/3592433]

Kim M S, Jeon S B, Cho I, Yoo M J and Kim S J. 2025. ExploreGS: explorable 3D scene reconstruction with virtual camera samplings and diffusion priors//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, HI, USA: IEEE: 27042-27051 [DOI: 10.1109/ICCV51701.2025.02510]

Knapitsch A, Park J, Zhou Q Y and Koltun V. 2017. Tanks and Temples: benchmarking large-scale scene reconstruction. ACM Transactions on Graphics, 36(4): 78: 1-78: 13 [DOI: 10.1145/3072959.3073599]

Kong H Y, Yang X Y and Wang X C. 2025. Generative sparse-view Gaussian splatting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 26745-26755 [DOI: 10.1109/CVPR52734.2025.02491]

Li L G, Zhang Z Y, Li Y W, Xu J L, Hu W B, Li X Y, et al. 2025. NVComposer: boosting generative novel view synthesis with multiple sparse and unposed images//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 777-787 [DOI: 10.1109/CVPR52734.2025.00081]

Ling L, Sheng Y C, Tu Z, Zhao W T, Xin C, Wan K, et al. 2024. DL3DV-10K: a large-scale scene dataset for deep learning-based 3D vision//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 22160-22169 [DOI: 10.1109/CVPR52733.2024.02092]

Liu F F, Sun W Q, Wang H Y, Wang Y K, Sun H W, Ye J L, et al. 2026. ReconX: reconstruct any scene from sparse views with video diffusion model. IEEE Transactions on Image Processing, 35: 2305-2319 [DOI: 10.1109/TIP.2026.3666733]

© 中国图象图形学报版权所有

Liu R S, Wu R D, Van Hoorick B, Tokmakov P, Zakharov S and Vondrick C. 2023. Zero-1-to-3: zero-shot one image to 3D object//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 9298-9309 [DOI: 10.1109/ICCV51070.2023.00853]

Liu X, Zhou C Y and Huang S Y. 2024a. 3DGS-Enhancer: enhancing unbounded 3D Gaussian splatting with view-consistent 2D diffusion priors. Advances in Neural Information Processing Systems, 37: 133305-133327 [DOI: 10.52202/079017-4237]

Liu X H, Chen J B, Kao S H, Tai Y W and Tang C K. 2024b. Deceptive-NeRF/3DGS: diffusion-generated pseudo-observations for high-quality sparse-view reconstruction//Computer Vision—ECCV 2024. Cham, Switzerland: Springer: 337-355 [DOI: 10.1007/978-3-031-72640-8_19]

Liu Y, Lin C, Zeng Z, Long X, Liu L J, Komura T, et al.2024c. SyncDreamer: generating multiview-consistent images from a single-view image//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net.

Long X X, Guo Y C, Lin C, Liu Y, Dou Z Y, Liu L J, et al. 2024. Wonder3D: single image to 3D using cross-domain diffusion//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 9970-9980 [DOI: 10.1109/CVPR52733.2024.00951]

Mildenhall B, Srinivasan P P, Ortiz-Cayon R, Kalantari N K, Ramamoorthi R, Ng R, et al. 2019. Local light field fusion: practical view synthesis with prescriptive sampling guidelines. ACM Transactions on Graphics, 38(4): 29: 1-29: 14 [DOI: 10.1145/3306346.3322980]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]

Müller T, Evans A, Schied C and Keller A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics, 41 (4): 102: 1-102: 15 [DOI: 10.1145/3528223.3530127]

Nagoor Kani B R, Lee H Y, Tulyakov S and Tulsiani S. 2024. UpFusion: novel view diffusion from unposed sparse view observations//Computer Vision—ECCV 2024. Cham, Switzerland: Springer: 179-195 [DOI: 10.1007/978-3-031-73116-7_11]

Ni J F, Liu Y, Lu R J, Zhou Z R, Zhu S C, Chen Y X, et al. 2025. Decompositional neural scene reconstruction with generative diffusion prior//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 6022-6033 [DOI: 10.1109/CVPR52734.2025.00565]

Ni J F, Chen Y X, Yang Z F, Liu Y, Lu R J, Zhu S C, et al. 2026. G4Splat: geometry-guided Gaussian splatting with generative prior//Proceedings of the 14th International Conference on Learning Representations. Rio de Janeiro, Brazil: OpenReview.net

