引用格式：朱方.3D场景表征—神经辐射场（NeRF）近期成果综述[J].中国传媒大学学报（自然科学版），2022，29(05)：64-77.
文章编号：1673-4793(2022)05-0064-14

3D场景表征—神经辐射场（NeRF）近期成果综述

朱方[1,2]
（1.中兴通讯微电子研究院，深圳 518057; 2.移动通讯与移动多媒体国家重点实验室，深圳 518055）

摘要：场景三维空间感知建模和基于先验的内容重现一直是信息技术围绕人类交互体验的重要努力方向，也是元宇宙和增强现实等数字和现实世界纽带构建技术的核心基础。当前，随着深度学习驱动的信息技术快速发展，特别是神经辐射场兴起，作为建模和内容重现的核心—场景表征技术得到了跨越式发展。本文首先总结了不同表征技术的应用背景和近期基于深度学习所取得的进展；其次，对神经辐射场近期的重要成果进行了梳理和阐述；然后，就神经辐射场的场景构建和交叉领域研究的分析，揭示了显性空间和语义挖掘对神经辐射场的重要价值；最后，结合近期显性空间和语义挖掘中的重要发展方向，单视图3D场景感知，面临的挑战和神经辐射场相关研究展现的裨益，揭示了基于神经辐射场对场景三维空间感知建模和基于先验的内容重现发展带来的机遇。

关键词：空间感知建模；内容重现；3D场景表征；神经辐射场；单视图3D场景感知
中图分类号：TP391　**文献标识码**：A

A review of 3D scene representation—NeRF: neural radiance fields

ZHU Fang[1,2]
(1.ZTE Microelectronics R&D Institute, Shenzhen 518057, China; 2.State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

Abstract: Three-dimensional scene modeling and prior-based content reproduction have long been cornerstones of human-centered information technology endeavors. Furthermore, they serve as the technological foundation for new technology ideas, such as meta-universe and augmented reality, which integrate the digital and physical worlds. Recently, due to the rapid development of deep learning-driven technology, scene representation technologies have attained rapid achievements, especially with the growth of the neural radiance fields (NeRF). This paper begins by reviewing the different representation technologies and is followed by the hot spot, NeRF. Secondly, the recent important results of NeRF are sorted out and described. Thirdly, based on the analysis of scenario building and cross-field research, the importance of explicit space and semantic mining for NeRF is disclosed. Finally, given the virtues brought by the latest achievements from NeRF to the challenges of single-view 3D scene perception, the opportunities for the development of scene modeling and content reproduction technologies with NeRF are revealed.

Keywords: scene modeling; content reproduction; 3D scene representation; neural radiance fields; single-view 3D scene perception

1　引言

自然场景的3D空间建模，以及基于空间建模先验的场景内容重现一直是信息技术围绕人类交互体验的重要努力方向。如图1所示，从1920年第一个数字图像完成对1866年的跨大西洋电报电缆在纽芬兰登陆场景的记录，到结合计算机图形学的构建和基于物理渲染营造逼真呈现，再到结合成像的投影几何去完成空间场景

作者简介：朱方，(1976-)，博士，教授级高工，IEEE高级会员，主要从事数字视频和视觉人工智能高速系统和专用芯片的研发工作。Email：zhu.fang@sanechips.com.cn

的几何建模，研究者一直尝试将自然真实场景转化为有效的数字资产。同时对于场景的3D建模和高逼真呈现与再编辑能力也是元宇宙和增强现实等构建数字和现实世界纽带技术的核心基础。

当前，随着深度学习驱动的信息技术快速发展，作为建模和内容重现的核心—3D场景表征相关技术（从点云，网格，体素，隐函数以及神经隐函数等）得到了澎湃发展，特别是当前神经辐射场相关技术（NeRF：Neural Radiance Fields）的兴起，为基于有限观测自由地生成虚拟视点内容，以及基于空间和时间维度的体积显示重采样提供了丰富应用基石。

为了更加透彻的了解基于深度神经网络技术加持下，3D场景表征相关技术的发展，特别是作为神经隐式表征一个突出代表的NeRF技术的应用潜力和内在关键机理，本文对近期相关研究成果进行了深入的回顾和研究。

本文首先总结了不同3D场景表征技术的应用背景，并回顾了近期针对不同表征技术基于深度学习处理所取得的进展，由此引出了隐表面和神经隐式表征这些有着巨大发展潜力的表征技术；其次，对于神经隐函数中的特别具有代表性的NeRF技术，就其近期发展和延伸研究展开了广泛的探讨，包括其空间和光照可编辑性方面的研究，以及针对动态场景和时序输入场景建模的发展，和如何加速其内容生成以方便实际部署方面的进展。

然后通过针对近期NeRF涉及场景构建及其交叉领域相关研究的深入分析，本文揭示了显性三维空间和语义信息对于NeRF的神经高维隐空间训练构建的重要价值。同时结合近期基于图像的NeRF的有意义的成果，和对当前单帧图像显性3D空间信息感知的挑战分析，揭示了NeRF这种基于输入信息连续高维建模能力为3D场景鲁棒表征和自适应扩展带来的优势。

最后本文对相关论述进行了总结，并进一步呼吁越来越多的后期3D空间多媒体研究以这种“记忆和重现能力”方式向前推进。

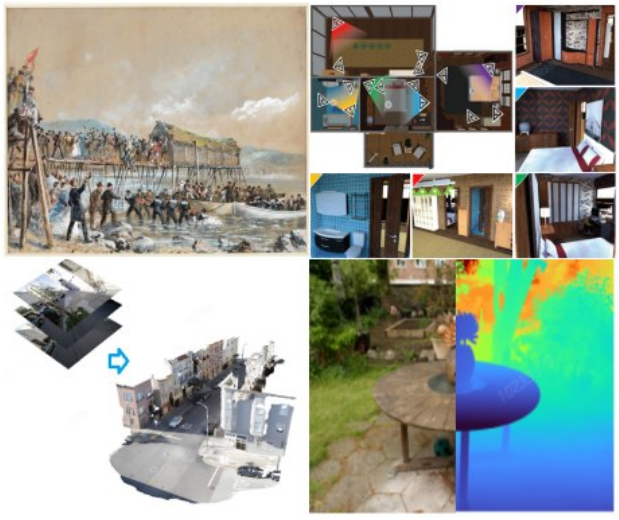

图1 自然场景的建模和重现发展历程

表1 3D场景表征相关技术

场景表征	优点	需要突破的缺点
点云（Point set）	●比较容易通过3D传感器直接获得	●稀疏和空间占用表示不规则 ●缺乏拓扑信息 ●不适合网格范式CNN处理
网格（Mesh & patch）	●有效反应几何信息 ●具备拓扑信息	●空间占用表示不规则 ●不适合网格范式CNN处理
体素（Dense voxels）	●直接对应像素推广到三维的情况 ●适合网格范式CNN处理	●内存占用随分辨率呈立方增长 ●比较高计算负荷
隐式表面（Implicit surface）	●可进行高质量的三维重建 ●参数包括局部几何形状和不确定性 ●内存占用少	●空间占用预测依赖等值面计算导致对应结果网格比较细碎
隐式神经（空间）表征（Implicit Neural Representation）	●不与空间分辨率耦合 ●内嵌直接在表示空间中工作的算法	●需要借助可微体积预测渲染生成可视数据，计算量较大

