基于三维高斯泼溅的动态场景重建研究综述

连振晗[1,2)]，王章野[1)]，牛炯[3)]，程乐超[4)*]

[1)] (浙江大学计算机辅助设计与图形系统全国重点实验室　杭州　310058)
[2)] (江西求是高等研究院　南昌　330038)
[3)] (浙江大学计算机科学与技术学院　杭州　310058)
[4)] (合肥工业大学计算机与信息学院(人工智能学院)　合肥　230009)
(chenglc@hfut.edu.cn)

摘　要：作为计算机图形学与视觉领域的重点研究方向，动态场景重建的核心目标是从单目或多视角的二维图像序列或视频中精确地捕获并恢复动态场景的三维几何结构及其时变外观属性。三维高斯泼溅（3D Gaussian splatting，3DGS）是一种基于三维高斯体的场景重建和渲染技术，高斯体的参数化特性使其既可以显式地表示场景元素，又可以隐式地优化局部细节，从而高效地模拟场景中的光线传播和散射，最终实现实时渲染。文中对基于 3DGS 的动态场景重建方法进行梳理，并将其分为 3 大类。首先介绍 3DGS 的基本原理和优势，并简述相关的应用；然后总结基于变形场、时空四维高斯体和帧间动态传递 3 类方法，并围绕场景表示、运动建模、训练策略等维度展开描述，通过对比分析了每类方法的优缺点；最后回顾该领域的发展现状，指出未来可能的研究方向，包括时空数据结构的设计、稀疏视角下的动态重建，以及大规模动态场景的实时重建。

关键词：动态场景重建；三维高斯泼溅；辐射场；显式表示
中图法分类号：TP391.41　　**DOI：**10.3724/SP.J.1089.2025-00251

A Review of Dynamic Scene Reconstruction Based on 3D Gaussian Splatting

Lian Zhenhan[1,2)], Wang Zhangye[1)], Niu Jiong[3)], and Cheng Lechao[4)*]

[1)] (*State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou*　310058)
[2)] (*Jiangxi Qiushi Advanced Institute, Nanchang*　330038)
[3)] (*College of Computer Science and Technology, Zhejiang University, Hangzhou*　310058)
[4)] (*School of Computer Science and Information Engineering (School of Artificial Intelligence), Hefei University of Technology, Hefei*　230009)

Abstract: Dynamic scene reconstruction, a key research direction in computer graphics and vision, aims to accurately capture and recover the three-dimensional geometric structure and time-varying appearance of dynamic scenes from monocular or multi-view two-dimensional image sequences or videos. 3D Gaussian splatting (3DGS) is a scene reconstruction and rendering technique based on three-dimensional Gaussian volumes. The parameterized nature of Gaussian volumes allows them to both explicitly represent scene elements and implicitly optimize local details, effectively simulating light propagation and scattering in the scene, ultimately enabling real-time rendering. This article reviews dynamic scene reconstruction methods based on 3DGS and categorizes them into three main categories. First, the basic principles and advantages of 3DGS are introduced, and related applications are briefly described. Second, three categories of methods are summarized: deformation fields, spatiotemporal four-dimensional Gaussian volumes, and inter-frame dynamic transfer. Various

收稿日期: 2025-07-04；修回日期：2025-09-25。**基金项目：**国家自然科学基金（62472139）。**连振晗**（2001—），男，硕士研究生，CCF 学生会员，主要研究方向为计算机图形学；**王章野**（1965—），男，博士，副教授，主要研究方向为计算机图形学、虚拟现实、机器学习；**牛炯**（1976—），男，博士，工程师，主要研究方向为机器学习，信息安全；**程乐超**（1992—），男，博士，副教授，论文通信作者，主要研究方向为视觉知识迁移。

works are described, focusing on scene representation, motion modeling, and training strategies. The advantages and disadvantages of each category are compared and analyzed. Finally, the current status of development in this field is reviewed, and potential future research directions are identified. This includes the design of spatiotemporal data structures, dynamic reconstruction under sparse viewports, and real-time reconstruction of large-scale dynamic scenes.

Key words: dynamic scene reconstruction; 3D Gaussian splatting; radiance field; explicit representation

1 研究背景

1.1 动态场景重建的背景

随着硬件条件和算力资源不断发展,以及混合现实、自动驾驶与元宇宙等概念逐渐成熟,高质量且实时渲染的画面越来越成为更多前沿领域追求的目标。目前,计算机图形学和视觉领域存在一套完善的渲染管线[1-3],可对复杂的静态场景进行细节捕捉和精确建模,并最终合成新视角下的逼真图像。但是,现实世界中的大多数应用场景涉及动态环境,其中的物体或人物会随时间变化。为了进一步提升在真实世界中建模与渲染的能力,研究人员逐渐将注意力转向动态场景的重建与合成[4-10]。动态场景重建指从多视角或时序观测数据中恢复动态三维场景的几何、外观和运动信息的过程[4-7],是计算机图形学和视觉领域的重要研究方向,其核心是同时建模场景的时空演化过程,通常涉及非刚性形变、运动分割、多目标跟踪、新视角合成等问题。与静态场景的重建相比,动态场景重建不仅需要理解不同时刻空间结构变化,还要求新视角下生成的图像符合物理一致性,因此它面临更多挑战[8-15]。这些挑战使得动态场景的重建成为当前研究的热点和难点。

动态场景重建的发展经历了从传统几何方法到深度学习驱动的演变。早期的研究对象以静态场景为主体[16-17],动态部分为稀疏点或简单刚性运动,这时提出的方法往往需要大量密集的相机位姿,并通过粒子轨迹建模稀疏动态点[18]或建立场景流[4]完成细致的重建工作。之后,研究人员引入参数化形变模型(如低维基、物理仿真)处理非刚性运动,如 Bregler 等[19]使用线性形状基表示非刚性形变。21 世纪第 2 个十年,RGB-D 传感器的普及催生了实时动态融合技术,同期研究进一步将深度传感器(如 Kinect)和多视角视觉系统相结合,以实现高精度且实时的重建。Dou 等[6]实现了多视角 RGB-D 数据实时融合,但该方法仍受限于传感器噪声和遮挡。近年来,NeRF[20]作为一种基于神经网络的三维场景建模与新视角合成方法取得了显著的成果,其核心思想是利用多层感知机(multilayer perceptron,MLP)隐式地表示一个连续的三维场景,并能够渲染出任意视角下的图像;该技术能够渲染生成高质量逼真的图像,极大地促进了新视角合成与场景重建领域的发展。之后,Kerbl 等[21]提出一种新兴的显式三维场景表示方法——三维高斯泼溅(3D Gaussian splatting,3DGS),它使用基于点云的三维高斯分布建模场景,并通过可微的光栅化实现高效渲染,在保持高质量渲染的同时,显著地提升训练速度,展现出巨大的潜力。在此基础上,Huang 等[22]提出重要变种 2DGS,并指出三维高斯体在不同视角下存在不同的视角-高斯体相交平面现象,可能导致渲染时多视角下的深度不一致情况,特别是对于薄表面的物体,为此,改用二维高斯盘对场景进行表面建模,保证重建时的几何一致性。

1.2 动态场景重建的挑战

与静态场景的重建相比,由于动态场景重建需要通过时序多模态观测数据恢复三维场景时空演化过程中的几何拓扑与表观属性[23-39],因此面临 3 个维度的核心挑战:(1)运动估计的时变不确定性,特别是高速运动导致的观测数据缺失问题[14,23,26-28];(2)非刚性形变的参数化表征难题,需要在特定复杂空间建立连续可微的形变场[7,9,29];(3)动态遮挡条件下的三维推理,要求算法具备遮挡区域的几何先验建模能力[13-15]。动态重建方法的复杂度显著高于静态场景情况:

(1)时空一致性的建模

现实中多数场景随时间不断变化,动态场景重建方法不仅需要捕捉场景每帧的三维结构,还要确保这些结构时间维度的连贯性和一致性[8-12,25-29]。要求重建方法能够有效地处理和衔接连续帧之间的数据,创建一个既精确又连贯的场景表示。

(2)处理遮挡和视角限制

与静态场景相比,动态场景中往往存在更复杂的遮挡问题。例如,某一时刻一个物体可能被另一个物体完全或部分遮挡,而下一个时刻遮挡关系可

能发生变化，造成突现或突隐现象；此外，从某些视角可能无法观察到场景的特定部分。因此，重建方法需要能够推断出被遮挡区域的外观和形状，这通常依赖于更加复杂的预测模型[13-15,30-34]或多视角合成技术。

（3）实时性能要求

对应用于增强现实或自动驾驶等实时系统的动态场景重建中，处理速度是一个关键因素，不仅要求算法要准确，而且还要足够快速，以便能够在几十毫秒内处理和更新场景信息并及时反馈给用户，实现交互的效果[5-6,21,34]。

（4）大规模和复杂度问题

通常，动态场景包括多个互动的复杂元素，如多个人物、动态的背景和变化的环境条件。重建方法需要在这种高度复杂的环境中有效工作，同时管理大量的数据输入并维护高度的计算效率[35]。这无疑造成了巨额的存储开销和复杂的数据处理工作，成为制约动态场景重建技术在实际应用中进一步发展的重大难题[36-40]。

（5）光照和反射的处理

真实世界中，光照条件可以极大地影响视觉感知，包括不同时间的光照变化和复杂的反射效果（如镜面反射、折射等）[8,36,38]。重建方法需要能够解析这些光照条件对场景外观的影响，并在模型中准确地复现。

（6）跨模态数据融合

理想的动态场景重建系统可能需要结合来自多种传感器的数据，如视频、深度传感器、激光雷达等[39-41]。要求重建方法能够整合和同步不同模态的数据，提高整体的重建质量和鲁棒性。

动态场景重建是一个前沿的研究领域，涉及图像处理、深度学习、计算机视觉、计算机图形学等多个领域的技术和方法[42]。随着技术的发展和算法的改进，这些挑战正在逐步得到解决，并推动相关技术的实际应用和发展。

