

多重采样抗锯齿和三维高斯点投影的约束优化

Zheng Zhou^{1†}, Jia-Chen Zhang^{1†}, Yu-Jie Xiong^{1*}, Chun-Ming Xia¹

^{1*}School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China.

*Corresponding author(s). E-mail(s): xiong@sues.edu.cn;

Contributing authors: m320123332@sues.edu.cn;

m325123603@sues.edu.cn; cmxia@sues.edu.cn;

[†]These authors contributed equally to this work.

Abstract

Recent advances in 3D Gaussian splatting have significantly improved real-time novel view synthesis, yet insufficient geometric constraints during scene optimization often result in blurred reconstructions of fine-grained details, particularly in regions with high-frequency textures and sharp discontinuities. To address this, we propose a comprehensive optimization framework integrating multisample anti-aliasing (MSAA) with dual geometric constraints. Our system computes pixel colors through adaptive blending of quadruple subsamples, effectively reducing aliasing artifacts in high-frequency components. The framework introduces two constraints: (a) an adaptive weighting strategy that prioritizes under-reconstructed regions through dynamic gradient analysis, and (b) gradient differential constraints enforcing geometric regularization at object boundaries. This targeted optimization enables the model to allocate computational resources preferentially to critical regions requiring refinement while maintaining global consistency. Extensive experimental evaluations across multiple benchmarks demonstrate that our method achieves state-of-the-art performance in detail preservation, particularly in preserving high-frequency textures and sharp discontinuities, while maintaining real-time rendering efficiency. Quantitative metrics and perceptual studies confirm statistically significant improvements over baseline approaches in both structural similarity (SSIM) and perceptual quality (LPIPS).

Keywords: Rendering, point-based models, rasterization, machine learning



Fig. 1: 渲染图像与真实图像之间的区别。与第一行的图像相比，3DGS 渲染的窗户失去了其几何形状，边缘变得不清晰。第二行图像的比较显示，3DGS 的渲染结果缺失了一些内容。

1 介绍

新视图合成 (NVS) 已成为 3D 计算机视觉的重要能力 [1]，能够从任意视角生成真实感图像——这是虚拟现实、计算摄影和沉浸式媒体等应用的重要需求。传统方法从多视角输入中重建明确的 3D 场景几何，以确保视图一致性，但在复杂几何或照明条件下常常面临保真度的限制 [2]。神经辐射场 (NeRF) [3] 的出现通过将隐式场景表示与可微体积渲染相结合，标志着一个范式转变，在视图合成质量上达到了前所未有的水平。随后衍生的方法通过 cone tracing [4, 5]、哈希网格编码 [6] 和稀疏体积表示 [7, 8] 等策略解决了效率问题。然而，这些方法在加速渲染时常常牺牲高频细节，特别是在高分辨率范围内。3D 高斯溅射 (3DGS) [9] 提供了一种引人注目的替代方案，通过自适应密度控制将场景显式建模为各向异性的 3D 高斯。与 NeRF 的计算密集型体积积分不同，3DGS 采用高效的基于溅射的投影渲染，实现实时性能同时保持几何保真度。然而，如图 1 所示，3DGS 在稀疏观测数据的区域表现出明显的局限性，因采样密度不足而表现为局部细节退化和模糊伪影 [9]。

为了解决这个问题，我们提出了一种新的 3DGS 框架，将多重采样抗锯齿 [10, 11] 和边缘感知约束 [12, 13] 进行整合，通过多重采样和特定的约束来减少渲染模糊并丰富局部细节。通常，3D 高斯投影仅通过与像素中心相对应的高斯椭球进行渲染，这可能导致僵硬的过渡和颜色跳跃 [14]。我们使用 4 个子像素点进行多重采样混合渲染。结合子像素点的信息进行综合渲染，将导致更平滑的颜色过渡。同时，考虑到不同高斯椭球覆盖区域的影响，多重采样渲染方法可以更加关注颜色细节，使图像渲染得更加逼真。在损失函数中，我们引入了基于误差大小的自适应权重 [15] 来解决预测结果与实际图像之间的差异。这样，当某些像素或区域的误差较大时，在训练过程中它们的影响将被放大，使网络更多地关注这些重建效果较弱或难以预测的区域，从而避免局部细节中过度平滑或纹理丢失的现象。为了增强物体边缘以及高频纹理等细节的重建效果，我们在原有损失之上额外添加了梯度差异损失 [16, 17]。通过在梯度域比较预测图像和真实图像之间的差异，模型更多地关注图像的局部梯度信息。与简单的 L1/L2 [18] 或 SSIM 损失 [19] 相比，梯度差异约束可以更好地捕捉图像中的高频纹理和结构，导致更尖锐和丰富的视觉细节。该工作的贡献可以总结为三个关键方面。

- 一种用于 3D 图形系统的 4—MSAA 光栅化管道，通过子像素感知的高斯混合来减少锯齿伪影。
- 结合误差自适应加权和梯度差约束的混合损失函数，以增强细节重建。
- 全面的基准测试在多个数据集上的定量指标和感知评估中展示了最新的性能。

