

顾及朗伯表面精确重建的2D-3D混合高斯模型

孙传凤¹⁾ 吴 军¹⁾ 徐 刚¹⁾ 王玲容²⁾ 罗 瀛¹⁾ 龙煜林¹⁾

¹⁾(桂林电子科技大学电子工程与自动化学院 广西 桂林 541004)

²⁾(广西交科集团有限公司 南宁 530007)

摘 要 针对现有3DGS方法在朗伯平面区域重建过程中的几何不连续性、结构无序性及深度信息缺失问题,提出一种顾及朗伯表面精确重建的2D-3D混合高斯溅射(LPGS)模型,主要涉及三方面内容:(a)朗伯平面跨源深度推理与混合基元初始化。结合单目深度网络估计与点云投影并通过分段统计映射精准重建朗伯区域深度场,进而将场景划分为朗伯平面区和非朗伯区并分别初始化为贴合平面的2D圆盘基元与表征复杂结构的3D椭球基元。(b)平面几何双正则化约束。为克服朗伯平面重建的结构无序性,通过最小化圆盘法向与平面法向的夹角误差的法向量一致性正则化处理确保基元朝向正确,通过强制圆盘中心贴合目标平面的距离贴合正则化处理确保基元位置精确。(c)渐进式基元融合与混合优化渲染。通过动态匹配代价函数将邻近符合条件的3D椭球同化为2D圆盘,以有效维持优化过程中朗伯平面区域的基元数量与几何连续性,在继承3DGS高效光栅化渲染框架基础上优化混合基元参数的同时,实现平面区域高精度重建与非平面区高质量渲染的协同提升。在Mip-NeRF360、Tanks & Temples、Replica和DTU等数据集上的实验结果表明,LPGS在保持高保真渲染效果(如Mip-NeRF360室内场景PSNR达31.30)的同时,显著提升朗伯平面几何重建效果(完整性评分IS较3DGS提升19.7%,平面贴合距离误差D_RMSE显著降低),有效提升复杂场景下的几何重建完整性与渲染质量。

关键词 朗伯平面重建;跨源深度推理;3DGS;法向量正则化;距离正则化

中图分类号 TP391 **DOI号** 10.11897/SP.J.1016.2026.01845

High-Quality Reconstruction of Lambertian Planes with 2D-3D Hybrid Gaussian Splatting (LPGS)

SUN Chuan-Feng¹⁾ WU Jun¹⁾ XU Gang¹⁾ WANG Ling-Rong²⁾ LUO Ying¹⁾ LONG Yu-Lin¹⁾

¹⁾(School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004)

²⁾(Guangxi Transportation Science and Technology Group Co., Ltd., Nanning 530007)

Abstract To address the problems of geometric discontinuity, structural disorder, and insufficient depth information in Lambertian planar region reconstruction encountered by existing 3D Gaussian Splatting (3DGS) methods, this paper proposes a Lambertian Plane-Guided 2D-3D hybrid Gaussian splatting model, termed LPGS, for accurate reconstruction of Lambertian planes. The proposed method is designed to improve both geometric completeness and rendering quality in complex scenes by introducing a hybrid primitive representation tailored to the structural characteristics of different regions. Specifically, the method consists of three closely related components, including cross-source depth inference with hybrid primitive initialization, dual regularization constraints for planar geometry, and progressive primitive fusion with hybrid optimization and rendering. First, a cross-source depth inference strategy and a hybrid primitive

收稿日期:2025-09-24;在线发布日期:2026-04-17。本课题得到国家自然科学基金项目(42361071)、2025年广西重点研发计划(桂科FN2504240020)、广西研究生教育创新计划项目(YCBZ2023136、YCBZ2024165)资助。孙传凤,博士研究生,主要研究方向为深度学习、三维重建。E-mail: 1691457253@qq.com。吴 军(通信作者),博士,教授,博士生导师,主要研究方向为摄影测量与遥感、计算机视觉与人工智能。E-mail: wujun@guet.edu.cn。徐 刚,博士研究生,主要研究方向为计算机视觉。王玲容,硕士,主要研究方向为智能交通。罗 瀛,博士研究生,主要研究方向为计算机视觉、图像处理。龙煜林,硕士研究生,主要研究方向为三维重建。

initialization scheme are developed for Lambertian planar regions. By integrating monocular depth estimation with projected depths derived from point clouds, the proposed method introduces a piecewise statistical mapping mechanism to reconstruct the depth field of Lambertian planar areas more accurately. This design helps alleviate the depth ambiguity and incompleteness that commonly appear when relying on a single depth source alone. Based on the reconstructed depth information, the scene is divided into Lambertian planar regions and non-Lambertian regions according to their geometric characteristics. These two types of regions are then initialized with different primitives, namely plane-aligned 2D disk primitives for Lambertian planes and 3D ellipsoidal primitives for structurally complex non-planar regions. Such an initialization strategy allows the representation to better match the underlying scene geometry, thereby providing a more suitable basis for subsequent optimization and rendering. Second, dual regularization constraints are introduced to strengthen planar geometric consistency during optimization. To alleviate the structural disorder that commonly occurs in Lambertian planar reconstruction, a normal consistency regularization is designed by minimizing the angular error between each disk normal and the corresponding plane normal. This regularization encourages the 2D disk primitives to maintain correct orientation with respect to the target plane and thus improves the local structural coherence of reconstructed planar regions. In addition, a distance-to-plane regularization is imposed to constrain the centers of the disk primitives to remain close to the target plane. This positional constraint further improves the geometric accuracy of planar reconstruction by reducing deviations of disk primitives from the expected planar structure. Through the joint action of these two regularization terms, the proposed method effectively enhances the structural integrity and geometric reliability of Lambertian plane reconstruction. Third, a progressive primitive fusion strategy together with a hybrid optimization and rendering scheme is proposed to improve geometric continuity while maintaining rendering efficiency. Specifically, a dynamic matching cost function is constructed to progressively assimilate neighboring qualified 3D ellipsoidal primitives into 2D disk primitives during the optimization process. This strategy effectively maintains the number of primitives and enhances geometric continuity in Lambertian planar regions throughout optimization. Meanwhile, built upon the efficient rasterization framework of 3DGS, the proposed method jointly optimizes the parameters of the hybrid primitives, enabling high-precision planar reconstruction and high-quality rendering in non-planar regions within a unified representation. Experimental results on Mip-NeRF360, Tanks & Temples, Replica, and DTU demonstrate the effectiveness of LPGA. While maintaining high-fidelity rendering quality, LPGA achieves a PSNR of 31.30 on indoor scenes of Mip-NeRF360 and significantly improves the geometric reconstruction quality of Lambertian planes. In particular, the Integrity Score (IS) is improved by 19.7% compared with 3DGS, and the plane-fitting distance error D_{RMSE} is markedly reduced. These results confirm that LPGA effectively improves both geometric reconstruction completeness and rendering quality in complex scenes.

Keywords Lambertian plane reconstruction; cross-source depth inference; 3D Gaussian Splatting; normal consistency regularization; distance-to-plane regularization

1 引 言

近年来,神经辐射场(Neural Radiance Fields,

NeRF)^[1]和高斯溅射(3D Gaussian Splatting, 3DGS)^[2]在影像多视图合成^[3-9]、场景几何重建^[10-14]等3D视觉任务方面显示出卓越性能,掀起了计算机图形学领域的一场革命性变革。相比于NeRF隐式表示在实

际应用中面临的高计算成本、场景编辑困难且模型性能依赖于相机位姿精确标定的多视角图像训练数据,3DGS使用大量3D高斯椭球显式表达场景,在数学层面上避免了NeRF中沿光线进行采样所需的巨大计算开销,可实现实时或交互式渲染且易于场景编辑、动态调整,尤其适合处理动态对象或场景变化,对于需要从不同角度全面理解、呈现真实世界且存在高实时性响应要求的场景如机器人技术、自动驾驶^[15-17]等具有重要价值。

尽管3DGS训练、渲染效率显著优于NeRF且对多视角图像相机标定误差的鲁棒性更强,但3DGS渲染质量依赖于高斯分布的离散近似,复杂几何或高光区域易丢失细节,如几何重建方面,3DGS采用椭球基元表征场景几何结构并在图像空间逐像素优化约束椭球的位置与形变参数时,该优化机制受限于初始椭球形状且为追求渲染质量,往往驱动椭球基元优先适配颜色与位置变化,导致椭球表面(曲率)重建结果无法准确描述平面区域;与此同时,场景中的朗伯平面在图像中因缺乏纹理信息使得3DGS难以有效提取特征点及获得深度信息,即输入3DGS的初始点云在朗伯平面区域存在大量“空洞”,而3DGS模型优化过程中缺乏针对椭球移动方向的强约束则会使椭球基元在空洞区域存在结构无序性问题,从而导致椭球基元与实际场景表面的对应不够精确。

针对现有3DGS模型在场景朗伯平面重建过程中的椭球基元几何表示能力受限问题,本文提出了一种新的高斯溅射LPGS(Lambertian Plane-Gaussian Splitting)模型,该模型采用2D-3D混合基元表征框架并将场景划分为复杂纹理区域和朗伯平面区域,通过使用不同场景表达方式(椭球/圆盘)并联合渲染优化,可在渲染质量保持不变的情况下大幅提升场景几何重建精度。如图1所示,LPGS基本思想是将场景中的朗伯平面区域初始化为2D圆盘基元,其他区域初始化为3D椭球基元,并在2D-3D混合椭球基元优化过程中为空洞区域的2D圆盘基元添加法向约束与距离约束以使其准确贴合于场景中的朗伯平面,同时通过渐进式基元融合算法来保证空洞区域(朗伯平面区域)基元数量及几何重建连续性,从而实现场景几何结构的全域高精度重建。本文主要贡献如下:

(1)提出一种朗伯平面跨源深度推理方法:结合单目深度网络估计与点云投影并利用分段统计精准重建朗伯区域深度场,为2D圆盘基元初始化提供几何先验。

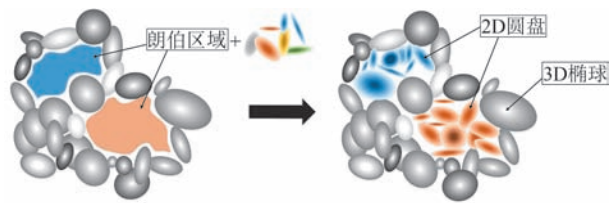


图1 3D椭球和2D椭圆联合优化(在保证复杂区域渲染质量的同时提高朗伯平面的几何重建精度)

(2)提出一种渐进式基元融合策略:通过匹配代价函数动态将邻近3D椭球同化为2D圆盘,解决优化过程中圆盘数量骤减导致的几何断裂问题,从而维持平面重建连续性。

(3)提出平面几何双正则化约束以克服朗伯平面几何重建中的结构无序性:法向量一致性约束-通过最小化圆盘法向与平面法向的夹角误差引导高斯圆盘的方向;平面距离贴合约束-通过构建距离损失项并通过强制圆盘中心贴合目标平面引导高斯圆盘位置。

(4)提出2D-3D混合高斯表征框架:将场景划分为朗伯平面区域(2D圆盘基元)与非朗伯区域(3D椭球基元),并通过联合优化实现几何重建精度与渲染质量的协同提升。

2 相关工作

在三维场景重建和新视图合成领域中,NeRF和3DGS代表了两种截然不同的技术路径,前者以高保真渲染质量而闻名,后者则以快速训练和高效渲染速度而受到关注。这里对两种技术的研究进展分别予以回顾。

2.1 隐式三维重建 NeRF

NeRF的基本思想是利用多层感知机(Multilayer Perceptron, MLP)隐式建模场景的连续体积密度、颜色并通过体绘制技术投影生成新视图,优点在于其隐式表示避免了显式网格或点云的拓扑限制,利于获得高保真、细节丰富的渲染效果,尤其擅长复杂光照和透明材质,但MLP中过多的参数及逐像素光线步进处理策略会导致模型训练时间过长,即庞大的参数数量会带来巨大的训练开销。

为加速模型训练速度,Liu等人^[18]在NeRF的基础上定义了一组以稀疏体素八叉树为组织形式的体素边界隐式场,通过可扩散光线来逐步学习底层体素结构,并利用这种结构来跳过不包含相关场景内容的体素,从而加快渲染新视图的速度。在此基础

上, Yu 等人^[19]通过将原始 NeRF 预采样的结果保存到一个与视图相关的表格体(PlenOctrees)中,将场景中所有采样点的外观和密度信息都存储在一个稀疏体素八叉树的叶子节点中,并使用球谐函数来解算任意视角的颜色值以避免 MLP 的大量计算,大大缩短了场景的重建时间,但这种方法不能从头开始训练,需要从训练好的隐式模型中进行转换。为此, Yu 等人^[20]进一步提出了 Plenoxels 模型,使用球谐函数对稀疏体素表示进行端到端的优化,并使用线性插值的方式来使整个场景连续化以避免神经网络的大量计算,最终将模型的训练速度加速至 15 分钟之内。借鉴该思路, Sun 等人^[21]使用密集的体素网格来直接模拟 3D 几何,在体素网格初始化时,将密度设置为极小值,并降低采样较少体素的学习率来避免出现次优几何解,再次加速模型训练。

为提高模型渲染质量, Barron 等人^[22]提出 Mip-NeRF 并使用圆锥体取代光线的方案来提高采样率,通过对三维位置及其周围的高斯区域进行集成位置编码并在频率周期小于被集成的频率周期时,集成位置编码的高频维度将向零收缩,从而具有更好的抗锯齿性能,不足之处在于无法准确捕获和再现物体光滑表面的外观。在此基础上, Verbin 等人^[23]提出了 Ref-NeRF,以视图向量对局部法向量的反射作为输入来替代视图向量本身并引入集成定向编码技术,将外向的亮度结构变成明确的漫射和高光组件以保持反射亮度平滑并有效捕获场景中材料和纹理的变化,从而再次提升模型渲染质量。

为减小模型体积, A Chen 等人^[24]提出 TensorRF,将场景建模为 4D 张量并分解为多组紧凑的低秩分量,在保证渲染质量的前提下大幅减少模型体积,但受限于简单的正交变换。在此基础上, A Chen 等人^[25]进一步提出 DiF-Grid 并使用具有周期变换的基场因子模拟信号域中共性模式、使用系数场因子表达信号域中局部的空间变化特征,然后利用可学习的投影函数从因子乘积中回归目标信号并通过联合建模双因子实现更紧凑的重建。受体素网格启发, T Müller 等人^[26]提出 Instant-NGP 并引入多分辨率哈希编码,通过使用小型网络显著减少浮点数的计算和内存访问,从而在不降低渲染质量前提下实现更小的带宽浪费和更高效的计算,但高质量重建和实时渲染仍然是一个重大的挑战。

2.2 显式三维重建 3DGS

与 NeRF 隐式建模方式不同, 3DGS 无须将场景所有外观、形状、透明度等信息存储在神经网络

中,而是通过直接优化 3D 高斯椭球几何(特征位置、尺度和旋转)、外观特征参数显式表达场景,并利用光栅化技术将这些各向异性的 3D 高斯椭球投影到 2D 屏幕上获得新视图,从而在数学层面上打破了 NeRF 光线采样对于渲染速度的限制,实现了训练时间和渲染速度上的显著提升。然而, 3DGS 在处理高频细节区域和场景边缘特征时会产生伪影和细节丢失,针对这一问题, Yu 等^[27]研究者引入多级渐进式高斯表示并提出 Mip-Splatting 方法,该方法通过在多个分辨率层次上表示高斯椭球并在渲染过程中允许根据不同场景区域动态选择最优分辨率,可在保持渲染效率的同时提升渲染质量,但其高斯致密化过程会出现图像中高方差区域被几个大高斯椭球同时覆盖的过度重建问题,从而产生伪影和模糊情况。为此, Zhang 等人^[28]提出了 FreGS,并利用从低到高的频率分量来执行从粗到细的高斯致密化、通过最小化渲染图像频谱与相应的地面实况间的差异实现高质量的高斯致密化,从而有效缓解高斯泼溅的过度重建。需要指出的是,无论是 Mip-Splatting 还是 FreGS 都高度依赖 SfM^[29]技术初始化点云,故无纹理区域表面重建面临挑战,据此 Cheng 等人^[30]提出 GaussianPro,通过渐进传播策略来指导 3D 高斯的致密化过程,并结合场景重建几何先验知识与匹配技术产生具有精确位置和方向的新高斯椭球。类似的, Jung 等人^[31]提出 RAIN-GS 并通过渐进高斯低通滤波控制和自适应边界扩展分裂策略,减少模型训练对 3D 高斯椭球精确初始化结果的依赖,从而显著提升模型鲁棒性和性能。

