深度学习的多视角三维重建技术综述

王文举[1],唐 邦[1+],顾泽骅[1],王 森[2]
1.上海理工大学,上海 200093
2.上海宝信软件股份有限公司,上海 200093
+ 通信作者 E-mail:223332932@st.usst.edu.cn

摘 要:为解决经典的多视角三维重建方法难以重建复杂物体、重建效果不佳以及在高分辨率上的扩展等问题,深度学习方法被引入用以重建具有更高精度的三维模型。系统地总结归纳、分析和比较了使用深度学习方法的多视角三维重建算法,并按照显式几何和隐式几何两种几何表示方式对近几年的多视角三维重建算法进行了分类与梳理。重点介绍了目前具有较高重建精度的将隐式函数以及体渲染相结合的神经隐式三维重建算法,并分别定量、定性分析了该类部分算法在数据集上的结果;另外列举了常用数据集和评价指标,并对未来的研究趋势和发展方向进行了展望。

关键词:深度学习;三维重建;神经隐式表示;体渲染
文献标志码:A **中图分类号:**TP391.41 doi:10.3778/j.issn.1002-8331.2405-0328

Overview of Multi-View 3D Reconstruction Techniques in Deep Learning

WANG Wenju[1], TANG Bang[1+], GU Zehua[1], WANG Sen[2]
1.University of Shanghai for Science and Technology, Shanghai 200093, China
2.Shanghai Baosight Software Co., Ltd., Shanghai 200093, China

Abstract: In order to solve the problems that classic multi-view 3D reconstruction methods are difficult to reconstruct complex objects and have poor reconstruction results, and to extend to high resolution, deep learning methods are introduced to reconstruct 3D models with higher accuracy. Thus multi-view 3D reconstruction algorithm using deep learning methods are systematically summarized, analyzed and compared, and the multi-view 3D reconstruction algorithms in recent years are classified and sorted out according to explicit geometry and implicit geometry representations. Neural implicit 3D reconstruction algorithms that combines implicit functions and volume rendering are mainly introduced, which currently have a high accuracy in reconstruction results, and the quantitative and qualitative analyses are conducted on some of these algorithms. In addition, commonly used datasets and evaluation indicators are listed, and the future research trends and development directions are discussed.

Key words: deep learning; 3D reconstruction; implicit neural representation; volume rendering

当前虚拟现实[1]、游戏开发[2],以及可视化医学成像[3]、文物数字化修复[4]等诸多行业领域都需要大量具有真实感的三维模型。这些模型主要依赖于手工制作或三维扫描仪精确扫描而获得。常规的手工制作需要耗费建模师大量的时间与精力从而影响到实际开发的工作效率和项目成本;三维扫描仪虽然能重建出高精度的三维模型,但存在采集仪器造价昂贵难以大量使用、操作流程复杂等问题。为实现低成本、高效的真实三维模型的制作,多视角三维重建技术应运而生。该技术能够根据单一物体或场景连续拍摄的多视角图像,使用机器学习方法重建出三维模型。当前多视角三维重建技术在实际应用中存在诸多问题与挑战:三维场景的复杂性会导致重建精度难以提升;对于动态场景或是光照、天气变化的场景的数据采集,以及采集数据中的不可见区域、自遮挡等情况,往往造成重建效果不稳定;三维重建过程需要多次人工介入,将耗时耗力。经典的三维重建方法如多视角立体视觉(multi-view stereo,MVS)方法[5-6]、体素重建方法[7-8]等难以解决这些问题。深度学习

基金项目:上海市自然科学基金面上项目(Z-2019-309-007)。
作者简介:王文举(1979—),男,博士,教授,CCF会员,研究方向为计算机视觉、深度学习、三维点云和计算光谱成像;唐邦(2000—),男,硕士研究生,研究方向为深度学习和三维重建。
收稿日期:2024-05-23 **修回日期:**2024-09-26 **文章编号:**1002-8331(2025)06-0022-14

方法由于其灵活性、可扩展性等特点能提供更多的解决思路：如通过超分辨率重建、生成对抗网络、深度融合等方式增加重建精度；通过数据增强、不确定性建模等方式可以提升算法鲁棒性；由于其“经验性”的特点，往往还可以端到端完成训练过程，从而减少手工调节和预处理的需求。尽管近年来有许多将经典算法与深度学习结合的优秀方法能一定程度进行优化，但仍然存在如累积误差等问题。目前随着深度学习以及神经体渲染（neural volume rendering，NVR）技术[9]的发展，这一类基于深度学习的多视角三维重建技术获得了高质量的重建结果，成为研究热点。

对于三维重建方法，按照相机的类型，可以分为使用事件相机拍摄的事件流作为输入的三维重建方法，以及使用标准相机拍摄的图像作为输入的三维重建方法，如图1所示。

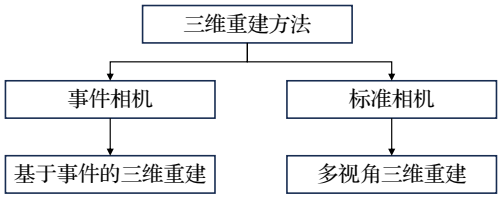

图1　三维重建方法分类

Fig.1　Classification of 3D reconstruction methods

事件相机是一种不同于标准相机的仿生传感器，它输出的并不是直接的像素值，而是异步地测量每个像素的变化，并输出编码这些变化的时间、位置和符号的事件流[10]。从事件相机的定义来看，事件就是事件相机输出的像素亮度变化情况。当场景中的物体运动或者光照改变从而导致大量的像素发生改变时，事件相机就会产生一系列事件，并以事件流的方式输出。这些事件具有时间戳、像素坐标和极性三个要素，分别表示事件发生的时间、发生改变的像素、像素亮度增加或者减少。图2展示了理想情况下标准相机和事件相机的区别，图中左侧表示一个圆盘，其上有一个黑点，右侧表示一段时间内圆盘的运动情况与两种相机的拍摄结果。当圆盘缓慢旋转时，传统相机输出明晰的多幅单帧图像，但由于每两幅相邻单帧图像之间存在时间间隔，会有一定

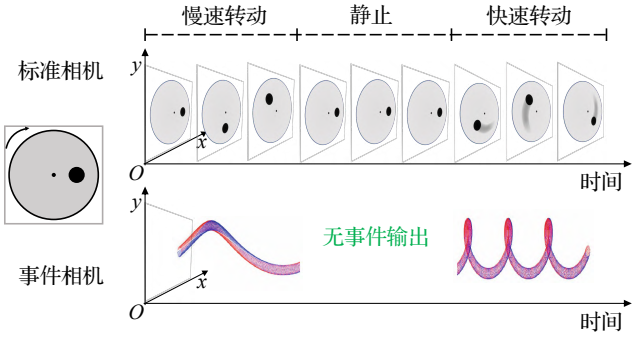

图2　标准相机与事件相机的对比

Fig.2　Comparison of standard cameras and event cameras

的延迟，而事件相机则输出一段连续的事件流；当圆盘静止时，事件相机没有事件输出，而传统相机则仍然产生图像，可能导致数据冗余；当圆盘快速旋转时，传统相机会出现运动模糊，可能导致数据错误，而事件相机则输出一段正常响应的连续事件流。事件相机具有高时间分辨率、高动态范围、低功耗和高像素带宽等特性，使其在机器人和计算机视觉领域具有巨大的潜力。

事件相机已经广泛应用于各个领域，如帧插值、语义分割和SLAM等[11]。但由于事件相机的输出与标准相机大相径庭，目前在三维重建领域的应用尚未得到充分的研究。基于事件的三维重建方法往往需要通过专门的算法去处理事件相机采集的事件数据，提取出与三维重建相关的信息，再利用计算机视觉和机器学习算法，构建出物体的三维模型。如Chen等人[12]使用3D ResNet作为Encoder，仅仅依靠单个事件相机就能进行密集3D体素重建，但作为该方向的早期尝试该方法存在计算资源受限和重建结果不佳等局限；Rudnev等人[13]首次尝试将事件流与神经辐射场结合进行密集重建，有效提升了过度曝光、低光照等情况下的鲁棒性；Li等人[14]提出E2DSNeRF，使用高动态范围和低延迟的事件相机代替常规的RGB-D相机作为NeRF的输入，其结果提高了稀疏视图的渲染质量，消除了动态输入造成的模糊和重影。由于事件流具有时间维度的信息，能一定程度上模拟物理效果，因此Wang等人[15]利用事件流数据来施加物理约束以增强NeRF的训练。该方法不仅加快了训练过程还有助于局部几何表达。但其性能并不稳定，受提取分支结果的影响。

基于事件的三维重建方法虽然在动态场景的三维重建具有一定的优势，但由于缺少相关研究存在诸多问题，因此目前主流的三维重建方法仍然使用标准相机输出的图像作为算法的输入。而其中又以多视角三维重建方法为主流的研究方向。多视角三维重建的方法主要分为经典方法和深度学习方法，如图3所示，其中3D Gaussian splatting（3DGS）是近年的新兴方向。