2 3D场景表征相关技术

现实场景3D建模和内容重现的核心—场景表征技术一直是研究者的重点关注领域,常用的包括了计算机图形学日常使用的网格技术,计算机视觉3D重建传统使用的点云技术,和沉浸式3D多媒体常规使用的3D体积表征(体素网格)技术[1]等。以上三种表征技术以不同的方式离散了输出空间。为了获得更好的空间表征,包括表征量的连续性和多尺度自适应性,以隐性表面为代表的隐式表征技术逐步获得更多研究者的关注[2,3]。特别是近期作为基于神经网络技术有机延伸的神经隐式表征技术,成为当前研究的热点并获得了广泛的探讨,如文献[4-6]。

表1总结了当前主流5种表征类别(点云,网格,体素,隐式表面以及隐式神经(空间)表征)及其局限性。这也是研究者积极寻求突破的重要方向[7-21]。本章节后续部分将就相应方向近期基于深度学习取得的研究进展进行阐述。并鉴于这些成果揭示空间表面隐式表征以及基于深度神经网络的延伸(神经隐式(空间)表征)对于场景三维空间感知建模和基于先验的内容重现发展带来的机遇。

总体而言,近期探索大体可以分成三个主要演进方向:

(1)解决不规则离散欧式空间数值的处理问题

在传统的三种常规表征中,如表1所示,网格和点集体现了实景采集数字化应用场景当前面临的挑战,即空间几何表现(具备清晰的拓扑关系和连续的空间占用表示)和感知工程实践(零散的确定性空间采样点)之间的差距。同时对于结合欧式距离和局部结构特征的拓扑性提取,深度学习网络(如卷积神经网络(CNN))可以提供较好地多层次特征提取能力,但是往往面临如下问题。即直接操作于原始数据(网格和点云),其欧式空间表示的不规则性严重阻碍了直接开展传统的深度学习技术。在欧式空间下3D网格规则化表示的体素网格,由于其空间表示规则化,成为传统卷积网络技术在三维空间下的直接扩展。然而,细粒度的几何信息最终会在3D网格量化中丢失,而且其可伸缩性也会受到高计算和内存成本的阻碍。

这一问题引发了两方面的努力,包括a)既维护基于体积表征的良好准确性,同时大幅降低处理计算复杂度,如近期文献中分别引入了自适应分辨率体积映射[7]和稀疏卷积网络[8]。前者利用空间八叉树数据结构对输出空间分层分解,而后者利用三维点云数据的固有稀疏性,通过只在输入数据的非空位置上保留和执行卷积来降低计算成本。另一方面,b)引入可以适应非欧临域关系或者基于流形的新型卷积计算方法也成为一个积极探索的方向,如球形分形卷积(SFC:Spherical Fractal Convolution)[9]、位置自适应卷积(PAC:Position Adaptive Convolution)[10]和点流算法(PointFlow)[11]以及基于细分结构的网格卷积网络算法(SBMC: Subdivision-Based Mesh Convolution)[12]。

对于缺乏拓扑性的点云表征数据,SFC方法将三维空间点映射到一个由基于分形的规则二十面晶格体创建的离散球体上,然后基于此球体构建具有多层次尺度的卷积神经网络。而PAC方法通过基于由基本权重矩阵构建的模板库构造动态卷积核来模拟三维点云的复杂空间变化和几何结构。其中卷积核对应的基本权重矩阵的组合系数由多层感知器(MLP)从相对点位置自适应学习。和以上基于流形映射以及动态组合来契合CNN的特性不同,PointFlow方法采用比较新颖的图数据表示来表征原始空间数据,以应对数据原始空间表达存在的不规则性。结合这种图表示,PointFlow算法采用了全新的神经网络学习框架,动态图卷积网络(DGCNN:Dynamic Graph CNN),来改进相邻位置之间的特征聚合计算。这种网络框架的彻底革新使得可以从空间数据点的各个空间角度分析其临近点来迭代优化本地三维空间特征的预测。

而对于本身具备拓扑信息的网格表征数据,SBMC方法通过将输入网格进行重网格化,将原本任意连接的局部网格构建为保持特定细分循环序列连接的网格模式。这种特定循环特质意味着一种类似于照片中像素的规则结构,方便了CNN对局部临域规则性的要求。

(2)解决自然场景真实连续性信号多尺度自适应表征的问题

虽然前文例举的相关研究,为将点集和网格引入基于深度网络学习框架,以及缓解基于体素表征的计算成本提供了很大的便利,但原始离散化数据(点集和网格)和基于原始数据的离散化(体素)仍然会限制多尺度密集输出空间的表达,也可能存在量化误差积累(如采用文献[7]中八叉树层次化表示体积表征计算引发的离散量化误差)。而且高质量的交互空间三维几何表示,需要能够描述足够精细的连续空间占用,并在较少存储要求下包含多尺度信息。也正是因此,隐式表面(Implicit surface)相关技术吸引了相关

研究者的关注。

使用隐式表面进行空间几何表征可以追溯到文献[13]，其通过将带符号距离函数（SDF：Signed Distance Function）数值存储在一组描述被占据表面的体素中。虽然这样隐函数描述的表面是连续的，但输入空间简单离散化会引入表面质量缺陷，如文献[14]中所述。为了克服这一缺点，后续的研究[15, 16]采用高斯过程对映射进行建模，将离散输入转化为数据先验问题，并增量地执行贝叶斯映射更新优化。

近期，随着深度神经网络所展现的强大数据驱动学习能力，利用深度学习融入相关数据先验的问题解决引发了神经隐式（空间）表征（INR：Implicit Neural Representation）研究热潮[4-6]，包括了非线性拟合[17]和元学习[18]等方面的研究尝试。其中比较经典的文献为近期的空间占用预测网络（Occupancy Networks）[4]和隐式移动最小二乘曲面算法（IMLS：Implicit moving least-squares surface）[19]。其中空间占用预测网络使用深度神经网络分类器隐式地将三维空间表面表征为连续决策边界。这样利用神经网络学习的连续决策边界不仅可以在固定的离散3D位置（如已有体素表示）推理空间表面占用率，而且在任何可能的空间3D点（$p\in R^3$）都可以实现占用率推理。因此这种创新方法可以在极小内存空间占用（学习的神经网络权重）并在无限输入分辨率下输出3D空间表面描述。

而IMLS方法则进一步体现了INR的优点。首先其和空间占用预测网络类似，即利用一个内嵌参数的神经网络的零水平集表述三维空间中的一个曲面。如前文所述，参数化信号所需的内存与输入空间分辨率无关。同时IMLS方法还拓展定义了所表征曲面所在的空间维度，即可以表示高维空间中的一个流形（超曲面）。这样INR即可不同于传统空间表征（点云、网格和体素）仅仅局限于空间占用或空间表面的表征，其还可以作为融合其他高维特性的重要表达。近期研究概率局部隐式体素（PLIVox: Probabilistic Local Implicit Voxel）[20]就是一个极好的例证。其不仅捕获场景空间几何描述，还通过单一深度神经网络捕获空间占用描述的不确定性属性。最近诸多的研究，如文献[21]，也不断地证明了INR源自深度神经网络的灵活性和良好的表达能力。其中特别需要强调的是其对可高维关联良好的归纳偏置以及隐式的正则化属性。

（3）解决多影响因素融合的光学被动采样信号的显式分离问题，如位置、角度、环境、材质和局部空间特征

如以上章节所述，常规INR使用离散空间点集作为输入对光滑连续空间曲面进行建模，可以为下游任务结合点集输入的灵活性和隐式曲面输出质量优异的特性。但是如果进一步提升自然场景空间表述的完整性和连续性，离散空间点集（对空间点可信感知）输入成为进一步制约。

