

三维高斯泼溅技术在场景重建中的 现状与挑战

朱东林, 陈淼, 毛宇岩, 张峻豪, 王忠立*
(北京交通大学自动化与智能学院, 北京 100044)

摘要: 三维场景重建是自动驾驶、机器人等领域的重要研究课题, 在导航建图、环境交互、虚拟和增强现实等任务中有着广泛应用。目前基于深度学习的重建方法从场景表示方式和核心建模技术角度主要分为五类: 基于代价体积的深度估计方法、基于截断的有符号距离函数(TSDF)的体素方法、基于 Transformer 架构的大规模前馈方法、基于多层感知机(MLP)的神经辐射场(NeRF)、三维高斯泼溅(3DGS)。每类方法都有其独特的优势和局限, 而新兴的 3DGS 方法通过高斯函数显式的表示场景, 并利用高效的光栅化操作实现场景的快速渲染和新视角合成。最大优势是它相比著名的神经辐射场方法采用 MLP 网络表示场景信息的建模方式不同, 能够在保证高效渲染的同时还具有可解释性且可编辑, 这为三维场景进行准确的重建铺平了道路。然而, 3DGS 在场景重建任务中的应用仍然面临着许多困难和挑战。基于此, 本文首先对 3DGS 的基本概念进行简单介绍, 并与上述其余四类方法进行特点比较。然后, 对现有 3DGS 重建算法进行系统性的调研后, 对这类方法要解决的关键问题进行总结, 并结合典型实例对相关核心难题的研究现状进行综述。最后, 对未来更有可能探索的新研究方向进行了展望。

关键词 深度估计; 3D Gaussian Splatting; 神经辐射场; 三维场景重建
中图分类号 TP391 **文献标志码** A

3D Gaussian Splatting: Research Status and Challenges in Scene Reconstruction

Zhu Donglin, Chen Miao, Mao Yuyan, Zhang Junhao, Wang Zhongli*

(School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China)
Corresponding Author: zlwang@bjtu.edu.cn

Abstract: 3D scene reconstruction is a critical research topic in autonomous driving, robotics, and related fields, with extensive applications in navigation mapping, environmental interaction, and virtual/augmented reality tasks. Current deep learning-based reconstruction methods can be primarily categorized into five groups from the perspectives of scene representation and core modeling techniques: cost volume-based depth estimation methods, truncated signed distance function (TSDF)-based voxel approaches, transformer architecture-based large-scale feedforward methods, multilayer perceptron (MLP)-based neural radiance fields (NeRF), and

来稿日期: 2024-11-27 修回日期: 2025-03-13

基金项目: 国家重点研发计划任务(2022ZD0205005), 研究生创新项目(2023YJS142)

作者简介: 朱东林, 博士研究生, 研究方向为交通系统感知与大数据; 陈淼, 硕士研究生, 研究方向为同步定位与建图; 毛宇岩, 硕士研究生, 研究方向为智能交通系统; 张峻豪, 硕士研究生, 研究方向为智能交通系统; 王忠立(通讯作者), 教授, 研究方向为智能交通系统、计算机视觉、机器学习和认知机器人学, E-mail: zlwang@bjtu.edu.cn.

3D Gaussian Splatting (3DGS). Each category exhibits unique strengths and limitations. The emerging 3DGS method distinguishes itself by explicitly representing scenes through Gaussian functions while achieving rapid scene rendering and novel view synthesis via efficient rasterization operations. Its most significant advantage lies in diverging from NeRF's MLP-based scene representation paradigm - 3DGS ensures both efficient rendering and interpretable editable scene modeling, thereby paving the way for accurate 3D scene reconstruction. However, 3DGS still faces numerous challenges in practical scene reconstruction applications. Based on this analysis, this paper first provides a concise introduction to 3DGS fundamentals and conducts comparative analysis with the aforementioned four categories. Following a systematic survey of existing 3DGS reconstruction algorithms, we summarize the key challenges addressed by these methods and review current research progress on core technical difficulties through representative case studies. Finally, we prospect potential future research directions worthy of exploration.

Key words: Depth Estimation; 3D Gaussian Splatting; Neural Radiance Field; 3D Scene Reconstruction

1 引言

近年来, 自动驾驶、虚拟与增强现实技术的快速发展对高保真、高效率的 3D 场景重建提出了迫切需求。传统计算机图形学技术对场景及物体进行三维建模时有一套完善的操作流程, 包括几何建模、纹理设计、材质定义和光照设置等。这些工作往往需要较多的人力物力, 凭借有美术丰富经验的人员完成, 时间的消耗延缓了项目的落地周期。

为了打破这一局面, 于 2020 年斯坦福大学和 DavidGAN 公司合作首次引出辐射场的概念, 主要为解决基于 2D 图像渲染方式下的 3D 场景重建问题。2021 年, 首个结合神经网络和机器学习技术的 NeRF 方法由 Mildenhall 等人^[1]提出并发表在 ECCV 会议上, 它将三维空间描述为一种连续函数, 达到能够描述场景中任意点处颜色和密度的目的, 涵盖了光线如何与物体表面、材质以及周围环境发生相互作用, 还原光在空间中的表现以重现现实世界。NeRF 的提出革新了场景表示方式, 通过隐式神经场建模实现了高质量的新视角合成, 一经出现便在建筑设计、虚拟现实、游戏开发、电影制作等多个领域得到人们的广泛关注。2022 年, Azinovic 等人^[2]又将 TSDF 与 NeRF 框架结合, 既解决了 NeRF 无法显式表示几何结构的问题, 又弥补了 TSDF 无法捕捉表面光照、颜色等信息的缺陷, 二者相辅相成。虽然二者的结合使得 NeRF 的表示问题得到解决, 但随着场景规模的扩大, NeRF 自身的渲染效率问题越来越凸显, 因其依赖光线的密集采样和神经网络查询的重复操作导致训练耗时 (通常数十小时) 且渲染速度低下 (数秒每帧), 难以满足实时应用需求和环境交互。例如, Mip-NeRF 360^[3]虽在无界场景中表现出色, 但其训练需多 GPU 协同, 且实时渲染仍无法实现, 严重限制了其在移动端和大范围场景中的应用。

随着对 3D 场景进行实时渲染需求的日益高涨, 2023 年 Kerbl 等人^[4]提出了三维高斯泼溅方法——3DGS, 与 Azinovic 等人的方法相比把体素表示改变为 3D 高斯函数。因其只需在相应位置放置基元并针对性的调节其协方差和朝向来拟合物体表面, 而无需过多关注无关区域的占用情况, 避免了大量计算量。另外, 与原始 NeRF 方法相比转变了渲染过程中的光线投射和合成方式, 转而采用光栅化渲染技术, 将 3D 高斯投影至 2D 图像平面,

直接合成得到像素颜色。这种设计避免了诸如 NeRF 方法在颜色渲染期间重复进行 MLP 网络推理产生的计算耗时，加快了渲染速度，使其在消费级 GPU 上仅需 30 分钟即可完成训练，并以 1080p 分辨率实现 30 FPS 以上的实时渲染。

在此背景下，3DGS 的显式几何特性为场景重建提供了天然优势。例如，VastGaussian^[5] 通过相机投影分布将城市级场景划分为多个区域，基于可见性准则动态扩展训练视角，成功实现了平方公里级场景的高效重建；StreetGaussians^[6] 则通过静态背景与动态对象分离建模，结合车辆姿态跟踪，实现了复杂城市场景的实时动态渲染。3DGS 在动态场景建模中也展现了较强的灵活性。例如，4D-GS^[7] 采用多分辨率 HexPlane 体素编码时空信息，通过轻量化 MLP 预测高斯属性偏移，在动态数据集上 PSNR 达 34.01，显著优于 Tensor4D^[8] 等 NeRF 衍生方法。此外，3DGS 在 SLAM 任务^[9,10] 中表现出色。例如，GS-SLAM^[11] 通过自适应高斯扩展策略，在 Replica 数据集上以 31.56 PSNR 超越 NeRF-based 方法 NICE-SLAM^[12] 的 24.42 dB，且重建速度提升两个数量级。这些实例表明，3DGS 不仅突破了 NeRF 的效率瓶颈，更在复杂场景的实用性上开辟了新路径。下图 1 的时间线展示了自 NeRF 方法出现至今，NeRF 和 3DGS 代表性方法的发展历程。

然而，现有这类重建方法的研究仍面临四大挑战：其一，静态场景模型中高斯球体的存储与计算开销仍待优化，如 LightGaussian^[13] 通过八叉树压缩将内存占用降低 15 倍，但重建精度和压缩比率之间的权衡尚未彻底解决；其二，复杂材质（如镜面反射）和光照条件的物理属性建模不足，GaussianShader^[14] 虽引入简化渲染方程，但在多光源交互场景中仍存在光路估计偏差；其三，虚拟世界建模中几何和语义信息不足导致结果紊乱，如一些综述^[15-17] 也指出通过大语言模型的语义理解和生成能力，赋予 3DGS 增强的空间语义感知能力，但语义丰富度和场景连贯性尚未彻底解决；其四，城市级动态场景建模的精度和渲染效率有待提高，如 DrivingGaussian^[18] 利用复合高斯泼溅方式，将动态驾驶场景中的车辆、行人等运动目标分解为独立的高斯集合，结合时序一致性约束，实现了多视角动态合成与碰撞模拟，但其在拓扑变化和多目标交互中的表现仍有待进一步提升。