经典方法主要有基于图像匹配的三维重建和基于体素三维重建两种方式。基于图像匹配三维重建方法主要是指寻找空间中具有图像一致性的点，对场景进行立体匹配完成稠密重建，最后使用筛选泊松重建[16]等方法进行表面重建。典型的算法有基于补丁的MVS算法[5]、非结构化MVS的像素级视图选择方法[6]。图像匹配的三维重建方法难以处理具有自遮挡等复杂的结构。而基于体素的三维重建方法通过从多视角图像中估计体素网格中的占用率和颜色，利用空间雕刻法或体素着色法去逼近满足光度一致性的三维模型，因此能规避图像匹配的一些困难。该类算法包括空间雕刻的概率框架[7]、基于体素着色的逼真场景重建算法[8]等。基于体素的三维重建方法直观但是可达到的体素分辨率

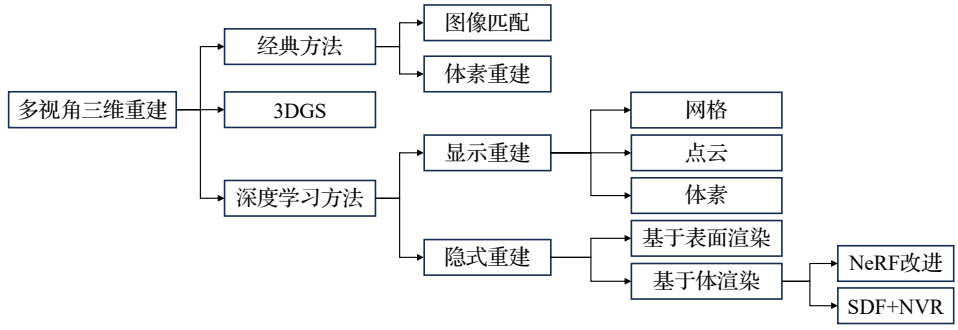

图3　多视角三维重建方法分类

Fig.3　Classification of multi-view 3D reconstruction methods

有限,并且存在无法对凹面建模等问题。经过多年的发展经典的三维重建方法已经趋于成熟,被广泛应用于实际工程中去,如Ebner等人[17]将MVS方法应用到虚拟与增强现实中去。但经典方法普遍存在难以重建复杂结构、处理分辨率低等问题。近年来,一些经典方法与深度学习相结合的方法能解决复杂结构和自遮挡的问题,如MVSNet[18]利用卷积网络进行特征匹配,扩展到多视图深度估计并重建三维点云。但该类方法对于不遵循Lambertian反射的场景,或是固定光照下的薄、细结构的重建仍然存在极大困难。

3D Gaussian splatting[19]是一种基于计算机图形学理论基础的渲染方法。该方法使用3D高斯分布准确地表示场景,并通过一种快速的可视性感知渲染算法实现了实时渲染。该类方法也能应用于三维重建领域,如SuGaR[20]提出了一个将3D高斯函数与三维场景表面对齐的方法,能够快速且精准地提取三维网格。3DGS作为一类新兴的方法,虽然能做到高质量的实时渲染,但是对于镜面、表面不明显(如云雾)等情况的三维重建仍然存在困难。

基于深度学习的多视角三维重建方法是指通过输入一组图像,通过深度学习网络直接或者间接地重建出图像中对的三维几何表示。相比于经典方法,该类型的深度学习方法具有更强的可扩展性、更简洁的技术路线、更逼真的视觉效果等优势,如Prokopetc等人[21]以及Lu等人[22]分别通过介绍从经典方法到深度学习方法在医学混合现实以及土木工程中的实际应用,来说明基于深度学习的多视角三维重建方法具有显著的优势,成为了当前的主流技术。

综上所述,三维重建的技术体系十分庞大,在近二十年间获得了广泛发展。2020年之前是多视角三维重建技术的早期探索阶段:如2010尝试对MVS方法进行改进的基于补丁的MVS算法[5];2018年的MVSNet[18]尝试在经典方法的基础上进行深度学习的改进;2016年的3D-R2N2[23]算法、2018年的Pixel2Mesh[24]算法、2019年的Pix2Vox[25]算法均尝试将深度学习与显式几何相结合。2021年,NeRF[26]将隐式几何表示与体渲染结合,席卷计算机视觉领域。同年的NeuS[27]与VolSDF[28]将神经体渲染技术应用于多视角三维重建领域,获得广泛关注。从2021—2023年,大量基于NeuS与VolSDF的多视角三维重建研究发表,典型的算法诸如Geo-NeuS[29]、NeuS2[30]、Neuralangelo[31]等。2023年诸多新兴的三维重建技术发表,如2023年首次尝试将事件相机与三维重建相结合的EventNeRF[13],2023下半年横空出世的3DGS[19],使用计算机图形学技术实现了高质量的实时渲染。2024年诞生了大量基于3DGS的多视角三维重建算法,其中以SuGaR[20]为典型代表。显然基于深度学习的三维重建技术是当前的研究主流。从中选取的典型方法按时间线绘制的发展脉络如图4所示。

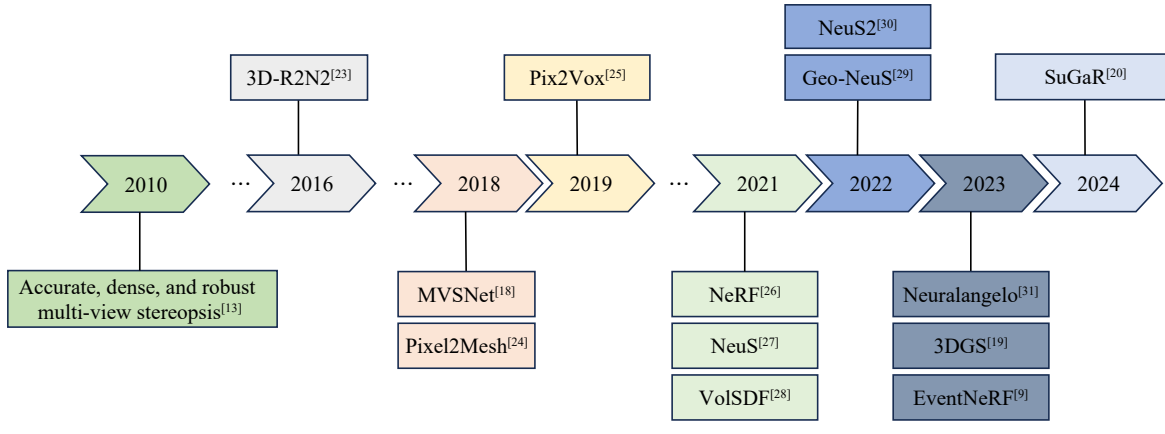

图4　典型算法时间线发展脉络

Fig.4　Timeline development of typical algorithms

文章后续将按照分类介绍基于深度学习的多视角三维重建方法,重点分析近年来提出的神经隐式三维重建方法,并提出面对的挑战以及对未来的展望。

1 深度学习的多视角图像三维重建方法

三维几何表示分为显示表示和隐式表示两种。显式表示是指通过体素、点云等方式,真正把物体上的点都表示出来。这种方式表示的物体具有明确的面,但难以判断点与物体的位置关系。隐式表示是指不表达点的具体位置,而通过函数关系表示点与点的关系。隐式几何方便判断点与物体的位置关系,并且存储和查询都很方便。按照三维几何的表示形式,基于深度学习的多视角三维重建方法可以分为显示重建方法和隐式重建方法。

1.1 显式三维重建

显式三维重建是指以显示几何为输出直接通过深度学习方法对场景或物体进行三维重建的方法。物体的显示几何表示方法有体素网格、点云、三角网格:体素网格(grid),其中体素是“体积元素”,是三维空间分割的最小单位;点云(point cloud)由大量空间的点组成,每个点具有三维坐标,可以包含颜色、强度、法线等附加的属性;三角网格(Mesh)由顶点、边和面构成,是最常用的三维模型表示方法。据此显式三维重建可以分为直接重建体素网格(grid)的方法、直接重建点云的方法、直接重建三角网格的方法。

直接重建体素网格(grid)的方法往往分三个步骤:首先单独对每幅图像进行2D特征提取,然后联合进行3D特征融合得到粗体素,最后通过一定的精细化方案得到最终的场景或物体的体素网格。如3D-R2N2[23]使用循环神经网络顺序地融合从输入图像中提取的多个特征图,并不断细化提取的体素网格;LSM[32]使用可微分的反投影操作结合3D CNN网络获得特征融合网格,并最终解码为体素网格;Pix2Vox[25]设计的编解码器以及上下文感知融合模块能从每个输入图像中生成粗糙的体素网格,并优化融合得到最终的3D体素网格;Pix2Vox++[33]由Pix2Vox改进而来,核心在于用一个多尺度上下文感知融合模块替换Pix2Vox中的上下文感知融合模块。