早期的新视图合成方法通常依赖于卷积神经网络 (CNNs) 来融合和扭曲多视图输入以生成未见过的视点。这些方法通常计算像素或特征级别的混合权重，将源图像组合成近似的目标视图。当视点变化适度时，它们表现良好，但在处理更复杂的几何形状和更大视点移动时常常遇到困难，导致伪影和细节模糊。后续研究转向了体积光线行进和储存在 3D 网格中的场景表示。例如，DeepVoxels 引入了一种持久的基于体素的表示，光线在其中采样 3D 特征以合成新视图，与基于 2D 图像的混合相比，空间一致性得到了改善。然而，这些显式的体积方法在高分辨率下可能变得计算和存储密集，从而限制了它们在大规模或高保真场景中的可扩展性。为了应对这些限制，研究转向了 NeRF，该方法通过多层感知机 (MLPs) 隐式地表示场景。NeRF 通过预测任意连续 3D 坐标的密度和颜色，取得了逼真的新视角，并具有强大的多视图一致性。自此，已经提出了许多 NeRF 扩展，以应对移动物体、未知相机姿态、少镜头设置、以及抗锯齿等挑战。然而，传统的基于 NeRF 的流程通常需要大量的训练时间，实现实时渲染仍然是一个挑战。目前的加速策略——包括更小的专业化 MLPs、张量分解、基于哈希的编码和各种压缩技术——有助于减轻这些瓶颈，但可能削弱高频细节，特别是在较高分辨率下。尽管取得了进展，隐式方法在训练速度和实时渲染方面面临着根本性的挑战，这激励了混合表示模式的出现。

为了在渲染速度、逼真度和数据效率之间取得更好的平衡，3D 高斯喷射作为一种有前途的替代方案出现。这种方法采用在 3D 空间中显式表示的各向异性高斯，通过端到端优化的高斯参数来建模局部几何和颜色。与依赖于大体素网格或完全隐式函数的方法不同，3D 高斯喷射利用可微分的喷射过程将这些高斯投射到图像平面上，从而在保持高保真度细节的同时，实现更快的训练和实时渲染。Mip-splatting 通过引入类似于 mipmapping 的多尺度表示进一步解决了 3D 高斯点绘中的别名问题 [20]。这种技术不是为整个场景渲染单一层次的高斯，而是将高斯点划分为不同分辨率的层。在渲染过程中，流水线会根据距离和视图条件自适应地选择最合适的层次，从而减少别名并提高整体渲染效率。多尺度 3D 高斯点绘的进一步发展引入了一种多分辨率策略，以在基于点绘的渲染中减轻别名现象 [21]。通过在不同尺度上通过分层高斯点表示几何和颜色信息，这种技术根据观看距离和细节要求自适应地选择最合适的分辨率水平，从而减少由欠采样或过采样引起的伪影。此方法实现了高保真、抗锯齿效果，同时保持了计算效率。Scaffold-GS 提供了一种以分层或基于网格的方式组织高斯点的结构化框架 [22]。通过将几何和颜色信息嵌入到类似脚手架的结构中，该方法根据每个观看条件自适应地选择最佳细节水平，有效地平衡视觉保真度和计算开销。尽管具有这些优势，3D 高斯点绘在某些区域高斯过多或冗余堆积时仍可能出现过度重建。这种重叠的高斯可能导致模糊和伪影，特别是在高对比度边缘或精细细节纹理附近 [23, 24]。

1.1 神经渲染中的细节保留

在神经渲染中追求高频细节的保留已催生了多种技术方法，每种方法都解决了别名与细节之间权衡的特定方面。传统图形学启发的抗锯齿技术，如 MSAA [25] 和时间抗锯齿 (TAA) [26]，通过超采样或帧累积实现了亚像素平滑，但由于梯度计算限制，它们在可微渲染流水线中的直接应用仍然具有挑战性。与此同时，频域优化策略已经

成为补充解决方案，方法如 FreGS [27] 通过利用在傅里叶空间中可以容易地用低通和高通滤波器提取的低到高频成分执行从粗到细的高斯密化。尽管取得了这些进展，目前的解决方案主要将别名伪影和细节侵蚀作为单独的挑战处理——MSAA 变体专注于信号空间平滑而忽略了纹理的锐度，而基于梯度的约束改善了局部对比度但难以处理亚像素不连续性。这种分化源于通过统一的损失景观共同优化几何稳定性和频谱保真度的固有困难，我们的方法通过双域正则化弥合了这一差距。

为了解决这些问题，我们提出了一种方法，该方法结合了 MSAA、自适应加权策略和梯度差约束。具体来说，我们将 MSAA 集成到光栅化过程中，以减少锯齿边缘并保留高频细节，介绍了一种逐像素加权机制，以强调具有较大重建误差的区域，并采用梯度感知损失来锐化边缘和纹理。通过同时针对反走样和细微细节还原，我们的方法提升了局部保真度，同时保持了 3D 高斯喷射技术的实时优势。

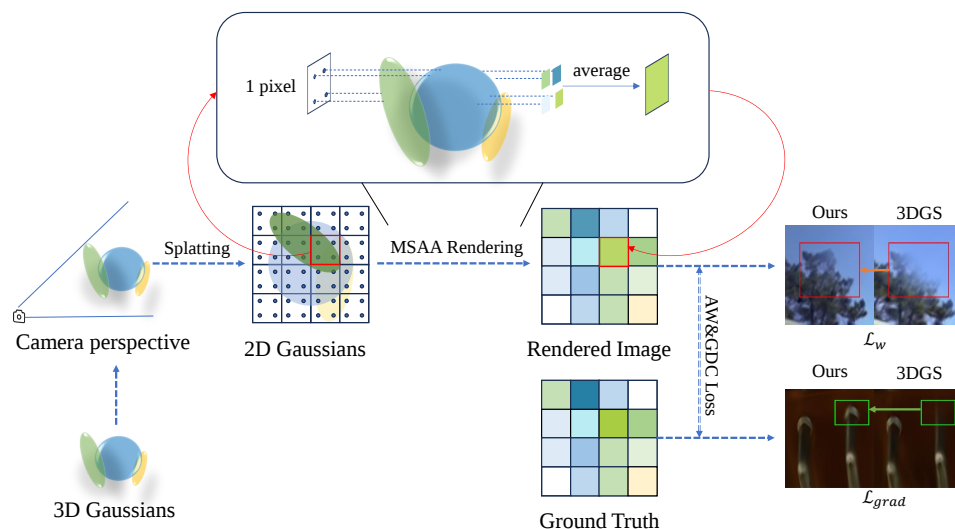


Fig. 2: 所提出方法的概述。我们的方法首先在光栅化阶段为每个像素设置四个偏移采样点。每个采样点独立执行 alpha 混合以渲染场景，然后将它们的颜色平均以生成每个像素的最终颜色。此过程有效地减少了锯齿边缘，同时增强了模型捕捉高频和低频细节的能力。在反向传播过程中，我们引入了两种互补的约束：自适应加权策略和梯度差异约束。自适应加权策略针对重建不足的区域，而梯度差异约束则着重于缺失的边界细节。通过引导模型关注这些不清晰的渲染区域，我们的方法最终提升了整体渲染质量。

