

中图法分类号: TP751; TP391 文献标识码: A 文章编号: 1006-8961(2026)03-0896-16

论文引用格式: Liu Z, Tang S J, Yao M M, Li Y and Guo R Z. 2026. Fast physically-based inverse rendering framework fusing 3D Gaussian splatting geometry enhancement and efficient material encoding. Journal of Image and Graphics, 31(3):0896-0911(刘铮, 汤圣君, 姚萌萌, 李游, 郭仁忠. 2026. 融合三维高斯泼溅几何增强与高效材质编码的快速物理逆渲染框架. 中国图象图形学报, 31(3):0896-0911)[DOI: 10. 11834/jig. 250257]

融合三维高斯泼溅几何增强与高效材质编码的快速物理逆渲染框架

刘铮^{1,2}, 汤圣君³, 姚萌萌¹, 李游^{1,3*}, 郭仁忠^{1,3}

1. 人工智能与数字经济广东省实验室(深圳), 深圳 518083; 2. 深圳大学计算机与软件学院, 深圳 518060;
3. 深圳大学建筑与城市规划学院、智慧城市研究院, 深圳 518060

摘要: 目的 逆渲染旨在从二维多视图图像中同时恢复场景几何、材质及光照。近期, 三维高斯泼溅(3D Gaussian splatting, 3DGS)因其高效渲染特性被引入该领域, 然而, 当前基于物理真实的逆渲染时面临两大核心挑战: 其一, 3DGS基元本身主要为新视角合成优化, 其提取的网格难以满足物理渲染的精度需求; 其二, 准确解耦材质与光照依赖对复杂光照传输和高频材质细节的精确建模, 但现有方法在估计具有复杂可见性的直接光照时常面临高方差与计算瓶颈, 影响了材质恢复的保真度和训练效率。方法 为此, 提出一种两阶段快速物理逆渲染框架: 首先, 在几何恢复阶段, 引入扁平高斯基元压缩与多视图双向重投影误差约束, 实现精度与速度的平衡, 生成可直接用于下游渲染引擎的高精度三角网格; 其次, 在材质与光照恢复阶段, 在提取的网格上采用基于单样本加权池采样的高效直接光照估计, 并基于多分辨率哈希网格的神经表示实现复杂高频材质细节的恢复, 在大幅降低渲染方差的同时显著缩短训练时间。结果 为全面验证本文方法的有效性, 本研究在基准数据集上开展了系统实验。在几何恢复方面, 本文方法在TensoIR(tensorial inverse rendering)数据集上的法线平均角误差相比次优方法降低了19.59%; 在DTU(Technical University of Denmark)数据集上, 生成的网格在倒角距离分数上与最优方法持平, 但训练速度提升了一倍。在材质恢复和新视角合成任务上, 本文方法同样表现出色: 在TensoIR数据集中, 材质恢复的峰值信噪比(peak signal-to-noise ratio, PSNR)值较次优方法提升了2.84%, 新视角合成的PSNR值提高了0.08。结论 本工作成功构建了从三维高斯泼溅表达到可物理渲染的网格与材质贴图快速、端到端逆渲染流程, 为逆渲染技术在实时交互与工业级场景中的应用提供了高效且鲁棒的新范式。

关键词: 逆渲染; 三维高斯泼溅(3DGS); 三维重建; 重投影误差; 加权池采样(WRS)

Fast physically-based inverse rendering framework fusing 3D Gaussian splatting geometry enhancement and efficient material encoding

Liu Zheng^{1,2}, Tang Shengjun³, Yao Mengmeng¹, Li You^{1,3*}, Guo Renzhong^{1,3}

1. Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen 518083, China; 2. College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China; 3. School of Architecture and Urban Planning, Research Institute for Smart Cities, Shenzhen University, Shenzhen 518060, China

收稿日期: 2025-06-09; 修回日期: 2025-09-10; 预印本日期: 2025-09-17

* 通信作者: 李游 liyou@gml.ac.cn

基金项目: 广东省自然科学基金项目(2023A1515010717, 2024A1515030061); 国家自然科学基金项目(41901329, 42471442); 国家重点研发计划资助(2022YFC3800602)

Supported by: Natural Science Foundation of Guangdong Province, China (2023A1515010717, 2024A1515030061); National Natural Science Foundation of China(41901329, 42471442); National Key R&D Program of China(2022YFC3800602)

Abstract: Objective Inverse rendering aims to recover the underlying three-dimensional geometry, surface material properties, and illumination parameters of a scene from a collection of two-dimensional images. This capability can be applied in a wide range of tasks—from accurate scene reconstruction in autonomous driving and robotic navigation, to photorealistic asset creation for virtual reality and game engines, and even post-production workflows in film and visual effects. Traditional inverse rendering methods, including multiview stereo, depth-based reconstruction, and neural radiance field (NeRF) approaches, typically face a trade-off between reconstruction accuracy, material fidelity, and computational cost. 3D Gaussian splatting (3DGS) has emerged as a highly efficient and flexible rendering approach, representing scenes as collections of anisotropic Gaussian primitives rather than dense volumetric grids or explicit meshes. However, when applied directly to inverse rendering, 3DGS-based techniques can suffer from misaligned normal directions on reconstructed surfaces and spurious artifacts in recovered material maps, limiting their practical utility. **Method** In this work, we propose a novel optimization framework that integrates multiview geometric consistency constraints with a weighted pool sampling strategy to address these limitations. Our approach comprises two complementary modules: a geometry enhancement pipeline that redefines surface normals through compressed Gaussian primitive analysis and enforces consistency across multiple camera views, and a material-lighting decomposition pipeline that accelerates direct illumination computation via a dynamic weighted sampling pool while employing multiple importance sampling for indirect lighting. By iteratively coupling these two modules, we achieve state-of-the-art performance in geometric accuracy and photorealistic material recovery, all without sacrificing the rendering efficiency of 3DGS representation. The first step in creating the geometry enhancement module is to observe that raw 3DGS reconstructions define normal vectors solely on the basis of the local Gaussian covariance axes, which can diverge from true surface orientations when primitives exhibit near-spherical shapes or when sampling is sparse. To remedy this issue, we propose a shape-driven normal redefinition strategy; we compress each Gaussian primitive along its shortest axis and define the normal direction as pointing along that shortest axis toward the camera view direction. This shape-driven normal initialization is then refined by incorporating a multiview reprojection consistency term, in which we measure the discrepancy between its projections under all available camera poses and enforce alignment through a robust optimization objective for each 3D point on the reconstructed surface. This constraint not only corrects directional drift in the normals but also suppresses outlier primitives that would otherwise introduce noise or spurious geometry. For material and lighting recovery, direct illumination estimation remains a bottleneck in inverse rendering, especially for large-scale or complex outdoor scenes with high-resolution textures. We therefore design a weighted pool sampling strategy that maintains a small, adaptive set of representative light rays—each weighted by its contribution to the final pixel intensity—and updates this pool dynamically during optimization. By reusing and re-weighting these samples across iterations, our method rapidly approximates the direct lighting integral with minimal memory overhead. To capture indirect illumination—including soft shadows, inter-reflection, and global light transport—we incorporate a multiple importance sampling framework that blends light-source sampling, cosine-weighted hemisphere sampling for diffuse components, and GGX microfacet sampling for specular highlights. The combined strategy yields accurate shading and material estimates while maintaining a throughput that is competitive with forward-only 3DGS rendering. **Result** We validate our approach on three widely adopted public benchmarks: the TensoIR dataset, the DTU multiview stereo collection, and the TNT scene suite. On TensoIR, our geometry module reduces the average angular error of surface normals by 19.6% compared with the next best 3DGS-based inverse rendering method, according to standard evaluation protocols. On DTU and TNT, we achieve mesh reconstruction quality—measured by Chamfer distance and F1 score—that matches or exceeds the top-performing baselines while boosting overall scene processing speed by approximately 50%. For material recovery and novel view synthesis on TensoIR, material PSNR increased by 2.8% and synthesis PSNR increased by 0.08 dB over the second-best approach. Compared with NeRF methods, our full pipeline delivers an end-to-end speedup of roughly 60%, demonstrating that high-fidelity inverse rendering can be both accurate and efficient. To isolate the contributions of each design choice, we conduct an extensive ablation study. **Conclusion** Despite these advances, certain limitations remain. Our current implementation assumes static scenes and fixed lighting during capture, excluding dynamic object interactions or time-varying illumination. Highly reflective or translucent materials can still challenge the weighted pool sampler's coverage, resulting in subtle bias in specular lobes. Furthermore, while our method scales well to moderately

complex scenes, extremely large-scale outdoor environments may require hierarchical partitioning or out-of-core streaming for real-time performance. We thus envision future extensions that incorporate temporal coherence constraints for video-based inverse rendering, adaptive sampling strategies for specialized material classes (e. g., subsurface scattering), and integration with custom GPU kernels to push toward interactive rates. In summary, we present a comprehensive inverse rendering pipeline built on the 3DGS foundation, enriched by multiview geometric consistency and a novel weighted sampling approach for lighting decomposition. Our method bridges the gap between accuracy and efficiency, delivering superior geometry, material, and illumination recovery on benchmark datasets while maintaining high rendering throughput. This work paves the way for practical, scalable inverse rendering solutions in applications ranging from autonomous perception to immersive content creation.