本文从上述存在的四大挑战为出发点，综述 3DGS 在场景重建中解决这些挑战的最新进展，从存储成本、计算效率、重建精度等维度剖析其技术优势，并结合实际案例验证其应用价值。下面将按照 3DGS 方法的基本概念、3DGS 与其他基于深度学习的重建技术的特点对比、基于 3DGS 的四个重建方向上的研究现状、面临挑战和未来展望在如下各个章节进行介绍。

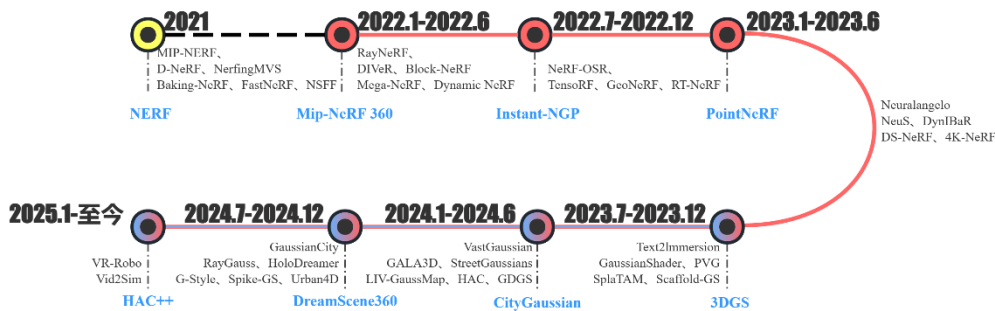


图 1 NeRF 和 3DGS 重建方法发展时间线

Figure 1 Development timeline of NeRF and 3DGS reconstruction methods

2 三维高斯泼溅方法基本概念及与其他重建技术的特点对比

2.1 三维高斯泼溅方法概念

3DGS 是一种新兴的 3D 场景表示方法，它通过在 3D 空间中使用高斯点云

$\mathbf{G} = \{g_1, \dots, g_N\}$ 表示三维场景，其中 N 是场景中高斯基元的数量。每个高斯点由多个属性组成，包括中心位置 $\boldsymbol{\mu} \in \mathbb{R}^3$ 、旋转四元数 $\mathbf{q} \in \mathbb{R}^4$ 、尺度 $\mathbf{s} \in \mathbb{R}^3$ 、球谐系数 $\mathbf{c} \in \mathbb{R}^{3(\text{SH}+1)^2}$ 和不透明度 $\alpha \in [0, 1]$ ，这些高斯点用于通过“泼溅技术”将 3D 高斯点投影到 2D 平面上，从而从特定视角渲染图像。这一过程通过将 3D 点投影到 2D，然后使用可微分块状光栅器将不同的高斯点混合在一起来完成。每个 3D 高斯的函数表达式为：

$$g(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (1)$$

其中，3D 协方差矩阵又可以通过缩放矩阵 \mathbf{S} 和旋转矩阵 \mathbf{R} 来定义，具体表达式为 $\boldsymbol{\Sigma} = \mathbf{R}\mathbf{S}\mathbf{S}^T\mathbf{R}^T$ 。矩阵 \mathbf{R} 和 \mathbf{S} 分别与旋转四元数 \mathbf{q} 和尺度向量 \mathbf{s} 有关。

对于一个具有投影变换 \mathbf{W} （世界到相机矩阵）的相机视角，以及 \mathbf{J} 作为投影变换的仿射近似的雅可比矩阵，相应的协方差矩阵投影到二维的表达式为： $\boldsymbol{\Sigma}' = \mathbf{J}\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T\mathbf{J}^T$ 和 $\boldsymbol{\mu}' = \mathbf{J}\mathbf{W}\boldsymbol{\mu}$ 。然后，通过混叠 2D 空间下与像素重叠的 N 个高斯点的颜色进行混合来计算像素颜色 \mathbf{C} ：

$$\mathbf{C} = \sum_{i \in \mathbf{N}} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (2)$$

$$\alpha_i = \alpha_i \exp\left(-\frac{1}{2}(\mathbf{x}' - \boldsymbol{\mu}'_i)^T \boldsymbol{\Sigma}'^{-1}(\mathbf{x}' - \boldsymbol{\mu}'_i)\right)$$

其中，每个高斯的颜色 \mathbf{c}_i 随后使用球谐系数计算。

同样，深度图也可以被渲染表示为：

$$D = \sum_{i \in \mathbf{N}} d_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (3)$$

其中， d_i 表示高斯中心到相机平面的距离。

高斯点是通过运动结构恢复（SfM）算法创建的稀疏点云初始化的。然后，使用随机梯度下降对属性进行进一步优化，因为渲染过程是完全可微分的。对于从训练数据集中采样的每个视图，使用上述前向过程进行投影和栅格化，然后计算重建损失，通过结合 L1 与 SSIM 损失来完成：

$$\mathbf{L} = (1 - \lambda)\mathbf{L}_1 + \lambda\mathbf{L}_{SSIM} \quad (4)$$

其中， λ 设置为 0.2。在高斯点的优化中，控制高斯点数量的另一个关键步骤是在预热阶段之后，每 100 次迭代移除不透明度值 α 低于阈值的高斯点。此外，大于相应几何形状的大型高斯点将被分割，而小型高斯点将被克隆，以便更好地适应物体几何表面。只有在每 100 次迭代后位置梯度高于阈值 τ 的高斯点才会被分割或克隆。3DGS 工作流程如下图 2 所示。

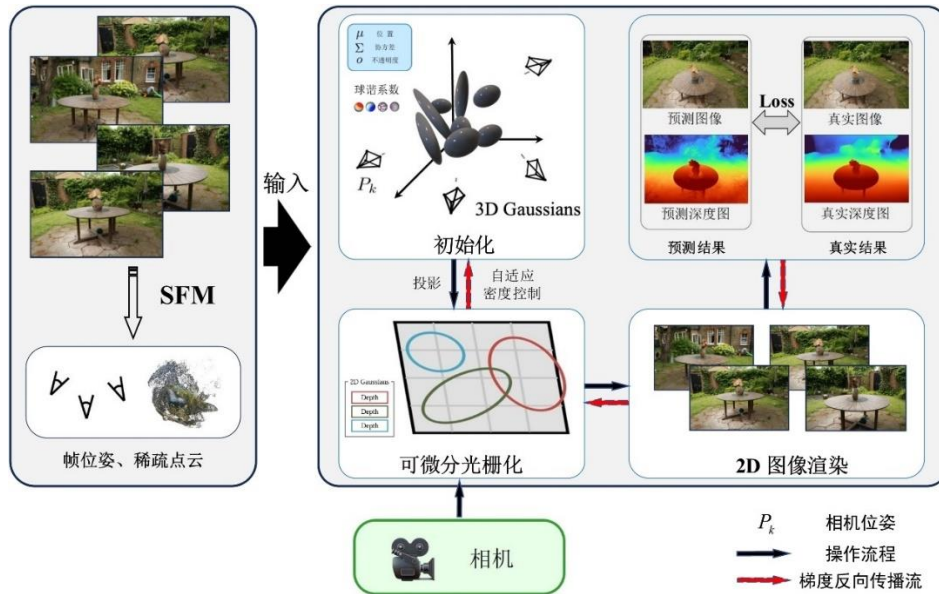


图 2 3DGS 工作流程图
Figure 2 The workflow diagram of 3DGS

2.2 三维高斯泼溅与其他重建技术的特点对比

基于深度学习的场景重建，利用深度学习模型对大量数据进行学习和分析^[19]，从而实现场景几何结构和纹理信息的精确重建。目前，基于深度学习的场景重建方法依据不同的技术路线和模型架构大致分为五类：基于代价体积的深度估计方法^[20]、基于截断的有符号距离函数的体素方法^[21]、基于注意力架构的大规模前馈方法^[22]、基于多层感知机的 NeRF 方法^[23]、3DGS 方法。

基于代价体积的深度估计方法又分为三类：监督、半监督和无监督，它们都通过构建多视图间的特征匹配代价体积进行深度估计。监督方法依赖密集深度真值，核心操作包括多尺度特征提取、可微分单应变换构建代价体积、基于 3D 卷积的代价体正则化及深度回归。最近，SimpleRecon^[24]通过引入时序一致性约束提升动态场景重建鲁棒性，而 SimpleMapping^[25]结合语义分割实现场景级语义重建。其优势在于高精度深度估计，但对标注数据依赖性强。半监督方法的典型代表有 SGT-MVSNet 模型^[26]，其通过融合稀疏 LiDAR 点云与稠密光度一致性约束，在有限标注下提升泛化能力。DepthFormer 模型^[27]利用 Transformer 建模语义上下文，显著改善弱纹理区域深度估计。无监督方法基于光度一致性假设与几何约束，通过最小化重投影误差实现自监督学习，典型工作有 MonoDepth2^[28]、Unsupervised-MVSNet^[29]。最新的研究 UniMVSNet^[30]引入跨视图注意力机制，有效缓解遮挡区域的深度模糊问题。此类方法计算复杂度高，且易受光照变化影响，多适用于小尺度场景重建。