在本节中，我们首先回顾 3D 高斯展开流水线，包括 3D 高斯如何被参数化和投影到图像平面上。然后我们描述我们提出的增强方法，包括三个主要组件：

这些技术共同解决了天真喷射的主要缺点，特别是细节的丢失和混叠伪影。完整的流程在图 2 中显示。

1.2 预备知识：三维高斯点撒

我们假设我们的场景由 N 各向异性 3D 高斯分布表示，记作 $\{G_n\}_{n=1}^N$ 。每个高斯分布 G_n 由一个中心 $\mu_n \in \mathbb{R}^3$ 、一个协方差矩阵 $\Sigma_n \in \mathbb{R}^{3 \times 3}$ 描述，该矩阵可分解为 $\Sigma_n = R_n S_n R_n^\top$ ，其中 R_n 是一个旋转矩阵， S_n 是一个对角缩放矩阵，一个颜色 $c_n \in \mathbb{R}^3$ ，以及一个不透明度 $\alpha_n \in [0, 1]$ 。这种显式表示允许每个高斯分布通过基于梯度的优化进行学习和更新。

1.2.1 投影与喷溅

设 $R_{\text{cam}}, t_{\text{cam}}$ 表示相机外部参数， K 是内参数矩阵。一个 3D 点 $x \in \mathbb{R}^3$ 通过小孔模型映射到图像平面上：

$$p = \pi(R_{\text{cam}}x + t_{\text{cam}}, K) \in \mathbb{R}^2, \quad (1)$$

其中 $\pi(\cdot)$ 使用 K 进行标准透视投影。对于每个高斯分布 G_n ，其中心 μ_n 被投影到 $p_n \in \mathbb{R}^2$ 。椭圆形二维特征的协方差 Σ_n^{2D} 从原始三维协方差 Σ_n 中推导而来，考虑了相机投影几何。

在二维像素坐标 $u \in \mathbb{R}^2$ 处， n -th 高斯的贡献可以通过一个高斯权重来近似：在实践中， u 处的颜色是通过加权或阿尔法混合所有重叠在该像素上的高斯来计算的：方程式 (??) – (??) 定义了前向喷溅步骤。



Fig. 3: 在普通的 3DGS 中观察到细节不足和模糊问题。特别是，渲染场景中的吊灯和框架显示出明显的失真和模糊，包括弯曲的墙面和边缘。这些伪像表明 3DGS 在处理锯齿边缘和边界时存在困难，同时难以保持整体场景几何形状，导致最终图像出现失真和模糊。

1.3 多重采样抗锯齿

尽管前向喷洒法维持了场景几何的连续逼近，但它沿高曲率边缘和细微结构引入了别名失真，主要表现为重构边界中的模糊或锯齿状不连续，请参见图 3。这种现象凸显了 3DGS 在保留尖锐几何不连续性和线性特征保真度方面的固有局限性。为了解决这些失真问题，我们将多重采样抗锯齿的原理应用于喷洒范式，通过扩展像素评估到四重子采样方案。具体来说，我们的方法在四个策略性偏移的子像素位置上计算颜色和几何贡献，有效地对边缘区域进行超采样，以减轻高频别名失真。该方法在通过优化子采样权重保持计算效率的同时，强制实施平滑的梯度过渡，在喷洒基础渲染中桥接了几何连续性和边缘清晰度之间的鸿沟。

1.3.1 亚像素采样

设 $n \in \mathbb{Z}^+$ 表示每像素的子像素采样数。为了在保持计算可行性的同时减轻混叠伪影，我们在单位正方形 $[0, 1] \times [0, 1]$ 内定义了一组子像素偏移 $\{\delta_k\}_{k=1}^n$ ，其中每个 $\delta_k = (\delta_{k,x}, \delta_{k,y})$ 表示从像素中心的分数位移。对于图像网格上的每个像素中心 $\mathbf{u}$ ，子像素采样坐标计算如下：在每个子像素 $\mathbf{u}_k$ ，我们执行公式 (??) 中定义的 alpha 合成，将重叠的高斯函数的颜色贡献累积，得到子采样的颜色 $\mathbf{C}_k(\mathbf{u})$ 。这种分层采样策略有效地捕获了高频几何和光度变化，否则在每像素单次采样的光栅化中会丢失。

1.3.2 可微多样本聚合

最终的抗锯齿像素颜色 $\mathbf{C}^{\text{MSAA}}(\mathbf{u})$ 是通过对所有子像素样本的归一化求和计算得出的：

$$\mathbf{C}^{\text{MSAA}}(\mathbf{u}) = \frac{1}{n} \sum_{k=1}^n \mathbf{C}_k(\mathbf{u}). \quad (2)$$

。关键的是，这个聚合运算符保持完全可微性，使得梯度可以通过所有 n 采样路径进行反向传播。链式法则分解了梯度 $\partial \mathbf{C}^{\text{MSAA}} / \partial \Theta$ ，其中 Θ 表示高斯参数。这个特性促使优化过程优先选择那些在子像素扰动情况下展示出视角一致行为的高斯，从而增强几何的稳定性和纹理的真实性。