为提高 3DGS 在复杂场景中的表面重建质量, Guédon 等人^[32]提出 SuGaR 并将高斯函数绑定到网格表面并在光栅化时进行联合优化,从而使高斯椭球紧贴于场景表面以提高几何重建精度,同时能从 3D 高斯椭球中精确、快速提取网格几何特征。可以简单认为, SuGaR 是将 3D 高斯椭球近似为高斯圆盘,据此 HUANG 等人^[33]提出 2DGS,直接将 3D 椭球压缩为 2D 圆盘并结合深度失真和正态一致性损失对表面进行建模,可精确恢复薄表面并提供视图一致的几何形状,从而大幅提升场景表面重建精度。受 SuGaR 和 2DGS 启发, Dai 等人^[34]结合 3DGS 的灵活优化过程提出一种体积切割方法以聚合高斯面元信息并去除 Alpha 混合生成深度图中的错误点,进而对深度图信息进行泊松重建以提取表面网格,最后借助准确的深度图再次提升表面重建精

度。尽管如此,上述方法中仍然存在表面网格提取质量不高等问题。为此,Chen 等人^[35]提出 PGSR,通过点云渲染生成相机原点到高斯平面的距离及对应的法线图并将两者相除获得无偏深度,同时引入单视图几何、多视图光度和几何正则化以保持全局几何精度,使表面重建质量和几何精度达到新的高度。

3 LPGS 模型

现有的 3DGS 模型及其变体模型中,多以椭球基元表征场景几何结构,而初始化高斯椭球基元的颜色/位置优先适配特性会通过牺牲椭球表面(曲率)与局部平面区域的贴合度来保证全局场景的渲染质量;另一方面,场景中的朗伯平面在图像中因缺乏纹理信息使得 3DGS 生成的初始点云存在大量“空洞”,而 3DGS 模型优化过程中缺乏针对椭球移动方向的强约束则会导致椭球基元与实际场景表面,尤其是朗伯平面区域的对应不够精确。针对现有 3DGS 模型在场景朗伯平面重建过程中的椭球基元几何表示能力受限问题,本文采用 2D-3D 高斯混合基元表征框架并将场景划分为复杂纹理区域和朗伯平面区域,其中:朗伯平面区域初始化为 2D 圆盘基元,其他区域初始化为 3D 椭球基元,两者通过联合渲染优化实现场景几何结构的全域高精度重建。

3.1 网络结构设计

为了解决现有 3DGS 模型在朗伯平面区域重建过程中的几何不连续性、结构无序性等问题,本文提

出 2D-3D 高斯基元混合表征下的 LPGS 框架,其总体结构如图 3 所示,由跨源深度推理、混合基元优化、渐进式基元融合、多正则化损失约束 4 个模块构成。整体重建流程如图 2 所示,具体步骤如下:

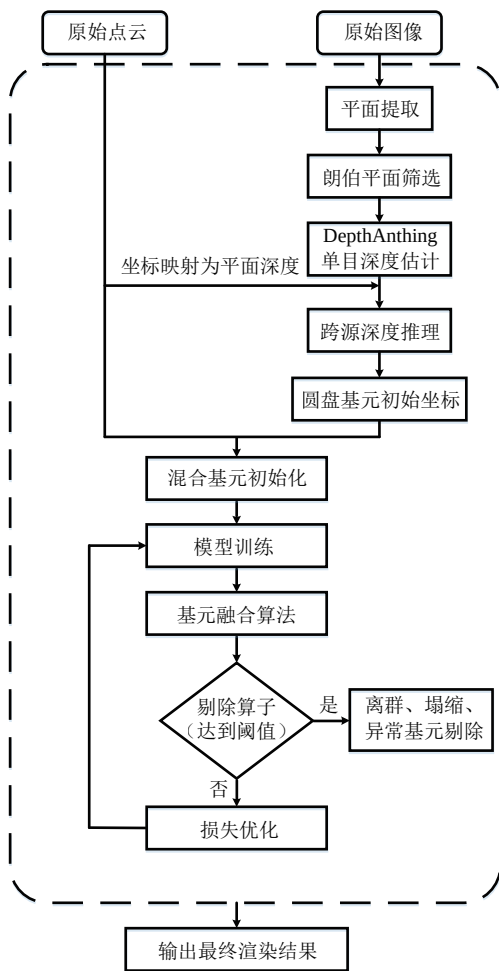


图2 LPGS 模型重建流程图

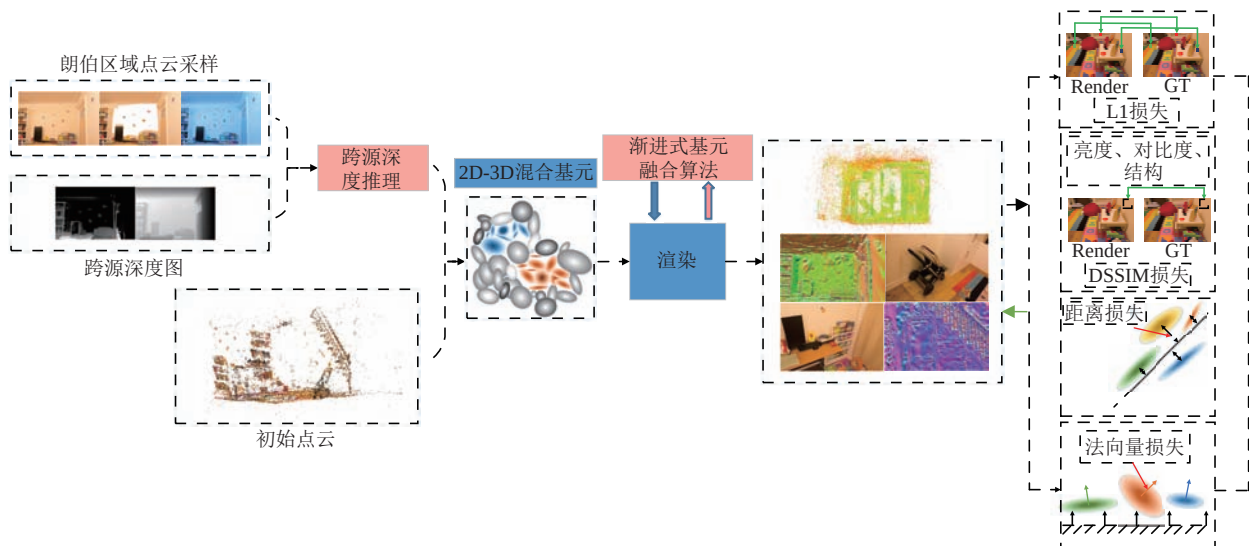


图3 LPGS 结构图

步骤1:对于输入图像信息,利用神经网络获取场景候选平面区域,得到平面概略图。

步骤2:结合 Sobel 算子量化平面概略图中像素梯度变化,过滤掉纹理丰富的区域,提取出最大连通域,得到精确的朗伯平面二值掩码。

步骤3:跨源深度推理中,通过将 DepthAnything 深度估计网络生成全局深度图和初始点云投影局部真实深度值进行校准融合,得到精准的朗伯平面深度值。

步骤4:精确定位朗伯平面上的采样点,将校准后的深度信息映射回点云空间,并据此初始化 2D 圆盘高斯基元。

步骤5:渐进式基元融合中,通过定义匹配代价函数,动态地将邻近的、符合条件的 3D 椭球基元同化为 2D 圆盘基元,提高朗伯平面区域基元数量的稳定性和几何连续性。

步骤6:对于 2D 圆盘基元使用两个额外的正则化损失约束:法向量正则化,利用 PCA 估计的朗伯平面法向量作为参考,约束圆盘基元的法向量与之对齐;平面距离正则化:基于估计的朗伯平面方程,约束圆盘基元的中心点贴合目标平面表面。从而避免朗伯区域内圆盘基元出现结构无序性(如方向偏离平面、位置偏移)。

3.2 朗伯平面影像检测与跨源深度推理

朗伯平面作为理想漫反射表面,其低纹理特性导致特征点稀疏,而均匀反射特性导致灰度变化(梯度)弱,使得从图像中准确检测朗伯平面区域较为困难。针对这一挑战,本文的基本思路是通过“粗筛-精炼-定位”三阶段实现朗伯平面区域的可靠检测,并通过跨源深度推理解决朗伯平面区域深度信息缺失问题,其整体流程图如图 4 所示,具体步骤如下:

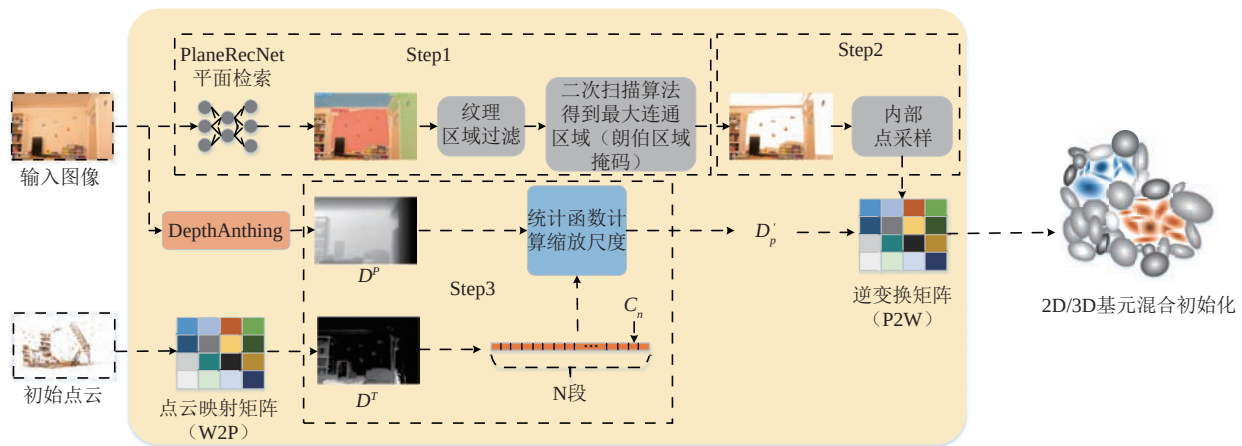


图4 朗伯平面影像检测与跨源深度推理流程图

步骤1:朗伯平面区域“粗-精”影像检测

旨在通过从“粗”到“精”的层级式检测框架从 2D 图像中准确获得朗伯区域,见图 5,具体步骤如下:首先利用 PlaneRecNet 网络^[36]提取场景特征并逐像素预测平面区域来得到整个场景的概略平面图,见图 5(b);其次,利用 Sobel 算子计算概略平面

图中所有像素的梯度幅值并基于给定的梯度阈值过滤纹理区域,进而从场景概略平面图中得到候选的朗伯平面区域,如式(1)和式(2)所示:

$$G = \sqrt{G_x^2 + G_y^2}, \quad \text{where, } G_x = I * K_x, \quad G_y = I * K_y \quad (1)$$

$$M_{\text{updated}} = \begin{cases} M, & \text{if } G < \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (2)$$



图5 朗伯平面检索效果图

其中, I 为原图像, K_x 和 K_y 分别是 Sobel 算子在水平和垂直方向上的核, $*$ 表示卷积操作, ϵ 为梯度阈值, M 为候选像素区域。最后, 采用二次扫描算法 (Two-Pass Algorithm) 对每个候选的朗伯平面内的连通区域进行识别与提取, 从中筛选出面积最大的连通区域并剔除其他小连通区域, 从而准确获得场景中的朗伯平面区域 (掩码图), 见图 5(c)。

步骤2: 朗伯平面区域采样

旨在从朗伯平面区域中获得用于初始化 2D 圆盘中心位置的采样点。本文首先利用随机采样策略获得朗伯平面掩码图中的内点集合 M , 考虑到当采样点位于朗伯平面区域边缘时, 边缘点的高斯椭圆包裹效应会使圆盘基元偏离真实平面而导致几何失真, 这里剔除平面中距离边缘点较近的点以确保后续优化的稳健性、准确性。令 N 为利用 Sobel 算子检测得到的朗伯平面的边缘点集合, $m(x_m, y_m)$ 和 $n(x_n, y_n)$ 分别表示 M 和 N 中任一点, 从 M 中剔除距离边缘点较近的点的计算过程如下:

$$D(m) = \min_{m \in M, n \in N} \|m - n\|_2 \quad (3)$$

$$\text{where, } \|m - n\|_2 = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}$$

其中, $D(m)$ 表示 M 中任一点 m 到其最近边缘的距离。对 D 中距离值从小到大排序并剔除前 60% 距离对应的点以获得候选点集 C , 在 C 中随机无重复地选取 N 个点即为最终的朗伯平面区域采样点集 S , 本文朗伯平面区域采样流程 (伪码) 如 Algorithm 所示:

Algorithm: Mapping algorithm for the initialisation of ‘hollow’ point clouds

I, L_masks : The original images, Lambertian pixel masks

$I_P, L_masks \leftarrow \text{Plane RecNet}(I)$ ▷ Positions

$G_I \leftarrow \text{greyscale}(I_P)$ ▷ Grayscale processing

$\text{Gard_Images} \leftarrow \text{calculate_gradient}(G_I)$

for all L **maskes in** I_P **do**

if $\text{Gard_Images.vaule} > \epsilon$ **then** ▷ Rejection of noise

$\text{Remove_masks}()$

end if

$\text{slices, nums} \leftarrow \text{Find_connected_area}(L_masks)$

▷ Two-Pass Algorithm

if $\text{slices.size} \leq 100 \times 100$ **then** ▷ Rejection of noise

$\text{Remove_masks}()$

end if

if $\text{nums} \leq 1$ **then**

$\text{Points} \leftarrow \text{Find_sample_in_slices}(\text{slices})$

▷ Internal Point Sampling Algorithm

else ▷ Finding the largest connected area

$L_slices \leftarrow \text{Finding_largest_connected_area}(\text{slices})$

$\text{Points} \leftarrow \text{Find_sample_in_slices}(L_slices)$

end if

end for

步骤3: 朗伯平面区域跨源深度推理

旨在获取朗伯平面区域采样点的深度信息。首先采用单目深度估计网络^[37]获得朗伯平面区域的相对深度图 D^P , 考虑到 D^P 自身误差及与实际深度间的空间基准差异, 这里以朗伯平面区域及其附近的稀疏点云为控制信息对 D^P 进行几何纠正, 具体步骤如下:

首先, 利用从世界坐标系到相机坐标系的变换矩阵 W_{w2p} 将初始点云投影到朗伯平面区域所在图像得到真实深度图 D^T , 由于平面特征点稀疏, 该 D^T 中存在缺乏有效深度信息的大量像素“空洞”;

其次, 对 D^T 中深度从小到大排序并划分为 N 段以分段计算尺度 S_n 。令 C_n 为第 n 段的特征点集合, $n \in \{1, 2, \dots, N\}$, 则对于每个深度段的 C_n , 尺度 S_n 计算如下:

$$S_n = f\left(\frac{D_n^T}{D_n^P}\right) \quad (4)$$

其中, $f()$ 表示 C_n 中特征点深度值的统计函数, 这里以 C_n 内对应特征点的深度比值 D_n^T/D_n^P 的中值来计算尺度 S_n 。

然后, 利用尺度 S_n 对朗伯平面区域采样点在 D^P 中的深度信息进行校准。令像素 q 为朗伯平面区域任意采样点, D_q 为该采样点在 D^P 中的深度值, D'_q 为校准之后的深度值, 即像素 q 位于深度段 C_n , 则有

$$D'_q = D_q \times S_n \quad (5)$$

最后, 利用 W_{w2p} 的逆变换矩阵及校准深度 D'_q 计算获得朗伯平面区域采样点 q 的真实点云信息, 从而为 2D 圆盘初始化及优化奠定基础。

需要强调的是, 上述平面检测模块给出的仅是一种“候选朗伯平面”的软划分, 而非对整个场景强制施加理想平面假设。其中小范围的局部平坦纹理块或具有明显弧度的表面往往难以同时满足“低梯度+大面积”的要求, 因此通常不会被标记为朗伯平面区域。这类区域在后续优化中仍由 3D 高斯椭圆基元表示, 来保证方法在非理想平面上的安全退化与泛化能力, 只有在确实存在尺度上足够大且局部可近似为单一主平面的区域时, 2D 圆盘分支才会被激活并施加额外的几何正则约束。