Key words: inverse rendering; 3D Gaussian splatting(3DGS); 3D reconstruction; re-projection error; weighted reservoir sampling(WRS)

0 引言

逆渲染通过多视角图像恢复场景几何、材质和光照,在影视、游戏、VR/AR(virtual reality/augmented reality)和自动驾驶中具有广泛应用前景。其核心挑战在于:有限观测可对应多种三维解释,不同几何、光照和材质可能产生相同结果,使问题高度不适定。传统方法常依赖强先验以缓解不确定性,限制了其适用性,且常面临效率瓶颈。

近年来,三维高斯泼溅(3D Gaussian splatting, 3DGS)(Kerbl等,2023)凭借高效渲染和高质量视图合成优势,大力推动了三维表示与重建技术发展,并在逆渲染中展现出潜力。与神经辐射场(neural radiance fields, NeRF)(Mildenhall等,2022)的体渲染不同,3DGS采用显式高斯基元,避免了高昂计算开销。然而,其基元属性直接编码视角相关外观,难以在逆渲染中实现光照与材质的物理解耦。现有工作(Liang等,2024;Gao等,2025)虽然提升了效率,但在几何精度、光照建模与材质解耦准确性方面仍有限:一方面,原始3DGS的隐式几何精度不足,难以作为物理渲染的可靠基底,从而影响材质估计并产生伪影;另一方面,逆渲染需同时估计复杂光照和固有材质,传统蒙特卡洛光照估计在有限采样下噪声大,简化模型则损失真实感。此外,高效学习空间变化的复杂材质也对表示方法提出更高要求,传统多层感知机(multilayer perceptron, MLP)结构往往训练缓慢、细节不足。

为此,本文提出两阶段快速物理逆渲染框架。第1阶段从多视图图像重建高质量几何基底,引入扁平高斯基元压缩与法线重定义,并结合多视图重

投影约束,显著提升几何精度与一致性。第2阶段在此几何基底上进行材质与光照恢复:采用单样本加权池采样降低光照估计方差,并以多分辨率哈希网格高效建模材质细节。两阶段协同有效减少噪声,加速收敛,并提升材质与光照的物理真实性。

本文贡献总结如下:1)提出一种新的两阶段快速物理逆渲染框架,该框架首先从3DGS优化并提取出高精度的表面网格,再在网格上进行高效、物理真实的材质与光照分解;2)针对3DGS几何估计,提出一种面向扁平高斯基元的法线估计新方法,并引入双向重投影误差进行多视图几何一致性约束,实现了精度与效率的平衡;3)在光照与材质恢复阶段,创新性地引入加权池采样机制引入直接光照估计,有效缓解低采样率下的方差问题,同时结合多分辨率哈希网格构建材质神经表达模型,实现对高频细节的精细建模与快速收敛。

1 相关工作

1.1 基于神经辐射场的逆渲染研究

神经辐射场(NeRF)自提出以来,因高精度新视角合成而广受关注。其可微体积渲染过程与逆渲染求解天然契合,应用于逆渲染已然成为趋势(曾志鸿等,2025)。NeRV(Srinivasan等,2021)在已知照明下估计法线;NeRD(Boss等,2021)结合微表面反射模型实现物理一致性解耦;NeRFactor(Zhang等,2021b)通过法线平滑约束显著提升解耦质量;PhysSG(Zhang等,2021a)引入球面高斯表示环境光照,虽然受各向同性假设限制,却开辟了新的光照建模思路。近期的TensorIR(Jin等,2023)基于张量分解表示并结合光线追踪,有效提升了质量与效率。

Nvdifrec (Munkberg 等, 2022) 及其拓展 Nvdifrecmc (Hasselgren 等, 2022) 构建了可微分渲染框架, 可从图像直接重建网格、估计 BRDF (bidirectional reflectance distribution function) 材质与环境光照。前者利用 DMTet (Shen 等, 2021) 提取高质量网格并通过近似加速渲染方程求解, 后者则引入物理真实的蒙特卡罗光线追踪, 并结合噪声抑制机制, 兼顾解耦精度与优化稳定性。这些工作为逆渲染提供了完整的端到端方案, 但在效率与细节恢复上仍有提升空间。本文在其基础上进一步改进材质与光照分解策略。

1.2 基于 3DGS 的逆渲染研究

近期, 3DGS 技术逐渐成为逆渲染研究的新热点 (吕恒焯 等, 2025)。其显式几何建模与高效渲染管线为逆渲染提供了新的可能, 但原始 3DGS 仅表示视角相关颜色, 缺乏对材质与光照的完整建模。GS-IR (Liang 等, 2024) 和 Relightable 3D (Gao 等, 2025) 率先将几何与 BRDF 参数绑定至高斯基元, 并通过可见性建模分离阴影, 但光照烘焙近似与伪法线监督限制了几何与材质精度。为提升几何重建质量, 2DGS (2D Gaussian splatting) (Huang 等, 2024) 引入法线一致性损失, PGSR (planar-based Gaussian splatting reconstruction) (Chen 等, 2025a) 通过引入单视图几何正则化、多视图几何一致性约束, 以及曝光补偿模型, 以提升几何一致性和重建精度。尽管其采用平面基元进行重建的思路富有启发, 但直接替换基元可能会牺牲原生 3DGS 的灵活性。本文方法受到这些工作启发, 但采用了不同的技术路径: 通过在优化过程中对标准三维高斯基元施加压缩约束, 引导其动态地形成扁平化结构, 从而在保持高效性的同时, 为后续物理渲染提供高质量的几何基础。

在光照建模方面, 3DGS 的间接光照处理能力有限, 早期方法常将其预烘焙至基元, 难以适应动态光照。GI-GS (Chen 等, 2025c) 通过延迟着色与路径追踪结合, 显著改善了间接光照还原。针对光泽材质, GlossyGS (Lai 等, 2025) 利用微表面几何先验与法线预过滤缓解材质歧义, 提高了镜面反射区域的分解准确性。

综上, 基于 3DGS 的逆渲染研究在几何、光照和材质方面不断改进, 已实现快速建模与直接光照分解, 并通过延迟着色、路径追踪及微表面建模提升了复杂光照与材质的重建质量, 为高保真逆渲染奠定

了基础。

2 方法

2.1 方法总述

如图 1 所示, 本文提出一种两阶段物理逆渲染方法: 第 1 阶段, 基于三维高斯泼溅 (3DGS) 快速拟合场景辐射场, 并在此过程中引入法线一致性约束与多视角几何正则化, 以增强内在几何结构的鲁棒性, 随后提取出三角网格; 第 2 阶段, 以三角网格作为基础, 采用基于物理的光线追踪框架求解渲染方程, 对表面材质 (BRDF 参数) 与全局光照环境进行联合估计与优化。鉴于原生 3DGS 缺乏显式几何约束, 直接用于逆渲染时难以保证法线精度与多视角一致性, 本文在高斯场训练中额外引入几何正则化项, 并在网格光追阶段重用优化所得的精确法线信息, 从而有效克服了 3DGS 在几何恢复上的天然局限。针对传统直接光照估计中的噪声与计算瓶颈问题, 本文进一步提出基于加权池采样的直接光照估计策略, 以集中采样于高贡献方向, 显著降低估计方差并提升计算效率, 最终实现高效、高精度的逆渲染。

2.2 几何估计

为解决前述 3DGS 基元直接提取的几何信息 (尤其是法线) 在精度上难以满足物理渲染需求的挑战, 本阶段提出了一套几何估计的联合优化方法。

2.2.1 基于扁平高斯基元的法线估计

首先, 针对原始 3DGS 基元因其各向异性或近乎各向同性的形状导致表面模糊和表面估算不稳定的问题, 本文引入了一种基于扁平高斯基元的法线估计方法。2DGS 方法由于基元设计本身在不同角度观察时, 目标基元始终是同一平面, 避免了这一问题, 几何渲染的结果较 3DGS 有所提升。然而, 纯 2D 高斯表示方法计算复杂度高, 降低了几何恢复速度。

为此, 本文引入了一种高斯基元的压缩方法。与 PGSR 等直接采用平面基元或 2DGS 采用纯 2D 高斯不同, 本文方法核心思想并非替换 3DGS 的基元, 而是在优化中动态地重塑 (reshape)。本文借鉴 NeuSG (Chen 等, 2025b) 和 PGSR (Chen 等, 2025a) 的思路, 将高斯基元沿着最短轴方向压缩, 并定义朝向入射光的最短轴方向是法线方向。通过沿特定轴压缩比例因子, 高斯椭球体可以被压平成与这些轴对