本文的结构如图 1 所示。

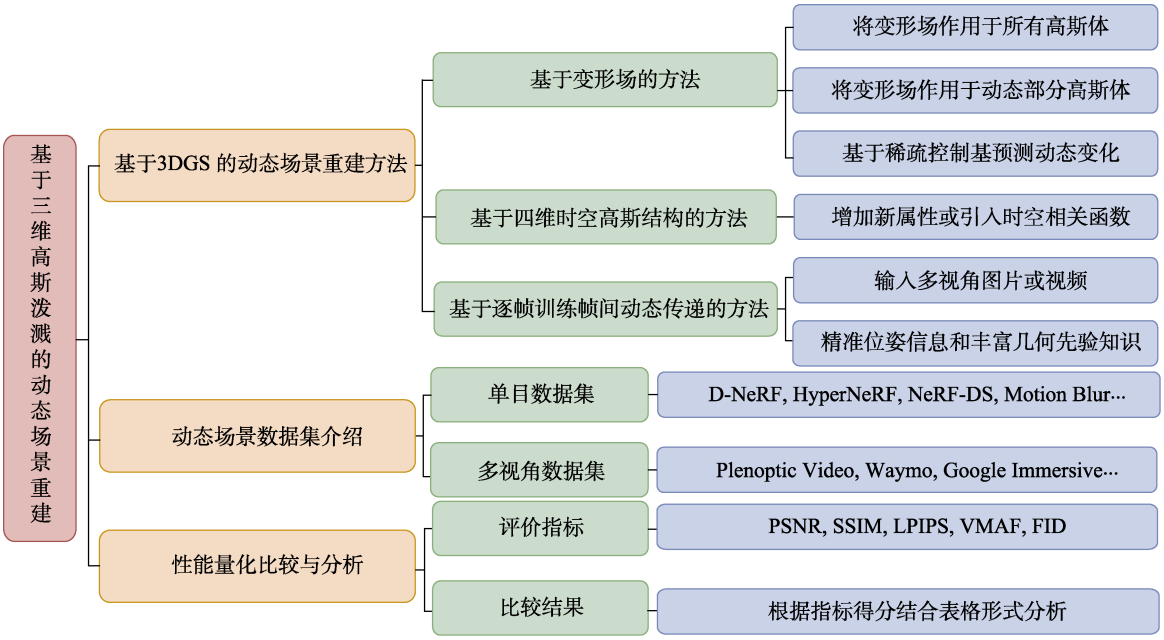


图 1　本文的结构

2　3DGS 基础理论

与传统的隐式表示方法（如 NeRF）不同，3DGS 利用三维高斯分布的特性将场景建模为一系列可学习的三维高斯体，并通过优化这些高斯体的参数捕捉场景的几何和外观信息，原始 3DGS 的工作管线如图 2 所示。本节从数学表示、基本原理及关键技术和优势与应用前景 3 个方面详细地介绍 3DGS 关键技术。

2.1　数学表示

在 3DGS 中，场景被表示为一系列的三维高斯分布，每个高斯体由以下参数定义：

（1）位置 $\boldsymbol{\mu}$：高斯体在三维空间中的位置；

（2）尺度 $\boldsymbol{\Sigma}$：高斯体的协方差矩阵，用于控制高斯体在空间中的形状和大小；

（3）不透明度 $\boldsymbol{\alpha}$：高斯体对光线的遮挡程度；

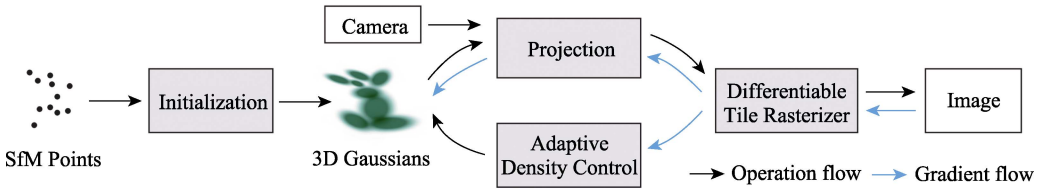


图 2 3DGS 的工作管线[21]

（4）颜色 $\boldsymbol{c}$：高斯体的辐射属性，由球谐函数（spherical harmonic，SH）表示为与视角相关的 RGB 值。

世界坐标系下，高斯体由 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 定义[21]为

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

为了保证协方差矩阵的半正定性质，采用缩放矩阵 $\boldsymbol{S}$ 和旋转矩阵 $\boldsymbol{R}$，相应的协方差矩阵表示为

$$\boldsymbol{\Sigma}=\boldsymbol{RSS}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}。$$

上述操作使得 3DGS 可以通过优化 $\boldsymbol{S}$ 和 $\boldsymbol{R}$ 代替直接优化 $\boldsymbol{\Sigma}$，保证了优化后的 $\boldsymbol{\Sigma}$ 的有效性；此外，为了方便独立优化及存储，采用三维向量 $\boldsymbol{s}$ 表示缩放，四元数 $\boldsymbol{q}$ 表示旋转。以上所有参数都是可学习的，并通过反向传播进行优化。

2.2 原理及技术

对于某一场景，3DGS 旨在高质量地生成任意给定镜头视角下的图像并保证渲染的实时性。与 NeRF 采用体积光线行进技术、通过沿视线路径密集采样三维空间点进行像素渲染的方法[20,43]不同，3DGS 使用基于泼溅的光栅化方法进行渲染[44]。在给定的相机视角下，渲染图像中每个像素的颜色 $\boldsymbol{c}$ 由其范围内的 N 个高斯体 $\left\{G_i\middle|i=1,2,\cdots,N\right\}$ 得到，即

$$\boldsymbol{c}=\sum_{i=1}^{N}\boldsymbol{c}_i\alpha_i'\prod_{j=1}^{i-1}\left(1-\alpha_j'\right)。$$

其中，$\boldsymbol{c}_i$ 表示第 i 个高斯体的颜色，由该点的球谐系数计算得到；α_i' 由第 i 个高斯点在投影平面上的二维高斯分布计算得到，即

$$\alpha_i'=\alpha_i\times\mathrm{e}^{\frac{-1}{2}(\boldsymbol{x}'-\boldsymbol{\mu}')^{\mathrm{T}}\boldsymbol{\Sigma}_i'^{-1}(\boldsymbol{x}'-\boldsymbol{\mu}')}。$$

其中，$\boldsymbol{x}'$ 和 $\boldsymbol{\mu}'$ 分别表示 $\boldsymbol{x}$ 和 $\boldsymbol{\mu}$ 在投影空间中的坐标；$\boldsymbol{\Sigma}_i'$ 表示 $\boldsymbol{\Sigma}_i$ 在投影空间上的投影，计算公式为

$$\boldsymbol{\Sigma}_i'=\boldsymbol{JW\Sigma}_i\boldsymbol{W}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}，$$

$\boldsymbol{W}$ 表示视图变换矩阵，$\boldsymbol{J}$ 表示投影变换的仿射近似的雅可比矩阵。上述过程被称为泼溅。不难看出，3DGS 的渲染过程与 NeRF 相反，避免计算密集的采样和查询过程，很大程度上加快了渲染的速度。

为提升渲染效率，3DGS 将渲染精度从像素级下降到块级，将整张图划分为多个块，每个块包含 16×16 个像素；通过判断每个块包含的高斯体，对覆盖多块的高斯体进行复制并分配唯一标识。随后将各高斯体（含副本）的投影深度与对应 ID 组合，利用基于 GPU 的基数排序进行整体排序，最终为每个块分配有序的高斯体子序列；计算时，块与块之间以及块内逐像素之间是独立的，此特性适合使用并行计算，从而极大地加快光栅化的速度；在训练过程中，3DGS 对高斯体进行分裂与克隆操作，保证空间中高斯体的密度能够与场景适配。对于协方差过大的高斯体，3DGS 将其分裂为 2 个较小的高斯体；对于协方差较小的高斯体，3DGS 对其进行复制使高斯体占据更多空间。此外，对于训练过程中透明度低于阈值的高斯体，直接将其去除。

2.3 优势与应用

3DGS 通过将三维场景表示为一系列可学习的三维高斯体分布，实现了高效的实时渲染[21]。该技术创造性地融合显式辐射场和隐式辐射场的双重优势。一方面，由于显式辐射场中每个空间元素单元，如体素网格[45]和点云[46]都存储了相应位置的辐射信息，因此可以根据不同的点坐标和观察位置直接查询光线的辐射信息，提供较快的渲染速度；但也会因一定的存储开销和空间元素结构潜在的设计缺陷产生分辨率低下的问题。另一方面，隐式辐射场通过神经网络表征一个连续的体积场景，其中最著名的就是 NeRF。NeRF 中，使用多个 MLP 将观察坐标和观察位置映射为颜色 $\boldsymbol{c}$ 和体密度 σ，通过体积光线行进技术合成新视角图像，虽然取得了逼真的渲染效果，但是不可避免地造成一定的计算负荷。

3DGS 本质上是使用一系列连续可学习的场景基本元素单元（三维高斯体）离散地表示某个场景中的光分布，也就是重建该场景的显式辐射场。不难发现，此技术借鉴并吸收了隐式辐射场的可取之处，即以连续的方式表示三维场景中的基本元素单元，避免了彻底离散化带来的精度损失，因此通过优化高斯体的参数，3DGS 能够捕捉复杂的几何细节和光照效果。与传统的基于体素或网格的方法

相比，3DGS 通过将三维几何显式地表征为参数化高斯分布，其紧凑的数学形式有效地降低了存储开销；而与 NeRF 等隐式表示方法相比，3DGS 在保持高质量的同时，利用泼溅式的光栅化方法极大地提升了渲染速度。此外，3DGS 的参数化表示形式天然适合与深度学习框架结合，通过端到端的优化，能够从稀疏的输入数据（如图像或点云）中学习高质量的场景表示，并支持多种下游任务。上述混合优势使得 3DGS 在复杂的场景中依然可以做到高质量地实时渲染并缩短训练时间。

随着 3DGS 技术的快速发展，实时三维内容生成与交互、自动驾驶、机器人与具身智能、影视与游戏产业、医学影像与生物建模等多个领域涌现出一系列基于该技术的创新型应用。3DGS 作为一种新兴的三维场景表示方法，以高质量、高效率和灵活性在多个领域展现出了巨大潜力[47-51]。

在实时三维内容生成与交互领域中，3DGS 支持从多视角图像或视频中快速地重建出高质量的三维场景，能够实现在消费级 GPU 上的实时渲染与简单交互，为虚拟展示、数字孪生等应用提供了建模选择。

在自动驾驶系统中，3DGS 高效的动态重建能力使得系统能够利用车载多目摄像头、雷达等硬件设施，实时生成周围环境的三维表示，提升障碍物识别、轨迹预测和决策规划的准确性；同时，也可以利用 3DGS 生成的高度真实的仿真驾驶场景开展训练和验证工作[52]。

