

DOI: 10.13973/j.cnki.robot.240088

## 基于 3D 高斯溅射的 3 维重建技术综述

曹振中<sup>1,2,3</sup>, 光金正<sup>1,2,3</sup>, 张千一<sup>1,2,3</sup>, 胡郑希<sup>1,2,3</sup>, 吴仕超<sup>1,2,3</sup>, 刘景泰<sup>1,2,3</sup>

(1. 南开大学人工智能学院机器人与信息自动化研究所, 天津 300350;

2. 天津市智能机器人技术重点实验室, 天津 300350; 3. 可信行为智能算法与系统教育部工程研究中心, 天津 300350)

**摘要:** 2023 年底推出的基于 3D 高斯溅射的 3 维重建技术相较于该领域主流传统算法在新视图合成质量和场景渲染速度两方面具有突出优势, 本文分析了 3D 高斯溅射的技术精髓和巨大的应用潜力, 全面梳理了基于 3D 高斯溅射的 3 维重建技术在视觉 SLAM (同步定位与地图构建)、动态场景重建、生成式 AI (人工智能) 以及自动驾驶领域的应用, 最后从科研产出和实际落地两方面展望了 3D 高斯溅射技术的潜力和技术迭代趋势。

**关键词:** 3D 高斯溅射; 3 维重建; 机器人; 视觉 SLAM (同步定位与地图构建); 动态场景重建; 生成式人工智能; 自动驾驶

### Survey of 3D Reconstruction Techniques Based on 3D Gaussian Splatting

CAO Zhenzhong<sup>1,2,3</sup>, GUANG Jinzheng<sup>1,2,3</sup>, ZHANG Qianyi<sup>1,2,3</sup>, HU Zhengxi<sup>1,2,3</sup>, WU Shichao<sup>1,2,3</sup>,  
LIU Jingtai<sup>1,2,3</sup>

(1. Institute of Robotics and Automatic Information System, College of Artificial Intelligence, Nankai University, Tianjin 300350, China;

2. Tianjin Key Laboratory of Intelligent Robotics, Tianjin 300350, China;

3. Engineering Research Center of Trusted Behaviour Intelligence, Ministration of Education, Tianjin 300350, China)

**Abstract:** The 3D reconstruction technology based on 3D Gaussian splatting (3D GS), launched at the end of 2023, has outstanding advantages in both the quality of new view synthesis and the speed of scene rendering compared to mainstream traditional algorithms in the field. This paper analyzes the essence of the 3D GS technology and its huge potential for application, comprehensively sorting out the application of 3D GS-based 3D reconstruction technology in the field of visual SLAM (simultaneous localization and mapping), dynamic scene reconstruction, generative AI (artificial intelligence), and autonomous driving. Finally, the potential and technological iteration trend of 3D GS technology are discussed from the perspectives of scientific research output and practical implementation.

**Keywords:** 3D Gaussian splatting; 3D reconstruction; robot; visual SLAM (simultaneous localization and mapping); dynamic scene reconstruction; generative artificial intelligence; autonomous driving

3 维重建指的是给定一个场景的多个视角图像, 重建出这个场景的 3 维模型, 是机器人感知模块中的关键技术之一。3 维重建使机器人能够更好地感知、理解所处的环境并与之交互, 在机器人领域的应用非常广泛, 包括环境感知与建模、导航与路径规划、物体识别与抓取、人机交互和人机共融等。这些机器人相关功能不仅展示了 3 维重建技术在机器人领域的广泛应用, 而且突显了其在提升机器人智能水平方面的关键作用。

传统的 3 维重建方法发展比较成熟, 通常分为 4 个步骤: 运动结构恢复 (SfM<sup>[1]</sup>)、多视立体视觉 (MVS)、表面重建以及纹理渲染。但传统重建方法在面对复杂场景时效率低且精度差, 而神经辐射场

(NeRF)<sup>[2]</sup> 的出现, 给 3 维重建领域带来了革新和生命力。NeRF 是一种利用稀疏视角图像重建场景的 3 维表示的隐式学习技术, 通过对场景的光照和颜色信息进行编码, 能够生成新的视角图像, 实现高质量的 3 维重建和视角合成效果。NeRF 的提出标志着 3 维重建领域的一次里程碑突破, 其高度逼真的渲染效果和对复杂场景的处理能力超过了以往的主流方法。

NeRF 的成功实质上是深度学习在 3 维重建领域的成功。但是使用深度神经网络来建模一个连续体场景, 对场景中的细节进行充分还原, 实现逼真的视觉效果, 需要付出巨大代价。一方面, 利用神经网络表达场景时需要大量的训练时间和渲染资

源,无法实现实时渲染;另一方面,采用隐式辐射场表达方式时很难对场景进行编辑。这些技术局限导致 NeRF 技术在一些对实时性要求很强的机器人应用中的发展受到极大的挑战与限制,例如环境建模、导航以及路径规划等。

在此背景下,3 维高斯溅射(3D GS)技术<sup>[3]</sup>应运而生。3D GS 技术改变了 NeRF 技术的底层范式,彻底颠覆了 NeRF 的整个架构,从场景表示和渲染方式上作出了革新。3D GS 技术一般使用数以万计的 3D 高斯基元来紧凑地表达某个场景,然后通过分块并行光栅化的方式实现高效的图像渲染。这种紧凑表达和高效渲染使得 3D GS 技术变得非常实用。在不影响重建质量和视觉效果的前提下,3D GS 实现了实时渲染。另外,3D GS 这种显式场景表达方式提高了场景的可编辑性,在 3 维重建和新视图合成领域达到了新的高度,基于 3D GS 的改进技术和下游应用也层出不穷<sup>[4]</sup>。

## 1 3D GS 技术框架概述(Overview of 3D GS technology framework)

本节主要从 3D GS 的核心技术原理出发,凝练出 3D GS 的整体架构,对 3D GS 的各个模块进行详细的解释和分析,并指出 3D GS 技术相较于 NeRF 技术的优势所在。

**定义:** 为了避免歧义,文中所提到的 **3D 高斯基元**指的是用于场景表达的基于 3 维高斯数学模型的基元,结合基元的形状信息,也称为 **3D 高斯椭球**。

### 1.1 3D GS 技术的核心原理

图 1 展示了 3D GS 技术的整体架构,主要分为正向误差计算和反向梯度传播 2 个数据流,包含 6 个模块: SfM 点云初始化、3D 高斯椭球集初始化、3D 椭球参数投影、光栅化图像渲染、损失计算以及自适应密度控制,整体流程就是正向计算和反向

优化的过程,接下来将详细介绍这 6 个模块。

#### 1.1.1 SfM 点云初始化

在获取到场景的一组 2D 图像后,使用 SfM 技术从这一组 2D 图像中恢复出场景的稀疏 3D 点云。SfM 是一种运动结构恢复技术,可以利用多视角图像估计 3D 结构,这里通过常用的 COLMAP 库<sup>[5]</sup>来实现。

通过一组图像估计点云的初始坐标值,一方面可以很好地保持连续体辐射场的理想特性,有利于后续的场景优化,另一方面也避免了空白空间中不必要的计算。

另外,3D GS 技术也可以使用随机初始化的点云,但是整个优化过程的收敛会比较慢,训练时间长。

#### 1.1.2 3D 高斯椭球集初始化

在 SfM 点云初始化步骤中得到的每个点云都将被初始化为一个 3D 高斯椭球,具体来说,对于每个点云,用一个包含位置信息和形状信息的 3D 高斯函数来生成 3D 高斯椭球:

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

式(1)的表示方式是 3D 高斯分布的概率密度函数表达形式,其中,位置信息通过 3D 高斯椭球的中心点即均值向量  $\boldsymbol{\mu}$  表达。形状信息(包含旋转和缩放)通过 3D 高斯椭球的协方差矩阵  $\boldsymbol{\Sigma}$  表达,可以拆解为旋转矩阵  $\mathbf{R}$  和缩放矩阵  $\mathbf{S}$  两个分量,用下列公式计算:

$$\boldsymbol{\Sigma} = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T \quad (2)$$

式(2)所表达的几何意义是先将 3D 高斯椭球旋转到与椭球世界平齐,然后沿着轴缩放,再旋转还原回去,这种拆解方式既能保证协方差矩阵的半正定性,又能降低矩阵之间的运算难度。

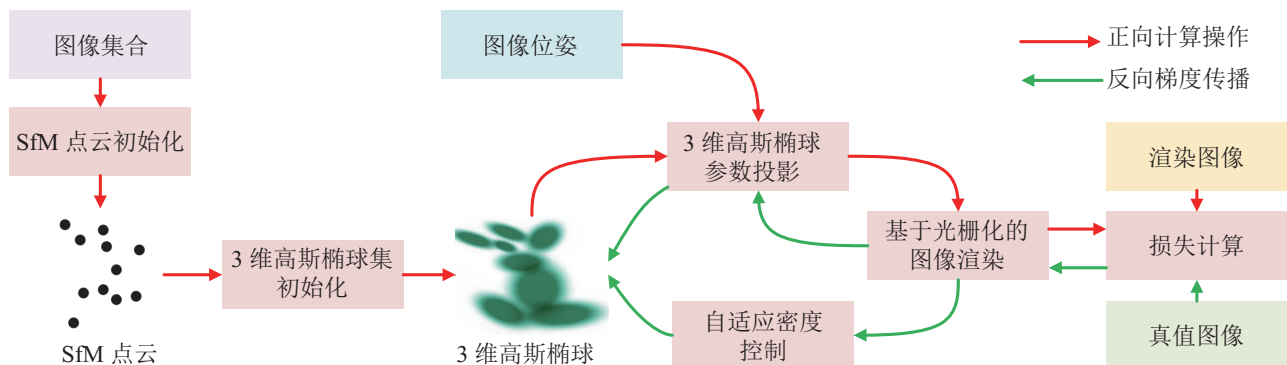


图 1 3D 高斯溅射技术整体架构

Fig.1 Overall architecture of 3D Gaussian splatting technology

对于 3D 高斯椭球的颜色信息  $(r, g, b)$  这种离散值, 使用连续的球谐 (spherical harmonics) 函数<sup>[6]</sup>来表示, 球谐函数本质上是一组能表达球面上不同位置值的基函数, 阶数越高, 表达能力越强, 使得点云在不同角度呈现出不同的颜色, 另外这种连续的函数表达形式也有利于后续的梯度计算和参数优化。对于 3D 高斯椭球的不透明度信息, 直接使用可学习优化的点云不透明度  $\alpha$  表示。

### 1.1.3 3D 椭球参数投影

得到 3D 高斯椭球后, 需要将其投影到 2 维图像平面进行后续渲染操作, 根据计算机图形学中的投影关系, 利用以下公式将空间中的 3D 高斯椭球投影到图像平面上:

$$\Sigma' = \mathbf{J}\mathbf{W}\Sigma\mathbf{W}^T\mathbf{J}^T \quad (3)$$

其中,  $\mathbf{W}$  是视图变换矩阵 (世界坐标系到相机坐标系的变换),  $\mathbf{J}$  是投影变换的仿射近似的雅可比矩阵。 $\Sigma'$  是投影在图像平面后的 2D 高斯椭球的协方差矩阵。

### 1.1.4 基于光栅化的图像渲染

3D GS 使用基于有序点叠加的渲染方法<sup>[7-10]</sup>, 即通过点云中一定半径范围内能影响的有序点集合  $N$  来计算一个像素的颜色  $C$ :

$$C = \sum_{i \in N} c_i \alpha'_i \prod_{j=1}^{i-1} (1 - \alpha'_j) \quad (4)$$

其中,  $c_i$  表示点颜色,  $\alpha'_i$  表示最终不透明度, 用原始不透明度和高斯函数的乘积计算:

$$\alpha'_i = \alpha_i e^{-\frac{1}{2}(\mathbf{x}'_i - \mathbf{u}_i)^T (\Sigma')^{-1} (\mathbf{x}'_i - \mathbf{u}_i)} \quad (5)$$

公式中累乘项表示当前点受排序后前面点不透明度的影响权重,  $\alpha'_j$  越大, 不透明度越大,  $1 - \alpha'_j$  越小, 权重越小, 当前点在颜色渲染时起到的作用越小, 反之, 作用越大。3D GS 采用一种高效并行的分块光栅化方法完成上述的渲染操作。具体的步骤为: 首先, 将要渲染的整个图像分为  $16 \times 16$  大小的图像块, 从每个图像块视锥内挑选可视的置信度大于一定阈值的 3D 高斯基元, 并按 3D 高斯基元到图像平面的深度值进行排序, 为每个图像块单独开一个线程, 然后并行地在每个图像块上从近到远对 3D 高斯基元进行溅射操作, 对溅射留下的痕迹进行堆叠累积, 然后作光栅化处理, 从堆叠的溅射痕迹中划分像素网格来生成像素值, 每个图像块一旦有像素的不透明度达到饱和, 就停止对应线程。

### 1.1.5 损失计算

3D GS 算法对渲染图像和真实图像之间误差的计算如下:

$$L = (1 - \lambda)L_1 + \lambda L_{\text{SSIM}} \quad (6)$$

其中,  $\lambda$  是平衡 2 个误差项的系数。  $L_1$  表示图像中  $M$  个像素的绝对值灰度误差, 用以下公式计算:

$$L_1 = \frac{1}{M} \sum_{i=1}^M |I(i) - \hat{I}(i)| \quad (7)$$

$L_{\text{SSIM}}$  表示图像之间的结构相似度 (structure similarity), 包含亮度 (luminance)、对比度 (contrast) 以及结构 (structure) 3 个指标上的差异。假设  $I$  为第 1 张图像数据,  $\hat{I}$  为第 2 张图像数据, 首先计算图像的均值  $\mu_I$  和  $\mu_{\hat{I}}$ , 方差  $\sigma_I$  和  $\sigma_{\hat{I}}$  以及图像之间的相关性  $\sigma_{\hat{I}I}$ 。进而, 用以下公式计算亮度相似度、对比度相似度以及结构相似度:

$$\begin{aligned} C_{\text{luminance}}(I, \hat{I}) &= \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1} \\ C_{\text{contrast}}(I, \hat{I}) &= \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2} \\ C_{\text{structure}}(I, \hat{I}) &= \frac{\sigma_{\hat{I}I} + C_3}{\sigma_I\sigma_{\hat{I}} + C_3} \end{aligned} \quad (8)$$

最终, 结合 3 种相似度, 整合得到了图像之间的 SSIM (结构相似度) 指标:

$$L_{\text{SSIM}} = C_{\text{luminance}}^\alpha(I, \hat{I}) C_{\text{contrast}}^\beta(I, \hat{I}) C_{\text{structure}}^\gamma(I, \hat{I}) \quad (9)$$

一般来说,  $\alpha, \beta, \gamma$  均取值为 1, 最终得到常用的衡量图像之间相似度的指标:

$$L_{\text{SSIM}} = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(\sigma_{\hat{I}I} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \quad (10)$$

### 1.1.6 自适应密度控制

计算完图像损失之后, 梯度会沿着数据流反向传播, 对 3D 高斯基元的参数进行优化, 而重建不充分的区域往往有着更大的梯度, 是需要优化的重点区域, 这里将重建不充分的情况划分为 2 种:

第 1 种情况: 欠重建 (under reconstruction), 指的是现有 3D 高斯基元不足以覆盖需要重建的区域, 要对已有 3D 高斯基元进行克隆操作。

第 2 种情况: 过重建 (over reconstruction), 指的是现有的 3D 高斯基元虽然能够覆盖所建区域, 但是不够细化, 要对已有 3D 高斯基元进行分割操作。

## 1.2 3D GS 和 NeRF 的区别

辐射场是 3D 空间中光分布的表示，它捕捉光如何与环境中的表面和材料相互作用。从数学上讲，辐射场可以描述为函数，它将空间中的一个点和由球面坐标指定的方向映射到非负辐射值。辐射场可以隐式表达，例如神经网络；也可以显式表达，例如体素网格或点集。

NeRF 是一种典型的隐式辐射场技术，主要使用隐式的基于坐标的模型将空间坐标映射到像素值，核心思想是用一个多层感知机 (MLP)<sup>[11]</sup> 来表示一个连续的场景，输入为空间点的 3D 坐标和观察方向，输出是空间点的颜色和密度信息。NeRF 的渲染过程是基于体渲染 (volume rendering) 的，它将每条从相机发出的射线上的点的颜色和密度进行加权求和，得到该射线的最终颜色。NeRF 的训练过程是通过最小化渲染结果和真实图像之间的 L2 损失来优化 MLP 的参数。

而 3D GS 是一种典型的显式辐射场技术，通过数以万计的 3D 高斯基元这样的显式场景表示，填充整个需要重建的场景，再通过并行的分块光栅化算法，实现高效的图像渲染效果。3D GS 技术的训练过程是通过最小化图像之间 L1 损失和 SSIM 损失来优化 3D 高斯基元的参数。比起 NeRF，3D GS 技术不仅可以实现场景的实时渲染，还提高了场景的可编辑性。

表 1 展示了 3D GS 和 NeRF 技术的多项指标对比。可以看出 3D GS 技术在训练和渲染时间上很有优势，在视图合成和 3D 重建方面已经完全可以替代 NeRF 技术。

表 1 3D GS 和 NeRF 技术细节对比  
Tab.1 Comparison of 3D GS and NeRF

对比项目	NeRF <sup>[2]</sup>	3D GS <sup>[3]</sup>
表达方式	隐式辐射场	显式辐射场
场景表征	神经网络	3D 高斯基元
渲染方法	体渲染	块渲染
渲染速度	慢	快
训练时间	长	短
新视图合成	支持	支持
3D 重建	支持	支持
3D 编辑	不支持	支持

## 2 基于 3D GS 的视觉 SLAM (Visual SLAM based on 3D GS)

视觉 SLAM 是机器人利用视觉传感器 (单目相机、双目相机、RGB-D 相机) 在未知环境中进行

同步定位和地图构建的一种技术，是机器人在未知环境中进行探索、导航和交互的必要技术<sup>[12]</sup>。视觉 SLAM 技术可以感知场景中丰富的纹理、结构和语义信息，凭借其鲁棒的定位能力和准确的建图效果，在室内服务机器人上应用广泛。视觉 SLAM 的核心在于创建环境地图，因此，完整准确地表达环境、构建高精度的稠密地图是视觉 SLAM 的关键问题。

传统视觉 SLAM 技术根据使用图像信息的不同，可以分为直接法和间接法。直接法指的是利用图像上每个像素的信息来估计相机位姿，基本原理是光度一致性，通过最小化光度误差来实现跟踪<sup>[13-16]</sup>。直接法在纹理较差的部分比间接法更鲁棒，但当场景中的光照变化后，直接法容易失效。间接法常用于稀疏视觉 SLAM 算法，只使用图像中显著的特征进行位姿估计，首先检测角点特征并计算包含方向和尺度信息的特征描述向量，然后通过特征匹配和最小化帧间重投影误差进行跟踪<sup>[17-20]</sup>。其中，最经典的间接法算法是 ORB-SLAM 系列算法，在 FAST (feature from accelerated segment test) 特征点<sup>[21]</sup>的基础上，提出了 ORB (oriented FAST and rotated BRIEF) 特征<sup>[22]</sup>，包含跟踪、局部建图和闭环检测 3 个线程。虽然传统的视觉 SLAM 算法取得了不错的效果，但是这类传统算法通常把定位放在首位，建图只是辅助定位的一种手段，现有的不管是稀疏级地图还是稠密级地图，都没有对环境有一个完整良好的表达。

随着 NeRF 的提出，场景的 3D 重建实现了非常高的精度和逼真的视觉效果<sup>[23]</sup>。因此，许多研究人员尝试将 NeRF 方法应用到 SLAM 算法中。基于 NeRF 的 SLAM 算法可以分为 3 种主要类型：基于 MLP 的方法、混合表示方法和显式方法。基于 MLP 的 iMAP<sup>[24]</sup> 算法提供了可扩展且内存负担小的地图表示法，但在较大场景中面临灾难性遗忘的挑战。混合表示方法结合了隐式 MLP 和结构特征的优点，显著增强了场景的可扩展性和精度。例如，NICE-SLAM 算法<sup>[25]</sup> 将 MLP 与多分辨率体素网格集成，从而实现大场景的重建。Vox-Fusion 算法<sup>[26]</sup> 采用八叉树扩展法实现地图可扩展性，而 ES-LAM<sup>[27]</sup> 和 Point-SLAM 算法<sup>[28]</sup> 分别利用三平面和神经点云来提高建图能力。至于文 [29] 中提出的显式方法，它将地图特征直接存储在体素中，无需任何 MLP，从而实现更快的优化。虽然上述基于 NeRF 的 SLAM 算法达到了不错的建图效果，但是使用神经隐式特征来表示地图，渲染速度慢，占用

资源多, 达不到实时渲染的效果。

3D GS 技术的出现解决了 NeRF 技术训练和渲染速度慢的缺点, 它使用 3D 高斯基元作为实时神经渲染的基元, 可在不牺牲图像质量的情况下显著提高渲染速度。因此, 很多研究人员转向基于 3D GS 的视觉 SLAM 算法研究。

**直接结合。**SplaTAM<sup>[30]</sup>、GS-SLAM<sup>[31]</sup>、Gaussian-SLAM<sup>[32]</sup>、Gaussian Splatting SLAM<sup>[33]</sup> 等算法是成功地将 3D GS 与 SLAM 结合起来的先驱作品。

这些方法实现了快速的渲染和高保真的重建, 但是实现细节有所不同。SplaTAM 算法提出了一种新的图像框架, 允许通过渲染轮廓来有效地区分哪些部分是新添加的场景内容, 并通过增加更多的 3D 高斯基元来提高地图容量。GS-SLAM 和 Gaussian-SLAM 算法都针对新探索的区域提出了高斯基元改变策略, 包括新高斯基元的增加和旧高斯基元的删除, 以便有效地重建新观察到的场景几何形状并改进先前观察到的区域的地图。Gaussian Splatting SLAM 算法通过直接优化 3D 高斯基元来实现相机跟踪, 并引入了几何验证和正则化方法来处理增量式 3D 密集重建中的歧义。这些先驱作品在一定程度上解决了 SLAM 在建模新区域时 3D 高斯基元的扩展问题以及位姿的优化收敛问题。

