

神经辐射场应用于大规模实景三维场景可视化研究进展

赵强^{1,2}, 余江峰¹, 万奇峰¹, 贺丽霞¹, 李思睿¹, 吴双品¹

1. 南京大学 地理与海洋科学学院 自然资源部国土卫星遥感应用重点实验室 江苏省地理信息技术重点实验室, 南京 210023;

2. 商丘师范学院 测绘与规划学院 河南省农业遥感大数据发展创新实验室 河南省黄河故道生态保护与治理工程技术研究中心, 商丘 476000

摘要: 地理实景三维场景是重要的国家数字基础设施, 其将地理信息从传统二维平面扩展到信息更丰富更全面的三维空间, 数据以显式三维模型的形式存储表达。然而, 经典的显式三维模型具有数据量大、可视化效果粗糙等问题, 在一定程度上限制了实景三维模型的实际应用。神经辐射场 NeRF (Neural Radiance Field) 是一种基于神经隐式立体表达 (Neural Implicit Volume Representing) 进行可微渲染 (Differentiable Rendering) 以实现高质量视图合成的新方法, 由 Mildenhall 等 (2020) 首次提出, 以其逼真的视图合成效果与新颖的实现方式成为计算机视觉领域的热点研究方向。自 NeRF 提出以来, 国内外爆发式涌现出大量有关神经辐射场的研究文献, 主要聚集于可视化效果的生成方法研究, 兼有少量将其用于大规模实景三维场景可视化研究探索。本文回顾了神经辐射场提出的背景, 概述了神经辐射场及其在大规模实景三维可视化方面的研究进展, 分析了目前利用神经辐射场进行大规模实景三维场景可视化研究中被关注的无边界场景、锯齿效果、瞬态遮挡、光度一致性、场景重照明与可见性场等问题, 指出了目前研究在多源数据融合、视觉效果优化、虚拟环境感知等方面面临的挑战, 对未来值得进一步深入探索的方向进行了展望。

关键词: 遥感, 神经辐射场, 视图合成, 隐式立体表达, 计算机视觉, 虚拟地理环境

中图分类号: TP391.9/TP18/P2

引用格式: 赵强, 余江峰, 万奇峰, 贺丽霞, 李思睿, 吴双品. 2024. 神经辐射场应用于大规模实景三维场景可视化研究进展. 遥感学报, 28(5): 1242-1261

Zhao Q, She J F, Wan Q F, He L X, Li S R and Wu S P. 2024. Progress in neural radiance field and its application in large-scale real-scene 3D visualization. National Remote Sensing Bulletin, 28(5): 1242-1261 [DOI: 10.11834/jrs.20232536]

1 引言

图像是对现实三维世界的二维投影, 具有易获取、细节丰富、效果逼真等优点。从二维图像中获取三维信息是计算机视觉研究的热点方向。基于二维图像重建三维几何模型是实现三维可视化的经典途径, 利用视图合成技术模拟三维漫游效果是实现三维空间可视化表达的又一重要方法。

随着神经网络与计算机视觉的深度融合, 基于神经隐式立体表达 (Neural Implicit Volume

Representing) 的视图合成技术以其产生的逼真三维效果备受关注。神经辐射场 NeRF (Neural Radiance Field) (Mildenhall 等, 2020) 使用神经隐式立体表达技术, 基于二维图像生成三维空间任意视点的观察结果视图, 可实现三维场景自由漫游效果的模拟, 在无需重建三维模型的情况下即能生成可连续自由观察的照片级逼真三维效果, 是对借助于几何模型实现三维可视化的传统技术路线的颠覆。

2022年5月 Google 全球开发者大会上, Google

收稿日期: 2022-10-13; 预印本: 2023-02-16

基金项目: 国家自然科学基金(编号:41871293)

第一作者简介: 赵强, 研究方向为虚拟地理环境与计算机视觉。E-mail: giser.zq@smail.nju.edu.cn

通信作者简介: 余江峰, 研究方向为虚拟地理环境与计算机视觉。E-mail: GISJF@nju.edu.cn

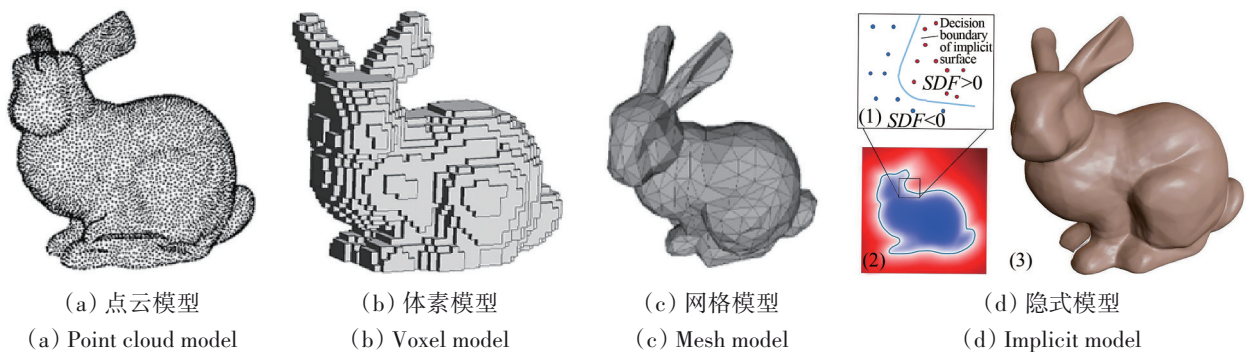
发布了沉浸式实景地图的最新研究成果：区别于传统借助于三维模型的可视化（龚建华等，1999）渲染路线，其基于机器学习与神经渲染（Neural Rendering）技术，利用伦敦市区数十亿张无人机航空影像、街景图像，以及室内图像，构建了世界上第一个基于视图合成技术的城市级大规模实景三维场景，可支持用户在不同天气和光照条件下的连续自由漫游，能够让用户体验照片级逼真的三维视觉效果，但其是否能够满足生产级的需要尚需进一步验证和探索。基于视图合成技术生成的大规模实景三维场景以其逼真的效果和新颖的路线引起学术界和产业界的极大关注，在实景三维建设的相关研究方面翻开了新的篇章。

2 神经辐射场(NeRF)

2.1 视图合成技术的发展

近几十年来，利用二维图像进行三维模型重建的研究一直备受关注（Früh和Zakhor，2004；

吴军等，2006；Agarwal等，2011；Raguram等，2011；颜深等，2021），出现了很多经典方法，例如：运动结构恢复SFM（Structure From Motion）（李德仁等，2016；Zhu等，2018；Schönberger和Frahm，2016；Özyeşil等，2017）、即时定位与地图构建SLAM（Simultaneous Localization and Mapping）（Davison等，2007；Engel等，2014；Mur-Artal等，2015）与光度误差法（Bundle Adjustment）（Newcombe等，2011）等。此类方法的基本思想都是基于二维图像重建显式三维模型（图1（a）点云模型、图1（b）体素模型与图1（c）网格模型）。显式模型由离散元素构成，导致三维模型表面不连续，模型数据量大且外观较粗糙，重建过程复杂且具有较高的生产成本。如何在同样使用二维图像作为信息源的前提下，摆脱对复杂的三维几何模型重建过程的依赖，且能获得与显式模型表现一致的三维可视化效果，是探索基于二维图像实现三维空间可视化表达的又一重要研究方向。



(1) 隐式立体表面 $SDF=0$, 外部 $SDF>0$, 内部 $SDF<0$; (2) 有符号距离场的二维剖面; (3) 从 $SDF=0$ 恢复渲染的三维隐式立体
 (1) Implicit stereo surface $SDF=0$, outside $SDF>0$, inside $SDF<0$; (2) 2D cross-section of the signed distance field;
 (3) Rendered 3D implicit stereo is recovered from $SDF=0$

图1 显式模型(Hoang等,2019)与隐式模型(Park等,2019)比较

Fig.1 Comparison of explicit models (Hoang et al., 2019) and implicit model (Park et al., 2019)

视图合成技术可在稀疏采样获取的二维图像基础上，无需重建三维模型即可生成不同于输入视图的任意新视点视图，模拟出三维空间的自由表达效果（Shan等，2013）。早期的视图合成方法以简单的图像拼接为主（Szeliski和Shum，1997；常远和盖孟，2021），此类方法要求输入包含场景重叠内容的大量图像，且合成的新视点非常有限。（Buehler等，2001；Hedman等，2018）将三维重建模型与视图合成相结合，先重建目标对象的点云或网格模型，再进行新视角重投影，实现新视

图合成效果，此类方法本质上仍依赖三维几何模型的重建，且整个过程计算量非常大。同一阶段，也出现了另一类针对目标对象立体表面进行三维外观模拟的视图合成方法，此类方法不依赖三维模型重建，且获得的视图合成效果相较于先前介绍的方法更逼真，如（Levoy和Hanrahan，1996；Davis等，2012）基于密集采样的图像，使用光场采样插值（Light Field Sample Interpolation）等简单方法即可合成逼真的新视图，但该方法要求非常大的源数据量。也有研究针对稀疏采样的图像源

数据进行基于立体表面模拟的视图合成方法，这些方法可分为两类：一类基于网格进行场景表示 (Loper 和 Black, 2014; Li 等, 2018; Chen 等, 2019)，利用梯度下降的方法实现新视图合成，但此类方法视觉干扰较多且不适用于无约束的户外现实场景；另一类基于立体表达 (Volumetric Representation) 技术进行复杂形状和纹理的表示，从使用图像直接为体素网格着色 (Szeliski 和 Golland, 1998; Seitz 和 Dyer, 1999; Kutulakos 和 Seitz, 2000) 到使用大型数据集训练深度网络 (Choi 等, 2019; Tucker 和 Snaveley, 2020; Penner 和 Zhang, 2017; Zhou 等, 2018)，此类视图合成方法的研究正取得越来越好的效果。

神经隐式立体表达方法通过神经网络以隐函数的方式计算目标三维立体对象 (图 1 (d))，再进行渲染以实现三维立体表达效果。区别于显式模型的离散表达原理，隐式立体表达实现了三维对象表面的连续逼真效果。神经隐式立体表达技术发展的初始阶段，主要是将目标对象的三维坐标映射到有符号距离函数 (Park 等, 2019; Curless 和 Levoy, 1996; Jiang 等, 2020; Takikawa 等, 2021) 或占用场的深度网络 (Occupancy Networks) (Peng

等, 2020; Genova 等, 2020; Mescheder 等, 2019)，以实现目标对象三维形状的隐式表达。此类研究针对几何结构简单的三维对象可取得较好的效果，但对于现实环境中复杂的三维对象来说，其效果的逼真度远不如基于网格或体素的离散模型。Niemeyer 等 (2020)、Sitzmann 等 (2019)、Lombardi 等 (2019) 通过优化神经隐式立体表达的可微渲染 (Differentiable Rendering) 函数，逐步支持具有更高复杂度的目标对象。总体而言，这些技术具有表示复杂几何立体的能力，但仅适用于室内小规模场景或实验室模型，对于复杂的现实三维环境尚无法取得理想的逼真效果。

2.2 NeRF 的提出

NeRF 利用稀疏采样的普通 RGB 图像作为监督信息，通过神经网络迭代拟合出隐式表达函数以神经网络参数的形式对隐式立体连续编码，通过对隐式立体采样，渲染高分辨率的几何和外观，能呈现出复杂对象逼真的高分辨率新视图 (图 2)。NeRF 以其逼真的视图合成效果、简单的结构和较少的存储成本，自 2020 年被提出以来即备受关注，成为实现神经隐式立体表达和新视图合成算法的经典代表 (Shen 等, 2021a)。