在机器人与具身智能领域，3DGS 可从机器人视觉与深度传感器实时构建高保真、可编辑的三维场景，支持精细的物体级分割与场景理解，显著地提升环境感知、预测模拟、抓取规划等核心能力。

在影视与游戏产业中，3DGS 正在改变传统三维内容的生成方式，支持快速地扫描真实场景，并生成可用于特效合成的逼真数字资产，显著地缩短了电影的制作周期。在游戏引擎中，也可以集成 3DGS 渲染管线，实现各种场景的高保真实时可视化，在不牺牲帧率的前提下，提升画面细节与真实感，为玩家提供更具沉浸感的体验。

在医学影像与生物建模领域，3DGS 为器官、组织和细胞结构的三维重建提供了新的解决方案。该技术能够从医学扫描数据（如 CT、MRI）中快速地生成精细化解剖模型[53]，为医学教学和远程诊疗时的可视化需求提供技术支持。

3　基于 3DGS 的动态场景重建方法

作为近几年新兴的三维重建技术，3DGS 通过显式表示与隐式优化的协同，突破了传统辐射场在渲染效率与动态建模上的局限性；通过灵活的高斯体参数化与时变建模策略，为复杂运动场景的高保真重建提供了新范式。如图 1 所示，当前主流的基于 3DGS 的动态场景重建方法可以归纳为 3 类[51]：

（1）通过变形场捕捉和预测不同时刻高斯体的形变、位移和颜色变化；

（2）通过增加时间维度，将三维高斯扩展到四维高斯，某一特定时刻的采样结果为其对应的空间高斯体表示；

（3）在特定关键帧（如初始帧）初始化高斯体，训练其动态传递到其他帧的过程。

基于 3DGS 的动态三维场景重建中一些关键工作的时间线如图 3 所示。

3.1　基于变形场的方法

基于变形场的方法中，通过学习一个随时间或者帧号变化的变形场，解耦动态场景中时间和辐射场之间的复杂关系，最终通过变形场将任意时刻 t 的点坐标映射到规范空间。通常，这类方法训练一个 MLP 作为变形网络，预测某个规范空间中的高斯体经过 Δt 发生的位移和形变，以捕捉场景随时间的动态变化。此外，上述方法大多假设只有高斯体的坐标位置、缩放和旋转情况会随着时间变化而变化，而其他的属性和参数保持不变[7,9]。这一假设显著地简化了模型设计。

3.1.1　将变形场作用于所有高斯体

Deformable 3DGS[9]首次将变形场引入原始

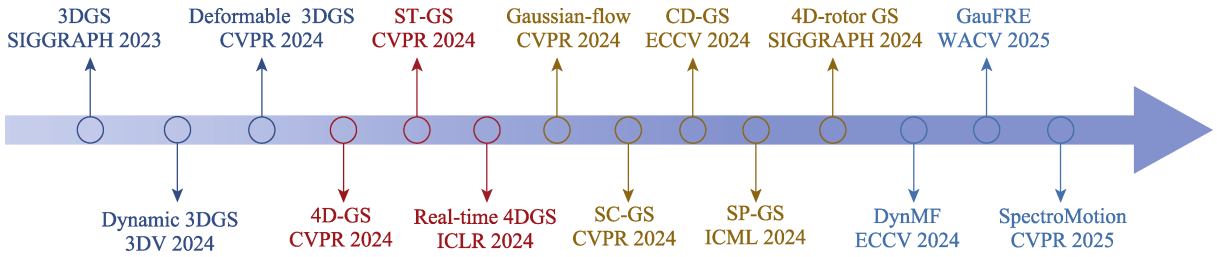


图 3　基于 3DGS 的动态三维重建领域的关键工作时间线

3DGS 方法中，在 3DGS 的基础上定义了规范空间，并利用变形网络预测经过时间 Δt 后空间高斯体发生的变化，将其作用回规范空间得到新的场景表示。变形网络将某一时刻 t 的规范空间中的高斯体的坐标位置 $\boldsymbol{\mu}$ 和时间间隔 Δt 进行编码后作为输入，预测位置、旋转和缩放的变化量；在训练时，使用退火平滑训练（annealing smooth training，AST）机制应对真实世界数据集中存在的相机位姿估计不够准确的问题，通过向时间编码添加线性递减的高斯噪声，减少姿势估计不准确造成的帧与帧之间的空间抖动现象。

4D-GS[10]创新地将 *K*-planes[54]引入变形场中，采用编解码器结构预测空间高斯体随时间的变化情况。对于动态场景，编码器首先用六角平面（HexPlane）[55]提取四维时空体素特征信息，然后经过一个微小的 MLP 处理后完成编码；解码器架构为多头高斯变形解码器，其本质上是多个独立的 MLP。该方法在编码时充分地利用周围高斯体的时空信息，可以较好地保持图像的时空一致性并减少抖动现象。

LoopGaussian[56]实现了三维空间中的动态静图。首先增加偏心正则化初始化静态场景中的点，生成指定区域的掩模；然后将该区域的高斯体通过自编码器映射到特征空间，并设计 SuperGaussian 对特征点进行聚类操作；根据聚类之间的自相似度进行稀疏速度场估计，得到稀疏速度场，再利用 Kriging 插值得到密集速度场，使用上述速度场作为监督，训练一个预测图像中每个位置运动情况的 MLP 作为欧拉运动场；最终通过欧拉运动场预测时间 t 某位置高斯体的位置改变量，实现图像的动态循环运动效果。

为了更好地表示场景随时间变化的阴影和光的反射等条件，CoGS[57]将颜色属性也作为可学习的参数进行训练。与上述方法将位置和时间编码同时输入 MLP 中训练不同，CoGS 在第 1 个阶段分别为每个属性单独训练一个 MLP；在第 2 个阶段中，为了提升动态场景的可控性，为相应的控制区域设置掩模以提取动态模型特征值，为每个区域训练 1 个 MLP 获得专属的属性变化值；最后与第 1 阶段得到的变化相结合，作用到最终的高斯体上。

CD-GS[58]采用比较紧凑的方式表示场景中的高斯体随时间的变化。该方法同样认为，一般情况下，某个场景中物体的大小在宏观情况下不会发生尺寸伸缩变化，只有场景中的高斯体位置或场景光流会发生改变，因此采用了傅里叶近似表示高斯体中心点位置随时间的变化；采用简单线性变化表示四元数随时间的变化，以改变高斯体的旋转情况；这些参数在整个时间域上共享，得到比较紧凑的表示，并能够从少量视频帧中重建某一动态场景。

在基于坐标和时间输入的变形场的基础上，PGED[59]为每个高斯体引入 32 维的、可学习的潜在嵌入，同时也为动态场景的每帧图像引入 2 种可学习的潜在嵌入，一种捕捉高频细节的动作，一种学习缓慢的动作；然后将高斯体潜在嵌入分别和 2 种帧潜在嵌入结合并输入到一个通用的浅层 MLP，再经过每个属性对应的 MLP 预测出各个属性的变化，将 2 部分输出相结合作为最终整个场景的变化。该方法还使用局部平滑正则化作用于每个高斯体及其邻近的高斯体的潜在嵌入，促进局部动作一致性。

Gaussian-flow[60]引入双域变形模型作为变形场作用于高斯体。双域指时域上的多项式表示和频域上的傅里叶级数表示方法，其无需训练神经网络模型，可以实现较快的渲染速度。

GPS-Gaussian[61]通过直接回归定义在源视图二维图像平面上的像素级高斯参数图（如位置、颜色、旋转、缩放和透明度），避免了传统三维高斯分布或神经隐式渲染方法中每个基本元素优化的需求；并且结合深度估计模块和高斯参数回归模块，能够将二维参数图提升到三维空间，利用高斯泼溅技术进行实时渲染。

MoDGS[62]用于随意捕获的单目视频的动态场景重建。该方法通过引入单视角深度估计器 GeoWizard[63]对视频帧进行二维深度图预测，并参考初始帧校准其余帧的相对深度信息和高斯点尺寸；然后结合二维光流场上升为三维光流场[64]；再通过帧间变化，训练一个可逆转的变形 MLP 完成变形场的初始化[65]；最后将每帧输入此变形场并下采样，得到初始化的规范空间。此外，在训练过程中还引入了原始深度损失。

为了表示复杂精细的动作，MoDec-GS[66]中引入全局到局部运动分解（global-to-local motion decomposition，GLMD）。首先为每个视频帧训练一个全局规范支架[67]（global canonical Scaffold-GS，Global CS），通过全局六角平面[54]对锚点的位置和属性进行变形；然后将 Global CS 转换为局部规范支架（local canonical Scaffold-GS，Local CS）捕捉场景中的全局运动。在 Local CS 中，通过显式变形调整每个高斯体的局部运动，进一步细化动态细节。此外，无需额外的运动数据，还可以动态地

调整时间间隔以优化不同运动区域的表示，并且通过可学习参数区分静态和动态部分，减少冗余计算。该方法使用基于锚点的表示和共享六面体结构，显著地降低了模型的大小。上述支架结构作为体素网格和高斯泼溅相结合的代表，通常在场景的稀疏体素网格中心处初始化锚点，并将锚点作为体素网格附近相关联的高斯体的基点，通过为每个锚点赋予高纬度特征向量，结合相机-锚点方向和距离预测相关联的高斯体特征和位置偏移，进一步优化整个场景表示，在减少存储消耗的同时增加实时渲染帧率。

在 Deformable 3DGS 的基础上，SpectroMotion[68]引入法线贴图增强几何一致性和训练过程中的反射阶段，保持变化的环境光照与规范环境光照的一致性。该方法分为 3 个阶段：第 1 阶段使用原始 3DGS 渲染静态图像，保证静态图像一致性；第 2 阶段是动态阶段，使用 Deformable 3DGS 预测高斯体的动态变化，通过优化物理法线估计法构建渲染图像的法线贴图和渲染结果深度图对应法线贴图之间的损失函数，进一步增强场景几何一致性；第 3 阶段是反射阶段，通过预计算规范环境光照图，并训练一个与入射角度和时间相关的 MLP，以计算发射光颜色，结合 0 阶 SH 表示的漫反射颜色得到最终颜色。