基于截断的有符号距离函数的体素方法将空间离散化为体素网格，通过 TSDF 隐式建模表面几何。截断的有符号距离函数作为隐函数不仅能提供点到最近表面的距离信息，还能考量此距离是朝向表面内部还是外部。典型框架如 3DVNet 模型^[31]将 TSDF 预测转化为体素分类任务。最近，大规模稠密语义快速三维重建系统 Voxblox++^[32]中提出的稀疏体素表示与 HVOFusion 框架^[33]中提出的分层体素优化都显著降低了内存占用。此类方法的几何重建精度高（可达毫米级），但体素分辨率与计算量呈立方级增长，且难以建模复杂材质与光照特性。

基于注意力架构的端到端大规模前馈方法通过全局注意力机制直接建模多视图几何关

系，典型工作包括 DUS3R^[34]与 Spann3R^[35]。DUS3R 利用交叉注意力实现像素级特征匹配，结合 transformer 解码器输出密集 3D 点云，在零样本泛化能力上表现突出。最新方法 GaussianTransformer^[36]引入可变形注意力模块，将 3D 高斯参数预测融入 transformer 框架，在保持高效率的同时提升重建质量。此类方法对显存需求高，且多聚焦于对象级几何重建，缺乏外观属性建模能力。

基于多层感知机的 NeRF 方法通过 MLP 隐式编码辐射场（密度与颜色），结合可微分体渲染实现照片级新视角合成。后续研究主要从几何优化、动态建模、加速策略三方面拓展其应用。几何优化方面，例如 NeuS^[37]将符号距离函数（SDF）引入体渲染方程，实现高质量表面提取；动态建模方面，例如 D-NeRF^[38]通过时变 MLP 参数化动态场景，4K-NeRF^[39]利用显式变形场提升时序一致性；加速策略方面，例如 Instant-NGP^[40]采用多分辨率哈希编码将训练时间缩短至数分钟，Mip-NeRF 360^[3]通过圆锥采样与在线蒸馏优化无界场景重建。此类方法需要密集采样光线，每条光线需至少 128 个样本点，这会导致训练耗时通常大于 24 小时，并且渲染过程存在延迟，每帧耗时可达数秒，限制了其在实时系统中的应用。各重建方法的特点对比汇总如表 1 所示。

表 1 基于深度学习的重建方法的分类对比

方法分类	典型工作	核心技术	优点	缺点	应用场景
基于代价体积分的深度估计方法	SimpleRecon[24] SimpleMapping[25] SGT-MVSNet[26] DepthFormer[27] MonoDepth2[28]	构建一个包含所有可能视差假设的代价体积，通过累积匹配信息和计算每个像素的深度。	可以与现有的深度学习框架和算法相结合，易于扩展和优化。	处理高分辨率图像时成本高，对纹理和光照变化敏感。	产品质量控制和缺陷检测、增强和虚拟现实。
基于截断的有符号距离函数的体素方法	3DVNet[31] Voxblox++[32] HVOFusion[33]	将三维空间离散化为体素网格，每个体素存储一个 TSDF 值，通常使用 marching cubes 操作提取等值面。	TSDF 值的累积更新机制对于噪声和不完整的数据具有一定的鲁棒性。	受分辨率影响，难以捕捉细节丰富的几何和颜色特征，尤其在初始化和表面提取阶段，处理速度较慢。	结构化场景可视化和建模、三维地图用于辅助机器人路径规划和避障。
基于注意力架构的大规模前馈方法	DUS3R[34] Spann3R[35] GaussianTransformer[36]	利用自注意力或多头注意力机制直接从图像特征中学习场景的三维表示。	端到端学习，可以直接从图像到三维表示的映射，无需复杂的中间步骤。	为了捕捉长距离上下文信息，搭建的深度网络结构往往复杂。	历史文物的数字化重建、电影和游戏三维模型生成。
基于多层感知机的 NeRF 方法	NeRF[1] Mip-NeRF 360[3] NeuS[37] D-NeRF[38] 4K-NeRF[39] Instant-NGP[40]	使用 MLP 网络来学习从 3D 坐标和视点方向到颜色和密度的映射，同时捕捉几何和外观，通过体积渲染技术合成新视图图像。	提供了一种连续的 3D 场景表示，能够在渲染和训练阶段同时捕捉几何和外观，可以适应不同的场景和条件。	在渲染和训练阶段需要大量的计算资源，训练过程缓慢。	三维地图构建、建筑和城市规划、文化遗产保护等。
3D Gaussian Splatting 方法	3DGS[4]	使用 3D 高斯分布表示三维场景，通过高效的光栅化和 α -blending 技术实现快速渲染。	在内存和计算上比 NeRF 等方法高效，对场景几何结构的理解和操作直观。	提取干净、无噪声的表面存在挑战。	三维地图构建、建筑和城市规划、文化遗产保护等。

3 基于三维高斯泼溅的四个重建方向的研究现状

虽然 3DGS 相比 NeRF 有较快的渲染速度和质量，能够满足实时重建的需求。但在应用过程中暴露了一些问题亟待解决，比如稀疏点云模型作为网络输入进行多轮迭代优化后得到最终稠密 3D 模型期间，模糊的特征匹配容易造成异常高斯基元的加入而加重存储成本、悬浮伪影等现象。因此，为了改善与这两个主要现象有关的地图精度下降问题，现有研究工作主要围绕静态模型成本压缩、模型外观色彩属性优化、结合大语言模型的虚拟世界重构、城市级动态场景建模这四个重点方向展开研究。除此之外，对于上述每项挑战的研究，下面对现有方法提出的解决方案进行更细致的划分，例如模型成本压缩问题，细分为特征编码、高斯修剪、变换域、模型量化等技术途径展开讨论。

3.1 面向静态模型成本压缩的三维高斯泼溅重建研究

在利用三维高斯泼溅进行重建的过程中，3DGS 模型通常包含大量的高斯椭球体，每个椭球体都附带了位置、旋转、缩放、不透明度和球谐系数等属性。这些属性在高分辨率场景重建中会迅速增加导致较大的存储需求，尤其是在大范围场景或动态场景重建中，存储成本可能变得不可承受，故解决模型存储成本的问题至关重要。为了应对这一问题，研究者们提出了多种压缩和正则化方法，旨在减少存储需求并保持重建质量。下面列举了现有工作在解决相关问题时采用的四种技术手段的研究情况。

(1) 基于特征编码的方法

三维环境下的重建需要大量的存储开销，这是因为地图表示需要将 3D 高斯数量扩展到数千万个，甚至是数十亿个，通常约占用数十 GB 的显存 (VRAM)。在面向大规模场景的 3D 高斯重建中，特征编码技术通过参数压缩有效缓解了存储压力。例如，Niedermayr 等人^[41]提出了一种灵敏度感知的向量聚类与量化感知训练方法，其核心在于建立参数敏感性度量模型。具体而言，该方法通过计算训练图像中各参数梯度的 L1 范数，量化其对渲染质量的贡献度，进而对球谐系数 (SH) 和高斯形状参数进行重要性排序。高敏感度参数被保留为码本中的关键特征，而冗余参数则通过 K-means 聚类生成紧凑码本。实验表明，该方法在真实场景中实现了 31 倍的压缩率，且 PSNR 损失仅为 0.23 dB。类似地，GaussianCity 框架^[42]提出基于鸟瞰图的网格特征表示 BEV-Point，通过将语义地图的像素映射到高度场生成三维锚点集。该方法采用分层编码策略，在局部 BEV 网格内对锚点进行空间关联性建模，结合轻量化卷积网络实现特征降维。相较于传统 3DGS，BEV-Point 在城市场景重建中可将存储开销降低至原方法的五分之一，但受限于高度场的离散化误差，其在复杂曲面重建中易产生伪影。

(2) 基于高斯修剪的方法

针对高斯冗余性问题，研究者提出基于显著性分析的动态修剪策略。Lee 等人^[43]设计了一种双阶段优化框架：首先引入体积-透明度联合掩码机制，定义高斯显著性函数 $S(p) = \alpha \cdot V(p) + (1 - \alpha) \cdot T(p)$ ，其中 $V(p)$ 为高斯体积占比， $T(p)$ 为透明度累积贡献度；其次采用残差向量量化 (R-VQ) 对几何参数进行多级码本编码，并利用 InstantNGP 哈希网格^[44]实现颜色属性的紧凑表示。实验表明，该方法在 Deep Blending 数据集上实现 28 倍压缩率，同时渲染速度提升 40%。LightGaussian^[13]进一步结合知识蒸馏技术，通过教师-学生网络架构将高阶球谐系数投影至低维隐空间，其关键创新在于设计基于 KL 散度的特征对齐损失，在保证视觉质量的同时将参数量缩减 17 倍。Scaffold-GS^[45]则提出稀疏体素锚点表示法，通过空间相关性分析选择代表性锚点，并学习其尺度偏移矩阵 ΔS 和颜色残差 ΔC 。该方法虽实现 4 倍压缩率，但因锚点密度不足导致高频细节丢失。HAC 方法^[46]对此改进，引入层次化锚点聚类算法，通过谱聚类生成多分辨率锚点网格，并建立网格内参数的概率分布模型。实验表明其压缩率可达 30 倍，但因引入马尔可夫随机场优化，训练耗时增加约 0.9 倍。