从反射模型的角度来看, 还需强调的是, 本文中使用的“朗伯平面”概念并非严格意义上的物理朗伯体, 而是从重建任务出发的一种工程化近似。按照经典定义, 朗伯体应当在各个观察方向上满足辐射亮度服从余弦定律, 原则上需要依赖多视角下的反射

分布来判定其是否为理想漫反射表面,仅凭单视角纹理特征显然是不充分的。因此,本文在第3.2小节中通过 PlaneRecNet 结合梯度幅值与连通域面积,仅从单视角图像中筛选出“大尺度、局部近似平坦且纹理相对均匀”的区域,作为朗伯候选平面,并在后续 2D-3D 混合优化中将其作为几何软先验而非刚性约束:一方面,这些标签仅用于初始化 2D 圆盘及施加法向/距离正则,并不会强制整块区域永久保持 2D 表达;另一方面,多视角光度一致性和几何损失会在训练过程中持续“审查”这些候选平面,对于存在明显视角依赖或镜面高光的区域,2D 圆盘难以同时兼顾光度与几何拟合,其不透明度会在渐进式基元融合与裁剪机制中逐步下降,对应结构最终由 3D 椭球主导表达。第4章关于平面先验噪声与消融实验的结果也表明,即便存在一定程度的平面误标,模型在该区域的最坏情况会平滑退化为接近 3DGS 的几何质量,而不会产生大面积明显伪影,从而在实践中实现了对“近似朗伯/非朗伯混合场景”的可容错建模。

3.3 顾及朗伯平面区域的 2D-3D 高斯混合优化

由于本文场景中的朗伯平面区域、非朗伯平面区域分别采用 2D 高斯圆盘、3D 高斯椭球进行表示,这里分别进行初始化,其中 3D 高斯椭球基元初始化仍遵循传统 3DGS 模型方法,关键在于 2D 圆盘基元的初始化。本文 2D 圆盘基元的中心点位置采用初始点云位置,考虑到 3.2 小节中映射得到的平面点为随机颜色,为加快模型收敛速度,这里将 2D 圆盘基元的初始颜色设置为与其最近的 3D 椭球基元一致, z 轴方向为圆盘基元的法向量方向,并将所有平面上椭球基元的 z 轴尺度因子初始化为一个极小值。

需要指出的是,本文使用 3D 椭球基元的透明度来初始化 2D 圆盘基元,但对于场景中复杂纹理区域,该种方式下高斯圆盘无法保持各向异性,可能导致高斯圆盘在优化过程中发生显著位置偏移,从而引发“坍缩现象”,即当初始化时对颜色参数优化的学习率低于透明度参数优化的学习率时,模型对于基元颜色的调整会慢于透明度的调整,从而导致模型倾向于将二维圆盘“坍缩”为一维线条,见图 6。为解决这一问题,本文调整 2D 圆盘基元的初始颜色与透明度参数的学习率并将透明度初始化为 0.99,以使 2D 圆盘基元能均匀覆盖朗伯区域,从而避免圆盘多次混叠。图 7 展示了混合基元在场景中的初始空间构型,其中,图 7(a)为通过三维渲染呈现的初始 3D-2D 基元的异构分布,图 7(b)则投影显示混合基元中心点云的空间拓扑。从图 7 可清晰观察到两类

基元的差异化特性:在复杂纹理区域,初始化的 3D 椭球表现为尺寸各异的半透明球体($\alpha=0.1$),其各向异性协方差结构通过 Alpha 混合机制高效重建高频细节;而在朗伯平面区域,由于其颜色和几何性质单一,为此初始化的 2D 圆盘表现为尺寸统一的高不透明圆盘($\alpha=0.99$),其连续覆盖特性消除了初始点云的空洞缺陷,为后续均匀覆盖朗伯平面区域提供前置基础。

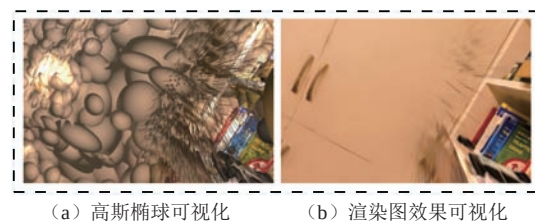


图6 “坍缩现象”实验结果图

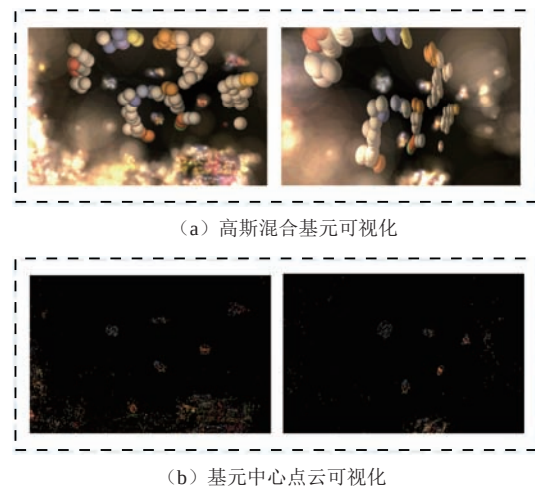


图7 混合 2D-3D 基元初始化图

在模型混合优化初期,由于圆盘基元初始化方向与朗伯区域平面法向并非一致,传统 3DGS 模型的自适应密度控制过程会将周围的 3D 椭球基元快速移动至平面来填充空缺,导致朗伯平面周围出现大量各向异性椭球,不仅破坏平面几何重建的连续性,还会因将 2D 圆盘视为噪声点而导致朗伯平面上的圆盘基元数量骤降。为解决这一问题,本文提出渐进式基元融合算法来增加 2D 圆盘数量并稳定朗伯平面几何重建连续性,其基本思想是通过建立椭球基元与参考圆盘基元的动态匹配机制实现几何属性的渐进式对齐,其原理如图 8 所示,总体上分为两个阶段:

同化点选取:旨在识别场景中适合被同化为当前朗伯平面上 2D 圆盘的邻近 3D 椭球基元,关键在于构建匹配代价函数来量化目标圆盘 p_i 与候选椭球

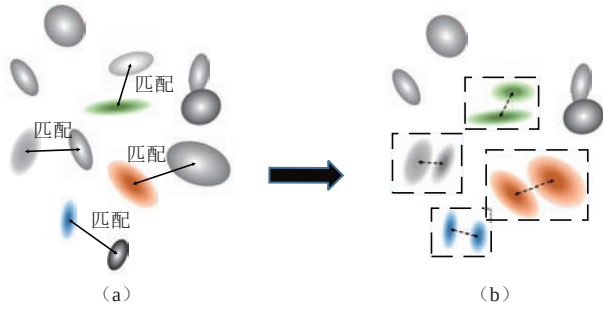


图8 渐进式基元融合算法示意图

e_j 之间在几何属性上的契合度。场景中所有高斯基元(包括2D圆盘 P_d 和3D椭球 P_e)均由其位置向量 $x \in \mathbb{R}^3$; 旋转矩阵 $R \in SO(3)$ ($SO(3)$ 为三维旋转群); 缩放向量 $s \in \mathbb{R}^3$; 不透明度 $\alpha \in [0, 1]$; 球谐系数 $c \in \mathbb{R}^k$ (k 为球谐函数的阶数)定义。特别地,朗伯平面区域上的2D圆盘基元,其缩放向量的 z 轴分量被初始化为并优化为一个极小值(近似于二维平面分布)。基于此,本文将场景基元明确划分为两个点集:

$$\begin{cases} P_d = \{p_i | L(p_i) \neq 0\} \\ P_e = \{e_j | L(e_j) = 0\} \end{cases} \quad (6)$$

其中, $L(\cdot)$ 为标签函数,等于0表示该点为椭球,非0则为圆盘。对于 $L(\cdot)$ 非0的点,本文设计了椭球 e_j 与圆盘 p_i 的匹配代价函数来进行同化点的选取。

$$D(p_i, e_j) = w_1 \|x_i - x_j\|_2 + w_2 |\theta_{ij}| + w_3 |d_i - d_j| \quad (7)$$

其中, $\|x_i - x_j\|_2$ 为空间距离, θ_{ij} 为方向向量夹角, d_i 为点到平面的代数距离。 w 为可调节权重,分别为0.5, 0.3, 0.2。即通过融合椭球基元和圆盘基元之间的空间邻近性、方向一致性、平面贴合度来得到综合距离度量 $D(p_i, e_j)$ 。

几何同化:旨在将上一步筛选出的候选椭球 e_j 的几何属性逐步变换为目标圆盘 p_i 的属性,从而实现平滑融合。具体步骤如下:

渐进式位置融合:

$$x_{j,new} = \alpha x_i + (1 - \alpha) x_j \quad (8)$$

其中, x_i 和 x_j 为第 i 个和第 j 个椭球基元的位置; $\alpha \in [0, 1]$ 为插值系数,用来平衡原始椭球位置与目标圆盘位置,避免突变导致的拓扑失真。

方向对齐与缩放调整:

$$\begin{cases} R_{j,new} = R_i \\ s_{j,new} = [s_x, s_y, -11.5129] \end{cases} \quad (9)$$

其中, R_i 为第 i 个椭球的旋转系数; s_x, s_y 为椭球缩放系数,保持原值, z 轴方向采用固定缩放值,为确保缩放系数始终为正值,需使用指数函数进行尺度激

活,因此 z 轴固定为-11.5129可确保为极小值。

属性继承:

$$\begin{cases} L(e_j) = L(p_i) \\ \alpha_{j,new} = \max(\alpha_i, \alpha_j) \end{cases} \quad (10)$$

其中, $L(\cdot)$ 为基元类型标签; α_i 为第 i 个椭球的不透明度;属性继承确保基元类型标签和不透明度的同步更新。

上述每个轮次的圆盘同化处理会导致圆盘数量急速增加且产生大量噪声。为防止过度同化,本文在匹配代价函数中定义了剔除因子和同化因子,计算公式如下:

$$E(p_i) = \Phi[(\alpha_i \leq \alpha_{th}) \vee (\min(s_{x,i}, s_{y,i}) \leq s_{th})] \times \Phi[L(p_i) \neq 0] \quad (11)$$

$$\begin{cases} A_{dist}(e_j) = \Phi[d(e_j, \Pi_i) \leq d_{th}] \\ \text{where, } d(e_j, \Pi_i) = |n_i^T x_j + D_i| \\ A_{angle}(e_j) = \Phi[\cos \theta_{ij} \geq \cos \theta_{th}] \\ \text{where, } \cos \theta_{ij} = \frac{v_j^T n_i}{\|v_j\| \cdot \|n_i\|} \end{cases} \quad (12)$$

其中,式(11)为透明度、尺度剔除因子,即符合条件则剔除。式(12)为距离、角度同化因子,即符合条件才同化。 $\Phi[\cdot]$ 为指示函数,条件满足取1,否则取0; α_{th} 为透明度阈值,取0.1; s_{th} 为最小尺度阈值,取0.01。 $\prod_i$ 为第 i 个平面, d_{th} 为距离阈值,取1; $\cos \theta_{ij}$ 为椭球向量 v_j 和平面法向的夹角余弦, $\cos \theta_{th}$ 为方向阈值,取0.85。因此,本文渐进式基元融合算法的优化目标可表示为

$$\begin{aligned} & \max_{s \subseteq P_e} \sum_{e_j \in S} D(p_i, e_j)^{-1} \\ & s.t. \\ & |S| \leq N_{\max} \\ & s_z^{(j)} = \gamma, \gamma = -11.5129 \\ & \alpha_i \geq \alpha_{th} \\ & \min(s_{x,i}, s_{y,i}) \geq s_{th} \\ & d(e_j, \Pi_i) \leq d_{th} \\ & \cos \theta_{ij} \geq \cos \theta_{th} \end{aligned} \quad (13)$$

其中, $s \subseteq P_e$ 为待优化椭球集合, $N_{\max}$ 为当前圆盘总数。至此,通过渐进式基元融合算法可顾及朗伯平面区域的2D-3D高斯基元混合优化。

3.4 多正则化损失约束

传统3DGS模型基元优化过程中依赖图像重建损失(如RGB光度或D-SSIM损失)来引导基元的位置、形状以及颜色变化。然而,对于朗伯平面区域的2D圆盘基元来说,由于这些区域纹理单一、颜色变化平缓,图像重建损失提供的梯度信号较弱且易导致

重建过程中出现局部过拟合现象,从而产生圆盘基元的位置和朝向与实际朗伯表面不一致的问题(例如圆盘可能偏离平面或法向不平行)。考虑到本文通过朗伯平面深度尺度校准初始化圆盘基元时,可利用PCA分析估计法获得朗伯平面的法向量估计,进而可通过该法向量估计与朗伯平面上初始化点坐标建立朗伯平面的平面方程。为有效约束圆盘基元贴附于朗伯平面,本文在3DGS核心渲染损失的基础上,引入了两个显式的几何正则化约束以引导朗伯平面区域的2D圆盘基元精确贴合目标平面,分别为:

法向量正则化:旨在最小化所有2D圆盘与朗伯平面法向量之间的角度误差以确保圆盘和平面之间的方向一致性,该正则项可以描述为

$$L_{norm} = \frac{1}{M} \frac{1}{2N_j} \sum_{i=1}^{N_j} (1 - (\hat{\mathbf{n}}_j \cdot \mathbf{n}_{ji})) \quad (14)$$

其中, M 为平面总数量, N_j 为第 j 个平面上的2D圆盘总数, $\hat{\mathbf{n}}_j$ 为第 j 个朗伯平面的法向量估计; $\mathbf{n}_{ji}$ 为第 j 个平面上第 i 个圆盘的法向量,2D圆盘初始化法向量垂直于 $x-y$ 平面,即为 $[0, 0, 1]^T$,故2D圆盘当前法向量可描述为: $\mathbf{n}_{ji} = R_{ji} \cdot [0, 0, 1]^T$, R_{ji} 为用于描述从一个2D圆盘的初始参考框架到其当前方向的 3×3 的旋转矩阵。

平面距离正则化:旨在最小化所有2D圆盘中心点坐标与朗伯平面之间的欧氏距离以确保圆盘和平面之间的几何一致性,该正则项可以描述为

$$L_{dist} = \frac{1}{M} \sum_{j=1}^M \frac{1}{2N_j} \sum_{i=1}^{N_j} \left(\frac{|A_j x_{ji} + B_j y_{ji} + C_j z_{ji} + D_j|}{\sqrt{A_j^2 + B_j^2 + C_j^2}} \right)^2 \quad (15)$$

其中, $A_j + B_j + C_j + D_j$ 定义了第 j 个朗伯面的方程

$A_j x + B_j y + C_j z + D_j = 0$, x, y, z 为空间中任意点的坐标, M 为场景中平面数量, N_j 为第 j 个平面上的2D圆盘总数。综上,本文模型的总体损失函数可定义为

$$L_{total} = (1 - \lambda - \beta - \gamma) L_1 + \lambda L_{DSSIM} + \beta L_{norm} + \gamma L_{dist} \quad (16)$$

其中, λ, β, γ 为损失权重占比超参数, λ 设置为0.2, β 和 γ 设置为0.15。 L_1 为像素级渲染损失, L_{DSSIM} 为结构相似性损失,均直接继承自3DGS,前者度量了颜色和亮度上的逐像素偏差,通过计算渲染图像上所有像素、颜色通道值与真实图像对应位置值的绝对差异,然后最小化所有差异的平均值来提高渲染图像质量。后者度量了渲染图像在结构、对比度和亮度感知层面上与真实图像的偏差,通过在多个局部窗口内计算两幅图像的亮度均值、对比度(标准差)以及结构(协方差)等统计量,并综合比较这些统计量的相似性来量化感知差异,从而驱动模型优化以提升视觉清晰度、边缘锐利度和纹理细节的保真度。 L_{norm} 为法向量正则化损失,度量了圆盘基元局部法向量与目标平面参考法向量在方向上的偏差,通过计算两者点积的负值(或直接计算角度差)并最小化该偏差,强制约束圆盘基元的朝向严格垂直于目标平面的宏观表面法向,确保其几何表示与平面结构对齐。 L_{dist} 为距离正则化损失,度量了圆盘基元中心点到目标平面几何表面的空间偏移偏差,通过计算该点到平面方程的有符号距离的绝对值(或平方)并最小化该偏差,强制约束圆盘基元紧密贴合在目标平面之上,消除不必要的悬浮或嵌入。所有损失函数的结构示意图如图9所示。