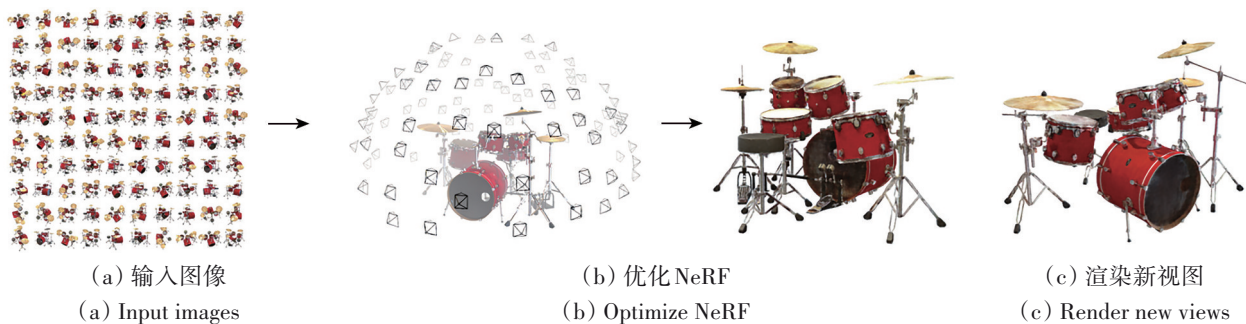


图 2 神经辐射场概念 (Mildenhall 等, 2020)
Fig. 2 Concept of NeRF (Mildenhall et al., 2020)

NeRF 将连续的三维场景表示为一个 5D 向量值函数 (图 3)，每个输入的 5D 坐标包括目标空间点的位置向量 $\mathbf{x} = (x, y, z)$ 和观察视角方向向量 $\mathbf{d} = (\theta, \varphi)$ ，将观察射线上目标空间点的 5D 坐标值输入多层感知机 MLP (Multilayer Perceptron) 网络，利用 MLP 网络 $F_{\theta}(\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$ 近似表达连续的 5D 场景，并优化权重 θ 以映射出目标空间点向观察射线逆方向发射的颜色向量 $\mathbf{c} = (r, g, b)$ 和立体密度 σ 。假设观察光线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ ，近边

界 t_n ，远边界 t_f ，则颜色值计算公式 (Mildenhall 等, 2020) 如下：

$$C(r) = \int_{t_n}^{t_f} e^{-\int_{t_n}^t \sigma(\mathbf{o} + s\mathbf{d}) ds} \cdot \sigma(\mathbf{o} + t\mathbf{d}) \cdot \mathbf{c}(\mathbf{o} + t\mathbf{d}, \mathbf{d}) dt \quad (1)$$

NeRF 基于经典的立体渲染算法将 MLP 计算输出的颜色和密度累积投影到图像中，因为立体渲染函数是可微分的，所以实现了移动视点的三维立体连续观察效果。NeRF 以目标空间点位置和观察方向作为输入，计算输出目标空间点的密度和

颜色以表达三维场景。在对复杂场景进行高分辨率建模时, 与传统三角网格或体素网格等离散模

型相比, 不仅可产生更好的渲染效果, 同时也可极大降低存储成本。

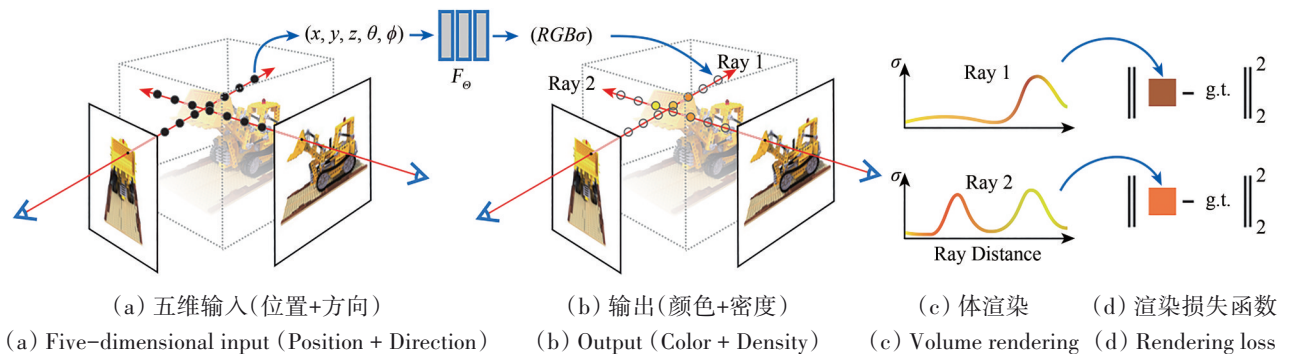


图3 神经辐射场算法(Mildenhall等,2020)

Fig.3 Algorithm of NeRF (Mildenhall et al., 2020)

2.3 NeRF面临的挑战

神经辐射场实现了效果逼真的复杂三维场景新视图合成, 但仍然面临源数据要求高、不支持动态对象、渲染效率低, 以及渲染对象类型单一等挑战。

(1) 源数据要求高。NeRF要求输入图像必须具有相机参数, 这限制了可用图像的范围, 有学者针对扩展NeRF数据输入类型进行相关研究(表1)。Wang等(2022b)、Lin等(2021)、Yen-Chen等(2021)、Meng等(2021)提出了对普通无序图像自动发现其相机参数的方法, 但增加了额外的计算负担。Deng等(2022)、Roessle等(2022)提出可以降低源图像的数量。Jang和Agapito(2021)、Gao等(2021)、Rematas等(2021)、Li等(2021a)甚至使用单目图像或单目视频(Li等, 2021b; Tretschk等, 2021; Su等, 2021; Xian等, 2021)实现NeRF, 但视图输入信息的减少, 导致输出效果变差。Martin-Brualla等(2021)、Boss等(2021a)、Pearl等(2022)、Mildenhall等(2022)尝试使用不同曝光效果图像或模糊图像(Ma等, 2022)。Dey等(2022)将深度图与彩色图像结合, Xu等(2022a)将三维点云数据或Wang等(2022c)、Luo等(2021)使用特殊硬件系统获取的图像作为NeRF的数据源, 这些数据获取难度相对较大, 相关研究取得的实际效果有限, 但其重要意义在于针对不同质量的源数据输入促进了NeRF鲁棒性的提升。

表1 不同类型数据的针对性算法

Table 1 Targeted algorithms for different types of data		
数据特征	算法特点	算法简称及文献来源
图像无位姿信息或位姿信息不准确		NeRF——(Wang等, 2022b)
	训练或优化相机参数	BARF(Lin等, 2021)
		INeRF(Yen-Chen等, 2021)
稀疏图像	几何感知生成模型	GNeRF(Meng等, 2021)
	语义一致性损失	NeRF-VAE(Kosiorek等, 2021)
	深度	DietNeRF(Jain等, 2021)
	计算或引入辅助数据	DS-NeRF(Deng等, 2022)
多视图图像	深度	(Roessle等, 2022)
	点云	Point-NeRF(Xu等, 2022a)
	多源	CG-NeRF(Jo等, 2023)
单目图像	文本	CLIP-NeRF(Wang等, 2022a)
		CodeNeRF(Jang和Agapito, 2021)
		Portrait-NeRF(Gao等, 2021)
		ShaRF(Rematas等, 2021)
单目视频		MINE(Li等, 2021a)
		(Li等, 2021b)
		NR-NeRF(Tretschk等, 2021)
低质量图像		A-NeRF(Su等, 2021)
		(Xian等, 2021)
	非结构化图像重建	NeRF-W(Martin-Brualla等, 2021)
	不同照明条件	NeRD(Boss等, 2021a)
	噪声图像处理	NAN(Pearl等, 2022)
专用拍摄设备	黑暗中拍摄的图像	RawNeRF(Mildenhall等, 2022)
	散焦或模糊图像	Deblur-NeRF(Ma等, 2022)
	球面反射镜阵列	MirrorNeRF(Wang等, 2022c)
	多视图获取系统	ConvNeRF(Luo等, 2021)

基于机器学习的生成式网络模型 (Generative Network Model) (Goodfellow 等, 2020; Kingma 和 Welling, 2022; Radford 等, 2021) 根据输入条件生成新图像, NeRF 与生成式网络的结合分两类, 一类涉及扩展数据辅助视图合成 (Jo 等, 2023)、稀疏视图重建 (Kosiorek 等, 2021; Jain 等, 2021) 和基于自然语言或示例图像的图像生成 (Wang 等, 2022a) 等源数据扩展方向; 另一类则将 NeRF 应用于生成式模型的三维感知图像生成式算法 (Niemeyer 和 Geiger, 2021; Chan 等, 2021; Schwarz 等, 2021), 属于对 NeRF 的应用研究。NeRF 与生成式网络模型的结合进一步提高了 NeRF 的智能化与自动化, 扩展了其应用场景。

(2) 静态对象限制。原始 NeRF 仅支持静态内容, 有研究对 NeRF 进行扩展以支持动态对象

(表 2)。动态是一个时间维度上的感知状态, (Pumarola 等, 2021; Fang 等, 2022; Xu 等, 2021; Guo 等, 2022a; Attal 等, 2021) 增加时间作为输入参数构建时空 NeRF, 对 NeRF 的源数据提出了包含连续时间信息的要求, 是当前实现动态 NeRF 的主要技术路线。除了增加时间信息以外, Liu 等 (2022)、Zhao 等 (2022)、Park 等 (2021) 基于变形场 (Deformation Field), Athar 等 (2021)、Kania 等 (2022)、Hong 等 (2022)、Zhuang 等 (2022) 增设变形参数, 以及 Niemeyer 和 Geiger (2021)、Wu 等 (2022b) 将前景动态对象与静态背景分离等方法都实现了 NeRF 对动态内容的扩展支持, 但这些方法都有具体的适用场景限制, 在普适性方面存在很大不足。

表 2 动态对象支持算法

Table 2 Dynamic object support algorithms

算法分类	重要技术点	算法简称及文献来源
时空 NeRF	将时间视为附加输入	D-NeRF(Pumarola 等, 2021)
	网络中增强时间信息	TiNeuVox(Fang 等, 2022)
	图像随时间的一致融合	H-NeRF(Xu 等, 2021)
	时间作为参数参与计算	NDVG(Guo 等, 2022a)
变形场	利用连续波飞行时间相机图像优化	TöRF(Attal 等, 2021)
	4D 体素变形场	DeVRF(Liu 等, 2022)
	非刚性变形场	HumanNeRF(Zhao 等, 2022)
变形参数	参数化变形场	HyperNeRF(Park 等, 2021)
	参数控制表情	FLAME-in-NeRF(Athar 等, 2021)
	属性视为潜在变量	CoNeRF(Kania 等, 2022)
动静分离	参数化头部模型	HeadNeRF(Hong 等, 2022)
	参数化面部外观及表情	MoFaNeRF(Zhuang 等, 2022)
	背景中分离对象	GIRAFFE(Niemeyer 和 Geiger, 2021)
场景编辑	动态对象与静态背景分离	D ² NeRF(Wu 等, 2022b)
	2D 图像颜色或形状编辑传播到 3D 空间	(Liu 等, 2021b)
	特征向量移动实现编辑效果	DIVeR(Wu 等, 2022a)
	双路径架构, 场景与对象单独编码	(Yang 等, 2021)
生成式网络模型	表面对齐 NeRF, 对人体体型与服装控制	Surface-Aligned NeRF(Xu 等, 2022b)
	改变潜在代码简单编辑对象形状、纹理和视点	CodeNeRF(Jang 和 Agapito, 2021)
	3D 感知生成器	FENeRF(Sun 等, 2021)
	基于 CLIP 模型的可操控能力	CLIP-NeRF(Wang 等, 2022a)

有研究针对目标对象的编辑操作开展探索。此类工作主要有修改神经网络输入参数 (Liu 等, 2021b; Wu 等, 2022a; Yang 等, 2021; Xu 等, 2022b) 和潜在代码 (Latent Codes) (Jang 和

Agapito, 2021), 与基于生成式网络模型 (Sun 等, 2021; Wang 等, 2022a) 两种技术路线, 前者是将编辑信息以参数形式传入网络, 修改体绘制积分, 后者是将编辑信息传入生成式网络, 基于三维感

知生成编辑后的新视图。目前,这类研究能对简单的对象形状和颜色进行编辑操作,还无法实现对NeRF场景的自由编辑。

(3) 渲染效率低。NeRF对于每条观察射线的渲染都需要查询MLP数百次,导致渲染效率较低。渲染效率的提升是NeRF亟待解决的问题。根据文献分析,目前算法效率的提升方法主要包括3类(表3):1)针对隐式立体进行优化,以体素模型(Wu等,2022a;Hedman等,2021;Liu等,2021a;Yu等,2021;Guo等,2022a)、深度信息(Lin等,2022),以及点云(Xu等,2022a)等显式数据或结构对隐式立体进行预估,通过降低对隐式立体生成的计算量以提高渲染效率。此类方

