

神经辐射场技术及应用综述

陈涛, 杨启亮*, 陈寅

(中国人民解放军陆军工程大学国防工程学院 南京 210007)
(yql@893.com.cn)

摘要: 神经辐射场(NeRF)是一种基于神经网络的三维重建技术, 它将场景定义为位置和观察视角的五维辐射场函数, 并通过隐式的神经网络来表示. 该技术仅需要单个场景在不同视角下的二维图像, 结合体渲染方程, 可以通过深度学习的方式训练得到该场景的神经辐射场模型, 利用该模型可以合成出高质量的新视角图像. 文中对 NeRF 的现有工作进行调查和分类, 从相参位姿估计、视图集合标准、渲染过程优化等方面介绍各种方法的基本原理及优势, 重点阐明不同方法之间的相似性和差异, 帮助理解它们之间的关系; 还讨论了各种受益于 NeRF 合成机理的应用场景; 最后通过对现有方法优势和局限性的分析, 对未来研究的趋势进行展望.

关键词: 神经辐射场; 神经网络三维重建; 体渲染; 新视角图像

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00759

Overview of NeRF Technology and Applications

Chen Tao, Yang Qiliang*, and Chen Yin

(School of Defense Engineering, Army Engineering University of PLA, Nanjing 210007)

Abstract: NeRF(neural radiance fields) is a neural network-based 3D reconstruction technology. It defines a scene as a five-dimensional radiance field function of position and viewing angle, represented through an implicit neural network. This technique only requires two-dimensional images of a single scene from different perspectives. Coupled with volume rendering equations, a neural radiance field model of the scene can be trained through deep learning. This model can be used to synthesize high-quality images from new perspectives. This paper surveys and categorizes existing work on NeRF, mainly introducing the basic principles and advantages of various methods from aspects such as relative pose estimation, view aggregation standards, and rendering process optimization. The focus is to elucidate the similarities and differences between different methods to help understand their relationship. The paper also discusses various application scenarios that benefit from the NeRF synthesis mechanism. Through the analysis of the advantages and limitations of existing methods, the paper provides a perspective on future research trends.

Key words: neural radiance fields; neural network 3D reconstruction; volume rendering; new viewpoint images

二维图像生成三维模型是计算机视觉的经典难题之一, 该问题的输入是特定场景稀疏视点下的二维 RGB 图像, 输出是该场景的三维模型, 在

此基础上, 可以生成新视角下的二维 RGB 图像. 三维模型分为显式和隐式 2 种表征形式, 显式表示包含点云^[1]、体素^[2-4]和网格^[5]等, 而隐式表示则是

收稿日期: 2023-11-25; 修回日期: 2024-09-24. 基金项目: 国家自然科学基金(52178307); 江苏省自然科学基金(BK20201335, BK20211226). 陈涛(2000—), 男, 硕士, 主要研究方向为计算机图形学、国防工程及智能化; 杨启亮(1975—), 男, 博士, 教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为自适应软件系统、国防工程及智能化; 陈寅(1986—), 男, 博士, 副教授, CCF 会员, 主要研究方向为计算机图形学、国防工程及智能化.

将模型通过函数表达. 2021 年, Mildenhall 等^[6]首次提出神经辐射场(neural radiance fields, NeRF), 它是一种典型的隐式表示方式, 输入稀疏的多视角附着位姿的图像, 即可训练得到一个 NeRF 模型, 基于该模型可以渲染出新视角下的 RGB 图像.

与传统的基于点云、网格和体素等显式的表达相比, NeRF 继承了神经网络的诸多优点, 具有更高的精度、更高的渲染质量、更快的训练速度, 以及应用场景更加广泛等显著优势, 启发了大量的后续工作. 截止到 2023 年, NeRF 模型在大规模三维城市建模、人体建模、照片修复编辑、视频处理和视图合成等多方向得到了应用. 本文首先对传统三维建模和渲染方法进行简要介绍; 然后阐述 NeRF 模型的基础原理, 并从相参位姿估计、视图集合标准和过程速度优化 3 个方面对相关工作进行分类介绍; 再针对不同应用场景下 NeRF 的变种, 如动态模型重建、大范围无界场景、生成式结合与模型标注、引导模型可编辑等, 概述不同方法的相似性和差异性, 并总结其优势和局限性; 最后对 NeRF 未来发展方向以及需要攻克各类问题提出展望, 以期对相关研究人员提供有益参考.

