

基于动态采样范围的神经隐式表面重建优化

钟娟, 谭诗瀚, 梁书凝, 陈杨仁, 王录涛

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘要: 神经隐式表面重建技术为三维重建带来了新的契机, 然而现有方法存在采样效率低、训练时间长等局限。为此, 提出一种基于动态采样范围的优化方案以提升重建的质量与效率。该方案聚焦采样策略, 通过占据栅格引导动态生成不断缩小的核心采样区域, 结合分层体积采样策略和采样点正则化约束, 引导采样点更接近物体表面。实验结果显示, 该方案在DTU数据集15个场景下的评价指标均优于基线模型, 峰值信噪比为29.79, 倒角距离为0.68, 与传统方法NeuS相比分别提升了6.39%和13.24%; 训练时间为20 min, 相较NeuS方法加速了24倍。基于动态采样范围的神经隐式表面重建优化方案可有效提升重建质量与效率, 为三维重建技术的发展提供了新思路。

关键词: 三维重建; 神经隐式表面重建; 符号距离函数; 占据栅格; 分层体积采样

DOI: 10.11907/rjdk.241615

扫描二维码阅读全文:



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)010-0088-09

Optimization of Neural Implicit Surface Reconstruction Based on Dynamic Sampling Range

ZHONG Juan, TAN Shihan, LIANG Shuning, CHEN Yangren, WANG Lutaotao

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Neural implicit surface reconstruction technology has brought new opportunities to the field of 3D reconstruction, but existing methods have limitations such as low sampling efficiency and long training time. Therefore, an optimization scheme based on dynamic sampling range is proposed to improve the quality and efficiency of reconstruction. This scheme focuses on the sampling strategy, which dynamically generates a shrinking core sampling area by occupying the grid, and combines the layered volume sampling strategy and sampling point regularization constraints to guide the sampling points closer to the surface of the object. The experimental results show that the evaluation metrics of this scheme are superior to the baseline model in all 15 scenarios of the DTU dataset, with a peak signal-to-noise ratio of 29.79 and a chamfer distance of 0.68. Compared with the traditional method NeuS, it has improved by 6.39% and 13.24%, respectively; The training time is 20 minutes, which is 24 times faster than the NeuS method. The neural implicit surface reconstruction optimization scheme based on dynamic sampling range can effectively improve the reconstruction quality and efficiency, providing new ideas for the development of 3D reconstruction technology.

Key Words: 3D reconstruction; neural implicit surface reconstruction; signed distance function; occupancy grid; hierarchical volume sampling

0 引言

随着计算机图形学与深度学习技术的深度融合, 从二维图像中恢复三维场景的重建技术取得了显著进展, 为虚拟现实、增强现实、数字孪生等前沿领域的发展带来了新

的契机^[1]。神经隐式表面重建技术突破了传统算法在几何上非连续表征的局限, 首先采用多层感知机 (Multi-Layer Perceptron, MLP) 或结合哈希编码对几何属性进行隐式表示, 如体密度、占用率和符号距离函数 (Signed Distance Function, SDF) 等^[2-9]; 其次, 借鉴 NeRF 中的逆渲染管线对场景进行可微体积渲染^[6,10]; 最后通过经典的行进立方体

收稿日期: 2024-07-18

基金项目: 四川省科技计划项目 (2023YFG0304); 四川省重大科技专项 (2022ZDZX0008)

作者简介: 钟娟 (1989-), 女, 成都信息工程大学计算机学院硕士研究生, 研究方向为三维重建; 谭诗瀚 (1986-), 男, 博士, 成都信息工程大学计算机学院讲师, 研究方向为计算机图形学、扩展现实; 梁书凝 (1998-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为三维重建; 陈杨仁 (2004-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为三维重建; 王录涛 (1979-), 男, 博士, 成都信息工程大学计算机学院副教授、硕士生导师, 研究方向为机器视觉、嵌入式人工智能。本文通讯作者: 王录涛。

(Marching Cubes, MC)算法按需导出指定分辨率的显式几何网格^[11]。在这一过程中,沿着每条光线的路径在多个位置上采样后综合这些采样点的特征来确定光线对应的像素点颜色。图像像素点的颜色理论上直接等于光线与物体表面交点处的颜色值,然而体积渲染却依赖于空间中多个采样点颜色的累加,这种差异导致了不可避免的渲染偏差^[12]。同时,真实场景中大量无前景物体的空区域,这些区域中的无效采样不仅影响重建的准确性,而且会降低训练效率。因此,提高采样点的有效性,尤其是在减少无效空间内采样的同时确保采样点接近物体表面,已成为当前基于神经隐式表达的三维重建领域中亟待解决的问题。

1 相关工作

1.1 神经隐式表面重建

神经隐式表面重建技术是计算机图形学领域的前沿研究方向,其核心在于利用神经网络对三维空间中的物体表面进行隐式建模,摒弃了传统的显式表示方法。例如,文献[6]提出的 NeRF 采用 MLP 将三维空间位置映射为体积密度和颜色,并通过神经体积渲染生成图像,缓解了表面渲染带来的重建缺陷,但 NeRF 因缺少表面约束而难以精确建模几何结构。为解决这一问题,文献[8]提出的 NeuS 和文献[9]提出的 VolSDF 将 SDF 的表征引入神经体渲染管线中,并基于此分别推导出 NeRF 中体密度到 SDF 的两种不同转换方式^[6]。文献[13]提出的隐式可微渲染(Implicit Differentiable Renderer, IDR)、文献[14]提出的可微体积渲染(Differentiable Volumetric Rendering, DVR)通过 SDF 和占用函数来表示物体,利用表面渲染技术恢复物体的几何信息和颜色,但表面渲染的方法在训练过程中容易陷入局部最优解。文献[15]提出的 UniSurf 则是将占用函数的表征融入到体积渲染中。为进一步提升重建质量,后续研究致力于减少几何与颜色之间的模糊性与耦合性。例如,文献[16]提出的 MonoSDF 通过单目深度预测获取深度和法线信息,并在训练中引入深度一致性损失和法线一致性损失来约束几何。文献[12]提出的 Geo-NeuS 利用 ColMap^[17]生成的稀疏点约束点云 SDF 值为 0,并引入多视图光度一致性损失提升几何一致性。尽管以上方法提高了重建质量,但多依赖于额外的输入数据,如点云和深度图。这不仅增加了数据预处理的复杂性,而且对神经网络的计算能力提出了更高要求,导致模型训练耗时增加。

1.2 基于体渲染的采样策略

体积渲染技术在空间内进行采样,并利用这些采样点合成对应图像像素点的颜色,因此采样策略的选择对于渲染的效率和质量至关重要。NeRF 引入的分层体积采样(Hierarchical Volume Sampling, HVS)技术被认为是提升采样效率的关键,此后的研究从多个角度对采样方法进行了

