

薛晨霞,孙英锴,陈占国,等.面向室内三维模型重建的神经辐射场网络[J].辽宁工程技术大学学报(自然科学版),2025,44(5):627-633.DOI:10.11956/j.issn.1008-0562.20240433
XUE Chenxia,SUN Yingkai,CHEN Zhanguo,et al.Neural radiation field network for indoor 3D model reconstruction[J].Journal of Liaoning Technical University(Natural Science),2025,44(5):627-633.DOI:10.11956/j.issn.1008-0562.20240433

面向室内三维模型重建的神经辐射场网络

薛晨霞¹,孙英锴¹,陈占国¹,王国亮¹,李金峰¹,许文洁¹,于保才^{2*}

(1. 国家能源集团宝日希勒能源有限公司 物供中心,内蒙古 呼伦贝尔 021000;
2. 辽宁工程技术大学 鄂尔多斯研究院,内蒙古 鄂尔多斯 017000)

摘要:针对室内三维重建弱纹理区域鲁棒性不足、重建效果不佳的问题,基于神经辐射场原理,提出一种新型的室内三维重建方法M-HashRecon。该算法采用点云选取模块提取关键点云信息,并引入多分辨率哈希编码实现点云的多尺度特征索引;设计残差模块优化模型性能,提升深层网络的训练效率。在ScanNet数据集的4个典型场景中开展实验,对实验结果和模型的收敛性进行分析。研究结果表明:该算法的F-score综合指标显著优于对比算法,对多个场景重建精度较高,稳定性较好。研究结论可为后续高精度室内三维重建系统的设计提供参考。

关键词:室内三维重建;深度引导采样;多分辨率哈希编码;残差模块;ScanNet数据集

中图分类号:TP391

文献标志码:A

文章编号:1008-0562(2025)05-0627-07

Neural radiation field network for indoor 3D model reconstruction

XUE Chenxia¹,SUN Yingkai¹,CHEN Zhanguo¹,WANG Guoliang¹,LI Jinyi¹,
XU Wenjie¹,YU Baocai^{2*}

(1. Material Supply Center of Baorixile Energy Company Limited, China Energy Investment Corporation, Hulunbuir 021000, China;2. Ordos Institute of Liaoning Technical University, Ordos 017000, China)

Abstract: Aiming at the problems of insufficient robustness and poor reconstruction effect of weak texture regions in indoor 3D reconstruction, a new indoor 3D reconstruction method M-HashRecon is proposed based on the principle of neural radiation field. The algorithm utilizes a point cloud selection module to extract the key point cloud information, and introduces multi-resolution hash coding to realize the multi-scale feature index of the point cloud. The residual module is designed to optimize the performance of the model and improve the training efficiency of the deep network. Experiments are carried out in four typical scenarios of ScanNet dataset, and the experimental results and the convergence of the model are analyzed. The research results show that the F-score comprehensive index of the algorithm is significantly better than that of the comparison algorithm, and the reconstruction accuracy of multiple scenes is high and the stability is good. The research conclusions can provide reference for the design of subsequent high-precision indoor three-dimensional reconstruction system.

Keywords: indoor 3D reconstruction; deep guided sampling; multi-resolution hash coding; residual module; ScanNet dataset

0 引言

随着科技的不断发展,智能化技术不断进步,利用计算机视觉和传感技术,将室内环境中的物体、结构和布局信息转换为高质量的三维模型已

成为研究的热点。这一技术在建筑设计、室内装修、虚拟现实(VR)、增强现实(AR)和智能家居等领域展现了巨大的应用潜力。

在三维重建领域,经典的多视图立体(multi-view stereo, MVS)技术通过捕获多视角图像,利

收稿日期:2024-11-29 修回日期:2025-02-24 接受日期:2025-03-02 责任编辑:焦丽

基金项目:辽宁省教育厅基本科研项目(LJ212410147003)