**紧凑表达。**这些直接结合的工作虽然实现了一个完整的基于 3D GS 的视觉 SLAM 系统, 但是 3D GS 技术原始的地图表征比较冗余, 导致算法的实时性并不好, 因此, 部分研究者对现有的基于 3D GS 的视觉 SLAM 进行了改进, 使得高斯基元的表达更加紧凑<sup>[34-36]</sup>。例如, 文 [34] 提出了一种紧凑的基于 3D GS 的 SLAM 系统, 利用基于滑动窗口的掩蔽策略来减少冗余椭球, 并使用一种新颖的几何密码本来压缩 3D 高斯几何属性。这些紧凑高斯方案减少了高斯椭球的数量, 提高了系统的效率和实时性。

**语义信息。**还有部分研究者在基于 3D GS 的视觉 SLAM 系统中额外引入语义信息, 对高斯特征进行语义编码, 以实现精确的语义场景表示<sup>[37-39]</sup>。例如, 文 [38] 提出了 SemGauss-SLAM 算法, 将语义特征嵌入到 3D 高斯基元表示中, 有效地在环境的空间布局内编码语义信息, 提出了用于更新 3D 高斯基元表示的特征级损失函数, 从而为 3D 高斯基元优化提供更高级别的指导。此外, 该方法还利用语义关联来联合优化 3D 高斯基元表示和相机姿势, 从而实现更稳健的跟踪和一致的建图。为了减轻引

入语义信息带来的额外内存消耗, 文 [39] 采用轻量级编码器-解码器将高维语义特征压缩为紧凑的 3D 高斯基元表示, 从而减轻过多内存消耗的负担。语义信息的引入, 让基于 3D GS 的视觉 SLAM 算法可以实现语义建图、语义定位以及物体估计等功能。

表 2 展示了基于 NeRF 和 3D GS 的视觉 SLAM 算法在 Replica 数据集<sup>[40]</sup>上的视图合成精度对比。表中 NICE-SLAM<sup>[25]</sup>、Vox-Fusion<sup>[26]</sup>、ESLAM<sup>[27]</sup> 以及 Point-SLAM<sup>[28]</sup> 算法的数据来自于文 [32] 的对比实验, GS-SLAM<sup>[31]</sup>、SplaTAM<sup>[30]</sup>、Gaussian Splatting SLAM<sup>[33]</sup> 以及 Gaussian-SLAM<sup>[32]</sup> 算法的数据均来自于其原文数据。

Replica 数据集是 SLAM 算法中常用的数据集, 该数据集具有 18 个照片级真实感 3D 室内场景, 存在密集网格、HDR 纹理、语义数据和反射表面。它跨越不同的场景类别, 包括 88 个语义类, 并结合了单个空间的 6 次扫描, 捕捉不同的家具布置和时间快照。重建中要使用定制的 RGB-D 捕捉设备, 该设备具有同步 IMU (惯性测量单元)、RGB 和广角灰度传感器, 通过 6 自由度位姿准确融合原始深度数据。用于 SLAM 评估的数据集部分是由重建过程中产生的精确网格综合生成的。

在指标方面, PSNR (峰值信噪比)、SSIM 以及 LPIPS (学习感知图像块相似度) 是评价新视图合成质量常用的 3 个指标。

PSNR 是一种用于衡量图像质量的指标, 是信号最大功率和噪声功率的比值。PSNR 先计算原始图像和失真图像之间的均方误差 (MSE), 然后将其转换为对数尺度。PSNR 的单位是分贝 (dB), 值越大表示图像质量越好。

SSIM 是一种全参考图像质量评估方法, 即它需要一个无损的原始图像作为参考, 然后对失真图像进行评估。SSIM 的值在  $-1 \sim 1$  之间, 值越大表示图像质量越好。

LPIPS 是一种基于学习的感知图像补丁相似性指标, 用于评估图像的感知质量。LPIPS 的设计灵感来自于人眼对图像的感知, 它通过学习一个神经网络模型来近似人类感知的视觉相似性。该模型使用卷积神经网络 (CNN) 对图像的局部补丁进行特征提取, 并计算补丁之间的相似性得分。

黄色高亮部分是当前基于 NeRF 的视觉 SLAM 算法中视图合成精度最高的, 绿色高亮部分是当前基于 3D GS 的视觉 SLAM 算法中视图合成精度最高的。可以看出, 基于 3D GS 的视觉 SLAM 方法

表2 基于 NeRF 和 3D GS 的视觉 SLAM 算法在 Replica 数据集上的视图合成精度对比  
Tab.2 Accuracy of view synthesis of visual SLAM algorithms based on NeRF and 3D GS on the Replica dataset

方法	类型	指标	序列名称								
			R0	R1	R2	Of0	Of1	Of2	Of3	Of4	Avg
NICE-SLAM <sup>[25]</sup>	NeRF	PSNR↑	22.12	22.47	24.52	29.07	30.34	19.66	22.23	24.94	24.42
		SSIM↑	0.689	0.757	0.814	0.874	0.886	0.797	0.801	0.856	0.809
		LPIPS↓	0.330	0.271	0.208	0.229	0.181	0.235	0.209	0.198	0.233
Vox-Fusion <sup>[26]</sup>	NeRF	PSNR↑	22.39	22.36	23.92	27.79	29.83	20.33	23.47	25.21	24.41
		SSIM↑	0.683	0.751	0.798	0.857	0.876	0.794	0.803	0.847	0.801
		LPIPS↓	0.303	0.269	0.234	0.241	0.184	0.243	0.213	0.199	0.236
ESLAM <sup>[27]</sup>	NeRF	PSNR↑	25.25	27.39	28.09	30.33	27.04	27.99	29.27	29.15	28.06
		SSIM↑	0.874	0.89	0.935	0.934	0.910	0.942	0.953	0.948	0.923
		LPIPS↓	0.315	0.296	0.245	0.213	0.254	0.238	0.186	0.210	0.245
Point-SLAM <sup>[28]</sup>	NeRF	PSNR↑	32.40	34.08	35.50	38.26	39.16	33.99	33.48	33.49	35.17
		SSIM↑	0.974	0.977	0.982	0.983	0.986	0.960	0.960	0.979	0.975
		LPIPS↓	0.113	0.116	0.111	0.100	0.118	0.156	0.132	0.142	0.124
GS-SLAM <sup>[31]</sup>	3D GS	PSNR↑	31.56	32.86	32.59	38.70	41.17	32.36	32.03	32.92	34.27
		SSIM↑	0.968	0.973	0.971	0.986	0.993	0.978	0.970	0.968	0.975
		LPIPS↓	0.094	0.075	0.093	0.050	0.033	0.094	0.110	0.112	0.082
SplaTAM <sup>[30]</sup>	3D GS	PSNR↑	32.86	33.89	35.25	38.26	39.17	31.97	29.70	31.81	34.11
		SSIM↑	0.98	0.97	0.98	0.98	0.97	0.97	0.95	0.95	0.97
		LPIPS↓	0.07	0.10	0.08	0.09	0.09	0.10	0.12	0.15	0.10
Gaussian Splatting SLAM <sup>[33]</sup>	3D GS	PSNR↑	34.83	36.43	37.49	39.95	42.09	36.24	36.7	36.07	37.5
		SSIM↑	0.954	0.959	0.965	0.971	0.977	0.964	0.963	0.957	0.96
		LPIPS↓	0.068	0.076	0.075	0.072	0.055	0.078	0.065	0.099	0.070
Gaussian-SLAM <sup>[32]</sup>	3D GS	PSNR↑	38.88	41.80	42.44	46.40	45.29	40.10	39.06	42.65	42.08
		SSIM↑	0.993	0.996	0.996	0.998	0.997	0.997	0.997	0.997	0.996
		LPIPS↓	0.017	0.018	0.019	0.015	0.016	0.020	0.020	0.020	0.018