优化^[6,18]。例如,文献[19]提出的 MipNeRF 沿着光线采样锥形区域而非单个样点,以降低混叠效应。文献[20]提出的 DDNeRF 均分区间后拟合每个区间内的高斯分布,以更准确的概率分布指导采样。文献[21]提出的 L0-Sampler 使用分段指数函数对初步采样点的权重进行插值,相较原始的 HVS 常数段函数指导更符合真实的几何特性。文献[22]将运动恢复结构(Structure from Motion, SFM)技术得到的稀疏点云应用于采样过程,使用体积采样与表面采样相结合的方式在物体表面附近采样。文献[23-27]利用深度先验来引导采样。文献[28-30]直接利用神经网络辅助预测采样点的位置。以上基于 HVS 的非均匀采样策略可以在一定程度上跳过无物体区域,但由于初步采样时没有信息源指导其跳过空区域,导致初步采样点的有效率不高,进而影响了重要性采样点的有效性。文献[31]提出的 Instant-NGP 和文献[32]提出的 NeuS2 通过引入占据栅格来有效划分空间并标记每个子空间的占用情况,进而可以在光线步进过程中跳过无物体占据的区域,显著提升光线和采样点的利用率。然而,基于占据栅格策略的采样点过多且分布均匀,不仅会增加网络训练的计算成本,而且会影响几何重建的精细程度。例如,文献[33]提出的 Neuralangelo 作为 Instant-NGP 的衍生技术,为了追求更高的重建质量而增加了网络复杂度和哈希特征数量,导致其难以处理占据栅格采样策略产生的大量采样点的计算需求。此外,均匀分布的采样点缺乏对物体表面的趋向性,导致在变化剧烈的区域采样不足,而在变化平缓的区域过度采样,从而无法精确捕捉场景的细节。

本研究在现有工作基础上创新性地提出一种基于动态采样范围的神经隐式表面重建方案。该方案聚焦于采样策略,其核心在于利用占据栅格动态生成采样区域,仅在该范围内分层采样生成更接近物体表面的少量采样点,通过采样点约束项优化整个采样过程。与现有工作相比,本研究的主要贡献为:①基于动态采样范围的优化方案提高了光线与采样点的有效性,提升了重建质量,契合下游市场的应用;②采样策略引导在较小的核心区域内进行分层采样,使得精细采样点之间更加紧凑,从而有效减少了渲染偏差中的采样误差;③创新性地引入采样点正则化项,进一步缩小了有效采样区域并减少了渲染偏差中的权重误差。不仅在理论上具有创新性,而且在实际应用中也具有显著价值,为神经隐式表面重建领域的发展提供了新的视角和实施方案。

2 方案构建

2.1 整体架构

基于动态采样范围的神经隐式表面重建方案整体架构如图 1 所示。给定 N 张不同视角的图像及对应相机参数,该方案可以高效且精细地重建出三维物体表面。首

先,在采样策略部分为输入的二维图像构建栅格并根据空间占用信息划定核心采样区域,输出非均匀分布的少量采样点;其次,通过网络将得到的采样点经训练后输出对应

的颜色值和 SDF 值;最后,在损失约束中对几何和颜色两方面的特征进行监督。本文提出的采样约束项将损失反向传播到采样策略部分以优化采样过程。

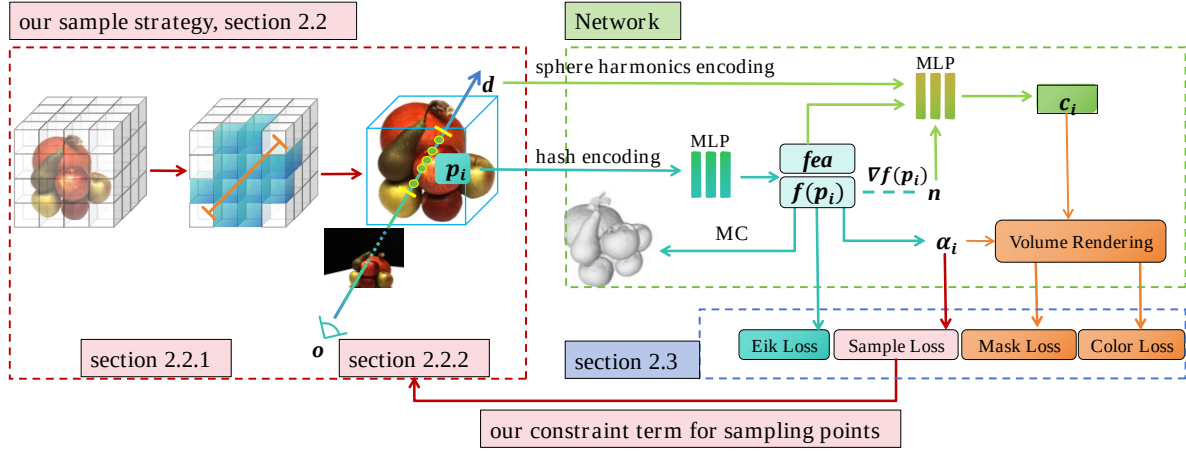


Fig. 1 Overall architecture

图1 整体架构

训练从随机选定的图像和像素开始,从该图像视角的相机原点 o 出发,沿方向 d 朝该像素上投射光线,然后在光线上进行体积采样。表示为:

$$p_i = o + t_i \cdot d \quad (1)$$

式中: p_i 为其中一个采样点的位置; t_i 为该采样点离原点 o 的距离。

基于 NeuS 的神经隐式重建技术随机生成的光线与采样范围是无约束的^[8]。而本文方案提出的采样策略(见图1左虚线框部分)利用占据栅格信息生成动态缩小的核心采样区域,随后仅在该范围分层采样生成较少且非均匀分布的采样点,在提升采样效率的同时使得精细采样点更加紧凑,为后续渲染质量奠定了良好基础。对采样点位置和方向向量分别进行哈希编码和球谐编码,将编码后的数据输入几何网络和颜色网络,预测得到采样点的 SDF 值 $f(p_i)$ 和颜色值 c_i ^[31]。在该过程中,颜色网络的输入还依赖于几何网络输出的几何特征 fea 与该采样点的法向量 n ,而法向量的计算源于 SDF 的梯度^[8]。然而,联合了多分辨率哈希编码的几何网络直接采用解析梯度(Analytic Gradient)会导致反向传播在整个空间中不连续,因此在进行具体计算时,本文采用数值梯度(Numerical Gradient)方式^[33]。随后,构建 SDF 与不透明度 α 的转换方程。表示为:

$$\alpha_i = \max \left(\frac{\Phi_s(f(p_i)) - \Phi_s(f(p_{i+1}))}{\Phi_s(f(p_i))}, 0 \right) \quad (2)$$