光学被动采样的空间采集完备性（仅仅受限于采集传感器精度）一般远远大于主动检测。但光学被动感知的后续运用，如传统的多视角立体几何稠密重建，一直受限于光学被动采样结果的多影响因素融合，如照明、相机参数、采集姿态和对象外观等。近期由有限观测自由生成虚拟视点内容驱动的INR升级，神经辐射场（NeRF：Neural Radiance Fields）[22]为消除以上局限性开辟了新的机遇。和常规INR的零水平集表征不同，基于NeRF的一个空间场景被表示为一个输入为5D向量的函数，用一个多层感知器（MLP）神经网络隐式表达，其输入包括视图采集射线的角度和场景中射线上的特定3D位置，其输出包括了3D位置对应的颜色和空间体积密度。其公式如式1所示，详细注解和计算过程可以参照文献[22]。

$$F_\theta:\big(x(x,y,z),d(\theta,\phi)\big)\rightarrow(c(r,g,b),\sigma) \tag{1}$$

对于一个已知场景和观测角度，F_θ对应的视图可视内容需要依赖数值积分方法来近似一个真实的体积渲染过程，如下式所示。

$$C=\int_{t_n}^{t_f}T(t)\sigma(r(t))c(r(t),d)dt \tag{2}$$

基于以上表征模型和可视内容生成模型（NeRF的核心构成），3D空间场景可以结合许多已知姿态的视图图像进行训练，对应场景体积表示（包含光照和对象外观等隐变量）存储为MLP的权值。

NeRF自2020年进入研究者视野，成为近期的一个重要技术方向，也为基于深度学习有机融合已有计算机图形学和计算机视觉的典型机理开辟了广阔的机遇。

3 神经隐式表征-NeRF近期发展

发展之初的NeRF伴随着如下问题，诸如：无论是训练（小时级）和渲染（几百毫秒）都很慢；只对静态场景表征；一个训练所得场景表征无法拆解和知识转移到类似场景/对象。这些问题的提出也体现了研究者和业界对NeRF表征对后续应用的期望：包括快速可部署性，和基于时序动态可变形场景内容建模以及后

期结果基于环境和空间的可编辑性。针对这些诉求，近期众多的NeRF研究成果[23-48]涌现，主要归纳为如下几个方向的开拓和尝试：

(1)针对可变形对象的建模

这个方向主要针对动态对象建模。这里的动态主要指时变观测下场景中对象外观存在非刚性形变，但同时这种形变存在很强隐变量约束。近期可变形对象研究方向的成果主要聚焦在人体的体积动画模型表征构建和相关自由视角合成方面。典型成果包括了基于像素对齐的人物化身体积动画建模研究(PVA：Pixel-aligned Volumetric Avatars)[23]，姿态可控的人物化身自由视角影像合成研究(Neural Actor)[24]，和用于动态人物化身的动画神经辐射场研究(ANeRF:Animatable Neural Radiance Fields)[25]。相关的包括了聚焦人脸面部化身4D动画重建应用的动态神经辐射场研究(D-NReF-Face)[26]，以及可以兼具场景和人物化身4D动画处理能力的动态神经辐射场研究(D-NeRF)[27]和显式构建神经辐射场拓扑可变高维表示的研究(Hyper-NeRF)[28]。

其中对可变形对象神经辐射场建模的一个基本思路往往是将一个动态神经辐射场(对应非刚性变形场景)分解为一组变形场和一个标准的静态神经辐射场。其中变形场负责将被观测变形空间点映射到标准空间，从而使它们能够从图像视图序列中学习动态可变形场景。比较典型的如D-NeRF和Hyper-NeRF，其主架构中都启用了不同的变形网络架构将动态场景中变形后的空间点映射到后续静态神经辐射场所包含的一个标准空间。所不同的是映射计算过程不同。其中D-NeRF直接将变形估计为3维空间位移推断，而Hyper-NeRF则在变形网络基础上并行一个高维辅助函数的切片推断网络，通过提升原有映射空间到一个高维变形场模拟(额外维度为环境维度)实现了对一系列拓扑变化的形状建模，并强化了拓扑可变中包含的上下文一致性。在时间和空间维度下的场景插值测试中，Hyper-NeRF方法都保持了很高的场景设定一致性和视觉合理性。

同时对于聚焦人物化身体积动画的研究，如Neural Actor和ANeRF，往往会强化添加与人的形体相关的特定约束。如ANeRF采用了基于骨架驱动的变形预测，利用可观测三维人体骨架分析赋能随后的权重混合计算，进而推动观察变形空间到标准空间准确映射。而Neural Actor则是利用结合形状参数和姿势参数的人体编码模型(SMPL:Skinned Multi-Person Linear Model)作为代理，并结合变形空间点周围纹理特征分析推动相关准确映射展开。

(2)针对连续时序内容的建模

和之前的侧重点不同，这个方向主要针对基于时空联系的场景表征建模。通过同时构建时空两个维度的建模表征，研究者后期就可以方便开展基于场景视频记录的时间插值、视点插值以及混合插值的应用探索。相关领域典型成果包括针对动态场景时空视图自由合成的场景流场算法(NSFF: Neural Scene Flow Fields)[29]，时空辐照度场算法(STNIF: Space-time Neural Irradiance Fields)[30]，动态视点合成算法(DVS: Dynamic View Synthesis from Dynamic Monocular Video)[31]和侧重于人体动态时空新视角合成的隐式神经人体表征研究(Neural Body)[32]。和Neural Body方法侧重于连续时刻稀疏多视图同步输入构建时空模型不同，NSFF等其他方法都侧重于单一视图的视频输入，既在任何时间点只包含对场景的一个观察结果。基于显性时空联系的场景表征建模方法，如具有代表性的NSFF和DVS方法，都将动态时变空间场景建模为场景元素的外观、空间几何属性和其三维场景中运动的时变连续函数表示。特别是NSFF通过明确地将时间纳入场景函数表征变量域内，将场景元素三维运动建模为密集的场景流场，并同时将正向和反向场景流都显式建模为密集的三维向量场来准确建模场景中元素三维运动。同时针对视频内动态空间场景元素涉及的采样特点：即运动元素通常会经历较大形变，无法可靠地推断出在较大时间间隙上的空间对应关系，而静态元素则能保持准确的对应关系，可利用框架下所有的可共视观察样本强化静态元素的表征构建。NSFF和DVS都采用基于以上机理的区域分别处理和再合成的处理策略。当前的研究取得了不错的进展，但在应对更加挑战的野外场景，如包括复杂薄结构和包含复杂运动程度等，还存在不足。

(3)表征建模的环境光照分离和编辑

这个方向主要针对场景元素建模的外观分量中光照隐变量的分解和重计算，包括了如何消除建模过程中不规则光照的影响，以及准确捕获新颖视角下的光照效果和重新构建场景中的光照效果。近期典型文献包括了基于不受约束环境下采样照片集合构建神经辐射场的算法(NeRF-W)[33]，对形状和反射率隐变量进行因子分解的算法(NeRFactor)[34]和NeRD[35]，以及用于视图和光照重新合成的神经反射和可见场

算法(NeRV)[36]。其中NeRF-W是NeRF的第一批后续工作之一,针对NeRF依赖光照保持不变的输入视图集合的缺陷,其运用生成式隐变量优化框架(GLO:Generative Latent Optimization),优化出每个输入图像的外观嵌入向量(apperance embedding),并以此学习到整个输入照片数据集中的共享外观表示。这使得相片相关的外观和光照变化解耦,并可以显式地建模。这种光照分离让NeRF-W在光照环境变化的场景下有很大的灵活性和鲁棒性,可以从较少环境约束的多视图集合中稳健地完成场景神经表征学习。为了更好实现NeRF辐射场隐函数空间对应隐空间变量(光照,法线,漫反射,空间表面表征)的分解,后续相关的研究都引入了类似式3的辐射合成计算模型,如NeRFactor和NeRV算法。这也借鉴了计算机图形学高逼真渲染的计算机理。

$$View = \int_s (Lvis \times Dill + Iill) \times BRDF d\omega_i \quad (3)$$

其中s表示围绕场景的球形环境图,$Lvis$表示场景可视性因素,$Dill$表示直接光照因素,$Iill$表示间接光照因素,$BRDF$为双向反射分布函数,ω为入射角度。