BARD-GS[69]专注于处理单目相机下拍摄视频重建后的模糊现象，将模糊现象分解为相机运动模糊和物体运动模糊，采用两阶段训练流程减少模糊。第 1 阶段优化相机轨迹，仅针对静态区域进行重建，消除相机运动模糊；第 2 阶段通过变形 MLP 建模动态高斯的运动轨迹，结合静态区域的优化结果完成动态区域的清晰重建。此外，还提出了首个真实世界动态模糊数据集，包含相机和物体运动模糊，支持模型训练与评估。

3.1.2　将变形场作用于动态部分高斯体

GauFRE[70]同样采用变形场表示高斯体随时间发生的变化。由于场景中存在静态部分和动态部分，因此通过一个可学习的过程将空间高斯体划分为静态和动态的部分，在训练时只将动态部分的高斯体参数形成变形网络，可大大减少计算开销，同时达到高质量的全局动态重建。

ReconDreamer++[71]用于重建动态的驾驶场景，该方法将动态场景中的物体分为 3 部分：地面的静态背景、非地面的静态背景和动态物体，并且引入新型轨迹可变形网络（novel trajectory deformable network，NTDNet）分别训练不同的部分[72-73]。对于原始运动轨迹即原始输入图像序列，该方法在渲染时并未使用 NTDNet。对于新型运动轨迹，地面高斯体对中心位置参数进行固定；背景部分的高斯体采用非地面的静态物体点云初始化，并随机增加额外点云以保证背景生成的正确性，在训练时采用 Chen 等[74]的方法；通过将轨迹偏移、高斯参数和时间戳输入到 NTDNet，预测动态物体的高斯体参数变化。

3.1.3　基于稀疏控制基预测动态轨迹

DynMF[75]也是通过变形网络预测高斯体的属性变化，但与 Deformable 3DGS 不同，DynMF 并没有将变形场作用到所有的空间高斯体上，而是定义了一组可学习的基运动轨迹点，并认为动态场景中的动作可以用一组基运动轨迹点线性表示。由于基运动轨迹点的数量通常远小于空间高斯体，因此将物体的运动转化成过约束的低秩分解问题，极大地减少了训练时的参数量，只需要训练一个较小的 MLP 预测基运动轨迹点随时间变化的系数矩阵，场景中其他的点即可以通过这组基轨迹点的线性组合表示。

与 DynMF 的思路类似，SC-GS[76]采用规范空间中的稀疏控制点作为场景高斯体的基，通过训练 1 个 MLP 预测 6 个自由度的变换矩阵；该矩阵包含 1 个场景的旋转矩阵和平移向量。在表示整个场景的高斯体上，该方法采用线性混合蒙皮（linear blend skinning，LBS）技术[77]将某个点周围邻近的 4 个点进行变换插值，最终得到整个场景中所有高斯体随时间变化的结果。

SP-GS[78]按照尽可能刚性的原则将场景中的三维高斯体聚集成超点，并为超点训练一个 MLP 预测其发生的刚性变化；然后训练一个可选择的非刚性变形场预测超点内部的高斯体的变化细节。

D-MiSo[79]用于动态三维场景的编辑和渲染，通过引入多高斯组件建模动态场景；其中，每个组件由一个核心高斯和多个子高斯组成，核心高斯负责全局变换，子高斯用于细节渲染。该方法通过将高斯参数化为三角形面片，实现了对动态场景的灵活编辑，包括局部修改、缩放和旋转，避免了传统方法中需要静态和动态标记的复杂性。D-MiSo 采用两阶段训练策略：第 1 阶段优化核心高斯，以捕捉运动；第 2 阶段添加子高斯，以提升渲染质量。

Video-3DGS[80]是一种两阶段方法，旨在通过 3DGS 技术增强动态单目视频的时序一致性。在第 1 阶段，利用改进的 MC-COLMAP 技术对视频进行空间和时间分解，分别提取前景和背景的三维点

云，并分别用 2 组高斯体进行表示，最终通过可学习的二维参数图合并渲染结果，实现高质量的视频重建。在第 2 阶段，将优化后的 Video-3DGS 作为即插即用的时序优化模块应用于零样本视频编辑器中，通过固定三维高斯结构并调整颜色参数，显著地提升了编辑视频的时序一致性和视觉连贯性。

Park 等[81]提出一种基于三次 Hermite 样条函数[82-83]的运动轨迹表示方法 SplineGS，其场景中包含一系列可学习的控制基点，这些控制基点的位置变化基于三次 Hermite 样条函数，通过基点变化快速、精确地表示场景高斯体随时间变化的动态轨迹。此外，该方法还采用了两步式训练方法，避免对相机位姿数据的任何预处理过程。

3.2 基于时空四维高斯体的方法

在三维高斯体的基础上，基于时空四维高斯体的方法拓展时间的维度，这种方法通常表现为显式地改变高斯体属性的维度，或通过引入随时间变化的属性函数表示四维的高斯体结构。在渲染时，固定四维高斯体的时间维度来表示某一时刻的辐射场，即通过采样不同时刻的四维高斯体表示动态的场景变化。

Real-time 4DGS[84]首次通过重定义高斯体属性维度将原始三维高斯体拓展到了四维，将时间维度引入高斯体的表示中，使用四维旋转矩阵和一对各向同性的旋转项表示旋转矩阵；然后将时间条件限制下的四维高斯分布函数与对应时刻的边际分布相乘来渲染动态场景。

Chen 等[74]将场景中的物体统一表示为周期振动高斯体（periodic vibration Gaussian，PVG），其中，位置和不透明度定义为时间相关函数，并引入寿命峰值 τ 和该时刻的瞬时速度 $\boldsymbol{v}$；通过查询给定时刻 t 将此时的 PVG 映射到三维空间，最后完成渲染操作。在训练时，该方法采用时间平滑机制，通过查询下一时刻的状态，利用平均速度与时间差的乘积计算每个高斯体的位移，以监督时刻 t 的训练效果。此外，还采用位置感知点自适应控制[85]，以更好地控制无界场景中远景和近景高斯体的增减。

ST-GS[86]中引入时空高斯体（spacetime Gaussian，STG）的概念，在原始三维高斯体的基础上，其引入时间径向基函数[87]表示高斯体中心点在不同时刻的不透明度，而高斯体内其他各点则沿用原始 3DGS 的计算方法；通过多项式运动轨迹和多项式旋转分别表示高斯体中心点和旋转随时间的变化[88-89]。此外，通过 2 层浅 MLP 将时间维度特征、视角方向特征和视角方向作为输入，将中间特征与基本特征相结合，在降低存储开销的同时不影响渲染速度。该方法在训练时还采用引导采样技术，以减少遥远区域的伪影和模糊现象。

4D-rotor GS[90]中引入基于转子的旋转因子表示法，将空间高斯体升维到了四维 STG，通过对时间维度进行切片，生成每个时刻的动态三维高斯体表示，最终将这些不同时刻的高斯体投影到二维平面上，实现了实时、高质量的渲染。此外，该方法还引入熵损失迫使高斯体的不透明度趋于 0 或 1（更符合理想情况），并利用四维一致性损失保证邻近高斯体的时空一致性。

通过对基于变形场或时空四维高斯体方法进行梳理与分析，本文总结出基于 3DGS 的动态场景重建方法的通用工作管线，涵盖从动态表征到最终渲染的核心模块与流程。如图 4 所示。

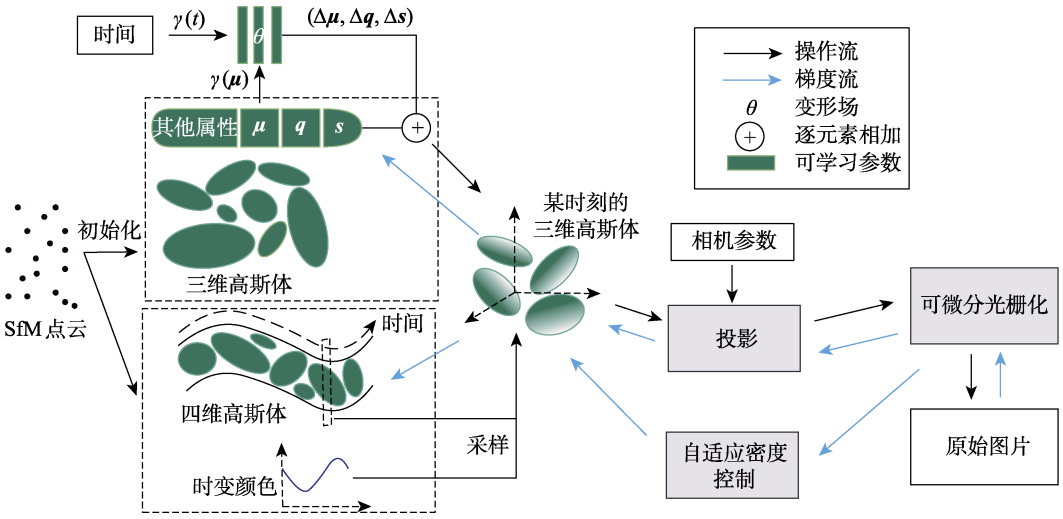


图 4 基于 3DGS 的动态场景重建方法的通用工作管线

3.3　基于逐帧训练帧间动态传递的方法

顾名思义，基于逐帧训练帧间动态传递的方法通常会选定一系列的关键帧，并在关键帧处初始化三维高斯体，在非关键帧时，基于前面的关键帧和已经训练好的帧训练场景中高斯体的动态变化。与基于变形场方法的本质不同，该方法不是直接利用变形 MLP 预测不同时间或帧间高斯体的动态变化，而是基于前面帧训练当前帧中的场景高斯体，通常需要多视角图像或视频作为输入，或更加精准的相机位姿信息和一些几何先验信息（如深度图等）。

3.3.1　输入多视角图像或视频

Dynamic 3DGS[91]在首帧初始化此时的静态三维高斯分布，即对其进行静态重建；除了位置和旋转之外，固定空间高斯体的其他参数，使该方法能够跨帧跟踪高斯体的动态轨迹；其优化是通过正则化进行的，并未使用神经网络。

3DGStream[92]同样在首帧初始化三维高斯体，它使用上一帧中的三维高斯体作为下一帧训练的初始化条件，并采用两阶段式的训练方法，先优化神经变换缓存（neural transformation cache，NTC）变换帧间高斯体位置和旋转，再优化局部新增高斯体表示帧间可能突现出的物体或遮挡关系。