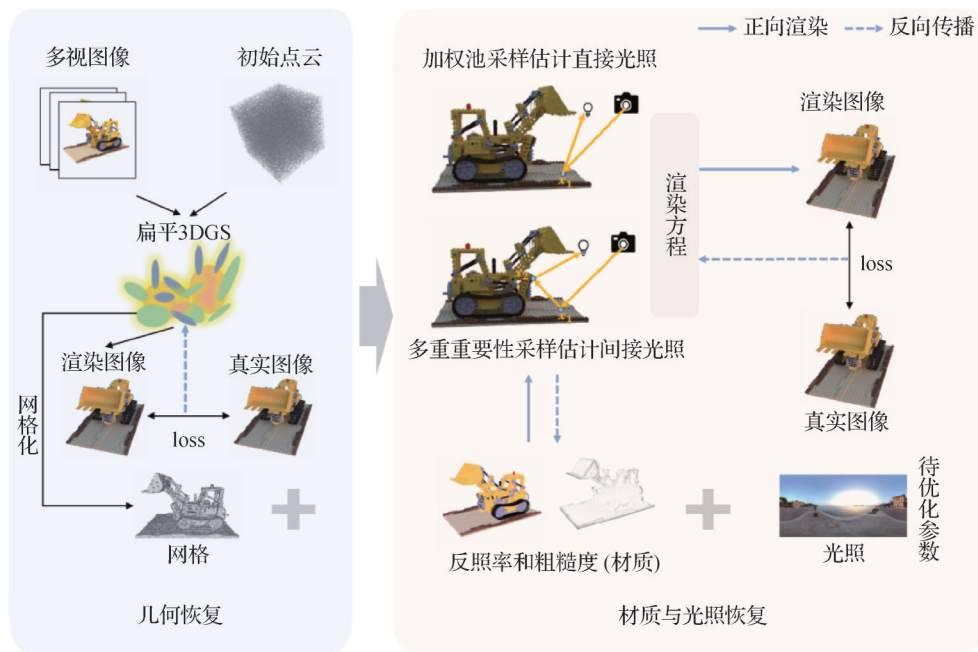


图1 本文提出的逆渲染框架整体流程图

Fig. 1 Overview of the proposed inverse rendering framework

齐的平面。最小化每个高斯基元的缩放矩阵 $S(s_i) = \text{diag}(s_1, s_2, s_3)$ 中的比例因子, 具体为

$$L_s = \|\min(s_1, s_2, s_3)\|_1 \quad (1)$$

式中, $\min(s_1, s_2, s_3)$ 表示选择其中最小的尺度因子, 使其趋向于零, 从而将高斯椭球压缩成一个极薄的平面状结构。然后计算其 L_1 范数, 即取绝对值的总和, 确保优化过程中该值逐渐减小。由于3DGS生成的点云通常不在表面上, 而是分布在表面周围, 缩小最小尺度可以使点云更加紧密地贴合目标表面, 提高几何精度。

同时, 最小比例因子所在的方向定义为法线 n_i 。为进一步消除法线双向不确定性, 本文假定观察方向与法线方向之间的角度始终大于 90° 。引用PGSR中的方法, 当前视点下的最终法线贴图通过阿尔法混合计算得到, 具体为

$$N = \sum_{i \in N} R_c^T n_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (2)$$

式中, R_c 是相机坐标到世界坐标的旋转矩阵, α 是不透明度。进一步, 还需使用下式计算深度图作为下一步的必要参数。具体为

$$D(p) = \frac{D}{N(p)K^{-1}\tilde{p}} \quad (3)$$

式中, D 表示视点到平面的距离, $p = [u, v]^T$ 是图像平面上的2D位置。 \tilde{p} 表示 p 的齐次坐标, K 是相机

内参。有了深度和法线贴图, 便于引入多视图正则化。

与2DGS不同, 本文并未采用纯2D平面高斯代替3D高斯, 而是压缩原本的3D高斯基元, 使其表现为近似平面的扁平状, 保持高精几何恢复的同时降低了计算量。这种对原生3DGS“微创式”的正则化方法, 避免了引入全新基元表示所带来的额外计算复杂性, 是框架在第1阶段的训练速度能够比PGSR快一倍的关键(具体效率对比见3.3节)。

2.2.2 基于双向重投影误差的多视图几何一致性约束

传统基于3DGS几何估计的另一个缺陷是缺乏多视图几何一致性约束。3DGS在像素射线和高斯基元之间的交点处评估高斯值, 但在不同视角下, 3DGS会得到不同的观察结果。同时, 其相交平面也不一致, 即缺乏多视图一致性, 导致法线和深度计算不准确。

为进一步确保从3DGS重建的几何具备跨视角的一致性, 克服单一视角下几何推断的歧义性, 本文针对性引入了基于双向重投影误差的多视图几何一致性约束。本文通过传统MVS(multi-view stereo)方法中的重投影误差为3D高斯基元添加几何一致性约束, 提高几何恢复精度。重投影误差法通过将三维点投影到不同视图的二维图像平面上, 并计算投

影点与真实观测点之间的误差(即重投影误差),以优化三维几何结构和相机参数,从而保证多视图之间的一致性。

重投影误差法原理如图2所示。场景中的点 p_1 被投影到两个不同的视图上,得到两个的像素位置 x_1 和 y_1 ,然后根据这两个像素位置可以重新计算得到一个新的点位置 p_2 ,这个点的位置由于多视图下的偏差,往往与 p_1 本身不重合。然后把 p_2 重新投影到其中一个视图上,得到像素位置 y_2 ,像素位置之间的误差称为重投影误差 L_{geo} ,而本文方法将利用这个像素重投影误差对训练结果施加约束。其过程如下:

在一组随机视角下,首先依据式(2)和式(3)为参考视角和相邻顺序视角计算法线 \mathbf{n} 和到相机的平面距离 d (即深度信息 D);对于参考视角中的特定像素 p_a ,需要为之施加一个齐次变换矩阵 \mathbf{H} ,使其映射到相邻视角的对应像素位置 p_b , \mathbf{H} 的计算方法为

$$\mathbf{H}_{ab} = \mathbf{K}_b (\mathbf{R}_{ab} - \frac{\mathbf{T}_{ab} \mathbf{n}_a^T}{d_a}) \mathbf{K}_a^{-1} \quad (4)$$

式中, \mathbf{R}_{ab} 和 \mathbf{T}_{ab} 分别是两个坐标系之间的齐次变换矩阵, \mathbf{K}_b 和 \mathbf{K}_a 分别是相机内参。使用同样的方法也可以计算出反方向的齐次矩阵 \mathbf{H}_{ba} 。最终可以得到每一对相邻视角的齐次变换矩阵。最小化前向和后向投影误差构成了多视图几何一致性正则化。具体为

$$L_{geo} = \frac{1}{V} \sum_{p_a \in V} \phi(p_a) \quad (5)$$

式中, V 表示所有有效像素的集合, $\phi(p_a) = \|p_a - \mathbf{H}_{ba} \mathbf{H}_{ab} p_a\|$ 表示像素的相邻顺序视角下的前后投影误差。由于 ϕ 有时会过大,可以认定为存在遮挡而被排除, V 中需要排除掉这些高误差。

2.2.3 损失设置与网格提取

本文方法优化过程中,除了前两个小节提到的两个几何损失,还保留了3DGS的原有RGB损失。具体为

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

式中, L_1 损失是保证渲染图像与真实图像一致约束,而 L_{D-SSIM} 损失要求渲染图像具有与真实图像相似的结构。

本文的总体损失函数 L 由图像重建损失 L_{rgb} 、压缩3D高斯损失 L_s 、几何损失 L_{geo} 共同组成。具体为

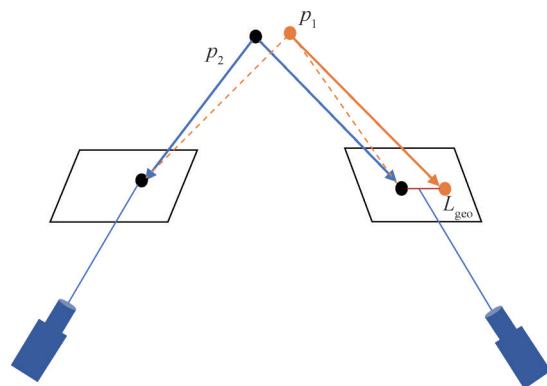


图2 重投影误差原理

Fig. 2 Schematic diagram of the fundamental principle of re-projection error

$$L = L_{rgb} + \lambda_1 L_s + \lambda_2 L_{geo} \quad (7)$$

本文设置 λ_1 为100, λ_2 为0.1, L_{rgb} 中的 λ 项为0.2。

在高斯渲染完成后,需从高斯场中提取出场景网格。传统方法包括泊松重建和TSDF融合(TSDF-fusion)(Zeng等,2017)等。泊松重建对输入噪声十分敏感,且在大规模场景下求解计算开销较大,容易引入过度平滑的伪影。为此,本文采用TSDF融合方法对场景进行网格提取,并用于下一阶段的材质与光照恢复。

2.3 光照计算与材质恢复

2.3.1 基于多分辨率哈希网格的材质编码

为了恢复场景的材质信息,本文引入Disney材质模型的简化版本表示材质,为描述方便,将其表示为

$$f_r = f_{diffuse}(\omega_i, \omega_o) + f_{specular}(\omega_i, \omega_o) \quad (8)$$

即总反射分布函数 f_r 由漫反射项 $f_{diffuse}$ 和镜面反射项 $f_{specular}$ 共同控制。漫反射项的物理意义是表面吸收光照并在各个方向上均匀散射的能力。漫反射光的强度可以表示为

$$f_{diffuse}(\omega_i, \omega_o) = \frac{k_d}{\pi} \quad (9)$$

式中, k_d 是表面漫反射的反射率(albedo), π 是常数。

镜面反射部分的关键组件包括法线分布函数(D)、几何遮蔽函数(G)以及菲涅尔项(F)。具体为

$$f_{specular}(\omega_i, \omega_o) = \frac{D(h) \cdot G(\omega_i, \omega_o) \cdot F(\omega_o)}{4 \cos \theta_i \cos \theta_o} \quad (10)$$