作者简介:薛晨霞(1980-),女,内蒙古呼伦贝尔人,工程师,主要从事机电管理、物资管理方面的研究。E-mail: mxh_wy2022@163.com

通信作者:于保才(1993-),男,河南新乡人,博士,工程师,主要从事机器视觉方面的研究。E-mail: 936036802@qq.com

用特征点匹配与三角测量技术重建三维模型。该技术通过分析多视角图像中的对应点关系,估计相机的相对位置和姿态,精确地复原场景中物体的三维坐标,构建出完整的三维空间结构^[1]。MVS方法对纹理丰富的区域重建效果较好,但对低纹理或重复纹理区域的重建效果欠佳。随着计算机视觉与图像处理技术的发展,基于深度传感器的三维重建方法逐渐成为研究的重点。SHOTTON等^[2]提出一种基于Kinect的低成本深度传感方法,通过捕捉RGB图像与深度信息生成三维点云,显著降低了硬件成本,为大规模三维重建提供可能。IZADI等^[3]提出KinectFusion方法,实时融合多个Kinect深度帧,实现高密度、实时的三维重建,有效提升重建的精度和效率。SCHÖNBERGER等^[4]开发了COLMAP工具,这一开源的跨平台软件集成了特征检测、匹配、相机姿态估计和三维点云重建等模块,为室内场景的三维重建提供了全面的解决方案。深度学习技术的迅猛发展为三维重建提供了全新的方法与思路。SITZMANN等^[5]提出DeepVoxels方法,通过学习持久的三维特征嵌入,有效解决了传统多视图立体重建中特征不稳定的问题。OECHSLE等^[6]进一步提出UNISURF方法,将神经隐式表面(implicit surface)与辐射场(radiation field)统一建模,不仅提升了多视图重建中表面表示的一致性,还增强了对复杂场景的适应能力。此外,Martin-Brualla等^[7]提出NeRF-W方法,通过改进神经辐射场(NeRF)的架构,解决了非受控环境下场景重建问题,在复杂动态场景中表现尤为突出。但在重建低纹理区域或平坦表面上,现有方法依然存在细节捕捉不足的问题。为此,研究者提出了一系列改进方法。VolSDF-G通过引入几何约束和稀疏深度信息,增强了对低纹理区域细节的捕捉能力,从而生成更为精细的重建结果^[8]。在此基础上,VolSDF-P进一步将平面约束融入VolSDF框架,显著提升了在平坦区域及弱纹理区域中的重建精度与稳定性^[9]。

在室内三维重建中,神经辐射场(NeRF)及其变种(如Multi-Scale NeRF、DeepNeRF)已成为主流方法。NeRF及其变体通过从单视角或多视角RGB图像中学习空间和光照特性,能够生成高质量的三维模型,特别是在光照变化、复杂纹理和透明材质等场景中展现出卓越性能^[10]。此外,多模态学习方法(如M3DNet)通过融合多模态数据(如RGB、深度图和激光雷达数据),结合深度学习优化技术,不仅提升了模型处理复杂室内结构的能力,还显著增强了对高维空间数据的适应性与鲁棒性^[11-12]。

尽管上述方法在大多数情况下可以获得理想的三维重建结果,但在处理弱纹理区域时仍存在鲁棒性不足的问题,为此本文提出一种改进的室内三维重建网络。在构建三维模型的过程中,首先通过点云选取模块对输入点云数据进行精准筛选,确保关键信息的高效提取;随后,采用多分辨率哈希编码(multi-resolution hash encoding)对点云数据进行多尺度索引,以增强对不同分辨率几何特征的建模能力;最后,在多层感知机(MLP)网络中引入残差模块(residual module),通过特征增强和梯度优化进一步提升网络的训练效果与重建精度。

1 本文方法

M-HashRecon网络总体结构见图1。通过旋转矩阵 R 和平移矩阵 T 获取视角信息。在射线方向上,借助点云选择模块从点云中筛选出具有代表性的点。利用基于深度矩阵的深度信息 d 引导采样,在深度值附近均匀地选择近地表的点。采用多分辨率哈希编码对筛选出的点云进行不同分辨率的索引处理。将这些编码后的点云输入至以下网络模块:符号距离预测网络,用于预测每个点的有符号距离(SDF);颜色渲染网络,用于预测每个点的颜色信息;几何约束网络,通过语义分割增强场景理解。综合这些信息,最终生成三维模型。