直接重建点云的方法与直接重建体素的方法类似,分别使用卷积生成器预测多个视角的3D结构并通过融合点云的方法得到最终的场景或物体的点云。如PointOutNet [34]使用多个卷积层、ReLU层和全连接网络组成点集预测网络来重建点云;密集三维物体重建的高效点云生成学习方法[35]使用2D卷积结构生成器来预测不同视角图像对应的3D点云,并联合伪渲染器生成深度图优化网络。

直接重建三角网格的方法类似于经典方法中的空间雕刻法,从一个椭球形网格开始逐步增加三角网格中的顶点以变形生成场景或物体对应的三角网格。神经三维网格渲染器[36]提出一种用于光栅化的梯度,可以结合网格生成器和神经渲染器生成三维网格模型;Pixel2Mesh[24]的网络由图像特征网络和级联网格变形网络组成,能根据单幅RGB图像逐步将椭球网格变形为对应的3D模型;基于Pixel2Mesh改进的Pixel2Mesh++[37]方法,提出了multi-view deformation网络,将Pixel2Mesh网络的输出网格作为输入并进一步细化该网格。

显式三维重建方法的输出结果能直接应用于大多数图形引擎,但该类方法受到分辨率的限制,难以做到高精度的三维重建,实际应用较少。对基于深度学习的各种显式三维重建方法的比较与分析见表1。

1.2 神经隐式三维重建

神经隐式三维重建是指通过学习的方法使用连续的数学函数来表示三维场景。相比于显示三维重建方法,由于使用连续数学函数,神经隐式三维重建方法能够在高分辨率下进行三维重建。神经隐式表示(implicit neural representation,INR)主要包括占用场、有符号距离函数等隐式函数。占用场通过将空间均匀切割为多个小方块,类似于像素值,使用小方块中的占用率来表示该空间位置是否有三维物体。如卷积占用网络[38]

表1 显式三维重建方法的比较

Table 1 Comparison of explicit 3D reconstruction methods

显式表示方法	名称	特点	缺点
体素网格(grid)	3D-R2N2[23]	使用RNN融合特征图,输入图像越多,体素越精细	普遍在分辨率上具有较大的限制,难以做到高精度三维重建
	LSM[32]	可微分反投影操作与3D CNN相结合	
	Pix2Vox[25]	从单幅图像生成模型,再利用上下文感知模块进行融合,最后细化为精细模型	
	Pix2Vox++[33]	多尺度上下文感知融合模块	
点云	PointOutNet[34]	在MLP前加上卷积层和ReLU层	
	密集三维物体重建的高效点云生成学习方法[35]	使用卷积网络来生成点云,结合一种伪渲染方法来优化网络	
三角网格(Mesh)	神经三维网格渲染器[36]	提出离散栅格化的近似梯度,进而能够反向传播	
	Pixel2Mesh[24]	使用GCN来对应Mesh的顶点、边和特征向量	
	Pixel2Mesh++[37]	multi-view deformation网络能进一步细化Pixel2Mesh的输出网格	

等。有符号距离函数(sign distance function,SDF)只使用三个值来表示场景,即-1、0、1分别表示物体外部、物体表面、物体内部,有符号距离函数的零水平集即是目标物体的表面函数,例如DeepSDF[39]。由于SDF能够很明确地确定三维表面,因此SDF已经成为主要的隐式几何表示方法。

随着渲染方法的发展,隐式三维重建方法开始结合渲染方法用以优化网络结构,这种端到端的方法逐渐成为了多视角三维重建方法的主流。这种方法的主要过程,是将多视角图像参数输入进入网络,对该网络预测的三维结果进行图像渲染并与真实图像比较计算损失,用于反向传播优化网络,不断迭代最终输出三维网格模型。方法概述见图5所示。按照渲染的方法,神经隐式三维重建方法可以大致分为使用表面渲染的方法和使用体渲染的方法两类。

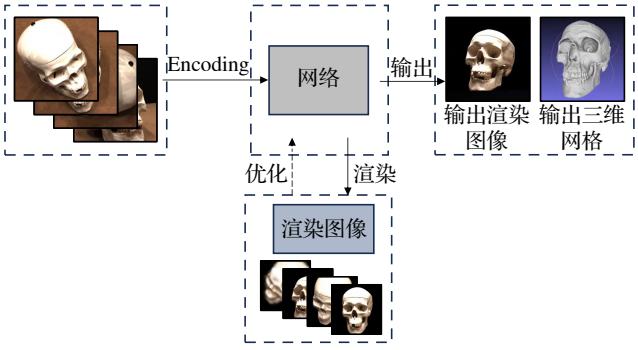

图5　隐式三维重建方法概述

Fig.5　Overview of implicit 3D reconstruction methods

1.2.1　表面渲染的神经隐式三维重建方法

以往的多视角三维重建方法大多使用基于表面的渲染方法,如DIST[40]提出一种完全可微分的球形跟踪算法用以弥补逆向渲染和基于隐式符号距离函数的深度学习算法之间的差距。IDR[41]使用类似于DeepSDF[39]的有符号距离函数的零水平集表示物体表面,并且结合DIST[40]中的可微分球形跟踪算法提出IDR前向模型用以计算几何表面的渲染颜色,实现了自监督的多视角三维重建。神经光场渲染[42]使用与IDR类似的框架来表示3D对象的形状和外观,但其网络架构建立在正弦表示网络[43]上,在相同数量的可学习参数内其表示明显具有更高的复杂性。RegSDF[44]提出两个几何正则化,结合适当点云监督能够进行高质量重建。表面渲染方法仅适用光线与场景几何交点的颜色来渲染,可能导致梯度只会反向传递到交点附近的局部区域,难以重建具有严重自遮挡或有深度突变的复杂物体。

1.2.2　体渲染的神经隐式三维重建方法

体渲染使用沿射线的所有采样点来渲染图像,能处理深度突变同时还能合成高质量图像。当NeRF[26]在新视角合成领域开始使用多层感知机来表示神经辐射场,越来越多的三维重建方法开始尝试多层感知机与体渲染相结合的方式。

(1)基于NeRF的三维重建方法

NeRF虽然是一种应用于新视角合成领域的算法,但能够表示真实的3D场景。NeRF方法首先通过Raymarching算法从每个像素计算出位置信息和视角信息,并通过位置编码分别输出高维向量用于MLP网络的训练。MLP输出场景几何密度和颜色值用于体渲染从而得到像素的渲染颜色。一个视角的所有像素的渲染颜色构成一幅新的渲染图像并与真实图像对比计算Loss,从而反向传播用于网络对场景几何和颜色的学习。经过多轮迭代,即可通过体渲染方式渲染出新视角的图像。其方法概述如图6所示[26]。

一些基于NeRF的方法使用体素网格特征来快速训练3D属性,如Plenoxels[45]将场景表示为带有球面谐波的稀疏3D网格,Direct voxel grid optimization[46]组合密度体素grid和特征体素grid,Instant-NGP[47]采用基于哈希搜索的多分辨率编码方式替代MLP。尽管这些方法在训练效率和渲染质量等方面取得了较好的效果,但由于几何表示方式缺乏对几何表面的约束,因此难以提取高质量的三维模型表面。

(2)SDF与NVR方法结合的三维重建方法

继NeRF之后,NeuS[27]和VolSDF[28]尝试将SDF作为隐式表示方法,并引入神经体渲染(NVR),使得基于体渲染的多视角三维重建成为可能。

NeuS利用Raymarching算法对图像中的像素计算出位置信息和射线的单位方向向量,在经过位置编码之后分别输入SDFNetwork(MLP)预测SDF值、ColorNetwork(MLP)预测颜色值。这些网络的预测值将用于体渲染,得到的渲染图像与真实图像计算Loss用于反向传

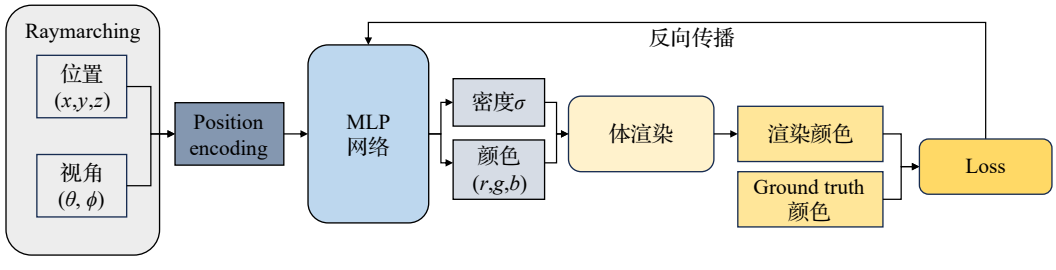

图6　NeRF方法概述

Fig.6　Overview of NeRF

播整个网络直至满足迭代停止条件。其方法概述如图7所示[27]。

NeRF方法中的MLP网络用于表示隐式函数，该隐式函数见公式(1)：

$$F:(x,d)\rightarrow(c,\sigma) \tag{1}$$

其中，F 用于近似连续5D场景表示，$x=(x,y,z)$ 表示观察位置，$d=(\theta,\phi)$ 表示观察方向，$c=(r,g,b)$ 表示发光颜色。可以看出NeRF的表示方式并没有对三维表面做出明确区分，这也是导致NeRF的三维重建结果表面不平整的原因之一。因此NeuS采用SDF作为隐式函数强化对表面的感知，通过多层感知器编码。该过程可以表示为公式(2)：

