

AGES: 各向异性高斯平滑增强的三维几何一致性重建方法

屈世泽*, 苏佳文, 张 明

内蒙古大学数学科学学院, 内蒙古 呼和浩特

收稿日期: 2026年1月3日; 录用日期: 2026年2月3日; 发布日期: 2026年2月11日

摘 要

在稀疏视角和低纹理区域条件下, 从多视角图像实现高保真三维重建仍然是一个具有挑战性的问题, 传统方法在此类场景中往往表现不稳定。尽管三维高斯溅射(3D Gaussian Splatting, 3DGS)能够实现实时渲染, 但其几何优化与各向异性外观建模相互解耦, 容易在无纹理或高反射区域产生伪影。针对上述问题, 本文提出一种各向异性高斯平滑增强方法(Anisotropic Gaussian Enhancement with Smoothness, AGES), 构建了一个基于概率路由的联合优化框架, 在单一训练流程中联合优化场景几何结构与视角相关外观建模。该方法引入两个关键组成部分: (1) 自适应几何-外观路由模块(Adaptive Geometry-Appearance Routing, AGAR), 基于学习得到的逐高斯不确定性度量, 动态地将高斯基元分配至跨视角几何细化分支或各向异性反射建模分支; (2) 深度平滑正则化项(Depth Smoothness Regularization, DSR), 通过约束渲染深度与几何优化深度之间的局部梯度一致性, 在抑制噪声的同时有效保持结构边缘。大量在Waymo和YouTube数据集上的实验结果表明, 所提出的AGES方法在全分辨率及降采样设置下均显著优于现有先进方法, 在复杂真实场景中实现了更高的几何一致性和视觉保真度。

关键词

三维高斯溅射, 新视角合成, 神经渲染, 跨视角几何增强, 深度平滑正则化

AGES: Anisotropic Gaussian Enhancement with Smoothness for Geometric-Consistent 3D Reconstruction Method

Shize Qu*, Jiawen Su, Ming Zhang

School of Mathematical Sciences, Inner Mongolia University, Hohhot Inner Mongolia

Received: January 3, 2026; accepted: February 3, 2026; published: February 11, 2026

*通讯作者。

文章引用: 屈世泽, 苏佳文, 张明. AGES: 各向异性高斯平滑增强的三维几何一致性重建方法[J]. 计算机科学与应用, 2026, 16(2): 289-302. DOI: 10.12677/csa.2026.162059

Abstract

High-fidelity 3D reconstruction from multi-view images remains a significant challenge, particularly under sparse viewpoints and in low-texture regions where conventional methods are often unreliable. While 3D Gaussian Splatting (3DGS) enables real-time rendering, the decoupling of its geometric optimization and anisotropic appearance modeling frequently leads to artifacts in textureless or specular areas. To address these limitations, we propose Anisotropic Gaussian Enhancement with Smoothness (AGES), which formulates a probability-guided joint optimization framework to jointly optimize scene geometry and view-dependent appearance within a single training pipeline. Our approach introduces two key components: (1) an Adaptive Geometry-Appearance Routing (AGAR) module, which dynamically routes each Gaussian to either a cross-view geometric refinement branch or an anisotropic reflectance modeling branch based on a learned per-primitive uncertainty measure; and (2) a Depth Smoothness Regularization (DSR) loss, which enforces local gradient consistency between the rendered and geometrically refined depth maps to preserve structural edges while suppressing noise. Extensive experiments on the Waymo and YouTube datasets demonstrate that AGES significantly outperforms state-of-the-art methods at both full and downsampled resolutions, achieving superior geometric consistency and visual fidelity in challenging real-world scenarios.

Keywords

3D Gaussian Splatting, Novel View Synthesis, Neural Rendering, Cross-View Geometric Enhancement, Depth Smoothness Regularization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

多视角图像的高保真三维重建与真实感新视角合成是计算机视觉、计算机图形学以及遥感领域中的核心研究问题。这两项技术在城市三维建模、环境监测、自动驾驶、增强现实与虚拟现实、数字人建模[1]、沉浸式视频合成[2]以及大规模城市场景重建[3]等诸多应用中具有重要支撑作用。在遥感应用场景中,精确的三维建模对于灾害评估、农业监测以及城市尺度地形分析等任务尤为关键。然而,受限于视角稀疏、低纹理区域以及光照条件变化等因素,现有方法在实际应用中仍面临较大挑战。

近年来,研究者通过引入显式几何先验或改进外观建模方式,在一定程度上提升了三维重建质量。例如,Jiang 等[4]将网格形变机制引入基于高斯表示的数字人建模中,以增强几何表达能力;Wen 等[5]则利用深度引导的一致性约束,提高了大规模场景重建的几何稳定性。神经辐射场(Neural Radiance Fields, NeRF)方法[6]通过多层感知机(MLP)参数化连续体渲染场,并结合可微体渲染机制,实现了高质量的新视角合成。随后,大量研究围绕光照建模[7]-[9]、多尺度优化策略[10]-[12]以及更高表达能力的具体表示方式[13] [14]等方面对 NeRF 进行了改进。与此同时,自适应采样技术[15] [16]在提升渲染效率的同时,有效保留了细粒度结构信息。

针对遥感场景中常见的复杂成像条件,相关研究也对 NeRF 在少视角重建[17]、低照度成像[18]以及大规模航拍或卫星影像建模[19] [20]等问题进行了探索。尽管取得了一系列进展,NeRF 方法在计算效率方面仍存在明显瓶颈。其基于射线行进的渲染过程需要进行密集采样并多次调用 MLP 网络,导致在大规

模遥感任务中难以满足高效处理的需求。尽管部分加速方法通过稀疏采样[21] [22]或优化数据结构[23]-[25]减少了计算开销,但基于射线的渲染框架仍然是主要限制因素。

近年来提出的三维高斯点渲染(3D Gaussian Splatting, 3DGS)方法[26]通过以显式的高斯基元替代隐式体渲染表示,并采用栅格化友好的渲染流程,在保证较高视觉质量的同时实现了实时渲染性能。然而,该方法在表达能力上仍受到两方面限制:一方面,低阶球谐函数难以准确刻画高频的视角相关效应,如尖锐高光等。尽管各向异性球高斯(ASG)模型[27]在一定程度上提升了外观建模能力,但其对几何误差仍较为敏感;另一方面,更为关键的是,3DGS在初始化阶段高度依赖于结构光束法(Structure-from-Motion, SfM)。在低纹理区域或视角覆盖不足的情况下,SfM往往生成不可靠的点云,进而导致高斯基元错位和几何结构不一致(见图1)。现有的致密化策略虽可部分缓解该问题,但由于缺乏跨视角的几何一致性约束,难以从根本上修正深层次的几何误差。

