Geo-NVS-w：基于有符号距离函数渲染器的真实场景几何感知新视角合成

Anastasios Tsalakopoulos[1*] AngelosKanlis[1*] EvangelosChatzis[1]
Antonis Karakottas[1] DimitriosZarpalas[1]1希腊研究与技术中心(CERTH),
塞萨洛尼基，希腊{tsalakop, a.kanlis, chatzise, ankarako, zarpalas}@iti.gr

*共同贡献

arXiv:2601.08371v1 [cs.CV] 13 Jan 2026

摘要

我们提出 *Geo-NVS-w*，这是一种几何感知框架，可从无结构的真实场景图像集合中实现高保真新视角合成。尽管现有的真实场景方法在新视角合成方面已表现出色，但它们往往缺乏对复杂表面的几何约束，有时会产生不一致的结果。

Geo-NVS-w 利用基于有符号距离函数（*SDF*）的底层几何表示来引导渲染过程，从而解决这一局限。此外，我们还提出了几何保持损失，以确保保留精细的结构细节。该框架实现了具有竞争力的渲染性能，同时与类似方法相比，能耗降低了 *4–5*× 倍。我们证明，*Geo-NVS-w* 是一种稳健的真实场景新视角合成方法，能够生成照片级真实感结果，并呈现清晰且几何一致的细节。

1. 引言

从不受约束的野外图像集合中进行新视角合成，已成为计算机图形学与视觉领域的核心问题。通过从任意视点渲染出照片级真实感视图来探索三维场景的能力，在虚拟现实、数字遗产和视觉特效领域具有深远的应用价值。然而，野外数据集（例如广泛使用的IMC-Phototourism数据集）[11]由地标的游客照片构成，带来了巨大的挑战。此类数据集的特点是光照不一致、相机设置各异，以及行人、车辆或临时结构等瞬态遮挡物。

以神经辐射场 [7] 及其针对野外数据的变体 [1, 6] 为代表的开创性方法已取得显著进展。然而，它们继承了体密度表示的一项根本局限，即存在几何歧义，容易产生半透明伪影，并使清晰的建筑细节变得模糊。

我们认为，要提升新视图合成的照片级真实感与一致性，渲染过程可以受益于场景底层几何体的显式引导。为此，我们提出Geo-NVS-w框架，该框架以高保真有符号距离函数（SDF）表示 [13]作为其几何骨干。精确的SDF为表面位置提供了强有力的先验，使我们的渲染器能够生成清晰的深度不连续性，进而产生更清晰的图像。

我们的核心贡献如下：

- **八叉树加速的基于特征的体**：我们将 SDF 引导的渲染器与基于八叉树的特征体相结合，在含几何区域内插值特征，以保留细节并加速训练。
- **几何保持损失（GPL）**。我们提出了一种新颖的损失函数，明确惩罚模型将几何上重要的区域错误地屏蔽为瞬态区域，从而确保纹理和特征在不同视角间保持一致。
- **量化能效分析：**我们在训练过程中记录图形处理器功耗，以量化能耗与质量之间的权衡，并记录显著缩短的训练时间和降低的能耗。

我们的方法基于这样一个原则：准确且稳健的几何表示对于高质量视图合成至关重要。通过以有符号距离函数为基础进行渲染，我们为表面提供了强有力的归纳偏置，这直接提升了合成视图的质量，尤其是在复杂场景中。因此，Geo-NVS-w 展示了先进渲染技术与坚实几何基础相结合，能够实现高质量的新视图合成。