$$f(x):\mathbb{R}^3\rightarrow\mathbb{R}=SDF(x) \tag{2}$$

其中，$f(x)$ 表示将空间中的坐标 $x\in\mathbb{R}^3$ 映射到它距离物体的SDF值。物体的表面 S 可以通过其SDF值的零水平集表示，见公式(3)：

$$S=\{x\in\mathbb{R}^3|f(x)=0\} \tag{3}$$

上述公式表示，只需要找到SDF值为0的点即可确定三维物体的表面。因此相比于NeRF，NeuS能够很好地确定物体表面，并且重建出的物体表面更加平滑。为了从隐式函数中渲染图像，需要使用到体渲染方法。对于一幅图像的任意像素，从该像素发射一条射线，则该射线的累积预测颜色(渲染图像中对应于该像素的预测颜色值 $C(o,v)$)是：

$$C(o,v)=\int_0^{+\infty}\omega(t)c(p(t),v)\mathrm{d}t \tag{4}$$

其中，$\omega(t)$ 是点 $p(t)$ 的权重，$c(p(t),v)$ 是点 $p(t)$ 沿方向 v 的所预测出的观测颜色。NeuS[27]通过优化体渲染中的权重 $\omega(t)$，能够重建出具有复杂结构的物体。但NeuS[27]仍然没有解决需要大量计算资源的问题；另外该方法对于纹理较为复杂的区域难以重建出高质量结果。VolSDF[28]采用与NeuS类似的方法，将体积密度转换为使用累积密度分布函数的SDF表示后再进行体渲染，相比于NeRF能提高重建质量。但是VolSDF[28]在不可见区域、无纹理区域难以进行准确地重建。后续的算法大都以这两个算法为baseline进行改进。根据改进的方式可以为以下三类：

① 直接添加约束优化算法：如MonoSDF[48]通过预训练引入深度和法向几何信息约束，显著提升了重建效果，但却十分依赖于预训练结果。Geo-NeuS[29]引入对geometry的直接约束，可以复杂结构和光滑区域实现高质量的重建。与以往的方法类似，Geo-NeuS的效率仍然十分有限。D-NeuS[49]在NeuS的基础上，使用预训练的CNN来进行特征提取并约束特征的差值。该方法能够在有阴影的区域达到较好的重建效果，但由于需要额外的预训练模型，增加了计算时间和内存占用。Neural-Warp[50]引入经典MVS方法中的warping-based loss，在VolSDF的体渲染公式基础上按图像patch进行计算。Warping-based loss的引入有效提高了重建效果，但也带来了渲染开销巨大、数据处理麻烦等问题。S-VolSDF[51]使用MVS方法来正则化神经体渲染优化过程，并使用全局一致性约束提高MVS性能。该方法能够在稀疏视角下实现高质量重建，但由于引入MVS方法，对于透明表面的重建质量有所降低。直接添加约束的方法往往能在重建质量上带来较大提升，但也会明显导致效率的下降。

② 对baseline中隐式函数、采样策略等的改进：如HF-NeuS[52]将SDF分解为基函数和位移函数，该优化策略能重建出具有细密纹理的表面细节。但由于该方法需要优化一个额外的隐式函数，因此需要更多的计算资源。神经隐式无偏体渲染方法[53]提出一种具有更小体渲染偏差的转换方程并使用MVS中的sparse points来约束表面，有效提升了重建质量。这种方法在高分辨率场景下的重建效率较低，同时对于具有镜面或半透明材质的重建也有较大困难。球体神经隐式渲染方法[54]使用球体来对射线进行采样，大大提高了体渲染过程中采样的有效性。该方法仅仅是提出了更优的采样方案，并未解决baseline中诸如细节纹理、训练效率等问题。与直接添加约束的方法类似，该类方法在提高重建质量的同时也会带来效率降低等问题。

③ 使用不同的编码方法或grid等来加速训练：如NeuS2[30]使用Instant-NGP[47]中提出的Hash encoding方法来替代MLP加速训练。NeuS2在加速训练的基础上，能够实现动态场景每一帧的高质量重建，但是每帧之间的重建结果不能很好地对应。Neuralangelo[31]使用与NeuS2类似的Hash encoding方法来加速NeuS的训练，并使用数值梯度来代替分析梯度，能够使重建得到的表

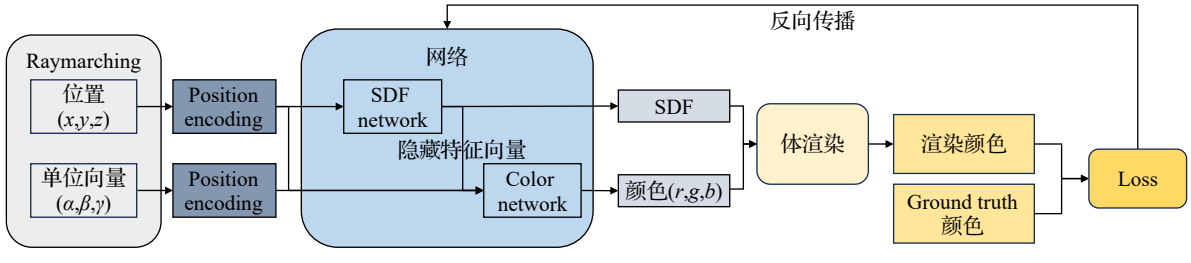

图7　NeuS方法概述

Fig.7　Overview of NeuS

面更加平滑。为了保证足够的细节采样,Neuralangelo的训练迭代次数过长,因此训练效率较低。NeuDA[55]使用立方体网格(grid)来加速训练,将grid的顶点映射之后进行位置编码以及插值得到采样点特征,这种方法能够重建精细的复杂结构。PermutoSDF[56]使用三角锥形grid来加速训练,并提出了两个新的几何约束,实现了快速训练和实时渲染。但该方法由于使用了较多的grid,对内存消耗巨大。

1.2.3 隐式三维重建方法的比较

从近来多视角三维重建领域的相关文献可以看出,SDF与NVR的组合无论是在重建精度还是对于复杂结构、自遮挡等问题的处理上,都具有良好的效果。但从总体来看,各个算法都难以同时在重建质量和效率上取得良好的结果,并且对于纹理细节的处理还存在诸多问题。基于深度学习的隐式三维重建方法的比较与分析见表2。从比较结果来看,要做到高质量的实时渲染,三维重建领域仍然任重道远。

1.2.4 神经隐式三维重建方法的应用

神经隐式三维重建凭借其逼真的视觉效果,被大量应用于AR/VR、医学、文化遗产、文物修复等领域。如Li等人[57]将神经隐式三维重建算法应用至虚拟/增强现实,能实现即时的3D重建,单个场景耗时不超过5 s;Croce等人[58]则提出将神经隐式三维重建算法与实际摄影测量相结合,应用于数字文化遗产领域;Ge等人[59]利用SDF与神经辐射场相结合的三维重建算法进行古建筑的重建,为测绘、可视化以及遗产保护等方面提供高质量的古建筑模型;Chen等人[60]将Instant-NGP[47]直接应用于医学膀胱镜三维重建,为泌尿系统疾病的观察和指导治疗提供了重要价值。

2 数据集和评价指标

2.1 数据集

多视角三维重建算法的重建质量受数据集中图像分辨率、三维场景的规模、光照、材质等的影响。图像质量高、场景数丰富的优质数据集能极大推动三维重建的发展,因此总结了目前常用的数据集,包括DTU[61]、BlendedMVS[62]、Tanks and Temples[63]、ETH3D[64]、LLFF[65]、ScanNet[66]等。各个数据集的图像采集分辨率、场景类型以及场景数量的比较见表3所示。

(1)DTU[61]是一个多视角立体(MVS)数据集,常被用来做MVS、NeRF方法的训练集。DTU包含受控实验室环境中的128个场景,在7种不同的照明条件下,由6轴工业机械臂获得。每个场景包含由49或64个精确的相机位置和参考结构光扫描,扫描得到的RGB图像的分辨率为1 200×1 600。数据集包括拍摄图像、相机参数、深度图、前景遮罩。

(2)BlendedMVS[62]是一个大规模的多视角立体数据集。该数据集使用三维重建算法从给定场景图像中恢复三维网格模型,然后将重建的三维网格模型渲染得到RGB图像和深度图。与DTU数据集使用固定机械臂获取相比,BlendedMVS数据集中的场景包含不同的相机轨迹,并且包括建筑、街景、雕塑等不同场景。