式中: $\Phi_s(\cdot)$ 为常见的 S 型(Sigmoid)函数。

通过经典的体渲染方案对整条光线所有采样点估计的颜色进行积分,得到该视角图像的像素颜色 C 。其数值积分的近似表达为:

$$C(o, d) = \sum_{i=1}^N w_i c_i; \quad w_i = T_i \alpha_i \quad (3)$$

式中: T_i 为累积透光率,表示到达相机的光的比例; N 为该光线的采样点数量。

采用 L1 范数的形式约束当前光线 k 的预测渲染颜色 $\hat{C}_k$ 与实际图像中该像素颜色 C_k 之间的差异。表示为:

$$L_{color} = \frac{1}{m} \sum_k \|\hat{C}_k - C_k\|_1 \quad (4)$$

网络训练完成后可采用 Marching Cubes 算法结合预测的 SDF 值导出指定分辨率的显式几何网格^[11]。值得一提的是,NeuS 采用的体渲染技术虽然可以弥补表面渲染极易陷入局部最优解的缺陷,但是由于光线传播到相机的颜色为相机与物体表面交点的颜色,体渲染的计算方式始终存在固有的渲染偏差^[8]。如图2所示,基于体渲染的采样方式在一条光线上有众多采样点,其中 t^* 表示光线与物体表面的最近交点, t_j 表示距离 t^* 最近的采样点。根据 Geo-NeuS 的分析,在体渲染过程中会将最高权重赋予 t_j ,因此 t^* 与 t_j 之间的距离差异会导致采样偏差^[12]。而剩余采样点 t_i 分走了部分本该属于表面点的权重,降低了 $w(t_j)$ 值,从而导致权重偏差。

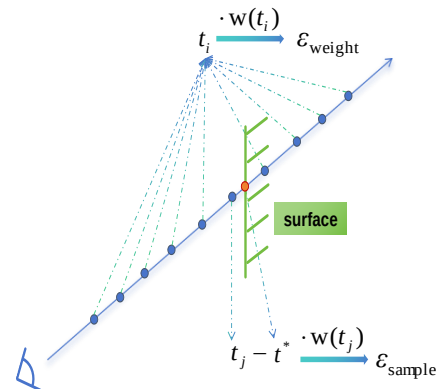


Fig. 2 Analysis of volume rendering deviation

图2 体积渲染的偏差分析

Geo-NeuS 通过引入额外的几何约束以减缓偏差, 而本文方案着眼于采样策略本身。除前文提到的网络训练前的采样策略优化外, 本文还在体渲染结束后创新性地提出针对采样点的约束项, 进一步缩小了有效采样区域并缓解了渲染过程中的权重偏差。

2.2 动态采样策略

2.2.1 占据栅格引导生成动态采样区域

在基于 NeuS 的神经隐式表面重建技术的采样策略中, 光线是在三维空间内随机发射的, 常导致大量光线并未与前景物体相交^[8]。此外, 采样点范围通常定义为光线与初始球体重合的全部区域, 这增加了在无效空间内采样的概率。为解决这一问题, 本文借鉴 Instant-NGP 的思路, 引入一种二值化的占据栅格, 该栅格可以标记子空间区域是否包含前景物体^[31]。该方式可引导光线在步进过程中仅在占据物体的核心区域内放置采样点, 从而显著提升采样有效性。如图 3 所示, 本文构建了一个分辨率为 Gr 的三维均匀栅格, 栅格中的每个体素 v_i 缓存其自身的二值化占用信息 b_i 以及不透明度值 α_i , 并在训练过程中动态更新。

对于单个体素 v_i , 每次随机选取该体素空间中的一点 p_i , 通过网络预测其 SDF 值并利用式 (2) 推断当前体素的不透明度 $\hat{\alpha}_i$, 然后根据缓存的不透明度值进行更新。

$$\alpha_i = \max(\hat{\alpha}_i \cdot \mu, \alpha_i) \quad (5)$$

式中: μ 为衰减系数。通过以上计算更新, 可以避免预测值大幅波动带来的训练不稳定性。

利用更新后的不透明度 α_i 判断当前体素的占用情况:

$$b_i = \begin{cases} \text{True}, & \alpha_i > \tau \\ \text{False}, & \alpha_i \leq \tau \end{cases} \quad (6)$$

当 α_i 大于设定阈值 τ 时将当前体素标记为“占用”, 否则标记为“未占用”。在实际训练过程中, 为了提升效率与鲁棒性, 仅在预热阶段 (warm-steps) 对栅格内所有体素进行预测与更新。预热阶段完成后, 参与更新的体素只包括全部占用体素以及随机抽取的未占用体素。

Instant-NGP 使用均匀密集采样点参与网络训练, 本文则利用过滤后的密集采样点信息生成核心采样区域, 用于后续采样策略^[31]。整体流程如图 4 所示。

具体而言, 从相机出发沿像素方向朝占据栅格内随机生成 N 条光线, 设定密集采样点步进值。计算公式为:

$$\delta = \sqrt{3} \cdot \frac{d}{Gr \cdot k} \quad (7)$$

式中: d 为占据栅格的边长; Gr 为占据栅格的分辨率; k 为缩放因子。

按照步进值在每条光线上生成间隔均匀的采样点 (图 4a), 保留落在标记为“占据”体素内的采样点 (图 4b)。根据过滤后的采样点位置计算生成当前核心采样区域 (图 4c)。公式为:

$$\begin{aligned} S_{near} &= \max(t_s - \delta, near) \\ S_{far} &= \max(t_e + \delta, far) \end{aligned} \quad (8)$$

式中: S_{near} 、 S_{far} 分别表示生成采样区域的起始位置和结束位置; $near$ 和 far 分别表示光线与占据栅格交点的最近值和最远值。为提升鲁棒性, 在生成采样区域时, 为过滤后的第一个采样点 t_s 和最后一个采样点 t_e 均扩展间隔 δ , 同时保证生成的采样区域落在 $near$ 与 far 之间的范围内。通过以上方式, 占据栅格可为每条光线生成符合其当前特征的采样范围, 随着占据栅格占用依据—不透明度的不断更新, 采样区域的生成将持续动态调整。