Dual-GS[93]通过分离运动感知的关节高斯和外观感知的皮肤高斯，实现了高效的人体动态建模与渲染。在训练时，采用从粗到细的优化策略，先通过关节高斯进行全局运动预测，再通过皮肤高斯优化细节外观，并结合残差向量量化、编解码压缩和持久码本等技术，实现高达 120 倍的压缩率，最终在虚拟现实设备上实现了实时高保真渲染。通过将动态三维高斯体的时空属性编码为多通道二维视频流，实现了移动端实时流式传输与高保真体视频渲染。Dual-GS 的核心是两阶段优化策略：首先利用哈希编码和浅层 MLP 快速地估计高斯基元的运动，通过残差熵损失和时间一致性损失微调外观细节，增强时间连续性；然后进行分组训练和关键帧剪枝，将长序列分割为 20 帧的组，首帧作为关键帧进行静态 3DGS 重建和低透明度高斯剪枝，后续帧基于运动估计生成紧凑表示。

STC-GS[94]采用一组高斯体优化每帧的场景，有效地捕捉它们在连续帧中的运动，用于预测各种动态气象雷达信号。该方法采用双向重建策略，首先从时间 T 到时间 0 预重建场景高斯体，该步骤只有位置参数进行优化，有效地提取了未来的信息；然后进行正向过程和全参数优化。此外，在训练时还采用局部细节调整和全局控制策略[95]。

Yan 等[96]提出一种流式动态场景重建框架 IGS，旨在应对现有方法在实时性和误差累积方面的挑战。该方法通过单次推理计算高斯体在相邻帧之间的运动，避免逐帧优化；利用锚点（即运动控制基）采样和三维运动特征增强，仅需要单次推理即可预测高斯点运动，避免逐帧优化；同时，采用关键帧引导策略、周期性优化关键帧并限制高斯点数量，以减少误差累积，进而提升复杂动态场景的重建质量。

3.3.2　精准位姿信息和丰富的几何先验

GFlow[97]仅通过图像的二维先验信息（深度和光流信息）动态重建单目视频拍摄的场景，其输入为图像序列、深度信息、光流信息和已知的相机内参，可以自动地将场景中的物体解耦为静态和动态的高斯体。在优化过程中，首先根据输入的先验信息和场景中的静态聚类优化相机位姿信息；然后利用优化后的位姿信息和先验信息进行耦合，联合优化静态与动态高斯体参数，以提升模型整体的一致性。

Seidenschwarz 等[98]提出一种针对单目无位姿视频的在线点跟踪方法 DynOMo，通过引入额外的语义标签和视觉特征提升跟踪效果。该方法首先基于预计算的深度图和对应的相机位姿，从视频首帧中提取场景的高斯体表示，将其作为后续帧的参考模型；然后利用静态区域信息通过重建损失优化相机位姿，并进一步将重建损失和三维正则项相结合，联合优化所有高斯参数；最后借助运动先验算法（如物理或语义运动先验）和特征相似性信息，预测下一帧的高斯体状态。

4　性能比较

首先介绍动态场景重建常见的数据集，它们的关键信息如表 1 所示，并将这些数据集分为单目数据集和多视角数据集 2 大类；然后介绍一些图像生成和视频评价领域的可量化的评价指标；最后在一些使用较为广泛的数据集基础上，对比文中介绍的一些具有代表性的基于 3DGS 的动态场景重建方法，并以表格的方式量化和分析各种重建方法的性能指标，若无特殊说明，这些数值均来源于原文献。

4.1　动态场景数据集

动态场景数据集分为单目数据集和多视角数据集。其中，单目数据集多为单摄像机或少量摄像机保持一致的移动轨迹所拍摄，多视角数据集通常

表 1 11 种动态场景数据集信息

数据集	类别	分辨率	数据集大小	分类数目	主要内容
D-NeRF[7]	单目数据集	800×800	560~1 760 帧	8	合成数据
HyperNeRF[11]	单目数据集	≥1 280×720	≥1 700 帧	17	真实世界复杂场景
iPhone[28]	单目数据集	720×480	4 180 帧	14	真实世界非重复动作
NeRF-DS[99]	单目数据集	480×270	5 134 帧	8	真实世界镜面反射场景
Motion Blur[69]	单目数据集			12	真实世界动态模糊场景
Waymo[100]	多视角数据集	1 920×1 280	230 000 帧*	1 150*	真实世界自动驾驶场景
Google Immersive[101]	多视角数据集			15	真实世界复杂场景
Meet Room[102]	多视角数据集	1 280×720	15 600 帧	4	复杂室内场景
Plenoptic Video[27]	多视角数据集	2 028×2 704		6	复杂室内场景
HiFi4G[103]	多视角数据集			6	人体动作
ParticleNeRF[104]	多视角数据集			6	合成数据

注：≥表示可能超过此数值或存在超过此数值的数据，*表示初始版本的数值。

为多个摄像机从不同角度同时拍摄某动态场景。

4.1.1 单目数据集

D-NeRF 数据集[7]是专为动态场景重建与视图合成设计的合成数据集，用于评估非刚性物体在单目视频中的重建性能。该数据集包含 8 个合成动态场景，如跳跃人偶、恐龙和乐高模型等，涵盖不同的运动模式，如肢体摆动、旋转或形变等；每个场景提供 50~200 张训练视图和 20 张测试视图，分辨率为 800×800 像素。

HyperNeRF 数据集[11]是专为动态场景中复杂非刚性形变建模设计的真实单目视频数据集，核心目标是通过高维表示捕捉时间连续性与拓扑变化。该数据集包含多个复杂动态场景，如人体动作、物体形变、遮挡变化等；其中，典型场景包含 100~300 帧，分辨率为 1 280×720 像素或更高，覆盖单目多视角下动态物体的连续变化。

iPhone 数据集[28]是一个真实世界多类别、非重复的物体运动数据集，包含宠物、水果、汽车等 14 种物体的非重复运动的视频，由 7 个单摄像机和 7 个多摄像机拍摄视频构成，多摄像机部署为一台手持移动摄像机用于训练，2 台静态摄像机用于评估；此外，还包含激光雷达传感器采集的深度信息。

NeRF-DS 数据集[99]是一个真实世界动态镜面反射场景数据集，包含 8 个场景，涵盖刚性（如金属球、静态反光物体）和非刚性变形（如液体表面、可弯曲镜面材料）的镜面反射物体。场景由 2 个固定安装的摄像头捕捉，一个摄像头用于训练，另一个用于测试。

Motion Blur[69]是首个真实世界动态模糊数据集，使用 2 台同步的摄像机捕获由相机运动和快速移动的物体引起的模糊。其中，一个摄像头的快门速度为 1/24 s，可拍摄 24 帧/s 的模糊视频；另一个摄像头的快门速度为 1/240 s，可拍摄 240 帧/s 的清晰视频，该摄像头仅用于评估。

4.1.2 多视角数据集

Plenoptic Video 数据集[27]又称为 N3DV 数据集，是基于 21 台摄像机的移动多视角光场视频数据集，以 2 028×2 704 像素的分辨率和 30 帧/s 同步采集动态场景，包含 6 个真实世界的场景（其中的场景 flame salmon 被分为 4 个视频片段）。通过 COLMAP 标定相机内外参数，其中，常规场景使用 18 个视角训练、1 个视角测试，多人复杂运动场景则采用 14 个训练和 1 个测试的划分策略。该数据集突破了传统静态光场的限制，提供了高分辨率、精确标定的动态多视角视频。

Waymo 数据集[100]提供了丰富的真实世界自动驾驶场景数据，旨在推动自动驾驶技术的研发和创新。车身分布 5 个高分辨率 LiDAR 传感器（顶部 1 个和四周 4 个）提供三维点云数据；顶部的 5 个高动态范围摄像头（前、左前、右前、左后和右后）拍摄长为 20 s 的分辨率为 1 920×1 280 像素的视频片段；5 个雷达传感器提供速度和距离信息。此外，还提供三维边界框、二维图像分割信息，以及多个版本迭代更加精细化的标注信息。

Meet Room 数据集[102]使用 13 个摄像头的多视角系统捕获，包含以 1 280×720 像素的分辨率和 30 帧/s 记录的动态场景。该数据集使用 12 台摄像机进行训练，1 台进行测试，并使用外部脉冲信号同步所有摄像机快门。

Google Immersive 数据集[101]是一个多视角动态场景数据集，包含 15 个场景，涵盖多样化的室内外环境（如舞蹈表演、运动场景等）；每个场景

的摄像机数量最大为 46 个。该数据集还提供了由 SfM 算法生成的详细摄像头校准信息[105]。

HiFi4G 数据集[103]是一个信息丰富的多视图数据集，专注于人体运动。该数据集包含 6 个场景，共 81 台摄像机，涵盖各种复杂的人体运动，如演奏乐器、跳舞和换衣服等。

ParticleNeRF 数据集[104]是一个使用 Blender 生成动态场景的合成数据集，包含 6 个场景，由分布在球体上半部的 20 台摄像机捕捉。

4.2　评价指标

基于 3DGS 的动态场景重建方法相关工作中，常用的评价指标有峰值信噪比（peak signal-to-noise ratio，PSNR）、结构相似度（structural similarity index measure，SSIM）、学习感知图像块相似度（learned perceptual image patch similarity，LPIPS）等，本文选取的评价指标[106-109]的关键信息如表 2 所示。

（1）PSNR

PSNR 是一种广泛用于评估信号重建质量（如图像、视频压缩或传输）的客观度量指标，尤其适用于图像压缩和恢复任务中，它通过比较原始信号与处理后信号之间的误差，用于衡量重建信号的保真度。PSNR 基于均方误差（mean squared error，MSE）计算，单位为 dB，其核心思想是衡量信号的最大可能功率（峰值）与噪声功率（误差）的比值，其值越高表示信号质量越好。然而，由于 PSNR 无法完全反映人类对图像质量的感知，因此在评估感知相似性方面存在一定的局限性。其计算公式为

$$\mathrm{PSNR}=10\times\lg\left(\frac{P_{\max}^2}{\mathrm{MSE}}\right)。$$

其中，$P_{\max}$ 表示图像上像素点的最大数值，通常定义为 1 或 255；MSE 定义为

$$\mathrm{MSE}=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I(i,j)-K(i,j)\right\|^2，$$