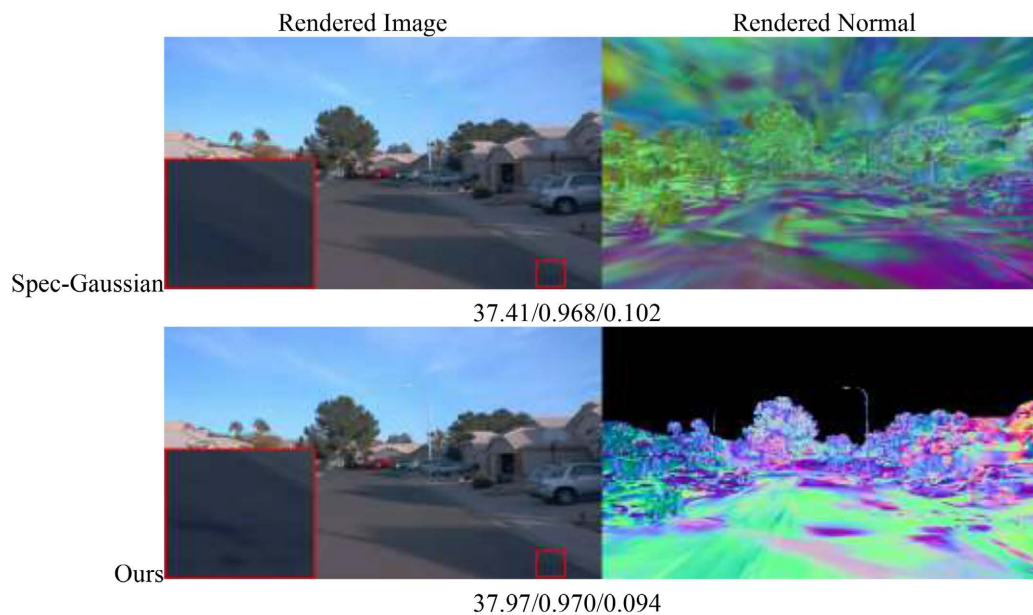


Figure 1. Comparison of rendered images and surface normals on the Waymo dataset (Segment-102751) at a 4× downsampled resolution, along with the corresponding PSNR/SSIM/LPIPS index

图 1. Waymo 数据集(Segment-102751)在 4×下采样分辨率下的渲染结果与表面法线对比结果及对应的 PSNR/SSIM/LPIPS 指标

基于上述分析,我们认为稳健的三维重建方法应当在几何结构与视角相关外观之间建立紧密耦合关系。为此,本文提出了一种各向异性高斯增强与平滑约束相结合的三维重建方法(Anisotropic Gaussian Enhancement with Smoothness, AGES),构建了一个基于概率路由的联合优化框架,在同一训练过程中联合优化几何结构与外观表示。AGES 包含两个关键组成部分:其一是自适应几何-外观路由模块(Adaptive Geometry-Appearance Routing, AGAR),该模块依据学习得到的不确定性度量,在几何增强分支与各向异性反射建模分支之间动态选择优化路径;其二是深度平滑正则项(Depth Smoothness Regularization, DSR),通过约束多视角深度梯度的一致性,在抑制噪声的同时有效保留结构边界。在初始化条件较差的情况下,该方法能够显著提升几何稳定性,并保持高频外观细节,在定量指标与视觉质量方面均优于现有方法。

本文的主要贡献总结如下:

(1) 提出了一种基于概率路由的几何-外观联合优化框架 AGES,实现了在复杂真实场景中的高鲁棒

性三维重建;

(2) 设计了自适应几何 - 外观路由模块 AGAR, 可在几何增强与各向异性外观建模之间动态切换, 从而兼顾深度精度与视角相关视觉细节;

(3) 提出了一种基于梯度一致性的深度平滑正则项 DSR, 在保持结构完整性的同时, 有效提升了大规模数据集上的自监督重建质量。

本文其余结构安排如下: 第二节回顾与三维高斯点渲染相关的研究工作; 第三节详细介绍 AGES 的整体方法; 第四节给出实验设置与结果分析; 第五节对全文进行总结, 并讨论未来研究方向。

2. 相关工作

3D 高斯溅射(3D Gaussian Splatting, 3DGS) [26]提出了一种显式场景表示方法, 通过可学习的高斯基元对场景进行建模。该方法采用可微分溅射渲染, 在保持较高视觉质量的同时, 实现了实时渲染, 从而有效克服了 NeRF [6]在渲染效率方面的不足。

在 3DGS 中, 场景被表示为一组三维各向异性高斯, 其参数形式为:

$$G_i = (\mu_i, \Sigma_i, c_i, \alpha_i), \quad (1)$$

其中, $\mu_i \in \mathbb{R}^3$ 表示第 i 个高斯的空间中心位置, $\Sigma_i \in \mathbb{R}^{3 \times 3}$ 为协方差矩阵, 用于控制高斯的形状与方向, c_i 表示视角相关颜色, α_i 为不透明度参数, 用于控制其透明程度。

每个三维高斯在相机视角下被投影到二维图像平面。对于像素坐标 $x \in \mathbb{R}^2$, 其对应的投影二维高斯为:

$$G_i(x) = \alpha_i \exp\left(-\frac{1}{2}(x - \mu_{p,i})^T \Sigma_{p,i}^{-1}(x - \mu_{p,i})\right), \quad (2)$$

其中 $\mu_{p,i}$ 和 $\Sigma_{p,i}$ 分别表示投影后的均值和协方差矩阵。

最终像素颜色通过 α 合成方式计算:

$$C(x) = \sum_{i \in \mathcal{N}(x)} T_i G_i(x) c_i, \quad (3)$$

其中 $\mathcal{N}(x)$ 表示对像素 x 有贡献的高斯集合, T_i 为第 i 个高斯对应的透射率, 其定义为:

$$T_i = \prod_{j=1}^{i-1} (1 - G_j(x)). \quad (4)$$

现有的 3DGS 扩展方法主要从多个关键方面对原始框架进行改进。多尺度渲染方法, 如 Multi-Scale GS [28]和 MIP-Splatting [29], 通过引入层次化与频率感知表示来有效缓解混叠问题。Mirror-3DGS [30]与 Spec-Gaussian [27]则引入更先进的反射率建模, 以提升对复杂光照条件和高光效应的表达能力。R3DG [31]通过引入光线追踪机制, 实现了可重光照的三维表示。几何优化类方法, 包括 GaussianPro [32]、FreGS [33]以及 CGC-GS [34], 则利用几何约束与结构信息提升了深度与法向量估计的精度。