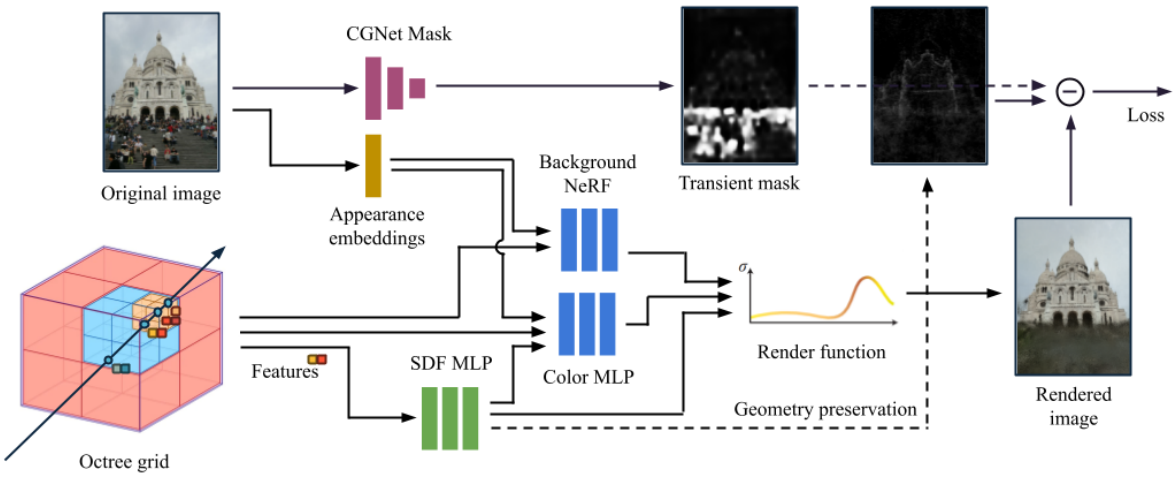

图1. **Geo-NVS-w框架概览**。对于给定的相机光线，我们沿着包含前景（基于SDF）和背景（基于NeRF）特征网格的八叉树进行光线步进。在前景单位球内，插值后的特征会输入多层感知机，以预测SDF值 $s(\mathbf{x})$ 和颜色 $c(\mathbf{x}, \mathbf{d})$。我们使用 NeuS 渲染公式将 SDF 值转换为 Alpha合成权重 w_i，确保渲染颜色与底层表面紧密关联。利用这些权重累积颜色，再将其与背景合成。几何保持损失（GPL）确保瞬态掩码不会侵蚀高曲率细节。

2. 相关工作

我们的工作建立在新视角合成和神经表面表示领域的最新进展之上。

2.1. 新视角合成

神经辐射场（NeRF） [7] 将场景建模为从 5D 坐标（位置和视角方向）到体密度和颜色的连续映射，能够实现照片级真实感渲染，但需要密集采样和较长的训练时间，且难以处理大规模场景。为加快渲染速度，PlenOctrees [17]将网络烘焙到显式八叉树中，实现实时渲染。为处理无界场景，NeRF++ [19]使用双多层感知机分别建模前景和背景。针对野外数据，NeRF-W [6] 为每张图像添加外观嵌入，并引入瞬态组件以应对光照变化和动态场景；而 Ha-NeRF [1] 则将外观幻觉卷积神经网络与瞬态二维掩码多层感知机相结合。

2.2. 神经表面表示

另一条研究路线通过隐式表示追求高质量几何体。NeuS [13] 引入了有符号距离函数体渲染，确保光线的透明度加权颜色与底层表面相对应。为了实现规模化，加速结构至关重要：Instant-NGP [8] 为 NeRF 采用多分辨率哈希网格，而 Neuralangelo [12] 则将其应用于有符号距离函数，以实现高保真、大规模重建，凸显了数值梯度和由粗到细的优化对于稳定哈希网格上的 SDF训练的重要性。

Geo-NVS-w 融合了这些研究方向：我们采用 NeuS 风格的有符号距离函数渲染，以实现几何保真度和网格效率；这与 Neuralangelo 类似，但使用的是八叉树特征网格。与专注于重建的工作不同，我们将这套流程用作高质量新视角合成的基础。

3. Geo-NVS-w框架

Geo-NVS-w主要面向高保真新视角合成。其架构建立在稳健的几何基础之上——有符号距离函数（SDF），这是渲染一致视图的关键。该框架结合了高效的八叉树特征体、SDF引导的渲染过程，以及我们提出的几何保持损失（GPL）。这一几何基础通过强制保持连贯的表面几何，避免了 NeRF 渲染中常见的伪影，例如由不一致的光线采样导致的密度云化。