整个分解过程/网络框架也遵循利用多个独立MLP对相应隐空间变量进行建模原则,包括对应表面法线、表面材质参数、体积密度、场景对于外部环境在任何方向的可见性等。即整个框架为借助于将标准NeRF表征(独立MLP)输出到后续的多个MLP之中,并利用这些MLP完成对应隐空间变量的因式分解。

为了有效训练对应空间表征的隐函数参数(MLP网络的权重),整个训练过程,如NeRFactor,采用了分步开展的流程。在其余MLP被固定的情况下,先训练好标准NeRF MLP,同时利用真实测量值训练隐变量空间进而获得BRDF先验模型。然后将NeRF初始估计完成的体积密度提取成空间表面表征(结合法线和可视性)并联合优化,再最终实现结合反照率(Albedo)和反射系数特征(BRDF latent code)以及光照环境(Light)的联合模型训练和全局优化。

以上处理使得相应研究算法,如NeRFactor,能够基于一系列不同位置的图片估算出物体形状和光场信息,并能在任意光照条件下,都可以从新的视点完成体场景空间的准确呈现。

(4)基于空间的表征建模可编辑性

这个方向主要针对多物体组成的大型场景的结构化表征,包括了如何将多物体组合成一个完整可体积渲染场景,以及场景内容再编辑方面的研究。其也对应可控图像合成任务。换言之,表征建模可编辑性着眼于生成新的图像和控制将要出现的内容、对象及其位置和方向、背景等。近期典型文献包括了针对可编辑场景表示的可组合生成特征算法(GIRAFFE:Compositional Generative Neural Feature Fields)[37],可组合场景对象算法(ObjectNeRF)[38],以及涉及动态场景构建的场景图算法(Neural Scene Graphs)[39],和可编辑条件辐射场算法(EditNeRF)[40]。其中GIRAFFE为国际计算机视觉与模式识别会议(CVPR)2021的最佳论文。

空间表征建模的结构化对应着3D体积表征和3D对象以及3D特征的关联构建过程(训练),同时可控图像合成也覆盖了结合特征空间的3D体渲染内容生成过程。早期工作生成辐射场(GRAF:Generative Radiance Fields)[41]开创性的引入了生成框架(GAN),和NeRF训练以及体渲染过程融合,实现了局限于单物体场景的高分辨率可控图像合成。GIRAFFE和EditNeRF方法多受其启发。

同时,为了进一步深入多物体大型场景,即需要从背景中分离出一个或多个物体以及能够表达单个物体的形状和外观,GIRAFFE突破性地将场景表示为可组合的神经特征场。其将不同物体从场景中分解出来,并引入了对应仿射变换来表示每个物体,从而可以对场景中单个物体的姿态、形状和外观进行控制。在后续处理中,GIRAFFE通过使用以对象为中心的NeRF模型输出特征向量而不是颜色来支持组合,并通过平均来开展组合计算,并最终通过神经渲染将2D特征向量图转化成高分辨率彩色可视图像。

(5)更快的可视视图内容生成推理

伴随着以上NeRF内容适应性和应用扩展性的研究,众多研究者也就快速生成显示内容(神经辐射场渲染)展开了大量探索工作。其中极具代表性的包括:起始于2020年的稀疏体素场算法SVF(Sparse Voxel Fields)[42],和2021年涌现出的快速高保真辐射场渲染相关研究:FastNeRF[43],SNeRG(Sparse Neural Radiance Grid)[44],和PlenOctrees(plenoptic octrees)[45]。

以上研究成果都围绕上文公式2所描述的依赖数值积分方法近似一个体积渲染过程。参照公式2和文献[22]相关描述,其中沿摄像机射线与场景空间几何表征的精确交叉查询,以及在场景描述精度上(对应网络容量)沿射线进行的体积积分都引发了较大计算

负荷。针对相关空间内数据的稀疏性,展开高效检索,如采用空间八叉树(Octrees)的数据组织结构,成为很多方法共同的考量,如NSVF和PlenOctrees方法。

同时就其初始的端到端计算过程,FastNeRF将原有过程拆解成2个步骤(位置相关和角度相关)。其中位置相关计算结果为包含深度信息的辐射度贴图,可以缓存下来供后期使用,而不用反复计算。SNeRG则更进一步,首先将辐射度计算按影响因素分解(如空间表面,漫反射和反射)。这些因素针对NeRF的输入(位置和射线角度)具有不同的可复用计算程度,如漫反射对于特定空间位置和领域就比较一致,而反射和空间表面特性有关,可以通过结构特征提取构建组合模板。进而,SNeRG将整个计算过程有机的区分和预计算,将一个端到端的计算过程转化为一个高效的查询和简单组合计算过程,如同计算机图形学常用的烘培技术。

除了将整个空间表征作为一个整体,参照式2进行流程优化,2021年也有很多研究者试图从辐射场体积表征数据分解的角度,探索加速可能性。这类研究以成果(DeRF)[46]和(KiloNeRF)[47]为代表。其中特别是KiloNeRF探讨了利用众多微小MLP替换原有MLP(NeRF空间表征隐函数权重)的神经辐射场加速创新的可行性,并在没有产生较高存储成本前提下,与原始的NeRF模型相比取得了三个数量级的渲染速度提升。

4 显性空间语义对NeRF的重要性

上文将近期NeRF诸多发展方向进行了详细阐述。同时也使得我们对基于NeRF体系的神经隐式空间表征模型特点有了一定认识。本节将结合近期如何更快完成表征参数空间训练的研究,即相应场景构建分析,以及包含显性使用空间和语义的NeRF相关交叉研究来揭示显性空间和语义信息及其预测对NeRF的重要性。

(1)高效表征的参数空间训练研究

在构建NeRF体积空间表征时,如前文所述,我们需要大量已知采集方向和位置的视图图像反复使用辐射场体积渲染,来训练对应MLP网络权重。

如何高效(利用少量稀疏输入以及高速训练)实现权重训练和最终结果的核心影响因素是什么是本节希望解析的要点。以下我们就两个方向的近期研究展开回溯:

首先是如何基于稀疏视图(单个或几个视图图像)来实现NeRF的MLP网络训练。这方面可以借鉴的典型论文包括:隐性构建统一空间几何先验的神经辐射场训练研究(pixelNeRF)[48],显性构建统一空间几何先验的神经辐射场训练研究(SRF:Stereo Radiance Fields)[49],和神经辐射场正则化的研究(RegNeRF)[50],以及360度无边界场景无歧义神经辐射场训练的研究(Mip-NeRF 360)[51]。

初始构建神经辐射场的方法是独立地优化对每个视图场景的表示,其中视图场景的生成依赖于输入射线的位置和方向。从前文对加速体积渲染的相关成果阐述中(如SNeRG方法),我们可以发现其场景内部的空间结构也是一个重要的隐变量,并具备一定的共视一致性和外观决定性。

pixelNeRF方法就引入了一种完全卷积架构,对视图图像输入序列进行跨多个场景的统一学习训练,以学习场景中的空间先验。而SRF方法更是直接借鉴计算机视觉的立体几何重建机理,即组合图像对可以构建基于几何一致性的显性外观匹配关系,同时表面空间占用信息(空间结构)会导致对应外观有明显可区分性。SRF方法对于输入的参考视图集合基于场景中空间点对应视图投影位置提取CNN特征并结合学习到的相似度函数构建对应匹配。然后用深度神经网络计算聚合的立体特征和对应编码。这个立体特征空间也对应了其神经隐式空间表征,其编码对应了显性外观颜色和空间密度,通过辐射场解码网络完成对应推理计算。