$I(i,j)$ 表示参考图像或原始图像在位置 (i,j) 处的像素值，$K(i,j)$ 表示待比较图像（即重建后的图像）在位置 (i,j) 的像素值。

（2）SSIM

SSIM[106]是一种衡量 2 幅图像相似度的指标，取值为−1~1；其中，1 表示 2 幅图像完全相同，−1 则表示 2 幅图像完全不同。SSIM 基于人眼视觉系统（human visual system，HVS）的特性，认为人眼对图像的结构信息更敏感，而非简单的亮度或对比度变化。SSIM 中包含多尺度 SSIM[110]（multiscale SSIM，MS-SSIM）和结构不相似性（structural dissimilarity，DSSIM），是一种更符合人类视觉感知的评价指标。

（3）LPIPS

LPIPS[107]是一种基于深度学习的图像相似性度量方法。与传统的图像质量评估指标（如 PSNR，SSIM）不同，LPIPS 通过预训练的神经网络提取图像特征，并通过计算特征空间中的距离衡量图像的感知相似性。在人类对图像质量的感性判断方面，LPIPS 已经被证明比传统的指标更准确，因此在图像生成、超分辨率、风格迁移等任务中被广泛使用。LPIPS 取值为 0~1，其值越低，表明 2 幅图像在人的感知上越相似，0 表示在感觉上完全相似，反之为 1。

（4）视频多方法评估融合

视频多方法评估融合（video multi-method assessment fusion，VMAF）[108]是一种全参考视频质量评估算法，旨在更准确地预测人类对视频质量的主观感受。VMAF 结合多个底层质量指标，并通过机器学习模型（如 SVM）提取信息，以更接近 HVS 的感知。VMAF 不仅考虑了画面细节、锐度、饱和度、压缩失真等多种视觉因素，还能够在复杂场景下保持较高的评估鲁棒性。

（5）Fréchet 初始距离

Fréchet 初始距离（Fréchet inception distance，FID）[109]是一种评估生成对抗网络[111]或其他生成模型性能的指标，用于衡量生成图像与真实图像分布之间的相似性。FID 使用预训练的 Inception-v3 网络提取图像特征，假设这些特征服从多元高斯分布，然后计算生成图像与真实图像特征分布的 Fréchet 距离（即 2 个高斯分布之间的 Wasserstein-2 距离）。FID 值越低，表示生成图像与真实图像在统计特性上越接近。与简单的像素级差异（如 MSE）或人工评估相比，FID 能够更深层次地反映生成图像的质量。

表 2　常用评价指标的关键信息

指标	取值范围	评价内容
PSNR	[0，+∞] ↑	像素精度
SSIM[106]	[−1，1] ↑	结构相似性
LPIPS[107]	[0，1] ↓	语义相似性
VMAF[108]	[0，100] ↑	综合视频质量
FID[109]	[0，+∞] ↓	语义相似性

注：↑表示指标越高越好，↓表示指标越低越好。

4.3 对比结果

高斯泼溅方法以其独特的显式表示方法和泼溅式的可微光栅化渲染流程为动态场景重建提供了全新的研究方向。在 D-NeRF[7]、HyperNeRF[11]和 NeRF-DS[99]这 3 个应用最广泛的数据集上，对单目动态场景进行重建，并采用原文献方法中对应的数据，对比 PSNR、SSIM、MS-SSIM、LPIPS 等指标，结果分别如表 3~表 5 所示。其中，对比时各种方法使用的 GPU 和最终渲染结果的分辨率不同，相关说明见表注。

从表 3 可以看出，SC-GS 方法的各项指标均为最优，这是由于该方法在变形场的基础上引入稀疏控制点作为高斯体的基，高效地将场景变化分解成 6 个自由度的矩阵，并采用 LBS 技术进行插值计算所有高斯体的变形，同时采用较为高效又符合物理规律的做法，因此取得了优异成绩。Deformable 3DGS 和 SP-GS 的表现也比较优秀，3 项指标均分别位居第 2 名和第 3 名。

表 3 在 D-NeRF 数据集[7]上 14 种方法的性能对比

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓
D-NeRF[7]	31.69	0.975	0.058
3DGS[21]	23.19	0.930	0.080
HexPlane[55]	31.04	0.970	0.040
K-planes[54]	32.32	0.973	0.038
*Real-time 4DGS[84]	34.09	0.980	0.020
CD-GS[58]	32.19	0.970	0.040
*4D-GS[10]	34.05	0.980	0.020
4D-rotor GS[90]	34.26	0.970	0.030
Gaussian-flow[60]	34.27	0.980	0.030
DynMF[75]	36.89	0.983	0.020
*SP-GS[78]	37.98	0.988	0.019
*Deformable 3DGS[9]	39.51	0.991	0.012
SC-GS[76]	43.31	0.997	0.006
GauFRE[70]	34.80	0.982	0.020

注：加粗的方法表示提出相应数据集时采用的方法；*表示该数据集最终渲染的分辨率为 800×800 像素；各项指标第 1 名~第 3 名的得分分别用红色、绿色和蓝色表示。

表 4 在 HyperNeRF 数据集[11]上 9 种方法的性能对比

方法	PSNR/dB ↑	MS-SSIM ↑	训练时间/min ↓
HyperNeRF[11]	23.47	0.856	
4D-GS[10]	25.20	0.845	30
DynMF[75]	25.27	0.750*	
PGED[59]	25.43	0.697	75
Gaussian-flow[60]	26.30	0.862	12
SP-GS[78]	26.78	0.892	
SpectroMotion[68]	22.22	0.639	
GauFRE[70]	24.10	0.830	90
MoDec-GS[66]	27.78	0.827*	

注：加粗的方法表示提出相应数据集时采用的方法；*表示该方法所用的指标为 SSIM；各项指标第 1 名~第 3 名的得分分别用红色、绿色和蓝色表示。

表 5 在 NeRF-DS 数据集[99]上 6 种方法的性能对比

方法	PSNR/dB ↑	MS-SSIM ↑	LPIPS ↓
NeRF-DS[99]	23.90	0.898	0.127
SP-GS[78]	23.15	0.834	0.206
Deformable 3DGS[9]	24.11	0.853*	0.177
SC-GS[76]	24.10	0.891	0.140
GauFRE[70]	23.90	0.887	0.148
SpectroMotion[68]	24.65	0.887	0.159

注：加粗的方法表示提出相应数据集时采用的方法；*表示该方法所用的指标为 SSIM；各项指标第 1 名~第 3 名的得分分别用红色、绿色和蓝色表示。

从表 4 可以看出，SP-GS 以次高的 PSNR 得分和最高的 MS-SSIM 得分表现出色，表明该方法在处理高维表示和生成准确的新视图方面比较优秀；Gaussian-flow 和 MoDec-GS 也取得不错的效果，且 Gaussian-flow 的训练时间极短，反映了渲染时的高效性。

从表 5 可以看出，在 NeRF-DS 数据集上各种方法的竞争比较激烈，指标值都比较接近，其中，SpectroMotion 的 PSNR 得分最高，反映其镜面反射场景渲染图像的高质量，MS-SSIM 的得分次高，反映其对动态镜面反射场景的优秀重建能力；SC-GS 的所有指标得分均位居第 2 名，表明了该方法在不同场景下的适用性。在提出该数据集时采用的 NeRF-DS 竞争力依然存在，说明在重建动态镜面反射场景的情况下，基于高斯泼溅的方法有待进一步优化和提升。

在多视角数据集上的动态场景重建中，本文采用 Plenoptic Video 数据集进行实验，对比 6 种方法的 PSNR、SSIM、LPIPS 和帧速指标，部分结果如表 6 所示。可以看出，4D-rotor GS 的帧速最高，展示了其出色的实时渲染能力，凸显该方法中四维转子高维表示法的高效性；Real-time 4DGS 具有最高的 PSNR 得分和较高的帧速，展示了其强大的性能；在图像质量指标方面，Gaussian-flow 的 SSIM 得分最高和 PSNR 得分次高，表明采用光流方法在捕捉复杂动态场景的图像合成方面也有突出效果。

表 6　在 Plenoptic Video 数据集[27]上 6 种方法的性能对比

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	帧速/（帧·s^{-1}）↑
Real-time 4DGS[84]	32.01		0.055	114.0
4D-GS[10]	31.15		0.049	30.0
PGED[59]	31.31	0.945	0.037	74.5
4D-rotor GS[90]	31.62	0.940	0.140	277.5
DynMF[75]	31.70	0.946	0.180	135.0
Gaussian-flow[60]	32.00	0.970		

注：各项指标第 1 名~第 3 名的得分分别用红色、绿色和蓝色表示。

5　结　语

自从 2023 年 3DGS 被提出后，与之前的模型相比，其在模型效果、训练速度和渲染速度上都有了极大的提升，后续工作也被拓展到各个领域中，并在许多传统方法存在局限的问题上取得了突破。例如，国内知名 XR 品牌 PICO①旗下的产品 PICO 4 Ultra 成功地实现了自主优化的 3DGS 渲染方法，并将这种优化的渲染设计提供给 Unreal Engine②作为插件、美国初创公司 Luma AI③提供基于 Web 的平台和 API，利用 3DGS 生成和渲染高质量图像。本文对该领域当前的研究现状进行综述，未来可能的研究方向如下。

（1）新型高斯结构和其他数据结构设计

目前，已有的方法除了精确地表示动态场景复杂的几何结构外，还能够将时间维度完美融合，在共同表征动态场景上还存在许多不足之处。一方面，动态的场景往往包含复杂的时空信息，与静态场景的三维信息不同，动态意味着随时间发生变化，如何设计高斯体结构，使其在不失物理意义的同时又能表示丰富的新增时间维度信息，是一个巨大的挑战；另一方面，与静态场景相比，动态场景中的信息量激增，如何高效地存储场景中的元素信息并且在一定程度上减少计算开销，也是一大难题。在此背景下，未来应着眼于新的适用于动态场景重建的高斯结构或新的场景数据结构设计，如将传统耦合的时空参数分离为空间属性（静态几何特征）和时间属性（动态运动参数）。通过建立时变关系模型，在保持物理合理性的同时，可采用线性运动轨迹和低维形变系数等简洁形式表征动态变化，避免复杂时空耦合带来的计算负担和优化困难。

（2）利用稀疏视角及微量位姿信息进行重建

当前，动态场景重建通常依赖密集的多视角图像或视频，以及精确的相机位姿标注信息，但实际应用中，输入数据往往是稀疏视角且位姿信息存在噪声或缺失。未来，可以设计对位姿误差不敏感的高斯泼溅优化算法，以及联合优化场景几何与相机位姿（类似 SLAM 中的 Bundle Adjustment），并引入隐式位姿校正模块；或者，利用动态场景的时间连续性，通过光流或时序 Transformer 预测缺失视角的几何先验，增强稀疏输入下的重建完整性；甚至可以开发无需真实三维标注的训练范式，降低对标注数据的依赖。