3.1. 八叉树特征体

为兼顾性能与效率，我们使用两个彼此独立的稀疏八叉树特征网格来表示场景：一个用于前景，以有符号距离函数建模；另一个用于背景环境，以传统 NeRF 建模。这种双网格八叉树结构将计算资源集中在包含几何体的区域，并剔除大片空旷空间。对于前景单位球内部的任意三维点 $\mathbf{x}$，我们通过对有符号距离函数网格进行三线性插值来查询特征。随后，由小型多层感知机解码这些特征，生成有符号距离函数值和颜色值。对于该球体外部的点，则从背景网格查询特征，并将其输入一个小型背景神经辐射场网络。

3.2. 有符号距离函数引导的体渲染

准确的表面定位对于实现锐利渲染至关重要。有符号距离函数通过零水平集 $s(\mathbf{x}) = 0$ 定义表面。

我们使用 NeuS [13]将有符号距离函数与 alpha 合成相联系。对于一个光线采样点 $\mathbf{x}_i$，其有符号距离函数为 $s_i = s(\mathbf{x}_i)$，其占用率为

$$\alpha_i = \text{sigmoid}(\zeta s_i + \beta_i) - \text{sigmoid}(\zeta s_{i+1} + \beta_i), \quad (1)$$

其中，ζ 是学习得到的全局偏差，$\beta_i = \langle \nabla s_i, \mathbf{d} \rangle \Delta t_i/2$ 用于校正光线$\mathbf{d}$ 与法线 ∇s_i之间的夹角。紧凑的偏差网络用于学习 ζ，法线/导数则通过有限差分 [12]计算。渲染权重遵循离散合成规则：

$$w_i = T_i \alpha_i, \quad T_i = \prod_{j<i} (1 - \alpha_j), \quad (2)$$

最终颜色 $C(\mathbf{r})$ 将前景累积到背景NeRF上。这样一来，颜色积分便集中在表面周围的狭窄带状区域内，从而生成几何一致且清晰的视图。

3.3. 外观与瞬态模块

每张图像都使用一个潜在外观编码来建模光照和相机参数的变化，其灵感来自 NeRF-W [6]。对于瞬态遮挡物，我们采用与 CR-NeRF [15] 相同的无监督分割方法，并训练轻量级 CGNet [14] 生成瞬态物体掩码。

3.4. 几何保持损失（GPL）

真实场景新视角合成面临的一项常见挑战，是将静态背景与瞬态物体（例如行人、汽车）分离开来。通常使用瞬态掩码网络来处理这一问题，该网络会学习降低动态元素对应像素的权重。然而，使用光度损失训练多个网络，往往会导致瞬态掩码网络移除静态结构边缘，使有符号距离函数表示的部分区域退化，尤其是几何体较为复杂的区域。

为解决这一问题，我们引入**几何保持损失（GPL）**。其直观思路是：对于穿过几何上重要区域的光线，惩罚瞬态掩码处于激活状态。我们将损失定义为：

$$\mathcal{L}_{\text{GPL}} = \mathbb{E}_{\mathbf{r}} \left[M(\mathbf{r}) \cdot \Phi(\mathbf{r}) \right], \quad (3)$$

其中，$M(\mathbf{r}) \in [0, 1]$ 表示光线$\mathbf{r}$的预测掩码值，而 $\Phi(\mathbf{r})$ 是根据前景有符号距离函数得到的边缘指示函数。对于与几何复杂区域相交的光线，该指示函数被设计为具有较高值。对于每条光线，我们计算每个采样点的 Eikonal 误差的逐光线平均值 $E_{\text{grad}}(r) = (\|\nabla s\| - 1)^2$, 以及绝对曲率估计值 $E_{\text{curv}}(r) = |\Delta s|$。随后，指示函数定义为：

$$\Phi(\mathbf{r}) = \sigma(\lambda_{\text{grad}} E_{\text{grad}}(\mathbf{r}) + \lambda_{\text{curv}} E_{\text{curv}}(\mathbf{r})), \quad (4)$$