式中,法线分布函数为

$$D(h) = \frac{\alpha^2}{\pi [(\cos \theta_h)^2 (\alpha^2 - 1) + 1]^2} \quad (11)$$

式中, α 为粗糙度(roughness)参数, 控制表面微结构的分布宽度。 θ_h 为半程向量与法线的夹角。 α 值越小, 表面越光滑, 反射高光越集中; α 值越大, 表面越粗糙, 高光越分散。其他项则是关于输入输出光线方向的函数, 此略。最终, 本文将计算中间值中的反射率 k_d 和粗糙度 α 作为可以以图像形式展示的被优化参数。

在材质恢复阶段, 本文以 Nvdiffrmc 为基础, 对场景材质进行重建。由于原文方法训练速度较慢且无法准确建模高频细节, 本研究引入多分辨率哈希表加速技术替代传统 MLP (Müller 等, 2022), 实现了以小型多层感知机结合哈希网格的形式对材质参数进行编码。

图3为本文采用的基于多分辨率哈希网格与小型多层感知机(MLP)相结合的材质编码结构示意图。输入为场景中某一三维空间位置, 该位置在多个分辨率层的哈希网格中查表获得对应特征向量, 并拼接后输入至浅层 MLP 中进行非线性映射, 最终输出包含反射率和粗糙度等材质参数。该编码结构同时兼具表达能力强与计算效率高的特点, 能有效支持复杂高频材质细节的恢复。

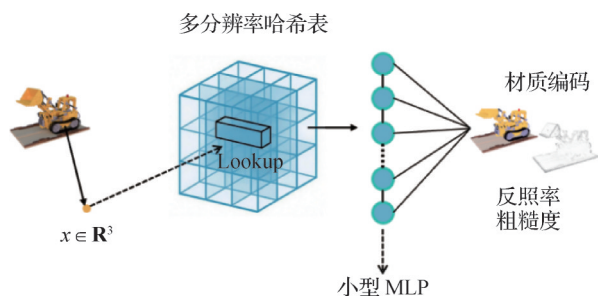


图3 材质参数编码结构

Fig. 3 Material encoding structure

2.3.2 基于加权池采样的直接光照计算

传统蒙特卡洛方法(包括多重重要性采样)在低样本数下估计直接光照时, 常因采样分布与被积函数匹配不佳而导致结果充满噪声且收敛缓慢。为克服这一难题并提升渲染效率, 本文采用了基于加权池采样(weighted pool sampling, WRS)的直接光照估计策略。

计算渲染方程是整个材质光照恢复流程中的核心步骤。计算渲染方程需要分为直接光照和间接光照分别计算, 最后相加得到最终的图像。具体为

$$I_{\text{final}}(x) = L_{\text{dir}}(x) + L_{\text{ind}}(x) \quad (12)$$

式中, $I_{\text{final}}(x)$ 是最终渲染图像在像素 x 处的颜色值。 L_{dir} 和 L_{ind} 分别是直接光照和间接光照, 需要分别计算。

对于每个场景中的点, 直接光照来自场景中的光源。对于一个表面点 x 的直接光照可以通过下式计算, 具体为

$$L_{\text{dir}}(x, \omega_o) = \int_{\Omega} V(x, \omega_i) L_{\text{env}}(\omega_i) \times f_r(x, \omega_i, \omega_o, m)(\omega_i \cdot n) d\omega_i \quad (13)$$

式中, V 为可见性项, 用于判断从表面点 x 沿入射方向 ω_i 发出的射线是否被场景中的其他物体遮挡。该射线可表示为: $r_i(t) = x + t\omega_i$ 。如果射线与场景中的其他物体相交, 则 $V(x, \omega_i) = 0$, 表示射线被遮挡; 如果射线没有相交任何物体, 则 $V(x, \omega_i) = 1$, 表示射线没有被遮挡, 可以接收到光源的光照。 L_{env} 被初始化为一个高动态范围噪声图像, 以“.exr”文件保存。 f_r 为材质, 由式(8)计算, m 为材质参数, 是粗糙度和反照率的组合。最后一项为入射光与法线夹角的余弦值, 表示光的投影。通过对半球方向 Ω 上所有入射光方向进行积分, 可以得到表面点 x 在出射方向 ω_o 上的直接光照结果。

由于该积分本身是一个复杂的非线性方程, 直接计算非常困难, 因此本文将积分转换为离散形式使用采样的方法对积分结果进行估计。离散形式表达为

$$L_{\text{dir}}(x, \omega_o) \approx \frac{1}{N} \sum_{i=1}^N \frac{V(x, \omega_i) L_{\text{env}}(\omega_i) f_r(x, \omega_i, \omega_o, m)(\omega_i \cdot n)}{p_{\text{dir}}(\omega_i)} \quad (14)$$

式中, N 表示总样本个数, 而 $p_{\text{dir}}(\omega_i)$ 则表示某种特定的概率分布函数, 如均匀分布或随机分布等。在基线方法 Nvdiffrmc 中使用的是多重重要性采样, 这种采样方法会组合多个简单分布采样, 从而降低单一分布采样带来的方差。

但重要性采样在低样本条件下依然具有较大的方差, 体现在图像上即会产生较大的噪声。如图4所示, 在低样本条件下, 使用多重重要性采样方法(左侧图像)相比真实值会有明显的噪点。

另外, 多重重要性采样需要多次采样代入公式计算并求平均值, 这也增加了时间成本。为此, 本文引入加权池采样方法(Bitterli 等, 2020), 降低方差并提高计算效率, 从而得到更加精确的直接照明估算。

本文方法在理论上具有良好的统计特性。基于

加权池的重采样在前向直接光照估计中是一种无偏(unbiased)估计器,其期望值等于真实的光照积分(Bitterli等,2020),并且通过降低估计方差而加快收敛。这里的“无偏与收敛”仅限于前向积分估计层面,并不直接意味着逆渲染参数估计的无偏性或一致性。在逆渲染语境下,关于梯度和参数估计的性质,Chang等人(2023)在文中进行了系统讨论。本文方法与该方法兼容互补,但目前主要通过提升前向估计的稳定性改善参数优化过程。

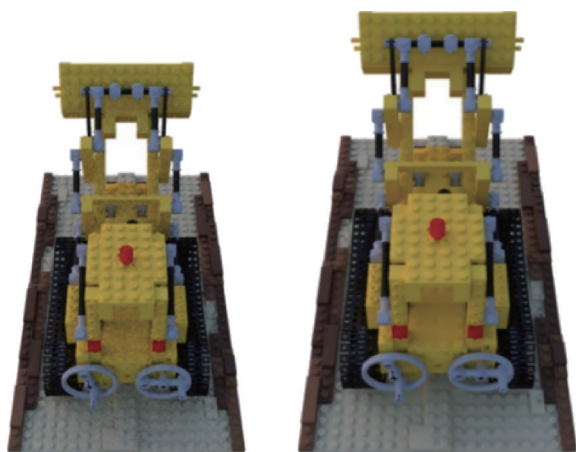


图4 重要性采样在低样本条件下渲染效果对比

Fig. 4 Rendering quality comparison with importance sampling under low sample counts

具体方法是:首先另外构造一个与 L_{env} 成比例的概率分布函数 $q_{\text{dir}}(\omega_i)$,并给这个分布创建 M 个样本 $S = \{\omega_1, \omega_2, \omega_3, \dots, \omega_M\}$,然后为每个样本 ω_j 分配一个权重,具体为

$$\beta_j = \frac{p_{\text{dir}}(\omega_j)}{q_{\text{dir}}(\omega_j)} \quad (15)$$

最后,依照权重来重新采样样本集合 S 。因此最终计算直接光照的表达式为

$$L_{\text{dir}}(x, \omega_o) \approx \frac{1}{N} \sum_{i=1}^N \left(\frac{f_r(\omega_i)}{p_{\text{dir}}(\omega_i)} \frac{1}{M} \sum_{j=1}^M \frac{p_{\text{dir}}(\omega_j)}{q_{\text{dir}}(\omega_j)} \right) \quad (16)$$

在计算时,并不会一次保存所有样本的权重然后排序,而是维护一个小型“池子”(reservoir),其中只存储当前抽中的样本,随着新样本流入,逐步以权重概率更新池子中的样本。这种加权池采样方法称为单样本重要性重采样,它使用选中的样本及其权重计算光照估计值,通过集中在高权重样本上采样,相比于基线方法的原始多重重要性采样,减少了低权重样本对估计值的影响,在降低方差的同时提高

采样效率。

2.3.3 基于多重重要性采样的间接光照计算

间接光照来自于其他表面反射的光。本文通过多次反弹路径追踪来模拟这些间接光照反射的过程。间接光照的计算式为

$$L_{\text{ind}}(x, \omega_i) = \int_{\Omega} L_i(\hat{x}, \omega'_i) f_r(\hat{x}, \omega'_i, \omega_o, \hat{m})(\omega'_i \cdot \hat{n}) d\omega'_i \quad (17)$$