在指标上有所提升。另外在不牺牲视图合成和 3D 重建质量的前提下, 3D GS 技术有效地提高了场景的渲染和训练速度, 同时提升了 SLAM 系统的运行实时性。

### 3 基于 3D GS 的动态场景重建 (Dynamic scene reconstruction based on 3D GS)

动态场景重建是一种利用多个视频帧数据重建出场景 3D 动态模型的技术, 也是支撑机器人众多应用的关键技术。动态场景是机器人进行导航规划和交互的主要挑战, 动态场景的感知和建模对于机器人是至关重要的, 有助于提高机器人对周围环境的理解和反应能力。

动态场景重建和一般的静态场景重建的区别在于动态场景中动态物体在图像中占据较大的画面且其运动变化复杂, 而且运动和变形在时序上难

以捕捉。因此, 动态场景重建技术的核心在于对动态场景中的运动和变形规律进行建模。NeRF 技术凭借其隐式辐射场高质量的渲染能力, 在一定程度上推动了动态场景重建技术的发展, 也出现了一系列优秀的基于 NeRF 的动态场景重建方法: D-NeRF<sup>[41]</sup>、HyperNeRF<sup>[42]</sup>、Nerfies<sup>[43]</sup>、TiNeuVox-B<sup>[44]</sup>、FDNeRF<sup>[45]</sup>、K-Planes<sup>[46]</sup>、HyperReel<sup>[47]</sup>、NeRF-DS<sup>[48]</sup>、HexPlane<sup>[49]</sup>、NeRFPlayer<sup>[50]</sup> 等。上述基于 NeRF 的方法虽然取得了一定的进展, 但是 NeRF 技术的训练时间和渲染速度使得动态场景实时重建仍然存在困难和挑战。

3D GS 技术的出现重构了 3D 重建和新视图合成领域的底层架构, 显式表达和可微分渲染的优势使得 3D GS 技术可以在不降低重建质量的情况下实现实时渲染。但是 3D GS 技术专注于静态场景重建, 直接将其应用到动态场景时效果很差。因此,

4D GS 技术应运而生, 它将 3D GS 技术在时序上进行扩展和改进, 为 3D 高斯基元增加了适应场景变形和运动的属性。

**4D GS。**文 [51-52] 首次引入 4D GS 的概念, 着重于将动态场景中的时空元素统一起来, 但是重点解决的问题有所不同。文 [51] 为了在实现动态场景实时渲染的同时保持高训练效率和高存储效率, 利用 4D 高斯基元和 4D 神经体素显式表示动态场景, 然后应用轻量级 MLP 来预测新时间戳下的高斯变形。文 [52] 为了解决动态场景中场景结构不足和缩放变形建模这两大困难, 将时空视为一个整体, 并提出通过优化一组 4D 高斯基元来近似动态场景的底层时空 4D 体。这些 4D 高斯基元具有明确的几何和外观建模能力, 使得模型能够通过定制的渲染流程在任何所需时间合成新视图。两种方法都重构了之前 3D GS 的空间属性, 将时间属性引入到 3D GS 的架构中, 增强 3D 高斯基元的时序变化能力。

**变形场。**大部分方法着重于改进 3D 高斯基元的变形属性来适应动态场景的变化, 将变形场引入了 3D GS, 提出了可变形的 3D GS<sup>[53-56]</sup>, 但是实现细节各有不同。文 [53-54] 提出使用可调参数的 MLP 来学习 3D 高斯基元的变形场, 以适应场景中动态元素的变化。文 [55] 引入了高斯基元光流方法, 提出了一种新颖的双域变形模型来显式地建模每个高斯点的属性变形, 从频域和时序 2 个方面建模 3D 高斯基元的动态属性。文 [56] 提出了一种可变形的 3D 高斯基元, 该方法使用 3D 高斯基元重建场景, 并在具有变形场的规范空间中学习它们, 重点针对单目动态场景进行建模。这些方法通过赋予 3D 高斯基元变形属性来适应动态场景的变化。

**运动建模。**一部分工作专注于解决动态场景中的运动变化<sup>[57-60]</sup>。文 [58] 提出了动态 3D 高斯基元的概念, 允许高斯基元随时间移动和旋转, 同时强制它们具有持久的颜色、不透明度和大小, 并利用局部刚度约束对高斯基元的运动和旋转进行正则化。文 [59] 提出了动态场景运动分解的概念, 将动态场景中各点的运动分解为一小组明确的或学习的轨迹, 从而实现快速、有效的优化。这些方法的重点是对场景中的高斯基元的运动变化进行拆解和模拟。

**高效存储。**少部分方案解决了使用 3D 高斯基元进行动态场景建模时的内存消耗问题<sup>[61-62]</sup>。文 [61] 提出了一种专为动态场景量身定制的高效 3D 高斯基元表示法, 它将位置和旋转定义为时间的函数, 同时保持静态 3D 高斯基元的其他属性不变,

这种表示有效减少了动态场景重建所需要的内存空间。

还有一小部分工作通过直接控制动态场景中的元素提高了 3D GS 的可编辑性<sup>[63-64]</sup>。

表 3 展示了基于 NeRF 和 3D GS 的动态场景重建算法在 D-NeRF 数据集<sup>[41]</sup>上的视图合成质量对比。表中 D-NeRF<sup>[41]</sup>、HexPlane<sup>[49]</sup>、K-Planes<sup>[46]</sup>、TiNeuVox-B<sup>[44]</sup>、3D-GS<sup>[3]</sup> 以及 4D-GS<sup>[51]</sup> 算法的各项数据均选取自文 [51] 对比实验中, 4D\_GS<sup>[52]</sup> 和 GauFR<sup>[53]</sup> 的表中数据均来自于其论文原文数据。

表 3 基于 NeRF 和 3D GS 的动态场景重建算法在 D-NeRF 数据集上的视图合成质量对比

方法	类型	PSNR↑	SSIM↑	LPIPS↓
D-NeRF <sup>[41]</sup>	NeRF	30.50	0.95	0.07
HexPlane <sup>[49]</sup>	NeRF	31.04	0.97	0.04
K-Planes <sup>[46]</sup>	NeRF	31.61	0.97	0.06
TiNeuVox-B <sup>[44]</sup>	NeRF	32.67	0.97	0.04
3D-GS <sup>[3]</sup>	3DGS	23.19	0.93	0.08
4D-GS <sup>[51]</sup>	4DGS	34.05	0.98	0.02
4D_GS <sup>[52]</sup>	4DGS	34.09	0.98	0.02
GauFR <sup>[53]</sup>	4DGS	34.80	0.98	0.02

D-NeRF 数据集常用于动态场景重建方法的评估, 是一个 360° 观察合成物体的单目视频数据集, 视频中每个时间戳下观察物体时的相机位姿是随机生成的。另外, 数据集中每个场景都包含动态帧, 数量从 50~200 不等。

蓝色高亮部分是当前基于 NeRF 的方法中视图合成质量最高的, 紫色高亮部分是当前基于 4D GS 的方法中视图合成质量最高的。可以看出基于 4D GS 的方法在 PSNR 指标上提升较大。

## 4 基于 3D GS 的生成式 AI (Generative AI based on 3D GS)

AIGC (人工智能生成内容) 指的是使用人工智能技术来自动创建内容, 包括文本、图像、视频、音乐等多种形式<sup>[65]</sup>。随着大模型的发展, AIGC 在机器人领域的应用日益广泛, 例如增强机器人交互能力、提升机器人自主性、优化机器人操作效率等。其中比较热门的应用是文本到 3D 场景的生成模型, 用户描述一个场景, AI 则可以根据其描述生成相应的 3D 模型。