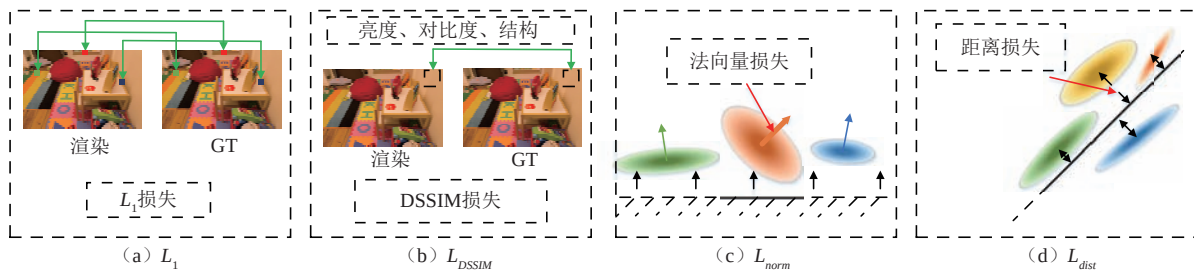


图9 模型损失函数设计图

4 实验与结果分析

本文实验平台配置如下:NVIDIA A100 40G GPU, Ubuntu 22.04, CUDA11.3, 深度学习框架Pytorch1.12。模型训练超参数设置如下:在2D高斯基元初始化阶

段直接赋予非零平面标签(label $\neq 0$),并将对应高斯的 z 轴尺度压缩到一个极小的值(10^{-3}),同时将不透明度初始化为0.99,朗伯平面点云映射算法中的平面采样点选取个数为100;训练早期(迭代次数 ≤ 6000)每200步执行一次“匹配代价函数同化”操作;在迭代 ≤ 15000 的阶段,每500步对所有label $\neq 0$ 的

圆盘进行一次透明度/尺度剔除：不透明度阈值为 $\alpha \leq 0.1$ ，尺度阈值为 $\min(s_x, s_y) \leq 0.01$ ，剩余圆盘则将不透明度 α 重置为 0.99；在 10 000-15 000 迭代区间，每隔 300 步对所有 label $\neq 0$ 的圆盘进行一次距离/角度剔除：距离阈值为 1.0m，角度余弦阈值为 0.85（约 32°）；场景级的密度梯度阈值 = $2e-4$ 、最小透明度 = 0.005 等保持与 3DGS 一致，用于控制整体高斯的增删，对于带平面标签的 2D 圆盘，固定 z 轴缩放并调低透明度的梯度。

超参数调节建议：在噪声较低、平面结构清晰的室内场景中，可以适当提高圆盘的透明度阈值（ $\alpha = 0.2$ ），只保留“非常显性”的薄片；在噪声较大的室外或 DTU 场景则建议维持 0.1，避免早期误删真实平面。尺度阈值 $\min(s_x, s_y) \geq 0.01$ 对应归一化尺度下约 1cm 的平面厚度，可根据场景单位比例在 0.01-0.05 范围内微调。距离阈值 1.0 与角度余弦阈值 0.85 需要与场景归一化尺度匹配，单位较小的数据集（如 DTU）可按场景尺度相应缩放距离阈值；对于法向较稳定的垂直平面可将余弦阈值提高至 0.9 以进一步抑制抖动。如果某些数据平面区域占比极大，可将同化触发上限从 6000 迭代延长到 8000-10 000，使更多靠近平面的 3D 椭球在早期被吸附为 2D 圆盘；反之，对于平面较少的场景则可保持默认设置。

实验数据集：为全面验证本文提出方法的有效性与泛化能力，本文在涵盖多种真实世界场景类型的广泛数据集上进行了系统性实验评估。具体包括：Mip-NeRF360 数据集：该数据集是广泛用于新视角合成和三维重建基准测试的综合性数据集，包含复杂多样的室内外场景（如自行车、花园、厨房、房间等）。其高分辨率的图像和具有挑战性的光照条件及几何结构，为评估方法在复杂环境下的鲁棒性提供了理想平台。本文使用其全部室内外场景进行实验。Replica 数据集^[38]：该数据集包含 18 个在房间和建筑尺度上重建的高度逼真室内三维场景。每个场景不仅提供密集网格和高分辨率高动态范围（HDR）纹理，还包含语义类别、实例信息以及反射表面（如镜子和玻璃）。本文选取了其中 8 个具有代表性的室内场景（如 office0, room0, apartment0 等）作为实验对象，以验证方法在结构化室内环境中的表现。Tanks and Temples (T&T) 数据集：这是一个专注于大型、复杂真实场景（如寺庙、博物馆、办公室等）的基准数据集。“中级”（Intermediate）和“高级”（Advanced）子集对三维重建算法的精度和完整性提出了严峻挑战。本文主要使用其包含大型室内外

结构的场景（如 Train, M60, Playground）来测试方法在大规模复杂环境中的重建能力。DTU 数据集：这是一个经典的以物体为中心的多视角立体视觉数据集。它提供在严格受控光照条件下、围绕物体从多个固定视角拍摄的高质量图像及对应的相机参数和激光扫描真值。数据集包含多种物体，具有不同的材质（包括非朗伯表面）和几何复杂度。本文利用该数据集评估方法在物体级精细几何重建方面的精度。

评价指标计算：对于新视图合成任务，本文采用三种广泛使用的图像评估指标来进行验证：峰值信噪比（PSNR）、结构相似性指数度量（SSIM）^[39]、学习感知图像块相似性（LPIPS）^[40]。而对于平面的重建质量，本文使用完整性评分（IntegrityScore）来评估平面重建的平滑度和完整性，使用椭球采样点到平面距离的 RMSE 来评估平面重建的贴合度。为体现平面重建评估指标的公正性及全面性，其中设计的 IntegrityScore 指标反映了 3D 高斯椭球与目标平面的两个关键方面。

（1）表面平滑度：通过对椭球表面进行均匀采样并计算其每个采样点的局部曲率来量化椭球是否在目标平面上形成了连续且平滑的平面结构。

（2）法向一致性：通过计算椭球最小轴与目标平面的法向量夹角来量化椭球与目标平面的对齐程度，从而反映椭球是否正确地“贴合”在目标平面上。

最后两者各占 50% 权重得到 IntegrityScore。而 D_RMSE 则表示采样点到平面的平均偏差距离，即椭球上采样点到目标平面的均方根误差，单位为米。其值越小，则表明椭球越拟合平面；其值越大，则表明椭球越远离平面。所有的评价指标计算方式分别为：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{N} \sum_{i=1}^N (I_i - \hat{I}_i)^2} \right) \quad (17)$$

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (18)$$

$$\text{LPIPS}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\phi_l(x)_{h,w} - \phi_l(y)_{h,w})\|_2^2 \quad (19)$$

$$D_{\text{RMSE}} = \sqrt{\frac{1}{M} \sum_{i=1}^M (n^T P_i + d)^2} \quad (20)$$

$$\text{IS} = 0.5 \cdot \frac{1}{M} \sum_{i=1}^M 1\{k_i < \tau\} + 0.5 \cdot \frac{1}{M} \sum_{i=1}^M |n_i^T n_\pi|,$$

$$\text{where, } k_i = \frac{\lambda_{i,1}}{\lambda_{i,1} + \lambda_{i,2} + \lambda_{i,3} + \epsilon}, (\lambda_{i,1} \leq \lambda_{i,2} \leq \lambda_{i,3}) \quad (21)$$

其中, PSNR 计算中, L 表示图像像素的最大值(如 255), I_i 和 $\hat{I}_i$ 分别为参考图像与合成图像在第 i 个像素的值, N 为像素总数; SSIM 计算中, x, y 表示参考图像与合成图像像素集合, μ_x, μ_y 为 x, y 的像素均值, σ_x^2, σ_y^2 为 x, y 的像素方差, σ_{xy} 为像素协方差, $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, 其中 K_1 和 K_2 为稳定常数, 默认为 0.01 和 0.03, L 同上; LPIPS 计算中, $\phi_l(\bullet) \in R^{H_l \times W_l \times C_l}$ 表示在预训练深度网络的第 l 层提取的特征, $w_l \in R^{C_l}$ 为通道加权系数。 H_l, W_l, C_l 为第 l 层特征的高, 宽, 通道数。 $\odot$ 为按通道逐元素相乘。 $\|\cdot\|_2$ 为欧几里得范数; D_{RMSE} 计算中, $n^T P_i + d = 0$ 为 π 的平面方程, P_i 为椭球基元中第 i 个采样点; IS 计算中, M 为采样点总数, k_i 为第 i 个采样点的曲率, $\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3}$ 为采样点 P_i 半径为 r 邻域内局部协方差的特征值, ϵ 为极小值, 防止除零。 τ 为阈值常数, 设为 0.05, n_i 和 n_π 为采样椭球的最小轴方向对应的法向量和目标平面法向量。

此外, 考虑到 IntegrityScore 的计算依赖于对目标朗伯平面上高斯椭球表面的离散采样, 为避免采样设置对评价结果产生偏差, 本文对采样策略作出统一约定并进行了鲁棒性分析。具体地, 对属于目标平面的每一个高斯椭球, 本文均在其表面进行均匀采样: 先在单位球面上均匀采样方向向量, 再通过椭球的旋转矩阵与尺度参数进行仿射变换, 从而获得分布近似均匀的表面采样点集合 $\{p_i\}$ 。在所有对比方法(包括 3DGS、2DGS、PGSR 及本文方法)中, 本文统一采用相同的采样策略, 并将单个椭球的采样点数固定为 100, 以保证评价的可比性。需要强调的是, IntegrityScore 本质上是在这些采样点上对局部曲率和平面法向一致性进行统计, 因此只要采样密度足以覆盖目标区域, 其数值就应当对采样参数保持稳定。为此, 本文在多个代表性平面

上进一步考察了采样数量 $\in \{50, 100, 150, 200\}$ 时的 IntegrityScore 变化, 如图 10 所示。

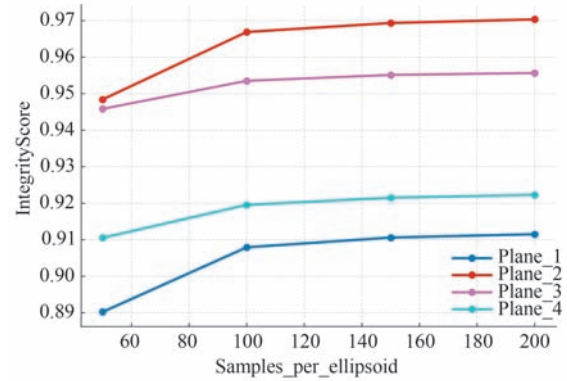


图 10 采样密度曲线图

从图 10 的采样密度曲线图不难看出, 四条曲线整体呈平滑、缓慢上升趋势, 且在 50-200 的范围内最大变化幅度仅在 1%-2% 量级; 不同平面之间的相对排序保持完全一致, 不存在因采样密度改变导致优劣次序翻转的情况。这说明在通常使用的采样密度区间内, IntegrityScore 对采样参数不敏感, 默认的 100 采样点已能得到稳定可靠的评价结论。

进一步地, 本文将不同采样密度下所有椭球的采样点投影到平面坐标系, 统计其二维栅格上的归一化对数密度, 并绘制密度热力图与覆盖率, 见图 11。当 samples=50 时, 平面上亮区已经覆盖了主要朗伯区域, 覆盖率约为 59.4%; 当采样密度从 50 增加到 100 时, 亮区范围和覆盖率仅有 1% 左右的轻微提升(约 59.4% → 61.2%), 主要表现为对边缘小空洞的细化填充, 而不会改变平面上的整体结构。这说明在默认采样密度下, 椭球在目标平面上的采样已经足以覆盖主区域, IntegrityScore 计算实际上是在一组已收敛的几何样本上进行统计, 进一步加密采样只会带来极小的数值扰动, 而不会改变总体评价。

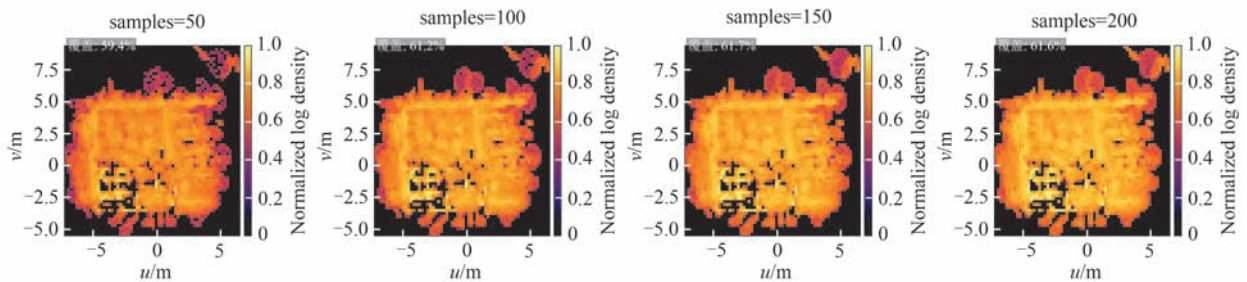


图 11 平面投影局部密度热力图

上述分析表明, 本文所采用的采样设置不会改变评价结论, IntegrityScore 对合理范围内的采样密

度具有良好的鲁棒性, 因此可作为几何重建质量的有效补充指标。

4.1 Mip-NeRF360数据集对比实验

将本文模型与之前最先进的新视图合成方法及同类型的表面重建算法在 Mip-NeRF360 数据集上进行比较,包括不同类型的神经辐射场模型 NeRF、DeepBlending^[41]、INGP、Neus^[42]和同类型的表面重建模型 3DGS、2DGS、PGSR。表1展示了各种模型在 Mip-NeRF360 数据集上得到的包括 PSNR、SSIM、LPIPS、IS 和 D_RMSE 在内的指标值,如表1所示。

神经辐射场模型中,NeRF使用深层MLP拟合场景中每个采样点的体密度和RGB颜色值,严重依

赖采样点数量,在 Mip-NeRF360 数据集上总体渲染质量较低。基于 NeRF 框架,Instant-NGP 模型采用多分辨率哈希编码的方法,在捕获高频局部细节的同时保持紧凑数据结构,大幅提升渲染质量(平均 PSNR 29.15)。而 Neus 摒弃 NeRF 的体积密度表示,转而采用符号距离函数(SDF)定义几何表面,通过可学习的概率密度函数关联 SDF 值与光线投射权重,实现表面几何的精准建模,但其新视图合成质量有所下降(PSNR 25.10)。Deep Blending 作为后处理模型,通过编码器-解码器网络隐式融合多视角纹理,解决光照差异与遮挡伪影问题,PSNR 达到 26.40。

表1 在 Mip-NeRF360 数据集上的定量结果(其中“粗体”表示最优结果,本文方法在室内场景中所有指标均获得最佳结果)

		Indoor scenes					Outdoor scenes				
		PSNR↑	SSIM↑	LPIPS↓	IS↑	D_RMSE↓	PSNR↑	SSIM↑	LPIPS↓	IS↑	D_RMSE↓
NeRF-based	NeRF	26.84	0.790	0.370	--	--	21.46	0.458	0.515	--	--
	INGP	29.15	0.880	0.216	--	--	22.90	0.566	0.371	--	--
	Neus	25.10	0.789	0.319	--	--	21.93	0.629	0.600	--	--
	Deep Blending	26.40	0.844	0.261	--	--	21.54	0.524	0.364	--	--
GS-based	3DGS	30.99	0.926	0.199	0.647	0.0388	24.24	0.705	0.283	0.770	0.0327
	2DGS	30.39	0.923	0.183	0.745	0.0631	24.33	0.709	0.284	0.750	0.0537
	PGSR	30.41	0.930	0.161	0.822	0.0296	24.45	0.730	0.224	0.791	0.0244
	Ours	31.30	0.930	0.160	0.844	0.0212	24.17	0.701	0.285	0.794	0.0229