(3) 基于变换域的方法

传统空域表示存在信息冗余现象，变换域方法通过特征解耦实现高效压缩。Gong 等人^[47]提出梯度域高斯表示 (GDGS)，其理论依据为梯度场的稀疏特性满足奈奎斯特-香农采样定理下界。具体实现中，首先对原始 RGB 图像进行 Laplacian 金字塔分解，构建多尺度梯度特征图；其次设计梯度感知的高斯初始化策略，通过优化目标函数最小化梯度域重建误差。该方法在 DTU 数据集上将存储需求降低至传统 3DGS 的八分之一，但因梯度反投影计算需迭代求解 Poisson 方程，导致实时渲染帧率下降约 22%。近期研究尝试结合小波变换与高斯混合模型^[48]，通过 Daubechies 小波基函数对高斯参数进行频带划分，保留低频分量并量化高频噪声，在保持 PSNR>32 dB 前提下实现 15 倍压缩率。

(4) 基于模型量化的方法

针对高维参数的空间相关性，结构化量化方法通过重排序提升压缩效率。Morgenstern 等人^[49]提出自组织高斯网格 (SOGG)，其核心贡献包括：(1)设计并行线性分配排序算法 (PLAS)，

通过块状匈牙利算法在 GPU 上实现百万级高斯的快速重排；(2)引入局部平滑正则项，强制相邻网格参数满足局部连续性；(3)采用 JPEG XL 编解码器对重排后的 2D 网格进行分层量化，其中位置参数采用 14 位对数量化，旋转参数使用 6 位均匀量化。实验表明，该方法在 Mip-NeRF360 数据集上实现 42 倍压缩率，且训练耗时仅增加 7%。对比传统点云压缩算法（如 Draco），SOGG 通过训练阶段的结构优化，在同等压缩率下 PSNR 提升 2.1dB。然而，该方法对球谐系数的各向异性特征敏感，在视角依赖渲染场景中易产生色带伪影。

总结，为了降低地图模型的占用成本，现有方法主要围绕编码、压缩、量化、修剪等操作展开研究。相比原始的 3DGS 重建方法，在不降低质量的情况下地图成本不断减小。下表 2 显示了在三个公开数据集下多个著名的 3DGS 算法在模型成本方面的压缩性能对比，指标包括峰值信噪比（PSNR）、结构相似性指数（SSIM）、学习感知图像块相似度（LPIPS）以及模型大小（以兆字节为单位）。每个指标的前三名分别用红色、绿色和黄色块标出。从表中可以看出，前三行方法比原始 3DGS 方法以牺牲重建质量为代价实现了高压缩率。HAC++ 算法[]在所有数据集的测试中压缩率最高，但其他指标却不是最优的。IGS 算法的 high 版本在 Deep Blending 和 Mip-NeRF 360 数据集上都表现出较高的 PSNR 水平，证明其适用于优先考虑质量而非压缩的场景。

表 2 三种数据集上多个重建算法的存储压缩性能比较

Table 2 Comparison of Storage Compression Performance of Multiple Reconstruction Algorithms on Three Datasets

方法	Tanks and Temples				Mip-NeRF 360				Deep Blending			
	PSNR↑	SSIM↑	LPIPS↓	Size↓	PSNR↑	SSIM↑	LPIPS↓	Size↓	PSNR↑	SSIM↑	LPIPS↓	Size↓
Niedermayr et al. [41]	23.58	0.851	0.191	17.65M	27.12	0.801	0.232	28.61M	29.45	0.912	0.254	23.87M
Lee et al. [43]	23.32	0.831	0.201	39.43M	27.08	0.798	0.247	48.80M	29.79	0.901	0.258	43.21M
LightGaussian [13]	22.83	0.822	0.242	22.43M	27.00	0.799	0.249	44.54M	27.01	0.872	0.308	33.94M
Scaffold-GS [45]	23.96	0.853	0.177	86.50M	27.50	0.806	0.252	253.9M	30.21	0.906	0.254	66.00M
HAC [46]	24.04	0.846	0.187	8.10M	27.53	0.807	0.238	15.26M	29.98	0.902	0.269	4.35M
HAC (highrate)	24.40	0.852	0.176	11.80M	27.77	0.811	0.230	22.9M	30.34	0.906	0.258	6.70M
HAC++ [50]	24.22	0.849	0.190	5.18M	27.60	0.803	0.253	8.34M	30.16	0.907	0.266	2.91M
HAC++ (highrate)	24.32	0.854	0.178	8.63M	27.82	0.811	0.231	18.48M	30.34	0.911	0.254	5.28M
Morgenstern et al. [49]	22.78	0.817	0.211	13.05M	26.01	0.772	0.259	23.90M	28.92	0.891	0.276	8.40M
3DGS [4]	23.69	0.844	0.179	431.0M	27.49	0.813	0.222	744.7M	29.42	0.899	0.247	663.9M
3DGS-30K	23.14	0.841	0.183	411.0M	27.21	0.815	0.214	734.0M	29.41	0.903	0.243	676.0M
IGS [51]	23.70	0.836	0.227	8.90M	27.34	0.811	0.255	13.40M	30.63	0.904	0.293	6.60M
IGS (high)	24.05	0.849	0.211	13.10M	27.62	0.820	0.245	27.00M	32.33	0.924	0.253	8.10M

3.2 面向模型外观色彩属性优化的三维高斯泼溅重建研究

场景的复杂性在于光照条件的不一致性以及物体表面的反射特性，这些因素对准确的场景重建带来了严峻挑战。尤其是亮度变化明显、阴暗或者反光场景中对重建算法提出了更高的要求。例如，场景中反射物体（如镜子）的表面特性常常导致色彩重建误差，原始的 3DGS 的球谐系数难以准确模拟这些物体表面的光照变化，容易将反射内容错误地视为实际存在的几何实体。面对这一难题，Mirror-3DGS 框架^[52]将镜子的反射属性纳入 3D 高斯表示中，利用平面镜成像原理，从镜子背面的虚拟视角出发，优化镜子内的反射内容。该方法通过分离反射与真实几何信息，显著提升了反射物体的色彩重建精度。此外，针对高反射表面，一些研究还引入了环境光遮蔽和全局光照技术，以更好地模拟反射光线的复杂交互。然而，如何在重建过程中高效处理包含高光照条件变化的区域或物体，并避免局部色彩与全局内容发生混乱，仍是值得研究的重要挑战。下面列举了现有工作在解决相关问题时常用的三种手段。

(1) 基于未知视角合成的方法

在 3D 重建中，视角的局限性常常导致色彩信息的不完整，尤其是在自动驾驶等场景中，车载摄像头通常只能捕捉到部分视角的数据。为了解决这一因素导致的模型外观质量下降，SGD 框架^[44]引入了扩散先验，通过 ControlNet 模块预测深度信息，增强了 3DGS 在训练过程中对未

知视角的合成能力，从而优化了色彩属性的表达。类似地，Gao 等人^[48]提出的 Bootstrap 3DGS 框架利用扩散模型生成缺失视角的几何和色彩信息，进一步提升了色彩重建的精度。然而，这类方法由于引入了扩散模型的生成操作，显著增加了计算复杂度，导致重建时间延长。HGS-Mapping 框架^[53]则通过融合激光雷达和图像数据，对场景中的天空、地面和路边物体分别采用球面高斯、平面高斯和 3D 高斯进行表示，最大程度地优化了缺乏几何信息的天空部分的色彩重建精度。尽管如此，如何在不显著增加计算负担的前提下提升未知视角的色彩合成能力，仍是未来值得研究的重要话题。

(2) 基于再照明和光线追踪的方法

现实场景中的光照强度、观察角度和表面散射等物理因素对物体表面的颜色表现具有显著影响。为更好地模拟这些复杂的光照效果，Gao 等人^[54]在传统的球谐系数基础上，引入了法线、双向反射分布函数 (BRDF) 参数和入射光线信息，以更精确地反映物体表面的色彩变化和阴影效果。这种方法通过物理光照模型优化了色彩属性，显著提升了场景的真实感。然而，每个高斯点需要优化的参数数量高达 11 个，显著增加了训练负担。RayGauss 方法^[55]则通过引入密度量作为光线追踪过程中的辐射吸收因素，分析了光线上所有贡献高斯对图像合成的影响程度，进一步优化了色彩表现。NVIDIA 的研究人员通过构建包围粒子的几何体^[56]，并利用硬件加速技术提升了光线追踪的效率。尽管这些方法在光照模拟方面取得了显著进展，但引入的 BRDF 等参数显著增加了计算复杂度。为此，Jiang 等人^[14]提出了一种简化的着色函数，通过采样最短轴方向来估计表面法线，从而在保证光照效果的同时降低了计算负担。然而，如何在高精度光照模拟与计算效率之间取得平衡，仍是当前研究的难点。