与上述方法主要侧重于外观建模或几何建模不同, 本文方法联合利用各向异性球面高斯与跨视角几何一致性, 在训练过程中对几何结构进行动态优化, 从而克服基于 SfM 初始化所带来的固有局限, 并在复杂外观条件下实现更加稳健的三维重建。

3. 方法

我们提出 AGES, 一个统一的三维高斯溅射(3D Gaussian Splatting, 3DGS)框架, 用于同时建模并优化真实场景中常见的两个关键挑战, 即几何不稳定性与过于简化的外观建模。通过协同增强视角相关的外

观表达能力与多视角几何一致性, AGES 能够实现高质量的三维重建。

本文方法的核心思想在于: 高保真三维重建依赖于各向异性反射建模与鲁棒几何增强之间的紧密耦合。如图 2 所示, AGES 框架融合了两项关键创新:

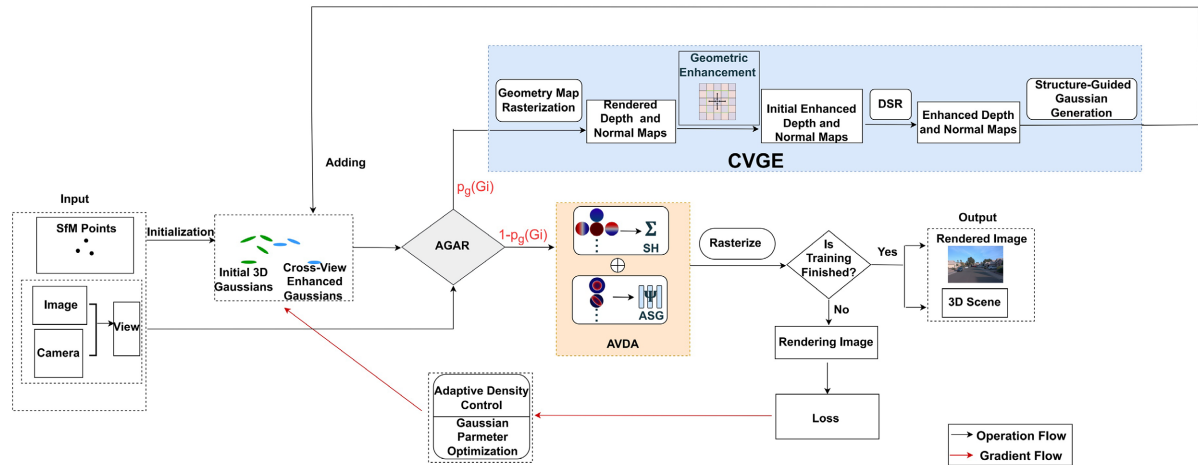


Figure 2. The AGES framework integrates anisotropic appearance modeling with geometry-aware optimization through cross-view constraints

图 2. AGES 框架通过跨视角约束将各向异性外观建模与几何感知优化相结合

(1) AGAR 模块: 该模块在几何约束不足的区域提供鲁棒的深度与法向增强, 同时在具有高光或各向异性材质的区域提升视角相关外观建模能力, 从而在提升几何一致性的同时改善整体视觉质量;

(2) DSR: 该正则化项通过在渲染深度与增强深度之间引入局部一致性约束, 平滑深度变化并保持几何连续性与局部结构完整性。

3.1. 自适应几何 - 外观路由模块(AGAR)

现有多视角三维重建与渲染方法通常将几何建模与外观建模割裂开来分别处理。例如, GaussianPro 主要依赖渐进式传播(progressive propagation)来细化由 SfM 初始化得到的稀疏几何结构。然而, 该策略仍然存在明显局限性: 一方面, 其所采用的 ASG 或 BRDF 表示能力有限, 虽然可以得到几何上较为完整的高斯分布, 但往往导致外观质量退化; 另一方面, 在存在高光反射或各向异性材质的区域, 仅依赖几何传播难以刻画真实的视角相关变化, 常常出现“几何准确但外观平淡”的现象。

相反, 侧重于视角外观建模的方法(如 AVDA)对几何误差极为敏感: 当几何不准确时, 纹理往往被强行“贴”到错误的位置, 从而引入重影; 在稀疏视角设置下, 会进一步产生形状与外观之间的错位。因此, 仅关注几何或仅关注外观的策略, 难以在稀疏视角与大规模场景下同时保证几何一致性与高质量外观。

为解决上述问题, 本文提出一种自适应几何 - 外观路由模块, 称为 AGAR。AGAR 通过一个可学习的判别函数, 对每个高斯基元进行逐高斯决策, 在其内部的跨视角几何增强分支(Cross-View Geometry Enhancement, CVGE)与各向异性视角相关外观建模分支(Anisotropic View-Dependent Appearance, AVDA)之间进行概率路由与加权融合, 从而在几何约束不足的区域与外观变化显著的区域分别采用最合适的建模策略。

不同于传统的几何传播方法, AGAR 中的 CVGE 分支基于本文提出的跨视角几何增强机制, 通过推断多视角深度与法向一致性, 自动从相邻视角中聚合可靠的几何线索, 并结合所提出的深度平滑正则项(DSR), 在训练过程中持续校正 SfM 初始化误差, 同时提升稀疏观测区域的几何稳定性。该分支能够在几何约束不足的区域提供更加鲁棒的深度与法向估计, 从而显著增强整体几何一致性。

相比之下, AGAR 中的 AVDA 分支侧重于在存在高光反射或各向异性材质的区域捕捉真实的视角相关外观变化。在判别函数输出的概率引导下, AGAR 模块对两个分支的结果进行自适应融合, 从而在不同区域实现几何与外观之间的最优权衡。

具体而言, 对于每一个高斯 G_i , 我们定义一个线性判别函数:

$$r_i = w_g U_g(G_i) - w_a U_a(G_i), \quad (5)$$

其中, w_g 与 w_a 为可学习的权重参数, 几何不确定性度量 $U_g(G_i)$ 由高斯的尺度各向异性计算得到, 具体定义为:

$$U_g(G_i) = \sigma \left(\beta_g \cdot \frac{\min(s_i)}{\max(s_i)} \right), \quad (6)$$