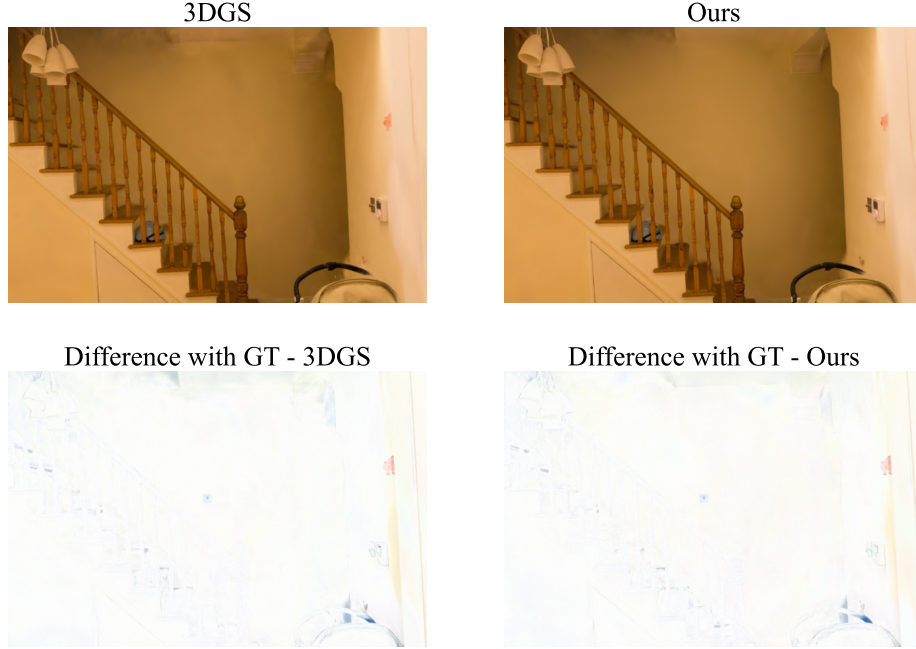


Fig. 4: 不清晰区域的渲染结果。3DGS 在处理不清晰区域时可能会出现重建不足，导致在渲染该区域时产生严重的失调和模糊。我们的方法为重建不足的区域分配更多的权重，这有效地增强了该区域的重建效果，为不清晰区域增加了细节，并产生了更逼真的渲染结果。

1.4 自适应加权策略

虽然 MSAA 能有效减轻锯齿伪影，但在高频区域仍然存在残余重建误差，如图 4 所示。通过将我们模型的输出与 3DGS 渲染进行比较分析，我们生成相对于真实情况的反色差异图 - 较白的区域表示与参考数据更接近。这些可视化揭示了在几何复杂区域（如天花板和楼梯）中显著的重建不足，3DGS 在这些区域表现出大量细节丢失和结构不连续性。为了解决这些局限性，我们在损失函数中实施了一种自适应加权机制，动态优先考虑显示出较高重建误差的像素。这种战略重点使我们的模型能够集中学习能力于常规渲染方法难以应对的区域，特别是在需要高频细节保留的密集几何配置中。

1.4.1 像素级权重定义

设 I_{pred} 和 I_{gt} 是大小为 $H \times W$ 的预测图像和真实图像。对于每个像素 (i, j) ，我们首先计算 ℓ_1 误差：

$$e_{i,j} = \left\| I_{\text{pred}}(i, j) - I_{\text{gt}}(i, j) \right\|_1. \quad (3)$$

为了优先考虑重建不足的区域，我们设计了一个自适应权重图：

$$w_{i,j} = \alpha + (1 - \alpha) \frac{e_{i,j}}{\max_{(p,q) \in \Omega} e_{p,q} + \varepsilon}, \quad (4)$$

其中 $\Omega = \{1, \dots, H\} \times \{1, \dots, W\}$ 表示空间域， $\alpha \in [0, 1]$ 控制基准权重下限， $\varepsilon > 0$ 确保数值稳定性。该公式保证 $w_{i,j} \in [\alpha, 1]$ 相对于局部误差幅度单调增加。

1.4.2 加权重建损失

误差自适应权重调节我们的重建目标：

$$\mathcal{L}_w = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W w_{i,j} \left\| I_{\text{pred}}(i, j) - I_{\text{gt}}(i, j) \right\|_1. \quad (5)$$

为此，我们通过分解的 SSIM 度量 $\mathcal{L}_{\text{D-SSIM}}$ 引入一个结构正则化项，从而得到组合损失：

$$\mathcal{L}_{\text{recon}} = \lambda_1 \mathcal{L}_w + \lambda_2 \mathcal{L}_{\text{D-SSIM}}. \quad (6)$$

在这里， λ_1 和 λ_2 控制加权的 ℓ_1 和 SSIM 的相对影响。通过选择性放大高误差区域的损失。这种选择性误差放大机制特别有利于高频细节恢复，在这种情况下，传统的 ℓ_1 损失表现较差。

1.5 梯度差约束

虽然自适应加权机制有效地优先处理高误差区域，但它缺乏对边缘一致性和高频保真度的显式执行——这一限制在图 5 中得到了体现，其中逐像素损失导致纹理细节的过度平滑。为了弥补这一关键缺陷，我们引入了一个梯度差约束（GDC），直接调节一阶图像变化，确保边界重建的几何保真度。



Fig. 5: 3DGS 边界区域重建误差，将重建的直线边界视为曲线边界。3DGS 对边界信息没有有效的约束，这导致对边界重建的关注不足，使得准确重建边界变得困难。

对于离散图像域 $\Omega = \{1, \dots, H\} \times \{1, \dots, W\}$ ，我们定义了捕捉水平和垂直强度变化的前向差分算子：

$$\nabla_x I_{\text{pred}}(i, j) = I_{\text{pred}}(i, j+1) - I_{\text{pred}}(i, j), \quad (7)$$

$$\nabla_y I_{\text{pred}}(i, j) = I_{\text{pred}}(i+1, j) - I_{\text{pred}}(i, j). \quad (8)$$