(3) 针对去除图像模糊的方法

在数据采集过程中，镜头散焦、物体运动和相机抖动等因素不可避免地导致图像模糊，从而影响重建时作为监督信号的色彩质量。Byeonghyeon Lee 等人^[57]提出了一种基于 MLP 网络的方法，通过调整 3D 高斯的协方差矩阵来模拟模糊的形成过程，从而优化色彩属性的表达。在推理阶段，该方法使用原始的 3DGS 组件渲染场景，避免了 MLP 的计算开销，从而有效消除散焦模糊。然而，该方法主要针对散焦模糊进行优化，对其他类型的模糊（如运动模糊）效果有限。CRiM-GS 框架^[58]则通过建模相机运动的成像过程，利用神经常微分方程模拟模糊图像的形成，并考虑了刚体运动和可变形体运动对模糊的影响，通过纠正相机位姿来消除运动模糊，从而优化色彩重建的精度。此外，针对 Spike 高速脉冲相机采集的连续时空数据，Guo 等人^[59]提出了首个面向该传感器的 3DGS 框架，包含累积光栅化和间隔监督两大核心模块。累积光栅化通过长时间积累 Spike 数据，提取清晰的纹理信息，减轻时间混叠和运动模糊；间隔监督则通过计算相邻尖峰的时间间隔，推断场景中的动态变化以恢复几何结构，从而优化色彩属性。然而，该方法在寻找最佳积累阈值和间隔参数时面临耗时和挑战性。Zhang 等人^[60]提出的 SpikeGS 框架则采用轻量级映射网络 (SIM) 从 Spike 数据中提取瞬时图像进行光栅化，避免了间隔监督的计算开销，进一步提升了色彩重建的效率。Weng 等人^[61]则针对事件相机提出了 EaDeblur-GS 框架，通过自适应偏差估计器 (ADE) 调整高斯位置的偏差，以应对曝光期间的相机抖动，从而优化色彩重建的鲁棒性。

总结，为了提高模型外观参数对环境光照的模拟质量，现有方法主要围绕未知视角合成、物体反射表面光照增强、环境光再照明、运动去模糊等手段进行考量。相比于原始的 3DGS 方法，在保证算法实时运行的前提下，地图精度得到改善。各实例的解决方案对比如表 3 所示。

表 3 面向模型色彩属性增强的算法实例汇总

Table 3 Collection of Algorithm Instances for Model-Oriented Color Attribute Enhancement

方法	解决的问题	解决方案
SGD [44]	道路物体因观测视角不完整导致重建出现空洞。	采用扩散模型生成伪视角图像充当未知视角+ControlNet 模块提供空间结构信息。
HGS-Mapping [53]	各类传感器都无法提供天空区域的距离信息导致重建模糊。	天空、地面和路边物体分别采用不同高斯基元表示，天空被约束在固定的圆球体表面。

Relightable 3D Gaussians [54]	原始的 3DGS 模型无法描述环境中的阴影光。	额外引入法线、BRDF 参数和入射光线来反映物体表面的色彩。
RayGauss [55]	原始的 3DGS 模型无法描述物体表面的反光效果。	基于密度量的光线追踪方法。
3D Gaussian Ray Tracing [56]	光线追踪计算效率慢。	基于几何体外形自动变形收敛的硬件加速方法。
GaussianShader [14]	原始的 3DGS 模型无法描述物体表面的反光效果。	着色函数建模物体的表面反射。
Mirror-3DGS [52]	镜子等反射物体的重建。	利用平面镜成像原理把镜子背面的视角作为观察视角进行参数优化。
Lee et al. [57]	相机散焦模糊导致的重建精度退化。	使用 MLP 网络模拟模糊的形成,从而消除。
CRiM-GS [58]	相机运动模糊导致的重建精度退化。	利用神经常微分方程模拟模糊的形成,从而消除。
SpikeGS [60]	针对 Spike 高速脉冲相机数据的重建。	累积光栅化和间隔监督组成的框架。
EaDeblur-GS [61]	针对使用事件相机用来消除抖动模糊的问题。	使用自适应偏差估计器解决。

3.3 结合大语言模型和三维高斯泼溅的虚拟世界重建研究

大语言模型在自然语言处理方面表现出色,能够提供丰富的语义信息。通过将大语言模型与 3DGS 结合,可以显著增强对 3D 场景的语义理解能力。例如,Shi 等人^[62]通过从 CLIP 和 DINO 中提取和离散化密集特征,将相应的连续语义向量引入到 3DGS 中,这些特征用于通过 VQ-VAE 中的 MLP 预测离散特征空间中的语义索引。

大语言模型能够处理多种模态的数据,为 3DGS 提供了多模态融合的可能性,使得重建框架具有较强的泛化能力,提高模型在不同场景和任务中的适用性。例如,Zhou 等人^[63]先利用大语言模型 GPT-3.5 生成初始布局,然后提出了一种对象-场景组合优化机制,结合条件扩散,协作更新布局引导的高斯参数获得具有一致几何形状、纹理的场景表示。Text2Immersion 框架^[64]预先定义一组锚定相机位置,以此作为参考点在后续阶段中构建和细化 3D 场景。这种方法通过结合大语言模型和 3DGS,实现了从文本到 3D 场景的高效转换,显著提升了虚拟世界的重构质量。

大语言模型可以提供高效的初始化和优化策略,从而显著提升 3DGS 的重建效率。例如,GaussianObject^[65]介绍了一种基于 Visual Hull 的初始化策略,并利用预训练模型来解决稀疏视点条件下的遮挡和信息丢失问题。此外,GaussianObject 设计了一个高斯修复模块,通过数据采集、训练和推理范式,利用预训练模型来优化 3DGS 的重建过程。Ma 等人^[66]使用预训练的 Diffusion360 模型,根据文本提示几秒钟内就能生成高分辨率全景图。接着 EGformer 模型进行深度估计,通过渐进式新视角修复策略(PNVI)分步并优先处理较小的图像区域,避免一次性处理大范围移动带来全局不一致问题。结果显示,该方法能够在大约 15 分钟内完成 3D 场景的生成。

通过结合大语言模型,3DGS 可以更好地模拟复杂光照条件下的场景,从而在重建结果中保留更多的细节。例如,DreamScene360^[67]算法利用扩散模型 MultiDiffuser 和大语言模型 GPT-4V 经过多轮质量评估生成全景图像。深度估计采用 DPT 模型,并配合配准操作保证视角间有着一致的几何结构。IM-3D 通过微调现有的图像到视频生成模型,支持生成多视点空间一致的图像(视频),这些生成的多视点图像然后被用作 3DGS 生成的监督。这些方法显著提升了虚拟世界的真实感,使其在渲染复杂、动态和高光场景方面超越了现有的方法。

总结,结合大语言模型重构虚拟世界不仅能够显著提升 3DGS 在语义理解、生成能力、重建效率和泛化能力等方面的表现,还能推动多模态融合的发展,为虚拟世界的重构提供了更强大的工具和方法。根据上述实例,可以概括出重建虚拟世界的基本工作框架如下图 3 所示。首先,利用多视角扩散模型根据提示词生成全景草图。然后,利用大语言模型进行草图内容质量的检查,经过多轮修正操作得到最佳候选的全景图。接着,利用深度估计器预测全景深度图。再者,通过反向等距柱状投影获得新生成的初始点云,确保点云的过渡平滑和全局一致性。最后,结合几何和语义损失优化构建高斯虚拟地图。

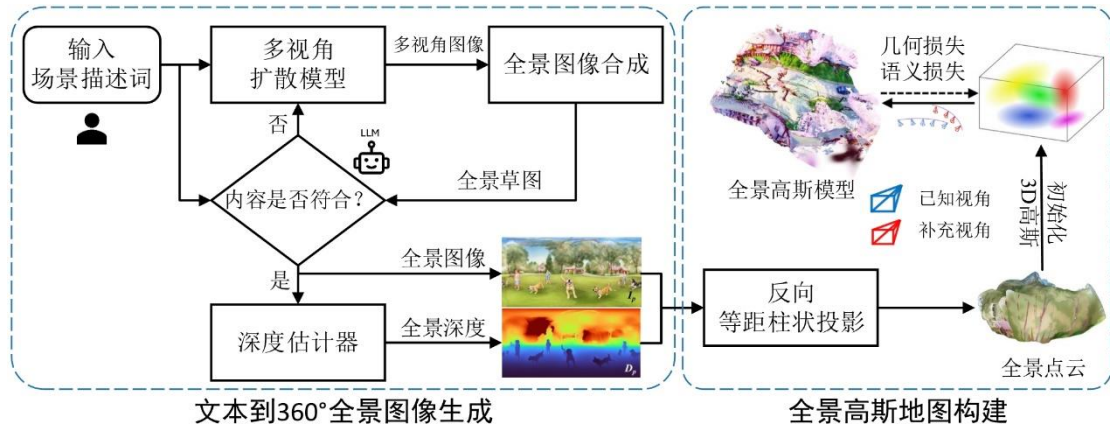


图 3 虚拟世界重建的基本流程图
Figure 3 Basic Flowchart of Virtual World Reconstruction

3.4 城市级动态场景建模的三维高斯泼溅重建研究

现实世界中许多场景具有时间变化的特性，如行驶车辆、人体运动、碰撞变形、河流流体等。大多数 3DGS 方法主要面向静态场景进行重建，无法有效捕捉和表示这些动态变化，故动态场景建模在 3DGS 研究中至关重要。