表2 隐式三维重建方法的比较

Table 2 Comparison of implicit 3D reconstruction methods

渲染方法	名称		特点	缺点
表面渲染	DIST[40]		完全可微分的球形跟踪算法	表面渲染仅关注光线与表面的交点,梯度会限制在交点局部,难以处理复杂结构
	IDR[41]		SDF的零水平集表示表面、IDR前向模型渲染颜色	
	神经光场渲染[42]		网络架构建立在正弦表示网络之上	
	RegSDF[44]		提出两个正则化方法,并使用点云监督	
体渲染	NeRF[26]	Plenoxels[45]	带有球面谐波的稀疏3D网格	缺乏对表面的约束,难以提取高质量三维模型
		Direct voxel grid optimization[46]	体素密度grid和特征体素grid相结合	
		Instant-NGP[47]	基于哈希搜索的多分辨率编码方式	
	SDF+NVR	NeuS[27]	一种基于体渲染改进的近似无偏公式	处理复杂纹理表面存在困难,并且难以同时保证高质量和高效
		VolSDF[28]	使用累积密度函数CDF表示SDF	
		MonoSDF[48]	引入深度和法向几何信息约束	
		Geo-NeuS[29]	引入对geometry的直接约束	
		D-NeuS[49]	预训练CNN进行特征提取	
		NeuralWarp[50]	引入warping-based loss	
		S-VolSDF[51]	带有全局一致性约束的MVS方法用于正则化	
		HF-NeuS[52]	将SDF分解为基函数和位移函数	
		神经隐式无偏渲染方法[53]	一种具有更小体渲染偏差的转换方程	
		球体神经隐式渲染方法[54]	球体采样方案	
		NeuS2[30]	Hash encoding加速训练	
		Neuralangelo[31]	Hash encoding加速,并用数值梯度代替分析梯度	
		NeuDA[55]	立方体网格(grid)加速训练	
		PermutoSDF[56]	三角锥形grid来加速训练	

表3 常用数据集的比较

Table 3 Comparison of common datasets

数据集		分辨率	场景	场景数	网址
DTU[61]		1 600×1 200	单个物体	124	https://roboimagedata.compute.dtu.dk/
BlendedMVS[62]	高分辨率	2 048×1 536	单个物体、室内外场景	113	https://github.com/YoYo000/BlendedMVS/
	低分辨率	768×576			
Tanks and Temples[63]		4 096×2 160	单个物体、室内场景	14	https://www.tanksandtemples.org/
ETH3D[64]		5 632×4 224	室内外场景	25	https://www.eth3d.net/
LLFF[65]		1 008×756	单个物体	24	https://bmild.github.io/llff/
ScanNet[66]		1 296×968	室内场景	1 500	http://www.scan-net.org/

(3)Tanks and Temples[63]是包含数个视频数据集的三维重建数据集,其3D数据通过高质量工业激光扫描仪捕获。该数据集总共14个场景,既包括“坦克”“火车”等单个物体,也包括“礼堂”“博物馆”等大型室内场景。

(4)ETH3D[64]是一个应用于多视角立体和三维重建领域的数据集,由多个子数据集组成,涵盖建筑物、自然景观、室内场景、工业和机械场景等各种室内外场景。每个子数据集中包含图像序列以及每个图像的相机参数和相机位姿。其中的图像通过数码单反相机以及具有不同视野的同步多相机装备拍摄,而真实的几何形状则通过高精度激光扫描仪获得。

(5)LLFF[65]包含24个场景,由手持手机摄像头拍摄,每张图像都面向中心对象。

(6)ScanNet[66]是大型RGB-D数据集,可应用于三维重建、语义分割等领域。该数据集拥有1 500个扫描场景,总计250万张图像,并且包含对应的相机参数、网格模型等。其中RGB图像分辨率为1 296×968,深度图分辨率为640×480。

2.2 评价指标

不同的数据集使用的评价指标不尽相同,其中使用最为广泛的是倒角距离和峰值信噪比两种。

(1)倒角距离(chamfer distance,CD),是用于测量两个点集之间相似性的度量方法。倒角距离非常适用于三维重建任务质量评估、计算效率、算法鲁棒性的评价。在重建质量评估方面,CD可以用于对比重建模型与真实模型之间的表面对齐程度,因此可以用来衡量三维重建模型的质量。在计算效率上,相比于其他度量方式CD的计算简单且高效,只需要进行距离计算求和。这一特性对于实时或大规模的三维重建任务尤为重要。从算法鲁棒性角度来看,CD对于有噪声的点集或模型时也能进行质量评估。因为它主要依赖于最短距离的计算,而并不会依赖于单一的精确匹配点或者点集的具体位置。

倒角距离在评估重建精度和人眼感官上存在一定的局限性。CD主要衡量点集之间的平均距离,关注形状的整体结构。因此,倒角距离不能很好地捕捉复杂的几何细节或局部微小形状,存在一定的精度限制。CD对细节感知能力有限,并且在对形状相似度量上没有考虑到纹理、光照等因素,导致其评价结果可能与人眼存在差异。

在三维重建中,CD计算点集和ground truth点云之间的平均最短点距离。它考虑了每个点的距离,对于每个点集中的每个点,CD都会在另一个点集中找到最近的点,并将距离的平方向上求和。计算公式如下:

$$CD(S_1,S_2)=\frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}\left\|x-y\right\|_2+\frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}\left\|y-x\right\|_2 \tag{5}$$

其中,S_1、S_2分别表示两组3D点集,第一项代表S_1中任意一点x到S_2的最小距离之和,第二项则表示S_2中任意一点y到S_1的最小距离之和。该距离越小,则说明重建效果越好。

(2)峰值信噪比PSNR(peak signal-to-noise ratio),是衡量图像质量的指标之一。PSNR也可以用于三维重建任务的评价,主要应用于评估合成视图。在多视角三维重建中,可以通过合成视图和真实视图之间的PSNR来间接评价重建模型的准确性。

当然PSNR也存在一定的局限性,如不能直接衡量三维模型的几何误差或拓扑结构的准确性,而偏重评估合成视图的质量;对于复杂的三维形状,PSNR可能不足以捕捉到细微的几何变化;与CD类似,PSNR的分数也无法和人眼看到的视觉品质完全一致。

PSNR是基于均方误差MSE定义的,其公式如下:

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \tag{6}$$

其中,I表示原始图像,K表示加入噪声后的图像,大小为$m\times n$。PSNR定义见公式(3):

$$PSNR=10\cdot\lg\frac{MAX_I^2}{MSE}=20\cdot\lg\frac{MAX_I}{\sqrt{MSE}} \tag{7}$$

其中,MAX_I表示图像像素的最大值。PSNR值越大,表示图像质量越好。由于基于体渲染的三维重建方法也能应用于新视角合成领域,因此当评价新视图的合成质量时使用PSNR作为评价指标。

3 算法性能比较与分析

对于三维重建算法的评价,可以使用倒角距离、

PSNR等评价指标来定量地比较算法的性能。其中倒角距离主要用于计算两个3D点集之间的差异,PSNR主要用于比较渲染图像与真实图像的相似性。但是这些评价指标存在不能完全准确体现人眼感知差异的问题。如PSNR主要计算图像间的像素差异,但没有考虑人类视觉系统的感知特性[67-69],因此可能不能准确反映图像的主观质量。另外,三维重建算法的重建结果往往可以直接在计算机软件中进行可视化观察。因此在定量分析(见3.1节)之外,还需要进行更为直观、直接观察重建结果差异的定性分析(见3.2节)。

3.1 定量分析

从表2对隐式重建算法的比较来看,基于体渲染的神经隐式三维重建方法解决了以往算法存在的诸多问题,并且发展与应用十分迅速,因此本文从表2中选择了NeRF[26]、NeuS[27]、VolSDF[28],以及基于NeuS和VolSDF改进的算法[29-31,50-52]等8篇文章从重建精度以及训练效率角度进行对比。如表4所示,使用DTU数据集(ScanID表示从数据集中选择的场景),不使用前景遮罩,使用倒角距离CD作为评价指标(倒角距离值越小重建精度越高,使用↓表示),并列出了部分算法的训练时间(runtime一行表示运行时间)。

从结果来看,NeRF虽然在场景55、106上的CD值分别为0.58、0.88的较低值,但是在每一个重建场景的倒角距离值都是最大值,平均倒角距离值高达1.49,甚至在场景63、97、110的CD值超过2.0。NeRF作为新视角合成领域的算法,关注的核心在于如何将场景渲染到屏幕上,其渲染结果虽然十分优秀,但没有对三维表面有明确的区分,因此导致重建精度较低。

最初尝试将SDF与NVR结合的NeuS和VolSDF平均倒角距离都近似于0.85,重建精度上相差不大。其中NeuS在场景114的CD值达到了0.35的优秀水平,但在场景83的CD值却高达1.48;VolSDF虽然大部分场景的CD值略低于NeuS,但整体表现相对稳定,没有出现较大的差异。这两个算法比较类似,都使用能很好表征三维表面的SDF作为隐式函数表示,并结合渲染质量高的体渲染进行三维重建,因此达到了较为优秀的重建结果。