其中 $s_i = (s_i^1, s_i^2, s_i^3)$ 表示高斯的三轴尺度, β_g 为可学习的标量参数。

外观不稳定性度量 $U_a(G_i)$ 由高斯在当前迭代中的颜色残差幅度定义:

$$U_a(G_i) = \sigma(\beta_a \cdot \|\Delta c_i\|_2), \quad (7)$$

其中 Δc_i 表示预测颜色与观测颜色之间的残差, β_a 为可学习的标量参数。

判别得分通过 Sigmoid 函数转换为选择几何增强分支的概率:

$$p_g(G_i) = \sigma(r_i). \quad (8)$$

最终输出由两个分支按照概率进行加权融合得到:

$$\text{AGAR}(G_i) = p_g(G_i) \text{CVGE}(G_i) + (1 - p_g(G_i)) \text{AVDA}(G_i). \quad (9)$$

该形式使得 AGAR 模块能够在高斯级别上自适应地平衡几何增强与视角相关外观建模。在几何约束不足的区域, 高斯更倾向于选择 CVGE 分支, 从而获得更加鲁棒的深度与法向估计; 而在存在高光或各向异性材质的区域, 高斯则更可能依赖 AVDA 分支, 以捕获真实的视角相关细节与反射变化。通过这种加权融合机制, AGES 在保证几何一致性的同时保持高质量外观, 有效克服了单独使用 CVGE 或 AVDA 所带来的局限性。



Figure 3. Illustrative visualization of AGAR routing probability $p_g(G_i)$ in a typical scene

图 3. 在典型场景中 AGAR 路由概率 $p_g(G_i)$ 的示意性可视化

为进一步分析 AGAR 模块的行为, 我们对不同类型区域的路由概率 $p_g(G_i)$ 进行了示意性可视化。如图 3, 在几何约束不足或低纹理区域, 高斯基元倾向于选择 CVGE 分支, 而在高光或各向异性材质区域,

则更偏向 AVDA 分支。这说明 AGAR 能够根据局部不确定性自适应调整几何增强与外观建模的权重，从而验证其自适应行为。

3.2. 深度平滑正则项(DSR)

深度平滑正则化(Depth Smoothing Regularization, DSR)是本文框架中的一项核心创新，旨在解决基于三维高斯的重建过程中一个根本性问题：视角相关渲染与多视角几何传播之间固有的几何不一致性。

传统 3DGS [26]方法由于外观优化与几何约束相互解耦，常在几何边界附近产生结构性伪影。尽管 CVGE 分支通过跨视角传播提供了更优的深度估计，但若直接强制渲染深度 D_{render} 与增强深度 D_{enh} 在数值上保持一致，往往会引入过度平滑或不自然的几何畸变。

我们的关键观察是：尽管渲染路径与几何传播路径所得到的绝对深度值可能存在差异，但它们的局部结构变化应当保持一致。基于这一认识，我们提出一种基于梯度的一致性正则项，用于惩罚渲染深度与增强深度之间的局部梯度差异：

$$\mathcal{L}_{depth-smooth} = \|\nabla_x (D_{enh} - D_{render})\|_1 + \|\nabla_y (D_{enh} - D_{render})\|_1, \quad (10)$$

其中， ∇_x 与 ∇_y 表示空间梯度算子。该形式在保持真实深度不连续性的同时，有效约束局部几何变化的一致性。

与现有的几何一致性约束相比，DSR 具有以下三方面优势：

- (1) 自监督性：不依赖外部深度传感器或预先计算的真值深度；
- (2) 边界保持性：对梯度采用 ℓ_1 范数可避免传统正则项常见的过度平滑问题；
- (3) 自适应调度：该正则项仅在训练的中期阶段(第 1000 至 12000 次迭代)激活，在不干扰早期收敛的同时，于最有效阶段提供几何引导。

通过约束梯度一致性而非强制绝对对齐，DSR 在可微渲染框架中从根本上弥合了外观建模与几何一致性之间的差距，从而显著提升了边界精度与整体结构完整性，相关效果已在消融研究中得到验证。

3.3. 训练策略

我们的框架从 $4\times$ 下采样的输入分辨率开始训练，并逐步提升分辨率，以更好地适应复杂场景。为应对不同场景的几何复杂度，CVGE 分支每 15~50 次迭代更新一次。整体优化目标由光度重建损失与本文提出的几何正则项共同构成：

$$L = L_{rgb} + \lambda_{depth-smooth} L_{depth-smooth}. \quad (11)$$

在所有实验中，我们将权重参数设为 $\lambda_{depth-smooth} = 0.01$ 。通过联合引入各向异性外观建模与鲁棒的几何增强，AGES 框架在多个具有挑战性的真实场景数据集上取得了领先的重建性能，相关结果在实验部分进行了验证。

4. 实验与结果

我们进行了系统而全面的实验评估，以验证所提出框架在高保真三维重建任务中的有效性。本节将详细介绍所使用的数据集与实现设置，并与当前先进方法进行定量与定性对比分析，同时通过消融实验进一步分析各关键模块的贡献。

4.1. 实验设置

4.1.1. 数据集与评价指标

我们在两个具有挑战性的真实世界数据集上评估所提出的方法：Waymo [35]数据集和 YouTube [32]

数据集。Waymo 数据集包含来自城市驾驶场景的高分辨率图像(1920×1280), 并提供精确的相机位姿; YouTube 数据集则涵盖多样化的户外场景, 具有不同的环境与光照条件。在实验中, 我们从 Waymo 数据集中随机选取 9 个场景进行测试, 并采用 YouTube 数据集的标准划分进行评估, 同时使用 7:1 的训练 - 测试划分比例以保证评估的稳健性。重建质量通过 PSNR (峰值信噪比)、SSIM (结构相似度) 和 LPIPS (学习感知图像块相似度) 三个指标进行量化评估。

4.1.2. 实现细节

我们采用两种不同的训练配置以进行全面评估。对于原始分辨率输入, 使用较小的优化参数(位置学习率为 0.000016, 尺度学习率为 0.001, 密度比例为 0.05%), 并每 50 次迭代更新一次 CVGE 分支, 以加速训练过程。对于 $4\times$ 下采样输入, 则采用较大的参数设置(位置学习率为 0.00016, 尺度学习率为 0.005, 密度比例为 1%), 并每 15 次迭代更新一次 CVGE 分支, 以保证训练的有效性与稳定性。所有实验均在 NVIDIA RTX 5000 Ada GPU 上完成。