2.2.2 基于动态范围的 HVS 优化

采用占据栅格引导采样的模型, 如 Instant-NGP、NeuS2 等, 通常将生成的均匀密集采样点中的空区间过滤后直接输入网络训练^[31-32]。这种方式虽然跳过了空区间, 但是过滤后的采样点仍然过多且分布均匀, 对物体表面没有趋向性。因此, 本文在生成的动态采样区域内引入 HVS 策略, 以便在物体表面附近产生密集采样点的同时保持整条光线采样点数量较少的特质。相较于基于占据栅格的常规采样策略, 本文方法生成的采样点更少且位置分布趋向于物体表面; 相较于常规 HVS 策略, 本文方法跳过了众多空区域且重要性采样点间距更小。具体如图 5 所示。

基于动态范围的 HVS 策略包含粗采样和细采样两个阶段。在粗采样阶段计算生成的采样区域 S_{near} 与 S_{far} 之间生成稀疏而均匀的 N_c 个采样点 S_c 。在细采样阶段计算所有 S_c 的权重 w_c 并进行归一化, 生成概率密度函数 (Proba-

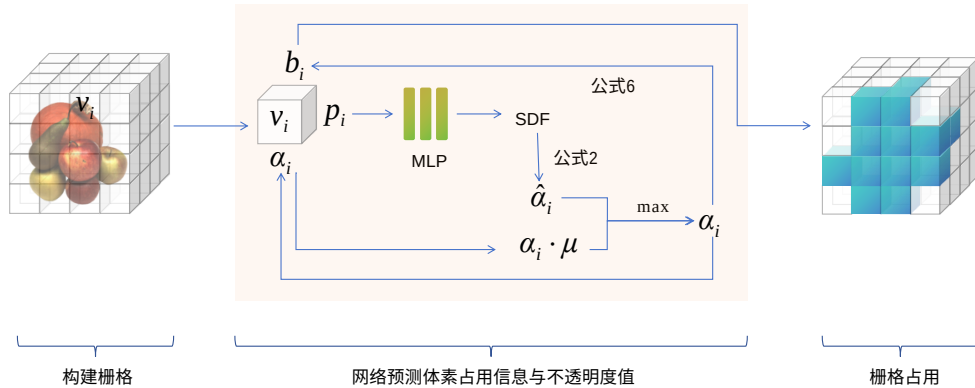


Fig. 3 Grid construction and binary information marking

图 3 栅格构建与二值化信息标记

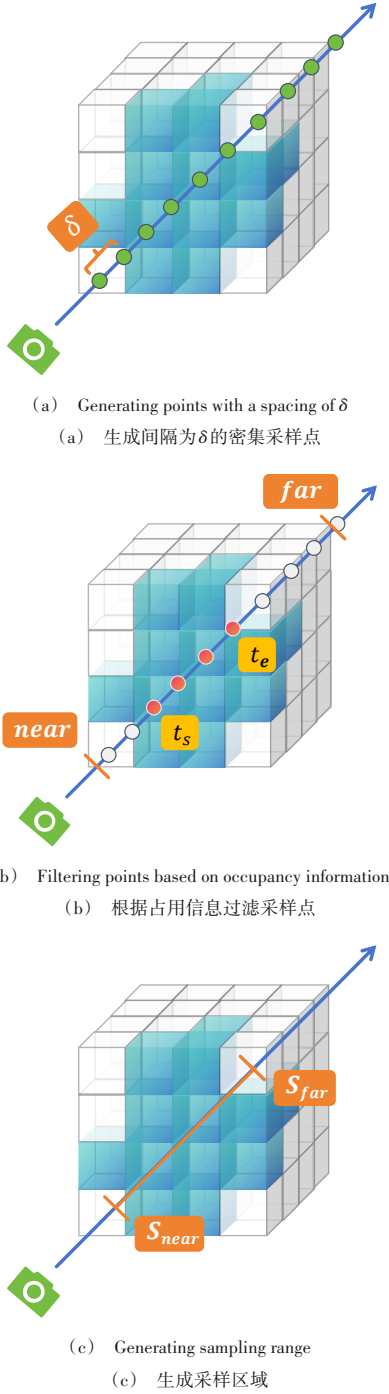


Fig. 4 Generating sampling range using occupancy information

图4 利用占用信息生成采样区域

bility Density Function, PDF), 然后根据累积分布函数(Cumulative Distribution Function, CDF)进行逆变换采样, 得到 N_f 个精细采样点 S_f 。细采样阶段通常可以重复 h 次, 每条光线上的总采样点数量为:

$$N = N_e + h \cdot N_f \quad (9)$$

根据前文分析, 体渲染包含由两种因素引起的误差, 分别为最靠近物体表面的采样点 t_j 与物体表面点 t^* 之间的距离差异带来的采样误差, 以及其余采样点 t_i 的高权重导致的权重误差。本文提出的缩小采样范围后的分层采样

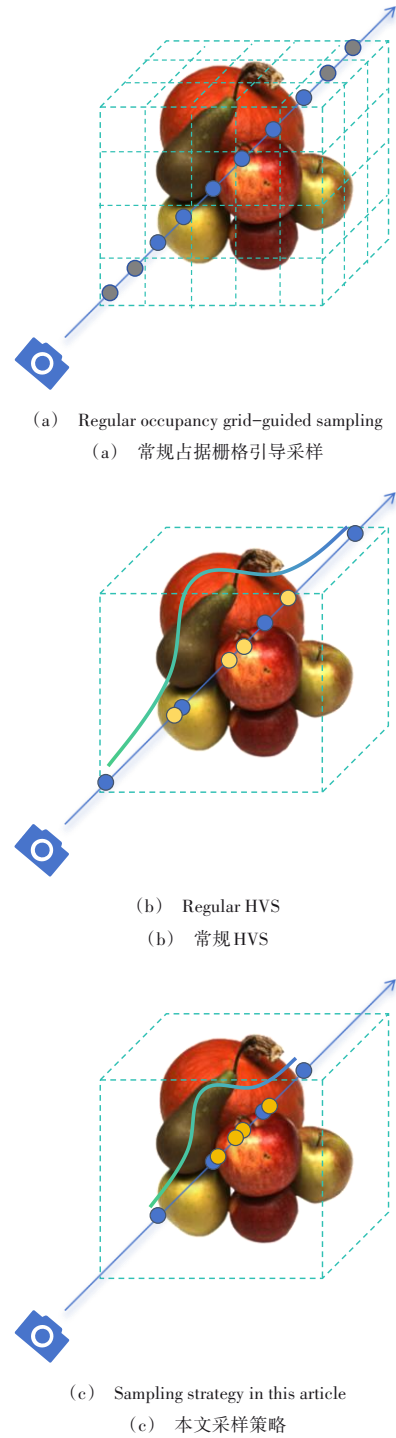


Fig. 5 Comparison of different sampling strategies

图5 不同采样策略比较

策略可减小精细采样点之间的间距, 在一定程度上降低了采样误差。此外, 本文还为采样点引入以下创新性约束:

$$L_{sample} = \frac{1}{N} \sum_{i=0}^N (\alpha_i \cdot \exp(1 - \alpha_i)) \quad (10)$$