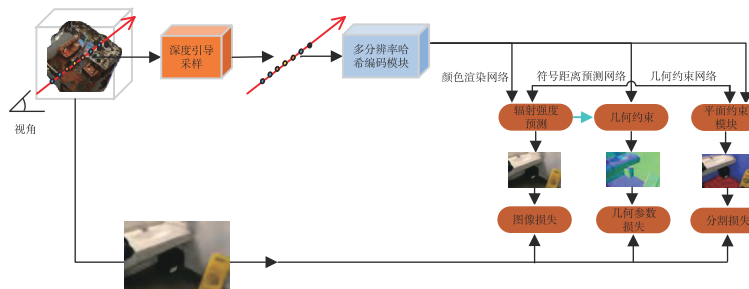


图1 M-HashRecon网络总体结构

Fig.1 overall architecture of the M-HashRecon network

1.1 点云选取模块

为有效提高训练精度与速度,采用点云选取模块通过深度值引导筛选出具有代表性的点云。在点云选取模块中,利用深度传感器对输入图像中的每个像素的深度值进行估算。该步骤能够捕捉到场景中物体与相机之间的距离信息。令 $z(\cdot)$ 为像素的深度值, $f(\cdot)$ 为深度传感器函数, $D(\cdot)$ 为像素距离信息,对应的深度值为

$$D(x,y)=f(z(x,y)). \quad (1)$$

对无效深度值进行筛选,将阈值小于0的射线去除,挑选后的点云表示为

$$X_{\text{inds}} = \begin{cases} \text{True} & D \leq 0 \\ \text{False} & D > 0 \end{cases}, \quad (2)$$

式中, D 为得到的深度值。

对筛选后获得的射线进行均匀采样,在射线 $[0,2]$ 内选取100个点云,构建初步点云集合 P_c 。为了更精确地捕获射线与场景中物体表面的交点信

息,在每条射线对应的深度值 D 附近,即在区间 $[D-0.01, D+0.01]$ 内,均匀选取25个点云,获得更为精细的点云集合 P_f ,最终点云集合为

$$P_{\text{final}} = P_f \cup P_c. \quad (3)$$

1.2 点云编码模块

多分辨率哈希编码在场景表示方面具有较强的可扩展性,能够为新视图合成等任务生成更多细节。为更好地学习和表示空间中的细节,提升渲染质量和精度,采用多分辨率哈希编码对点云进行多尺度特征索引,见图2。给定输入的空间坐标 x ,在不同尺度下选取周围的4个特征点。这些特征点以等间距的方式分布,尺度不同,特征点间距不同。为在每个尺度下提取对应的特征点,通过哈希编码表对这些特征点进行索引。通过2D线性插值方法计算目标坐标处的特征向量。将所有尺度下得到的特征向量与其他输入信息(如视角方向或纹理信息)进行拼接,送入多层感知机(MLP)网络进行下一步处理。