这构建了预测和真实图像的梯度向量场：

$$\mathbf{G}_{i,j}^{\text{pred}} = (\nabla_x I_{\text{pred}}(i, j), \nabla_y I_{\text{pred}}(i, j)), \quad (9)$$

$$\mathbf{G}_{i,j}^{\text{gt}} = (\nabla_x I_{\text{gt}}(i, j), \nabla_y I_{\text{gt}}(i, j)). \quad (10)$$

1.5.1 多尺度梯度对齐

我们的 GDC 通过 ℓ_1 - 范数惩罚实现多层次梯度一致性：

$$\mathcal{L}_{\text{grad}} = \frac{1}{(H-1)(W-1)} \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} \left\| \mathbf{g}_{i,j}^{\text{pred}} - \mathbf{g}_{i,j}^{\text{gt}} \right\|_1. \quad (11)$$

我们排除 $i = H$ 或 $j = W$ 的边界索引以避免索引越界。通过最小化水平和垂直图像梯度的差异，网络被引导去再现诸如边缘、角和纹理等的锐利变化。我们在图 6 中展示了使用小波变换的一个案例以获得更清晰的可视化效果。将边界分解为水平和垂直方向进行视觉观察，并比较低频和对角高频信息。GDC 特别增强了 LH/HL 波段的重建，而传统的损失无法保持边缘方向统计。

这种几何正则化通过双重机制补充了自适应加权策略：1) 边缘锐化，直接梯度匹配防止高曲率区域的边界模糊。2) 纹理保持，高频梯度对齐维持随机纹理模式。3) 尺度感知，小波分析显示在频带中的恢复有所改善。

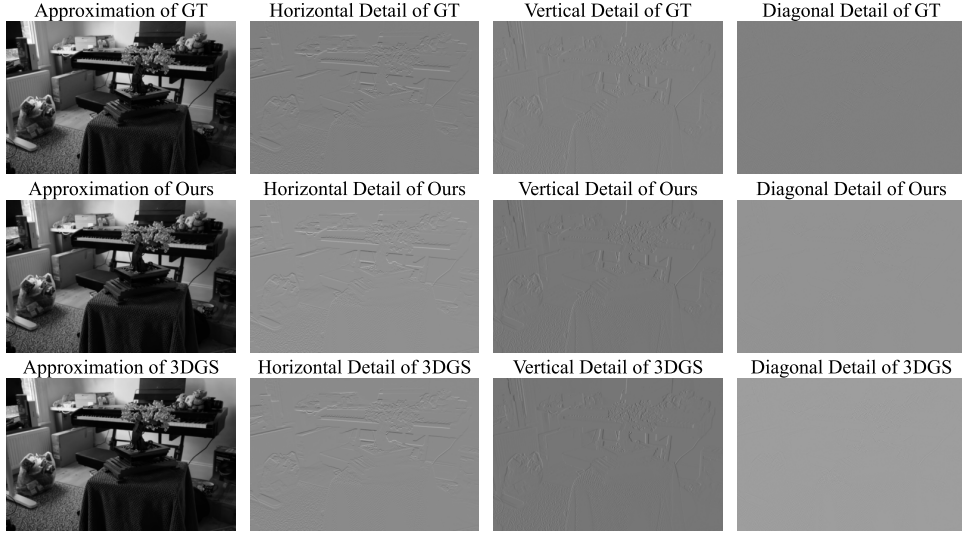


Fig. 6: 我们对图像进行小波变换，以可视化渲染结果的边界信息。小波变换后的水平和垂直细节部分分别显示渲染图像在水平方向和垂直方向的边界信息。显然，与 3DGS 相比，我们的方法具有更清晰、更锐利的边界、更丰富的细节以及更接近真实图像的低频部分。

1.6 总体损失框架

1.6.1 复合目标函数

我们完整的优化目标通过精心校准的耦合系数综合了三种互补机制：

$$\mathcal{L} = \lambda_1 \mathcal{L}_w + \lambda_2 \mathcal{L}_{D-SSIM} + \lambda_3 \mathcal{L}_{grad}, \quad (12)$$

其中三线加权方案 $\lambda_1, \lambda_2, \lambda_3$ 控制：

- 局部误差校正： $\mathcal{L}_w$ 的空间适应性以实现光度精确性
- 结构一致性： $\mathcal{L}_{D-SSIM}$ 的窗口相似性保持
- 几何保真度： $\mathcal{L}_{grad}$ 的边缘方向约束

这里， λ_1 、 λ_2 和 λ_3 平衡了加权重构项、SSIM 和梯度保留。我们在实践中发现，如果边缘仍然过于平滑， λ_3 可以在训练过程中逐渐增加。

当在每次迭代中计算 I_{pred} 时，我们用基于 MSAA 的过程替换了方程 (??) 中的标准像素评估：在像素坐标中， $\mathbf{u} = (j, i)$ 和 $\mathbf{C}^{MSAA}(\cdot)$ 如上所述定义。因此， I_{pred} 实际上是跨多个子像素偏移的平均颜色。

通过结合 MSAA 来减少锯齿，采用自适应加权策略针对高误差区域，以及梯度差约束以保留边缘，我们的方法缓解了 3D 高斯涂抹中的常见问题。这不仅在复杂场景中实现了更真实的局部细节重建，还提高了全局渲染质量。

在训练和测试中，我们遵循 3DGS [9] 的数据集设置，并在共 13 个真实世界的图像上进行了实验。具体来说，我们在 Mip-NeRF360 [4] 数据集的所有九个场景、Tanks & Templates 数据集 [28] 的两个场景和 Deep Blending 数据集 [29] 的两个场景上评估了我们的方法。所选场景展示了多种风格，从有界的室内环境到无限的室外环境。所有图像的分辨率也与 3DGS 中一样。在单个 RTX 4090 GPU 上，所有场景均训练了 30000 次迭代，超参数与 3DGS 一致。评估 3DGS 模型和我们提出的模型在所有数据集上的性能，包括定量和定性分析。