NeuralWarp、HF-NeuS、Geo-NeuS等优化策略的方法在重建精度上均有提升,平均CD值都低于NeuS。Geo-NeuS是所有算法中CD值表现最优的算法,除场景83外其余场景CD值都低于1.0,平均CD值低至0.5的水平,并在场景37、40、69、83、97、105、106、110、114、118、122上都是最佳CD值;但同时其训练时间相比于NeuS增加了2倍,效率不佳。这是因为Geo-NeuS引入多项几何约束,虽然能很好地约束三维表面,但却增加了大量的计算成本。

NeuS2和Neuralangelo都尝试引入Instant-NGP中的Hash encoding来优化网络,二者的平均重建精度均不超过0.7。其中NeuS2虽然CD值相比于其他基于NeuS和VolSDF的优化策略算法没有明显优势,但其训练时间大幅缩减至分钟级别。NeuS2有效地将Hash encoding应用到了基于SDF的体渲染过程,因此起到了优秀的加速算法效果。

综合来看,Geo-NeuS的重建精度极佳,但同时效率也很低;NeuS2训练效率最佳,并在重建精度上也较为优秀。但值得注意的是,从数据来看同一个算法在不同

表4 基于体渲染的神经隐式三维重建方法比较

Table 4 Comparison of neural implicit 3D reconstruction methods based on volume rendering

方法	NeRF[26]	NeuS[27]	VolSDF[28]	NeuralWarp[50]	HF-NeuS[52]	Geo-NeuS[29]	NeuS2[30]	Neuralangelo[31]
Runtime	—	8 h	—	—	—	16 h	5 min	—
ScanID	CD↓	CD↓	CD↓	CD↓	CD↓	CD↓	CD↓	CD↓
24	1.90	1.00	1.14	0.49	0.76	0.38	0.56	**0.37**
37	1.60	1.37	1.26	0.71	1.32	**0.54**	0.76	0.72
40	1.85	0.93	0.81	0.38	0.70	**0.34**	0.49	0.35
55	0.58	0.43	0.49	0.38	0.39	0.36	0.37	**0.35**
63	2.28	1.10	1.25	**0.79**	1.06	0.80	0.92	0.87
65	1.27	0.65	0.70	0.81	0.63	0.45	0.71	**0.54**
69	1.47	0.57	0.72	0.82	0.63	**0.41**	0.76	0.53
83	1.67	1.48	1.29	1.20	1.15	**1.03**	1.22	1.29
97	2.05	1.09	1.18	1.06	1.12	**0.84**	1.08	0.97
105	1.07	0.83	0.70	0.68	0.80	**0.54**	0.63	0.73
106	0.88	0.52	0.66	0.66	0.52	**0.46**	0.59	0.47
110	2.53	1.20	1.08	0.74	1.22	**0.47**	0.89	0.74
114	1.06	0.35	0.42	0.41	0.33	**0.29**	0.40	0.32
118	1.15	0.49	0.61	0.63	0.49	**0.36**	0.48	0.41
122	0.96	0.54	0.55	0.51	0.50	**0.35**	0.55	0.43
Mean	1.49	0.84	0.86	0.68	0.77	**0.51**	0.70	0.61

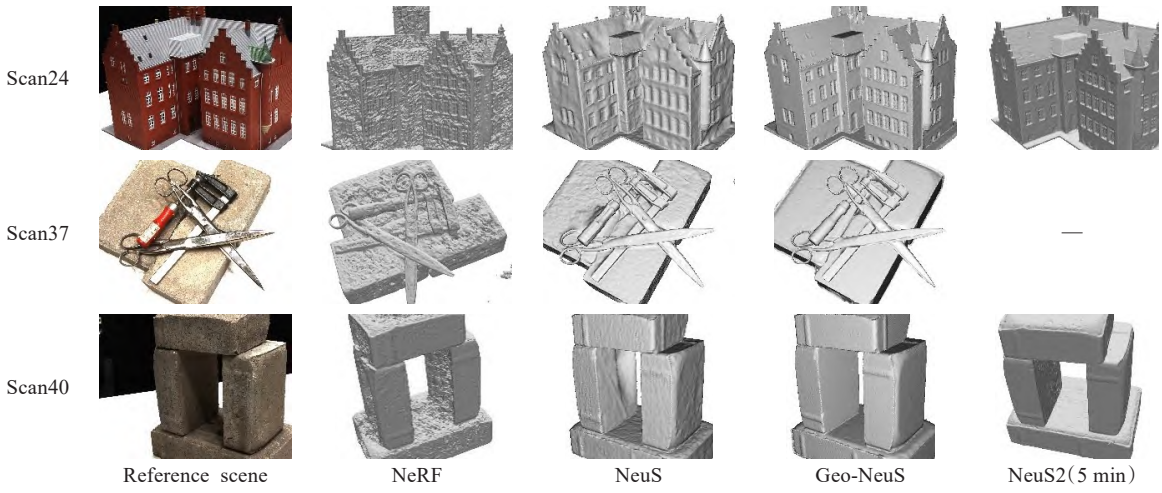

图8　DTU数据集上的定性分析

Fig.8　Qualitative evaluation on DTU dataset

场景的重建结果的CD值差异可能很大，如NeuS在DTU数据集第83号场景的CD值为1.48、在DTU数据集第114号的CD值则是0.35。显然算法在CD值上有较大差异，但人眼对于这些场景的观察结果差异性不大，这将不便于比较算法在不同场景中重建结果的稳定性。因此，为了强化同一算法或不同算法在不同场景中重建质量比较，需要引入定性分析比较。

3.2　定性分析

对于定性的比较，本文选择DTU、BlendedMVS数据集从三维网格角度来比较小物体的重建效果(均使用前景遮罩)，选择Tanks and Temples数据集从渲染图像的角度来比较大场景的重建效果。

其中NeRF[26]、NeuS[27]、Geo-NeuS[29]、NeuS2[30]四个算法在DTU数据集的三个场景上进行比较，如图8所示。对于NeRF，其重建结果不具有完整平滑的表面，并且可能出现与重建主体的非连通块，这与定量分析中对NeRF的描述表现一致。NeuS引入能明确区分三维表面的SDF，从图8第三列可以看出，其三个场景的重建结果相比于NeRF有了质的飞跃；但对于被遮挡区域的重建，可能出现不正确突起，如图9所示；而在光照/阴影影响的区域则可能出现凹陷，如图10所示，因此其鲁棒性较差。在定性分析中表现最优的Geo-NeuS，如图8第四列所示，其三个场景的重建结果有着更加光滑的表面，并且在各个场景下都保持极高的重建质量，极大程度解决了NeuS的鲁棒性问题，重建结果已经相当接近于手工制作的三维模型；注意图8中的最后一列，NeuS2呈现的是其中两个场景在训练5 min时的重建结果，其重建结果虽然相比于Geo-NeuS还有差距，但已经明显优于NeuS的重建结果。

NeRF[26]、NeuS[27]、Geo-NeuS[29]三个算法在BlendedMVS数据集上进行的比较分析，如图11所示。其中NeRF在两个数据集上都只能重建出大致形体，重建表面出现大

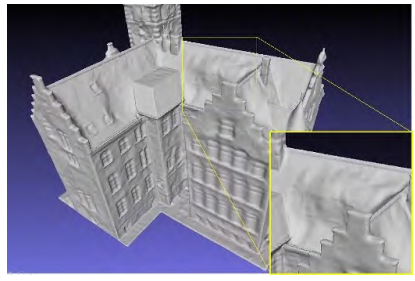

图9　NeuS的缺陷(DTU Scan24)

Fig.9　Defects of NeuS(DTU Scan24)

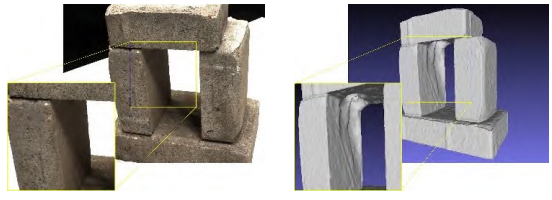

图10　NeuS的缺陷(DTU Scan40)

Fig.10　Defects of NeuS(DTU Scan40)

量空洞、凹陷，无法重建出平滑连续的三维表面；NeuS在两个数据集的五个场景都出现不平滑的表面、不正确的凸起和凹陷等问题；Geo-NeuS在这五个场景的重建质量则都是最优。虽然这三个算法在不同场景的重建质量存在差异，但总体来看这类算法在不同数据集的重建质量较为稳定。

另外对于针对大场景重建的Tanks and Temples数据集，本文选择NeuS[27]以及Neuralangelo[31]两个算法在三个场景上进行比较，如图12所示，其中第一列为对应的场景名称，第二列为渲染图像的真实参考图像，第三、四列分别为NeuS和Neuralangelo的渲染结果。可以看出，Neuralangelo的渲染精细度明显优于NeuS。

总体来看，体渲染的神经隐式三维重建方法的重建视觉效果已经非常优秀。但通过图8、图11、图12分析得出，该类方法对每个场景的重建都需要经过从头开始的一次训练过程和提取三维网格过程，每个场景的重建