这是一个递归的方程,每一次的结果 $L_{\text{ind}}(x, \omega_i)$ 将会代入到下一个方程中的 $L_i(\hat{x}, \omega'_i)$ 。理论上计算间接光照需要无限次的反弹,会导致无限递归。但由于内存和时间限制,本文只计算前3次反弹。与上一小节所述相同,这个积分需要使用蒙特卡洛积分方法计算。在这一节为了计算间接光照,本文直接使用了基线方法中的多重重要性采样方法进行估算。具体为

$$L_{\text{ind}}(x, \omega_i) \approx \frac{1}{N} \sum_{k=1}^N \frac{L_{\text{ind}}(x, \omega_i^k) f_r(x, \omega_i^k, \omega_o, m)(\omega_i^k \cdot n)}{p_{\text{ind}}(\omega_i^k)} \quad (18)$$

式中, k 表示第 k 次反弹, $k \in 1, 2, 3$ 。多重重要性采样的分布 $p_{\text{ind}}(\omega_i)$ 的构建方法与Nvdiffrecmc相同,方法如下:

首先,分别构建3个分布。光源重要性采样分布 $p_{\text{light}}(\omega_i)$,漫反射余弦采样分布 $p_{\text{diffuse}}(\omega_i)$,以及镜面反射重要性采样分布 $p_{\text{specular}}(\omega_i)$ 。然后,3种采样分布分别采样 M 个样本,并使用多重重要性采样的Balance Heuristic权重。具体为

$$w_i(\omega) = \frac{n_i p_i(\omega)}{\sum_k n_k p_k(\omega)} \quad (19)$$

$$i \in \{\text{light, diffuse, specular}\}$$

将3种采样分布组合成 $p_{\text{ind}}(\omega_i)$ 。由于场景中往往同时存在漫反射、镜面反射、环境贴图3种特征,而单个分布难以覆盖所有情况,导致方差大,因此多重重要性采样的做法是综合上述3种分布,令整体采样效果更稳定。

2.3.4 损失设置

最终通过式(16)和式(18)计算后分别得到结果 $L_{\text{dir}}(x, \omega_o)$ 和 $L_{\text{ind}}(x, \omega_i)$,代入式(12)可以计算得出最终的渲染图 $I_{\text{final}}(x)$,这是整个逆渲染方法中前向传播的核心过程。在反向传播过程中将优化几何参数点偏移量 v 、材质参数反照率 k_d 、粗糙度 α 和光照

参数 L_{env} 。

损失函数分为4个部分,分别是几何项损失、材质项损失、光照项损失和图像损失。

首先介绍几何项损失。本文方法中,不再增减网格点的数量,也不直接优化网格点位置本身,而是优化每个点的偏移 v 。与基线相同,为了保持顶点之间的局部平滑性,本文施加了常见的平滑度损失,具体为

$$L_{\text{smooth}} = \sum_i \sum_{j \in \mathcal{X}_i} \frac{1}{|\mathcal{X}_i|} \left\| (v_i + \Delta v_i) - (v_j + \Delta v_j) \right\|^2 \quad (20)$$

式中, $|\mathcal{X}_i|$ 表示顶点 i 的邻居集合, $\left\| (v_i + \Delta v_i) - (v_j + \Delta v_j) \right\|^2$ 表示经过形变后,顶点 i 与它的邻居 j 之间的距离平方。该损失是通过惩罚所有形变后顶点之间的不平滑度,让一个顶点移动 Δv_i 后,它与邻居顶点的相对位置不要变化太剧烈,避免出现突兀的变形或锯齿状的结构。另外,为了防止偏移量过大,从而维持网格或点云的整体结构,并避免出现不合理的几何变形,本文对顶点施加了 L_2 损失。具体为

$$L_{\text{offset}} = \sum_i \left\| \Delta v_i \right\|^2 \quad (21)$$

该损失惩罚大的偏移量,鼓励每个顶点的变形尽量小。对于材质和光照部分,本文的损失函数和基线 Nvdiffrmc 相同。即

$$L_k = \frac{1}{|X|} \sum_{x_i \in X} |k(x_i) - k(x_i + \varepsilon)|, k \in k_d, \alpha \quad (22)$$

$$L_{\text{light}} = |Y(c_d + c_s) - V(I_{\text{ref}})| \quad (23)$$

式中, $|X|$ 是表面上的世界空间位置集, ε 是一个小的随机偏移向量,式(22)是 k_d 和 α 通用的平滑正则化器。 $Y(x) = (x_r + x_g + x_b) / 3$ 是一个简单的亮度运算符, c_d 和 c_s 分别表示漫反射和镜面反射项,而 $V(x) = \max(x_r, x_g, x_b)$ 是 HSV (hue, saturation, value) 颜色分量。

最后,由于本文使用了基于物理渲染的渲染器并且会生成具有高动态范围的图像,因此目标函数必须对浮点值的整个范围具有鲁棒性。根据先前的基于可微分渲染的工作,本文的图像空间损失 L_{image} 通过 L_1 范数计算渲染图像和目标图像的差异,但不是直接计算,而是先经过对数变换后应用 sRGB 传输函数(Gamma 曲线)进行色调映射。即

$$T(x) = \Gamma(\log(x + 1)) \quad (24)$$

式中, $\log(x + 1)$ 主要用于压缩 HDR (high dynamic range compression) 值,以避免极端值的影响。sRGB 传输函数 $\Gamma(x)$ 负责将线性颜色值转换为感知均匀

的颜色空间。在色调映射之后的颜色空间中,计算 L_1 损失。具体为

$$L_{\text{image}} = \left\| T(x) - T(y) \right\|_1 \quad (25)$$

式中, x 是渲染图像的像素值, y 是目标图像的像素值。

最终,本节优化过程的损失函数为

$$L = \lambda_{\text{smooth}} L_{\text{smooth}} + \lambda_{\text{offset}} L_{\text{offset}} + \lambda_k L_k + \lambda_{\text{light}} L_{\text{light}} + L_{\text{image}} \quad (26)$$

式中,超参数的初始取值与 Nvdiffrmc 相同,分别为 $\lambda_{\text{smooth}} = 0.025$, $\lambda_{\text{offset}} = 0.25$, $\lambda_k = 0.1$, $\lambda_{\text{light}} = 0.15$, 大多数实验都取这些默认权重。

3 实验

3.1 数据集

本文主要用到的逆渲染数据集来源于开源数据集 TensorIR (tensorial inverse rendering) 和 DTU (Technical University of Denmark)。TensorIR 合成数据集使用 Blender 渲染,由4个复杂的合成场景(乐高、热狗、榕树和狢狢)组成。相机设置与 NeRFactor 相同,数据划分设置有100个训练视图、200个测试视图和10个验证视图。对于每个视图, TensorIR 数据集提供了法线贴图、反照率贴图和11个不同光照条件下渲染结果图。DTU 数据集是由丹麦技术大学(DTU)发布的高质量、多视角三维重建数据集,专为计算机视觉和三维重建研究设计。该数据集包含高分辨率的多视角图像、精确的相机参数,以及场景和物体的三维点云和网格真值,为研究者提供了全面的基准。DTU 数据集涵盖多种物体类型和材质,支持在不同光照条件下进行测试,尤其适合验证复杂场景下的算法性能。

3.2 几何估计结果对比

为验证本文几何估计性能,在 TensorIR 数据集上与3种代表性方法进行对比实验,包括基于神经辐射场的 Nvdiffrmc 及基于 3DGS 的逆渲染方法 GS-IR 和 R3DG。鉴于基于 3DGS 的工作无法直接生成网格,实验对其在特定视角下生成的法线图质量进行评估,评估指标为法线平均角误差(mean angular error, MAE)。

定量结果如表1所示,本文法线 MAE 指标上表现最优,误差为 3.98,显著优于其他方法,证明了本

文提出的扁平高斯法线估计与多视图一致性约束有效克服了原始3DGS几何在法线精度和稳定性上的缺陷。图5展示了各方法的定性对比结果。其中Nvdiffrmc方法的重建效果较为粗糙,GS-IR方法和R3DG方法的法线恢复表现优于Nvdiffrmc方法。然而,这两种方法存在平滑度不足的问题,例如,在热狗场景中,光滑盘子表面法线图的颜色分布不均匀,表明与真实图像存在一定偏差。本文方法