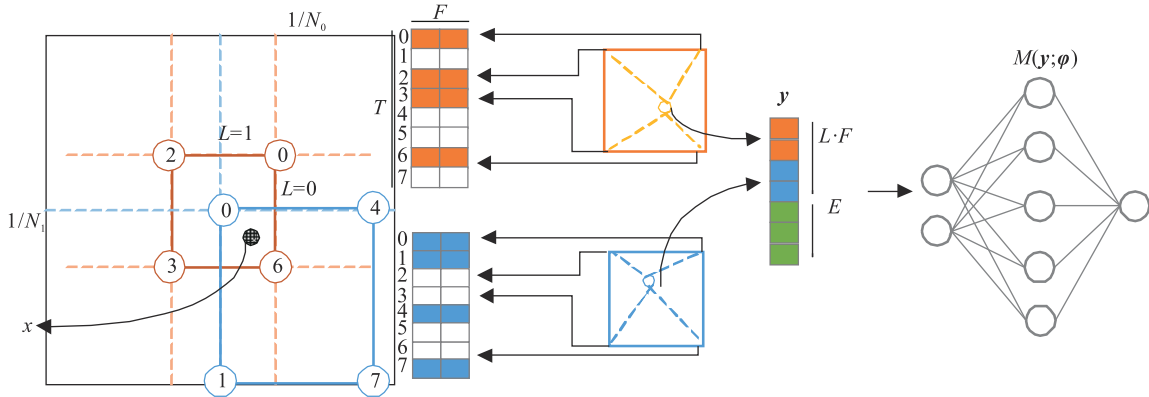


图2 多分辨率哈希编码

Fig.2 multi resolution hash encoding

假设不同的空间分辨率集合为 $\{V_1, V_2, \dots, V_L\}$,给定一个输入位置 x_i ,将 x_i 映射到每个网格分辨率 V_L 的相应位置,此时 $x_{iL} = x_i \cdot V_L$ 。 V_L 一定时,特征向量 $y_l(x_{iL}) \in \mathbf{R}^C$ 可以通过对网格单元角点的哈希条目进行三线性插值得到^[13]。所有分辨率下的编码特征最终会被拼接成一个综合的特征向量,即

$$y_l(x_i) = (y_1(x_{i1}), y_2(x_{i2}), \dots, y_L(x_{iL})). \quad (4)$$

该方法为每个网格角点分配了唯一对应的编码,虽然避免了特征混淆或重叠的问题^[14],但在使用体积网格表示特征时,随着分辨率提高,需要同时处理的参数数量增长较快,内存消耗较大。通常采用层次化的空间分解结构(如八叉树)解决此问题,但这可能会导致粗分辨率下的曲面表

示误差无法被精细分辨率修复^[15]。与此相比,哈希编码方法不依赖于空间层次结构,通过梯度平均来缓解此问题^[16],但受位置的影响,优化更新只能在局部哈希网格中传播,缺乏全局平滑性^[17-18]。

为解决局部更新的缺陷,引入数值梯度。当数值梯度的步长小于哈希编码网格的尺寸时,数值梯度与解析梯度相当;而当步长较大时,多个网格单元的哈希项共同参与曲面法线的计算。通过此方式,曲面法线的反向传播机制允许多个哈希条目同时更新,从而改进了优化过程^[19]。

采用数值梯度计算表面法线,需要额外的SDF样本。对于给定的采样点 $x_i = (x_i, y_i, z_i)$,以步长 ϵ 为半径,沿 x_i 邻域规范坐标轴,对每个轴进行两次额外采样。例如,令 $\epsilon_x = [\epsilon, 0, 0]$,则表面法线

的 x 分量为

$$\nabla_x f(x_i) = \frac{f(\gamma(x_i + \epsilon_x)) - f(\gamma(x_i - \epsilon_x))}{2\epsilon}, \quad (5)$$

式中, γ 为位置编码函数。

此外,采用从粗到细的优化策略,逐步重建表面的细节。通过数值梯度计算高阶导数,可在粗分辨率和精细分辨率两个层面上同时进行优化。数值梯度作为一种平滑操作,通过步长 λ 来控制重建的分辨率和细节恢复的程度。在数值表面法线的计算中,较大的 λ 可以保证表面法线在较大尺度上保持一致,从而生成平滑连续的表面;较小的 λ 会影响更小的区域,产生更多的细节但较为粗糙。

通常,哈希网格的大小初始化为最粗的网格,并在优化过程中以指数方式减小,以适应不同的哈希值分布。为捕捉几何细节,将精细哈希网格从粗优化过程中分离,采用大步长的“重新学习”策略,逐步过渡到小步长的“再学习”阶段,从而调整更细的哈希网格。在遇到几何细节丢失的情况时,可以启用初始粗哈希网格集,并逐步激活更精细的哈希网格进行优化修正。

为进一步提高重建表面的平滑性,引入平均曲率的正则化先验。平均曲率通过离散拉普拉斯

算子进行计算,其正则化损失函数为

$$L_{\text{curv}} = \frac{1}{N} \sum_{i=1}^N |\nabla^2 f(x_i)|, \quad (6)$$

式中: N 为样本点数; $|\nabla^2 f(x_i)|$ 为在位置 x_i 处拉普拉斯算子(即二阶梯度)的绝对值。

采用多分辨率哈希编码结合数值梯度计算表面法线及优化策略,能有效提高三维重建模型的分辨率和精度,同时保持表面光滑性并解决传统方法中可能遇到的局部优化问题。

1.3 残差连接模块与整体网络

为解决深层神经网络训练中的梯度消失和梯度爆炸问题,引入残差模块提高网络训练的效率和性能。残差模块引入跳跃连接(skip connection),使输入能直接跳跃至输出,得到网络每一层学习输入与输出的残差(residual),标准残差模块的输出为