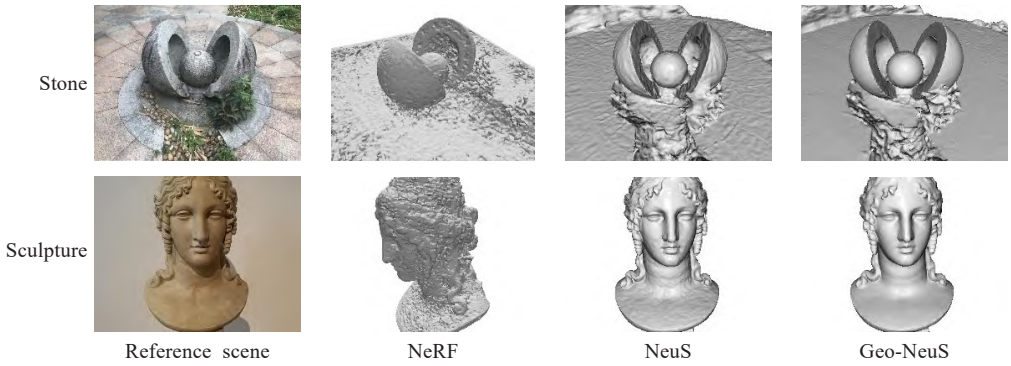

图11　BlendedMVS数据集上的定性分析

Fig.11　Qualitative evaluation on BlendedMVS dataset

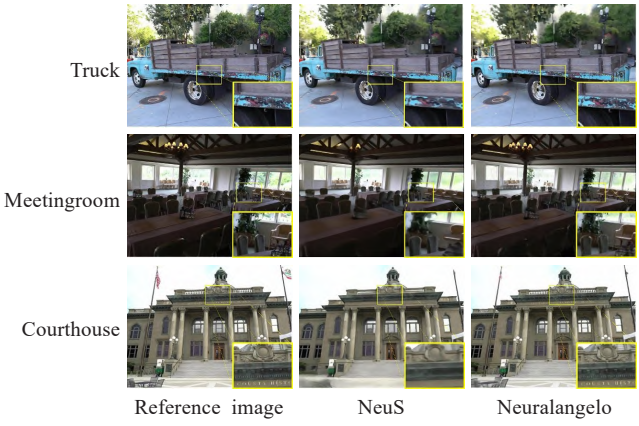

图12　Tanks and Temples数据集上的定性分析

Fig.12　Qualitative evaluation on Tanks and Temples dataset

都是独立的。因此,该类算法的泛化能力较弱,即时性成为了这类算法的一大发展趋势。而从NeuS2训练5 min的重建结果来看,这类算法在即时性角度具备进一步研究的价值。

4　面临的挑战与未来发展趋势

近年来三维重建领域的算法百花齐放,其中基于体渲染的多视角三维重建方法在重建精度上取得了良好的成果。该领域的方法从最初的理论推导验证探索,到目前已经逐渐开始向工程化、应用化转变。但是目前研究人员在探求更优的多视角三维重建解决方案时,仍然存在许多挑战。本文从未来的研究方向和发展趋势入手,提出以下几个值得深入探索的问题:

(1)大规模场景的数据集较少,有待进一步丰富。多视角三维重建领域对数据集要求较高,而目前应用于多视角三维重建的数据集相对较少,主要有以单个物体为主的DTU数据集、以室内外场景为主的ETH3D等。单个小物体的数据集可以通过现有方法[70]制作DTU格式或者LLFF格式的自建数据集,但大规模场景的数据集仍然存在困难。一方面大规模场景的数据采集需要耗费大量的人力物力,对采集设备的要求也很高;另一方面大场景的采集还涉及隐私等合法性的问题。一种解决方案是通过已有的新视角合成算法或者游戏引擎等渲染出大规模场景需要的图像数据。

(2)如何使三维重建技术适应大规模场景也是未来的研究热点。目前的三维重建技术对于小规模场景的已经能得到较好的重建质量,但训练效率仍然不够理想。以NeuS为例,其对于DTU数据集上单个场景的重建,数据量不到百万级别的同时运算量已经到达百亿级别。这是由于其网络主要采用多层感知器(MLP)实现,并且会对输入通过位置编码进行维度扩张,因此其运算量将会爆炸式地增加。如果将NeuS直接应用到大规模场景的数据集,其效率和重建结果将会很不理想,因此可以从计算效率的角度考虑对算法的网络框架进行调整。例如使用InstantNGP[47]中提出的Hash encoding来代替MLP的计算,可以有效规避在大规模场景下运算量的爆炸式增长。

(3)如何快速重建三维物体,稀疏视角图像的三维重建是重要的方向。目前主要的三维重建方法分为两类:一类是多视角三维重建,即利用多个视角的图像来进行单一物体或场景的三维重建;另一类是仅使用单幅图像进行三维重建的方法。多视角三维重建的优势在于,随着视图数量的增多,能获取更多的三维信息,并且能处理诸如遮挡、薄细结构等复杂情况,但相应的计算量很大;单视角三维重建能够根据单幅图像迅速重建出三维物体,但由于单幅图像蕴含的三维信息过少,往往难以重建复杂物体或者大规模场景。一个折中的方案是稀疏视角三维重建,即使用较少的图像进行三维重建,在保留足够重建精度的前提下,有效加快重建过程,并且可以通过将重建过程拆分为子问题的方法,使得重建过程更加稳定。

(4)为了实现高质量的实时渲染,需要平衡重建精度和训练效率并重。不同于其他图像任务,三维重建既需要考虑二维的图像特征,还需要考虑三维特征,同时基于体渲染的方法还需要实现高分辨率的图像渲染,因此训练开销巨大。从近几年的算法比较来看,为了实现更高质量的重建,需要引入更多的约束,这必然会导致训练效率上的降低,典型的算法如Geo-NeuS[29];而一些加速训练的算法能做到训练效率的显著提升,但在重

建精度上与针对于高质量重建的算法仍有差距,如NeuS2[30]。因此,高效率、高精度的多视角三维重建方法将是该领域长久的发展方向。

5 结束语

本文对基于深度学习的多视角三维重建技术方法进行了研究。本文重点介绍了将SDF与NVR相结合的基于体渲染的多视角隐式三维重建方法,总结了这类算法的最新研究进展,并进行了比较与分析。另外还整理了较为常用的三维重建数据集以及评价指标。最后,本文讨论了多视角三维重建技术的研究展望,并对可行的研究方向进行了分析。根据上述研究与分析,基于深度学习的多视角三维重建技术具有广阔的应用前景。本文所做的研究综述,期望能够为多视角三维重建的研究工作提供参考,促进多视角三维重建技术在理论与工程上的发展。

参考文献:

[1] KAMRAN-PISHHESARI A, MONIRI-MORAD A, SATTARVAND J. Applications of 3D reconstruction in virtual reality-based teleoperation: a review in the mining industry [J]. Technologies, 2024, 12(3): 40.

[2] HUANG T Y. Research on three-dimensional reconstruction [J]. Science and Technology of Engineering, Chemistry and Environmental Protection, 2024, 1(5): 1-4.

[3] ZI Y, WANG Q, GAO Z J, et al. Research on the application of deep learning in medical image segmentation and 3D reconstruction[J]. Academic Journal of Science and Technology, 2024, 10(2): 8-12.

[4] CALLET P, CALLET P. 3D reconstruction from 3D cultural heritage models[C]//Proceedings of the Roadmap in Digital Heritage Preservation on 3D Research Challenges in Cultural Heritage. New York: ACM, 2014: 135-142.

[5] FURUKAWA Y, PONCE J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376.

[6] SCHÖNBERGER J L, ZHENG E L, FRAHM J M, et al. Pixelwise view selection for unstructured multi-view stereo[C]// Proceedings of the 14th European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 501-518.

[7] BROADHURST A, DRUMMOND T W, CIPOLLA R. A probabilistic framework for space carving[C]//Proceedings of the Eighth IEEE International Conference on Computer Vision. Piscataway: IEEE, 2001: 388-393.

[8] SEITZ S M, DYER C R. Photorealistic scene reconstruction by voxel coloring[J]. International Journal of Computer Vision, 1999, 35(2): 151-173.

[9] DELLAERT F, YEN-CHEN L. Neural volume rendering: NeRF and beyond[J]. arXiv:2101.05204, 2021.

[10] GALLEGO G, DELBRÜCK T, ORCHARD G, et al. Event-based vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 154-180.

[11] KIM H, LEUTENEGGER S, DAVISON A J. Real-time 3D reconstruction and 6-DoF tracking with an event camera [C]//Proceedings of the 14th European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 349-364.

[12] CHEN H D, CHUNG V, TAN L, et al. Dense voxel 3D reconstruction using a monocular event camera[C]//Proceedings of the 2023 9th International Conference on Virtual Reality. Piscataway: IEEE, 2023: 30-35.

[13] RUDNEV V, ELGHARIB M, THEOBALT C, et al. EventNeRF: neural radiance fields from a single colour event camera[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 4992-5002.