表1 法线平均角误差的定量比较

Table 1 Quantitative comparison of MAE of normals

方法	法线 MAE ↓
Nvdiffrmc	5.05
R3DG	4.97
GS-IR	4.95
本文	3.98

注:加粗字体表示最优结果。

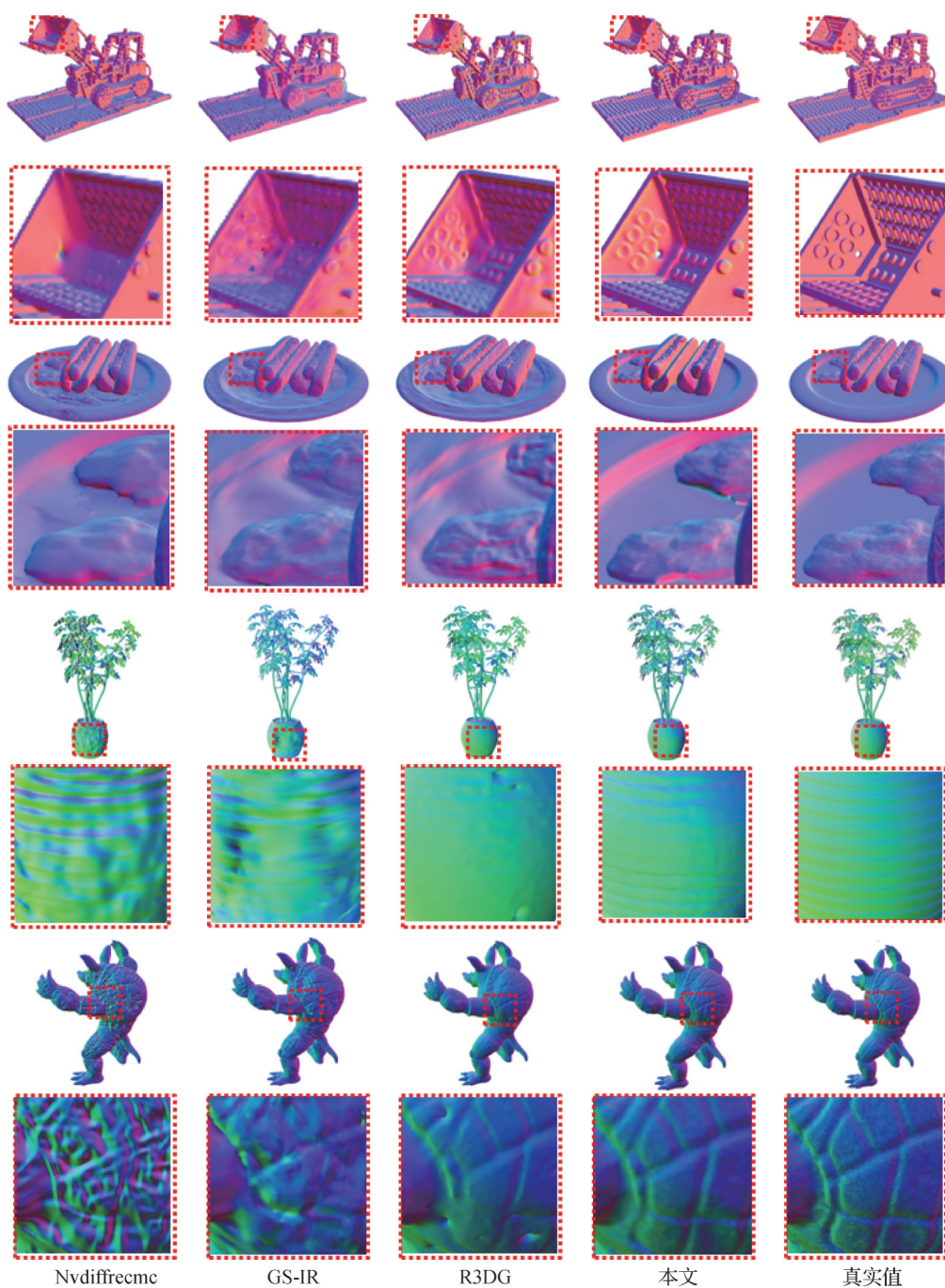


图5 TensorIR数据集下法线图的定性比较

Fig. 5 Qualitative comparison of normal on TensorIR dataset

在图中细节部分都展示了与真实值的一致性。

综上所述,本文方法在 TensoIR 数据集上的表现显著优于现有方法,在细节保持能力、表面光滑度和精细度方面取得了良好的平衡。尤其是在物体边缘、表面细节以及复杂纹理的恢复上,本文方法均展现出了更优的性能,进一步证明了其有效性和适用性。

3.3 网格质量对比

由于本文框架不仅依赖高质量的几何估计,还需从 3DGS 表达中提取网格以支撑后续材质与光照的恢复,因此网格重建的精度尤为关键。为此,本文进一步评估和对比不同方法生成网格质量。考虑到 GS-IR 和 R3DG 原始方法都没有直接生成网格结果,

本文转而通过与当前基于 3DGS 生成网格最先进的方法 2DGS 和 PGSR 进行对比,同时参与对比的方法还有基于神经辐射场的方法 NeuS (learning neural implicit surfaces by volume rendering for multi-view reconstruction) (Wang 等, 2021)。

在 DTU 数据集上的定量比较结果如表 2 所示,本文方法在倒角距离指标上表现优越。相较于 NeuS 和 2DGS,倒角距离值表现优异,而与基于 3DGS 的方法 PGSR 相当。此外,本文方法的训练时间为 0.5 h,介于 2DGS 的 0.32 h 和 PGSR 的 1 h 之间,表明本文方法在保持较高重建精度的同时,具有较快的训练速度。

表 2 DTU 数据集上倒角距离的定量比较

Table 2 Quantitative comparison of chamfer distance on DTU dataset

方法	s24	s37	s40	s55	s63	s65	s69	s83	s97	s105	s106	s110	s114	s118	s122	均值	训练时间/h
NeuS	1.00	1.37	0.93	0.43	1.10	1.29	0.57	1.48	1.09	0.83	0.52	1.20	0.35	0.49	0.54	0.84	6
2DGS	0.48	0.91	0.39	0.39	1.01	0.83	0.81	1.36	1.27	0.76	0.7	1.40	0.40	0.76	0.52	0.80	0.32
PGSR	0.34	0.58	0.29	0.29	0.78	0.58	0.54	1.01	0.73	0.51	0.49	0.69	0.31	0.37	0.38	0.53	1
本文	0.33	0.60	0.30	0.29	0.74	0.63	0.54	0.99	0.74	0.52	0.47	0.65	0.34	0.35	0.41	0.53	0.5

注:加粗字体表示各列最优结果。s24 指场景 scan 24,其余类同。

DTU 数据集的部分定性比较结果如图 6 所示。可以看出,与 2DGS 方法相比,本文的几何估计结果有很大的提升。2DGS 的训练结果并不稳定,细节处可能出现空洞或异常突起,而本文结果则少有这些问题。与最近的基于 3DGS 的几何恢复方法 PGSR 相比,本文方法结果与之差异不大,细节上本文结果的平滑度较高,在 24、37、83 等场景下较为明显。总体而言,本文方法在 DTU 数据集上表现出较高的重建精度和较短的训练时间,相较于其他方法具有明显的优势。

3.4 材质恢复与渲染质量对比

本文在 TensoIR 数据集上与近年来多个逆渲染方法进行比较,包括基于神经辐射场的 NeRFactor、InvRender、Nvdiffrmc 和 TensoIR,以及基于三维高斯泼溅的 GS-IR 和 R3DG 方法。

由于逆渲染问题的本身是不适定问题,同一图像可能对应多组不同的材质与光照组合,而不同方法在材质建模上也存在差异,因此直接对比中间变量如粗糙度的准确性缺乏意义,且 TensoIR 数据集也未提供粗糙度的真实标签。为此,本文仅对材质、

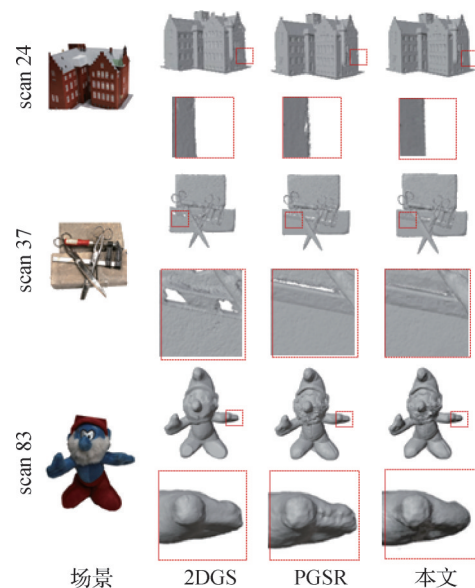


图 6 DTU 数据集部分定性比较结果

Fig. 6 Qualitative comparison of different methods on selected DTU scenes

光照的中间表示进行定性观察,不参与定量分析。而在定量评估中,则采用这些中间参数重新渲染生成图像,并通过峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity index

measure, SSIM)和学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)等指标,分别评估其在新视角合成与新环境光渲染任务中的表现,以综合反映方法在材质与光照估计方面的有效性。

表3给出了反照率、新视角合成(novel view synthesis, NVS)和训练时间的定量比较结果。本文方法在反照率和新视角合成上分别取得31.15 dB和35.36 dB的最高分,超越所有基于NeRF或3DGS的现有方法,证明了本框架材质恢复与全局光照分解的优势;从SSIM指标看,本文方法在反照率和新视角合成中的得分高于所有基于3DGS的方

法。虽然反照率的LPIPS指标结果落后于GS-IR(0.087 vs. 0.084),但在新视角合成下的LPIPS值最优,为0.038,说明本文方法合成的新视角图像在感知质量上更接近真实图像,细节保留更佳,视觉体验更佳。与Nvdiffrmc基线模型相比,本文方法显著提升了训练效率。第1阶段利用3DGS进行几何恢复,将处理时间从约2 h缩短至0.5 h;第2阶段通过加权池采样策略,减少了光照采样量,进一步节省了约1 h的计算开销。另外,由于各方法材质模型差异较大且缺乏粗糙度真值标签,本文未对粗糙度进行定量对比,仅通过定性结果辅助验证方法的综合表现。

表3 反照率、新视角合成和训练时间的定量比较

Table 3 Quantitative comparison of albedo, novel view synthesis, and training time