当前文本到 3D 模型的生成工作大多是基于分数蒸馏采样的方法和基于扩散模型的方法。但是这

些方法通常存在渲染过程缓慢、训练和优化时间过长的的问题，而且其生成的模型细节不足，整体性较差。3D GS 技术凭借显式表达法提升了场景的可编辑性，同时保持了实时渲染的优势，因此，基于 3D GS 的文本到 3D 模型的生成应用随之出现，为场景生成和场景编辑领域注入了新的活力。

**场景生成。**许多研究者着力于研究基于 3D GS 的场景生成<sup>[66-80]</sup>。文 [66-70] 是对传统的 SDS (score distillation sampling) 方法的改进。例如，文 [66] 提出了一种新颖的文本到 3D 场景的生成方法，通过结合 3D GS 技术和分阶段优化策略，有效生成具有精细细节和准确几何形状的 3D 模型。图 2 展示了 GSGEN 算法的场景生成效果（图片选取自 GSGEN<sup>[66]</sup> 原文）。文 [69] 专注于人体生成，提出了一种新颖的结构感知 SDS 算法，可以生成具有细粒度几何形状和逼真外观的高质量 3D 人体。

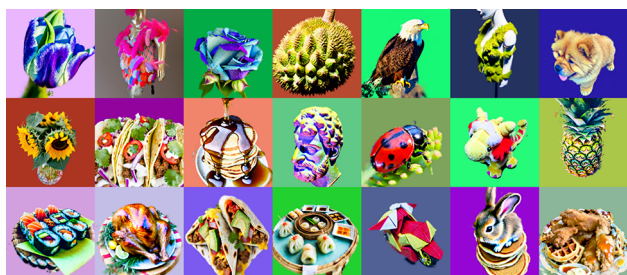


图 2 GSGEN<sup>[66]</sup> 的场景生成效果

Fig.2 Scene generation performance of GSGEN<sup>[66]</sup>

文 [71-77] 是对扩散模型的改进。例如，文 [71] 首次提出了使用 3D GS 技术结合 3D 扩散模型和 2D 扩散模型的优势，使用 3D 扩散模型为场景的初始化提供先验信息，2D 扩散模型提供丰富的几何形状和外观，快速实现文本到 3D 模型的生成。文 [78-80] 是对 4D 场景生成的探索。例如，文 [78] 提出对齐高斯方法，利用动态 3D 高斯基元和变形场作为 4D 表示，提出了一种新颖的方法来规范移动 3D 高斯基元分布，还提出了运动放大机制以及新的自回归合成方案来生成和组合多个 4D 序列，以生成更长时间序列的场景。

**场景编辑。**部分研究者专注于研究基于 3D GS 的场景编辑<sup>[81-87]</sup>。文 [81-82] 都提出了高斯基元编辑框架，但是侧重点有所不同，文 [81] 通过 3D 高斯基元和文本指令精细地编辑 3D 场景。首先提取与文本指令相对应的 RoI (region of interest)，并将其与 3D 高斯基元对齐。进一步的，利用高斯 RoI 控制编辑过程。文 [82] 通过高斯语义跟踪法来增强编辑的精度和控制力，还提出了分层高斯溅射方法，实现了更加稳定且精细的场景编辑结果。图 3

展示了该方法在不同场景下通过文本指令进行场景编辑的渲染效果（图片选取自 GaussianEditor<sup>[82]</sup> 原文）。文 [83-87] 侧重于场景的语义分割，例如，文 [83] 提出了一种新颖的 3D 交互式分割方法，将 2D 分割基础模型与 3D GS 无缝融合，将分割基础模型生成的多粒度 2D 分割结果有效地嵌入到 3D 高斯基元点特征中。这些方法提高了场景编辑的效率和精度，使得编辑后的场景一致性更好。



图 3 GaussianEditor<sup>[82]</sup> 场景编辑效果

Fig.3 Scene editing performance of GaussianEditor<sup>[82]</sup>

## 5 基于 3D GS 的自动驾驶 (Autonomous driving based on 3D GS)

自动驾驶又称无人驾驶，是指通过计算机系统、传感器和人工智能技术，使车辆能够在没有人类驾驶员干预的情况下自动行驶。当前自动驾驶的核心功能之一是识别和感知周围的环境，包括实时准确地重建周围场景（静态背景和动态物体），这对于车辆的导航、决策、控制以及安全都是至关重要的。

NeRF 技术虽然可以实现逼真的视图合成和 3D 重建效果，但训练和渲染速度太慢，与自动驾驶所需的实时性相悖。因此，3D GS 技术的出现可以解决实时渲染的问题，但是由于自动驾驶场景非常复杂，规模也非常大，对其进行实时重建很有挑战，需要大量计算资源，因此当前非常少的研究使用 3D GS 进行自动驾驶场景的建模<sup>[88-91]</sup>。

**动静态元素统一。**文 [88] 引入了街道级的高斯基元表示，将动态城市街道表示为一组配备语义逻辑和 3D 高斯参数的点云，每个点云与前景车辆或背景相关联。为了对前景车辆的动力学进行建模，

对每个物体所关联的点云集合赋予可优化的跟踪位姿以及动态的球谐函数, 实现了动态物体和静态背景的准确分割和合成效果。文 [89] 针对存在移动物体的复杂场景, 使用增量静态 3D 高斯基元对整个场景的静态背景进行建模。然后, 利用复合动态高斯图来处理多个移动物体, 单独重建每个物体并恢复它们在场景中的准确位置和遮挡关系, 实现了动态物体和静态背景的统一。

**高斯属性优化。**文 [90] 设计了一种新型紧耦合激光雷达—相机高斯溅射方法, 用 LiDAR 数据生成彩色 3D 点云和分层八叉树特征来丰富 3D 高斯基元的属性, 提供更完整的 3D 形状和颜色信息。文 [91] 提出了一种名为混合高斯的优化方法, 消除了对 SfM 点初始化的依赖, 可渲染城市场景, 此外, 还引入高斯基元方向编码作为渲染管道中球谐函数的替代方案, 从而实现依赖于视图的颜色表示。

## 6 未来发展趋势 (Future trends)

### 6.1 受限视角下的 3D GS 研究

当前 3D GS 在训练前需要高质量的场景初始 3D 点云先验信息, 没有这些信息会出现训练时间变长和效果变差的问题。但是在实际应用中, 视角受限的场景是很常见的, 无法为 3D GS 提供多视图先验 3D 点云。因此, 研究这种视角受限甚至是单视图条件下的 3D GS 技术是必要的。

### 6.2 内存紧张条件下的 3D GS 研究

3D GS 采用数以万计的 3D 高斯基元填充场景, 是显式的表征, 比起 NeRF 使用的深度神经网络, 参数量比较大并且有很多参数冗余, 从而导致内存的占用量也比较大。因此, 在内存紧张的环境下优化高斯基元的表征和渲染的密度, 在保证精度的同时紧凑地表达场景是未来的一个研究方向。

### 6.3 3D GS 和语义分割结合

3D GS 的可编辑性为其进行场景编辑提供了可能。精准地操控编辑区域、实现比较好的场景一致性效果是机器人对场景进行编辑所需要的。语义分割是机器人感知模块中的一项重要技术, 它的目标是将图像分割成若干区域, 并为每个区域赋予一个类别标签。因此, 使用语义分割作为先验信息, 融合 3D GS, 可以为机器人场景语义建图和编辑提供助力。

### 6.4 3D GS 和机械臂抓取结合

机械臂抓取任务是机器人应用中的一种动态任务, 需要在抓取过程中不断动态获取局部的场景信息。因此, 如何快速更新局部场景表达, 从而执行

准确的操作是机械臂抓取中重要的一环。而 3D GS 凭借其密度控制和协方差优化的性质, 对场景重建有很强的适应能力, 通过 3D 高斯基元的扩展和优化, 可以快速有效地对场景进行重新渲染。