[14] LI Q X, WANG Z, JIE L P, et al. Dynamic wind turbine blade 3D model reconstruction with event camera[C]//Proceedings of the UNified Conference of DAMAS, InCoME and TEPEN Conferences (UNIfied 2023). Cham: Springer Nature Switzerland, 2024: 863-875.

[15] WANG J X, HE J H, ZHANG Z Y, et al. Physical priors augmented event-based 3D reconstruction[J]. arXiv:2401.17121, 2024.

[16] KAZHDAN M, HOPPE H. Screened Poisson surface reconstruction[J]. ACM Transactions on Graphics, 2013, 32(3): 1-13.

[17] EBNER T, FELDMANN I, RENAULT S, et al. Multi-view reconstruction of dynamic real-world objects and their integration in augmented and virtual reality applications[J]. Journal of the Society for Information Display, 2017, 25(3): 151-157.

[18] YAO Y, LUO Z X, LI S W, et al. MVSNet: depth inference for unstructured multi-view stereo[C]//Proceedings of the 15th European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 785-801.

[19] KERBL B, KOPANAS G, LEIMKUEHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 1-14.

[20] GUÉDON A, LEPETIT V. SuGaR: surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 5354-5363.

[21] PROKOPETC K, DUPONT R. Towards dense 3D reconstruction for mixed reality in healthcare: classical multi-view stereo vs deep learning[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE, 2019: 2061-2069.

[22] LU Y J, WANG S, FAN S S, et al. Image-based 3D reconstruction for multi-scale civil and infrastructure projects: a

review from 2012 to 2022 with new perspective from deep learning methods[J]. Advanced Engineering Informatics, 2024, 59: 102268.

[23] CHOY C B, XU D F, GWAK J, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction [C]//Proceedings of the 14th European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 628-644.

[24] WANG N Y, ZHANG Y D, LI Z W, et al. Pixel2Mesh: generating 3D mesh models from single RGB images[C]//Proceedings of the 15th European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 55-71.

[25] XIE H Z, YAO H X, SUN X S, et al. Pix2Vox: context-aware 3D reconstruction from single and multi-view images [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 2690-2698.

[26] 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.

[27] WANG P, LIU L J, LIU Y, et al. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction[J]. arXiv:2106.10689, 2021.

[28] YARIV L, GU J T, KASTEN Y, et al. Volume rendering of neural implicit surfaces[C]//Advances in Neural Information Processing Systems, 2021: 4805-4815.

[29] FU Q, XU Q, ONG Y S, et al. Geo-NeuS: geometry-consistent neural implicit surfaces learning for multi-view reconstruction[C]//Advances in Neural Information Processing Systems, 2022: 3403-3416.

[30] WANG Y M, HAN Q, HABERMANN M, et al. NeuS2: fast learning of neural implicit surfaces for multi-view reconstruction[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 3272-3283.

[31] 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. Piscataway: IEEE, 2023: 8456-8465.

[32] KAR A, HÄNE C, MALIK J, et al. Learning a multi-view stereo machine[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 364-375.

[33] XIE H Z, YAO H X, ZHANG S P, et al. Pix2Vox++: multi-scale context-aware 3D object reconstruction from single and multiple images[J]. International Journal of Computer Vision, 2020, 128(12): 2919-2935.

[34] FAN H Q, SU H, GUIBAS L. A point set generation network for 3D object reconstruction from a single image[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2463-2471.

[35] LIN C H, KONG C, LUCEY S. Learning efficient point cloud generation for dense 3D object reconstruction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 7114-7121.

[36] KATO H, USHIKU Y, HARADA T. Neural 3D mesh renderer[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3907-3916.

[37] WEN C, ZHANG Y D, LI Z W, et al. Pixel2Mesh++: multi-view 3D mesh generation via deformation[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1042-1051.

[38] PENG S Y, NIEMEYER M, MESCHEDER L, et al. Convolutional occupancy networks[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 523-540.

[39] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: learning continuous signed distance functions for shape representation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 165-174.

[40] LIU S H, ZHANG Y D, PENG S Y, et al. DIST: rendering deep implicit signed distance function with differentiable sphere tracing[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 2016-2025.

[41] YARIV L, KASTEN Y, MORAN D, et al. Multiview neural surface reconstruction by disentangling geometry and appearance[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 2492-2502.

[42] KELLNHOFER P, JEBE L C, JONES A, et al. Neural lumigraph rendering[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 4285-4295.

[43] SITZMANN V, MARTEL J N P, BERGMAN A W, et al. Implicit neural representations with periodic activation functions[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 7462-7473.

[44] ZHANG J Y, YAO Y, LI S W, et al. Critical regularizations for neural surface reconstruction in the wild[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 6260-6269.

[45] FRIDOVICH-KEIL S, YU A, TANCIK M, et al. Plenoxels: radiance fields without neural networks[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 5491-5500.

[46] SUN C, SUN M, CHEN H T. Direct voxel grid optimization: super-fast convergence for radiance fields reconstruction[C]//Proceedings of the 2022 IEEE/CVF Conference on

Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 5449-5459.

[47] 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.

[48] YU Z, PENG S, NIEMEYER M, et al. MonoSDF: exploring monocular geometric cues for neural implicit surface reconstruction[C]//Advances in Neural Information Processing Systems, 2022: 25018-25032.

[49] CHEN D C, ZHANG P, FELDMANN I, et al. Recovering fine details for neural implicit surface reconstruction[C]// Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 4319-4328.

[50] DARMON F, BASCLE B, DEVAUX J C, et al. Improving neural implicit surfaces geometry with patch warping[C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 6250-6259.

[51] WU H Y, GRAIKOS A, SAMARAS D. S-VolSDF: sparse multi-view stereo regularization of neural implicit surfaces [C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 3533-3545.

[52] WANG Y, SKOROKHODOV I, WONKA P. HF-NeuS: improved surface reconstruction using high-frequency details[C]//Advances in Neural Information Processing Systems, 2022: 1966-1978.

[53] ZHANG Y Q, HU Z P, WU H Q, et al. Towards unbiased volume rendering of neural implicit surfaces with geometry priors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 4359-4368.

[54] DOGARU A, ARDELEAN A T, IGNATYEV S, et al. Sphere-guided training of neural implicit surfaces[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 20844-20853.

[55] CAI B W, HUANG J C, JIA R F, et al. NeuDA: neural deformable anchor for high-fidelity implicit surface reconstruction[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 8476-8485.

[56] ROSU R A, BEHNKE S. PermutoSDF: fast multi-view reconstruction with implicit surfaces using permutohedral lattices[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 8466-8475.

[57] LI S X, LI C J, ZHU W B, et al. Instant-3D: instant neural radiance field training towards on-device AR/VR 3D reconstruction[C]//Proceedings of the 50th Annual International Symposium on Computer Architecture. New York: ACM, 2023: 1-13.

[58] CROCE V, BILLI D, CAROTI G, et al. Comparative assessment of neural radiance fields and photogrammetry in digital heritage: impact of varying image conditions on 3D reconstruction[J]. Remote Sensing, 2024, 16(2): 301.

[59] GE Y W, GUO B X, ZHA P S, et al. 3D reconstruction of ancient buildings using UAV images and neural radiation field with depth supervision[J]. Remote Sensing, 2024, 16(3): 473.

[60] CHEN P C, GUNDERSON N M, LEWIS A, et al. Enabling rapid and high-quality 3D scene reconstruction in cystoscopy through neural radiance fields[C]//Proceedings of the Medical Imaging 2024: Image-Guided Procedures, Robotic Interventions, and Modeling, 2024: 56.

[61] JENSEN R, DAHL A, VOGIATZIS G, et al. Large scale multi-view stereopsis evaluation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 406-413.

[62] YAO Y, LUO Z X, LI S W, et al. BlendedMVS: a large-scale dataset for generalized multi-view stereo networks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 1787-1796.

[63] KNAPITSCH A, PARK J, ZHOU Q Y, et al. Tanks and Temples: benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics, 2017, 36(4): 1-13.

[64] SCHÖPS T, SCHÖNBERGER J L, GALLIANI S, et al. A multi-view stereo benchmark with high-resolution images and multi-camera videos[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2538-2547.

[65] MILDENHALL B, SRINIVASAN P P, ORTIZ-CAYON R, et al. Local light field fusion: practical view synthesis with prescriptive sampling guidelines[J]. ACM Transactions on Graphics, 2019, 38(4): 1-14.

[66] DAI A, CHANG A X, SAVVA M, et al. ScanNet: richly-annotated 3D reconstructions of indoor scenes[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2432-2443.

[67] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[68] HORE A, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]//Proceedings of the 2010 20th International Conference on Pattern Recognition. Piscataway: IEEE, 2010: 2366-2369.

[69] SARA U, AKTER M, UDDIN M S. Image quality assessment through FSIM, SSIM, MSE and PSNR: a comparative study[J]. Journal of Computer and Communications, 2019, 7(3): 8-18.

[70] SCHÖNBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4104-4113.