表面重建模型方面,3DGS引入各向异性3D高斯椭球作为非结构化辐射场表达,通过 Splatting 渲染策略有效规避了传统体渲染的效率瓶颈,在 Mip-NeRF360 基准测试中取得 PSNR 30.99 的突破性成果。然而该模型在几何重建质量方面存在明显缺陷,尤其对平面区域的表达能力不足,其 IS 完整性指标仅 0.647,暴露出各向异性高斯分布与平面几何先验的匹配失当问题。针对这一关键缺陷,2DGS 提出各向同性高斯约束,通过将 3D 高斯投影到 2D 图像平面生成几何先验,强制高斯椭球退化为圆形分布,从而提升薄壁结构和低纹理平面的重建完整性(IS 提升至 0.745)。PGSR 模型进一步引入平面几何正则化,联合优化高斯分布参数与显式平面方程,利用多视角几何一致性损失约束平面区域的法向量对齐,显著降低几何误差(D_RMSE 从 3DGS 的 0.0388 降至 0.0296),同时保持高渲染效率(PSNR 30.41),其工作揭示了平面几何约束对表面重建质量的关键作用,为后续研究提供了重要启示。

在含有大量规则结构和平面构件的室内场景中,本文方法在各项指标上均优于其他表面重建模

型。而在户外场景中,朗伯平面面积较小,平面先验的作用主要集中在少量物体表面、道路等区域,其余大部分区域(植被、天空、远景结构等)仍由 3D 椭球主导;在这种情况下,局部几何优先的策略会导致模型在整体 PSNR/SSIM 上与 3DGS 非常接近甚至略有牺牲(24.17 vs 24.24/0.701 vs 0.705),但换来的是更优的平面完整性与几何精度(0.794 vs. 0.770/0.0229 vs. 0.0327)。

本文方法在 3DGS 模型的基础上引入朗伯平面 2D-3D 混合优化。通过将场景中朗伯平面(如“墙壁”、“天花板”)上的三维高斯动态扁平化为二维分布,实现了平面几何的平滑重建与高质量渲染的协同优化。表1数据和场景法向贴图可视化充分体现了本方法的场景适应性优势。在室内场景基准测试中(包含大量标准朗伯平面),本方法 PSNR 达到 31.30 的当前最优水平,IS 指标较 3DGS 提升 19.7%。值得注意的是,在 Mip-NeRF360 室外场景测试中(如图12(b)、(d)所示),尽管标准朗伯平面占比显著降低,本方法仍保持几何重建精度优势。图12的对比可视化清晰显示:在室内标准平面(图12(a)、(c))和室外非标准朗伯平面(图12(b)、(d))

“桌面”)两种典型场景下,本方法均展现出更精确的几何重建效果。此外,由于Mip-NeRF360数据集只提供了多视角RGB图像及相机位姿,并不包含与本文渲染结果同分辨率、同视角对齐的表面法向或显

式几何网格作为严格意义上的几何真值。为此,本文当前的实验设置与现有绝大部分表面高斯重建方法在真实数据上的评测协议是一致的,使用对应视角的原始RGB图像作为GT参考。

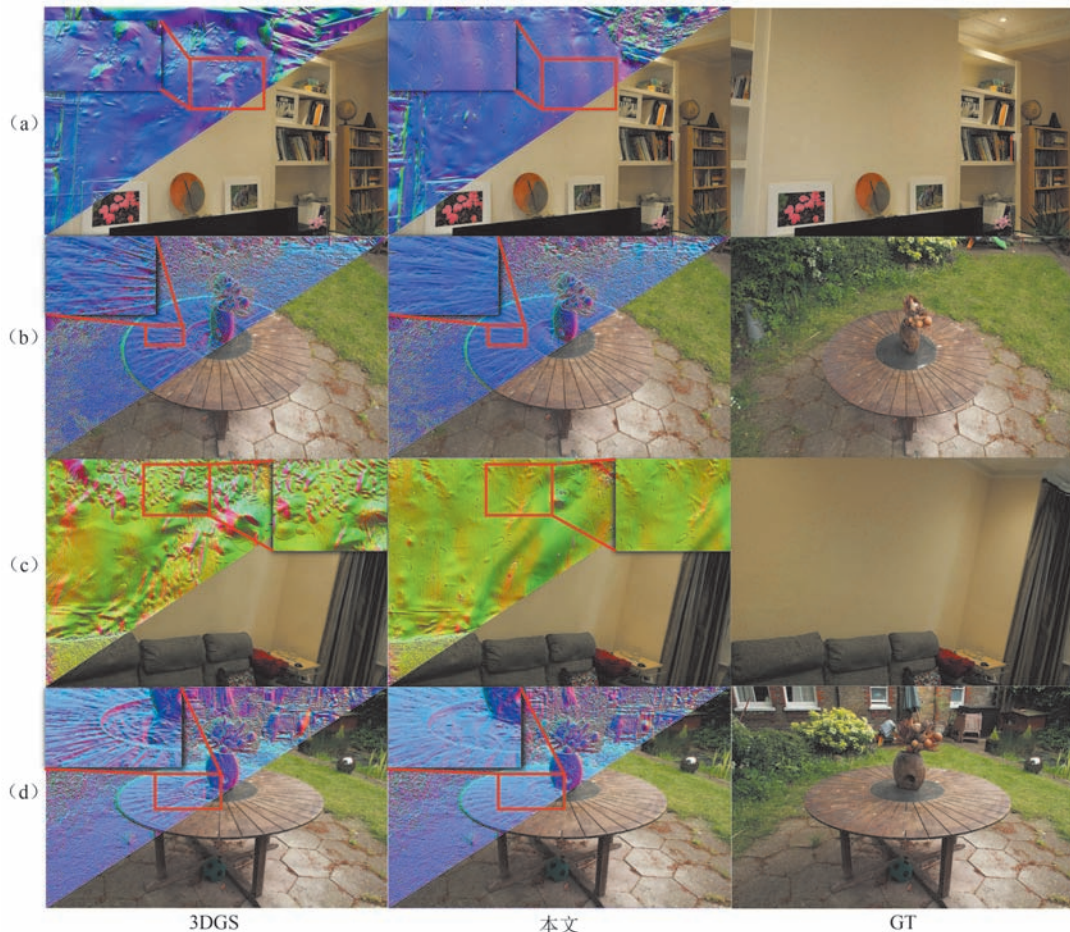


图12 Mip-NeRF360数据集室内外场景法向贴图可视化

4.2 T&T数据集对比实验

基于T&T数据集的法向贴图可视化分析(图13)与定量评估结果(表2)共同揭示了不同模型的性能特征。在典型室内场景(图13(a)(b))中,3DGS展现出对高纹理区域(如“照片密集墙面”)的强重建能力,但其各向异性高斯分布在低纹理的“伪朗伯平面”(如天花板)产生几何重建不完整问题——椭球无序分布导致平面几何严重失真(平面区域D_RMSE达0.0377)。对比而言,2DGS和PGSR模型虽然对于平面的重建效果较好,但其过度平滑了纹理区域,丢失了大量细节特征。如在场景(c)(d),2DGS和PGSR虽然对于朗伯平面的几何重建效果较为优质,但对于“墙面”上“涂鸦图案”的细节特征几乎完全丢失,PSNR仅为29.03和29.29。而本文模型则采用混合高斯的方式,在保

持复杂纹理细节区域高质量渲染的同时大幅度提升了场景中朗伯区域的几何重建精度。PSNR达到同类模型中最优(29.45),IS(平面完整性评分)相较于3DGS提升19.0%,并且高斯椭球到平面的距离均方误差值减少1.54 cm。

4.3 DTU数据集对比实验

针对DTU数据集大部分场景不存在朗伯平面的特性,因此本研究选取包含平面结构的场景11(电源箱)、24(红房子)、40(砖墙)、47(波形发生器)进行专项评估,定性比较与法向贴图可视化结果分别呈现在表3与图14中。值得注意的是,2DGS和PGSR在场景47(b)电源箱、11(c)波形发生器的表面几何重建中展现出结构优势,但如图14(b)中波形发生器面板型号、图14(c)中电源箱表面“DC POWER SUPPLY”提示文字的严重模糊所示,其

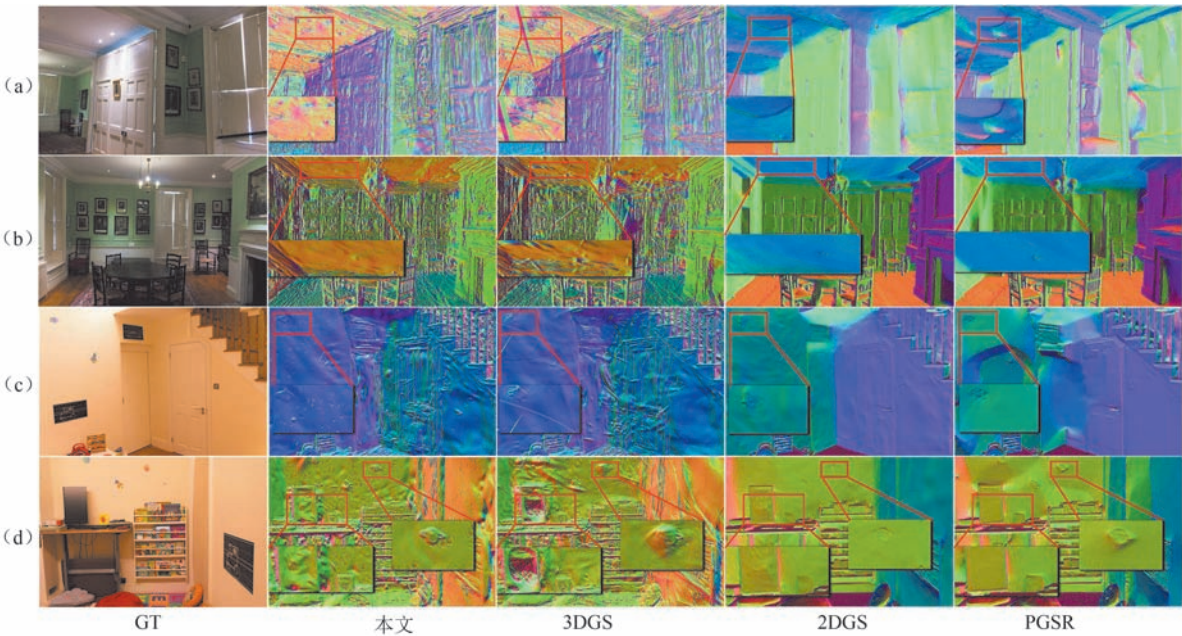


图 13 T&T数据集定性比较(本文使用重建网格生成的法线贴图来可视化场景的重建质量。其中3DGS算法在纹理区域表现较好,几何特征较为明显,但在平面区域重建质量较差。同系列表面重建模型2DGS和PGSR的重建效果则过于平滑,丢失了大量纹理信息。本文方法结合两者优势,在保留纹理区域的细节同时大幅提升了平面区域的重建效果。)

表 2 在 T&T 数据集上的定量结果					
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	IS $\uparrow$	D_RMSE $\downarrow$
NeRF	25.78	0.864	0.198	--	--
INGP	24.96	0.817	0.390	--	--
NeuS	24.11	0.832	0.179	--	--
3DGS	29.41	0.903	0.242	0.644	0.0377
2DGS	29.03	0.914	0.233	0.811	0.0276
PGSR	29.29	0.922	0.237	0.817	0.0269
Ours	29.45	0.904	0.238	0.834	0.0223

纹理细节保留能力存在显著缺陷。本方法通过二维/三维高斯混合优化策略,在复杂非朗伯场景中实现了渲染质量与几何精度的协同提升。由表 3 可知,在维持同系列模型最高新视图合成质量(PSNR 22.55)的前提下,平面重建精度显著优于 3DGS 基线;IS(完整性评分)提升 10.2%,高斯椭球表面采样点到平面距离的均方误差降低 10.58 cm。图 14(c)中电源箱旋钮周边的清晰文字重建(如“DC POWER SUPPLY”标识),直观验证了本方法在细节保留方面的优势。

表 3 在 DTU 数据集上的定量结果					
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	IS $\uparrow$	D_RMSE $\downarrow$
3DGS	21.66	0.773	0.379	0.757	0.1451
2DGS	21.78	0.783	0.367	0.813	0.0667
PGSR	22.10	0.784	0.371	0.855	0.0219
Ours	22.55	0.774	0.371	0.859	0.0393

4.4 Replica室内场景数据集对比实验

为测试本方法在室内场景的重建效果,本文使用 Replica 室内场景数据集进行定量测试,结果如表 4 所示,法向贴图可视化如图 15 所示。

实验分析表明(图 15,表 4),Replica 数据集室内场景以规则平面结构为主且复杂纹理区域较少的主导特性使表面重建方法获得显著优势。基准方法 3DGS 受限于无序椭球表征,其几何重建精度与图像渲染质量方面表现欠佳,平面距离误差达到 11.07 cm。2DGS 通过平面约束改进虽提升 PSNR 至 36.09,但由于其几何简化所导致的细节损失问题,致使其渲染质量在同类型表面重建模型中表现不佳。PGSR 通过多视角几何一致性损失约束平面区域的法向量对齐以及使用无偏深度降低距离误差,使其在 D_RMSE(0.0341)上保持优势。而本文方法在通过法向约束和距离约束在 D_RMSE(0.0370)仅增加 8.5% 的代价下,实现了 PSNR(+0.15)与 IS(+0.006)的双重突破。值得注意的是,在朗伯表面区域,本方法与 PGSR 的 IS 差异小于 1%(0.676 vs. 0.670),D_RMSE 偏差仅 8.5%,表明其继承了表面重建框架的几何保真特性。这种精度-鲁棒性平衡策略为实际应用提供了更优解。

4.5 朗伯平面跨源深度推理测试实验

为验证朗伯平面跨源深度推理中的深度校准的准确性和鲁棒性,以 T&T 数据集中复杂室内场景

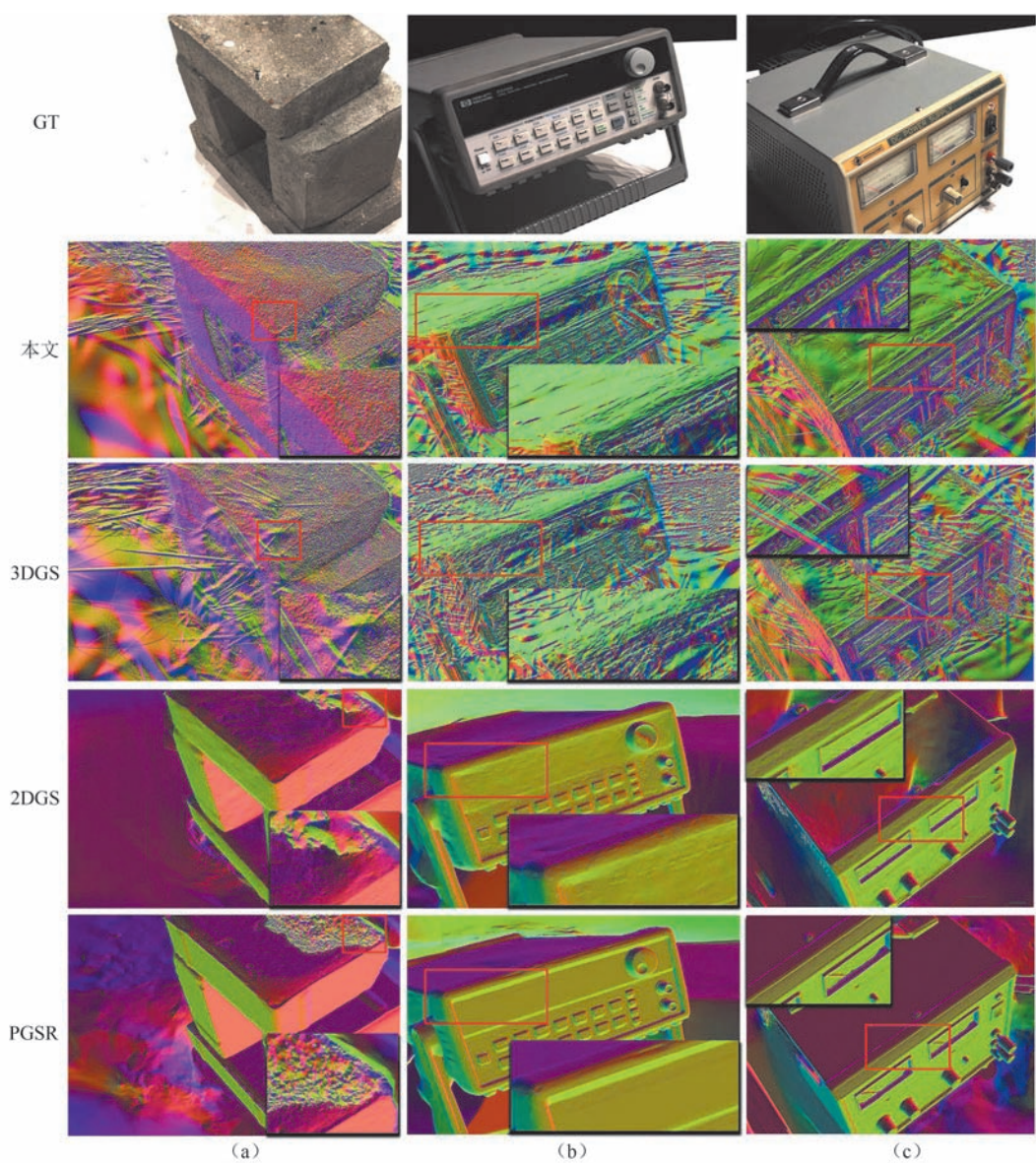


图 14 DTU 数据集定性比较(使用法线贴图来可视化真实物体场景的重建质量,本文方法对于部分区域的细节特征保留效果明显优于其他模型。)