(1) 多模态数据融合

在多模态数据融合领域，技术的发展呈现出从刚性配准到动态自适应学习的演进脉络。早期工作如 LiDAR-GS^[68]通过密度梯度对齐实现激光雷达与相机数据的联合高斯参数化，重点解决传感器间刚性配准问题，但其固定权重策略在动态场景中表现受限。后续 HybridGS^[69]引入跨模态注意力机制，首次实现了 RGB 与深度信息的动态权重分配，在复杂城市场景中显著提升了语义理解能力。最新进展 DeSiRe-GS^[70]进一步突破时空异步限制，通过可微分传感器残差建模，成功整合激光雷达、相机和毫米波雷达的异构数据流，标志着多模态融合从空间对齐向时空联合建模的跨越式发展。

DHGS^[71]开创性地将语义分割网络与高斯散射过程联合优化，通过 $L_{joint} = \lambda_1 L_{geo} + \lambda_2 L_{sem}$ 实现几何误差与语义一致性同步收敛，在 nuScenes 数据集上语义标注传播准确率达到 81.7%。LiHi-GS^[72]进一步融合 LiDAR 反射强度与高光谱信息，构建了面向城市基础设施的多层级语义表征体系（如区分玻璃幕墙与混凝土墙面），将材质分类精度提升至 94.3%。该方向技术路径呈现几何主导 → 语义辅助 → 联合表征的演进特征。

(2) 动态物体的建模表示与几何失真

动态场景建模方法的技术突破则沿着运动预测精度提升与复杂运动模式刻画两条主线展开。SplatFlow^[73]基于光流引导的高斯点运动预测开创了动态建模新范式，但其线性运动假设难以捕捉周期性动作。为此，Periodic Vibration Gaussian^[74]引入傅里叶基函数参数化高斯核为每个高斯椭球体的均值和不透明度属性赋予生命周期，成功刻画行人步态等周期性运动，从而统一表示场景中的静态和动态元素。Momentum-GS^[75]进一步将物理动量守恒约束融入高斯分布演化，通过 $\partial \Sigma / \partial t = \alpha \cdot F(v) + \beta \cdot M(p)$ 这一微分方程有效抑制突发运动导致的几何失真，形成了从简单运动预测到复杂物理规律建模的技术深化路径。此外，DrivingGaussian 方法^[18]采用了一种分层架构，将场景分解为静态背景和动态对象，并分别重建每一部分。这种方法通过增量静态 3D 高斯构建全面的背景场景，并使用复合动态节点图单独重建每类移动对象，最后将它们动态地集成到静态背景中。这种分层方法不仅提高了重建的效率，还增强了对多种动态对象的跟踪和理解能力。该方向发展逻辑呈现出阶梯式：单类物体运动预测和表征 → 多类对象运动规律表示和建模 → 系统化分层管理架构设计。

(3) 面对多样化环境重建的鲁棒性

在点云稀疏性挑战驱动下，AutoSplat^[76]率先提出基于密度感知的自适应点云聚合算法，通过动态调整高斯核半径实现了城市道路细节的精细化重建，相比传统方法提升 23% 的局部点云覆盖率。DENSER^[77]在此基础上引入概率密度评估函数，构建了基于信息熵的点云筛选机制，

在 KITTI-360 数据集上将无效点云剔除率提升至 68%，同时保持关键结构完整性。针对雨雾雪等恶劣天气场景，HUGS^[78]提出基于对抗训练的高斯散射稳定性框架，通过生成对抗网络模拟极端天气下的传感器噪声分布，使重建系统在能见度小于 50m 时的场景完整性保持率达 58%。该方法突破传统物理建模局限。这类方法的发展逻辑体现为：主动优化数据质量→自适应动态调整→对抗增强，为大规模场景重建奠定了数据基础。

(4) 城市级动态场景的渲染效率

大规模城市场景优化技术经历了从粗粒度压缩到精细化表征的迭代过程。CityGaussian^[79]提出的分块 LOD 策略首次实现平方公里级场景的高效渲染，但其均匀分块机制导致细节丢失。Street Gaussians 方法^[6]在 3DGS 模型的基础上为 3D 高斯配备了语义 logits 属性，区分前景车辆和背景，并对动态前景目标和静态背景分别表示。动态目标按照实例 ID 管理位置、旋转和外观参数，当需要渲染图像时会从局部坐标系切换到世界坐标系下联合静态背景进行计算。此外，为了减小动态目标每一时刻下球谐系数的存储开销，使用一组傅里叶变换系数代替。最新提出的 Urban4D^[80]融合时序变化检测技术，不仅实现城市级场景的高精度重建，更支持数字孪生体的增量更新，推动 3DGS 从静态场景重建向动态城市演化的跨越。这一演进过程体现了从空间规模扩展到时空维度深化的技术发展逻辑。

总结，城市级动态场景建模能够大大扩展 3DGS 的应用范围，使其能够处理时间维度上的变化。通过引入基于物理属性的变形场、时间参数化等方法，3DGS 能够有效捕捉和表示动态场景中的变化，为数字人、动画制作、自动驾驶等领域的应用提供了强有力的技术支持。下表 4 显示了在两个公开数据集下多个著名重建算法在重建自动驾驶场景中的性能对比，除了比较已知相机视角的重建精度外，还添加了对未知视角的重建精度比较，具体指标为 PSNR、SSIM、LPIPS 以及模型渲染效率（以 FPS 为单位）。每个指标的前三名分别用红色、绿色和黄色块标出。表格中的符号“-”表示原文献缺少相应的场景测试。Waymo 数据集下重建精度最优的前三名依次是：OmniRe、SplatFlow 和 DeSiRe-GS 算法。KITTI 数据集下重建精度最优的前三名依次是：DeSiRe-GS、SplatFlow 和 PVG 算法。渲染效率最快的前三名是：3DGS、PVG 和 SplatFlow。虽然 DeSiRe-GS 重建精度较高，但由于需要额外渲染法线等物理属性，导致计算速度比 PVG 和 SplatFlow 略慢。整体看来，SplatFlow 算法权衡了效率和精度，在多个指标下均排名靠前。

表 4 两个数据集上多个重建算法的精度和效率比较

Table 4 Comparison of Accuracy and Efficiency of Multiple Reconstruction Algorithms on Two Datasets

方法	Waymo Dataset							KITTI Dataset						
	重建精度			新视角合成精度			FPS↑	重建精度			新视角合成精度			FPS↑
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
3DGS[4]	27.99	0.866	0.293	25.08	0.822	0.319	63	21.02	0.811	0.202	19.54	0.776	0.224	125
PVG[74]	32.46	0.910	0.229	28.11	0.849	0.279	50	32.83	0.937	0.070	27.43	0.896	0.114	59
SplatFlow[73]	33.64	0.951	0.198	28.71	0.874	0.239	40	33.37	0.943	0.057	28.32	0.932	0.089	44
SUDS[81]	28.83	0.805	0.317	25.36	0.783	0.384	0.008	28.83	0.917	0.147	26.07	0.797	0.131	0.29
DeSiRe-GS[70]	33.61	0.919	0.204	29.75	0.878	0.213	36	33.94	0.949	0.04	28.87	0.901	0.106	41
OmniRe[82]	34.25	0.954	-	32.57	0.942	-	-	-	-	-	-	-	-	-
DeformGS[83]	28.40	0.929	-	27.72	0.922	-	-	-	-	-	-	-	-	-
StreetGS[6]	29.08	0.936	-	28.54	0.928	-	-	-	-	-	-	-	-	-

4 讨论和挑战

如前面第 2 节和第 3 节所述，尽管 3DGS 重建方法在解决“静态模型存储成本、模型外观色彩属性优化、虚拟世界重构、城市级动态场景建模”四个主要方向取得了显著进展，但在研究过程中出现了一些新的的问题亟待解决。本节中，对上述内容进行讨论分析，概括出不同框架和算法出现的新挑战有哪些，以便为初入该领域的研究人员研究这一迅速发展的技术提供更多的研究方向。

(1) 几何表面与高斯分布的错位

3DGS 通过显式的高斯点云表示场景，但高斯点的 3D 协方差矩阵难以精确对应真实表面几何。例如，SuGaR^[84]通过几何正则化约束高斯点分布，但其提取的网格仍存在噪声和表面起伏。这种错位在锐利边缘或薄结构等复杂几何结构中尤为明显，导致重建表面细节模糊或断裂。此

外，高斯点的参数优化对初始化敏感，若初始点云存在偏差，可能导致表面不连续或孔洞。例如，在 ETH3D 数据集的反射地板场景中，初始点云因镜面反射难以捕捉真实几何，导致重建表面出现异常波动。

(2) 多样材质表面

面对多样材质表面时，色彩还原度与真实世界的贴合度仍不理想。比如在室内装修场景重建中，对于多种材质混合且光照复杂的空间，模型的外观色彩容易出现偏差，一些金属材质的光泽感和反射效果无法精准呈现，木质纹理的色彩过渡不自然，这使得重建场景无法真实还原人眼所观察到的物理世界色彩，影响后续的视觉分析和利用。

(3) 物理和语义感知的场景表示

当前的 3DGS 重建方法主要关注几何和外观建模，缺乏对物理属性的整合，例如材料特性和动态行为。建模弹性或可变形物体需要更深入地理解物理原理，而现有的 3DGS 重建框架尚未完全解决这一问题。尽管一些工作如 PhysGS^[85]引入了弹簧-质量系统来模拟物体的物理行为，每个 3D 高斯体都被视为一个质量点，而高斯体之间的相互作用则通过弹簧来建模。这种方法能够有效地模拟物体的弹性变形和动态行为，但它们仅限于特定场景，缺乏通用性。