4.2. 原始分辨率下的实验结果

如表 1 所示, 在原始分辨率输入条件下, 我们的方法在两个数据集上均取得了当前最优性能。在具有挑战性的 Waymo 数据集上, 我们的方法取得了 PSNR = 34.20、SSIM = 0.945 和 LPIPS = 0.204, 相比于性能最接近的 Spec-Gaussian 分别提升了 0.17 dB、0.002 和 0.006。在 YouTube 数据集上, 我们同样在重建质量方面占据领先优势, PSNR 达到 36.21, 优于 Spec-Gaussian 的 36.13。值得注意的是, 我们的方法在几何一致性指标方面表现尤为突出。在 YouTube 数据集上, 相较于 GaussianPro, 我们在 LPIPS 指标上取得了约 14.3% 的提升, 这充分体现了 AVDA 模块在复杂户外场景下的有效性。在效率方面, 如表 2 所示, 我们的方法在保持更优重建质量的同时, 仍具有较高的计算效率。相较于 Spec-Gaussian 和 GaussianPro, AGES 在训练时间和推理开销上均未引入显著额外负担, 整体效率保持在可比甚至更优的水平。这一优势主要得益于 AVDA 模块的轻量化设计, 其通过各向异性外观建模与几何一致性约束的联合优化, 在避免复杂后处理步骤的同时, 实现了更稳定的收敛行为。因此, AGES 在质量与效率之间取得了良好的平衡, 使其在实际大规模场景重建任务中具备更高的应用价值。此外, 我们进一步分析了在极端低质量 SfM 初始化条件下的方法鲁棒性。如图 4 所示, 我们选取了若干 SfM 初始化质量极差的挑战性场景进行实验, 此时初始点云稀疏且存在明显几何错位, 给后续重建带来了极大的不确定性。在这种极端条件下, 基线方法往往出现几何结构塌缩或噪声显著放大的问题, 而引入 AVDA 模块的 AGES 仍能够在一定程度上稳定优化过程, 并保持相对一致的几何结构。尽管在部分区域仍存在不可避免的失败情况, 但整体结果表明, AGES 在面对严重退化的初始化时具有更强的鲁棒性, 其各向异性外观建模与几何一致性约束能够有效抑制由初始化误差引发的伪影传播。这一实验结果明确揭示了方法的适用边界, 同时也验证了 AGES 在极端条件下的稳定性优势。这一性能优势主要源于我们的方法能够在纹理稀疏区域保持稳定的几何结构, 而这些区域往往是基线方法容易失败的地方。如图 5 所示的可视化对比进一步验证了我们方法在细节重建方面的优势。在道路表面区域, 我们的方法能够保持清晰的几何边界和准确的表面法向, 而对比方法则出现了明显伪影: 3DGS 在路面标线处产生模糊重建, GaussianPro 引入了过度平滑, Spec-Gaussian 则存在残余噪声模式。得益于各向异性外观建模与几何一致性约束的协同作用, 我们的方法能够更加真实地复现材质属性与表面几何。放大对比结果表明, 我们的方法在不同材质(如沥青与路面标线)之间保持了更干净的过渡, 并生成了更加符合物理规律的高光反射。这些视觉改进与 SSIM 和 LPIPS 的定量优势高度一致, 验证了我们所提出的统一框架有效缓解了传统 3DGS 方法中外观与几何不一致的问题。



Figure 4. Performance under extremely poor SfM initialization
图 4. 极端低质量 SfM 初始化条件下的性能表现

Table 1. Comparison of visual quality metrics at full resolution
表 1. 原始分辨率下的视觉质量指标比较

Dataset	Waymo			YouTube		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS [26]	33.68	0.938	0.228	34.81	0.960	0.084
Stop-the-Pop [36]	33.38	0.936	0.230	34.87	0.960	0.083
GaussianPro [32]	34.01	0.943	0.205	35.53	0.961	0.070
Spec-Gaussian [27]	34.03	0.943	0.210	36.13	0.969	0.061
Ours	34.20	0.945	0.204	36.21	0.970	0.060

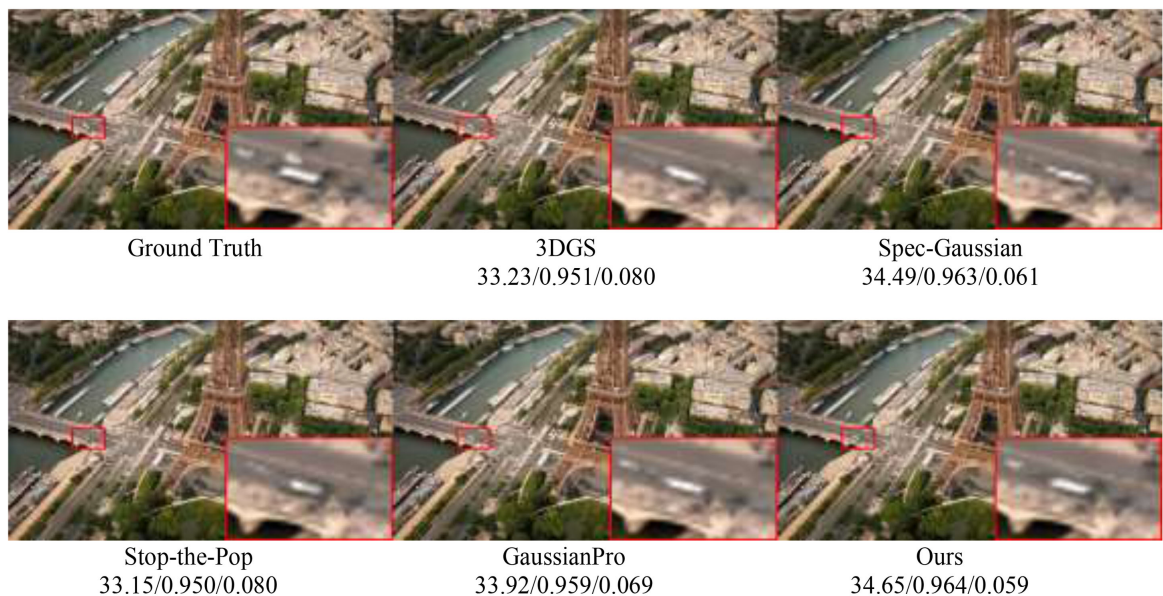


Figure 5. Comparison results of fine-grained local structures on the YouTube dataset (YouTube04: Eiffel Tower) at full resolution, along with the corresponding PSNR/SSIM/LPIPS index

图 5. 在全分辨率下 YouTube 数据集(YouTube04: 埃菲尔铁塔)细粒度局部结构的对比结果及对应的 PSNR/SSIM/LPIPS 指标

Table 2. Comparison of computational efficiency at full resolution
表 2. 原始分辨率下的效率指标比较