虽然不管是运用隐性或显性的场景空间几何先验都可以有效降低原始训练对输入样本数量上的需求,但过于稀疏的输入视图数据仍然会导致场景空间辐射场估计的误差,并最终导致新颖视点视图合成输出的伪影。RegNeRF方法针对这种情况,设计了一套正则化机制来规范化未观察到的视点颜色。其核心思想就包括了外观正则化和空间几何正则化两个部分。其空间几何正则化过程通过设计重建损失优化项,即对渲染图形片段的深度强制执行平滑性损失,并通过在训练过程中对射线采样空间进行退火,进而提升了过于稀疏的输入导致的质量下降问题。除了以上视点聚焦的中心场景及其对象,在360全景自由视点构建时,其360度背景也会呈现稀疏输入且场景无边界的特点。近期论文Mip-NeRF 360亦和RegNeRF方法相似的构建了空间几何失真正则化器(基于

不同场景参数化形式)。通过此正则化器,场景空间几何属性训练结果可以更有效地纠正悬浮物和背景坍塌等缺陷。

其次,我们在保证最终新颖视点视图质量的前提下,聚焦场景表征的快速构建方法,并尝试对相关核心要素进行剖析。

这个领域相关核心力作包括两方面的探索,第一类当属如何对MLP构建的权重空间进行分解和并行构建方面的研究。这个方面前文已有初步涉及,如KiloNeRF方法,但最具代表性的文献为近期英伟达研究团队的Instant NeRF/Instant Neural Graphics Primitives[52]和谷歌研究团队的Block NeRF[53]。

其中Instant NeRF相关研究区别于之前的权重空间分离(KiloNeRF)和检索方法(NSVF)最突出的是体积渲染和检索所依赖体积空间索引通过特征可学习的参数编码,即不仅公式1的映射函数用学习驱动的隐式特征表征(MLP),而且公式2的组织方式也用学习驱动的特征向量协助构建。其网络框架转化为以MLP为核心,由包含特征向量组成的多分辨率哈希表增强的参数编码框架。由于引入此种位置参数编码机制和巧妙设计了低计算复杂度的哈希算法,Instant NeRF的训练学习完成时间缩小到了秒级。这种通过输入参数特征空间引发计算简化提升效率,也在一定程度上体现了背后空间和语义信息的重要性。

第二类是关于结合辐射显示计算机理更新隐式参数表征的物理意义方面的研究,如基于基函数隐式组合扩展的多平面图像(MPI)场景表征的研究(NeX)[54],以及依据辐射场和光场的关联性,探讨光场神经隐式表征(LFN:Light Field Networks)的研究,如近期麻省理工团队的研究成果[55]和卡内基梅隆的团队相关研究成果[56]。

其中NeX采用了混合隐显式建模策略,即和NeRF原始采用隐式空间几何表征对比,其利用了MPI这种显示的空间几何表征作为基础,但吸取了NeRF对于视角依赖隐式表征的优势。这样的有机混合不但加速了相应的生成速度(有点和FastNeRF相似),而且使得相比于原始NeRF对于更具挑战性的场景视觉效果(比如CD上的彩虹反射)取得了更一致和逼真的效果。

而LFN相关研究则揭示了光场和辐射场对于视图合成和场景建模的优缺点。其中光场可以表示沿光线的辐射合成,其渲染过程比较辐射场计算(多次计算完成一条射线的近似体积积分)简单。但其对空间几何场景的映射方式(沿可观测射线),由于并不直接对应空间占用的显性信息(三维世界坐标),导致其不是直接保证多视图的一致性。而相反,基于3维世界坐标系的NeRF通过射线和空间的匹配计算可以确保多视图的一致性。也基于此NeRF可以直接通过最小化已知相机姿态下的真实视图与对应基于表征重建之间的差异来充分优化。为此,麻省理工的研究者和卡内基梅隆的研究者都通过引入元计算(Meta Learning)来学习LFN的空间先验信息,既三维场景的空间分布。并基于此,相关LFN研究在生成质量和NeRF齐平的情况下,实现了表征紧凑和生成迅速的目标。

(2)显性空间和语义结合的NeRF交叉研究

本节,我们将就包含显性使用空间和语义的NeRF相关交叉研究展开探讨,相关研究可以归纳为三个主要类型:

第一类相关研究借助于显性使用多视图的一致性信息,提升NeRF的训练和显示计算的鲁棒性和准确性。典型研究包括借助多视图立体几何计算优化神经辐射场的成果,如MVSNeRF[57], NerfingMVS[58],和BARF[59]。此类研究或者提高了稀疏输入的鲁棒性,如MVSNeRF,或消除了NeRF构建过程可能存在的几何形状与生成内容的不匹配模糊,如NerfingMVS,或提升了NeRF训练过程对相机姿势缺失的鲁棒性。

其中MVSNeRF运用3D CNN,基于多视图多深度平面扫描计算,构建了由体素神经特征组成的神经编码空间,进而获得了可微分学习的场景空间几何显式表达,并将其与体积渲染相结合。而NerfingMVS则利用运动结构恢复(SFM:Structure from motion)对每个视图输入的单帧稠密深度估计进行微调,进而通过优化的视图深度先验来监测和优化NeRF体积渲染的采样过程。BARF则是聚焦位置编码(NeRF构建过程的核心)局限性:即没有位置编码在重建中缺乏保真度,而完全位置编码容易导致空间注册次优。通过建立与经典图像对齐理论的联系,BARF构建了从粗到细的NeRF配准流程,实现了三维神经表示和相机帧注册问题的联合学习。

除了以上经典成果,近期研究Point-NeRF[60]则更是将空间先验索引构建和引导NeRF训练优化推到了一个新的高度。其利用基于成本空间的3DCNN基于多视图空间一致性生成视图稠密深度估计,并利用2DCNN生成平面片段特征。其中特征矢量和显性空

间占用的点集合并构成初始神经点云(每个点都有一个空间位置、一个置信度和反投影的图像特征)。然后其利用三维点云显性空间信息索引构建基于空间点临域图像特征矢量的隐性辐射场(由MLP构建隐性表征参数)。由于神经点云可以借助显性空间信息,使得其构建过程可以利用通常点云处理工具实现点云的剪枝和补全以提高质量。同时基于此神经点云,以及由此构建的局部点特征输入,MLP更容易优化,这一点可以借鉴上文提到的方法KiloNeRF。以上显性的空间信息构建和利用使得Point-NeRF相较于初始的NeRF在构建速度和生成视觉质量方面都有较大的提升。

第二类相关研究直接利用显性深度信息提升NeRF的相关训练和可视化内容生成。这类研究包括了近期利用单视图稠密深度信息预测网络辅助实时绘制视点视图的研究(DONeRF)[61],和稀疏输入训练的研究(DS-NeRF)[62],以及结合连续主动深度检测信息实现动态场景视图生成的研究(TöRF)[63]。

其中DONeRF的原理基于当样本采样积聚在场景空间表面周围时,视图渲染中每个视图射线计算所需的样本数量可以显著减少。DS-NeRF则是揭示了稠密的深度信息(空间几何信息)提供了整个视图重建基于像素级的空间和颜色反向传播优化的途径,这一点在Point-NeRF方法中也有明显的体现。

而TöRF方法则创新性的探讨了结合主动深度检测结果指导动态NeRF构建的意义和局限性。其基于飞行时间(ToF)相机测量数据的NeRF建模,和仅使用彩色摄像头相比,减少了场景建模所需的图像数量。同时也进一步验证了直接编码有关场景空间几何信息令基于单视图的动态NeRF建模更容易处理。