Paliwal A, Zhou X L, Ye W, Xiong J H, Ranjan R and Kalantari N K. 2025. RI3D: few-shot Gaussian splatting with repair and inpainting diffusion priors//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, HI, USA: IEEE: 25094-25103 [DOI: 10.1109/ICCV51701.2025.02327]

Poole B, Jain A, Barron J T and Mildenhall B. 2023. DreamFusion: text-to-3D using 2D diffusion//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net

Qureshi A, Jung D, Choi J H and Manocha D. 2026. PanoPlane: plane-aware panoramic completion for sparse-view indoor 3D Gaussian splatting[EB/OL]. [2026-08-04]. https://arxiv. org/abs/2605.14135 [DOI: 10.48550/arXiv. 2605.14135]

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, 139: 8748-8763.

Reizenstein J, Shapovalov R, Henzler P, Sbordone L, Labatut P and Novotný D. 2021. Common Objects in 3D: large-scale learning and evaluation of real-life 3D category reconstruction//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE: 10901-10911 [DOI: 10.1109/ICCV48922.2021.01072]

Roberts M, Ramapuram J, Ranjan A, Kumar A, Bautista M A, Paczan N, et al. 2021. Hypersim: a photorealistic synthetic dataset for holistic indoor scene understanding//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE: 10912-10922 [DOI: 10.1109/ICCV48922.2021.01073]

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE: 10684-10695 [DOI: 10.1109/CVPR52688.2022.01042]

Sargent K, Li Z Z, Shah T, Herrmann C, Yu H X, Zhang Y Z, et al. 2024. ZeroNVS: zero-shot 360-degree view synthesis from a single image//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 9420-9429 [DOI: 10.1109/CVPR52733.2024.00900]

Schönberger J L and Frahm J M. 2016. Structure-from-motion revisited//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]

Shen Y Y, Yang Y, Zhou K and Shao T J. 2026. Sparse-to-complete: from sparse image captures to complete 3D scenes//Proceedings of the ACM SIGGRAPH 2026 Conference Papers. Los Angeles, CA,

© 中国图象图形学报版权所有

USA: ACM: 1-11 [DOI: 10.1145/3799902.3811099]

Shi R X, Chen H S, Zhang Z Y, Liu M H, Xu C, Wei X Y, et al. 2023. Zero123++: a single image to consistent multi-view diffusion base model[EB/OL]. [2026-08-04]. https://arxiv. org/abs/2310.15110 [DOI: 10.48550/arXiv. 2310. 15110]

Shih M L, Ma W C, Boyice L, Holynski A, Cole F, Curless B L, et al. 2024. ExtraNeRF: visibility-aware view extrapolation of neural radiance fields with diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 20385-20395 [DOI: 10.1109/CVPR52733. 2024.01927]

Sohl-Dickstein J, Weiss E A, Maheswaranathan N and Ganguli S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR: 2256-2265

Straub J, Whelan T, Ma L, Chen Y F, Wijmans E, Green S, et al. 2019. The Replica dataset: a digital replica of indoor spaces[EB/OL]. [2026-08-04]. https://arxiv. org/abs/1906.05797 [DOI: 10.48550/arXiv. 1906. 05797]

Tang J M, Zhang W Y, Zhou J S, Huang Z A, Shi K L, Xu S K, et al. 2026. VidSplat: Gaussian splatting reconstruction with geometry-guided video diffusion priors//Proceedings of the ACM SIGGRAPH 2026 Conference Papers. Los Angeles, CA, USA: ACM: 1-11 [DOI: 10.1145/3799902.3811138]

Tang J X, Ren J W, Zhou H, Liu Z W and Zeng G. 2024a. DreamGaussian: generative Gaussian splatting for efficient 3D content creation//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net.

Tang S T, Chen J C, Wang D L, Tang C Z, Zhang F Y, Fan Y C, et al. 2024b. MVDiffusion++: a dense high-resolution multi-view diffusion model for single or sparse-view 3D object reconstruction//Computer Vision—ECCV 2024. Cham, Switzerland: Springer: 175-191 [DOI: 10.1007/978-3-031-72640-8_10]

Tatarchenko M, Richter S R, Ranftl R, Li Z W, Koltun V and Brox T. 2019. What do single-view 3D reconstruction networks learn? // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE: 3405-3414 [DOI: 10.1109/CVPR.2019.00352]

Voleti V, Yao C H, Boss M, Letts A, Pankratz D, Tochilkin D, et al. 2024. SV3D: novel multi-view synthesis and 3D generation from a single image using latent video diffusion//Computer Vision – ECCV 2024. Cham: Springer: 439-457 [DOI: 10.1007/978-3-031-73232-4_25]

Wang C, Peng H Y, Liu Y T, Gu J T and Hu S M. 2025a. Diffusion models for 3D generation: a survey. Computational Visual Media, 11(1): 1-28 [DOI: 10.26599/CVM.2025.9450452]

Wang J, Chen M, Karaev N, Vedaldi A, Rupprecht C and Novotny D. 2025b. VGGT: visual geometry grounded transformer//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 5294-5306.