Dataset	Waymo			YouTube		
Method	训练时间	推理时间/帧	GPU 显存	训练时间	推理时间/帧	GPU 显存
3DGS [26]	45 min	1.3 s	357 mb	37 min	1.1	328 mb
Stop-the-Pop [36]	41 min	1.2 s	342 mb	35 min	0.9	317 mb
GaussianPro [32]	37 min	1.2 s	339 mb	31 min	0.9	309 mb
Spec-Gaussian [27]	35 min	1.1 s	334 mb	27 min	0.8	295 mb
Ours	32 min	0.8 s	325 mb	25 min	0.7	287 mb

4.3. 4×下采样分辨率下的实验结果

表 3 给出了在更具挑战性的 4×下采样训练设置下的实验结果，在该条件下，各方法需要应对显著的信息损失。我们的方法表现出卓越的鲁棒性，在 Waymo 数据集上取得了 PSNR = 36.00、SSIM = 0.962、LPIPS = 0.084 的最优成绩，三项指标均排名第一。相比 Spec-Gaussian，我们在 PSNR 上提升了 0.18 dB，而相较于 GaussianPro 更是取得了 0.80 dB 的显著优势，这表明其基于渐进传播的策略在分辨率降低时更容易受到影响。在 YouTube 数据集上的结果进一步验证了我们方法的泛化能力。尽管 GaussianPro 与 Spec-Gaussian 在 SSIM 指标上均达到 0.986，我们的方法在 PSNR (38.70)和 LPIPS (0.017)上仍明显优于它们，说明我们在结构细节和感知质量方面都具有更好的保持能力。这种在不同数据集上的一致性提升表明，外观与几何之间的双向互补约束能够在多样化的数据条件下稳定提升重建质量。如图 6 所示，即使在强烈下采样的情况下，我们的方法仍能保持较高的视觉质量。相比所有基线方法，我们的重建结果具有更清晰的边缘和更少的伪影，建筑立面细节与植被结构也得到了更好的保留，这进一步说明由外观驱动的几何补偿与几何感知的外观建模共同构成的双域融合策略，可以有效缓解低分辨率输入带来的退化问题。这种鲁棒性具有重要的实际意义：在保持高质量重建性能的同时，我们的方法能够显著降低计算成本。在计算效率方面，如表 4 所示，在 4×下采样分辨率的训练设置下，AGES 进一步展现出显著的效率优势。相较于原始分辨率训练，各方法的计算开销均明显降低，而我们的方法在保持最优重建质量的同时，实现了更为可观的训练与推理效率提升。具体而言，AGES 在 4×下采样条件下的训练时间较原始分辨率减少约 65%，显著优于对比方法。这一结果表明，所提出的双域融合策略不仅在低分辨率输入下具备更强的鲁棒性，同时也能够有效降低计算成本，使其在资源受限或大规模场景重建任务中具有更高的实用价值。

Table 3. Comparison of visual quality metrics at 4× downsampled resolution
表 3. 4×下采样分辨率下的视觉质量指标比较

Dataset	Waymo			YouTube		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS [26]	35.94	0.961	0.090	38.48	0.985	0.021
Stop-the-Pop [36]	35.28	0.957	0.098	37.82	0.982	0.023
GaussianPro [32]	35.20	0.956	0.093	38.56	0.986	0.020
Spec-Gaussian [27]	35.82	0.958	0.090	38.66	0.986	0.018
Ours	36.00	0.962	0.084	38.70	0.986	0.017

Table 4. Comparison of computational efficiency at 4× downsampled resolution
表 4. 4×下采样分辨率下的效率指标比较

Dataset	Waymo			YouTube		
Method	训练时间	推理时间/帧	GPU 显存	训练时间	推理时间/帧	GPU 显存
3DGS [26]	15 min	0.8 s	297 mb	11 min	0.8	172 mb
Stop-the-Pop [36]	27 min	0.9 s	305 mb	14 min	0.8	204 mb
GaussianPro [32]	13 min	0.8 s	207 mb	9 min	0.7	150 mb
Spec-Gaussian [27]	12 min	0.7 s	198 mb	8 min	0.6	145 mb
Ours	10 min	0.5 s	174 mb	6 min	0.4	132 mb



Figure 6. Comparison of restored scenes on the Waymo dataset (Segment-102751) at a 4× downsampled resolution, together with the corresponding PSNR/SSIM/LPIPS index

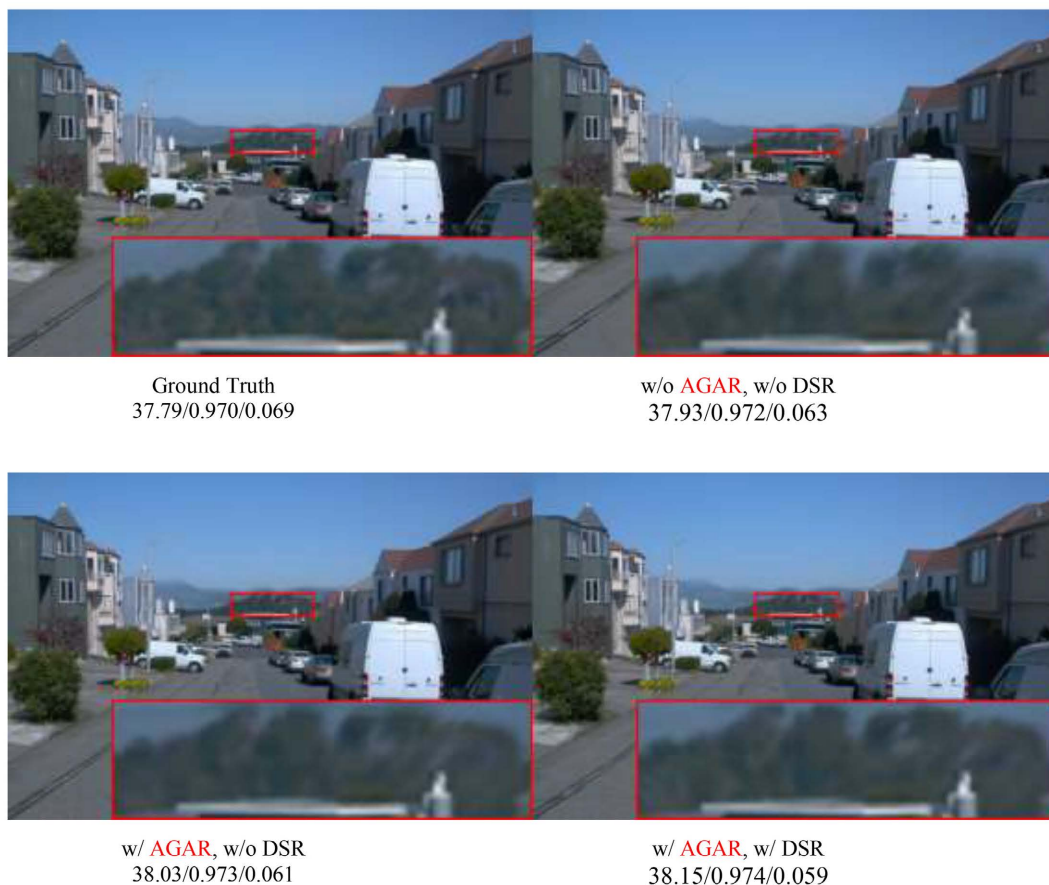
图 6. Waymo 数据集(Segment-102751)在 4×下采样分辨率下的重建场景对比及对应的 PSNR/SSIM/LPIPS 指标