第三类相关研究主要聚焦语义信息和NeRF隐性表征的互动,包括基于语义一致性稀疏输入训练的研究(DietNeRF)[64],和通过语义信息嵌入将NeRF隐空间维度提升的研究(Semantic-NeRF: Semantic Neural Radiance Fields)[65],以及直接由语义图生成NeRF表征的尝试(Sem2NeRF)[66]。

其中DietNeRF和之前基于多视图一致性研究相比,提出了高层语义一致性的思路,提升了多视图一致性的应用层面。而Semantic-NeRF相比于前文高维NeRF表示研究的Hyper—NeRF方法,强化了语义背后对于外观和几何形状的表征,这也被Sem2NeRF研究进一步揭示。同时Semantic-NeRF利用自然场景空间临域固有(由几何空间信息决定)的一致性和平滑性,强化了稀疏语义标签的空间有效传播。这为诸多视觉语义空间感知的相关应用,如新颖的语义视图合成、标签去噪、超分辨率、标签插值和多视图语义标签融合,提供了一种高效和鲁棒的方法。

5 单视图场景感知融合NeRF的机遇

在上一章节中,我们可以清晰捕捉到显性场景空间和语义信息是有效提升神经隐式表征的核心环节。同时在当前3D空间场景感知研究领域,也如文献[67]所述,单视图空间感知(稠密深度估计)然后融合通常比直接多视图配置具有更高的鲁棒性。

由于基于神经网络的单视图3D场景感知,其早于NeRF的出现已经经历了一定的发展阶段。本章我们将从其近期发展和面临的挑战入手,和NeRF相关研究展现的裨益,探讨融合NeRF的单视图3D场景感知面临的机遇。

(1)当前单视图3D场景空间和语义感知的挑战

单视图3D场景空间和语义是基于神经网络计算的三维重建和场景理解处理框架的一种重要领域,其具备潜质可以避免现有基于有源深度传感器密集测量的诸多缺点,包括操作范围有限、空间分辨率低、传感器多源和多径干扰和功耗过高等。

近期很多新颖的研究成果展示了基于神经网络的单视图图像稠密深度感知的潜力。其主要围绕2个主题展开:

a)提高单视图稠密深度预测的性能

近期典型研究成果对性能方面的追求包括了对单视图场景结合高分辨率输入提升预测精度的探索(MergNet)[68],通过辅助可信度信息提升准确性的探索(Neural RGB®D)[69],和结合图像中的结构信息和纹理信息解耦,降低基于深度学习的被动感知纹理依赖性的探索(S2R-DepthNet)[70],以及轻量化应用网络架构的探索(FuSaNet)[71]。

其中,对于单视图每像素深度估计存在的挑战,即由于给定网络模型容量和接收域大小限制引发的准确性缺失。Neural RGB®D将基于单张图像的一次深度值估计转变为单次深度值的概率分布后验,并利用多次估计基于时间聚合优化(通过贝叶斯滤波框架)来提高准确性。而MergNet对这个问题的解决则通过利用图像的近似边缘图(对RGB梯度进行阈值处理获得)构建结构一致性传递的重要线索,将单次网络推理对应的不同图像分块的不同分辨率估计进

行合并,来构建一个具有一致整体结构高频细节的高分辨率估计。FuSaNet则是通过对应视图显著点的提取和对应空间信息来规范化深度预测结果来提升对应网络模型有效容量。

和之前挖掘网络容量,利用全局结构一致性和时间一致性规范预测输出结果提升质量不同,S2R-DepthNet则针对深度预测训练中深度网络比较聚焦纹理特征的提取,对结构特征关注不够的缺点,提取深度相关结构信息强化网络对深度预测的准确性和网络泛化能力。这一研究也揭示了当前很多方法过分关注纹理信息,也会导致深度信息的数据领域存在场景依赖,加重了网络容量负荷。同时现实场景下纹理信息容易受光照、噪音和运动模糊等因素的干扰,结构信息往往更加重要。同时,这也体现出当前阶段的稠密深度预测依然存在很大的提升空间。

b)提升单视图深度预测的自监督学习能力

自监督单视图深度预测是实际部署相关感知能力非常重要的环节。而且单视图深度估计的自监督方法本质上是利用三维场景中对象在投影成像后结合空间结构信息和相机位姿存在的光度一致性,基于内在几何关系(主要是多视图一致性)监督网络对深度信息/相机位姿估计的学习。其中配合自监督训练过程,构建最小重构误差的规范项,和提升重构光度计算的准确性,以及有效搜寻对应光度匹配就显得尤其重要。

在这方面探索的典型成果包括了近期的成果基线MonoDepth2[72]和对目标细节更好特征封装的PackNet[73],以及同时应对刚性和非刚性部件的Non-Rigid-DepthNet[74]。其中MonoDepth2方法引入了在输入/目标图像之间对边缘敏感的平滑度损失规范项,鼓励模型学习到尖锐边缘并有效抑制噪声。而PackNet方法则添加了对相机位姿平移分量的约束,避免了之前方法存在的尺度不清晰的问题。同时PackNet使用3D卷积替换了传统使用的pooling和线性upsample操作,从而使得图片中的目标细节能够更好的保留下来,提升了重构光度计算的准确性。Non-Rigid-DepthNet方法则通过针对像素构建运动内嵌隐变量,并利用结合光流计算的结构边缘提取来提取有效匹配。同时其通过在CNN训练中尽可能利用刚性变换先验作为监督,对非刚性单目深度实现了有效无监督学习。

从以上研究结果可以看到像素级甚至亚像素级的结构细节提取,以及与噪音区别的准确光度计算都将为后期相关研究提供重要的提升空间。

同时针对单视图3D场景空间和语义联合感知方面,聚焦核心挑战,即如何提升显性融合和辅助同步语义理解,很多研究也做了积极尝试。其中就包括近期经典研究,单目三维空间语义场景完全感知(MonoScene)[75]。针对挑战,MonoScene方法提出从单个RGB图像中通过对特征进行视线投影计算(FLoSP:Features Line of Sight Projection),即由光学投影启发的二维-三维特征转换,构建了一种图像三维特征体素空间计算范式。其中体素特征通过反投影图像坐标临近的多尺度特征构建。同时这种计算范式为后继基于3D卷积的空间上下文关系先验挖掘提供了一种独特的损失函数约束基础,即视锥空间和投影平面语义一致性损失。

通过文献自身的结果分析,可以看到基于图像的稠密语义提取往往是不稳定的,容易受到视点焦平面,光照环境和环境噪音的诸多影响。前文中的研究Semantic-NeRF中提示的NeRF语义固有多视图一致性和平滑性使非常嘈杂环境下稀疏可信采集信息能有效传播也给我们提供了不小的想象空间。

(2)融合NeRF的单帧图像空间语义探索的机遇

近期融合NeRF的单帧图像空间语义探索已经引发研究社区的兴趣,初步涌现的研究成果包括了最近的为新颖视图合成应用结合NeRF的连续深度MPI研究(MINE:Continuous Depth MPI with NeRF)[76]。

其中MINE的研究,在一个单一图像输入基础上,通过引入神经辐射场思想构建了一个可表达连续深度的多平面图像(MPI:Multiple Image)扩展三维空间表达方式。在弱监督的系统设置下,MINE在单目深度估计任务上取得了大幅超越其他弱监督设置方法的性能,甚至非常接近全监督设置最先进的方法性能。