其中，σ 是经过校准的 Sigmoid 函数。直观而言，$\Phi(\mathbf{r})$在穿过锐利边缘或高曲率区域的射线上达到峰值。将掩码值与该指示量相乘后，我们的



(a) (b) (c) (d)

图2。**Phototourism 场景上的定性结果**。从上到下：圣心大教堂、特雷维喷泉、勃兰登堡门。（a）输入图像。（b）使用来自相同视点和外观嵌入的 Geo-NVS-w 渲染结果。（c）估计的瞬态掩码。（d）几何保持图的可视化结果（沿每条光线累积的 SDF梯度）。

几何保持损失会促使掩码在对定义场景结构至关重要的射线上保持为零（即完全属于前景）。这样设计旨在保留锐利特征，从而生成视觉上更清晰、更连贯的新视图。我们在所有实验中均将损失权重设为 $\lambda_{\text{GPL}} = 0.05$、$\lambda_{\text{grad}} = 10.0$ 和 $\lambda_{\text{curv}} = 2.0$。

3.5. 学习目标

通过最小化复合损失，对完整框架进行端到端训练：

$$\begin{aligned} \mathcal{L} = {} & \lambda_{\text{rgb}} \mathcal{L}_{\text{rgb}} + \lambda_{\text{eik}} \mathcal{L}_{\text{eik}} + \lambda_{\text{curv}} \mathcal{L}_{\text{curv}} \\ & + \lambda_{\text{mask}} L_{\text{reg}} + \lambda_{\text{GPL}} \mathcal{L}_{\text{GPL}} + \lambda_{\text{lps}} \mathcal{L}_{\text{lps}}. \end{aligned} \quad (5)$$

关键组成部分包括：主要光度损失$\mathcal{L}_{\text{rgb}}$（L1差值）；Eikonal损失 $\mathcal{L}_{\text{eik}} = (\|\nabla s\| - 1)^2$；我们提出的新型 $\mathcal{L}_{\text{GPL}}$；沿用先前工作的颜色网络 Lipschitz损失 $\mathcal{L}_{\text{lps}}$ [4] [10]；曲率损失 $\mathcal{L}_{\text{curv}} = |\Delta s|$；以及离面惩罚项。正则化的瞬态掩码$L_{\text{reg}}$结合瞬态CNN掩码，以及针对强度和双峰性的惩罚项，以促进静态和动态元素的清晰分离。在预热阶段，我们还使用球体初始化损失 $\mathcal{L}_{\text{sphere}}$，以提供稳定的初始几何体。

4. 实验

我们在 IMC-PT数据集当前可用的全部四个场景上评估 Geo-NVS-w[11] ；这些场景也用于 NeRF-W 和 Ha-NeRF，其余场景因数据不一致而被移除 [16]。我们还认为 CR-NeRF [15] 可作为互补方法，并计划在未来工作中进行受控对比。

表 1. **新视图合成（NVS）质量**。我们报告峰值信噪比（↑）、结构相似性指数（↑）和学习感知图像块相似度（↓）。Geo-NVS-w 在 IMC-PT数据集的所有场景中均取得了出色的结果。

方法	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
勃兰登堡门			
Ha-NeRF	23.45	0.811	0.247
NeRF-W	23.98	0.915	0.198
Ours	**25.40**	**0.944**	**0.158**
圣心大教堂			
Ha-NeRF	**25.40**	**0.877**	**0.124**
NeRF-W	25.11	0.859	0.141
Ours	23.23	0.850	0.160
特雷维喷泉			
Ha-NeRF	22.15	0.695	**0.117**
NeRF-W	23.01	0.751	0.109
Ours	**24.50**	**0.831**	0.203
泰姬陵			
Ha-NeRF	22.72	0.767	0.301
NeRF-W	**25.15**	0.833	0.195
Ours	24.19	**0.860**	**0.194**
平均值			
Ha-NeRF	23.43	0.788	0.197
NeRF-W	24.31	0.840	**0.161**
Ours	**24.33**	**0.871**	0.179