表 4 在 Replica 数据集上的定量结果					
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	IS $\uparrow$	D_RMSE $\downarrow$
3DGS	35.51	0.963	0.069	0.587	0.1107
2DGS	36.09	0.969	0.073	0.664	0.0930
PGSR	36.41	0.970	0.070	0.670	0.0341
Ours	36.56	0.970	0.058	0.676	0.0370

“PlayRoom”为例(室内场景深度变化较室外场景更为剧烈),对其进行深度校准。图 16 给出了朗伯平面区域深度估计在分段纠正前后的误差分布情况,其中图 16(a)为在同一视角下未进行跨源深度推理的真实点云深度与对应像素位置估计深度的比值 Ratio 分布。16(a)左边为特征点对应位置比值可视化图,

横纵坐标为对应图像大小,所有特征点按照真实深度和估计深度比值差共分为 5 类颜色,其中绿色为准确估计点(比值差在 0.05 之间),粉色、灰色为较准估计点(比值差在 0.2 之间),红色与黄色为较差估计点(比值差大于 0.2)。16(a)右边为特征点深度比值差统计图,横坐标为特征点数量,纵坐标为比值区间,共分为 10 段。图 16(b)为跨源深度推理之后真实点云深度与校准深度的比值大小。

由图 16 可知,朗伯平面区域的相对深度估计值与点云投影的真实深度存在显著尺度偏差(见图 16a),其中红色比值点产生较多,表明单目深度估计网络整体估计的深度值较小,直接应用会导致初始化圆盘基元会朝向相机方向偏移;校正后的尺

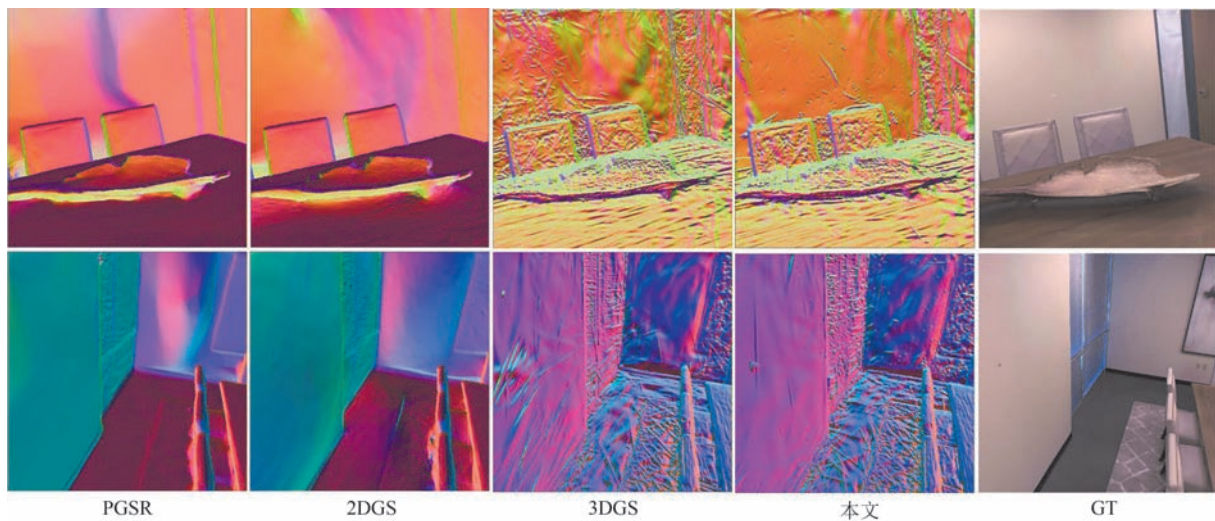


图15 Replica数据集的定性比较

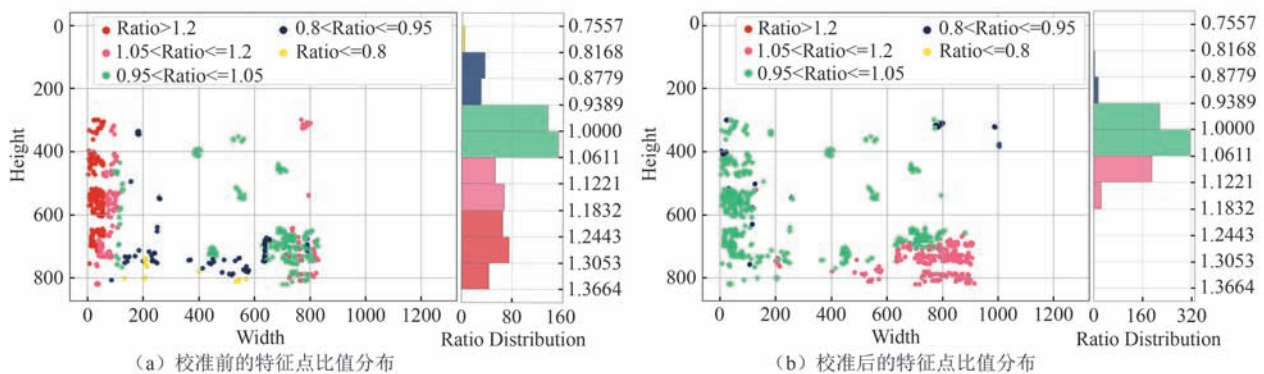


图16 朗伯平面跨源深度推理及校准

度偏差得到有效抑制(见图16(b)),所有深度比值都为绿色、粉色和灰色,即深度估计较为准确,为后续圆盘基元初始化提供稳定基础。

4.6 高斯重建基元数及时间效率分析

为进一步分析2D-3D混合高斯表示的参数规模与收敛行为,本文在 Tanks&Temples数据集上统计了各方法在训练过程中的基元数量变化及训练耗时。所有模型在相同配置下训练:深度学习环境为 PyTorch 1.12.0+cu113,硬件为 NVIDIA RTX 4080 Laptop GPU,总训练步数为30 k,并按0~5 k、5 k~10 k、10 k~15 k、15 k~20 k、20 k~25 k、25 k~30 k六个阶段进行记录。

图17(a)展示了本文方法在不同阶段的2D基元、3D基元以及总基元数量的极坐标统计结果,可以看到:在0-5 k阶段,平面区域的2D基元迅速收敛到约 4×10^4 个,之后在各阶段维持在 4.0×10^4 左右,说明渐进式平面检测与2D基元融合机制能够在早期就稳定捕获主要朗伯平面,并通过同化与剔除策

略避免2D基元数量的无界膨胀。与此同时,3D基元数量从初始阶段的约 9.3×10^5 随训练逐渐增加,在10 k-15 k阶段达到约 1.86×10^6 并基本保持不变,对应的总基元数约为 1.94×10^6 。值得注意的是,2D基元在最终阶段仅占总基元数的约2%,表明本文方法在总体参数规模上仍以3D基元为主,而2D分支以较小的额外代价显著提升了平面区域的几何一致性。

图17(b)对比了本文方法与3DGS、2DGS、PGSR在各训练阶段的总基元数量(纵轴为基元总数量,横轴为5 k-30 k的训练步数)。可以观察到,在早期5 k步时,本文方法的总基元数约为 9.8×10^5 ,明显少于3DGS和PGSR(均在 1.3×10^6 量级),略高于仅采用2D表示的2DGS。随着迭代次数增加,3DGS的高斯数量在30 k步前后稳定在约 2.36×10^6 ,而本文方法最终稳定在约 1.94×10^6 ,高斯数量相对减少约18%,表明引入平面先验和融合策略可以在保证重建质量的同时有效压缩3D基元

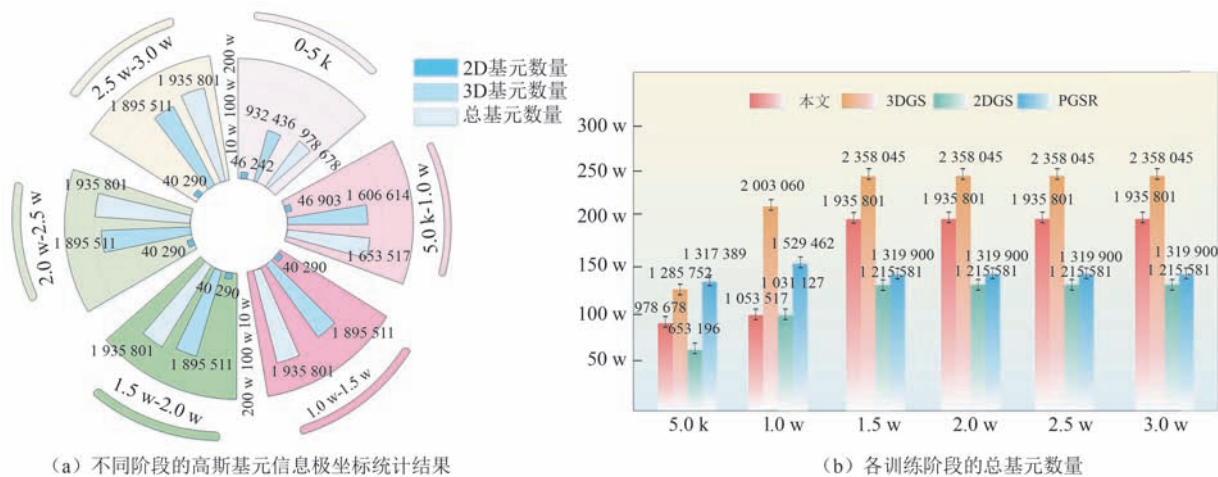


图 17 模型训练各阶段基元数量对比图

规模。相比之下,2DGS 和 PGSR 在后期阶段的总基元数分别稳定在约 1.22×10^6 和 1.32×10^6 ,整体规模略低于本文方法,但在前文的定量与定性结果中,它们在朗伯平面几何精度和结构完整性方面明显逊于本文,这说明本文模型以适度增加基元数量为代价换取了更高的几何表达能力。

图 18 和表 5 统计了四种方法在 Tanks&Temples 数据集上随训练步数变化的累计训练时间(纵轴为分钟,横轴为 5 k~30 k 的训练步数)。整体来看,除 PGSR 外,其余三种方法的时间曲线均近似线性增长,说明在训练过程中单步计算代价基本稳定。PGSR 的时间开销最高,在 5 k、10 k、15 k、20 k、25 k、30 k 步时的累计耗时分别为 9.3、20.3、40.9、61.9、79.7 和 111.6 min;而本文方法在相同阶段的耗时为 4.3、13.2、29.2、40.0、48.7 和 55.3 min,最终总训练时间约为 PGSR 的 49.5%(55.3 vs. 111.6 min),相当于在相同步数下获得约 2.01 倍的训练效率。与基础的 3DGS 和 2DGS 相比,本文方法在 0-5 k 阶段由于需要完成 2D 基元的初始化及其匹配/同化最优的 3D 基元,耗时略高于二者(5 k 时分别为 4.3、4.1、3.7 min);但在随后的各阶段,每增加 5 k 步的增量时间稳定在 4.4~16.0 min,与 3DGS(9.0~14.5 min) 和 2DGS(7.1~14.6 min) 处于同一量级;在 1.5 w 步之后,由于不需要进行基元之间的渐进式融合,本文方法中的 2D 基元计算量少于 3D 基元,因此训练效率明显提高,超过了 3DGS 模型。最终在 30 k 步时,本文方法的总训练时间比 3DGS 降低约 16.5%(55.3 vs. 66.2 min),比 2DGS 增加约 3.5%(55.3 vs. 53.4 min)。综合基元数量与训练时间两个维度,可以看出在引入 2D-3D 混合基元及渐进融合机制后,

本文方法不仅在时间成本上超越部分主流高斯重建算法,并在前文实验中表现出更优的几何重建质量与朗伯平面结构完整性。

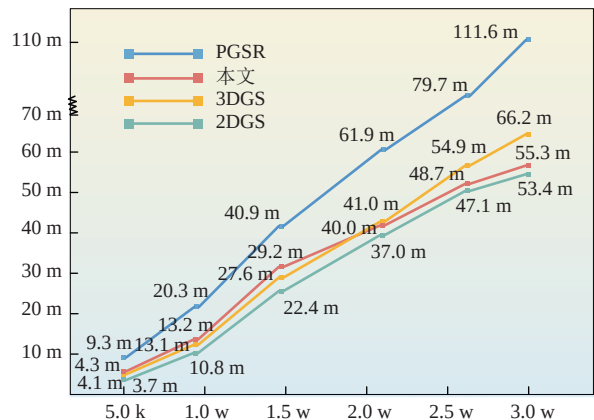


图 18 模型训练各阶段训练时间效率对比图

表 5 在 T&T 数据集上的定量结果					
	Time ↓	Total-num ↓	3D-num ↓	2D-num ↓	Device
3DGS	66.2 m	2 358 045	2 358 045	--	4080
2DGS	53.4 m	1 215 581	--	1 215 581	4080
PGSR	111.6 m	1 319 900	1 319 900	--	4080
Ours	55.3 m	1 935 801	1 895 511	40 290	4080

4.7 模型消融实验

为验证各个关键模块对朗伯平面重建质量的贡献,本节在 Tanks&Temples 数据集的典型室内场景 Playroom 上进行了系统的消融实验。除特别说明外,训练配置与前述实验完全一致:训练步数为 30 k,优化器与超参数保持不变,均在同一台 NVIDIA RTX 4080 Laptop GPU、PyTorch 1.12.0+cu113 环境下完成,所有方法的“训练时间”统计均不包含

SfM/COLMAP 以及深度网络等离线预处理的耗时。本文对比了以下几种设置：

(1)Ours (full):完整的2D-3D混合高斯模型,包含平面先验检测、渐进式基元融合/剔除以及2D基元的距离与法向量几何约束;

(2)w/o Plane Prior:移除平面候选检测,2D基元仅随机初始化其位置与法向量;

(3)w/o Progressive Fusion:保留平面检测,但不再使用渐进式融合与剔除策略,2D基元不与3D基元进行协同优化;

(4)w/o Dist. +Normal Loss:取消2D基元到目标平面的距离约束和法向一致性约束,仅依赖重投影的照片一致性损失;

(5)3DGS:仅使用3D高斯基元的标准3DGS作为几何基线模型。

表6给出了各方法在该场景上的PSNR、SSIM、IntegrityScore (IS)与D_RMSE结果。表7给出了各方法在该场景上的各类高斯基元数量及比例结果。

表6 在T&T数据集上的质量指标定量结果				
	PSNR↑	SSIM↑	IS↑	D_RMSE↓
3DGS	29.93	0.904	0.644	0.0377
Ours(full)	30.29	0.911	0.824	0.0234
w/o Plane Prior	29.87	0.903	0.789	0.0311
w/o P F	29.28	0.895	0.677	0.0378
w/o Dist. +Normal	30.10	0.905	0.738	0.0308

从表6可以看出,相较于3DGS模型,本文方法在外观指标上取得了略优于3DGS的结果(PSNR提升0.36 dB、SSIM提升0.007),而在几何相关指标上获得显著优势:IS从0.644提升到0.824,平面