类别	方法	反照率			NVS			训练时间/h
		PSNR ↑ /dB	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓	
基于NeRF的方法	NeRFactor	25.13	0.940	0.109	24.68	0.922	0.120	> 100
	InvRender	27.34	0.933	0.100	27.37	0.934	0.089	15
	Nvdiffrmc	29.17	0.908	0.115	34.41	0.967	0.059	4
	TensoIR	29.27	0.950	0.085	35.09	0.976	0.040	5
基于3DGS的方法	R3DG	29.40	0.932	0.094	34.85	0.963	0.049	1
	GS-IR	30.29	0.941	0.084	35.33	0.974	0.039	1
	本文	31.15	0.948	0.087	35.36	0.973	0.038	1.5

注:加粗字体表示各类别各列最优结果。“↑”和“↓”分别表示值越高越好、值越低越好。

为了更直观地展示实验结果,本文选取评估指标数值较高的两项工作 TensoIR 和 GS-IR,与本文方法进行定性对比,结果如图7所示。从法线(normal)图对比来看,与第1阶段实验所得规律类似,本文方法产生的法线结果更加平滑,且细节恢复程度更好。在反照率方面,GS-IR整体颜色较暗,未能准确实现材质与光照的分离,这尤其体现在高光和阴影边缘,光照残留导致了材质细节的模糊;而本文反照率图则无明显光照干扰,色彩自然,视觉上与 TensoIR 的结果较为接近,这主要得益于在直接光照估计阶段采用的加权池采样策略。其高效与低方差的特性,使得即使面对“热狗”场景中盘子光滑表面的强高光,或是“乐高”挖掘机复杂结构产生的软阴影,算法依然能以较少的样本获得准确的光照估计,从而避免了光照信息“泄露”到反照率中,实现了更彻底和干净的解耦。

在粗糙度方面,本文与GS-IR的结果图像偏亮,

表明粗糙度较高,光线反射更加散射;而 TensoIR 的粗糙度灰度图偏暗,表明其预测表面较为光滑,反射更集中。新视角合成结果在乐高场景下无明显区别,在热狗场景下可以较为明显地看出盘子表面更光滑,与真实值更加接近。

本文方法在多个关键指标上均优于 TensoIR 和 GS-IR。得益于本文提出的基于物理的光照分离策略。与GS-IR易受环境光干扰、在反照率中产生伪影不同,本文方法通过对直接光照和间接光照的合理建模,有效解耦了光照与材质信息,从而显著提升了反照率估计的准确性。

3.5 消融实验

为了评估几何恢复部分与材质光照恢复部分关键部件对整体逆渲染方法的影响,本文开展两个实验验证。两项实验的未修改部分均使用基线工作 Nvdiffrmc 的对应部分替代,结果如表4所示。从表4可以看出,若移除重投影误差约束,新视角合成

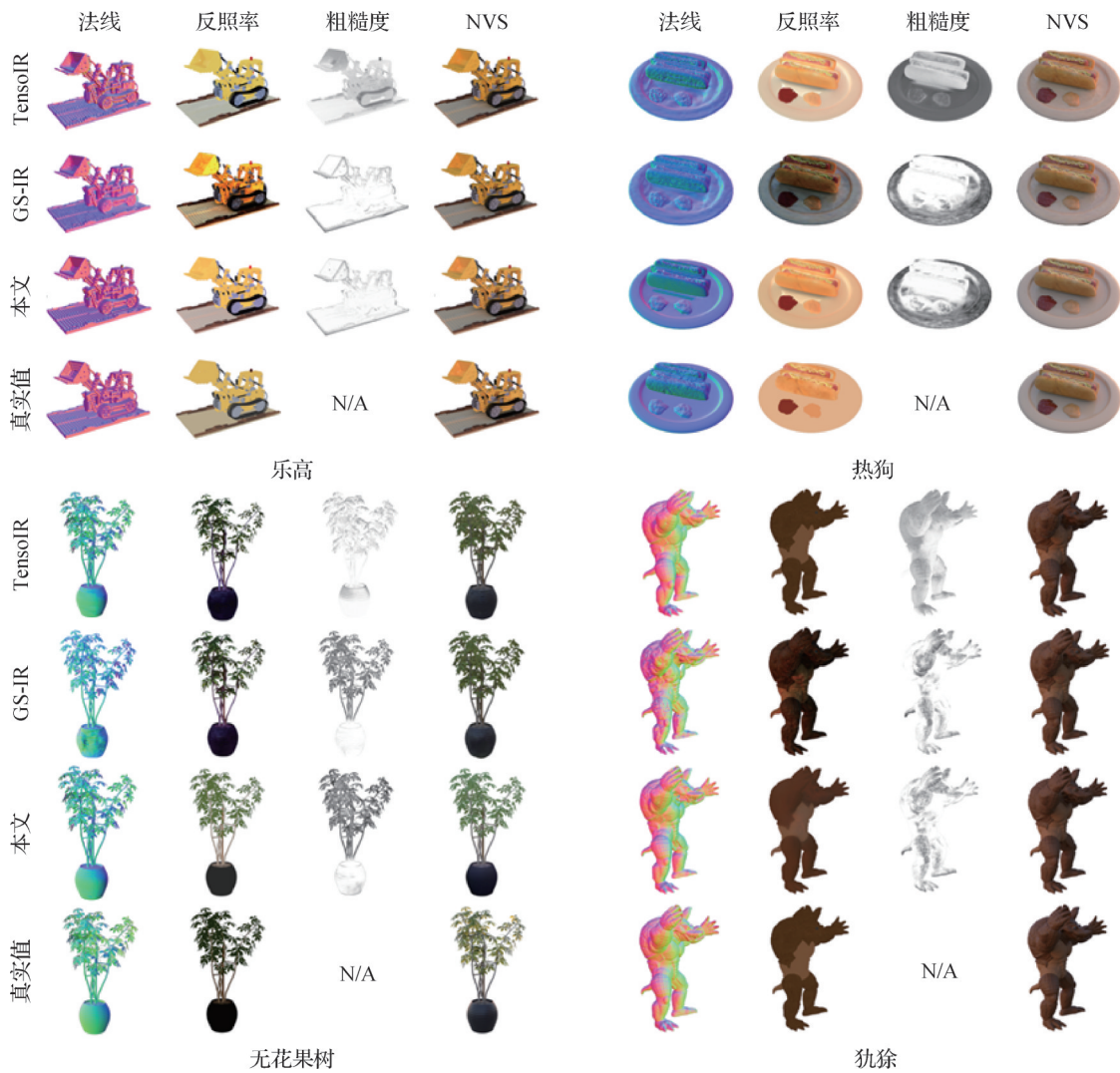


图7 TensoIR数据集下的定性比较

Fig. 7 Qualitative comparison on the TensoIR dataset

的PSNR从35.36 dB下降至34.10 dB,说明了高质量、一致性的几何对于最终渲染的重要性。同样,若不使用加权池采样,PSNR大幅下降至31.67 dB, LPIPS指标也显著变差,这凸显了高效低方差光照估计在材质恢复和渲染中的关键作用。

表4 重投影误差约束与加权池采样的消融研究

Table 4 Ablation study on re-projection error constraint and weighted reservoir sampling

方法	新视角合成		
	PSNR ↑ /dB	SSIM ↑	LPIPS ↓
减重投影误差约束	34.10	0.968	0.046
减加权池采样	31.67	0.963	0.050
本文	35.36	0.973	0.038

注:加粗字体表示各列最优结果。“↑”和“↓”分别表示值越高越好、值越低越好。

为分析单个像素采样数(samples per pixel, SPP)对渲染质量和训练效率的影响,本文在不同的SPP设置下进行对比实验,定量结果如表5所示,定

表5 单个像素采样数量的消融研究

Table 5 Ablation study on the number of samples per pixel

方法	新视角合成			训练时间/min
	PSNR ↑ /dB	SSIM ↑	LPIPS ↓	
Nvdiffrecmc (SPP = 128)	34.41	0.967	0.059	240
本文 (SPP = 4)	34.27	0.964	0.048	63
本文 (SPP = 8)	34.85	0.969	0.043	65
本文 (SPP = 32)	35.19	0.971	0.040	77
本文 (SPP = 64)	35.36	0.973	0.038	92

注:加粗字体表示各列最优结果。“↑”和“↓”分别表示值越高越好、值越低越好。

性对比结果如图8所示。

从定量结果可以看出,随着SPP的增加,PSNR和SSIM指标逐步提高,LPIPS逐步降低,表明渲染质量得到了明显改善。结合图7可以看出,在SPP从4提升到32的过程中,渲染质量提升最为显著,噪点逐渐减少,最终在SPP = 64时趋于收敛。此外,实验表明,训练时间与SPP成近似线性关系,随着采样数

量增加,训练时间也相应增长。例如,从SPP = 4到SPP = 64,训练时间从63 min增长至92 min,但相比Nvdiffrmc(SPP = 128)的240 min训练时间,展现了显著的效率优势。同时,在更低采样率下,本文已可达到更优的渲染指标(PSNR = 35.19, LPIPS = 0.040),在保持高质量合成的同时大幅降低了计算成本。