以上正则化项的强度随着采样点不透明度 α_i 的增大而增强, 驱动整条光线上的 α_i 分布趋向稀疏化。根据式(3)可知, 采样点的权重随着 α_i 的降低而减小, 因此以上正则化项可有效减少远离物体表面采样点的权重, 最终缓解

权重误差。此外, 由于更多的 α_i 趋向 0, 在训练过程中还使得动态生成的采样区域不断缩小, 进一步提升了采样的有效性。

2.3 联合正则化

本文通过占据栅格引导生成核心采样区域, 然后经 HVS 得到采样点, 将每一个采样点输入到网络中预测出 SDF 值和颜色值, 最小化损失函数以反向传播优化网络。总损失函数为多个损失项的加权和。表示为:

$$L = L_{color} + \lambda_1 L_{eik} + \lambda_2 L_{mask} + \lambda_3 L_{sample} \quad (11)$$

式中: 前三项均借鉴 NeuS^[8]; 第四项为前文提及的采样点正则化项。渲染损失 L_{color} 的计算方式见式(4)。加入程函损失 L_{eik} 的目的是使 MLP 的预测满足 SDF 的性质, 约束每个采样点 SDF 梯度的模值为 1。 L_{eik} 的计算公式为:

$$L_{eik} = \frac{1}{n \cdot m} \sum_{k,i} \left(\left\| \nabla f(p_i) \right\|_2 - 1 \right)^2 \quad (12)$$

本研究面向的是采样策略优化问题, 实验中仅针对物体前景进行训练, 因此还使用掩膜损失以约束背景。表示为:

$$L_{mask} = BCE(M_k, \hat{O}_k) \quad (13)$$

式中: $BCE(\cdot)$ 为二值交叉熵; M_k 为第 k 条光线的背景掩膜值; $\hat{O}_k$ 为网络预测后计算出的第 k 条光线的透明度, 即采样点的权重之和。

这些损失项共同作用于网络的训练过程中, 通过设定各自的权重系数以确保采样点的网络预测值既符合几何特性, 又满足视觉效果的要求。渲染损失 L_{color} 作为基于神经体渲染过程中最主要和最直观的约束项, 在训练过程中的约束贡献最大, 因此其权重系数直接设定为 1, 其余三项的权重均小于 1。为了与现有工作保持一致性, 消除后续比较实验的歧义, 程函损失 L_{eik} 与掩膜损失 L_{mask} 的权重系数 λ_1 与 λ_2 遵循 NeuS 中的方案, 均设置为 0.1^[8]。本文提出的采样点约束项 L_{sample} 的计算值比程函损失和掩膜损失均大 100 倍左右, 考虑到该三项的约束贡献相差不大, 将采样点约束项的权重系数 λ_3 相应地缩小 100 倍, 将其基础值设置为 0.001。同时, 由于采样点约束会将更多采样点集中在局部区域, 为避免训练前期全局信息不足导致采样点分布错误, 本文还引入渐进式策略, 即训练开始时 λ_3 初始值为 0, 然后随迭代次数逐渐增加至 0.001。

3 实验方法与结果分析

3.1 实验环境与参数设置

所有实验均在单个 GeForce RTX3090 GPU 上执行。所用网络架构在 NeuS 的基础上融入 Instant-NGP 的多分辨率哈希编码和轻量级多层感知器配置, 具体包括以下组成部分: ① 16 级多分辨率哈希网格, 每级具有 2 个特征点, 分辨率从 2^5 逐渐扩展至 2^{16} ; ② 1 层 MLP 构成的几何网络, 其隐藏层维度为 64; ③ 2 层 MLP 构成的外观网络, 同样具有

64 维的隐藏层, 采用渐进式学习策略, 从粗略到精细逐步细化场景特征^[8,31]。

每经过 1 000 次迭代, 哈希网格的分辨率提升一级并激活相应的哈希特征。优化过程采用 AdamW 算法, 其 betas 参数设定为 0.9 和 0.99。几何网络与外观网络的初始学习率均设定为 0.01。前 5 000 次迭代采用固定衰减因子, 之后转为指数衰减策略。

超参数配置: 衰减系数 μ 设置为 0.95, 栅格分辨率 Gr 设为 128, 边长 d 为 2, 阈值 τ 设为 0.001, 步进缩放因子 k 为 8; 采样相关参数 N_c, N_f, h 分别设置为 64、16、4; 损失函数权重参数 λ_1 和 λ_2 均设置为 0.1, λ_3 则从 0 开始随迭代次数逐渐增加至 0.001。

3.2 数据集

采用 DTU 数据集集中的 15 个场景进行实验^[34]。该数据集的场景涵盖了各种不同材质和不同拓扑结构的物体, 每个场景包含不同视角下的 49 或 64 张分辨率为 $1\,600 \times 1\,200$ 的 RGB 图像, 并且自带真实点云数据用于后续几何质量评估。本文遵循文献[32]中 NeuS2 的设置, 对于新视角图像生成任务, 将每个场景分成训练集和测试集, 其中测试集由索引位置为 8、13、16、21、26、31、34 和 56 的图像组成, 其余图像归于训练集^[32]。在几何重建任务中则使用全部图像进行训练。此外, 本文专注于采样策略研究, 因此直接采用 IDR 提供的前景掩膜来训练和测试每个场景中的背景部分^[3]。

3.3 基准模型

选取 NeuS^[8]、ColMap^[17]、Instant-NGP^[31]、NeuS2^[32] 和 Voxurf^[35] 5 个基准模型与本文方法进行比较。其中, NeuS 是基于体渲染与 HVS 的代表性表面重建算法。ColMap 是一种应用广泛的经典多视图几何重建方法。Instant-NGP 是占据栅格引导采样的神经辐射场重建算法, 通过多分辨率哈希编码降低 MLP 网络体量。NeuS2 在 Instant-NGP 的基础上融入以 SDF 为表征的表面重建算法, 并加入多个在计算统一设备架构 (Compute Unified Device Architecture, CUDA) 上的优化技巧以提升重建效果。Voxurf 是采用体素-隐式混合架构的神经表面重建算法, 在保持重建质量的同时提升了训练速度。

3.4 评价指标

采用峰值信噪比 (Peak Singal-to-Noise Ratio, PSNR)、倒角距离 (Chamfer Distance, CD) 定量地评价各方法的新视角图像合成能力和几何表面重建质量。对未参与训练的测试集验证 PSNR, 值越高则表示新视角合成的图像质量越高, CD 值越低表示重建的几何质量越高。此外, 导出分辨率为 $512 \times 512 \times 512$ 的显式几何网格, 在不同视角下对几何细节进行定性评价。