法以显式模型为先验,增加了数据或计算成本,且对无边界场景(Unbounded Scene)(见3.3节)支持效果较差。2)仅对目标对象表面可见有效点优化采样(Piala和Clark,2021;Neff等,2021;Kondo等,2021),减少无效采样计算。此类方法对渲染对象的几何精度要求较高,不适用于现实场景;3)基于并行计算思想,将整个目标场景分解成若干个子场景(Rebain等,2021;Reiser等,2021),再为每个子场景指定较小的网络,子网络的数据量比整个场景显著减少且支持并行计算,整体训练效率明显提升。此类方法对于现实无边界场景中的子场景划分,需要平衡效果与成本开销,相关目标的达成还有待进一步探索。

表3 渲染效率提升算法

Table 3 Rendering efficiency improvement algorithms

算法分类	重要技术点	算法简称及文献来源
体素模型 引入显式模型数据	基于体素的特征场	DIVeR(Wu等,2022a)
	稀疏神经辐射场网络	sNeRG(Hedman等,2021)
	神经稀疏体素场	NSVF(Liu等,2021a)
	体素八叉树结构	PlenOctrees(Yu等,2021)
	体素网格优化的快速可变形辐射场	NDVG(Guo等,2022a)
深度	深度信息	ENeRF(Lin等,2022)
点云	神经点云	Point-NeRF(Xu等,2022a)
优化采样	相机光线取最可能影响外观的位置映射采样	TermiNeRF(Piala和Clark,2021)
	深度预言机网络预测每个视图光线的样本位置	DONeRF(Neff等,2021)
	与视觉外壳结合仅对边界体积内均匀密集采样	VaxNeRF(Kondo等,2021)
场景分解	基于空间分解场景	DeRF(Rebain等,2021)
	场景细分为3D网格,微型MLP训练一个单元格	KiloNeRF(Reiser等,2021)

(4) 渲染对象类型单一。NeRF对试验小场景渲染效果较好,而现实场景内容复杂,特殊类型对象的逼真渲染需要特别的处理(表4)。一些研究工作将NeRF与多视图立体(Wei等,2021;Chen等,2021;Rosu和Behnke,2022)、几何推理器(Geometry Reasoner)(Johari等,2022)等技术方法相结合,以预估的几何先验结果来提高隐式立体精度,或者训练神经网络从二维图像直接获取三维信息(Trevithick和Yang,2021;Xie等,2021),以上这些方法均对真实世界无边界场景支持较差。Huang等(2022b)、Nguyen-Phuoc等(2022)开展对场景特定风格迁移的渲染研究,增强了NeRF对不同类型对象的渲染支持,但其渲染效率低是待解决的问题。

传统三维重建技术针对透射、反射、边缘模

糊对象,以及流体对象等特殊场景对象的重建效果非常差,NeRF对于此类特殊对象的表现及效果上相比传统三维重建技术具有优势。Verbin等(2022)利用正则化法向量方法实现较逼真的反射效果,但计算量有增多。Guo等(2022b)将场景中的透射与反射分割成两个分量,取得较好的效果,但仅适用于视点固定的平面反射。Wang等(2022d)构建了神经折射场,利用物理折射原理计算模拟流体折射效果,但未考虑液体表面流动和视点位置移动的情况。Luo等(2021)利用不透明监督的方法实现对现实世界中头发、羽毛、树叶等细微局部透明效果的对象重建,但该算法依赖精确的相机姿态和高质量的图像,难以扩展到野外获取的数据中。

表4 特殊对象类型可视化效果改善渲染算法

Table 4 Rendering algorithms for improving visualization of special object types

算法分类	重要技术点	算法简称及文献来源
优化采样	SfM重建和基于学习的先验优化隐式体积	NerfingMVS(Wei等,2021)
	平面扫描成本体进行几何感知场景推理	MVSNeRF(Chen等,2021)
	3D场景几何与高分辨率图像恢复为距离函数	NeuralMVS(Rosu和Behnke,2022)
几何推理器	几何推理器为源视图构建级联成本量,渲染器推断形状与外观	GeoNeRF(Johari等,2022)
神经网络从2D图像直接学习3D信息	将3D几何与外观建模为神经网络,学习2D源图像中像素特征投影到3D点	GRF(Trevithick和Yang,2021)
	图像集中学习高质量3D对象类别模型	FIG-NeRF(Xie等,2021)
风格迁移	预训练标准NeRF,将颜色预测模块替换为风格网络	StylizedNeRF(Huang等,2022b)
	交替NeRF与风格化优化步骤	SNeRF(Nguyen-Phuoc等,2022)
特殊光学效果	法向量正则化器	Ref-NeRF(Verbin等,2022)
	场景拆分为透射与反射分量,分别建模	NeRFReN(Guo等,2022b)
	恢复具有相应密度场的体积法线场	NeReF(Wang等,2022d)
	显式不透明度监督与卷积机制结合到NeRF	ConvNeRF(Luo等,2021)

综上,目前NeRF相关研究多基于理想的实验室环境或人工合成的规则数据,针对复杂现实世界实景三维场景可视化的研究工作相对较少,其实际应用仍面临较多困难。

3 隐式模型实现大规模实景三维可视化

3.1 大规模场景NeRF

由于NeRF的实现原理、方法、难度、规模,以及现阶段计算机硬件性能的制约,目前将NeRF应

用于大规模实景三维可视化研究相对较少(表5)。

S-NeRF(Derksen和Izzo,2021)算法利用已知位姿的多视图高分辨率卫星影像进行训练,模拟直接或间接光源,基于阴影对地面建筑物进行三维形状估计,实现视图合成,但其在模型泛化性以及光照的处理仍具有较大的不足。Sat-NeRF(Marí等,2022)是一种适用于多时间多视图卫星影像的NeRF变体方法,以学习静态地物的几何和外观为主,可以处理阴影及瞬态对象(Transient Objects),其生成的表面模型在局部不规则性上仍需进一步优化。

表5 大规模场景NeRF算法

Table 5 NeRF for large-scale scenes algorithms

数据源	重要技术点	算法简称及文献来源
卫星影像	模拟直接或间接光源,基于阴影进行三维形状估计	SNeRF(Derksen和Izzo,2021)
	以有理多项式系数函数为代表的卫星相机模型与NeRF结合	Sat-NeRF(Marí等,2022)
航空影像	将训练图像划分为可并行训练的子模块	Mega-NeRF(Turki等,2022)
	渐进式增长模型与训练集	BungeeNeRF(Xiangli等,2022)
室内模型	多细节级别场景	Recursive-NeRF(Yang等,2023)
	场景以细节级别的方式呈现	
街景	结合全景图与激光雷达数据重建街景场景	URF(Rematas等,2022)
	利用NeRF场景几何训练相机姿态回归模型	LENS(Moreau等,2021)
	ω -net渲染网络从稀疏点云中学习神经描述符	READ(Li等,2022)

Mega-NeRF(Turki等,2022)首先分析大场景的可见性,然后将场景划分为可以并行训练的子模块,对子模块的并行训练实现了大规模场景的快速渲染。但其要求源图像具有较准确的位姿信息,且尚未实现实时交互。BungeeNeRF

(Xiangli等,2022)将NeRF的场景从单一细节层次扩展到多细节层次(图4),使用不同细节层次的图像进行逐级训练,在观察过程中随着视点的距离变化,渲染不同细节层次的视图,实现了连续LOD(Level Of Detail)(Kada等,2015,2016;

Ohori等, 2015; She等, 2022)的效果。Recursive-NeRF (Yang等, 2023)以多层次级别场景渲染的形式,降低场景训练量与渲染成本,渲染的结果是各级神经网络结果的组合。性能是这类算法在大规模场景实际应用中最大的阻碍。

URF (Rematas等, 2022)基于城市街景的全景图和三维点云数据进行新视图的合成和准确的三维重建(图5),该算法受数据限制,适用场景规模有限。LENS (Moreau等, 2021)生成分布在场景中规则网格上的虚拟相机位置,使用合成的或真实的图像来训练相机姿态回归模型,以生成高质量的建筑物场景新视图。READ (Li等, 2022)从稀疏点云中学习神经描述符,合成逼真的驾驶

场景,还可以对场景进行拼接和编辑,但其对输入数据要求较高,有一定局限性。

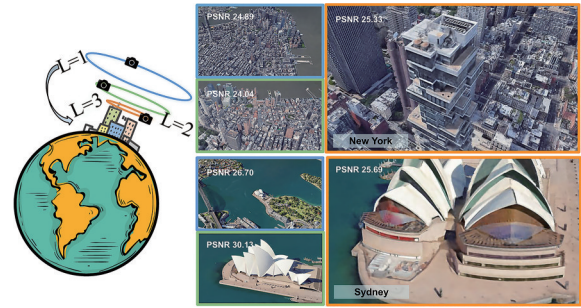


图4 BungeeNeRF 数据获取与不同级别渲染效果 (Xiangli等, 2022)

Fig.4 BungeeNeRF data acquisition and different levels of rendering effects (Xiangli et al., 2022)

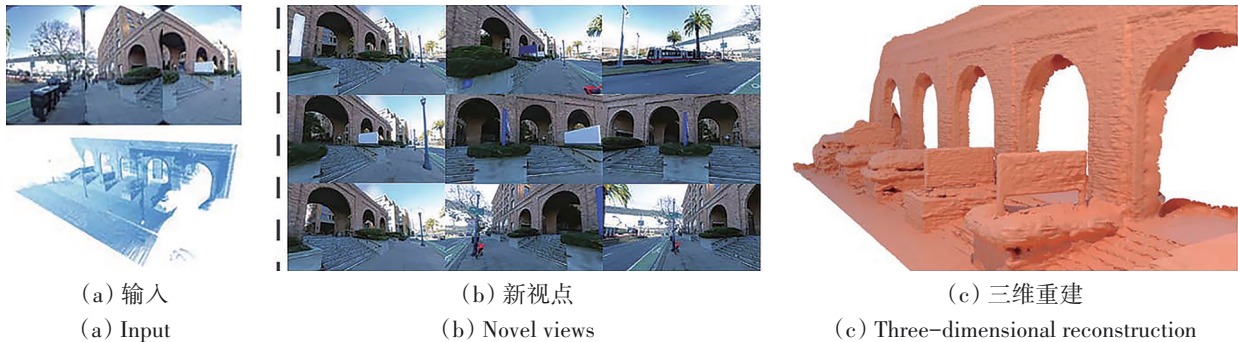


图5 基于全景图与点云的视图合成和三维重建(Rematas等, 2022)

Fig. 5 View synthesis and 3D reconstruction based on panoramas and point clouds (Rematas et al., 2022)

3.2 Block-NeRF

2022年2月加利福尼亚大学伯克利分校、Waymo与Google等多家单位联合(Tancik等, 2022)针对大规模场景发布了基于神经辐射场的变体算法Block-NeRF,将大规模场景按设定的规则分割为规模较小的局部子场景Block-NeRF,这些Block-NeRF可以独立且并行地训练,再以平滑的方式进行组合,实现场景连续性。Block-NeRF可以通过添加子场景实现城市级规模大场景的构建(图6)。这种结构不受场景规模限制,在添加新子场景时无需重新训练整个场景。

Block-NeRF的实现思想包括场景分割、选择合并、处理优化3部分:

(1) 场景分割。Block-NeRF介绍了两种场景分割算法,第一种分割方法以旧金山阿拉莫广场社区为例,将整个社区场景以道路为网络,每个路口的中心为单个Block-NeRF的圆心(图6),相

邻两个路口之间道路长度的75%为半径设置独立Block-NeRF的范围,可以保证相邻两个路口之间不低于50%的场景重叠率,必要时增加额外独立训练的Block-NeRF子场景块作为两个路口场景的过渡,使相邻两个Block-NeRF的外观表现更接近。据此原则,每个独立Block-NeRF所表达的范围半径是可变的,可根据基础地理数据(如OpenStreetMap (Haklay和Weber, 2008))进行自动分割。第二种分割方法(图7)是以固定距离为半径,将子场景块的圆心放置在街道中心线上,相邻场景块之间不低于50%的重叠度进行分割。