同时基于单张图片的NeRF构建进展,如用于图像超分应用的,基于局部隐式图像函数(LIIF:Local Implicit Image Function)学习的连续图像表示研究[77],为后续基于图像的自适应多尺度空间理解,提供了高度结构一致性的新解决线索。其中LIIF的相关研究受隐式神经表征的启发,通过自监督方式在图像超分任务上训练了一个提取特征编码器网络和图像表征,局部隐式图像函数(LIIF)。所学习的连续表达因坐标连续性,能够表示成任意分辨率形式,甚至对自然图像和复杂图像可进行30倍放大插值。

近期基于NeRF,对于高噪音低照度输入图像的单视图场景构建和后继动态光度的高质量高精度合

成的研究(RAWNeRF:NeRF in the Dark)[78],也为相关基于图像的空间及语义理解的实际落地应用提供了一定新颖的思路。其中RAWNeRF初始需要应对输入为基于受损相机信号采集管道的低动态范围(LDR)原始传感器数据,其伴随着噪音扭曲和细节平滑等质量问题,且信号采集在高噪音低照度环境下。为了应对这一挑战,RAWNeRF在训练流程中结合这种原始传感器数据(保留了场景的全动态范围信息),并采用了由高动态范围(HDR)新颖视图合成驱动的网络学习。研究结果发现RAWNeRF网络体现的基于积累噪音输入优化的场景信号保留能力要超过原始学习流程上采用专用去噪器所产生的效果,可以应对接近黑暗的学习场景。同时建模完成的NeRF甚至具备了操纵对焦、曝光和色调映射能力。

6 总结

现实场景3D建模和内容重现的核心—场景表征技术一直是研究者重点关注的领域。伴随着对3D虚拟场景、真实场景以及虚实融合场景的构建/呈现/编辑的不断尝试,3D场景表征技术涵盖了从计算机图形学日常使用的网格技术,以及计算机视觉3D重建传统使用的点云技术,和沉浸式3D多媒体常规使用的3D体素网格技术等。

为了实现对已观测采样的自然场景3D内容更自由和智能地呈现与再编辑,研究者对于有限采样下获得更高效的空间表征(隐式表面技术等)以及基于已建表征如何快速生成高逼真度的可视内容展开了积极探索,特别是在当前快速发展的深度神经网络技术加持下。在这个背景下,能基于低存储空间实现空间连续性表征和基于体积渲染实现高质量内容生成的神经辐射场(NeRF)技术及其延申研究获得了众多研究者的关注。

本文针对NeRF相关3D场景表征近期研究进行了回顾,包括:a)NeRF近期针对空间和光照的编辑方法;b)基于时序输入的表征构建方法;c)基于动态内容的表征构建方法;c)基于表征的可视内容快速生成方法。这些不断涌现的优秀成果,也一定会激发研究者对NeRF构建和生成核心影响要素的渴求。为了揭示这一奥秘,本文结合对近期如何更快完成表征参数空间训练的研究,和包含显性使用空间和语义的NeRF相关交叉研究的回顾,揭示了显性空间和语义信息及其预测对NeRF的核心重要性。

最后,结合近期显性空间和语义挖掘中的重要发展方向,单视图深度估计,面临的挑战和神经辐射场相关研究展现的裨益,揭示了基于神经辐射场对场景三维空间感知建模和基于先验的内容重现发展带来的机遇。本文专注于基于神经网络的3D空间场景高维表征,特别是NeRF的研究,并进一步呼吁越来越多的后期3D空间多媒体研究以这种"记忆和重现能力"方式向前推进。

参考文献(References):

[1] Steinbach E, Girod B, Eisert P and Betz A. 3-D object reconstruction using spatially extended voxels and multi-hypothesis voxel coloring[C]. International Conference on Pattern Recognition (ICPR), 2000: 774-777.

[2] Li S, Yan D, Li X, Hao A and Qin H. Detail-Preserving 3D Shape Modeling from Raw Volumetric Dataset via Hessian-Constrained Local Implicit Surfaces Optimization[C]. International Conference on Cyberworlds (CW), 2016: 25-32.

[3] Singla K and Astya P. Enhancing 3D implicit shape representation by leveraging periodic activation functions[C]. International Conference on Signal Processing, Computing and Control (ISPCC), 2021: 101-106.

[4] Mescheder L, Oechsle M, Niemeyer M, et al. Occupancy Networks: Learning 3D Reconstruction in Function Space[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019:4455-4465.

[5] Schirmer L, et al. Neural Networks for Implicit Representations of 3D Scenes[C]. Conference on Graphics, Patterns and Images (SIBGRAPI), 2021:17-24.

[6] Peng S, et al. Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 9050-905.

[7] Funk N, et al. Multi-Resolution 3D Mapping With Explicit Free Space Representation for Fast and Accurate Mobile Robot Motion Planning[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3553-3560.

[8] Han L, Zheng T, Zhu Y, Xu L and Fang L. Live Semantic 3D Perception for Immersive Augmented Reality[J]. Transactions on Visualization and Computer Graphics, 2020 26 (5):2012-2022.

[9] Rao Y, Lu J and Zhou J. Spherical Fractal Convolutional Neural Networks for Point Cloud Recognition[C]. IEEE/

CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019:452-460.

[10] Xu M, Ding R, Zhao H and Qi X. PAConv: Position Adaptive Convolution with Dynamic Kernel Assembling on Point Clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:3172-3181.

[11] Chen R, Han S, Xu J and Su H. Point-Based Multi-View Stereo Network[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1538-1547.

[12] Hu S, Liu Z, Guo M, et al. Subdivision-based Mesh Convolution Network[J]. ACM Transaction on Graphics, 2022, 4 (3):1-16.

[13] Brian C and Marc L. A volumetric method for building complex models from range images[C]. Conference on Computer graphics and interactive techniquesIn(SIGGRAPH), 1996:303-312.

[14] Chen J, Bautembach D, and Izadi S. Scalable real-time volumetric surface reconstruction[J]. ACM Transaction on Graphics, 2013, 32(4):1-16.

[15] Martens W, Poffet Y, et al. Geometric Priors for Gaussian Process Implicit Surfaces[J]. IEEE Robotics and Automation Letters, 2017, 2(2):373-380.

[16] Lee B, Zhang C, Huang Z. Online Continuous Mapping using Gaussian Process Implicit Surfaces[C]. International Conference on Robotics and Automation (ICRA), 2019: 6884-6890.

[17] Sitzmann V, et al. Implicit neural representations with periodic activation functions[C]. Advances in Neural Information Processing Systems(NeurIPS), 2020, 33.

[18] Sitzmann V, et al. Metasdf: Meta-learning signed distance functions[C]. Advances in Neural Information Processing Systems(NeurIPS), 2020.

[19] Liu S, Guo H, Pan H, et al. Deep Implicit Moving Least-Squares Functions for 3D Reconstruction[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:1788-1797.

[20] Huang J, S S Huang, H Song and S M Hu. DI-Fusion: Online Implicit 3D Reconstruction with Deep Priors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:8928-8937.

[21] Lipman Y. Phase Transitions, Distance Functions, and Implicit Neural Representations[C]. International Conference on Machine Learning(PMLR), 2021:6702-6712.

[22] Mildenhall B, Srinivasan P, Tancik M, Barron J, Ramamoorthi R, and Ng R. NeRF: Representing scenes as neural radiance fields for view synthesis[C]. European Conference on Computer Vision (ECCV), 2020:405-421.

[23] Raj A, et al. Pixel-aligned Volumetric Avatars[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:11728-11737.

[24] Liu L, et al. Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control[J]. ACM Transaction on Graphics, 2021, 40(6):1-16.

[25] Peng S, et al. Animatable Neural Radiance Fields for Modeling Dynamic Human Bodies[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 14294-14303.