3.5 实验结果与分析

3.5.1 与基准模型的比较

在定量评估中, 由于 ColMap 没有合成新视图的能力,

而 Voxurf 的测试集与本文设置不一致,只对这两个基准模型进行了几何质量评估,其他模型则分别使用 PSNR、CD 对新视角图像合成与几何重建两方面的效果进行评价。为保证数据集设置与数据比较的一致性,本文直接采用文献[8]中对于 ColMap 评价指标的报告,而 NeuS、Instant-NGP、NeuS2 的数据均来自于文献[32],Voxurf 的实验结果来自于文献[35]。表 1 为各方法的定量指标比较,各场景中最好的结果已加粗并用下划线标记,次优结果用双下划线标记。实验结果表明,无论是新视图合成能力还是几何重建质量,本文方法的平均表现均为最优且在大部分场景中取得最佳效果。

从时间效率方面分析,NeuS 单场景的训练时间为 8 h; Instant-NGP 和 NeuS 由于直接在 CUDA 中高效运行,均将训练时间缩短至 5 min; Voxurf 和本文方法均采用 pytorch 框架进行训练,用时分别为 16 min 和 20 min。与 NeuS 相比,本文方法在训练时间上加速了 24 倍。

选取 4 个不同类别的场景对几何重建质量进行定性评估,包括纹理细节丰富的 Scan24、平坦表面的 Scan40、规整曲面的 Scan63 以及金属光泽较大的 Scan110。对照模型选择基础的 NeuS 及其进阶版 NeuS2 作为代表,通过复现以上研究的官方代码来得到重建的几何网格模型结果。如图 6 所示,与 NeuS 相比,本文方法重建出的几何网格不仅保持表面平滑,而且保留了更多几何细节。NeuS2 在平坦

Table 1 Comparison of quantitative indicators among various methods

表 1 各方法定量指标比较

Scan ID	ColMap	Neus		Instant-NGP		Voxurf	Neus2		本文方法	
	CD ↓	PSNR ↑	CD ↓	PSNR ↑	CD ↓	CD ↓	PSNR ↑	CD ↓	PSNR ↑	CD ↓
24	0.81	26.49	0.83	28.32	1.68	0.65	<u>28.44</u>	<u>0.56</u>	<u>28.47</u>	<u>0.57</u>
37	2.05	26.17	0.98	<u>27.19</u>	1.93	<u>0.74</u>	27.14	0.76	<u>28.14</u>	<u>0.69</u>
40	0.73	27.66	0.56	<u>30.45</u>	1.57	<u>0.39</u>	29.70	<u>0.49</u>	<u>30.22</u>	<u>0.39</u>
55	1.22	27.78	<u>0.37</u>	<u>29.81</u>	1.16	<u>0.35</u>	<u>29.67</u>	<u>0.37</u>	29.65	0.38
63	1.79	30.63	1.13	31.22	2.00	<u>0.96</u>	<u>31.75</u>	<u>0.92</u>	<u>32.65</u>	0.99
65	1.58	27.42	<u>0.59</u>	27.78	1.56	<u>0.64</u>	<u>27.83</u>	0.71	<u>28.36</u>	0.79
69	1.02	<u>25.83</u>	<u>0.60</u>	24.79	1.81	0.85	24.84	0.76	<u>27.11</u>	<u>0.64</u>
83	3.05	30.00	1.45	31.23	2.33	1.58	<u>31.24</u>	<u>1.22</u>	<u>32.20</u>	<u>1.31</u>
97	1.40	26.40	<u>0.95</u>	<u>26.96</u>	2.16	<u>1.01</u>	26.86	1.08	<u>27.68</u>	1.15
105	2.05	29.63	0.78	<u>30.62</u>	1.88	<u>0.68</u>	30.57	<u>0.63</u>	<u>31.22</u>	0.69
106	1.00	25.87	<u>0.52</u>	25.62	1.76	0.60	<u>26.05</u>	0.59	<u>27.30</u>	<u>0.54</u>
110	1.32	28.82	1.43	28.60	2.32	1.11	<u>28.93</u>	<u>0.89</u>	<u>30.14</u>	<u>0.81</u>
114	0.49	28.80	<u>0.36</u>	<u>29.50</u>	1.86	<u>0.37</u>	28.98	0.40	<u>30.30</u>	<u>0.36</u>
118	0.78	27.36	<u>0.45</u>	<u>27.91</u>	1.80	<u>0.45</u>	27.82	<u>0.48</u>	<u>29.14</u>	<u>0.45</u>
122	1.17	31.19	<u>0.45</u>	<u>32.93</u>	1.72	0.47	32.48	0.55	<u>34.30</u>	<u>0.46</u>
mean	1.36	28.00	0.77	<u>28.86</u>	1.84	0.72	28.82	<u>0.70</u>	<u>29.79</u>	<u>0.68</u>

表面和规整曲面的表现与本文方法差异不大,但在光泽度大和纹理丰富的场景中,其重建细节稍逊。综上所述,本文方法在多场景重建中恢复出更多细节,更具泛化性,得益于其不断缩小的动态采样范围以及采样点与物体表面更近的距离。