(2) 子场景处理:获取的室外实景源图像受时间、天气、拍摄角度等因素影响,其光照效果不一致。Block-NeRF遵循NeRF-W (Martin-Brualla等, 2021)的原理并优化外观嵌入(Appearance Embeddings)向量,允许Block-NeRF解析光照等外观变化的条件,操纵这些外观嵌入因素,在训练数据中的不同条件之间进行插值(例如天空的

阴晴(图8(a)、(b)),白天与夜晚(图8(c)),可实现渲染时对不同光照条件以及天气条件的自由控制。



图6 基于Block-NeRF构建旧金山阿拉莫广场社区实景三维场景(Tancik等,2022)

Fig.6 Construction of real three-dimensional scene of the Alamo Square neighborhood in San Francisco based on Block-NeRF (Tancik et al., 2022)

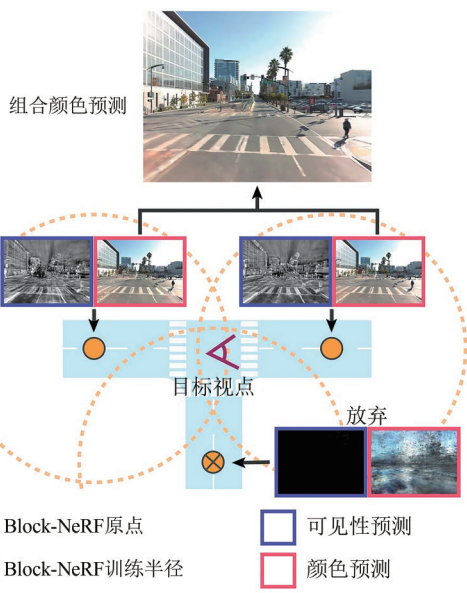


图7 目标视图与场景分割(Tancik等,2022)

Fig.7 Target view and scene segmentation(Tancik et al., 2022)



图8 外观代码控制场景表现不同的光照效果(Tancik等,2022)

Fig.8 Appearance codes control the scene to represent different lighting (Tancik et al., 2022)

在外观表现方面,Block-NeRF扩展了mip-NeRF(Barron等,2021)的模型,每个函数代表MLP的一层网络。在 f_σ 中预测空间位置 \mathbf{x} 的密度 σ ,输出一个特征向量,该向量与观察方向 \mathbf{d} 、曝光级别(Exposure Level)和外观嵌入相配合,这些因子都将被作为参数输入到颜色预测网络 f_c 中,最终计算输出RGB颜色值(图9)。对于场景中的动态对象,Block-NeRF使用语义分割模型Panoptic-deeplab(Cheng等,2020)生成常见移动对象的蒙版,在训练阶段忽略蒙版区域,以实现动态对象的遮蔽。

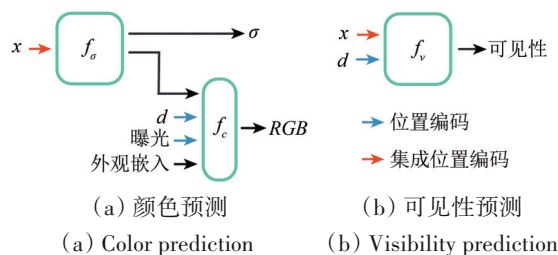


图9 颜色预测及可见性预测(Tancik等,2022)

Fig.9 Color prediction and visibility prediction (Tancik et al., 2022)

(3)子场景选择与合并:Block-NeRF增加了一个以视点位置 \mathbf{x} 和观察方向 \mathbf{d} 为输入的可见性网络 f_v (图9), f_v 返回该点的透射率,判断视点位置对于特定空间区域的可见性。透射率 T_i 的计算公式如下:

$$T_i = \exp\left(-\sum_{j < i} \Delta_j \sigma_j\right), \Delta_i = t_i - t_{i-1} \quad (2)$$

沿着观察方向自视点出发,自由空间中的点或第一个与对象表面相交的点 T_i 接近1,该对象内部或射线后方的点 T_i 接近0;如果一个点仅有部分视点可以观察到,则其透射率 T_i 为0—1,表明该点是部分可见。Block-NeRF对特定视点的可见性预测类似于NeRV(Srinivasan等,2021)算法中对可见性场的处理方法。当合并多个子场景Block-NeRF时,可见性判断不仅可以确定哪些子场景需要合并,而且还可以用于确定两个Block-NeRF之间进行外观匹配的位置。当选择的位置透射率很高时,可以快速优化目标场景的外观代码,以统一合并场景间的色彩、光照等全局属性,达到场景间的平滑过渡。

3.3 特殊问题处理

Block-NeRF 对 NeRF 应用于大规模实景三维场景建设的一些特殊问题给出了解决方案, 这些问题主要包括无边界场景、锯齿效果、光度一致性 (Photometric Consistency)、瞬态遮挡 (Transient Occluders)、重照明 (Relighting), 以及可见性场等。

(1) 无边界场景处理与 NeRF++。无边界场景指在户外可 360° 观察的无限空间场景。在这种室外大场景视图合成工作中, 视野边缘部分由于获取数据量小, 导致合成效果较差。NeRF++ (Zhang 等, 2020) 是一种针对大规模或无边界场景提高视图合成保真度的优化处理算法。形状—辐射歧义性 (Shape Radiance Ambiguity) 是 NeRF++ 中的一个重要的概念: 指在没有显式或隐式正则化的情况下, 基于一组图像利用与视图相关的辐射来拟合正确的外观细节 (图 10 (a))。当 NeRF 应用

于 360° 无边界场景时, 如果仅对局部前景建模, 背景会模糊 (图 10 (b) 红框), 对全局建模, 则整体细节会有损失 (图 10 (b) 橙框)。NeRF++ 将场景空间划分为内部球体和外部包围体 (图 10 (c)), 内部球体包含前景对象, 外部包围体包含除前景对象以外的部分, 两部分使用不同参数, 将前景与背景分离, 分别计算渲染, 再合成最终结果。但是, NeRF++ 的训练需要大量的时间与内存开销, 且对源数据位姿要求极高, 微小的相机标定误差就可能对合成效果。Turki 等 (2022) 引入一种简单的几何聚类算法, 在 NeRF++ 基础上对前景进一步分解形成若干空间单元, 每个单元学习单独的 NeRF, 可提高渲染速度以及保真度。Khademi 和 Ventura (2021) 基于 NeRF++, 使用曝光补偿参数来解释不同曝光条件的训练图像, 以渲染 360° 无边界场景的新视图。训练及渲染时间与内存开销大是这类方法共存的问题。

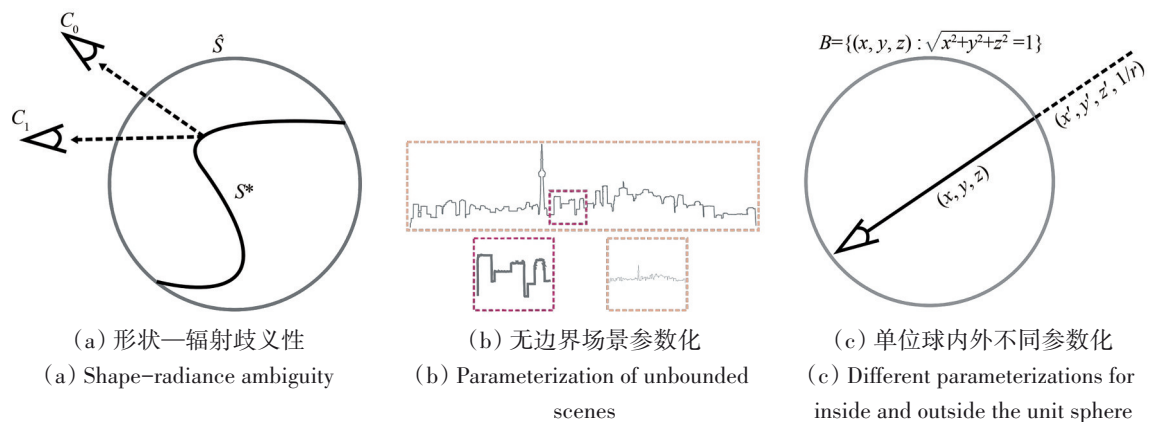


图 10 形状—辐射歧义性、无边界场景参数化以及单位球内外不同参数化 (Zhang 等, 2020)

Fig.10 Shape-radiance ambiguity, parameterization of unbounded scenes and different parameterizations for inside and outside the unit sphere (Zhang et al., 2020)

(2) 锯齿效果与 Mip-NeRF。NeRF 基于单射线对场景进行采样, 当图像分辨率不同时, 会产生模糊或锯齿效果, 超采样方法可以解决这个问题, 但会大幅增加计算成本。Mip-NeRF (Barron 等, 2021) 受计算机图形学中的 Mipmapping 方法启发, 使用预过滤策略, 将抗锯齿计算负担从渲染阶段转移到了预计算阶段。NeRF 沿着自相机投影中心出发通过每个像素追踪的射线对点 x 进行采样 (图 11 (a)), 使用位置编码 (PE) 获取点的特征 $\gamma(x)$ 。Mip-NeRF 以圆锥截头体代替射线 (图 11 (b)), 使用集成位置编码 (IPE) 对这些圆锥截头体进行特征化, 利用多元高斯函数逼近

截锥, 并在坐标位置编码上计算积分 $E[\gamma(x)]$ 。Mip-NeRF 通过渲染反锯齿圆锥台进行采样计算 (图 11 (c)), 显著提高了 NeRF 表示细节的能力。Mip-NeRF 较好地解决了 NeRF 中的锯齿效果问题, 但其在效率上仍具有进一步提高的空间。

在 360° 无边界场景中, 中心区域数据的信息量远大于边缘或远距离区域, 导致远、近场景在视图合成结果中产生较大的效果差异。(Barron 等, 2022) 基于扭曲的正则化来解决无边界场景中远景模糊的问题, 但会漏掉一些细节内容, 且处理时间过长。

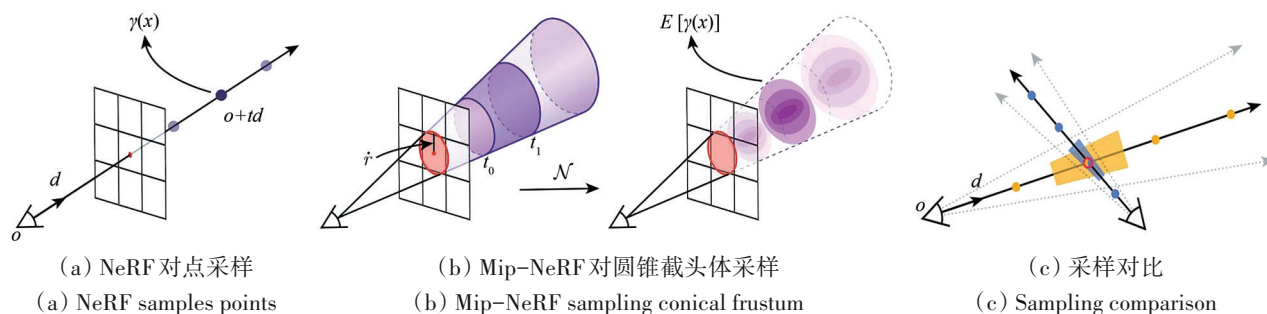


图 11 NeRF 对点采样、Mip-NeRF 对圆锥截头体采样以及采样对比(Barron 等, 2021)

Fig. 11 NeRF samples points, Mip-NeRF sampling conical frustum and sampling comparison (Barron et al., 2021)

(3) 光度一致性及瞬态遮挡与 NeRF-W。现实世界中获取的图像受各种条件影响，色彩表现不可能完全一致。NeRF-W (Martin-Brualla 等, 2021) 解决在非结构化的图像中处理可变光照或