$$y = F(x) + x, \quad (7)$$

式中: x 为输入特征; $F(x)$ 为主分支的变换函数。

室内三维重建网络结构见图3。主体网络主要分为:符号距离函数、颜色渲染、几何约束网络。

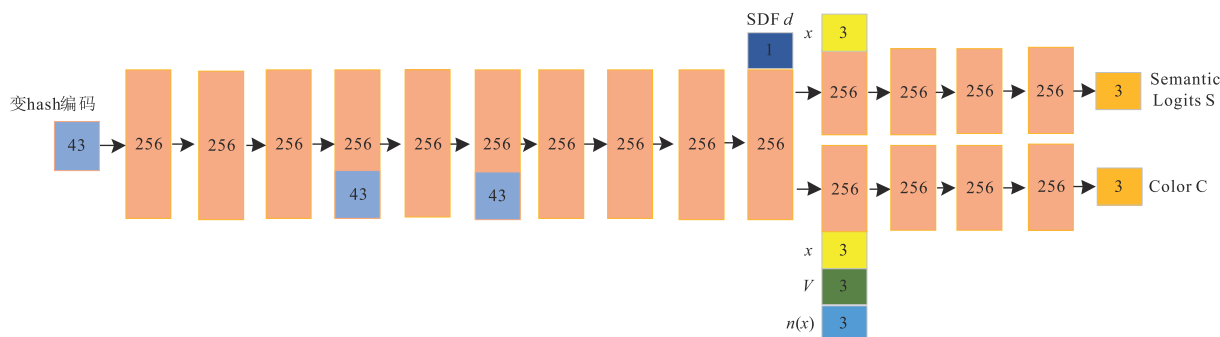


图3 室内三维重建网络结构

Fig.3 indoor 3D reconstruction network structure

符号距离函数的输入为经点云编码模块处理后得到的特征。该网络由9层MLP网络构成,为解决网络层数过大而造成的梯度消失与爆炸问题,在该网络的第4层与第6层引入残差模块,给定一个三维点 x ,通过符号距离函数 F_d 得到有符号距离SDF值 $d(x)$,几何特征 $z(x)$,表示为

$$(z(x), d(x)) = F_d(x). \quad (8)$$

颜色渲染网络旨在训练出给定点云的RGB颜色值,以近似辐射函数,选用空间点 x ,视图方向 v ,法线 $n(x)$, $z(x)$ 作为输入,颜色 $c(x)$ 作为输出,表示为

$$c(x) = F_c(x, v, n(x), z(x)). \quad (9)$$

为增强地板及墙体等低纹理区域的重建质量,引入针对这些区域的几何约束网络。为精确监督地板区域的重建,采用基于曼哈顿世界假设的方法。地板表面通常与垂直于 z 轴的平面平行,基于该假设给出计算地板像素的法向损耗方法,确保地板区域在重建过程中能够遵循预期几何约束,表示为

$$L_f(r) = |1 - n(x_r) \cdot n_f|, \quad (10)$$

式中: x_r 是相机光线表面交点; $n(x_r)$ 是交点处的法线; $n_f = \langle 0, 0, 1 \rangle$ 为地板区域的假设法线方向。

引入一个可学习向量 n_w 重建墙体区域,要求墙体表面点的法线方向之间保持平行或正交的关

系。在初始化时,将 n_w 设置为 $\langle 0,0,1 \rangle$ 并垂直于 n_f ,墙体区域的几何约束损失函数为

$$L_w(r) = \min_{i \in \{-1,0,1\}} |i - n(x_r) \cdot n_w|, \quad (11)$$

式中: $n(x_r)$ 为点 x_r 处法向量; $i \in \{-1,0,1\}$ 表示3个候选方向之间的几何关系, $i=0$ 表示法向量垂直或正交, $i=\{1,-1\}$ 表示法向量平行(方向一致或相反)。