（3）大规模场景的重建及实时渲染

虽然现有的高斯泼溅方法在小型场景中表现良好，但是扩展到城市级场景或长时序动态环境时面临巨大挑战，这一局限性主要源于场景规模的扩展带来的计算与存储需求显著增长，而当前硬件条件无法满足此类极端情况的实时处理需求。因此，未来需要关注如何突破效率瓶颈：a. 在计算效率方面，可以结合场景语义理解（如建筑物、车辆、行人等类别分割）对静态背景和动态物体实施差异化建模，通过运动特征感知的计算资源分配显著地减少冗余操作；同时，借鉴游戏引擎中的 LOD 技术[112]，基于视角距离和运动速度自适应地调节高斯体的密度和半径，实现实时的交互级渲染。b. 在存储优化方面，通过构建层次化高斯金字塔结构[113]，配合视锥体剔除与动态加载机制，智能地管理不同细节层次的高斯体，实现存储开销的渐进式控制。

本文聚焦基于 3DGS 的动态场景重建领域，对现有的工作进行梳理，较为详细地介绍了其中最具代表性的方法并进行分类，在渲染质量、效率、实时性等方面对其进行比较。此外，指出了该方法当前面临的挑战和局限性，并展望了未来的研究方向。希望本文可以让更多的研究者了解这一领域，并为后续工作提供有价值的参考。最后，衷心地期望动态场景重建领域能够涌现出更多基于 3DGS 的创新工作。

① https://www.picoxr.com/cn
② https://www.unrealengine.com
③ https://lumalabs.ai

参考文献（References）：

[1] Foley J D. Computer graphics: principles and practice[M]. Boston: Addison-Wesley, 1995

[2] Szeliski R. Computer vision: algorithms and applications[M]. 2nd ed. Cham: Springer, 2022

[3] Pharr M, Jakob W, Humphreys G. Physically based rendering: from theory to implementation[M]. 4th ed. Cambridge: MIT Press, 2023

[4] Vedula S, Rander P, Collins R, *et al*. Three-dimensional scene flow[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 475-480

[5] Newcombe R A, Fox D, Seitz S M. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 343-352

[6] Dou M S, Khamis S, Degtyarev Y, *et al*. Fusion4D: real-time performance capture of challenging scenes[J]. ACM Transactions on Graphics, 2016, 35(4): Article No.114

[7] Pumarola A, Corona E, Pons-Moll G, *et al*. D-NeRF: neural radiance fields for dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 10313-10322

[8] Li Z Q, Niklaus S, Snavely N, *et al*. Neural scene flow fields for space-time view synthesis of dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 6494-6504

[9] Yang Z Y, Gao X Y, Zhou W, *et al*. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 20331-20341

[10] Wu G J, Yi T R, Fang J M, *et al*. 4D Gaussian splatting for real-time dynamic scene rendering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 20310-20320

[11] Park K, Sinha U, Hedman P, *et al*. HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields[J]. ACM Transactions on Graphics, 2021, 40(6): Article No.238

[12] Garg R, Roussos A, Agapito L. Dense variational reconstruction of non-rigid surfaces from monocular video[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1272-1279

[13] Yang G S, Sun D Q, Jampani V, *et al*. LASR: Learning articulated shape reconstruction from a monocular video[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 15975-15984

[14] Lin W B, Zheng C W, Yong J H, *et al*. OcclusionFusion: occlusion-aware motion estimation for real-time dynamic 3d reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 1726-1735

[15] Waechter M, Moehrle N, Goesele M. Let there be color! Large-scale texturing of 3D reconstructions[C] //Proceedings of the 13th European Conference on Computer Vision. Heidelberg: Springer, 2014: 836-850

[16] Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D[C] //Proceedings of the ACM SIGGRAPH Conference. New York: ACM Press, 2006: 835-846

[17] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376

[18] Sand P, Teller S. Particle video: long-range motion estimation using point trajectories[J]. International Journal of Computer Vision, 2008, 80(1): 72-91

[19] Bregler C, Hertzmann A, Biermann H. Recovering non-rigid 3D shape from image streams[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2000: 690-696

[20] Mildenhall B, Srinivasan P P, Tancik M, *et al*. NeRF: representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2022, 65(1): 99-106

[21] 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): Article No.139

[22] Huang B B, Yu Z H, Chen A P, *et al*. 2D Gaussian splatting for geometrically accurate radiance fields[C] //Proceedings of the ACM SIGGRAPH 2024. New York: ACM Press, 2024: Article No.32

[23] Innmann M, Zollhöfer M, Nießner M, *et al*. VolumeDeform: real-time volumetric non-rigid reconstruction[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 362-379

[24] Levoy M, Hanrahan P. Light field rendering[M] //Whitton M C. Seminal Graphics Papers: Pushing the Boundaries, Volume 2. New York: Association for Computing Machinery, 2023: Article No.47

[25] Chen S E, Williams L. View interpolation for image synthesis[M] //Whitton M C. Seminal Graphics Papers: Pushing the Boundaries, Volume 2. New York: Association for Computing Machinery, 2023: Article No.45

[26] Gafni G, Thies J, Zollhöfer M, *et al*. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 8645-8654

[27] Li T Y, Slavcheva M, Zollhoefer M, *et al*. Neural 3D video synthesis from multi-view video[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5511-5521

[28] Gao H, Li R L, Tulsiani S, *et al*. Monocular dynamic view synthesis: a reality check[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM Press, 2022: Article No.2447

[29] Zhang L, Curless B, Seitz S M. Spacetime stereo: shape recovery for dynamic scenes[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2003:

367-374

[30] Lombardi S, Simon T, Saragih J, *et al*. Neural volumes: Learning dynamic renderable volumes from images[J]. ACM Transactions on Graphics, 2019, 38(4): Article No.65

[31] Chen Y Y, Wu W M, Peng Y S, *et al*. ESA-GS: elongation splitting and assimilation in Gaussian splatting for accurate surface reconstruction[J]. Computer Aided Geometric Design, 2025, 119: Article No.102434

[32] Müller T, Evans A, Schied C, *et al*. Instant neural graphics primitives with a multiresolution hash encoding[J]. ACM Transactions on Graphics, 2022, 41(4): Article No.102

[33] Zhang C B, Yan J C, Wei Y, *et al*. OccNeRF: self-supervised multi-camera occupancy prediction with neural radiance fields[OL]. [2025-07-04]. https://arxiv.org/abs/2312.09243v1

[34] Rünz M, Agapito L. Co-fusion: real-time segmentation, tracking and fusion of multiple objects[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2017: 4471-4478

[35] Turki H, Ramanan D, Satyanarayanan M. Mega-NeRF: scalable construction of large-scale NeRFs for virtual fly-throughs[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12912-12921

[36] Yu A, Li R L, Tancik M, *et al*. PlenOctrees for real-time rendering of neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5732-5741

[37] Boss M, Braun R, Jampani V, *et al*. NeRD: neural reflectance decomposition from image collections[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 12664-12674

[38] Zeng J Y, Bao C, Chen R, *et al*. Mirror-NeRF: learning neural radiance fields for mirrors with whitted-style ray tracing[C] //Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM Press, 2023: 4606-4615

[39] Senel N, Kefferpütz K, Doycheva K, *et al*. Multi-sensor data fusion for real-time multi-object tracking[J]. Processes, 2023, 11(2): Article No.501

[40] Rematas K, Liu A, Srinivasan P, *et al*. Urban radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12922-12932

[41] Liu A J, Tan Z C, Li X, *et al*. Static and dynamic fusion for multi-modal cross-ethnicity face anti-spoofing[OL]. [2025-07-04]. https://arxiv.org/abs/1912.02340

[42] Ingale A K, J. D U. Real-time 3D reconstruction techniques applied in dynamic scenes: a systematic literature review[J]. Computer Science Review, 2021, 39: Article No.100338

[43] Kajiya J T, Von Herzen B P. Ray tracing volume densities[J]. ACM SIGGRAPH Computer Graphics, 1984, 18(3): 165-174

[44] Zwicker M, Pfister H, van Baar J, *et al*. Surface splatting[C] //Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 371-378

[45] Brebin R A, Carpenter L, Hanrahan P. Volume rendering[M] //Wolfe R. Seminal Graphics: Pioneering Efforts That Shaped the Field. New York: Association for Computing Machinery, 1998: 363-372

[46] Pfister H, Zwicker M, van Baar J, *et al*. Surfels: surface elements as rendering primitives[C] //Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2000: 335-342

[47] Chen G K, Wang W G. A survey on 3D Gaussian splatting[OL]. [2025-07-04]. https://arxiv.org/abs/2401.03890

[48] Bao Y Q, Ding T Y, Huo J, *et al*. 3D Gaussian splatting: survey, technologies, challenges, and opportunities[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(7): 6832-6852

[49] Fei B, Xu J Y, Zhang R, *et al*. 3D Gaussian splatting as a new era: a survey[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(8): 4429-4449

[50] Liu Yebin, Su Hao, Gao Lin, *et al*. Research trends and major developments in 3D vision in 2024[J]. Journal of Image and Graphics, 2025, 30(6): 1717-1743(in Chinese)
(刘烨斌, 苏昊, 高林, 等. 2024 年度三维视觉前沿趋势与十大进展[J]. 中国图象图形学报, 2025, 30(6): 1717-1743)

[51] Zhu J X, Tang H. Dynamic scene reconstruction: recent advance in real-time rendering and streaming[OL]. [2025-07-04]. https://arxiv.org/abs/2503.08166

[52] Huang N, Wei X B, Zheng W Z, *et al*. S^3Gaussian: self-supervised street gaussians for autonomous driving[OL]. [2025-07-04]. https://arxiv.org/abs/2405.20323

[53] Cai Y, Liang Y, Wang J, *et al*. Radiative Gaussian splatting for efficient x-ray novel view synthesis[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2024: 283-299

[54] Fridovich-Keil S, Meanti G, Warburg F R, *et al*. *K*-planes: explicit radiance fields in space, time, and appearance[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 12479-12488

[55] Cao A, Johnson J. HexPlane: a fast representation for dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 130-141