## 7 结论 (Conclusion)

3D GS 技术的出现, 颠覆了 NeRF 技术在新视图合成和 3D 重建方面的底层架构, 3D GS 技术凭借其显式辐射场的场景表征以及分块并行的高效光栅化, 可以在不降低重建质量的前提下实现场景的实时渲染。此外, 3D GS 技术的训练时间非常短, 3D 高斯基元表征也提高了场景的可编辑性。然而, 3D GS 技术本身在视角受限和内存消耗方面还有待改进, 并且当前基于 3D GS 的应用中, 一部分展示意义大于实际价值, 一部分的扩展性较差, 距离实际应用落地还有一段距离, 仍然存在很大挑战, 需要后续研究者们进一步解决。另外, 机器人与人的协作共融是机器人领域的热点方向, 旨在使得机器人理解和预测人类行为, 从而更好地与人协作, 这对机器人场景感知和场景建模的精度和实时性要求比较高, 而 3D GS 技术兼顾了精度和实时性, 因此, 3D GS 技术在机器人协作共融方面具有发掘潜力。通过本文的梳理和展望, 希望可以为研究者们进一步改进 3D GS 技术, 为 3D GS 在机器人领域的应用和创新提供帮助。

## 参考文献 (References)

- [1] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 4104-4113.
- [2] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis[C]//European Conference on Computer Vision. Cham, Switzerland: Springer, 2020: 405-421.
- [3] KERBL B, KOPANAS G, LEIMKÜHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 1-14.
- [4] CHEN G K, WANG W G. A survey on 3D Gaussian splatting [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2401.03890>.
- [5] FISHER A, CANNIZZARO R, COCHRANE M, et al. ColMap: A memory-efficient occupancy grid mapping framework[J]. Robotics and Autonomous Systems, 2021, 142. DOI: 10.1016/j.robot.2021.103755.
- [6] MÜLLER C. Spherical harmonics[M]. Berlin, Germany: Springer, 2006.
- [7] ALIEV K-A, SEVASTOPOLSKY A, KOLOS M, et al. Neural point-based graphics[C]//16th European Conference on Computer Vision. Cham, Switzerland: Springer, 2020: 696-712.
- [8] RÜCKERT D, FRANKE L, STAMMINGER M. ADOP: Approximate differentiable one-pixel point rendering[J]. ACM Transactions on Graphics, 2022, 41(4): 1-14.

- [9] LAINE S, KARRAS T. High-performance software rasterization on GPUs[C]//ACM SIGGRAPH Symposium on High Performance Graphics. New York, USA: ACM, 2011: 79-88.
- [10] SCHÜTZ M, KERBL B, WIMMER M. Software rasterization of 2 billion points in real time[J]. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2022, 5(3): 1-17.
- [11] TAUD H, MAS J F. Multilayer perceptron (MLP)[M]// *Geomatic Approaches for Modeling Land Change Scenarios*. Cham, Switzerland: Springer, 2018: 451-455.
- [12] KAZEROUNI I A, FITZGERALD L, DOOLY G, et al. A survey of state-of-the-art on visual SLAM[J]. *Expert Systems with Applications*, 2022, 205. DOI: 10.1016/j.eswa.2022.117734.
- [13] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time[C]//International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2320-2327.
- [14] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM[C]//European Conference on Computer Vision. Cham, Switzerland: Springer, 2014: 834-849.
- [15] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(3): 611-625.
- [16] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 15-22.
- [17] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2007: 225-234.
- [18] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163.
- [19] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255-1262.
- [20] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. *IEEE Transactions on Robotics*, 2021, 37(6): 1874-1890.
- [21] VISWANATHAN D G. Features from accelerated segment test (FAST)[DB/OL]. [2024-04-01]. <https://api.semanticscholar.org/CorpusID:17031649>.
- [22] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2564-2571.
- [23] GAO K, GAO Y, HE H, et al. NeRF: Neural radiance field in 3D vision, a comprehensive review[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2210.00379>.
- [24] SUCAR E, LIU S, ORTIZ J, et al. iMAP: Implicit mapping and positioning in real-time[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 6229-6238.
- [25] ZHU Z, PENG S, LARSSON V, et al. NICE-SLAM: Neural implicit scalable encoding for SLAM[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: 2022: 12786-12796.
- [26] YANG X R, LI H, ZHAI H J, et al. Vox-Fusion: Dense tracking and mapping with voxel-based neural implicit representation [C]//IEEE International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2022: 499-507.
- [27] JOHARI M M, CARTA C, FLEURET F. ESLAM: Efficient dense SLAM system based on hybrid representation of signed distance fields[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 17408-17419.
- [28] SANDSTRÖM E, LI Y, VAN GOOL L, et al. Point-SLAM: Dense neural point cloud-based SLAM[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2023: 18387-18398.
- [29] TEIGEN A L, PARK Y, STAHL A, et al. RGB-D mapping and tracking in a plenoxel radiance field[C]//IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2024: 3330-3339.
- [30] KEETHA N, KARHADE J, JATAVALLABHULA K M, et al. SplatAM: Splat, track & map 3D Gaussians for dense RGB-D SLAM[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.02126>.
- [31] YAN C, QU D L, XU D, et al. GS-SLAM: Dense visual SLAM with 3D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.11700>.
- [32] YUGAY V, LI Y, GEVERS T, et al. Gaussian-SLAM: Photorealistic dense SLAM with Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.10070>.
- [33] MATSUKI H, MURAI R, KELLY P H J, et al. Gaussian splatting SLAM[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.06741>.
- [34] DENG T C, CHEN Y H, ZHANG L Y, et al. Compact 3D Gaussian splatting for dense visual SLAM[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.11247>.
- [35] SUN S, MIELLE M, LILIENTHAL A J, et al. High-fidelity S-LAM using Gaussian splatting with rendering-guided densification and regularized optimization[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.12535>.
- [36] HU J R, CHEN X H, FENG B Y, et al. CG-SLAM: Efficient dense RGB-D SLAM in a consistent uncertainty-aware 3D Gaussian field[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.16095>.
- [37] LI M R, LIU S H, ZHOU H. SGS-SLAM: Semantic Gaussian splatting for neural dense SLAM[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2402.03246>.
- [38] ZHU S T, QIN R J, WANG G M, et al. SemGauss-SLAM: Dense semantic Gaussian splatting SLAM[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.07494>.
- [39] JI Y M, LIU Y, XIE G, et al. NEDS-SLAM: A novel neural explicit dense semantic SLAM framework using 3D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.11679>.
- [40] STRAUB J, WHELAN T, MA L, et al. The Replica dataset: A digital replica of indoor spaces[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/1906.05797>.