瞬态遮挡等问题。图 12 (a) 展示对同一建筑物进行光照一致性处理，图 12 (b) 分离静态与瞬态对象，并对瞬态对象进行遮挡处理。

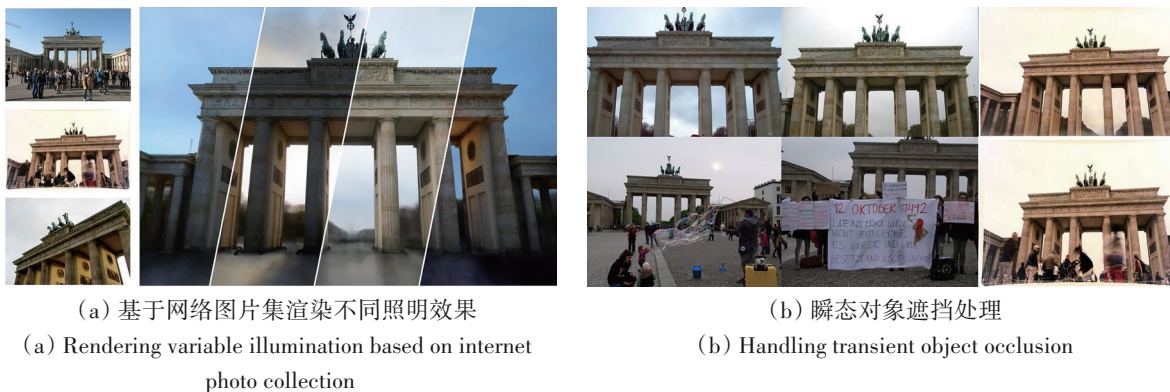


图 12 基于网络图片集渲染不同照明效果、瞬态对象遮挡处理(Martin-Brualla 等, 2021)

Fig. 12 Rendering variable illumination based on internet photo collection and handling transient object occlusion (Martin-Brualla et al., 2021)

NeRF-W 优化输入图像的外观嵌入，通过学习整个图像数据集的共享外观表示来解释图像之间的光度和环境变化，学习到的潜在空间提供了对输出渲染外观的控制 (图 12 (a))。将场景建模为共享元素 (静态组件) 和图像相关元素 (瞬态组件) 的联合 (图 13)，当给定空间位置、观察方向、外观嵌入和瞬态嵌入，NeRF-W 产生静态和瞬态颜色与密度以及不确定性的预估，以合成最终的结果。NeRF-W 可以从非结构化图像中渲染逼真的场景，但对于场景中采集密度低的区域或相机位姿不正确的源图像，会影响最终的渲染效果。

除了 NeRF-W 以外，也有研究基于 HDR 和 LDR 图像 (Huang 等, 2022a)、神经瞬态场 (Neural Transient Fields) (Shen 等, 2021b)、抗遮挡模块

(Anti-occlusion Module) (Chen 等, 2022)，以及高帧率事件相机 (Hwang 等, 2023) 等思路，实现静态与瞬态场景的分离，这些方法多针对特定的应用场景，具有较高的局限性。

(4) 重照明及可见性场与 NeRV。当输入的图像照明条件不统一时，视图合成算法需要对场景进行重照明，以统一不同图像中的照明效果。NeRV (Srinivasan 等, 2021) 将场景表示为一个连续体积函数，当输入任意一个空间位置，可以输出这个位置下的体积密度、表面法线、材料参数、该位置到任何方向上的第一个表面交叉点的距离，以及任何方向上的外部环境的可见性，通过这些参数可以进行场景重照明以及空间位置可见性判断。在给定的三维位置上，不需要频繁地评估 MLP 中相应射线对给定方向密度体的可见性，只

需要简单地查询可见性网络就能对任何方向上的能见度和视线终止深度进行预估(图14)。NeRV利用学习函数降低传统渲染算法计算开销,在任意直射与一次反射光照的模拟效果表现优秀,但对于复杂的光照环境需要更进一步研究。

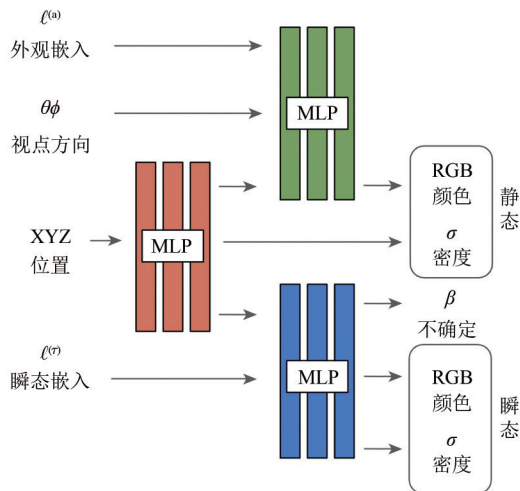


图13 NeRF-W模型架构(Martin-Brualla等,2021)
Fig. 13 NeRF-W model architecture
(Martin-Brualla et al., 2021)

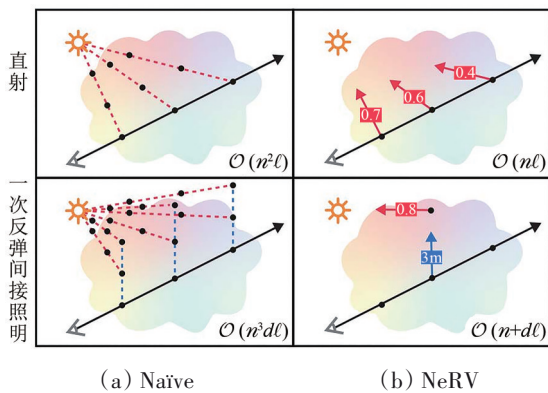


图14 Naïve与NeRV直接和一次反弹间接照明射线计算复杂性(Srinivasan等,2021)
Fig. 14 Naïve and NeRV directly and one-bounce indirect illumination ray computational complexity
(Srinivasan et al., 2021)

户外场景光线影响因素众多,重照明是实现户外视图合成外观统一的必要手段。重照明算法(Boss等,2021a,2021b;Zhang等,2021;Rudnev等,2022)一般均是通过解算物体表面法线、光线能见度、反射率和双向反射分布函数BRDF(Bidirectional Reflectance Distribution Function)等参数,在新照明条件下实现场景重照明。由于光照效果模拟涉及因素(光源、几何、材质、纹理、

反射、折射等)较多,这类方法通常对几何重建精度要求较高,且仅能模拟简单的光照效果。

4 总结

NeRF基于神经隐式立体表达方式进行可微渲染,通过视图合成实现逼真的三维效果。针对NeRF的各种扩展研究从增加数据源、动态对象渲染、效率提升以及增强NeRF对不同类型对象的渲染支持等方面,对已有算法进行了优化和提升,取得重要成果。目前将NeRF应用于大规模实景三维场景可视化所开展的研究工作相对较少,但这些成果在各种条件约束下实现了最基本的实景三维可视化效果,尽管其尚不具备可量测性且通用性还有待提高,但仍然起到了开拓、引领和示范的作用。由于神经辐射场在大规模实景三维可视化方面尚远不能满足生产意义上的实际应用需要,对其实用化所开展的研究探索皆可能成为持续研究的起点。考虑目前已经取得的研究成果,针对实景三维建设需求,未来的研究可关注以下方面:

(1) 多源数据融合。传统实景三维建设一般以有限类型数据作为数据源进行三维建模,包括密集获取的二维图像、三维激光扫描仪或激光雷达获取的点云数据等。不同类型数据的精度不同,建模质量差异较大。原始NeRF算法要求源数据为带有位姿信息的二维RGB图像,利用图像以外的数据辅助进行NeRF视图合成的相关研究表明,点云、深度图、网格模型、体素模型等辅助数据的引入可以在一定程度上提高NeRF的计算效率或提升模型精细度。

除此之外,个人移动终端、车载设备、街景监控、航空航天以及互联网等领域,时刻都在生产包含丰富地理环境信息的数据,将多源数据融合用于NeRF开展实景三维建设是一个值得进一步探索的方向。

(2) 视觉效果优化。传统大规模实景三维场景以显式模型表达为主,由于显式模型建模原理的限制,在表达透射、反射、植被、流体等特殊类型对象时,会出现大量空洞或伪影,导致较差的可视化效果。基于隐式模型的视图合成技术针对这些特殊对象的渲染效果会更逼真和更完整。将隐式模型用于对显式模型的渲染优化研究工作,将会为实景三维场景可视化效果带来质的飞越。

根据视点位置与观察目标的距离渲染不同精细程度的模型,是传统渲染大规模实景三维场景时避免资源过载的主要处理方式。随着视点的逼近或拉远,不同精细级别模型变换时普遍存在跳变感。而NeRF基于隐式模型连续可微渲染,每一帧渲染仅对视景范围内内容进行重新计算与绘制以构建精细度无级变化的模型可视化效果,将显著提高实景三维场景浏览的连贯性,大幅优化视觉效果。

(3) 虚拟环境感知。神经辐射场真正用于实景三维建设,不仅要在视觉效果方面做进一步研究,特别需要在计算性能以及实现场景空间定量解释性方面展开大力研究。国土实景三维场景作为国家数字基础设施不仅需要逼真的视觉效果,也需要对真实地理空间位置的映射。隐式模型的渲染在逼真程度上优于显式模型,但对于场景的三维量测、地物对象单体化等空间定量解释性上目前远不如显式模型,相关问题还需要展开大力研究。

实时渲染是各种虚拟地理环境应用场景中最重要的要求之一,而NeRF计算成本较高,目前无法满足实时渲染的要求,已有的研究以显式模型优化、并行计算或高性能芯片为主要提速技术路线,在提高训练速度与渲染效率方面取得一些进展,但距实时渲染的目标还很远,在相当长时期内,实时渲染都将成为NeRF重要的研究方向。

(4) 自动场景构建。三维场景构建一直是虚拟地理环境领域研究的主要内容,场景中地物对象模型的构建多以半自动化或人工建模为主,存在效率低成本高的问题,全自动化的三维建模技术是自动场景构建的难点。NeRF以神经隐式表达的方式进行三维场景的重构渲染,无需构建显式三维模型,理论上不受场景复杂度的限制,自动化的图像获取与基于NeRF的全自动大规模隐式三维场景渲染技术相结合,可以实现真正意义上的全自动化实景三维场景构建。对于3D GIS、自动驾驶、数字孪生城市,以及元宇宙等以虚拟地理环境三维场景作为基础数据底座的各种应用场景都具有特殊的意义。

神经辐射场作为一项新兴技术,可能会存在依赖源数据、计算周期长、空间定量解释性差等方面的约束和不足,在其进一步发展的过程中,将面临极大的挑战和机遇,其用于实景三维建设

已经初露锋芒,必将为虚拟地理环境建设注入新的技术动力。

参考文献(References)

- Agarwal S, Furukawa Y, Snavely N, Simon I, Curless B, Seitz S M and Szeliski R. 2011. Building Rome in a day. *Communications of the ACM*, 54(10): 105-112 [DOI: 10.1145/2001269.2001293]
- Athar S, Shu Z X and Samaras D. 2021. FLAME-in-NeRF: neural control of radiance fields for free view face animation. *arXiv*: 2108.04913 [DOI: 10.48550/arXiv.2108.04913]
- Attal B, Laidlaw E, Gokaslan A, Kim C, Richardt C, Tompkin J and O'Toole M. 2021. TöRF: time-of-flight radiance fields for dynamic scene view synthesis. *arXiv*: 2109.15271 [DOI: 10.48550/arXiv.2109.15271]
- Barron J T, Mildenhall B, Tancik M, Hedman P, Martin-Brualla R and Srinivasan P P. 2021. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE: 5835-5844 [DOI: 10.1109/ICCV48922.2021.00580]
- Barron J T, Mildenhall B, Verbin D, Srinivasan P P and Hedman P. 2022. Mip-NeRF 360: unbounded anti-aliased neural radiance fields//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE: 5460-5469 [DOI: 10.1109/CVPR52688.2022.00539]
- Boss M, Braun R, Jampani V, Barron J T, Liu C and Lensch H P A. 2021a. NeRD: neural reflectance decomposition from image collections//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE: 12664-12674 [DOI: 10.1109/ICCV48922.2021.01245]
- Boss M, Jampani V, Braun R, Liu C, Barron J T and Lensch H P A. 2021b. Neural-PIL: neural pre-integrated lighting for reflectance decomposition. *arXiv*: 2110.14373 [DOI: 10.48550/arXiv.2110.14373]
- Buehler C, Bosse M, McMillan L, Gortler S and Cohen M. 2001. Unstructured lumigraph rendering//*Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. Los Angeles: ACM: 425-432 [DOI: 10.1145/383259.383309]
- Chan E R, Monteiro M, Kellnhofer P, Wu J J and Wertzstein G. 2021. pi-GAN: periodic implicit generative adversarial networks for 3D-aware image synthesis//*Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE: 5795-5805 [DOI: 10.1109/CVPR46437.2021.00574]
- Chang Y and Gai M. 2021. A review on neural radiance fields based view synthesis. *Journal of Graphics*, 42(3): 376-384 (常远, 盖孟. 2021. 基于神经辐射场的视点合成算法综述. *图学学报*, 42(3): 376-384) [DOI: 10.11996/JG.j.2095-302X.2021030376]
- Chen A P, Xu Z X, Zhao F Q, Zhang X S, Xiang F B, Yu J Y and Su H. 2021. MVSNeRF: fast generalizable radiance field reconstruction from multi-view stereo//*Proceedings of the 2021 IEEE/CVF Inter-*