Table 1: 对三个数据集进行定量比较。SSIM $\uparrow$ 和 PSNR $\uparrow$ 的值越高越好；LPIPS $\downarrow$ 的值越低越好。为了公平比较并平衡整体质量和内存消耗之间的权衡，我们使用与 3DGS 相同的设置训练了这些数据集。所有方法使用相同的训练数据进行训练。最佳分数、第二高分和第三高的分数分别为红色、橙色和黄色。

Datasets	Mip-NeRF360			Tanks & Templates			Deep Blending		
Methods	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
Plenoxels	0.626	23.08	0.463	0.719	21.08	0.379	0.795	23.06	0.510
INGP	0.671	25.30	0.371	0.723	21.72	0.330	0.797	23.62	0.423
Mip-NeRF360	0.792	27.69	0.237	0.759	22.22	0.257	0.901	29.40	0.245
3DGS	0.815	27.21	0.214	0.841	23.14	0.183	0.903	29.41	0.243
Pixel-GS	0.823	27.67	0.193	0.856	23.74	0.151	0.896	28.91	0.248
AbsGS	0.820	27.49	0.191	0.853	23.73	0.162	0.902	29.67	0.236
Ours(3DGS)	0.819	27.62	0.207	0.851	23.79	0.165	0.900	29.52	0.246
Ours(Pixel-GS)	0.830	27.68	0.189	0.865	23.75	0.145	0.907	28.92	0.242
Ours(AbsGS)	0.829	27.51	0.188	0.862	23.74	0.157	0.906	29.69	0.230

在我们的渲染重建框架中，我们集成了三个增强功能。首先，为了减少混叠和锯齿边缘，在渲染阶段采用了 $4\times$ MSAA，每像素收集多个样本，这些样本随后在输入层或中间特征融合模块中进行融合。其次，我们的自适应加权策略根据像素或区域级重建误差动态调节损失权重，从而强调错误较高的具有挑战性的区域。最后，为了保留高频细节和锐利的边界，我们在预测图像和真实图像上应用梯度差异约束，并将所得差异融入整体损失函数中。在 Pytorch 框架中使用 Adam 优化器进行训练，集成 MSAA 到 3DGS 的光栅化中，并设置 $\lambda_1, \lambda_2, \lambda_3$ 为 0.8, 0.2 和 0.1。

Table 2: 在 Mip-NeRF360 数据集上的量化比较。SSIM $\uparrow$ 和 PSNR $\uparrow$ 是越高越好，使用 + 来表示我们的方法相比 3DGS 的改进；LPIPS $\downarrow$ 是越低越好，使用 - 来表示我们的方法相比 3DGS 的改进。为了公平比较并在整体质量和内存消耗之间进行平衡，我们使用与 3DGS 相同的设置训练这些数据集。所有方法使用相同的训练数据进行训练。最佳得分 是红色的。

Datasets	bicycle			bonsai			counter		
Methods	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
3DGS	0.745	25.08	0.245	0.947	32.36	0.179	0.915	29.11	0.182
Ours	0.758 (+0.013)	25.23 (+0.15)	0.222 (-0.023)	0.949 (+0.002)	32.44 (+0.08)	0.173 (-0.006)	0.918 (+0.003)	29.12 (+0.01)	0.175 (-0.007)
Datasets	flowers			garden			kitchen		
Methods	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
3DGS	0.589	21.40	0.358	0.856	27.28	0.122	0.931	31.32	0.116
Ours	0.606 (+0.017)	21.69 (+0.29)	0.339 (-0.019)	0.863 (+0.007)	27.41 (+0.13)	0.109 (-0.013)	0.934 (+0.003)	31.45 (+0.13)	0.111 (-0.005)
Datasets	room			stump			treehill		
Methods	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
3DGS	0.926	31.70	0.196	0.768	26.63	0.243	0.635	22.52	0.346
Ours	0.929 (+0.003)	31.78 (+0.08)	0.186 (-0.010)	0.778 (+0.010)	26.81 (+0.18)	0.224 (-0.019)	0.643 (+0.008)	22.63 (+0.11)	0.327 (-0.019)



Fig. 7: 我们提出的方法与当前最先进的方法在新视图合成中的定性对比。对比是在多个不同风格的场景中进行的，包含来自 Mip-NeRF360 的多个场景、Tank & Template 的火车和卡车，以及 Deep Blending 的游乐室。图中的 GT 代表用于参考的真实图像，我们将放大明显不同的细节以便于查看。从这些结果可以看出，我们的方法实现了出色的图像呈现，具有较少的伪影和更丰富的细节。

我们将我们的方法与 3DGS 以及其他视图生成方法进行了比较，包括 MipNerf360 [4]、InstantNGP [6]、Plenoxels [7]、Mip-NeRF360 [5]、3DGS [9]、AbsGS [30] 和 Pixel-GS [31]，在来自三个数据集的 14 个场景中进行了比较。为了确保公平的比较并在内存使用和性能之间取得平衡，我们使用了与 3DGS 相似数量的高斯体。所有方法均使用相同的数据和硬件配置进行训练。表 1 展示了所有方法的定量指标，表 2 详细列出了 MipNerf360 中的九个场景的结果。正如这些分析所示，我们的方法在所有现实世界场景中，在 PSNR、SSIM 和 LPIPS 方面始终优于最先进的方法。此外，图 7 中的定量结果进一步说明了我们卓越的渲染质量，突出显示了我们方法在多样环境中的稳健性。此外，我们使用点云和高斯椭球来证明在图 8 和图 9 中所提出的双重约束的有效性。