表7 在T&T数据集上的基元数量定量结果				
	Total(k)	3D(k)	2D(k)	2Dratio(%)
3DGS	2358	2358	--	0.0%
Ours(full)	1929	1889	39.77	2.1%
w/o Plane Prior	972	602k	370.6	38.1%
w/o P F	2193	2192	0.298	0.01%
w/o Dist. +Normal	1843	1816	27.2	1.5%

完整性指标提升约18%,D_RMSE从0.0377降至0.0234,误差降低约38%。结合表7可知,此时模型的总基元数约为1929 k,其中2D基元仅有约39.77 k,占比约2.1%。也就是说,在总体参数低于3DGS的前提下,少量2D平面基元已经足以接管主要朗伯平面区域,并有效提升几何完整性。

对比w/o Plane Prior可以更清楚地看到平面先验的作用。该设置中去除了显式平面候选,2D基元使用随机中心位置与法向量来进行初始化。表7显示,其最终总基元数量仅为972 k,但2D基元数量却大幅上升至370.6 k,占比高达38.1%。这是因为在缺乏平面几何约束的情况下,渐进式融合会“盲目”地将3D基元同化进相邻的2D基元,致使2D分支不再聚焦于朗伯平面,而是沿着局部纹理和深度变化扩散开来。尽管这种大量的2D基元在局部上能够提高对纹理的拟合能力,因而PSNR、SSIM仅略低于完整模型(29.87 vs. 30.29 dB,0.903 vs. 0.911),但由于2D基元缺乏统一的平面法向与距离约束,整体几何趋于破碎和起伏:D_RMSE从0.234退化到0.0311,IS也从0.824降至0.789。结合法向图对比可视化可以观察到(见图19),在墙面等本应平滑的区域出现大面积波纹和“浮雕”结构,这印证了大量无约束2D基元更倾向于追随图像纹理而非真实朗伯平面的现象。

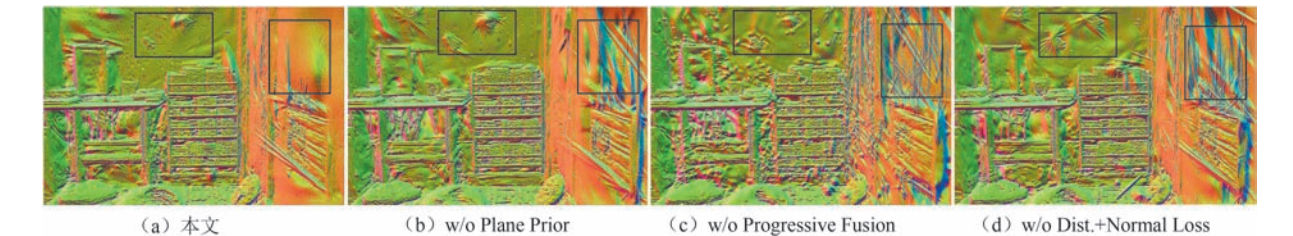


图19 模型消融实验法向量可视化对比图

在w/o Progressive Fusion中,虽然保留了平面先验检测,但去掉了渐进式基元融合与剔除机制。表7显示其最终总基元数约为2193 k,2D基元仅剩不到300个,占比不足0.02%。由于缺乏融合机制

对2D基元数量进行稳定控制,2D与3D基元之间不能形成“接管-优化”的正反馈,大多数2D基元在训练早期就被梯度抑制和密度控制所剔除,2D分支几乎完全失活,模型退化回近似纯3DGS的表示形式。

对应地,表6中该设置的D_RMSE和IS与3DGS几乎持平甚至略差(0.0378 vs. 0.0377, 0.677 vs. 0.644),PSNR/SSIM也明显低于完整模型。这说明,如果缺失渐进式融合,即便有平面先验,2D基元也难以维持足够数量去覆盖朗伯平面,最终沦为零散的“噪点基元”,对整体几何质量几乎没有贡献。

w/o Dist. + Normal Loss进一步体现了几何约束本身的重要性。该设置下,2D基元仍然具备平面先验和渐进式融合机制,但去掉了到平面距离和法向一致性两项损失,仅依赖照片一致性进行优化。表7中可以看到,它的2D基元数量略少于完整模型(27.2 k, 1.5%),原因是缺乏几何约束的2D基元在训练过程中更容易产生“乱动”,其位置和法向在深度方向上频繁震荡,部分基元因此被密度调度和剪枝策略移除。具体到训练过程,10 k-15 k阶段单阶段耗时从完整模型的约16.0 min急剧增加到28.8 min,反映出优化器在该阶段需要频繁调整这些不稳定的2D基元。虽然在渲染质量上,PSNR仍接近完整模型(30.10 vs. 30.29 dB),但几何指标明显退化:IS降到0.738, D_RMSE增大到0.0308(见表6)。法向图对比也显示(见图19),墙面上出现大量高频“拉扯”纹理,平面结构被细碎纹理撕裂。由此可见,距离和法向约束是确保2D基元真正“贴”在朗伯平面上、而不是仅拟合投影颜色的关键因素。

图19给出了完整模型与三种消融设置在同一视角下的法向图对比,其中(b)为w/o Plane Prior设置,(c)为w/o Progressive Fusion设置,(d)为w/o Dist. + Normal Loss设置。可以观察到:完整模型在墙面以及背景平面等朗伯区域呈现大面积色彩一致的法向分布,边缘轮廓清晰,几乎不存在条纹状伪影;w/o Plane Prior中,背景墙及右侧墙面出现明显波纹和局部扭曲,法向在矩形框区域内呈现大面积起伏,说明随机初始化的2D基元难以自动聚合成全局平面;w/o Progressive Fusion由于2D基元数量极少,图像整体呈现出与3DGS相似的条纹状与块状噪声,平面区域被不规则地切割;w/o Dist. + Normal Loss中,虽然色彩细节较丰富,但平面区域法向出现大量高频“拉扯”,对应的D_RMSE和IS也明显恶化,验证了几何约束对平面结构完整性的必要性。

综合定量与定性结果可以看出,平面先验、渐进式基元融合以及距离+法向约束三者共同作用,使2D基元在参数量较小的前提下,能够稳定接管朗伯平面区域并显著提升其几何一致性;任何一项缺失

都会导致几何质量明显退化或训练行为不稳定,从而验证了本文技术设计的有效性与必要性。

4.8 平面先验噪声消融实验

如前所述,本文通过“粗筛-精炼-定位”三阶段平面检测模块在多视图几何约束下估计朗伯平面,并据此初始化2D圆盘基元并施加距离与法向几何约束。为了定量分析平面检测误差对模型性能的影响,在Tanks&Temples数据集的代表性室内场景上,我们进一步设计了“平面先验噪声消融实验”。

实验设置:

在保持平面检测掩膜不变的前提下,对检测得到的主平面参数($\hat{n}, \hat{d}$)人为加入不同强度的噪声:将 $\hat{n}$ 绕随机轴旋转固定角度 $\Delta\theta$,并在距离 $\hat{d}$ 上施加偏移 Δd 。本文考虑三种噪声级别:

(1) $5^\circ/-2$ cm:轻度噪声,模拟平面法向和距离存在小偏差的情况;

(2) $10^\circ/-5$ cm:中度噪声,对应较为明显但仍可在可接受范围内的误差;

(3) $20^\circ/-10$ cm:重度噪声,近似于将平面误判到另一近邻几何层面的极端情况。

除平面参数之外,其余训练超参数和流程与完整模型完全一致,仍然使用原有的距离+法向几何损失以及渐进式基元融合/剔除策略。表8给出了不同噪声设置下的PSNR、SSIM、IntegrityScore(IS)和D_RMSE结果。

表8 在T&T数据集上的噪声消融实验定量结果

Setting	PSNR $\uparrow$	SSIM $\uparrow$	IS $\uparrow$	D_RMSE $\downarrow$
3DGS	29.93	0.904	0.644	0.0377
Ours(full)	30.29	0.911	0.824	0.0234
$5^\circ/-2$ cm	30.17	0.909	0.812	0.0276
$10^\circ/-5$ cm	30.06	0.905	0.767	0.0306
$20^\circ/-10$ cm	29.90	0.903	0.734	0.0374

从表8可以看到,随着平面先验噪声的增大,模型的外观指标和几何指标呈现出渐进退化而非突变崩溃的趋势:

在轻度噪声 $5^\circ/-2$ cm下,PSNR从30.29 dB轻微下降到30.17 dB,SSIM从0.911降至0.909;IS仅从0.824降到0.812, D_RMSE从0.0234增至0.0276。几何误差有一定上升,但相比3DGS(IS=0.644, D_RMSE=0.0377)和w/o Plane Prior(IS=0.789, D_RMSE=0.0311)仍具有明显优势,说明在实际场景中常见的轻度平面检测偏差下,2D圆盘基元仍能较好地贴合真实平面。

在中度噪声 $10^\circ/-5\text{ cm}$ 下, PSNR/SSIM 仍稳定在 30 dB/0.905 左右, 表明整体渲染质量基本不受影响; 但 IS 下降到 0.767, D_RMSE 提升到 0.0306, 几何误差逐渐接近 w/o Plane Prior 的水平。这一现象说明当平面偏差变大时, 几何正则项与光度项开始在 2D 基元的更新方向上“拉扯”, 部分平面区域由理想的平滑结构转变为局部起伏, 但整体仍优于完全不使用平面先验的情形。

当噪声进一步增大到 $20^\circ/-10\text{ cm}$ 时, IS 进一步下降至 0.734, D_RMSE 上升到 0.0374, 已经非常接近 3DGS 的 0.0377。此时, 平面先验的偏差足以使 2D 几何约束与真实平面严重不一致, 几何误差主要由光度一致性和密度调度平衡, 使得模型在该区域逐渐退化为接近 3D 表达的效果。其 PSNR 和 SSIM 仍保持在 29.90 dB 和 0.903, 说明错误平面并未导致整体重建崩溃, 而是在几何层面承担了“向 3DGS 下界退化”的角色。

从“近似平面”的角度来看, 表 8 中的噪声实验也可以理解为对现实中轻微弯曲或起伏平面的模拟。在实际场景中, 墙面、地板等结构往往并非严格平整, 而是在局部存在若干度的法向偏差以及厘米

级的高度起伏, 这在几何上等价于在理想平面参数上叠加 $5^\circ/-2\text{ cm}$ 或 $10^\circ/-5\text{ cm}$ 量级的扰动。实验结果表明, 在上述轻度和中度扰动下, 模型的 PSNR 和 SSIM 基本保持在 30 dB、0.905 左右, 渲染质量几乎不受影响; 几何指标 IS 与 D_RMSE 虽然较完整模型有所退化, 但整体仍显著优于 3DGS, 并通常优于或接近完全去除平面先验(w/o Plane Prior)的情形。这说明 2D 圆盘分支并不要求目标表面是严格意义上的理想平面, 只要该区域在局部尺度上仍能被 PCA 估计为单一主平面, 平面几何正则项便能够稳定提升其整体完整性, 而不会因为轻微弯曲而导致几何崩溃。

当扰动增大到 $20^\circ/-10\text{ cm}$ 时, IS 和 D_RMSE 进一步退化并接近 3DGS 的水平, 表明当平面假设与真实几何严重不一致时, 模型会逐渐削弱来自平面先验的约束, 更多依赖光度一致性与密度调度进行优化, 使得该区域的表达形式自然退化为接近 3D 椭球的表示。

图 20 展示了不同噪声设置下的高斯基元可视化结果(仅显示 2D 与临近 3D 基元), 红框标出的是主要朗伯平面区域。可以观察到:

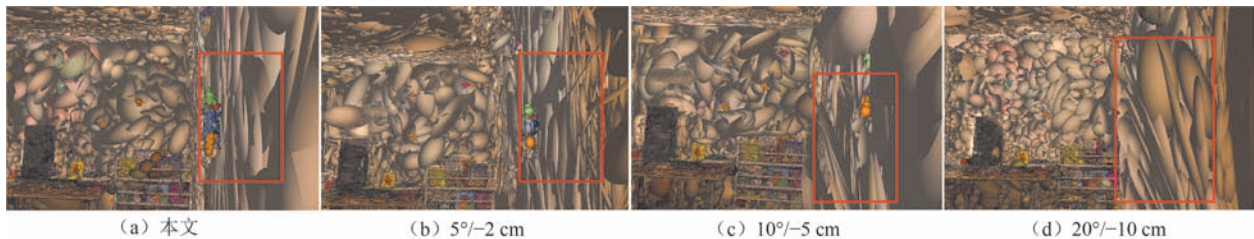


图 20 平面先验噪声消融实验可视化对比图

在无噪声的完整模型中, 该墙面由一条连续、厚度较小的 2D 基元带构成, 与周围 3D 高斯明显区分, 整体几何平整、过渡自然。在 $5^\circ/-2\text{ cm}$ 的轻度噪声下, 红框内的圆盘带仍然保持大体平直, 仅在局部出现轻微起伏, 验证了定量结果中“几何指标轻微退化”的结论。 $10^\circ/-5\text{ cm}$ 时, 墙面上开始出现明显的凸起与凹陷, 2D 基元与邻近 3D 基元混杂, 形成局部块状凸起, 相应地 IS 和 D_RMSE 也出现了较明显的退化。在 $20^\circ/-10\text{ cm}$ 的极端噪声下, 平面区域几乎完全被不规则团簇填充, 已难以辨识清晰的 2D 圆盘带, 其几何表现接近 3DGS 的结果, 印证了 D_RMSE 与 3DGS 接近的定量表现。

4.9 点云稀疏性与噪声偏差实验

如第 3.2 节所述, 本文在训练前通过 SfM 点云

对单目深度图进行一维尺度校准。具体做法是, 将 SfM 点云投影到图像平面, 对于每个同时拥有单目深度 D^p 与 SfM 真实深度 D^T 的像素, 计算二者之比。然后在预测深度轴上等间隔划分若干区间, 并以其中值作为该深度段的缩放因子。由于使用的是分段中值而非均值, 该估计对局部外点和随机噪声具有较高的抗干扰能力。

为分析 SfM 点云的稀疏性与噪声对上述校准过程的影响, 本文在 Mip360 数据集的代表性场景(room)上设计了如下对比实验:

- (1) 完整点云(100%): 使用原始 SfM 点云作为参考;
- (2) 稀疏 50%/25%: 随机保留 50% 和 25% 的参考点;
- (3) 噪声 3 cm/5 cm: 在完整点云的深度值上叠

加零均值高斯噪声,标准差分别为 $\sigma=0.03\text{ m}$ 和 $\sigma=0.05\text{ m}$ 。

在每一种设置下,本文重新计算缩放因子曲线。图 21 给出了不同设置下的曲线对比:横轴为预测深度值(m),纵轴为缩放因子。蓝色实线表示完整点云的结果,橙色和绿色虚线分别对应稀疏 50%和 25%的情况,红色和紫色点线则表示在原始点云上叠加 3 cm 和 5 cm 噪声。

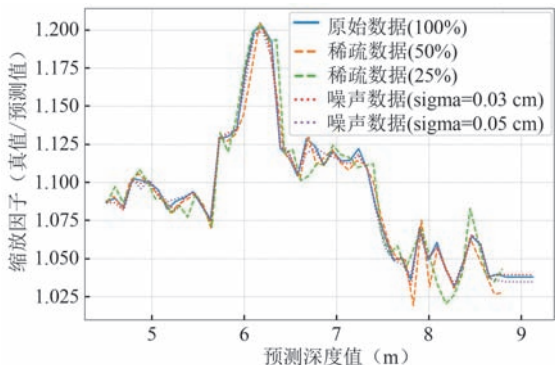


图 21 缩放因子曲线鲁棒性分析可视化对比图

从图中可以观察到以下几点现象：

(1) 对噪声的鲁棒性:红色和紫色点线几乎与蓝色实线完全重合,只在局部存在轻微偏差。这意味着即便 SfM 点云存在 3-5 cm 级别的随机误差,分段中值估计仍能恢复出与无噪声情形几乎一致的缩放因子曲线。其原因在于:噪声在不同像素上的分布近似独立,正负波动在同一 bin 内相互抵消,而中位数统计天然对少量离群值不敏感。

(2) 对中等稀疏性的适应性:当随机丢弃 50%的点时,橙色虚线在整个深度范围内依然紧贴蓝色实线,仅在个别位置产生细微偏移。这表明本文的校准算法并不依赖高度稠密的 SfM 点云,只要在每个深度段中仍有足够数量的点样本,分段中值就能够稳定地估计真实的尺度因子。