- national Conference on Computer Vision (ICCV). Montreal: IEEE: 14104-14113 [DOI: 10.1109/ICCV48922.2021.01386]
- Chen W Z, Gao J, Ling H, Smith E J, Lehtinen J, Jacobson A and Fidler S. 2019. Learning to predict 3D objects with an interpolation-based differentiable renderer//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 9609-9619
- Chen X Y, Zhang Q, Li X Y, Chen Y, Ying F, Wang X and Wang J. 2022. Hallucinated neural radiance fields in the wild//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12933-12942 [DOI: 10.1109/CVPR52688.2022.01260]
- Cheng B W, Collins M D, Zhu Y K, Liu T, Huang T S, Adam H and Chen L C. 2020. Panoptic-DeepLab: a simple, strong, and fast baseline for bottom-up panoptic segmentation//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 12472-12482 [DOI: 10.1109/CVPR42600.2020.01249]
- Choi I, Gallo O, Troccoli A, Kim M H and Kautz J. 2019. Extreme view synthesis//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE: 7780-7789 [DOI: 10.1109/ICCV.2019.00787]
- Curless B and Levoy M. 1996. A volumetric method for building complex models from range images//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New Orleans: ACM: 303-312 [DOI: 10.1145/237170.237269]
- Davis A, Levoy M and Durand F. 2012. Unstructured light fields. *Computer Graphics Forum*, 31: 305-314 [DOI: 10.1111/j.1467-8659.2012.03009.x]
- Davison A J, Reid I D, Molton N D and Stasse O. 2007. MonoSLAM: real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6): 1052-1067 [DOI: 10.1109/TPAMI.2007.1049]
- Deng K L, Liu A, Zhu J Y and Ramanan D. 2022. Depth-supervised NeRF: fewer views and faster training for free//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12872-12881 [DOI: 10.1109/CVPR52688.2022.01254]
- Derksen D and Izzo D. 2021. Shadow neural radiance fields for multi-view satellite photogrammetry//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville: IEEE: 1152-1161 [DOI: 10.1109/CVPRW53098.2021.00126]
- Dey A, Ahmine Y and Comport A I. 2022. Mip-NeRF RGB-D: depth assisted fast neural radiance fields. arXiv: 2205.09351 [DOI: 10.48550/arXiv.2205.09351]
- Engel J, Schöps T and Cremers D. 2014. LSD-SLAM: large-scale direct monocular SLAM//13th European Conference on Computer Vision - ECCV 2014. Zürich: Springer: 834-849 [DOI: 10.1007/978-3-319-10605-2_54]
- Fang J M, Yi T R, Wang X G, Xie L X, Zhang X P, Liu W Y, Nießner M and Tian Q. 2022. Fast dynamic radiance fields with time-aware neural voxels//SIGGRAPH Asia 2022 Conference Papers. Daegu: ACM: 1-9 [DOI: 10.1145/3550469.3555383]
- Früh C and Zakhor A. 2004. An automated method for large-scale, ground-based city model acquisition. *International Journal of Computer Vision*, 60(1): 5-24 [DOI: 10.1023/B:VISI.0000027787.82851.b6]
- Gao C, Shih Y, Lai W S, Liang C K and Huang J B. 2021. Portrait neural radiance fields from a single image. arXiv: 2012.05903 [DOI: 10.48550/arXiv.2012.05903]
- Genova K, Cole F, Sud A, Sarna A and Funkhouser T. 2020. Local deep implicit functions for 3D shape//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 4856-4865 [DOI: 10.1109/CVPR42600.2020.00491]
- Gong J H, Lin H, Xiao L B and Xie C J. 1999. Perspective on geo-visualization. *Journal of Remote Sensing*, 3(3): 236-244 (龚建华, 林琿, 肖乐斌, 谢传节. 1999. 地学可视化探讨. *遥感学报*, 3(3): 236-244) [DOI: 10.11834/jrs.19990314]
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139-144 [DOI: 10.1145/3422622]
- Guo X, Chen G Y, Dai Y C, Ye X Q, Sun J D, Tan X and Ding E R. 2022a. Neural deformable voxel grid for fast optimization of dynamic view synthesis//Proceedings of the 16th Asian Conference on Computer Vision (ACCV). Macau, China: Springer: 450-468 [DOI: 10.1007/978-3-031-26319-4_27]
- Guo Y C, Kang D, Bao L C, He Y and Zhang S H. 2022b. NeRFReN: neural radiance fields with reflections//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18388-18397 [DOI: 10.1109/CVPR52688.2022.01786]
- Haklay M and Weber P. 2008. OpenStreetMap: user-generated street maps. *IEEE Pervasive Computing*, 7(4): 12-18 [DOI: 10.1109/MPRV.2008.80]
- Hedman P, Philip J, Price T, Frahm J M, Drettakis G and Brostow G. 2018. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6): 257 [DOI: 10.1145/3272127.3275084]
- Hedman P, Srinivasan P P, Mildenhall B, Barron J T and Debevec P. 2021. Baking neural radiance fields for real-time view synthesis//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5855-5864 [DOI: 10.1109/ICCV48922.2021.00582]
- Hoang L, Lee S H, Kwon O H and Kwon K R. 2019. A deep learning method for 3D object classification using the wave kernel signature and a center point of the 3D-triangle mesh. *Electronics*, 8(10): 1196 [DOI: 10.3390/electronics8101196]
- Hong Y, Peng B, Xiao H Y, Liu L G and Zhang J Y. 2022. HeadNeRF: a realtime NeRF-based parametric head model//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 20342-20352 [DOI:

- 10.1109/CVPR52688.2022.01973]
- Huang X, Zhang Q, Feng Y, Li H D, Wang X and Wang Q. 2022a. HDR-NeRF: high dynamic range neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18377-18387 [DOI: 10.1109/CVPR52688.2022.01785]
- Huang Y H, He Y, Yuan Y J, Lai Y K and Gao L. 2022b. Stylized-NeRF: consistent 3D scene stylization as stylized NeRF via 2D-3D mutual learning//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18321-18331 [DOI: 10.1109/CVPR52688.2022.01780]
- Hwang I, Kim J and Kim Y M. 2023. Ev-NeRF: event based neural radiance field//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE: 837-847 [DOI: 10.1109/WACV56688.2023.00090]
- Jain A, Tancik M and Abbeel P. 2021. Putting NeRF on a diet: semantically consistent few-shot view synthesis//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5865-5874 [DOI: 10.1109/ICCV48922.2021.00583]
- Jang W and Agapito L. 2021. CodeNeRF: disentangled neural radiance fields for object categories//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 12929-12938 [DOI: 10.1109/ICCV48922.2021.01271]
- Jiang C Y, Sud A, Makadia A, Huang J W, Nießner M and Funkhouser T. 2020. Local implicit grid representations for 3D scenes//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 6000-6009 [DOI: 10.1109/CVPR42600.2020.00604]
- Jo K, Shim G, Jung S, Yang S and Choo J. 2023. CG-NeRF: conditional generative neural radiance fields for 3D-aware image synthesis//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE: 724-733 [DOI: 10.1109/WACV56688.2023.00079]
- Johari M M, Lepoittevin Y and Fleuret F. 2022. GeoNeRF: generalizing NeRF with geometry priors//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18344-18347 [DOI: 10.1109/CVPR52688.2022.01782]
- Kada M, Wichmann A, Filipovska Y and Hermes T. 2016. Animation strategies for smooth transformations between discrete LODs of 3D building models. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B2: 413-420 [DOI: 10.5194/isprsarchives-XLI-B2-413-2016]
- Kada M, Wichmann A and Hermes T. 2015. Smooth transformations between generalized 3D building models for visualization purposes. Cartography and Geographic Information Science, 42(4): 306-314 [DOI: 10.1080/15230406.2015.1039588]
- Kania K, Yi K M, Kowalski M, Trzeciński T and Tagliasacchi A. 2022. CoNeRF: controllable neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18602-18611 [DOI: 10.1109/CVPR52688.2022.01807]
- Khademi W and Ventura J. 2021. View synthesis in casually captured scenes using a cylindrical neural radiance field with exposure compensation//ACM SIGGRAPH 2021 Posters. Virtual Event: ACM: 28 [DOI: 10.1145/3450618.3469147]
- Kingma D P and Welling M. 2022. Auto-encoding variational bayes. arXiv: 1312.6114 [DOI: 10.48550/arXiv.1312.6114]
- Kondo N, Ikeda Y, Tagliasacchi A, Matsuo Y, Ochiai Y and Gu S S. 2021. VaxNeRF: revisiting the classic for voxel-accelerated neural radiance field. arXiv: 2111.13112 [DOI: 10.48550/arXiv.2111.13112]
- Kosiorek A R, Strathmann H, Zoran D, Moreno P, Schneider R, Mokrá S and Rezende D J. 2021. NeRF-VAE: a geometry aware 3D scene generative model//Proceedings of the 38th International Conference on Machine Learning (ICML). Virtual: [s. n.]: 5742-5752
- Kutulakos K N and Seitz S M. 2000. A theory of shape by space carving. International Journal of Computer Vision, 38(3): 199-218 [DOI: 10.1023/A:1008191222954]
- Levoy M and Hanrahan P. 1996. Light field rendering//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New Orleans: ACM Press: 31-42 [DOI: 10.1145/237170.237199]
- Li D R, Xiao X W, Guo B X, Jiang W S and Shi Y R. 2016. Oblique image based automatic aerotriangulation and its application in 3D city model reconstruction. Geomatics and Information Science of Wuhan University, 41(6): 711-721 (李德仁, 肖雄武, 郭丙轩, 江万寿, 时月茹. 2016. 倾斜影像自动空三及其在城市真三维模型重建中的应用. 武汉大学学报(信息科学版), 41(6): 711-721) [DOI: 10.13203/j.whugis20160099]
- Li J X, Feng Z J, She Q, Ding H H, Wang C H and Lee G H. 2021a. MINE: towards continuous depth MPI with NeRF for novel view synthesis//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 12558-12568 [DOI: 10.1109/ICCV48922.2021.01235]
- Li T M, Aittala M, Durand F and Lehtinen J. 2018. Differentiable Monte Carlo ray tracing through edge sampling. ACM Transactions on Graphics, 37(6): 222 [DOI: 10.1145/3272127.3275109]
- Li Z P, Li L, Ma Z Y, Zhang P, Chen J B and Zhu J K. 2022. READ: large-scale neural scene rendering for autonomous driving. arXiv: 2205.05509 [DOI: 10.48550/arXiv.2205.05509]
- Li Z Q, Niklaus S, Snavely N and Wang O. 2021b. Neural scene flow fields for space-time view synthesis of dynamic scenes//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 6494-6504 [DOI: 10.1109/CVPR46437.2021.00643]
- Lin C H, Ma W C, Torralba A and Lucey S. 2021. BARF: bundle-adjusting neural radiance fields//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5721-5731 [DOI: 10.1109/ICCV48922.2021.00569]
- Lin H T, Peng S D, Xu Z, Yan Y Z, Shuai Q, Bao H J and Zhou X W.