图8 不同数量的单像素采样的定性比较

Fig. 8 Qualitative comparison with varying numbers of samples per pixel

这一组消融实验的结果,强有力地印证了2.3.2节中讨论的加权池采样的理论优势。SPP较小时,噪点主要来源于采样不足导致的光照估计误差;而随着采样数增加,光照估计更加准确,噪点显著减少。本文提出的加权池采样策略,通过有效复用高质量样本,即使在SPP极低时(如4或8)也能大幅抑制方差,因此其渲染质量的“起点”远高于传统方法,并且能更快地收敛到视觉无噪点的状态。然而,当SPP超过64时,指标提升趋于饱和,这充分说明本文方法能够以远低于基线方法的计算成本,达到甚至超越其渲染质量,展现了在效率与质量平衡上的巨大潜力。

通过上述消融研究可以看出,本文在几何恢复与光照建模两个阶段中引入的关键模块均对提升整体渲染性能具有显著作用。其中,几何一致性约束在提升法线和粗糙度恢复方面效果显著,而加权池采样则主要优化光照估计,使得反照率分解更加准确。二者结合,可在保证高质量渲染结果的同时,显著降低计算成本。此外,本文还系统研究了单个像素采样数量对渲染质量和训练时间的影响,验证了即使在较低采样率(SPP = 32或SPP = 64)下,本文方法亦可获得接近最优的新视角合成质量,且训练时间远低于Nvdiffrmc,体现了效率与质量的平衡,展现了更高的实用性。

4 结论

本文针对当前基于3DGS的逆渲染方法在几何层面因缺乏精确法线、多视角一致性以及在光照材质解耦层面因直接光照计算效率和方差问题而导致的性能瓶颈,提出了一套两阶段快速物理逆渲染框架。通过在第1阶段引入基于扁平高斯基元的法线估计和双向重投影误差约束,显著提升了几何重建的鲁棒性和精度,为后续物理渲染奠定了高质量基础。在TensoIR和DTU数据集上的实验结果表明,本文方法在几何精度和训练效率方面均优于当前代表性方法。在第2阶段,本文创新性地将加权池采样方法应用于直接光照估计中,大幅降低了噪声方差并提高了采样效率。此外,本文还通过基于哈希神经表示实现复杂高频材质细节的恢复,通过引入结合多重重要性采样和基于物理渲染的精确路径追踪技术,最终实现了精确的全局光照模拟和准确材质恢复。基于TensoIR数据集的定量与定性对比,以及消融实验结果,验证了方法整体及关键模块的有效性。未来工作将主要考虑结合大场景数据集进行更广泛的模型训练,探索更多适用于大规模场景的几何和材质建模方法以支持实际应用,并结合参数空间的ReSTIR(resampled importance sampling)思路以获得更严格的理论保证。

参考文献 (References)

- Bitterli B, Wyman C, Pharr M, Shirley P, Lefohn A and Jarosz W. 2020. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting. *ACM Transactions on Graphics (TOG)*, 39(4): #148 [DOI: 10.1145/3386569.3392481]
- Boss M, Braun R, Jampani V, Barron J T, Liu C and Lensch H P A. 2021. NeRD: neural reflectance decomposition from image collections//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 12664-12674 [DOI: 10.1109/ICCV48922.2021.01245]
- Chang W, Sivaram V, Nowrouzezahrai D, Hachisuka T, Ramamoorthi R and Li T M. 2023. Parameter-space ReSTIR for differentiable and inverse rendering//Proceedings of 2023 ACM SIGGRAPH 2023 Conference Proceedings. Los Angeles, USA: ACM: #18 [DOI: 10.1145/3588432.3591512]
- Chen D P, Li H, Ye W C, Wang Y F, Xie W J, Zhai S J, et al. 2025a. PGSR: planar-based Gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 31(9): 6100-6111 [DOI: 10.1109/TVCG.2024.3494046]
- Chen H L, Li C, Wang Y S and Lee G. 2025b. NeuSG: neural implicit surface reconstruction with 3D Gaussian splatting guidance [EB/OL]. [2025-06-09]. <https://arxiv.org/pdf/2312.00846.pdf>
- Chen H Z, Lin Z H and Zhang J. 2025c. GI-GS: global illumination decomposition on Gaussian splatting for inverse rendering//Proceedings of the 13th International Conference on Learning Representations. Singapore, Singapore: OpenReview.net
- Gao J, Gu C, Lin Y T, Li Z H, Zhu H, Cao X, et al. 2025. Relightable 3D Gaussians: realistic point cloud relighting with BRDF decomposition and ray tracing//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 73-89 [DOI: 10.1007/978-3-031-72995-9_5]
- Hasselgren J, Hofmann N and Munkberg J. 2022. Shape, light, and material decomposition from images using Monte Carlo rendering and denoising//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #1661 [DOI: 10.5555/3600270.3601931]
- Huang B B, Yu Z H, Chen A P, Geiger A and Gao S H. 2024. 2D Gaussian splatting for geometrically accurate radiance fields//Proceedings of 2024 ACM SIGGRAPH 2024 Conference Papers. Denver, USA: ACM: #32 [DOI: 10.1145/3641519.3657428]
- Jin H A, Liu I, Xu P J, Zhang X S, Han S F, Bi S, et al. 2023. Tensoir: tensorial inverse rendering//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 165-174 [DOI: 10.1109/CVPR52729.2023.00024]
- Kerbl B, Kopanas G, Leimkuehler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4): #139 [DOI: 10.1145/3592433]
- Lai S C, Huang L T, Guo J, Cheng K, Pan B W, Long X X, et al. 2025. GlossyGS: inverse rendering of glossy objects with 3D Gaussian splatting. *IEEE Transactions on Visualization and Computer Graphics*, 31(10): 7478-7491 [DOI: 10.1109/TVCG.2025.3547063]
- Liang Z H, Zhang Q, Feng Y, Shan Y and Jia K. 2024. GS-IR: 3D Gaussian splatting for inverse rendering//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 21644-21653 [DOI: 10.1109/CVPR52733.2024.02045]
- Lyu H Y, Liu Y L, Li H, Yuan X and Xing G Y. 2025. Multi-view intrinsic decomposition of indoor scenes under a 3D Gaussian splatting framework. *Journal of Image and Graphics*, 30(7): 2514-2527 (吕恒焯, 刘艳丽, 李宏, 袁霞, 邢冠宇. 2025. 三维高斯泼溅框架下的室内场景多视图本征分解. *中国图象图形学报*, 30(7): 2514-2527) [DOI: 10.11834/jig.240505]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Müller T, Evans A, Schied C and Keller A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4): #102 [DOI: 10.1145/3528223.3530127]
- Munkberg J, Chen W Z, Hasselgren J, Evans A, Shen T C, Müller T, et al. 2022. Extracting triangular 3D models, materials, and lighting from images//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 8270-8280 [DOI: 10.1109/cvpr52688.2022.00810]
- Shen T C, Gao J, Yin K X, Liu M Y and Fidler S. 2021. Deep marching tetrahedra: a hybrid representation for high-resolution 3D shape synthesis//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s.l.]: Curran Associates Inc.: #466 [DOI: 10.5555/3540261.3540727]
- Srinivasan P P, Deng B Y, Zhang X M, Tancik M, Mildenhall B and Barron J T. 2021. NeRV: neural reflectance and visibility fields for relighting and view synthesis//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7491-7500 [DOI: 10.1109/CVPR46437.2021.00741]
- Wang Y, Liu Y, Chen H, Liu J and Wang X. 2021. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction[EB/OL]. [2025-06-09]. <https://arxiv.org/pdf/2106.10689.pdf>
- Zeng A, Song S R, Nießner M, Fisher M, Xiao J X and Funkhouser T. 2017. 3DMatch: learning local geometric descriptors from RGB-D reconstructions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 199-208 [DOI: 10.1109/CVPR.2017.29]

- Zeng Z H, Wang Z J, Zhang Y B, Cai W N, Zhang L L, Guo Y, et al. 2025. Semantic and consistent neural radiance field reconstruction method based on intrinsic decomposition via classification. *Journal of Image and Graphics*, 30(2): 559-574 (曾志鸿, 王宗继, 张源奔, 蔡伟南, 张利利, 郭岩, 等. 2025. 利用本征属性分类的神经辐射场视角及语义一致性重建. *中国图象图形学报*, 30(2): 559-574) [DOI: 10.11834/jig.240140]
- Zhang K, Luan F J, Wang Q Q, Bala K and Snavely N. 2021a. PhysSG: inverse rendering with spherical Gaussians for physics-based material editing and relighting//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 5449-5458 [DOI: 10.1109/cvpr46437.2021.00541]
- Zhang X M, Srinivasan P P, Deng B Y, Debevec P, Freeman W T and Barron J T. 2021b. NeRFactor: neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on*

Graphics (TOG), 40(6): #237 [DOI: 10.1145/3478513.3480496]

作者简介

刘铮,男,硕士研究生,主要研究方向为逆渲染。

E-mail: liuliubian@163.com

李游,通信作者,男,副研究员,硕士生导师,主要研究方向为3D视觉。E-mail: liyou@gml.ac.cn

汤圣君,男,副教授,主要研究方向为视觉定位。

E-mail: shengjuntang@szu.edu.cn

姚萌萌,男,工程师,主要研究方向为点云数据处理。

E-mail: yaomengmeng@gml.ac.cn

郭仁忠,男,教授,中国工程院院士,主要研究方向为地图制图、地理信息科学和智慧城市。E-mail: guorz@szu.edu.cn