4.4. 消融实验

表 5 对我们框架中的核心组件进行了系统性的评估。在 Waymo 数据集上, 基线模型(不包含 AGAR 与 DSR)取得了 PSNR = 35.82、SSIM = 0.958、LPIPS = 0.090 的结果。仅引入 AGAR 后, 各项指标均得到稳定提升(PSNR = 35.91、SSIM = 0.961、LPIPS = 0.086), 表明外观与几何之间的交互式更新能够显著提升重建质量。在此基础上进一步加入 DSR 后, 性能继续提升至 PSNR = 36.00、SSIM = 0.962、LPIPS = 0.084, 说明该正则项通过强化局部几何一致性提供了互补性的收益。图 7 给出的可视化消融结果直观展示了这些逐步改进。基线模型在几何边界处存在明显噪声; 引入 AGAR 后, 外观与几何之间增强的信息流有效抑制了这些伪影; 进一步加入 DSR 则使边界更加锐利, 表面结构更加连贯一致。综合来看, 这些结果验证了联合优化外观建模与几何一致性对于获得高质量三维重建至关重要。

Table 5. Ablation experiment on Waymo dataset**表 5.** Waymo 数据集上的消融实验

Method		PSNR	SSIM	LPIPS
w/o AGAR	w/o DSR	35.82	0.958	0.090
w/ AGAR	w/o DSR	35.91	0.961	0.086
w/ AGAR	w/ DSR	36.00	0.962	0.084

**Figure 7.** Ablation study on the Waymo dataset (Segment-100613) at a 4× downsampled resolution, along with the corresponding PSNR/SSIM/LPIPS values**图 7.** 在 4×下采样分辨率下 Waymo 数据集(Segment-100613)的消融实验结果及对应的 PSNR/SSIM/LPIPS 指标

5. 结论

本文提出了 AGES 框架，一个通过将各向异性的视角相关外观建模与鲁棒的几何约束相结合来解决 3D Gaussian Splatting 中几何不一致性问题的统一框架。我们的方法通过两个关键创新实现了显著性能提升：一是 AGAR 模块，其引入了方向感知的外观建模以加强外观与几何之间的耦合；二是提出了新的 DSR 正则项，用于约束局部几何一致性。实验结果表明，我们的方法在定量指标和视觉质量方面均优于现有最先进方法，尤其在保持细粒度几何细节和复杂表面结构方面表现突出。然而，当前框架在处理高度动态场景时仍存在局限，并且依赖于较为准确的相机位姿估计，这在存在大幅运动的实时场景中可能会限制其适用性。未来工作将致力于将该框架扩展到动态环境，开发对位姿误差更鲁棒的变体，并探索在复杂光照条件下的自适应几何增强策略。此外，引入语义理解还有望进一步提升方法在遮挡区域和反射表面上的建模能力。

参考文献

- [1] Weng, C., Curless, B., Srinivasan, P.P., Barron, J.T. and Kemelmacher-Shlizerman, I. (2022) HumanNeRF: Free-View-point Rendering of Moving People from Monocular Video. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 16189-161999. <https://doi.org/10.1109/cvpr52688.2022.01573>
- [2] Li, T., Slavcheva, M., Zollhoefer, M., Green, S., Lassner, C., Kim, C., et al. (2022) Neural 3D Video Synthesis from Multi-View Video. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5511-5521. <https://doi.org/10.1109/cvpr52688.2022.00544>
- [3] Lu, F., Xu, Y., Chen, G., Li, H., Lin, K. and Jiang, C. (2023) Urban Radiance Field Representation with Deformable Neural Mesh Primitives. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 165-176. <https://doi.org/10.1109/iccv51070.2023.00049>
- [4] Jiang, Y., Liao, Q., Li, X., Ma, L., Zhang, Q., Zhang, C., et al. (2025) UV Gaussians: Joint Learning of Mesh Deformation and Gaussian Textures for Human Avatar Modeling. *Knowledge-Based Systems*, **320**, Article ID: 113470. <https://doi.org/10.1016/j.knosys.2025.113470>
- [5] Wen, X., Sun, K., Chen, T., Wang, Z., She, J., Zhao, Q., et al. (2025) A Nerf-Based Technique Combined Depth-Guided Filtering and View Enhanced Module for Large-Scale Scene Reconstruction. *Knowledge-Based Systems*, **316**, Article ID: 113411. <https://doi.org/10.1016/j.knosys.2025.113411>
- [6] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R. (2020) NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer, 405-421. https://doi.org/10.1007/978-3-030-58452-8_24
- [7] Zhu, F., Guo, S., Song, L., Xu, K. and Hu, J. (2023) Deep Review and Analysis of Recent Nerfs. *APSIPA Transactions on Signal and Information Processing*, **12**, 1-32. <https://doi.org/10.1561/116.00000162>
- [8] Zhang, X., Fanello, S., Tsai, Y., Sun, T., Xue, T., Pandey, R., et al. (2021) Neural Light Transport for Relighting and View Synthesis. *ACM Transactions on Graphics*, **40**, 1-17. <https://doi.org/10.1145/3446328>
- [9] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T. and Srinivasan, P.P. (2022) Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5481-5490. <https://doi.org/10.1109/cvpr52688.2022.00541>
- [10] Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P. and Hedman, P. (2022) Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5460-5469. <https://doi.org/10.1109/cvpr52688.2022.00539>
- [11] Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B.P., Srinivasan, P., et al. (2022) Block-NeRF: Scalable Large Scene Neural View Synthesis. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 8238-8248. <https://doi.org/10.1109/cvpr52688.2022.00807>
- [12] Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R. and Srinivasan, P.P. (2021) Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5835-5844. <https://doi.org/10.1109/iccv48922.2021.00580>
- [13] Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., et al. (2021) Nerfies: Deformable Neural Radiance Fields. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5845-5854. <https://doi.org/10.1109/iccv48922.2021.00581>
- [14] Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A. and Duckworth, D. (2021) NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 7206-7215. <https://doi.org/10.1109/cvpr46437.2021.00713>
- [15] Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J.H., Chaitanya, C.R.A., et al. (2021) DONeRF: Towards Real-time Rendering of Compact Neural Radiance Fields Using Depth Oracle Networks. *Computer Graphics Forum*, **40**, 45-59. <https://doi.org/10.1111/cgf.14340>
- [16] Garbin, S.J., Kowalski, M., Johnson, M., Shotton, J. and Valentin, J. (2021) FastNeRF: High-Fidelity Neural Rendering at 200FPS. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 14326-14335. <https://doi.org/10.1109/iccv48922.2021.01408>
- [17] Zhang, Y., Wei, J., Zhou, B., Li, F., Xie, Y. and Liu, J. (2024) TVNeRF: Improving Few-View Neural Volume Rendering with Total Variation Maximization. *Knowledge-Based Systems*, **301**, Article ID: 112273. <https://doi.org/10.1016/j.knosys.2024.112273>
- [18] Wang, F., Yin, L., Qin, Y., Gao, X., Tang, X. and Zhou, H. (2025) Ray-Decomposed and Gradient-Constrained Nerf for Few-Shot View Synthesis under Low-Light Conditions. *Knowledge-Based Systems*, **330**, Article ID: 114568.