2022. Efficient neural radiance fields for interactive free-viewpoint video//SIGGRAPH Asia 2022 Conference Papers. Daegu: ACM: 39 [DOI: 10.1145/3550469.3555376]
- Liu J W, Cao Y P, Mao W J, Zhang W Q, Zhang D J, Keppo J, Shan Y, Qie X H and Shou M Z. 2022. DeVRF: fast deformable voxel radiance fields for dynamic scenes. arXiv: 2205.15723 [DOI: 10.48550/arXiv.2205.15723]
- Liu L J, Gu J T, Lin K Z, Chua T S and Theobalt C. 2021a. Neural sparse voxel fields. arXiv: 2007.11571 [DOI: 10.48550/arXiv.2007.11571]
- Liu S, Zhang X M, Zhang Z T, Zhang R, Zhu J Y and Russell B. 2021b. Editing conditional radiance fields//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5753-5763 [DOI: 10.1109/ICCV48922.2021.00572]
- Lombardi S, Simon T, Saragih J, Schwartz G, Lehrmann A and Sheikh Y. 2019. Neural volumes: learning dynamic renderable volumes from images. *ACM Transactions on Graphics*, 38(4): 65 [DOI: 10.1145/3306346.3323020]
- Loper M M and Black M J. 2014. OpenDR: an approximate differentiable renderer//13th European Conference on Computer Vision. Zürich: Springer International Publishing: 154-169 [DOI: 10.1007/978-3-319-10584-0_11]
- Luo H M, Chen A P, Zhang Q X, Pang B, Wu M Y, Xu L and Yu J Y. 2021. Convolutional neural opacity radiance fields//2021 IEEE International Conference on Computational Photography (ICCP). Haifa: IEEE: 1-12 [DOI: 10.1109/ICCP51581.2021.9466273]
- Ma L, Li X Y, Liao J, Zhang Q, Wang X, Wang J and Sander P V. 2022. Deblur-NeRF: neural radiance fields from blurry images//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12851-12860 [DOI: 10.1109/CVPR52688.2022.01252]
- Marí R, Facciolo G and Ehret T. 2022. Sat-NeRF: learning multi-view satellite photogrammetry with transient objects and shadow modeling using RPC cameras//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR). New Orleans: IEEE: 1310-1320 [DOI: 10.1109/CVPRW56347.2022.00137]
- Martin-Brualla R, Radwan N, Sajjadi M S M, Barron J T, Dosovitskiy A and Duckworth D. 2021. NeRF in the Wild: neural radiance fields for unconstrained photo collections//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 7206-7215 [DOI: 10.1109/CVPR46437.2021.00713]
- Meng Q, Chen A P, Luo H M, Wu M Y, Su H, Xu L, He X M and Yu J Y. 2021. GNeRF: GAN-based neural radiance field without posed camera//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 6351-6361 [DOI: 10.1109/ICCV48922.2021.00629]
- Mescheder L, Oechsle M, Niemeyer M, Nowozin S and Geiger A. 2019. Occupancy networks: learning 3D reconstruction in function space//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 4455-4465 [DOI: 10.1109/CVPR.2019.00459]
- Mildenhall B, Hedman P, Martin-Brualla R, Srinivasan P P and Barron J T. 2022. NeRF in the Dark: high dynamic range view synthesis from noisy raw images//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 16169-16178 [DOI: 10.1109/CVPR52688.2022.01571]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis. arXiv: 2003.08934 [DOI: 10.48550/arXiv.2003.08934]
- Moreau A, Piasco N, Tsishkou D, Stanculescu B and de la Fortelle A. 2021. LENS: localization enhanced by NeRF synthesis//Proceedings of the 5th Conference on Robot Learning. London: PMLR: 1347-1356 [DOI: 10.48550/arXiv.2110.06558]
- Mur-Artal R, Montiel J M M and Tardós J D. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5): 1147-1163 [DOI: 10.1109/TRO.2015.2463671]
- Neff T, Stadlbauer P, Parger M, Kurz A, Mueller J H, Chaitanya C R A, Kaplanyan A and Steinberger M. 2021. DONeRF: towards real-time rendering of compact neural radiance fields using depth oracle networks. *Computer Graphics Forum*, 40(4): 45-59 [DOI: 10.1111/cgf.14340]
- Newcombe R A, Lovegrove S J and Davison A J. 2011. DTAM: dense tracking and mapping in real-time//Proceedings of the 2011 International Conference on Computer Vision. Barcelona: IEEE: 2320-2327 [DOI: 10.1109/ICCV.2011.6126513]
- Nguyen-Phuoc T, Liu F and Xiao L. 2022. SNeRF: stylized neural implicit representations for 3D scenes. *ACM Transactions on Graphics*, 41(4): 142 [DOI: 10.1145/3528223.3530107]
- Niemeyer M and Geiger A. 2021. GIRAFFE: representing scenes as compositional generative neural feature fields//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 11448-11459 [DOI: 10.1109/CVPR46437.2021.01129]
- Niemeyer M, Mescheder L, Oechsle M and Geiger A. 2020. Differentiable volumetric rendering: learning implicit 3D representations without 3D supervision//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 3501-3512 [DOI: 10.1109/CVPR42600.2020.00356]
- Ohuri K A, Ledoux H, Biljecki F and Stoter J. 2015. Modeling a 3D city model and its levels of detail as a true 4D model. *ISPRS International Journal of Geo-Information*, 4(3): 1055-1075 [DOI: 10.3390/ijgi4031055]
- Özyeşil O, Voroninski V, Basri R and Singer A. 2017. A survey of structure from motion. *Acta Numerica*, 26: 305-364 [DOI: 10.1017/S096249291700006X]
- Park J J, Florence P, Straub J, Newcombe R and Lovegrove S. 2019. DeepSDF: learning continuous signed distance functions for shape representation//Proceedings of the 2019 IEEE/CVF Confer-

- ence on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 165-174 [DOI: 10.1109/CVPR.2019.00025]
- Park K, Sinha U, Hedman P, Barron J T, Bouaziz S, Goldman D B, Martin-Brualla R and Seitz S M. 2021. HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields. *ACM Transactions on Graphics*, 40(6): 238 [DOI: 10.1145/3478513.3480487]
- Pearl N, Treibitz T and Korman S. 2022. NAN: noise-aware NeRFs for burst-denoising//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12662-12671 [DOI: 10.1109/CVPR52688.2022.01234]
- Peng S Y, Niemeyer M, Mescheder L, Pollefeys M and Geiger A. 2020. Convolutional occupancy networks//16th European Conference on Computer Vision (ECCV). Glasgow: Springer International Publishing: 523-540 [DOI: 10.1007/978-3-030-58580-8_31]
- Penner E and Zhang L. 2017. Soft 3D reconstruction for view synthesis. *ACM Transactions on Graphics*, 36(6): 235 [DOI: 10.1145/3130800.3130855]
- Piala M and Clark R. 2021. TerminiNeRF: ray termination prediction for efficient neural rendering//2021 International Conference on 3D Vision (3DV). London: IEEE: 1106-1114 [DOI: 10.1109/3DV53792.2021.00118]
- Pumarola A, Corona E, Pons-Moll G and Moreno-Noguer F. 2021. D-NeRF: neural radiance fields for dynamic scenes//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 10313-10322 [DOI: 10.1109/CVPR46437.2021.01018]
- Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G and Sutskever I. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning (ICML). Virtual: [s.n.]: 8748-8763
- Raguram R, Wu C C, Frahm J M and Lazebnik S. 2011. Modeling and recognition of landmark image collections using iconic scene graphs. *International Journal of Computer Vision*, 95(3): 213-239 [DOI: 10.1007/s11263-011-0445-z]
- Rebain D, Jiang W, Yazdani S, Li K, Yi K M and Tagliasacchi A. 2021. DeRF: decomposed radiance fields//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 14148-14156 [DOI: 10.1109/CVPR46437.2021.01393]
- Reiser C, Peng S Y, Liao Y Y and Geiger A. 2021. KiloNeRF: speeding up neural radiance fields with thousands of tiny MLPs//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 14315-14325 [DOI: 10.1109/ICCV48922.2021.01407]
- Rematas K, Liu A, Srinivasan P, Barron J, Tagliasacchi A, Funkhouser T and Ferrari V. 2022. Urban radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12922-12932 [DOI: 10.1109/CVPR52688.2022.01259]
- Rematas K, Martin-Brualla R and Ferrari V. 2021. ShaRF: shape-conditioned radiance fields from a single view//Proceedings of the 38th International Conference on Machine Learning (ICML). Virtual: [s.n.]: 8948-8958
- Roessle B, Barron J T, Mildenhall B, Srinivasan P P and Nießner M. 2022. Dense depth priors for neural radiance fields from sparse input views//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12882-12891 [DOI: 10.1109/CVPR52688.2022.01255]
- Rosu R A and Behnke S. 2022. NeuralMVS: bridging multi-view stereo and novel view synthesis//2022 International Joint Conference on Neural Networks (IJCNN). Padua: IEEE: 1-7 [DOI: 10.1109/IJCNN55064.2022.9892024]
- Rudnev V, Elgharib M, Smith W, Liu L J, Golyanik V and Theobalt C. 2022. Neural radiance fields for outdoor scene relighting//17th European Conference on Computer Vision (ECCV). Tel Aviv: Springer: 615-631 [DOI: 10.1007/978-3-031-19787-1_35]
- Schönberger J L and Frahm J M. 2016. Structure-from-Motion revisited//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Schwarz K, Liao Y Y, Niemeyer M and Geiger A. 2021. GRAF: generative radiance fields for 3D-aware image synthesis. arXiv:2007.02442 [DOI: 10.48550/arXiv.2007.02442]
- Seitz S M and Dyer C R. 1999. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2): 151-173 [DOI: 10.1023/A:1008176507526]
- Shan Q, Adams R, Curless B, Furukawa Y and Seitz S M. 2013. The visual turing test for scene reconstruction//2013 International Conference on 3D Vision (3DV). Seattle: IEEE: 25-32 [DOI: 10.1109/3DV.2013.12]
- She J F, Chen B, Tan J Z, Zhao Q and Ge R C. 2022. 3D building model simplification method considering both model mesh and building structure. *Transactions in GIS*, 26(3): 1182-1203 [DOI: 10.1111/tgis.12907]
- Shen J X, Ruiz A, Agudo A and Moreno-Noguer F. 2021a. Stochastic neural radiance fields: quantifying uncertainty in implicit 3D representations//2021 International Conference on 3D Vision (3DV). London: IEEE: 972-981 [DOI: 10.1109/3DV53792.2021.00105]
- Shen S Y, Wang Z, Liu P, Pan Z Q, Li R Q, Gao T, Li S Y and Yu J Y. 2021b. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7): 2257-2268 [DOI: 10.1109/TPAMI.2021.3076062]
- Sitzmann V, Zollhöfer M and Wetzstein G. 2019. Scene representation networks: continuous 3D-structure-aware neural scene representations//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 1121-1132 [DOI: 10.48550/arXiv.1906.01618]
- Srinivasan P P, Deng B Y, Zhang X M, Tancik M, Mildenhall B and Barron J T. 2021. NeRV: neural reflectance and visibility fields for relighting and view synthesis//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition

- (CVPR). Nashville: IEEE: 7491-7500 [DOI: 10.1109/CVPR46437.2021.00741]
- Su S Y, Yu F, Zollhoefer M and Rhodin H. 2021. A-NeRF: articulated neural radiance fields for learning human shape, appearance, and pose. arXiv: 2102.06199 [DOI: 10.48550/arXiv.2102.06199]
- Sun J X, Wang X, Zhang Y, Li X Y, Zhang Q, Liu Y B and Wang J. 2021. FENeRF: face editing in neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 7662-7672 [DOI: 10.1109/CVPR52688.2022.00752]
- Szeliski R and Golland P. 1998. Stereo matching with transparency and matting//Sixth International Conference on Computer Vision. Bombay: IEEE: 517-524 [DOI: 10.1109/ICCV.1998.710766]
- Szeliski R and Shum H Y. 1997. Creating full view panoramic image mosaics and environment maps//Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles: ACM Press: 251-258 [DOI: 10.1145/258734.258861]
- Takikawa T, Litalien J, Yin K X, Kreis K, Loop C, Nowrouzezahrai D, Jacobson A, McGuire M and Fidler S. 2021. Neural geometric level of detail: real-time rendering with implicit 3D shapes//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 11353-11362 [DOI: 10.1109/CVPR46437.2021.01120]
- Tancik M, Casser V, Yan X C, Pradhan S, Mildenhall B P, Srinivasan P, Barron J T and Kretschmar H. 2022. Block-NeRF: scalable large scene neural view synthesis//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 8238-8248 [DOI: 10.1109/CVPR52688.2022.00807]
- Tretschk E, Tewari A, Golyanik V, Zollhöfer M, Lassner C and Theobalt C. 2021. Non-rigid neural radiance fields: reconstruction and novel view synthesis of a dynamic scene from monocular video//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 12939-12950 [DOI: 10.1109/ICCV48922.2021.01272]
- Trevithick A and Yang B. 2021. GRF: learning a general radiance field for 3D representation and rendering//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 15162-15172 [DOI: 10.1109/ICCV48922.2021.01490]
- Tucker R and Snavely N. 2020. Single-view view synthesis with multiple images//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 548-557 [DOI: 10.1109/CVPR42600.2020.00063]
- Turki H, Ramanan D and Satyanarayanan M. 2022. Mega-NeRF: scalable construction of large-scale NeRFs for virtual fly-throughs//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12912-12921 [DOI: 10.1109/CVPR52688.2022.01258]
- Verbin D, Hedman P, Mildenhall B, Zickler T, Barron J T and Srinivasan P P. 2022. Ref-NeRF: structured view-dependent appearance for neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 5481-5490 [DOI: 10.1109/CVPR52688.2022.00541]
- Wang C, Chai M L, He M M, Chen D D and Liao J. 2022a. CLIP-NeRF: text-and-image driven manipulation of neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 3825-3834 [DOI: 10.1109/CVPR52688.2022.00381]
- Wang Z R, Wu S Z, Xie W D, Chen M and Prisacariu V A. 2022b. NeRF——: neural radiance fields without known camera parameters. arXiv: 2102.07064 [DOI: 10.48550/arXiv.2102.07064]
- Wang Z Y, Wang L, Zhao F Q, Wu M Y, Xu L and Yu J Y. 2022c. MirrorNeRF: one-shot neural portrait radiance field from multi-mirror catadioptric imaging//2021 IEEE International Conference on Computational Photography (ICCP). Haifa: IEEE: 1-12 [DOI: 10.1109/ICCP51581.2021.9466270]
- Wang Z Y, Yang W, Cao J M, Xu L, Yu J Q and Yu J Y. 2022d. NeReF: neural refractive field for fluid surface reconstruction and implicit representation. arXiv: 2203.04130 [DOI: 10.48550/arXiv.2203.04130]
- Wei Y, Liu S H, Rao Y M, Zhao W, Lu J W and Zhou J. 2021. Nerfing-MVS: guided optimization of neural radiance fields for indoor multi-view stereo//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5590-5599 [DOI: 10.1109/ICCV48922.2021.00556]
- Wu J, Zhang Z X and Zhang J Q. 2006. 3D city modeling based on videogrammetry from helicopter. *Journal of Image and Graphics*, 11(8): 1161-1170 (吴军, 张祖勋, 张剑清. 2006. 基于机载视频量测的3维城市建模. *中国图象图形学报*, 11(8): 1161-1170) [DOI: 10.11834/jig.200608196]
- Wu L W, Lee J Y, Bhattad A, Wang Y X and Forsyth D. 2022a. DiVeR: real-time and accurate neural radiance fields with deterministic integration for volume rendering//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 16179-16188 [DOI: 10.1109/CVPR52688.2022.01572]
- Wu T H, Zhong F C, Tagliasacchi A, Cole F and Oztireli C. 2022b. D²NeRF: self-supervised decoupling of dynamic and static objects from a monocular video. arXiv: 2205.15838 [DOI: 10.48550/arXiv.2205.15838]
- Xian W Q, Huang J B, Kopf J and Kim C. 2021. Space-time neural irradiance fields for free-viewpoint video//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 9416-9426 [DOI: 10.1109/CVPR46437.2021.00930]
- Xiangli Y B, Xu L N, Pan X G, Zhao N X, Rao A Y, Theobalt C, Dai B and Lin D H. 2022. Bungeenerf: progressive neural radiance field for extreme multi-scale scene rendering//17th European Conference on Computer Vision (ECCV). Tel Aviv: Springer: 106-122 [DOI: 10.1007/978-3-031-19824-3_7]
- Xie C, Park K, Martin-Brualla R and Brown M. 2021. FiG-NeRF: fig-

- ure-ground neural radiance fields for 3D object category modeling//2021 International Conference on 3D Vision (3DV). London: IEEE: 962-971 [DOI: 10.1109/3DV53792.2021.00104]
- Xu H Y, Alldieck T and Sminchisescu C. 2021. H-NeRF: neural radiance fields for rendering and temporal reconstruction of humans in motion. arXiv: 2110.13746 [DOI: 10.48550/arXiv.2110.13746]
- Xu Q G, Xu Z X, Philip J, Bi S, Shu Z X, Sunkavalli K and Neumann U. 2022a. Point-NeRF: point-based neural radiance fields//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 5428-5438 [DOI: 10.1109/CVPR52688.2022.00536]
- Xu T H, Fujita Y and Matsumoto E. 2022b. Surface-aligned neural radiance fields for controllable 3D human synthesis//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 15862-15871 [DOI: 10.1109/CVPR52688.2022.01542]
- Yan S, Zhang M J, Fan Y C, Tan X H, Liu Y, Peng Y and Liu Y X. 2021. Progress in the large-scale outdoor image 3D reconstruction. Journal of Image and Graphics, 26(6): 1429-1449 (颜深, 张茂军, 樊亚春, 谭小慧, 刘煜, 彭杨, 刘宇翔. 2021. 大规模室外图像3D重建技术研究进展. 中国图象图形学报, 26(6): 1429-1449) [DOI: 10.11834/jig.200842]
- Yang B B, Zhang Y D, Xu Y H, Li Y J, Zhou H, Bao H J, Zhang G F and Cui Z P. 2021. Learning object-compositional neural radiance field for editable scene rendering//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 13759-13768 [DOI: 10.1109/ICCV48922.2021.01352]
- Yang G W, Zhou W Y, Peng H Y, Liang D, Mu T J and Hu S M. 2023. Recursive-NeRF: an efficient and dynamically growing NeRF. IEEE Transactions on Visualization and Computer Graphics, 29(12): 5124-5136 [DOI: 10.1109/TVCG.2022.3204608]
- Yen-Chen L, Florence P, Barron J T, Rodriguez A, Isola P and Lin T Y. 2021. INeRF: inverting neural radiance fields for pose estimation//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague: IEEE: 1323-1330 [DOI: 10.1109/IROS51168.2021.9636708]
- Yu A, Li R L, Tancik M, Li H, Ng R and Kanazawa A. 2021. PlenOc-trees for real-time rendering of neural radiance fields//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5732-5741 [DOI: 10.1109/ICCV48922.2021.00570]
- Zhang K, Riegler G, Snavely N and Koltun V. 2020. NeRF++: analyzing and improving neural radiance fields. arXiv: 2010.07492 [DOI: 10.48550/arXiv.2010.07492]
- Zhang X M, Srinivasan P P, Deng B Y, Debevec P, Freeman W T and Barron J T. 2021. NeRFactor: neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics, 40(6): 237 [DOI: 10.1145/3478513.3480496]
- Zhao F Q, Yang W, Zhang J K, Lin P, Zhang Y L, Yu J Y and Xu L. 2022. HumanNeRF: efficiently generated human radiance field from sparse inputs//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 7733-7743 [DOI: 10.1109/CVPR52688.2022.00759]
- Zhou T H, Tucker R, Flynn J, Fyffe G and Snavely N. 2018. Stereo magnification: learning view synthesis using multiplane images. ACM Transactions on Graphics, 37(4): 65 [DOI: 10.1145/3197517.3201323]
- Zhu S Y, Zhang R Z, Zhou L, Shen T W, Fang T, Tan P and Quan L. 2018. Very large-scale global SfM by distributed motion averaging//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4568-4577 [DOI: 10.1109/CVPR.2018.00480]
- Zhuang Y Y, Zhu H, Sun X S and Cao X. 2022. MoFaNeRF: morphable facial neural radiance field//17th European Conference on Computer Vision. Tel Aviv: Springer: 268-285 [DOI: 10.1007/978-3-031-20062-5_16]

Progress in neural radiance field and its application in large-scale real-scene 3D visualization

ZHAO Qiang^{1,2}, SHE Jiangfeng¹, WAN Qifeng¹, HE Lixia¹, LI Sirui¹, WU Shuangpin¹

1. Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China;

2. Henan Agricultural Remote Sensing Big Data Development and Innovation Laboratory, Henan Engineering Technology Research Center of Ecological protection and management of the Old Course of Yellow River, College of Surveying and Planning, Shangqiu Normal University, Shangqiu 476000, China

Abstract: Geographical real-scene 3D scenes are an important national digital infrastructure, which extends geographic information from 2D to 3D. Real-scene 3D data are stored and expressed in the form of an explicit 3D model, which has the problems of large amount of data and rough visualization effect. Neural Radiance Field (NeRF), realizing differentiable rendering based on neural implicit volume

representation, is an innovative approach of high-quality view synthesis. First proposed by Mildenhall et al. (2020), NeRF has become one of the hottest research direction in the field of computer vision due to its realistic view synthesis effect. A large amount of literature about NeRF have been published since NeRF was proposed, and the application of NeRF in large-scale real-scene 3D visualization has begun to attract the attention of some published papers.

View synthesis, which uses sparse 2D images to generate realistic new views at any viewpoint in 3D space without the reconstruction of 3D models, is a novel way to realize the representation of 3D scenes. The development of view synthesis technology has gone through several stages: image mosaicking, 3D model reprojection, view interpolation, and volume representing technology. NeRF, as an innovative approach of view synthesis, samples 5D coordinates (location and viewing direction) along camera rays, feeds those locations into a multilayer perceptron network to produce color and volume density, and uses volume rendering techniques to composite these values into a new image. NeRF not only produces remarkably higher-quality rendering than prior volumetric approaches but also requires just a fraction of the storage cost of other sampled volumetric representations. However, it faces problems such as requirements for high quality of source data, failure to support dynamic objects, low efficiency in processing, and single type of render target. Moreover, NeRF-related research are mostly conducted based on laboratory environment or standardized data at present. Due to these drawbacks, many obstacles need to be overcome before applying NeRF to large-scale real-scene 3D visualization.

This paper reviews the workaround of unbounded scene, aliasing, luminosity consistency, scene relighting, and visibility field in the Block-NeRF algorithm, a variant of NeRF that can represent large-scale environments (Tancik et al. 2022). The Block-NeRF algorithm splits the environment into a set of Block-NeRFs that can be independently trained in parallel and composited during inference, and it selects relevant Block-NeRFs for rendering, which are then composited smoothly when traversing the scene. To aid with this compositing, Block-NeRF optimizes the appearance codes to match lighting conditions. It trains individual Block-NeRFs using techniques such as appearance embeddings, learned pose refinement, exposure input, transient objects, and visibility prediction.

Nowadays, the studies of applying NeRF to large-scale real-scene 3D visualization are being conducted extensively, which has attracted much attention and plays a pioneering, leading role in further research. The results of these studies achieve the most basic 3D visualization large-scale real-scene effect, but they are limited by some conditions and their universality needs to be strongly improved. Because NeRF is still far from practical applications in producing large-scale real-scene 3D visualization, any slight progressive exploration is likely to become a continuous research hotspot. This paper identifies the challenges of NeRF research, including multisource data fusion, visual effect optimization, and virtual environment perception, which need more research.

Key words: remote sensing, neural radiance field, view synthesis, implicit volume representing, computer vision, virtual geographic environment

Supported by National Natural Science Foundation of China (No. 41871293)