[26] Gafni G, Thies J, Zollhöfer M and Nießner M. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:8645-8654.

[27] Pumarola A, Corona E, Pons-Moll G and Moreno-Noguer F. D-NeRF: Neural Radiance Fields for Dynamic Scenes [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:10313-10322.

[28] Park K, et al. Hyper NeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields [J]. ACM Transaction on Graphics, 2021, 40(6):1-12.

[29] Li Z, Niklaus S, Snavely N and Wang O. Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:6494-6504.

[30] Xian W, Huang J, Kopf J and Kim C. Space-time Neural Irradiance Fields for Free-Viewpoint Video[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 9416-9426.

[31] Gao C, Saraf A, Kopf J and Huang J. Dynamic View Synthesis from Dynamic Monocular Video[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5692-5701.

[32] Peng S, et al.Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 9050-9059.

[33] Martin-Brualla R, Radwan N, Sajjadi M, Barron J, Dosovitskiy A and Duckworth D. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 7206-7215.

[34] Zhang X, Srinivasan, Pratul P and Deng, Boyang and Debevec, Paul and Freeman, William T and Barron, Jonathan T. NeRFactor: neural factorization of shape and reflectance under an unknown illumination[J]. ACM Transactions on

Graphics, 2021, 40

[35] Boss M, Braun R, Jampani V, Barron J, Liu C and Lensch H. NeRD: Neural Reflectance Decomposition from Image Collections[C]. IEEE International Conference on Computer Vision (ICCV), 2021:12664-12674

[36] Srinivasan P, Deng B, Zhang X, Tancik M, Mildenhall B and Barron J. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:7491-7500

[37] Niemeyer M and Geiger A. GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:11448-11459.

[38] Yang B, et al. Learning Object-Compositional Neural Radiance Field for Editable Scene Rendering[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 13759-13768.

[39] Ost J, Mannan F, Thuerey N, Knodt J and Heide F. Neural Scene Graphs for Dynamic Scenes[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:2855-2864.

[40] Liu S, Zhang X, Zhang Z, Zhang R, Zhu J and Russell B. Editing Conditional Radiance Fields[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5753-5763.

[41] Schwarz K, Liao Y, Niemeyer M, and Geiger A. GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis [C]. Advances in Neural Information Processing Systems (NeurIPS), 2020:33.

[42] Liu L, et al. Sparse Voxel Fields[C]. Neural Information Processing Systems(NeurIPS), 2020

[43] Garbin S, Kowalski M, Johnson M, Shotton J and Valentin J. Fast NeRF: High-Fidelity Neural Rendering at 200FPS [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021:14326-14335.

[44] Hedman P, Srinivasan P, Mildenhall B, Barron J and Debevec P. Baking Neural Radiance Fields for Real-Time View Synthesis [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021:5855-5864.

[45] Yu A R Li, M Tancik, H Li, R Ng and A Kanazawa. PlenOctrees for Real-time Rendering of Neural Radiance Fields [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5732-5741.

[46] Rebain D, Jiang W, Yazdani S, Li K et al. DeRF: Decomposed Radiance Fields [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 14148-14156.

[47] Reiser C, Peng S, Liao Y and Geiger A. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021:14315-14325.

[48] Yu A, Ye V, Tancik M and Kanazawa A. pixelNeRF: Neural Radiance Fields from One or Few Images[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 4576-4585.

[49] Chibane J, Bansal A, Lazova V and Pons-Moll G. Stereo Radiance Fields (SRF): Learning View Synthesis for Sparse Views of Novel Scenes[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 7907-7916.

[50] Michael N, et al. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022

[51] Barron J, et al. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022

[52] Thomas M, et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding[OL]

[53] Matthew T. Block-NeRF Scalable Large Scene Neural View Synthesis[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022

[54] Wizadwongsa S, Phongthawee P, Yenphraphai J and Suwajanakorn S. NeX: Real-time View Synthesis with Neural Basis Expansion[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 8530-8539.

[55] Vincent S et al. Light Field Networks[C]//2021 Neural Information Processing Systems(NeurIPS), NIPS, 2021

[56] Attal B and Huang J. et al. Learning Neural Light Fields with Ray-Space Embedding Networks[J]. arXiv preprint, 2021,2112.01523

[57] Chen A, et al. MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021:14104-14113.

[58] Wei Y, Liu S, Rao Y, Zhao W, Lu Jand Zhou J. NerfingMVS: Guided Optimization of Neural Radiance Fields for Indoor Multi-view Stereo[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5590 -5599.

[59] Lin C, Ma W, Torralba A and Lucey S. BARF: Bundle-Adjusting Neural Radiance Fields[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5721-5731.

[60] Xu Q, et al. Point-NeRF: Point-based Neural Radiance Fields[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[61] Thomas N et al. DO NeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields using Depth Oracle Networks[J]. Computer Graphics Forum, 2021, 40

[62] Kangle Deng, Andrew Liu, Jun Yan Zhu, and Deva Ramanan. Depth-supervised NeRF: Fewer Views and Faster Training for Free[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[63] Benjamin A, et al. TöRF: Time-of-Flight Radiance Fields for Dynamic Scene View Synthesis[C]. Neural Information Processing Systems(NeurIPS), NIPS, 2021

[64] Jain A, Tancik M and Abbeel P. Putting NeRF on a Diet: Semantically Consistent Few-Shot View Synthesis[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021:5865-5874.

[65] Zhi S, et al. In-Place Scene Labeling and Understanding with Implicit Scene Representation[C].IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 15818-15827.

[66] Chen Y, et al. Sem2NeRF: Converting Single-View Semantic Masks to Neural Radiance Fields[OL]. https://deepai.org/publication/sem2nerf-converting-single-view-seman - tic-masks-to-neural-radiance-fields

[67] Bae G, Budvytis I, and Cipolla R. Multi-View Depth Estimation by Fusing Single-View Depth Probability with Multi-View Geometry[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[68] Miangoleh S, Dille S, Mai L, Paris S and Aksoy Y. Boosting Monocular Depth Estimation Models to High-Resolution via Content-Adaptive Multi-Resolution Merging[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 9680-9689.

[69] Liu C, Gu J, Kim K, Narasimhan S and Kautz J. Neural RGB®D Sensing: Depth and Uncertainty From a Video Camera[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 10978-10987.

[70] Chen X, Wang Y, Chen X and Zeng W. S2R-DepthNet: Learning a Generalizable Depth-specific Structural Representation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 3033-3042.

[71] Huynh L. et al. Monocular Depth Estimation Primed by Salient Point Detection and Normalized Hessian Loss[C]. International Conference on 3D Vision (3DV), 2021: 228-238.

[72] Godard C, Aodha O, Firman M and Brostow G. Digging Into Self-Supervised Monocular Depth Estimation[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3827-3837.

[73] Guizilini V, Ambruș R, Pillai S, Raventos A and Gaidon A. 3D Packing for Self-Supervised Monocular Depth Estimation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 2482-2491.

[74] Takmaz A, Paudel D, Probst T, Chhatkuli A, Oswald M and Gool L. Unsupervised Monocular Depth Reconstruction of Non-Rigid Scenes[C]. International Conference on 3D Vision (3DV), 2021: 825-836.

[75] Cao A and Charette R. MonoScene: Monocular 3D Semantic Scene Completion[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[76] Li J, Feng Z, She Q, Ding H, Wang C and Lee G. MINE: Towards Continuous Depth MPI with NeRF for Novel View Synthesis[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 12558-12568.

[77] Chen Y, Liu S and Wang X. Learning Continuous Image Representation with Local Implicit Image Function[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 8624-8634.

[78] Ben M, et al. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022.

编辑:龙学锋