- <https://doi.org/10.1016/j.knosys.2025.114568>
- [19] Hermann, M., Kwak, H., Ruf, B. and Weinmann, M. (2024) Leveraging Neural Radiance Fields for Large-Scale 3D Reconstruction from Aerial Imagery. *Remote Sensing*, **16**, Article 4655. <https://doi.org/10.3390/rs16244655>
 - [20] Xie, S., Zhang, L., Jeon, G. and Yang, X. (2023) Remote Sensing Neural Radiance Fields for Multi-View Satellite Photogrammetry. *Remote Sensing*, **15**, Article 3808. <https://doi.org/10.3390/rs15153808>
 - [21] Liu, L., Gu, J., Zaw Lin, K., Chua, T.S. and Theobalt, C. (2020) Neural Sparse Voxel Fields. *Advances in Neural Information Processing Systems*, **33**, 15651-15663
 - [22] Yu, A., Li, R., Tancik, M., Li, H., Ng, R. and Kanazawa, A. (2021) PlenOctrees for Real-Time Rendering of Neural Radiance Fields. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5732-5741. <https://doi.org/10.1109/iccv48922.2021.00570>
 - [23] Hu, T., Liu, S., Chen, Y., Shen, T. and Jia, J. (2022) EfficientNeRF—Efficient Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 12892-12901. <https://doi.org/10.1109/cvpr52688.2022.01256>
 - [24] Reiser, C., Peng, S., Liao, Y. and Geiger, A. (2021) KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 14315-14325. <https://doi.org/10.1109/iccv48922.2021.01407>
 - [25] Hedman, P., Srinivasan, P.P., Mildenhall, B., Barron, J.T. and Debevec, P. (2021) Baking Neural Radiance Fields for Real-Time View Synthesis. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5855-5864. <https://doi.org/10.1109/iccv48922.2021.00582>
 - [26] Kerbl, B., Kopanas, G., Leimkuehler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, **42**, 1-14. <https://doi.org/10.1145/3592433>
 - [27] Gao, X., Huang, Y., Jiao, S., Jin, X., Lyu, X., Qi, X., et al. (2024) Spec-Gaussian: Anisotropic View-Dependent Appearance for 3D Gaussian Splatting. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 61192-61216. <https://doi.org/10.52202/079017-1956>
 - [28] Yan, Z., Low, W.F., Chen, Y. and Lee, G.H. (2024) Multi-Scale 3D Gaussian Splatting for Anti-Aliased Rendering. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 20923-20931. <https://doi.org/10.1109/cvpr52733.2024.01977>
 - [29] Yu, Z., Chen, A., Huang, B., Sattler, T. and Geiger, A. (2024) Mip-Splatting: Alias-Free 3D Gaussian Splatting. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 19447-19456. <https://doi.org/10.1109/cvpr52733.2024.01839>
 - [30] Meng, J., Li, H., Wu, Y., Gao, Q., Yang, S., Zhang, J., et al. (2024) Mirror-3dgs: Incorporating Mirror Reflections into 3D Gaussian Splatting. 2024 *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, Tokyo, 8-11 December 2024, 1-5. <https://doi.org/10.1109/vcip63160.2024.10849936>
 - [31] Gao, J., Gu, C., Lin, Y., Li, Z., Zhu, H., Cao, X., et al. (2024) Relightable 3D Gaussians: Realistic Point Cloud Relighting with BRDF Decomposition and Ray Tracing. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Computer Vision—ECCV 2024*, Springer, 73-89. https://doi.org/10.1007/978-3-031-72995-9_5
 - [32] Cheng, K., Long, X., Yang, K., Yao, Y., Yin, W., Ma, Y., Wang, W. and Chen, X. (2024) GaussianPro: 3d Gaussian Splatting with Progressive Propagation. arXiv: 2402.14650.
 - [33] Zhang, J., Zhan, F., Xu, M., Lu, S. and Xing, E. (2024) FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 21424-21433. <https://doi.org/10.1109/cvpr52733.2024.02024>
 - [34] Yu, Z., Chen, Z., Zhou, Z. and Cao, H. (2025) CGC-GS: Cross Geometric Cues Constrained Gaussian Splatting. *Knowledge-Based Systems*, **330**, Article ID: 114630. <https://doi.org/10.1016/j.knosys.2025.114630>
 - [35] Sun, P., Kretschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., et al. (2020) Scalability in Perception for Autonomous Driving: Waymo Open Dataset. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 2443-2451. <https://doi.org/10.1109/cvpr42600.2020.00252>
 - [36] Radl, L., Steiner, M., Parger, M., Weinrauch, A., Kerbl, B. and Steinberger, M. (2024) StopThePop: Sorted Gaussian Splatting for View-Consistent Real-Time Rendering. *ACM Transactions on Graphics*, **43**, 1-17. <https://doi.org/10.1145/3658187>