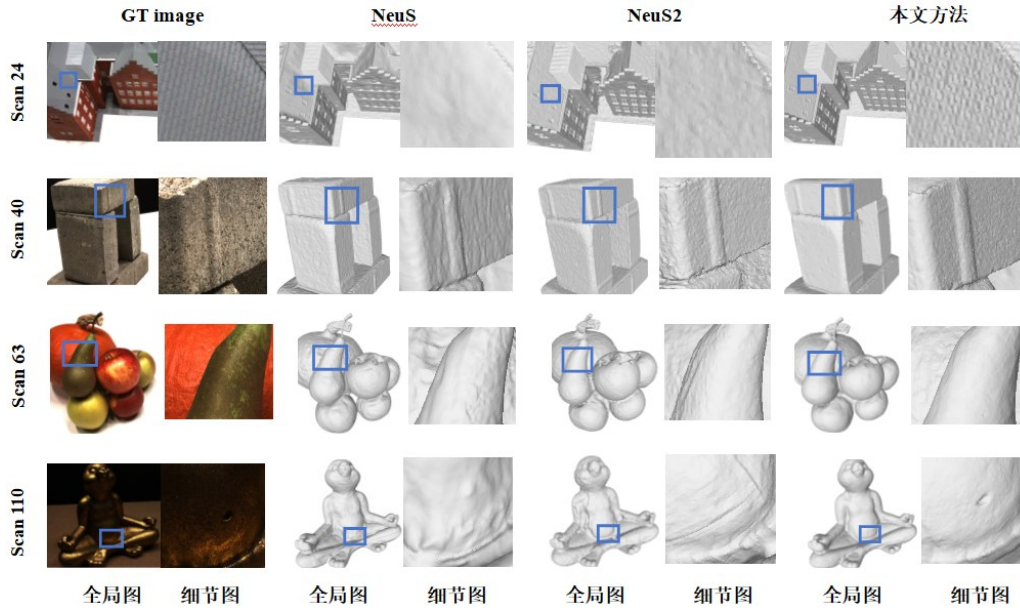


Fig. 6 Qualitative comparison of geometric reconstruction quality

图 6 几何重建质量定性比较

3.5.2 消融实验

为更好地评估本文方法各部分的有效性,对所提栅格引导生成核心采样区域(Occupancy Grid, OG)、基于动态范围的HVS和采样点正则化(L_{sample})策略分别进行消融实验。表 2 为各消融模型的消融方案设置,将采用常规 HVS 策略的模型定义为 Model A,由占用栅格引导采样的模型

定义为 Model B,将 HVS 策略融入到基于占据栅格跳过空区间方案的模型定义为 Model C,加入采样点正则化的无消融设置的模型定义为 Model Ours。

表 3 为各消融模型在 DTU 数据集中 15 个场景下的新视图合成能力与几何重建质量评价指标数据,最优结果已加粗并用下划线标记。可以看出,与 Model A 和 Model B 相

Table 2 Ablation scheme setup

表 2 消融方案设置

模型	模块		
	OG	HVS	L_{sample}
Model A	×	√	×
Model B	√	×	×
Model C	√	√	×
Model Ours	√	√	√

比, Model C 在新视图合成和几何重建两个维度的平均定量评价指标均得到提升。

Table 3 Ablation experiment result

表 3 消融实验结果

Scan ID	Model A		Model B		Model C		Model Ours	
	PSNR ↑	CD ↓	PSNR ↑	CD ↓	PSNR ↑	CD ↓	PSNR ↑	CD ↓
24	28.68	0.77	27.81	0.67	27.92	0.67	28.47	0.57
37	28.44	0.78	27.83	0.78	28.22	0.70	28.14	0.69
40	27.96	0.57	30.23	0.48	29.91	0.41	30.22	0.39
55	29.47	0.35	29.21	0.36	29.10	0.38	29.65	0.38
63	32.28	1.08	32.13	1.02	32.51	0.96	32.65	0.99
65	28.77	0.90	27.74	0.81	28.20	0.82	28.36	0.79
69	26.67	0.66	26.64	0.66	27.12	0.64	27.11	0.64
83	32.01	1.28	31.22	1.33	32.25	1.43	32.20	1.31
97	27.75	1.31	27.70	1.28	27.78	1.16	27.68	1.15
105	30.97	0.71	30.73	0.83	31.14	0.75	31.22	0.69
106	27.31	0.59	27.05	0.61	26.96	0.57	27.30	0.54
110	30.17	1.16	29.99	1.07	30.19	0.85	30.14	0.81
114	29.89	0.37	30.12	0.51	30.25	0.36	30.30	0.36
118	28.97	0.54	29.07	0.46	29.11	0.49	29.14	0.45
122	34.21	0.74	33.60	0.52	33.22	0.46	34.30	0.46
mean	29.57	0.79	29.40	0.76	29.63	0.71	29.79	0.68

为了更进一步分析其中的原因,图 7 以 scan24 场景为例展示了 OG 与 HVS 对于精细采样点间最小距离的影响。基于占据栅格引导采样的 Model B 的采样点间距是恒定的;采用常规 HVS 策略的 Model A 由于多重重要性采样而缩短了精细采样点的间距;Model C 的数据则证明了在栅格引导生成的动态区域内采用 HVS 技术可以进一步减少精细采样点间的距离,这也是 Model C 重建质量最高的原因。

同样以 scan24 场景为例,图 8 则说明随着训练的进

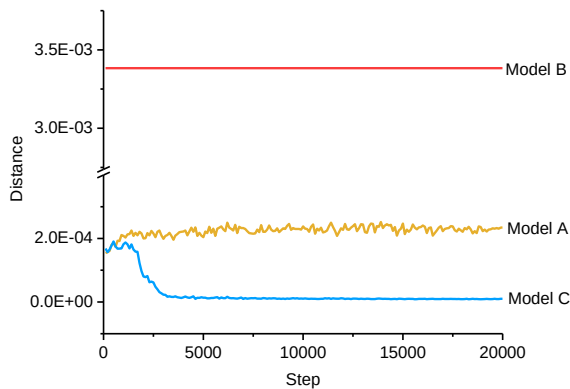


Fig. 7 The minimum distance between fine sampling points of each ablation model in scan24

图 7 scan24 场景中各消融模型精细采样点之间的最小距离

行, Model B、Model C 和 Model Ours 由于采用了占据栅格的引导,生成的采样区域与原始采样范围之间的占比逐渐减小,而引入了采样点正则化约束的 Model Ours 的缩小幅度最大。因此,与 Model C 相比,Model Ours 的新视角合成能力和几何重建质量均进一步提升。

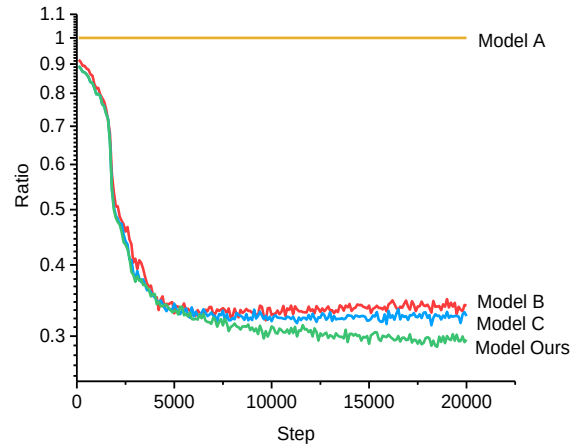


Fig. 8 The sampling range ratio of each ablation model in scan24

图 8 scan24 场景中各消融模型的采样区域占比

通过消融实验对这 3 种策略进行了评估,结果表明基于动态采样范围的 HVS 策略能高效地将采样点集中在物体表面附近,而对采样点引入的损失机制可以缩小核心采样范围并使采样点更靠近物体表面。以上结果证实了本文方法在新视角合成与几何重建两方面任务中的优化策略均是有效的。