4.1. 实现细节

我们的模型使用 PyTorch 实现 [9]，并在单个 NVIDIA A10G GPU 上采用混合精度计算进行训练，以加快训练速度。我们将 GPU 功耗测量直接纳入训练流程，通过记录一段时间内的累积用量来测量能耗。

4.2. 新视角合成基准

我们的方法在多个指标上达到或超过了基线方法 [1, 6]，如表1所示；同时，在大多数场景中实现了更优的图像质量指标。图 3 进一步印证了这些结果，表明我们的方法在更高处理速度和更低能耗的同时，能够提供更高的视觉保真度。

4.3. 能耗分析

除了原始性能之外，实际可用性和可扩展性还取决于计算效率。如图3所示，Geo-NVS-w 不仅速度更快，而且能耗更低。我们的方法完成 300,000 次迭代运行仅需约 2.05kWh，而 NeRF-W 消耗 9.7 kWh，Ha-NeRF 消耗 7.71kWh。除了实现更高的几何保真度之外，我们的基于 SDF 的方法还展现出效率优势：在我们的运行中，仅使用约 2.05kWh即可达到峰值信噪比。这得益于其架构能够大幅缩减多层感知机的规模。

5. 结论

我们提出了 Geo-NVS-w，这是一种以几何保持为渲染流程核心、推动野外新视角合成发展的框架。通过将有符号距离函数融入架构，并引入几何保持损失，我们的方法能够生成清晰度稳定且几何一致的渲染结果。与以往方法相比，该框架提供了一种具有优势的几何一致替代方案，在效率与质量之间实现了良好权衡。尽管高斯泼溅等新范式已经兴起，我们的研究结果仍证实，几何感知方法在高保真野外视图合成中依然有效。

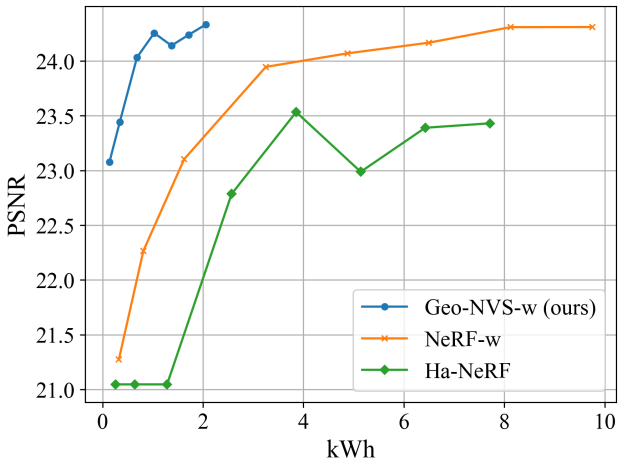

图 3。**能耗与质量之间的权衡**。Geo-NVS-w 以显著更短的训练时间和更低的累计能耗（kWh）实现了较高的峰值信噪比，相较于基线 NeRF-W，是一种更高效、更具可扩展性的解决方案。

未来的工作可以探索将三维高斯泼溅技术 [2]集成到野外场景中 [3, 18],，利用其速度优势，在保持我们的 SDF 方法所实现的表面一致性的同时，进一步降低渲染延迟和计算开销，如 [5]所示。

6. 致谢

本研究得到了欧盟委员会资助的 XReco 项目支持，该项目隶属于地平线欧洲计划，资助协议编号为 101070250。计算资源由 GRNET 提供支持。

参考文献

[1] Xiaoshuai Chen、Zizhang Cai、Jia Li 和 Guofeng Zhang。HA-NeRF：野外的幻觉神经辐射场。在 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第 12652–12661 页，2022 年 1 、 2 、 4– 。[2] Bernhard Kerbl、Georgios Kopanas、Thomas Leimk¨uhler 和 George Drettakis。用于实时辐射场渲染的三维高斯泼溅。*ACM*图形学汇刊，42(4):139:1–139:1，2023年。4[3] Jonas Kulhanek、Songyou Peng、Zuzana Kukelova、Marc Pollefeys 和 Torsten Sattler。Wildgaussians：野外的三维高斯泼溅。*NeurIPS*，2024年。4