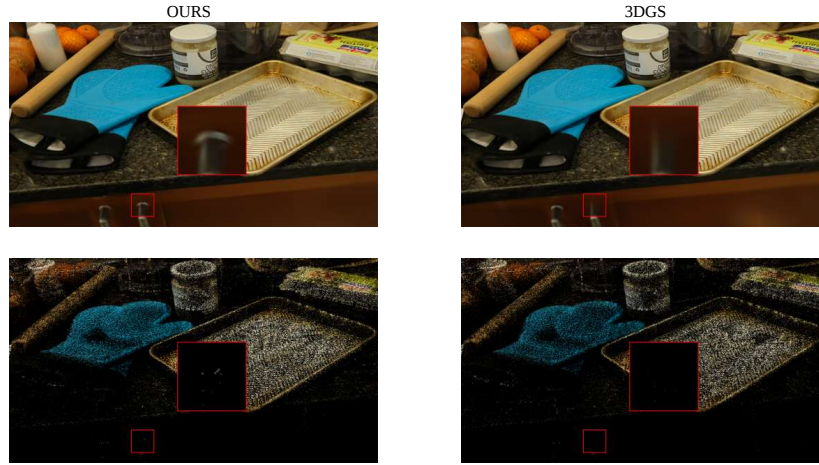


Fig. 8: 点云效果比较图。我们的方法在重建详细点云方面优于 3DGS，特别是在模糊或无纹理区域，能提高局部渲染质量。

为了研究我们提出的改进模块的各自贡献，我们对 MSAA、自适应加权策略 (AWS) 和梯度差约束 (GDC) 进行了消融实验。具体来说，我们从完整模型配置中依次移除或选择性保留每个模块，然后使用标准重建指标对每个变体进行评估和比较。

1.6.2 多重采样抗锯齿

我们首先研究了我们的 MSAA 如何影响高斯分布中的局部模糊。基于 3DGS，我们训练了一个模型 3DGS+MSAA，该模型在光栅化阶段执行 MSAA 策略。正如图 10 所示，3DGS+MSAA 模型在局部细节上更加清晰，MSAA 有效缓解了渲染过程中锯齿边缘和混叠失真，使得网络在处理锐利边缘或高频纹理时不易出现噪声和失焦问题。

1.6.3 自适应加权策略 & 梯度差异约束

为了验证我们提出的约束优化框架的有效性，我们通过训练一个混合模型，即 3DGS+MSAA+AWS+GDC，进行了全面的消融研究，该模型将 MSAA 与我们的自

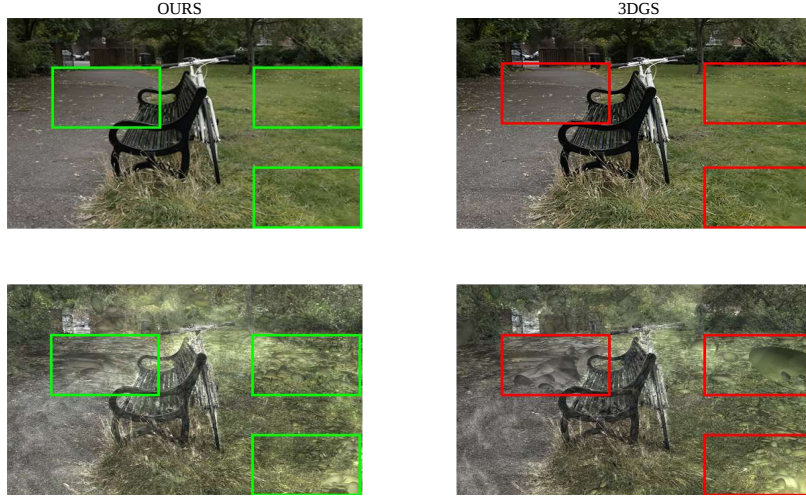


Fig. 9: 高斯椭球覆盖比较图。我们的方法通过将大高斯椭球分解为更细致、细节丰富的组件来增强重建，从而解决了 3DGS 的模糊渲染问题。



Fig. 10: MSAA 消融实验的结果。3DGS 在重建和渲染肖像时生成了一些不存在的部分，降低了渲染结果的清晰度。加入 MSAA 方法后，我们在重建和渲染过程中有效地区分了肖像、空白空间和边框的区域，使每部分的内容更清晰。

适应约束优化策略相结合。具体来说，我们的方法在关键指标上优于 3DGS+MSAA 和 3DGS+AWS+GDC，表明自适应加权和梯度域正则化结合的协同效益。AWS 根据每像素重建误差动态调整困难区域的损失权重。这一机制在训练中优先考虑未优化区域，显著增强了细纹理细节的恢复，同时缓解由不均匀优化焦点引起的视觉伪影，如局部模糊和过度平滑。补充 AWS，GDC 约束在梯度空间中强制相邻像素之间的连贯性，有效抑制过度平滑，同时保留高频特征和锐利结构边缘。如表 3 中定量验证，我们的完整框架在新颖视图合成质量上优于 3DGS+MSAA，同时也超过了独立的 3DGS+AWS+GDC 配置。这些结果证实我们的约束优化模式不仅解决了 MSAA 在处理复杂纹理时的局限性，还有比孤立约束策略更均衡和健壮的优化格局。

2 结论

总结来说，我们的方法利用像素级加权约束、MSAA 和梯度差约束来应对在 3D 高斯喷溅中重建细局部细节和减少混叠的挑战。通过根据像素梯度指定自适应权重，我

Table 3: 对 Mip-NeRF360 和 Tank & Temples 数据集进行消融实验。SSIM $\uparrow$ 和 PSNR $\uparrow$ 的值越高越好；LPIPS $\downarrow$ 的值越低越好。最佳得分 和 第二高分 分别是红色和橙色。