最终的几何约束损失函数为

$$L_{geo} = \sum L_f(r) \sum L_w(r), \quad (12)$$

式中: $r \in F$ 表示属于地板或天花板; $r \in w$ 表示属于墙面,须满足与地板或天花板正交的垂直约束。

2 实验分析

2.1 评价指标

实验中,采用5种标准指标评估方法的性能,分别是准确性 I_{ACC} 、完整性 I_{Comp} 、精度 I_{Prec} 、召回率 I_{Recal} 和 $I_{F-score}$ 。

I_{ACC} 衡量预测点云中最差匹配点的误差,即所有预测点与其最近真实点距离的最小值。

$$I_{ACC} = \min_{p \in P} \left(\min_{p^* \in P^*} \|p - p^*\| \right), \quad (13)$$

式中: P 为算法预测的点云集合,其中的点表示为 p ; P^* 为真实点云集合,其中的点表示为 p^* 。

I_{Comp} 衡量真实点云中每个点到预测点云的平均最近距离,评估预测点云对真实场景的覆盖程度。值越小表示缺失的区域越少,重建结果越完整。

$$I_{Comp} = \text{mean}_{p^* \in P^*} \left(\min_{p \in P} \|p - p^*\| \right). \quad (14)$$

I_{Prec} 表示预测点云中有多少比例的点与真实点的距离小于阈值(如0.05),衡量重建结果的纯净性,值越高表示冗余或噪声点越少。

$$I_{Prec} = \text{mean}_{p \in P} \left(\min_{p^* \in P^*} \|p - p^*\| < 0.05 \right). \quad (15)$$

I_{Recal} 表示真实点云中有多少比例的点与预测点的距离小于阈值,反映算法对真实场景的覆盖率。值越高表示遗漏的真实点越少。

$$I_{Recal} = \text{mean}_{p^* \in P^*} \left(\min_{p \in P} \|p - p^*\| < 0.05 \right). \quad (16)$$

$I_{F-score}$ 表示精度和召回率的调和平均数,综合评估算法的整体性能,平衡精确性与覆盖率,避免单一指标的片面性。

$$I_{F-score} = \frac{2I_{Prec}I_{Recal}}{I_{Prec} + I_{Recal}}. \quad (17)$$

$I_{F-score}$ 能够同时考虑精度和召回率,因此选择 $I_{F-score}$ 作为总体指标,以此评估模型重建的准确性。

2.2 实验设置

实验使用的计算机基本配置如下:处理器为IntelCore i7-8700@3.2GHz, CPU 为 NVIDIA RTX 3090。进行网络训练时,本实验选用Adam优化器,Adam优化器结合了动量方法和RMSProp算法的优点,通过自适应学习率加速神经网络的训练收敛。

2.3 数据集

ScanNet数据集是一个大规模三维扫描数据集,由便携式RGB-D传感器(如微软Kinect)对多种室内场景进行扫描获得^[20]。该数据集包含数千个高质量三维网格模型、对应的RGB-D图像序列、相机位姿信息以及详细的语义标注,包括客厅、厨房、办公室和浴室等多种室内场景,广泛应用于三维重建、语义分割、场景理解和机器人导航等任务。

2.4 实验结果

基于上述配置,采用 I_{ACC} 、 I_{Comp} 、 I_{Prec} 、 I_{Recal} 和 $I_{F-score}$ 五项指标,将COLMAP算法、NeRF算法、NeuS算法、VolSDF算法、VolSDF-P算法、VolSDF-G算法进行对比分析,结果见表1。

表1 不同算法的评价指标对比结果
Tab.1 comparison results of evaluation indexes of different algorithms

算法	评价指标				
	I_{ACC}	I_{Comp}	I_{Prec}	I_{Recal}	$I_{F-score}$
COLMAP	0.047	0.235	0.711	0.441	0.537
NeRF	0.735	0.177	0.131	0.290	0.176
Neus	0.179	0.208	0.313	0.275	0.291
VolSDF	0.414	0.120	0.321	0.394	0.346
VolSDF-P	0.100	0.118	0.493	0.427	0.459
VolSDF-G	0.106	0.101	0.507	0.469	0.487
本文算法	0.059	0.065	0.699	0.664	0.672