[4] Hsueh-Ti Derek Liu、Francis Williams、Alec Jacobson、Sanja Fidler 和 Or Litany。通过 Lipschitz 正则化学习光滑神经函数。载于 ACM SIGGRAPH 2022 会议论文集，第 31:1–31:13 页。ACM，2022年。3[5] Jianheng Liu 等。GS-SDF：激光雷达增强的高斯泼溅与神经符号距离函数，实现几何一致的渲染和重建。arXiv 预印本 arXiv:2503 10170，2025年 4 。.[6] Ricardo Martin-Brualla、Noha Radwan、Mehdi S. M. Sajjadi、Jonathan T. Barron、Alexey Dosovitskiy 和 Daniel Duck- worth。NeRF-W：面向无约束照片集的神经辐射场。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 7210–7219 页，2021年。1、2、3、4[7]Ben Mildenhall、Pratul P. Srinivasan、Matthew Tancik、Jonathan T. Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF：将场景表示为用于视图合成的神经辐射场。载于欧洲计算机视觉会议（ECCV）论文集，第 405–421 页，2020年。1、2[8] Thomas M¨uller、Alex Evans、Christoph Schied 和 Alexan- der Keller。采用多分辨率哈希编码的即时神经图形基元。ACM图形学汇刊（TOG），41(4):1–15，2022年。2[9] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Rai- son、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。PyTorch：一种命令式风格、高性能的深度学习库。载于《神经信息处理系统进展》第 32 卷，第 8024–8035 页。Curran Associates Inc.，2019年。4[10] Radu Alexandru Rosu 和 Sven Behnke。PermutoSDF：利用置换格和隐式曲面进行快速多视图重建。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 8466–8475 页，2022年。3[11] Noah Snavely、Steven M. Seitz 和 Richard Szeliski。照片旅游：探索三维照片集。ACM图形学汇刊，25(3):835–846，2006年。1、3[12]Matthew Tancik、Jonathan T. Barron、Ben Mildenhall、Thomas M¨uller、Alex Evans、Ricardo Martin-Brualla、Pratul P. Srinivasan、Mehdi S. M. Sajjadi、Daniel Duckworth 和 Tero Karras。Neuralangelo：高保真神经曲面重建。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 8548–8558 页，2023年。2、3[13] Peng Wang、Lingjie Liu、Yuan Liu、Christian Theobalt、Taku Komura 和 Wenping Wang。NeuS：通过体渲染学习用于多视图重建的神经隐式曲面。载于《神经信息处理系统进展》（NeurIPS），第 26998–27010 页，2021 年。1 、 2 、 3[14] Tianyi Wu 、Sheng Tang、Rui Zhang 和 Yongdong Zhang。CGNet：用于语义分割的轻量级上下文引导网络。IEEE图像处理汇刊，30：1169–1179，2021年。3[15] Yifan Yang、Shuhai Zhang、Zixiong Huang、Yubing Zhang 和 Mingkui Tan。用于新颖视图的交叉射线神经辐射场

从无约束图像集合进行视图合成，2023年。3[16] Kwang Moo Yi。IMC-PT 2020 数据集，2020年。不列颠哥伦比亚大学。访问日期：2025年9月26日。3[17] Alex Yu、Ruilong Li、Matthew Tancik、Hao Li、Ren Ng 和 Angjoo Kanazawa。PlenOctrees：神经辐射场的实时渲染。载于 *IEEE/CVF*计算机视觉国际会议（*ICCV*）论文集，第 5752–5761 页，2021年。2[18] Dongbin Zhang、Chuming Wang、Weitao Wang、Peihao Li、Minghan Qin 和 Haoqian Wang。野外高斯：面向无约束图像集合的 3D 高斯泼溅。欧洲计算机视觉会议（*ECCV*），2024年。4[19] Kai Zhang、Gernot Riegler、Noah Snavely 和 Vladlen Koltun。NeRF++：神经辐射场的分析与改进。*arXiv* 预印本 *arXiv:2010.07492*，2020年。2