Methods	Mip-NeRF360			Tank & Temple			Deep Blending		
	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
3DGS	0.815	27.21	0.214	0.841	23.14	0.183	0.903	29.41	0.243
3DGS+MSAA	0.812	27.50	0.221	0.844	23.69	0.178	0.899	29.48	0.246
3DGS+AWS+GDC	0.816	27.45	0.210	0.846	23.66	0.170	0.901	29.44	0.244
Ours	0.819	27.62	0.207	0.851	23.79	0.165	0.900	29.52	0.246

们的方法有效地聚焦于重建难度较高的区域，同时 MSAA 缓解了锯齿状边界。此外，结合梯度差约束能够保留高频信号和锐利边缘。因此，我们的方法相比于简单的喷溅表现出了显著的改进，并在局部细节重建中达到了最新的性能。

3

声明

- 利益冲突/竞争性利益（请查看期刊的具体指南以确定使用哪个标题）：作者声明无与本文内容相关的竞争性利益。
- 数据可用性：本文中使用的数据已在实验部分声明和引用 ??。

References

- [1] Shum, H., Kang, S.B.: Review of image-based rendering techniques. Visual Communications and Image Processing 2000 **4067**, 2–13 (2000)
- [2] Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 519–528 (2006). <https://doi.org/10.1109/CVPR.2006.19>
- [3] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
- [4] Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5855–5864 (2021)
- [5] Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5470–5479 (2022). <https://doi.org/10.1109/CVPR52688.2022.00539>
- [6] Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives

with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* **41**(4), 1–15 (2022)

- [7] Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance fields without neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5501–5510 (2022). <https://doi.org/10.1109/CVPR52688.2022.00542>
- [8] Garbin, S.J., Kowalski, M., Johnson, M., Shotton, J., Valentin, J.: Fastnerf: High-fidelity neural rendering at 200fps. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14346–14355 (2021)
- [9] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
- [10] Jimenez, J., Gutierrez, D., Yang, J., Reshetov, A., Demoreuille, P., Berghoff, T., Perthuis, C., Yu, H., McGuire, M., Lottes, T., *et al.*: Filtering approaches for real-time anti-aliasing. *SIGGRAPH Courses* **2**(3), 4 (2011)
- [11] Akeley, K.: Reality engine graphics. In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 109–116 (1993). <https://doi.org/10.1145/166117.166131>
- [12] Xu, L., Ren, J., Yan, Q., Liao, R., Jia, J.: Deep edge-aware filters. In: *International Conference on Machine Learning*, pp. 1669–1678 (2015)
- [13] Hua, M., Bie, X., Zhang, M., Wang, W.: Edge-aware gradient domain optimization framework for image filtering by local propagation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2838–2845 (2014). <https://doi.org/10.1109/CVPR.2014.363>
- [14] Mai, A., Hedman, P., Kopanas, G., Verbin, D., Futschik, D., Xu, Q., Kuester, F., Barron, J.T., Zhang, Y.: Ever: Exact volumetric ellipsoid rendering for real-time view synthesis. *arXiv preprint arXiv:2410.01804* (2024)
- [15] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, pp. 694–711 (2016)
- [16] Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3262–3271 (2018). <https://doi.org/10.1109/CVPR.2018.00344>
- [17] Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 136–144 (2017).

<https://doi.org/10.1109/CVPRW.2017.151>

- [18] Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. In: *IEEE Transactions on Computational Imaging*, vol. 3, pp. 47–57 (2017). <https://doi.org/10.1109/TCI.2016.2644865>
- [19] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
- [20] Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456 (2024). <https://doi.org/10.1109/CVPR52733.2024.01839>
- [21] Yan, Z., Low, W.F., Chen, Y., Lee, G.H.: Multi-scale 3d gaussian splatting for anti-aliased rendering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20923–20931 (2024). <https://doi.org/10.1109/CVPR52733.2024.01977>
- [22] Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20654–20664 (2024). <https://doi.org/10.1109/CVPR52733.2024.01952>
- [23] Deng, K., Liu, A., Zhu, J.-Y.: Depth-supervised nerf: Fewer views and faster training for free. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12735–12744 (2022). <https://doi.org/10.1109/CVPR52688.2022.01254>
- [24] Zhang, K., Barron, J.T., Tancik, M., Hedman, P., Srinivasan, P.P.: Nerf-gan: Learning implicit generative representations of 3d scenes. *arXiv preprint arXiv:2010.08422* (2020)
- [25] Molnar, S., Cox, M., Ellsworth, D., Fuchs, H.: A sorting classification of parallel rendering. *IEEE computer graphics and applications* **14**(4), 23–32 (1994)
- [26] Akenine-Moller, T., Haines, E., Hoffman, N.: *Real-time Rendering*, (2019)
- [27] Zhang, J., Zhan, F., Xu, M., Lu, S., Xing, E.: Fregs: 3d gaussian splatting with progressive frequency regularization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21424–21433 (2024). <https://doi.org/10.1109/CVPR52733.2024.02024>
- [28] Knapitsch, A., Park, J., Zhou, Q.-Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)* **36**(4), 1–13 (2017)

- [29] Hedman, P., Philip, J., Price, T., Frahm, J.-M., Drettakis, G., Brostow, G.: Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)* **37**(6), 1–15 (2018)
- [30] Ye, Z., Li, W., Liu, S., Qiao, P., Dou, Y.: Absgs: Recovering fine details in 3d gaussian splatting. In: *Proceedings of the 32nd ACM International Conference on Multimedia (ACM MM)*, pp. 1053–1061 (2024)
- [31] Zhang, Z., Hu, W., Lao, Y., He, T., Zhao, H.: Pixel-gs: Density control with pixel-aware gradient for 3d gaussian splatting. In: *European Conference on Computer Vision (ECCV)*, pp. 326–342 (2024). Springer