(3) 极端稀疏场景下的退化模式:在 25% 点云的设置下,大部分深度范围内绿色虚线仍与蓝线比较接近,但在约 6.5 m 和 8.5 m 两个深度附近出现明显的尖峰和振荡。这些位置恰好对应于该深度段附近几乎没有 SfM 点落入的区域:由于每个 bin 中只剩下极少数甚至单个样本,中位数不再具有统计意义,缩放因子被少数噪声点主导。该现象一方面揭示了方法的物理边界—在极端稀疏情况下,任何基于统计的校准都不可避免地不稳定。

4.10 损失正则项灵敏度分析实验

为讨论本文新增的法向约束与距离约束正则项权重对于实验结果的影响,本文对实验部分进行了补充和重组:在原有模块级消融(平面先验、渐进式基元融合、几何约束)以及平面先验噪声、SfM 稀疏/噪声鲁棒性分析的基础上,新增了对公式(16)中几何损失权重 β (法向项)、 γ (距离项)的灵敏度实验。

具体地,本文在 Tanks&Temples 数据集上的代表性场景中,保持其他训练配置和 DSSIM 权重 λ 不变(λ 一直固定为 0.2,与 3DGS 一致),仅改变几何损失的权重,测试了四组配置,并统计 PSNR、SSIM、IntegrityScore (IS) 与 D_RMSE 的变化。实验结果如表 9 所示。

表 9 在 T&T 数据集上的灵敏度分析定量结果

$\beta=\gamma$	PSNR $\uparrow$	SSIM $\uparrow$	IS $\uparrow$	D_RMSE $\downarrow$
0.00	30.10	0.905	0.738	0.0308
0.05	30.08	0.906	0.745	0.0297
0.15	30.29	0.911	0.824	0.0234
0.30	30.25	0.909	0.831	0.0232

可以看到,当完全去掉几何约束($\beta=\gamma$)时,重建的外观质量仍然合理(30.10 dB/0.905),但几何指标明显退化 (IS 降至 0.738, D_RMSE 升至 0.0308),验证了几何损失对提升平面完整性的必要性。随着 β,γ 从 0 增大到 0.05 和 0.15,IS 单调上升, D_RMSE 单调下降,说明适度增大几何权重可以持续改善几何重建,而对 PSNR/SSIM 几乎没有负面影响;当 $\beta=\gamma=0.15$ 时,PSNR 和 SSIM 达到最高 (30.29 dB/0.911),几何指标也已显著优于 $\beta=\gamma=0$ 的情形。进一步增大到 0.30 时,几何指标继续略有提升 (IS=0.831, D_RMSE=0.0232),但 PSNR/SSIM 相比 0.15 略有下降 (30.25 dB/0.909),表明过强的几何约束会轻微挤压外观重建的自由度。

总体来看,模型在 $\beta=\gamma\in[0,0.30]$ 范围内性能随权重变化平滑且稳定,不存在“轻微改动就导致训练失败”的现象;在 $\beta=\gamma=0.15$ 附近可以取得外观质量与几何完整性的较优折中。因此在全文实验中统一采用 $\beta=\gamma=0.15$ 。

5 总结与展望

针对三维椭球基元对朗伯平面的几何表达能力有限以及传统 3DGS 模型优化过程中结构无序性导

致其与实际场景表面的对应不够精确的问题, 本方法结合3DGS在复杂区域渲染质量优势与2DGS在平面区域重建中的高几何精度特点, 提出顾及朗伯表面精确重建的LPGS, 通过混合优化2D-3D高斯基元实现了复杂纹理区域各向异性高斯表达与平面区域各向同性圆盘表征的协同优化, 在保持3DGS卓越渲染效率(Mip-NeRF360数据集PSNR 31.30)的同时, 显著提升平面几何重建精度(IS指标较3DGS提升19.7%)。然而, 本文并未充分考虑场景中部分噪声基元产生的伪影现象, 应进一步对场景中混合基元进行噪声评估并进行降噪处理; 此外, 本文也未针对提出的新框架进行分布式优化处理, 未来工作将聚焦于上述改进方面, 以进一步优化模型使其符合实际应用。

参 考 文 献

- [1] Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 2021, 65(1): 99-106
- [2] Kerbl B, Kopanas G, Leimkühler T, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 2023, 42(4): 139:1-139:14
- [3] Yang H, Chen R, An S P, et al. Review of advances in image-based 3D reconstruction technology under the back-ground of deep learning. *Journal of Image and Graphics, China*, 2023, 28(8): 2396-2409 (in Chinese)
(杨航, 陈瑞, 安仕鹏, 等. 深度学习背景下的图像三维重建技术进展综述. *中国图象图形学报*, 2023, 28(8): 2396-2409)
- [4] Wang F, Yin Y, Wang JY, et al. Object segmentation in 3D reconstructed scenes based on gaussian splatting. *Chinese Journal of Computers*, 2025, 48(5): 1232-1243 (in Chinese)
(王锋, 银莹, 王佳炎, 等. 基于高斯泼溅的轻量级重建场景分割方法. *计算机学报*, 2025, 48(5): 1232-1243)
- [5] Max N. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1995, 1(2): 99-108
- [6] Hou F, Sun Q, Fang Z, et al. Poisson vector graphics (PVG). *IEEE Transactions on Visualization and Computer Graphics*, 2018, 26(2): 1361-1371
- [7] Tang S, Zeng WL, Yang SL, et al. Intra-block and inter-block dual aggregation transformer for single image super-resolution. *Chinese Journal of Computers*, 2024, 47(12): 2783-2802 (in Chinese)
(唐述, 曾琬凌, 杨书丽, 等. 基于Transformer的块内块间双聚合的单图像超分辨率重建网络. *计算机学报*, 2024, 47(12): 2783-2802)
- [8] Cheng H, Wang S, Li M, et al. Review of neural radiation fields for autonomous driving scenes. *Journal of Graphics*, 2023, 44(6): 1091. (in Chinese)
(成欢, 王硕, 李孟, 等. 面向自动驾驶场景的神经辐射场综述. *图学报*, 2023, 44(6): 1091)
- [9] Wang Y, Yan G, Zhu H, et al. VC-Net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 27(2): 1301-1311
- [10] Tian X, Günther T. A survey of smooth vector graphics: Recent advances in representation, creation, rasterization, and image vectorization. *IEEE Transactions on Visualization and Computer Graphics*, 2022, 30(3): 1652-1671
- [11] Zheng T X, Huang S, Li Y F, et al. Review of key technologies for visual-based 3D reconstruction. *Acta Automatica Sinica*, 2020, 46(4): 631-652 (in Chinese)
(郑太雄, 黄帅, 李永福, 等. 基于视觉的三维重建关键技术研究综述. *自动化学报*, 2020, 46(4): 631-652)
- [12] Kern M, Neuhauser C, Maack T, et al. A comparison of rendering techniques for 3D line sets with transparency. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 27(8): 3361-3376
- [13] Zhang JF, Yan T, Wang KQ, et al. 3D shape reconstruction from multi depth of field images: Datasets and models. *Chinese Journal of Computers*, 2023, 46(8): 1734-1752 (in Chinese)
(张江峰, 闫涛, 王克琪, 等. 多景深图像聚焦信息的三维形貌重建: 数据集与模型. *计算机学报*, 2023, 46(8): 1734-1752)
- [14] Li ZX, Jiang H, Liu YQ, et al. Research on multi-view stereo 3D reconstruction in virtual reality system of silk road cultural inheritance. *Chinese Journal of Computers*, 2022, 45(3): 500-512 (in Chinese)
(李兆歆, 蒋浩, 刘衍青, 等. 丝路文化虚拟体验中的多视角立体重建技术研究. *计算机学报*, 2022, 45(3): 500-512)
- [15] Li H, Wang J Y, Cheng Y Z, Liu J, Zhao G W, Chen S M, et al. 3D object detection method guided by image semantic features and cross-modal fusion with point clouds. *Journal of Computer-Aided Design & Graphics*, 2024, 36(5): 734-749 (in Chinese)
(李辉, 王俊印, 程远志, 刘健, 赵国伟, 陈双敏, 等. 图像语义特征引导与点云跨模态融合的三维目标检测方法. *计算机辅助设计与图形学学报*, 2024, 36(5): 734-749)
- [16] Hou Y, Wang C, Wang J, et al. Visual evaluation for autonomous driving. *IEEE Transactions on Visualization and Computer Graphics*, 2021, 28(1): 1030-1039
- [17] Guo K Y, Liu Y, et al. A review of visual SLAM technology combined with implicit mapping. *Journal of Computer-Aided Design & Graphics*, 2025, 37(5): 725-743 (in Chinese)
(郭恺悦, 刘越, 等. 结合隐式建图的视觉SLAM技术综述. *计算机辅助设计与图形学学报*, 2025, 37(5): 725-743)
- [18] Liu L, Gu J, Zaw Lin K, et al. Neural sparse voxel fields. *Advances in Neural Information Processing Systems*, 2020, 33: 15651-15663
- [19] Yu A, Li R, Tancik M, et al. Plenotrees for real-time rendering of neural radiance fields//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, Canada, 2021: 5752-5761
- [20] Fridovich-Keil S, Yu A, Tancik M, et al. Plenoxels: Radiance fields without neural networks//*Proceedings of the IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 5501-5510
- [21] Sun C, Sun M, Chen H T. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA, 2022: 5459-5469
- [22] Barron J T, Mildenhall B, Tancik M, et al. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021: 5855-5864
- [23] Verbin D, Hedman P, Mildenhall B, et al. Ref-nerf: Structured view-dependent appearance for neural radiance fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 5481-5490
- [24] Chen A, Xu Z, Geiger A, et al. Tensorf: Tensorial radiance fields//Proceedings of the European Conference on Computer Vision. Cham: Springer Nature Switzerland, Tel Aviv, Israel, 2022: 333-350
- [25] Chen A, Xu Z, Wei X, et al. Factor fields: A unified framework for neural fields and beyond. arXiv preprint, 2023: 01226
- [26] Müller T, Evans A, Schied C, et al. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (TOG), 2022, 41(4): 1-15
- [27] Yu Z, Chen A, Huang B, et al. Mip-splatting: Alias-free 3d gaussian splatting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 19447-19456
- [28] Zhang J, Zhan F, Xu M, et al. Fregs: 3d gaussian splatting with progressive frequency regularization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 21424-21433
- [29] Jiang S, Jiang C, Jiang W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 167: 230-251
- [30] Cheng K, Long X, Yang K, et al. GaussianPro: 3D Gaussian splatting with progressive propagation//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria, 2024: 8123-8140
- [31] Jung J, Han J, An H, et al. Relaxing accurate initialization constraint for 3d gaussian splatting. arXiv preprint arXiv: 2403.09413, 2024
- [32] Guédon A, Lepetit V. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 5354-5363
- [33] Huang B, Yu Z, Chen A, et al. 2d Gaussian splatting for geometrically accurate radiance fields//Proceedings of the ACM SIGGRAPH 2024 conference papers. Denver, USA, 2024: 1-11
- [34] Dai P, Xu J, Xie W, et al. High-quality surface reconstruction using gaussian surfels//Proceedings of the ACM SIGGRAPH 2024 conference papers. Denver, USA, 2024: 1-11
- [35] Chen D, Li H, Ye W, et al. Pggr: Planar-based Gaussian splatting for efficient and high-fidelity surface reconstruction. IEEE Transactions on Visualization and Computer Graphics, 2024, 31(9): 6100-6111
- [36] Xie Y, Shu F, Rambach J, et al. Planerecnet: Multi-task learning with cross-task consistency for piece-wise plane detection and reconstruction from a single rgb image. arXiv preprint arXiv:2110.11219, 2021
- [37] Yang L, Kang B, Huang Z, et al. Depth anything: Unleashing the power of large-scale unlabeled data//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 10371-10381
- [38] Straub J, Whelan T, Ma L, et al. The replica dataset: A digital replica of indoor spaces. arXiv preprint arXiv:1906.05797, 2019
- [39] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM//Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Türkiye, 2010: 2366-2369
- [40] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 586-595
- [41] Hedman P, Philip J, Price T, et al. Deep blending for free-viewpoint image-based rendering. ACM Transactions on Graphics (ToG), 2018, 37(6): 1-15
- [42] Wang P, Liu L, Liu Y, et al. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. arXiv preprint arXiv:2106.10689, 2021



SUN Chuan-Feng, Ph. D. candidate. His research interests include deep learning and 3D reconstruction.

WU Jun, Ph. D., professor, Ph. D. supervisor, corresponding author. His research interests include photogrammetry

and remote sensing, computer vision, and artificial intelligence.

XU Gang, Ph. D. candidate. His research interests include computer vision.

WANG Ling-Rong, M. S. Her research interests include intelligent transportation.

LUO Ying, Ph. D. candidate. His research interests include computer vision and image processing.

LONG Yu-Lin, M. S. candidate. His research interests include 3D reconstruction.

Background

This paper focuses on 3D scene reconstruction and novel view synthesis, with particular attention to the reconstruction challenges of Lambertian planar regions using explicit Gaussian splatting methods. As an ideal diffuse reflection surface, a Lambertian plane often suffers from insufficient texture information, leading to holes in the initial point cloud, geometric discontinuities, and structural disorder. These issues degrade both reconstruction accuracy and rendering quality. The root cause lies in the limited capability of existing methods to uniformly represent complex textures and planar regions, which poses significant challenges in applications such as robotic navigation, autonomous driving, and virtual reality.

Globally, research on this problem has evolved from implicit neural radiance fields (NeRF) to explicit 3D Gaussian splatting (3DGS). NeRF leverages multilayer perceptrons to implicitly model volume density and color, achieving high-fidelity rendering but at the cost of heavy computation and long training times. Improvements include Mip-NeRF, which introduces multiscale representations for anti-aliasing, and Ref-NeRF, which enhances reflection modeling. TensoRF and Instant-NGP accelerate training via tensor decomposition and hash encoding. However, these implicit methods remain limited in geometric accuracy for planar regions. In contrast, 3DGS explicitly represents scenes with anisotropic ellipsoidal primitives, significantly improving rendering efficiency. Yet, Lambertian planes are still prone to detail loss and discontinuities. Recent advances include Mip-Splatting, which employs progressive multilevel representations to mitigate artifacts; FreGS, which incorporates frequency regularization for densification; SuGaR, which binds primitives to mesh surfaces to enhance accuracy; 2DGS, which compresses primitives into 2D disks for thin-surface modeling; and PGSR, which introduces geometric regularization to achieve unbiased depth estimation. Despite these efforts, existing methods still face depth deficiencies and primitive collapse during optimization

in planar regions. Consequently, overall geometric completeness (IS score) typically remains between 0.6 and 0.8, while depth error (D_RMSE) is about 0.03-0.06 m, falling short of high-precision requirements.

To address this, we propose a 2D-3D hybrid Gaussian splatting framework (LPGS) tailored for accurate Lambertian surface reconstruction. LPGS integrates monocular estimation with point cloud projection through cross-source depth inference to reconstruct a coherent depth field. The scene is partitioned into 2D disks (Lambertian planes) and 3D ellipsoids (complex regions). A dual-plane geometric regularization, consisting of normal consistency and distance adherence, is introduced to constrain structural disorder. Furthermore, a progressive primitive fusion strategy dynamically assimilates ellipsoids to preserve continuity. Experiments on Mip-NeRF360, Tanks & Temples, Replica, and DTU datasets demonstrate that LPGS maintains high rendering quality (PSNR of 31.30), improves IS by 19.7% over 3DGS, and significantly reduces D_RMSE, offering an efficient solution for planar reconstruction.

The contribution of this work lies in enhancing the robustness of scene understanding, thereby supporting the development of digital twins and intelligent perception technologies within China's 14th Five-Year Plan.

This work was supported by the National Natural Science Foundation of China (Grant No. 42361071), 2025 Guangxi Key Research and Development Program (Gui Ke FN2504240020), the Graduate Education Innovation Program of Guilin University of Electronic Science and Technology "Research on Implicit 3D Reconstruction and Application of Stereo Vision Algorithms" (YCBZ2024165), and the Graduate Education Innovation Program of Guilin University of Electronic Science and Technology "Online Processing and Real-Time Distribution of UAV Images for Dynamic Natural Resource Monitoring" (YCBZ2023136).