(4) 存储和计算资源的消耗

为了应对复杂光照条件下的场景重建，3DGS 重建框架需要引入额外的描述参数来增强其对场景的描述能力。这些参数包括漫反射颜色、BRDF 参数、物体材质等，它们对于捕捉场景中的光照变化和物理特性至关重要。然而，这也意味着优化过程中需要处理更多的变量，无疑增加了计算的复杂性和训练的时间成本。Lin 等人提出的 VastGaussian^[5]通过渐进式数据分区策略来优化大规模场景的重建，但在处理超大规模场景时，仍然存在存储空间和渲染速度的问题。

(5) 动态建模与长期跟踪的挑战

动态 3DGS 重建方法在长序列跟踪中易出现运动不连续问题。例如，Yang 等人^[86]的 MLP 形变场在时间插值任务中因姿态误差累积导致伪影。基于扩散先验的 4D 生成在运动合理性与物理一致性上存在不足。例如，AYG^[87]生成的变形场缺乏碰撞约束，导致物体穿透等非物理现象。

5 总结和展望

根据上述发现的新挑战，结合相关领域知识，在场景重建方向上未来可能做的研究方向有：

(1) 独立且高效的模型编辑

开发先进的渲染技术，以独立提取几何、纹理和照明，实现 3DGS 的高效且独立编辑。这可以通过利用基于物理的渲染和逆向渲染技术来实现。例如像 GS-IR^[88]和 GIR^[89]这样的方法通过将 3DGS 分解为几何和照明组件，可以实现独立编辑，这表明进一步研究可以导致更高效的编辑工作流程。

(2) 分布式、层次化计算与存储

探索如何通过分布式计算和存储技术来处理超大规模场景的重建。例如，通过将场景分割为多个子区域，并在多个计算节点上并行处理，可以显著提高重建效率。引入层次化的场景表示方法，可以在不同细节层次上动态调整高斯分布的密度，从而在保证渲染质量的同时减少存储和计算开销。

(3) 模型跨平台部署和传输

3DGS 使用三维高斯函数作为基本元素，这种显式表示使得 3DGS 能够直接操作和编辑这些高斯函数，而不需要像 NeRF 方法通过隐式场的体积渲染来合成像素颜色。为了减少模型对内存和传输带宽的需求，研究针对 3D 高斯进行分批分块处理，以实现渐进渲染。还可以通过并行计算、实时光线追踪技术、空间索引等方式，提高 3DGS 在复杂动态场景下的渲染效率。

(4) 物理模拟与 3DGS 集成搭建通用仿真框架

从先前的工作中发现，除了 3D 高斯原本的 5 个基本属性外还可以拓展其他属性以满足特定任务的需要。因此，可以探索将其他物理因素集成到 3DGS 中，如材料属性、力学响应和环

境交互，从而搭建近似物理世界的仿真环境供其他现实任务使用。这涉及到开发新的算法来模拟和渲染物理上准确的物体行为，如布料的飘动、液体的流动和物体的碰撞响应等。

参考文献

- [1] Mildenhall B, Srinivasan PP, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. *Communications of the ACM*, 2021, 65(1): 99-106.
- [2] Azinovic D, Martin-Brualla R, Goldman DB, et al. Neural RGB-D Surface Reconstruction[C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 6280-6291.
- [3] Barron JT, Mildenhall B, Verbin D, et al. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields[C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 5460-5469.
- [4] Kerbl B, Kopanas G, Leimkuhler T, et al. 3d gaussian splatting for real-time radiance field rendering[J]. *ACM Transactions on Graphics*, 2023, 42(4): 1-14.
- [5] Lin J. VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2402.17427, 2024.
- [6] Yan YZ, Lin HT, Zhou CX, et al. Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2401.01339, 2024.
- [7] Wu G. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering[Z/OL]. arXiv preprint arXiv: 2310.08528, 2023.
- [8] Shao R, Zheng Z, Tu H, et al. Tensor4D : Efficient Neural 4D Decomposition for High-fidelity Dynamic Reconstruction and Rendering[Z/OL]. arXiv preprint arXiv: 2211.11610, 2022.
- [9] Zhu DL, Xu GH, Wang XT, et al. PairCon-SLAM: Distributed, Online, and Real-Time RGBD-SLAM in Large Scenarios[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70(10): 1-14.
- [10] Zhu DL, Wang ZL, Lu TF, et al. PMF-SLAM: Pose-Guided and Multiscale Feature Interaction-Based Semantic SLAM for Autonomous Wheel Loader[J]. *IEEE Sensors Journal*, 2024, 24(7): 11625-11638.
- [11] Yan, C. GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2311.11700, 2023.
- [12] Zhu Z, Peng SY, Larsson V, et al. NICE-SLAM: Neural Implicit Scalable Encoding for SLAM[C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 12776-12786.
- [13] Fan Z, Wang K, Wen K, et al. LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS[Z/OL]. arXiv preprint arXiv:2311.17245, 2023.
- [14] Jiang Y. GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces[Z/OL]. arXiv preprint arXiv:2311.17977, 2023.
- [15] Wu T. Recent Advances in 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2403.11134, 2024.
- [16] Bao Y. 3D Gaussian Splatting: Survey, Technologies, Challenges, and Opportunities[Z/OL]. arXiv preprint arXiv: 2407.17418, 2024.
- [17] Fei B, Xu J, Zhang R, et al. 3D Gaussian Splatting as New Era: A Survey[J]. *IEEE Transactions on Visualization and Computer Graphics*, Early Access, 2024.
- [18] Zhou XY, Lin ZW, et al. DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes[Z/OL]. arXiv preprint arXiv: 2312.07920, 2023.
- [19] Learning-based Multi-View Stereo: A Survey
- [20] 洪忠铖. 基于代价体积的自监督单目深度估计方法研究[D]. 华南理工大学, 2023.
- [21] 余瑛, 万贻前, 洪瀚, 等. 基于符号距离函数的体渲染表面重建研究综述[J]. *计算机应用研究*, 2024, 41(12): 3533-3542.
- [22] 朱德榕, 贺赛先. 基于深度学习的三维体素重建改进方法[J]. *激光杂志*, 2021, 42(08): 39-44.
- [23] 侯耀斐, 黄海松, 范青松, 等. 基于改进多层感知机的神经辐射场三维重建方法[J]. *激光与光电子学进展*, 2024, 61(04): 297-307.
- [24] Mohamed S, John G, Jamie W, et al. SimpleRecon: 3D Reconstruction Without 3D Convolutions[C] // 2022 European Conference on Computer Vision (ECCV), 2022: 1-19.
- [25] Xin YY, Zuo XX, Lu DY, et al. SimpleMapping: Real-Time Visual-Inertial Dense Mapping with Deep Multi-View Stereo[C] // 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2023: 273-282.
- [26] Kim T, Choi J, Choi S, et al. Just a Few Points are All You Need for Multi-view Stereo: A Novel Semi-supervised Learning Method for Multi-view Stereo[C] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 6158-6166.