Wang P, Liu L, Liu Y, Theobalt C, Komura T and Wang W. 2021. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction. Advances in Neural Information Processing Systems, 34: 27171-27183.

Wang Q S, Zhao Y F, Ma J W and Li J. 2024a. How to use diffusion priors under sparse views? Advances in Neural Information Processing Systems, 37: 30394-30424 [DOI: 10.52202/079017-0957]

Wang S Z, Leroy V, Cabon Y, Chidlovskii B and Revaud J. 2024b. DUSt3R: geometric 3D vision made easy//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 20697-20709 [DOI: 10.1109/CVPR52733.2024.01956]

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Wang Z R, Gu Y, Zhou D M and Xu R J. 2025c. FixingGS: enhancing 3D Gaussian splatting via training-free score distillation[EB/OL]. [2026-08-04]. https://arxiv. org/abs/2509.18759 [DOI: 10.48550/arXiv. 2509. 18759]

Wang Z Y, Lu C, Wang Y K, Bao F, Li C X, Su H, et al. 2023. ProlificDreamer: high-fidelity and diverse text-to-3D generation with variational score distillation. Advances in Neural Information Processing Systems, 36: 8406-8441

Wu C H, Chen Y C, Solarte B, Yuan L and Sun M. 2025a. iFusion: inverting diffusion for pose-free reconstruction from sparse views// Proceedings of the 2025 International Conference on 3D Vision. Singapore: IEEE: 813-823 [DOI: 10.1109/3DV66043.2025. 00080]

Wu J Z J, Zhang Y X, Turki H, Ren X C, Gao J, Shou M Z, et al. 2025b. DIFIX3D+: improving 3D reconstructions with single-step diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 26024-26035 [DOI: 10.1109/CVPR52734.2025.02424]

Wu R, Mildenhall B, Henzler P, Park K, Gao R, Watson D, et al. 2024. ReconFusion: 3D reconstruction with diffusion priors//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 21551-21561 [DOI: 10.1109/CVPR52733.2024.02036]

Wu S B, Xu C R, Huang B B, Geiger A and Chen A P. 2025c. GenFusion: closing the loop between reconstruction and generation via videos//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 6078-6088 [DOI: 10.1109/CVPR52734.2025.00570]

Wu T, Zhang J R, Fu X, Wang Y X, Ren J W, Pan L, et al. 2023.

© 中国图象图形学报版权所有

OmniObject3D: large-vocabulary 3D object dataset for realistic perception, reconstruction and generation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 803-814 [DOI: 10.1109/CVPR52729.2023.00084]

Wynn J and Turmukhambetov D. 2023. DiffusioNeRF: regularizing neural radiance fields with denoising diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 4180-4189 [DOI: 10.1109/CVPR52729.2023.00407]

Xiong H, Muttukuru S, Xiao H Y, Upadhyay R, Chari P, Zhao Y J, et al. 2025. SparseGS: sparse view synthesis using 3D Gaussian splatting//Proceedings of the 2025 International Conference on 3D Vision. Singapore: IEEE: 1032-1041 [DOI: 10.1109/3DV66043.2025.00100]

Xu C, Li A, Chen L H, Liu Y L, Shi R X, Su H, et al. 2024. SpaRP: fast 3D object reconstruction and pose estimation from sparse views//Computer Vision—ECCV 2024. Cham, Switzerland: Springer: 143-163 [DOI: 10.1007/978-3-031-73039-9_9]

Xu Y, Wang Y X and Yu S X. 2025. Novel view synthesis from a few glimpses via test-time natural video completion. Advances in Neural Information Processing Systems, 38: 46541-46561

Xu Z S and Oishi T. 2026. PanoImager: geometry-guided novel view synthesis and reconstruction from sparse panoramic views [EB/OL]. [2026-08-04]. https://arxiv. org/abs/2606.27071 [DOI: 10.48550/arXiv. 2606.27071]

Yang C, Li S K, Fang J M, Liang R F, Xie L X, Zhang X P, et al. 2024. GaussianObject: high-quality 3D object reconstruction from four views with Gaussian splatting. ACM Transactions on Graphics, 43(6): 199:1-199:13 [DOI: 10.1145/3687759]

Yin X Y L, Zhang Q, Chang J H, Feng Y, Fan Q N, Yang X, et al. 2026. GSFixer: improving 3D Gaussian splatting with reference-guided video diffusion priors//Proceedings of the 43rd International Conference on Machine Learning. Proceedings of Machine Learning Research, 306.