- [41] PUMAROLA A, CORONA E, PONS-MOLL G, et al. D-NeRF: Neural radiance fields for dynamic scenes[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021: 10313-10322.
- [42] PARK K, SINHA U, HEDMAN P, et al. HyperNeRF: A higher-dimensional representation for topologically varying neural radiance fields[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2106.13228>.
- [43] PARK K, SINHA U, BARRON J T, et al. Nerfies: Deformable neural radiance fields[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 5845-5854.
- [44] FANG J M, YI T R, WANG X G, et al. Fast dynamic radiance fields with time-aware neural voxels[C]//SIGGRAPH Asia 2022 Conference Papers. New York, USA: ACM, 2022. DOI: 10.1145/3550469.3555383.
- [45] ZHANG J B, LI X Y, WAN Z Y, et al. FNeRF: Few-shot dynamic neural radiance fields for face reconstruction and expression editing[C]//SIGGRAPH Asia 2022 Conference Papers. New York, USA: ACM, 2022. DOI: 10.1145/3550469.3555404.
- [46] FRIDOVICH-KEIL S, MEANTI G, WARBURG F R, et al. K-planes: Explicit radiance fields in space, time, and appearance [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 12479-12488.
- [47] ATTAL B, HUANG J B, RICHARDT C, et al. HyperReel: High-fidelity 6-DoF video with ray-conditioned sampling [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 16610-16620.
- [48] YAN Z, LI C, LEE G H. NeRF-DS: Neural radiance fields for dynamic specular objects[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 8285-8295.
- [49] CAO A, JOHNSON J. HexPlane: A fast representation for dynamic scenes[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 130-141.
- [50] SONG L C, CHEN A P, LI Z, et al. NeRFplayer: A streamable dynamic scene representation with decomposed neural radiance fields[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(5): 2732-2742.
- [51] WU G J, YI T R, FANG J M, et al. 4D Gaussian splatting for real-time dynamic scene rendering[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2310.08528>.
- [52] YANG Z Y, YANG H Y, PAN Z J, et al. Real-time photo-realistic dynamic scene representation and rendering with 4D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2310.10642>.
- [53] LIANG Y Q, KHAN N, LI Z Q, et al. GauFR: Gaussian deformation fields for real-time dynamic novel view synthesis [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.11458>.
- [54] SHAW R, SONG J, MOREAU A, et al. SWAGS: Sampling windows adaptively for dynamic 3D Gaussian splatting [DB/OL]. (2023-12-20) [2024-04-01]. <http://export.arxiv.org/abs/2312.13308v1>.
- [55] LIN Y T, DAI Z Z, ZHU S Y, et al. Gaussian-Flow: 4D reconstruction with dynamic 3D Gaussian particle[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.03431>.
- [56] YANG Z Y, GAO X Y, ZHOU W, et al. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2309.13101>.
- [57] DAS D, WEWER C, YUNUS R, et al. Neural parametric Gaussians for monocular non-rigid object reconstruction[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.01196>.
- [58] LUITEN J, KOPANAS G, LEIBE B, et al. Dynamic 3D Gaussians: Tracking by persistent dynamic view synthesis[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2308.09713>.
- [59] KRATIMENOS A, LEI J, DANILIDIS K. DynMF: Neural motion factorization for real-time dynamic view synthesis with 3D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.00112>.
- [60] HUANG Y H, SUN Y T, YANG Z, et al. SC-GS: Sparse-controlled Gaussian splatting for editable dynamic scenes [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.14937>.
- [61] KATSUMATA K, VO D M, NAKAYAMA H. An efficient 3D Gaussian representation for monocular/multi-view dynamic scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.12897v1>.
- [62] LI Z, CHEN Z, LI Z, et al. Spacetime Gaussian feature splatting for real-time dynamic view synthesis[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.16812>.
- [63] YU H, JULIN J, MILACSKI Z Á, et al. CoGS: Controllable Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.05664>.
- [64] SHAO R Z, SUN J X, PENG C, et al. Control4D: Dynamic portrait editing by learning 4D GAN from 2D diffusion-based editor[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2305.20082v1>.
- [65] WU J Y, GAN W S, CHEN Z F, et al. AI-generated content (AIGC): A survey[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2304.06632>.
- [66] CHEN Z L, WANG F, LIU H P. Text-to-3D using Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2309.16585v1>.
- [67] TANG J X, REN J W, ZHOU H, et al. DreamGaussian: Generative Gaussian splatting for efficient 3D content creation [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2309.16653>.
- [68] LIANG Y X, YANG X, LIN J T, et al. LucidDreamer: Towards high-fidelity text-to-3D generation via interval score matching [DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.11284>.
- [69] LIU X, ZHAN X H, TANG J X, et al. HumanGaussian: Text-driven 3D human generation with Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.17061>.
- [70] ZHANG J W, TANG Z Y, PANG Y T, et al. Repaint123: Fast and high-quality one image to 3D generation with progressive controllable 2D repainting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.13271>.
- [71] YI T R, FANG J M, WU G J, et al. GaussianDreamer: Fast generation from text to 3D Gaussian splatting with point cloud priors[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2310.08529v1>.
- [72] LI X H, WANG H B, TSENG K-K. GaussianDiffusion: 3D Gaussian splatting for denoising diffusion probabilistic models with structured noise[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.11221>.
- [73] CHUNG J, LEE S, NAM H, et al. LucidDreamer: Domain-free generation of 3D Gaussian splatting scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.13384>.
- [74] VILESOV A, CHARI P, KADAMBI A. CG3D: Compositional generation for text-to-3D via Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.17907>.

- [75] YANG X F, CHEN Y W, CHEN C, et al. Learn to optimize denoising scores for 3D generation: A unified and improved diffusion prior on NeRF and 3D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.04820>.
- [76] ZOU Z X, YU Z P, GUO Y C, et al. Triplane meets Gaussian splatting: Fast and generalizable single-view 3D reconstruction with transformers[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.09147>.
- [77] OUYANG H, HEAL K, LOMBARDI S, et al. Text2Immersion: Generative immersive scene with 3D Gaussians[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.09242>.
- [78] LING H, KIM S W, TORRALBA A, et al. Align your Gaussians: Text-to-4D with dynamic 3D Gaussians and composed diffusion models[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.13763>.
- [79] REN J W, PAN L, TANG J X, et al. DreamGaussian4D: Generative 4D Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.17142>.
- [80] YIN Y Y, XU D J, WANG Z Y, et al. 4DGen: Grounded 4D content generation with spatial-temporal consistency[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.17225>.
- [81] FANG J M, WANG J J, ZHANG X P, et al. GaussianEditor: Editing 3D Gaussians delicately with text instructions[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.16037v1>.
- [82] CHEN Y W, CHEN Z L, ZHANG C, et al. GaussianEditor: Swift and controllable 3D editing with Gaussian splatting[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.14521>.
- [83] YE M Q, DANELLJAN M, YU F, et al. Gaussian grouping: Segment and edit anything in 3D scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.00732>.
- [84] CEN J Z, FANG J M, YANG C, et al. Segment any 3D Gaussians[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.00860>.
- [85] LAN K, LI H R, SHI H L, et al. 2D-guided 3D Gaussian segmentation[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.16047>.
- [86] HUANG J J, YU H C. Point'n Move: Interactive scene object manipulation on Gaussian splatting radiance fields[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2311.16737>.
- [87] ZHOU S J, CHANG H R, JIANG S C, et al. Feature 3DGS: Supercharging 3D Gaussian splatting to enable distilled feature fields[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.03203>.
- [88] YAN Y Z, LIN H T, ZHOU C X, et al. Street Gaussians for modeling dynamic urban scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2401.01339v1>.
- [89] ZHOU X Y, LIN Z W, SHAN X J, et al. DrivingGaussian: Composite Gaussian splatting for surrounding dynamic autonomous driving scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2312.07920>.
- [90] ZHAO C, SUN S, WANG R Y, et al. TCLC-GS: Tightly coupled LiDAR-camera Gaussian splatting for surrounding autonomous driving scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2404.02410v1>.
- [91] LI Z P, ZHANG Y L, WU C M, et al. HO-Gaussian: Hybrid optimization of 3D Gaussian splatting for urban scenes[DB/OL]. [2024-04-01]. <https://arxiv.org/abs/2403.20032>.

#### 作者简介:

曹振中 (1998-), 男, 博士生。研究领域: 人机共融, 视觉 SLAM, 3 维重建, 深度学习。

刘景泰 (1964-), 男, 博士, 教授。研究领域: 智能机器人技术, 智能科学与技术, 人机共融, 网络机器人等。