-
- [27] Li Z, Chen Z, Liu X, et al. DepthFormer: Exploiting Long-Range Correlation and Local Information for Accurate Monocular Depth Estimation[Z/OL]. arXiv preprint arXiv: 2203.14211, 2022.
- [28] Godard C., Mac Aodha O., Firman M., et al. Digging Into Self-Supervised Monocular Depth Estimation[Z/OL]. arXiv preprint arXiv: 1806.01260, 2018.
- [29] Chang D. RC-MVSNet: Unsupervised Multi-View Stereo with Neural Rendering[Z/OL]. arXiv preprint arXiv: 2203.03949, 2022.
- [30] Peng R, Wang R, Wang Z, et al. Rethinking Depth Estimation for Multi-View Stereo: A Unified Representation[Z/OL]. arXiv preprint arXiv: 2201.01501, 2022.
- [31] Rich A, Stier N, Sen P, et al. 3DVNet: Multi-View Depth Prediction and Volumetric Refinement[C] // 2021 International Conference on 3D Vision (3DV), 2021: 700-709.
- [32] Oleynikova H, Taylor Z, Fehr M, et al. Voxblox: Incremental 3D Euclidean Signed Distance Fields for on-board MAV planning[C] // 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017: 1366-1373.
- [33] Liu S, Chen J, Zhu, J. HVOFusion: Incremental Mesh Reconstruction Using Hybrid Voxel Octree[Z/OL]. arXiv preprint arXiv: 2404.17974, 2024.
- [34] Wang S, Leroy V, et al. DUST3R: Geometric 3D Vision Made Easy[C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 20697-20709.
- [35] Wang H, Agapito, L. 3D Reconstruction with Spatial Memory[Z/OL]. arXiv preprint arXiv: 2408.16061, 2024.
- [36] Guo M, Zhang Y, Liu, T. Gaussian Transformer: A Lightweight Approach for Natural Language Inference[C] // 2019 AAAI Conference on Artificial Intelligence, 2019: 6489-6496.
- [37] Wang P, Liu L, Liu Y, et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction[Z/OL]. arXiv preprint arXiv:2106.10689, 2021.
- [38] Pumarola A, Corona E, Pons-Moll G, et al. D-NeRF: Neural Radiance Fields for Dynamic Scenes[Z/OL]. arXiv preprint arXiv: 2011.13961, 2020.
- [39] Wang Z, Li L, Shen Z, et al. 4K-NeRF: High Fidelity Neural Radiance Fields at Ultra High Resolutions[Z/OL]. arXiv preprint arXiv: 2212.04701, 2022.
- [40] Müller T, Evans A, Schied C, et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding[Z/OL]. arXiv preprint arXiv: 2201.05989, 2022.
- [41] Niedermayr S, Stumpfegger J, Westermann, R. Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis[Z/OL]. arXiv preprint arXiv:2401.02436, 2023.
- [42] Xie H, Chen Z, Hong F, et al. GaussianCity: Generative Gaussian Splatting for Unbounded 3D City Generation[Z/OL]. arXiv preprint arXiv:2406.06526, 2024.
- [43] Lee JC, Rho D, Sun X, et al. Compact 3D Gaussian Representation for Radiance Field[Z/OL]. arXiv preprint arXiv:2311.13681, 2023.
- [44] Yu Z. SGD: Street View Synthesis with Gaussian Splatting and Diffusion Prior[Z/OL]. arXiv preprint arXiv:2403.20079, 2024.
- [45] Lu T. Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering[Z/OL]. arXiv preprint arXiv:2312.00109, 2023.
- [46] Chen Y, Wu Q, Lin W, et al. HAC: Hash-grid Assisted Context for 3D Gaussian Splatting Compression[Z/OL]. arXiv preprint arXiv:2403.14530, 2024.
- [47] Gong Y. GDGS: Gradient Domain Gaussian Splatting for Sparse Representation of Radiance Fields[Z/OL]. arXiv preprint arXiv:2405.05446, 2024.
- [48] Gao Y, Ou J, Wang L, et al. Bootstrap 3D Reconstructed Scenes from 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2404.18669, 2024.
- [49] Morgenstern W, Barthel F, Hilsmann A, et al. Compact 3D Scene Representation via Self-Organizing Gaussian Grids[Z/OL]. arXiv preprint arXiv:2312.13299, 2023.
- [50] Chen Y, Wu Q, Lin W, et al. HAC++: Towards 100X Compression of 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2501.12255, 2025.
- [51] Wu M. and Tuytelaars T. Implicit gaussian splatting with efficient multi-level tri-plane representation. arXiv preprint arXiv:2408.10041, 2024.
- [52] Meng J. Mirror-3DGS: Incorporating Mirror Reflections into 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2404.01168, 2024.
- [53] Wu K. HGS-Mapping: Online Dense Mapping Using Hybrid Gaussian Representation in Urban Scenes[Z/OL]. arXiv preprint arXiv:2403.20159, 2024.
- [54] Gao J. Relightable 3D Gaussians: Realistic Point Cloud Relighting with BRDF Decomposition and Ray Tracing[Z/OL]. arXiv preprint arXiv:2311.16043, 2023.
- [55] Blanc H, Deschaud JE, Paljic A. RayGauss: Volumetric Gaussian-Based Ray Casting for Photorealistic Novel View Synthesis[Z/OL]. arXiv preprint arXiv:2408.03356, 2024.
- [56] Moenne-Loccoz N. 3D Gaussian Ray Tracing: Fast Tracing of Particle Scenes[Z/OL]. arXiv preprint arXiv:2407.07090, 2024.

-
- [57] Lee B, Lee H, Sun X, et al. Deblurring 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2401.00834, 2024.
- [58] Lee J, Kim D, Lee D, et al. CRiM-GS: Continuous Rigid Motion-Aware Gaussian Splatting from Motion Blur Images[Z/OL]. arXiv preprint arXiv:2407.03923, 2024.
- [59] Guo Y, Hu L, Ma L, et al. SpikeGS: Reconstruct 3D scene via fast-moving bio-inspired sensors[Z/OL]. arXiv preprint arXiv:2407.03771, 2024.
- [60] Zhang J, Chen K, Chen S, et al. SpikeGS: 3D Gaussian Splatting from Spike Streams with High-Speed Camera Motion[Z/OL]. arXiv preprint arXiv:2407.10062, 2024.
- [61] Weng Y, Shen Z, Chen R, et al. EaDeblur-GS: Event assisted 3D Deblur Reconstruction with Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2407.13520, 2024.
- [62] Shi Y, Wang P, Ye J, et al. Mvdream: Multi-view diffusion for 3d generation[Z/OL]. arXiv preprint arXiv: 2308.16512, 2023.
- [63] Zhou X. GALA3D: Towards Text-to-3D Complex Scene Generation via Layout-guided Generative Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2402.07207, 2024.
- [64] Ouyang H, Heal K, Lombardi S, et al. Text2Immersion: Generative Immersive Scene with 3D Gaussians[Z/OL]. arXiv preprint arXiv:2312.09242, 2023.
- [65] Yang C. GaussianObject: High-Quality 3D Object Reconstruction from Four Views with Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2402.10259, 2024.
- [66] Ma Y, Zhan D, Jin Z. FastScene: Text-Driven Fast 3D Indoor Scene Generation via Panoramic Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2405.05768, 2024.
- [67] Zhou S. DreamScene360: Unconstrained Text-to-3D Scene Generation with Panoramic Gaussian Splatting[Z/OL]. arXiv preprint arXiv:2404.06903, 2024.
- [68] Chen Q, Yang S, Du S, et al. LiDAR-GS:Real-time LiDAR Re-Simulation using Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2410.05111, 2024.
- [69] Lin J. HybridGS: Decoupling Transients and Statics with 2D and 3D Gaussian Splatting[Z/OL]. arXiv preprint arXiv: 2412.03844, 2024.
- [70] Peng C. DeSiRe-GS: 4D Street Gaussians for Static-Dynamic Decomposition and Surface Reconstruction for Urban Driving Scenes[Z/OL]. arXiv preprint arXiv: 2411.11921, 2024.
- [71] Shi X. DHGS: Decoupled Hybrid Gaussian Splatting for Driving Scene[Z/OL]. arXiv preprint arXiv: 2407.16600, 2024.
- [72] Kung PC, Zhang X, Skinner KA, et al. LiHi-GS: LiDAR-Supervised Gaussian Splatting for Highway Driving Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2412.15447, 2024.
- [73] Wang B. SplatFlow: Learning Multi-frame Optical Flow via Splatting[Z/OL]. arXiv preprint arXiv:2306.08887, 2023.
- [74] Chen YR, Gu C, Jiang JZ, et al. Periodic Vibration Gaussian(PVG): Dynamic Urban Scene Reconstruction and Real-time Rendering[Z/OL]. arXiv preprint arXiv: 2311.18561, 2023.
- [75] Fan J, Li W, Han Y, et al. Momentum-GS: Momentum Gaussian Self-Distillation for High-Quality Large Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2412.04887, 2024.
- [76] Khan M. AutoSplat: Constrained Gaussian Splatting for Autonomous Driving Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2407.02598, 2024.
- [77] Mohamad MA, Elghazaly G, Hubert A, et al. DENSER: 3D Gaussians Splatting for Scene Reconstruction of Dynamic Urban Environments[Z/OL]. arXiv preprint arXiv: 2409.10041, 2024.
- [78] Zhou HY, Shao JH, Xu L, et al. HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting[C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 21336-21345.
- [79] Liu Y, Luo C, Mao Z, et al. CityGaussianV2: Efficient and Geometrically Accurate Reconstruction for Large-Scale Scenes[Z/OL]. arXiv preprint arXiv: 2411.00771, 2024.
- [80] Li, Z. Urban4D: Semantic-Guided 4D Gaussian Splatting for Urban Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2412.03473, 2024.
- [81] Turki H, Zhang JY, Ferroni F, et al. SUDS: Scalable Urban Dynamic Scenes[Z/OL]. arXiv preprint arXiv: 2303.14536, 2023.
- [82] Chen Z. OmniRe: Omni Urban Scene Reconstruction[Z/OL]. arXiv preprint arXiv: 2408.16760, 2024.
- [83] Duisterhof BP. DeformGS: Scene Flow in Highly Deformable Scenes for Deformable Object Manipulation[Z/OL]. arXiv preprint arXiv: 2312.00583, 2023.
- [84] Guedon A and Lepetit V. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering[C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 5354-5363.
- [85] Xie T, et al. PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics[C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 4389-4398.
- [86] Yang Z, Wang J, Li L, et al. Idea2img: Iterative self-refinement with gpt-4v (ision) for

-
- automatic image design and generation[Z/OL]. arXiv preprint arXiv: 2310.08541, 2023.
- [87] Ling H, Kim SW, Torralba A, et al. Align Your Gaussians: Text-to-4D with Dynamic 3D Gaussians and Composed Diffusion Models[C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 8576-8588.
- [88] Liang Z, Zhang Q, Feng Y, et al. GS-IR: 3D Gaussian Splatting for Inverse Rendering," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [C] // 2024: 21644-21653.
- [89] Shi Y. GIR: 3D Gaussian Inverse Rendering for Relightable Scene Factorization[Z/OL]. arXiv preprint arXiv: 2312.05133, 2023.