Yoo P, Guo J X, Matsuo Y and Gu S X S. 2023. DreamSparse: escaping from Plato's cave with 2D diffusion model given sparse views. Advances in Neural Information Processing Systems, 36: 3307-3324 [DOI: 10.52202/075280-0146]

Younis T and Cheng Z L. 2025. Sparse-view 3D reconstruction: recent advances and open challenges [EB/OL]. [2026-08-04]. https://arxiv. org/abs/2507.16406 [DOI: 10.48550/arXiv. 2507.16406]

Yu A, Ye V, Tancik M and Kanazawa A. 2021. pixelNeRF: neural radiance fields from one or few images//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 4578-4587 [DOI: 10.1109/CVPR46437.2021.00455]

Yu W B, Xing J B, Yuan L, Hu W B, Li X Y, Huang Z P, et al. 2025. ViewCrafter: taming video diffusion models for high-fidelity novel view synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence: 1-18 [DOI: 10.1109/TPAMI. 2025.3613256]

Yuan Z L, Li Z H, Chen K H, Mao T L, Jiang H and Wang Z Q. 2026. Multiview stereo 3D reconstruction research: a survey. Journal of Image and Graphics, 31(3): 0657-0685 (袁祯泷, 李泽昊, 陈科桦, 毛天露, 蒋浩, 王兆其. 2026. 多视角立体匹配三维重建研究综述. 中国图象图形学报, 31(3): 0657-0685) [DOI: 10.11834/jig.250348]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]

Zhang S C, Xu H Y, Guo S T, Xie Z W, Bao H J, Xu W W, et al. 2025. SpatialCrafter: unleashing the imagination of video diffusion models for scene reconstruction from limited observations//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, HI, USA: IEEE: 27794-27805 [DOI: 10.1109/ICCV51701.2025.02580]

Zhao Q T and Tulsiani S. 2024. Sparse-view pose estimation and reconstruction via analysis by generative synthesis. Advances in Neural Information Processing Systems, 37: 111899-111922 [DOI: 10.52202/079017-3553]

Zhong Y J, Li Z H, Chen D Z Y, Hong L Q and Xu D. 2025. Taming video diffusion prior with scene-grounding guidance for 3D Gaussian splatting from sparse inputs//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 6133-6143 [DOI: 10.1109/CVPR52734.2025.00575]

Zhou T H, Tucker R, Flynn J, Fyffe G and Snavely N. 2018. Stereo magnification: learning view synthesis using multiplane images. ACM Transactions on Graphics, 37(4): 65:1-65:12 [DOI: 10.1145/3197517.3201323]

Zhou Z and Tulsiani S. 2023. SparseFusion: distilling view-conditioned diffusion for 3D reconstruction//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 12588-12597 [DOI: 10.1109/CVPR52729.2023.01211]

Zhu L Y, Narayana M, Stary M, Hutchcroft W, Wetzstein G and Armeni I. 2026. GaussFusion: improving 3D reconstruction in the wild with a geometry-informed video generator//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, CO, USA: IEEE: 15432-15442.

Zou Z X, Cheng W H, Cao Y P, Huang S S, Shan Y and Zhang S H. 2024. Sparse3D: distilling multiview-consistent diffusion for object reconstruction from sparse views//Proceedings of the 38th AAAI

© 中国图象图形学报版权所有

Conference on Artificial Intelligence. Vancouver, BC, Canada: AAAI: 7900-7908 [DOI: 10.1609/aaai.v38i7.28626]

作者简介

万贻前,男,助教,主要研究方向为三维计算机视觉。E-mail: 1833925375@qq.com
张永宏,男,副教授,主要研究方向为计算机应用技术、三维重建。E-mail:zyhaaa_cq@163.com
余鑫,男,助教,主要研究方向为大数据技术。E-mail: 28014003@qq.com

© 中国图象图形学报版权所有