从表1可见,本文算法在各项指标上均表现出色,其准确性为0.059,仅次于COLMAP算法,优于NeRF算法和NeuS算法;其完整性是所有对比方法中的最优值,显著提升了几何结构的覆盖程度;其 I_{Prec} 和 I_{Recal} 接近或超过COLMAP算法,远超NeRF算法、NeuS算法;其 $I_{F-score}$ 值最大。综合看来,本文算法的整体性能最优。

为了进一步验证本文方法的有效性,将不同算法进行对比,见图4。



图4 不同算法的实验结果对比

Fig.4 comparison chart of different methods in recent years

从图4可见,与其他算法相比,采用本文算法重构的图像对原图像的复杂纹理区域、低纹理区域(如墙面和地板)还原最优,具有较好的平滑性与完整性。对原图像的几何边界(如墙体和门框)清晰度和整体性重构效果较好,具有更高的准确性和鲁棒性,充分证明了本文算法在3D重建任务中的有效性。

为验证本文算法的收敛性,将本文算法在ScanNet数据集的四个场景(ScanNet 0050、ScanNet 0084、ScanNet 0580、ScanNet 0616)开展实验。结果表明,本文算法约30轮迭代后开始收敛,并在综合指标上有所提升,验证了方法的鲁棒性和高效性。通过对比四个场景的 $I_{F-score}$ 变化趋势,进一步证明了该方法在不同场景下均具有稳定的收敛特性和良好的鲁棒性。

2.5 消融实验

以VoISDF为基准,在模型的不同阶段添加模块,通过选定的评价指标分析添加相应模块后模型的性能,结果见表2。

表2 消融实验结果
Tab.2 results of ablation experiments

模块引入 位置	引入模块 名称	评价指标				
		I_{ACC}	I_{Comp}	I_{Prec}	I_{Recal}	$I_{F-score}$
初始采样	VoISDF	0.414	0.120	0.321	0.394	0.346
采样优化	深度引导采样	0.089	0.086	0.564	0.514	0.538
特征编码	点云编码模块	0.067	0.071	0.702	0.656	0.669
网络输出	残差模块	0.059	0.065	0.699	0.664	0.672

(1) 在采样过程中,初始阶段采用VoISDF方法,引入点云选取模块。在射线的[1,2]内均匀选取100个点作为初始采样点。为了提升采样在关键深度区域的密度,在预估深度值 d 的范围 $[d-0.01, d+0.01]$ 内均匀采样25个点云。最后将两阶段的采样点集成在一起,获得最终点云集合。实验结果表明,引入该模块后,准确性由0.414降至0.089,完整性由0.120降至0.086,精度由0.321提升至0.564,召回率由0.394提升至0.514, $I_{F-score}$ 提升0.192,达到0.538,表明该模块有效提升了重建质量和关键区域还原能力。

(2) 在特征编码阶段,进一步引入点云编码模块,采用多分辨率哈希编码对采样点云进行表达并输入网络^[21]。该模块进一步优化了各项指标:准确性从0.089降至0.067,完整性从0.086降至0.071,精度提升至0.702,召回率提升至0.656, $I_{F-score}$ 提升至0.669,显著增强了点云表达能力和网络学习效果。

(3) 在网络输出阶段引入残差模块,以增强特征流动与梯度传播。实验结果显示,该模块在已有基础上继续优化整体性能:准确性从0.067降至0.059,完整性从0.071降至0.065,精度略降至0.699,但召回率提升至0.664,最终 $I_{F-score}$ 升至0.672,进一步提升了网络的综合重建表现。

3 结论

为缓解室内三维重建中弱纹理区域鲁棒性不足以及重建效果不佳的问题,本文提出一个新型

室内三维重建网络 Manhattan-HashRecon。

采用深度引导采样对点云进行筛选,通过多分辨率哈希编码模块对点云进行编码,得到不同分辨率上的索引值,提高点云处理和分析的速度和准确性。在MLP中引入残差模块,减轻梯度消失问题,提高网络的性能和收敛速度。

实验结果表明,所提出的方法在多个评价指标上表现出较强的竞争力和鲁棒性。

参考文献(References):