4 结语

本文针对神经隐式表面重建技术中的采样策略进行了深入分析与优化,所提出的基于动态采样范围的方案在新视角合成和几何重建任务中均取得一定程度的质量提升。该方案基于占据栅格生成逐步缩小的动态采样范围,减少了无效空间内的采样,提高了光线与采样点的利用效率;通过在动态采样范围内使用 HVS 策略,进一步集中采样点于物体表面附近,减少了采样偏差;而引入的采样点正则化约束减少了渲染时的权重偏差。仅使用二维图像作为输入时,应用本文优化的采样策略可以在较短时间内得到表面平滑且细节丰富的三维几何网格。然而,本研究仍存在不足之处,如实验涉及的重建对象局限为物体级、未考虑真实户外场景的光照影响等。未来将探索将所提方法应用于更广泛的数据集和更复杂的大型场景中,进一步优化其性能,并探索与其他先进技术融合的可能性。

参考文献:

- [1] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision [M]. Cambridge: Cambridge University Press, 2003.
- [2] DE BONET J S, VIOLA P. Voxels: probabilistic voxelized volume reconstruction [C]//Proceedings of International Conference on Computer Vi-

- sion, 1999:1–8.
- [3] SEITZ S M, DYER C R. Photorealistic scene reconstruction by voxel coloring [J]. *International Journal of Computer Vision*, 1999, 35: 151–173.
- [4] KAZHDAN M, HOPPE H. Screened poisson surface reconstruction [J]. *ACM Transactions on Graphics*, 2013, 32(3):1–13.
- [5] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016:4104–4113.
- [6] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: representing scenes as neural radiance fields for view synthesis [J]. *Communications of the ACM*, 2021, 65(1): 99–106.
- [7] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: learning continuous signed distance functions for shape representation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 165–174.
- [8] WANG P, LIU L, LIU Y, et al. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction [J]. *Advances in Neural Information Processing Systems*, 2021, 34: 27171–27183.
- [9] YARIV L, KASTEN Y, MORAN D, et al. VolSDF: volume rendering of signed distance functions for 3D scene reconstruction [J]. *Advances in Neural Information Processing Systems*, 2021, 34:22078–22089.
- [10] DREBIN R A, CARPENTER L, HANRAHAN P. Volume rendering [J]. *ACM SIGGRAPH Computer Graphics*, 1988, 22(4):65–74.
- [11] LORENSEN W E, CLINE H E. Marching cubes: a high resolution 3D surface construction algorithm [J]. *ACM SIGGRAPH Computer Graphics*, 1987, 21(4):163–169.
- [12] FU H, LIU S, WANG J, et al. Geo-Neus: geometry-consistent neural implicit surfaces learning for multi-view reconstruction [J]. *Advances in Neural Information Processing Systems*, 2022, 35:12112–12123.
- [13] YARIV L, KASTEN Y, MORAN D, et al. Multiview neural surface reconstruction by disentangling geometry and appearance [J]. *Advances in Neural Information Processing Systems*, 2020, 33:2492–2502.
- [14] NIEMEYER M, MESCHEDER L, OECHSLE M, et al. Differentiable volumetric rendering: learning implicit 3d representations without 3d supervision [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020:3504–3515.
- [15] OECHSLE M, PENG S, GEIGER A. Unisurf: unifying neural implicit surfaces and radiance fields for multi-view reconstruction [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021:5589–5599.
- [16] YU Z, PENG S, NIEMEYER M, et al. Monosdf: exploring monocular geometric cues for neural implicit surface reconstruction [J]. *Advances in Neural Information Processing Systems*, 2022, 35:25018–25032.
- [17] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016:4104–4113.
- [18] LEVOY M. Efficient ray tracing of volume data [J]. *ACM Transactions on Graphics*, 1990, 9(3):245–261.
- [19] BARRON J T, MILDENHALL B, TANCIK M, et al. Mip-nerf: a multi-scale representation for anti-aliasing neural radiance fields [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021:5855–5864.
- [20] DADON D, FRIED O, HEL-OR Y. Ddnerf: depth distribution neural radiance fields [C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023:755–763.
- [21] LI L, ZHANG J. L0-Sampler: an L0 model guided volume sampling for NeRF [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024:21390–21400.
- [22] SUN J, CHEN X, WANG Q, et al. Neural 3d reconstruction in the wild [C]// *ACM SIGGRAPH 2022 Conference Proceedings*, 2022:1–9.
- [23] XU J X. A neural radiance field view synthesis method based on depth prior optimization [D]. Wuhan: Huazhong University of Science and Technology, 2023.
徐嘉希. 基于深度先验优化的神经辐射场视点合成方法 [D]. 武汉: 华中科技大学, 2023.
- [24] RAU A, AKLILU J, HOLSINGER F C, et al. Depth-guided NeRF training via earth mover's distance [DB/OL]. <https://arxiv.org/abs/2403.13206>.
- [25] WANG B, ZHANG D, SU Y, et al. Enhancing view synthesis with depth-guided neural radiance fields and improved depth completion [J]. *Sensors*, 2024, 24(6): 1919.
- [26] QIU J, ZHU Y, JIANG P T, et al. RDNeRF: relative depth guided NeRF for dense free view synthesis [J]. *The Visual Computer*, 2024, 40(3):1485–1497.
- [27] WANG J, XIAO J, ZHANG X, et al. Depth-based dynamic sampling of neural radiation fields [J]. *Electronics*, 2023, 12(4): 1053.
- [28] FANG J, XIE L, WANG X, et al. Neusample: neural sample field for efficient view synthesis [DB/OL]. <https://arxiv.org/abs/2111.15552>.
- [29] KURZ A, NEFF T, LV Z, et al. Adanerf: adaptive sampling for real-time rendering of neural radiance fields [C]// *European Conference on Computer Vision*, 2022:254–270.
- [30] NEFF T, STADLBAUER P, PARGER M, et al. DONeRF: towards real-time rendering of compact neural radiance fields using depth oracle networks [J]. *Computer Graphics Forum*, 2021, 40(4):45–59.
- [31] MÜLLER T, EVANS A, SCHIED C, et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. *ACM Transactions on Graphics*, 2022, 41(4):1–15.
- [32] WANG Y, HAN Q, HABERMANN M, et al. Neus2: fast learning of neural implicit surfaces for multi-view reconstruction [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 3295–3306.
- [33] LI Z, MÜLLER T, EVANS A, et al. Neuralangelo: high-fidelity neural surface reconstruction [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023:8456–8465.
- [34] JENSEN R, DAHL A, VOGIATZIS G, et al. Large scale multi-view stereopsis evaluation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014:406–413.
- [35] WU T, WANG J, PAN X, et al. Voxurf: voxel-based efficient and accurate neural surface reconstruction [C]// *Kigali: International Conference on Learning Representations*, 2023.

(责任编辑:尹晨茹)