[56] Li J Y, Cheng L C, Wang Z Y, *et al*. LoopGaussian: creating 3D cinemagraph with multi-view images via Eulerian motion field[C] //Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM Press, 2024: 476-485

[57] Yu H, Julin J, Milacski Z Á, *et al*. CoGS: controllable Gaussian splatting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 21624-21633

[58] Katsumata K, Vo D M, Nakayama H. A compact dynamic 3D Gaussian representation for real-time dynamic view synthesis[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2024: 394-412

[59] Bae J, Kim S, Yun Y, *et al*. Per-Gaussian embedding-based deformation for deformable 3D Gaussian splatting[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2024: 321-335

[60] Lin Y T, Dai Z Z, Zhu S Y, *et al*. Gaussian-flow: 4D reconstruction with dynamic 3D Gaussian particle[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 21136-21145

[61] Zheng S Y, Zhou B Y, Shao R Z, *et al*. GPS-Gaussian: generalizable pixel-wise 3D Gaussian splatting for real-time human novel view synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 19680-19690

[62] Liu Q M, Liu Y, Wang J P, *et al*. MoDGS: dynamic Gaussian splatting from casually-captured monocular videos with depth priors[C] //Proceedings of the 13th International Conference on Learning Representations. Washington DC: ICLR, 2025: 4568-4594

[63] Fu X, Yin W, Hu M, *et al*. GeoWizard: unleashing the diffusion priors for 3D geometry estimation from a single image[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2024: 241-258

[64] Teed Z, Deng J. RAFT: recurrent all-pairs field transforms for optical flow[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 402-419

[65] Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using real NVP[C] //Proceedings of the 5th International Conference on Learning Representations. Toulon: Curran Associates, 2017: 503-534

[66] Kwak S, Kim J, Jeong J Y, *et al*. MoDec-GS: global-to-local motion decomposition and temporal interval adjustment for compact dynamic 3D Gaussian splatting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 11338-11348

[67] Lu T, Yu M L, Xu L N, *et al*. Scaffold-GS: structured 3D Gaussians for view-adaptive rendering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 20654-20664

[68] Fan C D, Chang C W, Liu Y R, *et al*. SpectroMotion: dynamic 3d reconstruction of specular scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 21328-21338

[69] Lu Y R, Zhou Y L, Liu D S, *et al*. BARD-GS: blur-aware reconstruction of dynamic scenes via Gaussian splatting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 16532-16542

[70] Liang Y Q, Khan N, Li Z Q, *et al*. GauFRE: Gaussian deformation fields for real-time dynamic novel view synthesis[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2025: 2642-2652

[71] Zhao G S, Wang X F, Ni C J, *et al*. ReconDreamer++: Harmonizing generative and reconstructive models for driving scene representation[OL]. [2025-07-04]. https://arxiv.org/abs/2503.18438

[72] Ni C J, Zhao G S, Wang X F, *et al*. ReconDreamer: crafting world models for driving scene reconstruction via online restoration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 1560-1569

[73] Keetha N, Karhade J, Jatavallabhula K M, *et al*. SplaTAM: splat, track & map 3D Gaussians for dense RGB-D slam[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 21357-21366

[74] Chen Y R, Gu C, Jiang J Z, *et al*. Periodic vibration Gaussian: dynamic urban scene reconstruction and real-time rendering[OL]. [2025-07-04]. https://arxiv.org/abs/2311.18561

[75] Kratimenos A, Lei J H, Daniilidis K. DynMF: neural motion factorization for real-time dynamic view synthesis with 3D Gaussian splatting[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2024: 252-269

[76] Huang Y H, Sun Y T, Yang Z Y, *et al*. SC-GS: sparse-controlled Gaussian splatting for editable dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 4220-4230

[77] Sumner R W, Schmid J, Pauly M. Embedded deformation for shape manipulation[C] //Proceedings of the ACM SIGGRAPH Conference. New York: ACM Press, 2007: 80-86

[78] Wan D W, Lu R J, Zeng G. Superpoint Gaussian splatting for real-time high-fidelity dynamic scene reconstruction[C] //Proceedings of the 41st International Conference on Machine Learning. Vienna: PMLR, 2024: 49957-49972

[79] Waczynska J, Borycki P, Kaleta J, *et al*. D-MiSo: editing dynamic 3D scenes using multi-Gaussians soup[C] //Proceedings of the 38th International Conference on Neural Information Processing Systems. New York: ACM Press, 2024: Article No.3425

[80] Shin I, Yu Q H, Shen X H, *et al*. Enhancing temporal consistency in video editing by reconstructing videos with 3D Gaussian splatting[OL]. [2025-07-04]. https://arxiv.org/pdf/2406.02541

[81] Park J, Bui M Q V, Bello J L G, *et al*. SplineGS: robust motion-adaptive spline for real-time dynamic 3D Gaussians from monocular video[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 26866-26875

[82] Ahlberg J H, Nilson E N, Walsh J L. The theory of splines and their applications: mathematics in science and engineering: a series of monographs and textbooks[M]. Amsterdam: Elsevier, 2016

[83] De Boor C. A practical guide to splines[M]. New York: Springer, 1978

[84] Yang Z Y, Yang H Y, Pan Z J, *et al*. Real-time photorealistic dynamic scene representation and rendering with 4D Gaussian splatting[C] //Proceedings of the 12th International Conference on Learning Representations. Vienna: Curran Associates, 2024: 1934-1951

[85] Xie E Z, Wang W H, Yu Z D, *et al*. SegFormer: simple and efficient design for semantic segmentation with transformers[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: ACM Press, 2021: Article No.924

[86] Li Z, Chen Z, Li Z, *et al*. Spacetime Gaussian feature splatting for real-time dynamic view synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 8508-8520

[87] Chen Z, Li Z, Song L C, *et al*. NeuRBF: a neural fields repre-

sentation with adaptive radial basis functions[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 4159-4171

[88] Fang L J, Jiang Q H, Shi J P, *et al*. TPNet: trajectory proposal network for motion prediction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 6796-6805

[89] Houenou A, Bonnifait P, Cherfaoui V, *et al*. Vehicle trajectory prediction based on motion model and maneuver recognition[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2013: 4363-4369

[90] Duan Y X, Wei F Y, Dai Q Y, *et al*. 4D-rotor Gaussian splatting: towards efficient novel view synthesis for dynamic scenes[C] //Proceedings of the ACM SIGGRAPH Conference. New York: ACM Press, 2024: Article No.87

[91] Luiten J, Kopanas G, Leibe B, *et al*. Dynamic 3D Gaussians: tracking by persistent dynamic view synthesis[C] //Proceedings of the International Conference on 3D Vision. Los Alamitos: IEEE Computer Society Press, 2024: 800-809

[92] Sun J K, Jiao H, Li G Y, *et al*. 3DGStream: on-the-fly training of 3D Gaussians for efficient streaming of photo-realistic free-viewpoint videos[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 20675-20685

[93] Jiang Y H, Shen Z H, Hong Y, *et al*. Robust dual Gaussian splatting for immersive human-centric volumetric videos[J]. ACM Transactions on Graphics, 2024, 43(6): Article No.265

[94] Wang Z Y, Qin Y R, Zeng L, *et al*. High-dynamic radar sequence prediction for weather nowcasting using spatiotemporal coherent Gaussian representation[C] //Proceedings of the 13th International Conference on Learning Representations. Singapore: Curran Associates, 2025: 25879-25898

[95] LeCun Y, Chopra S, Hadsell R, *et al*. A tutorial on energy-based learning[OL]. [2025-07-04]. https://www.researchgate.net/profile/Marcaurelio-Ranzato/publication/216792742_A_Tutorial_on_Energy-Based_Learning/links/0912f50c6862425435000000/A-Tutorial-on-Energy-Based-Learning.pdf

[96] Yan J B, Peng R, Wang Z Y, *et al*. Instant Gaussian stream: Fast and generalizable streaming of dynamic scene reconstruction via Gaussian splatting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2025: 16520-16531

[97] Wang S Z, Yang X Y, Shen Q H, *et al*. GFLow: recovering 4D world from monocular video[C] //Proceedings of the 39th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2025: 7862-7870

[98] Seidenschwarz J, Zhou Q J, Duisterhof B P, *et al*. DynOMo: online point tracking by dynamic online monocular Gaussian reconstruction[C] //Proceedings of the International Conference on 3D Vision. Los Alamitos: IEEE Computer Society Press, 2025: 1012-1021

[99] Yan Z W, Li C, Lee G H. NeRF-DS: neural radiance fields for dynamic specular objects[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 8285-8295

[100] Sun P, Kretzschmar H, Dotiwalla X, *et al*. Scalability in perception for autonomous driving: Waymo open dataset[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2443-2451

[101] Broxton M, Flynn J, Overbeck R, *et al*. Immersive light field video with a layered mesh representation[J]. ACM Transactions on Graphics, 2020, 39(4): Article No.86

[102] Li L Z, Shen Z, Wang Z S, *et al*. Streaming radiance fields for 3D video synthesis[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM Press, 2022: Article No.980

[103] Jiang Y H, Shen Z H, Wang P H, *et al*. HiFi4G: high-fidelity human performance rendering via compact Gaussian splatting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 19734-19745

[104] Abou-Chakra J, Dayoub F, Sünderhauf N. ParticleNeRF: a particle-based encoding for online neural radiance fields[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2024: 5963-5972

[105] Schonberger J L, Frahm J M. Structure-from-motion revisited[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 4104-4113

[106] Wang Z, Bovik A C, Sheikh H R, *et al*. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612

[107] Zhang R, Isola P, Efros A A, *et al*. The unreasonable effectiveness of deep features as a perceptual metric[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 586-595

[108] Rassool R. VMAF reproducibility: validating a perceptual practical video quality metric[C] //Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Los Alamitos: IEEE Computer Society Press, 2017: 1-2

[109] Heusel M, Ramsauer H, Unterthiner T, *et al*. GANs trained by a two time-scale update rule converge to a local nash equilibrium[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 6629-6640

[110] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C] //Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers. Los Alamitos: IEEE Computer Society Press, 2003: 1398-1402

[111] Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144

[112] Overton S R. LODs and Nanite within unreal engine 5: the future of 3D asset creation for game engines[EB/OL]. [2025-07-04]. https://dc.etsu.edu/digitalmedia-culminating-experience/6

[113] Hyun S, Heo J P. Adversarial generation of hierarchical Gaussians for 3D generative model[OL]. [2025-07-04]. https://arxiv.org/abs/2406.02968v1