- [1] FURUKAWA Y, PONCE J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376.
- [2] SHOTTON J, FITZGIBBON A, COOK M, et al. Real-time human pose recognition in parts from single depth images[M]//CIPOLLA R, BATTIATO S, FARINELLA G. Machine Learning for Computer Vision. Berlin: Springer Berlin Heidelberg, 2013: 119-135.
- [3] IZADI S, KIM D, HILLIGES O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera[C]//The 24th Annual ACM Symposium on User Interface Software and Technology, October 16-19, 2011, Santa Barbara, California, USA. New York: ACM, 2011: 559-568.
- [4] SCHÖNBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 4104-4113.
- [5] SITZMANN V, THIES J, HEIDE F, et al. DeepVoxels: learning persistent 3D feature embeddings[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 2432-2441.
- [6] OECHSLE M, PENG S Y, GEIGER A. UNISURF: unifying neural implicit surfaces and radiance fields for multi-view reconstruction [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 5569-5579.
- [7] MARTIN-BRUALLA R, RADWAN N, SAJJADI M S M, et al. NeRF in the wild: neural radiance fields for unconstrained photo collections[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 7206-7215.
- [8] 谢浩哲. 多源多视的三维场景和物体重建[D]. 哈尔滨: 哈尔滨工业大学, 2021: 29-30.
- [9] WU H Y, GRAIKOS A, SAMARAS D. S-VolSDF: sparse multi-view stereo regularization of neural implicit surfaces[C]//2023 IEEE/CVF International Conference on Computer Vision, October 1-6, 2023, Paris, France. IEEE, 2023: 3533-3545.
- [10] GAO Y M, CAO Y P, SHAN Y. SurfelNeRF: neural surfel radiance fields for online photorealistic reconstruction of indoor scenes[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 108-118.
- [11] 何鸿添, 陈晗, 刘洋, 等. 面向多模态交互式融合与渐进式优化的三维视觉理解[J]. 计算机应用研究, 2024, 41(5): 1554-1561.
HE Hongtian, CHEN Han, LIU Yang, et al. 3D visual understanding oriented towards multimodal interactive fusion and progressive refinement[J]. Application Research of Computers, 2024, 41(5): 1554-1561.
- [12] LU C S, YIN F K, CHEN X, et al. A large-scale outdoor multimodal dataset and benchmark for novel view synthesis and implicit scene reconstruction[C]//2023 IEEE/CVF International Conference on Computer Vision, October 1-6, 2023, Paris, France. IEEE, 2023: 7523-7533.
- [13] SUN J K, JIAO H, LI G Y, et al. 3DGStream: on-the-fly training of 3D Gaussians for efficient streaming of photo-realistic free-viewpoint videos[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 20675-20685.
- [14] MESCHEDER L, OECHSLE M, NIEMEYER M, et al. Occupancy networks: Learning 3d reconstruction in function space [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 4455-4465.
- [15] CARLBOM I, CHAKRAVARTY I, VANDERSCHEL D. A hierarchical data structure for representing the spatial decomposition of 3-D objects[J]. IEEE Computer Graphics and Applications, 1985, 5(4): 24-31.
- [16] DENG Z, XIAO H Y, LANG Y N, et al. Multi-scale hash encoding based neural geometry representation[J]. Computational Visual Media, 2024, 10(3): 453-470.
- [17] SOMMER C, SANG L, SCHUBERT D, et al. Gradient-SDF: a semi-implicit surface representation for 3D reconstruction[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 6270-6279.
- [18] FRIDOVICH-KEIL S, YU A, TANIĆ M, et al. Plenoxels: radiance fields without neural networks[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 5491-5500.
- [19] WALKER T, MARIOTTI O, VAXMAN A, et al. Spatially-adaptive hash encodings for neural surface reconstruction[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision, February 26-March 6, 2025, Tucson, AZ, USA. IEEE, 2025: 2963-2972.
- [20] DAI A, CHANG A X, SAVVA M, et al. ScanNet: richly-annotated 3D reconstructions of indoor scenes[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 2432-2443.
- [21] ZHU L W, ZHANG Y, PAN Z Q, et al. Binary and multi-class learning based low complexity optimization for HEVC encoding [J]. IEEE Transactions on Broadcasting, 2017, 63(3): 547-561.