当前, 已有的 NeRF 综述类文献主要针对特定领域展开. Yao 等^[7]详细描述初始 NeRF 模型的元素构成, 包括其网络框架、损失函数和呈现方法, 并总结了当前常用的 NeRF 评估指标; Niu 等^[8]基于对图像的三维重建技术的调查分析, 介绍了部分 NeRF 算法在以深度学习作为基底的三维重建中的应用优势, 但对 NeRF 多类支流导向的全局汇总相对匮乏; 李吉洋等^[9]以 NeRF 的优化与扩展、渲染结果优化、模型压缩、人脸和场景建模等为分类依据, 对各类 NeRF 优化算法进行归纳总结和特性阐释, 并介绍了 NeRF 在国内工业界发展现状, 但缺少对同应用领域各类 NeRF 算法的量化分析与效果对比; He 等^[10]对面向自动驾驶背景下的 NeRF 应用进行调查和分类, 包括感知、同步定位和地图绘制(simultaneous localization and mapping, SLAM)等, 并对该领域的未来发展方向提出见解和讨论; Nguyen 等^[11]探究语义感知 NeRF 在视觉场景解释中的进展和各种应用; Lin^[12]详细阐述动态 NeRF 的开发基础和实现原理, 并对各种动态模型的不同特点进行比较和分析, 总结了实现动态 NeRF 的关键方法. 上述综述文献大多介绍了 NeRF 模型在定向领域的交叉搭建过程及应用前景, 但在全面洞悉 NeRF 发展的整体脉络方面仍显不足.

本文从传统三维建模与渲染方式入手, 分析

其特点及缺陷, 总结 NeRF 在二维图像生成三维模型领域的相对优势; 对 NeRF 中的各角度优化算法, 如相参位姿估计、视图集合标准(质量)、非刚性运动场景、人体及细节特征建模等进行量化分析, 凸显其实际应用效果; 基于对 NeRF 技术的最新研究成果进行综述, 探讨各类方法间的脉络关系、优势和局限性, 依托不同变体的技术核心进行分类介绍, 并通过可视化图像形式呈现, 贯穿近年来 NeRF 在各领域动态发展趋势, 总结其未来发展方向. 本文旨在为相关领域研究人员提供相关算法关键信息与数据的参考, 为感兴趣研究者提供一个清楚、详细的入门指南.

1 传统建模与渲染方式

三维建模^[13-16]指利用计算机软件或硬件创建三维模型的过程, 包括通过几何学和数学方法定义物体的形状、大小和位置等. 渲染指将三维模型转化为多视角图像的过程, 包括通过模型的几何信息、材质属性和光照条件等综合计算每个像素的颜色和亮度, 最终生成逼真的图像. 三维建模提供了渲染所需的几何信息和材质属性, 典型的重建过程涉及从二维图像的集合中恢复出三维模型, 而渲染则将这些信息转化为最终的图像, 决定了物体在图像中的外观和质感, 是典型的三维到二维的变换.

1.1 传统三维建模

从获取的源数据组合及处理方式(云配准、运动结构(structure from motion, SFM)等)上, 传统三维建模可以分为基于多视角的 RGB-D 与 RGB 图像 2 大类.

基于多视角的 RGB-D 图像的三维重建技术如图 1 所示. 首先从不同角度拍摄待重建物体的深度图像, 由于彩色图像有助于相机定位与添加模型颜色, 因此可以在采集深度图像同时获取对应的彩色图像; 然后利用迭代最近点(iterative closest point, ICP)^[17]等技术计算相机姿态, 将场景隐式表达为符号距离函数(signed distance function, SDF)^[18]模型; 最后将 SDF 转化为体素格式的三维表征, 使用光线投射等三维渲染算法重建视角, 输出到增强现实(augmented reality, AR)设备进行显示, 或者通过 MC 算法转换为网格模型.

基于多视角 RGB 图像的三维重建技术如图 2 所示. 针对输入图像为彩色图的三维建模方式, 首先通过 SFM^[19-20]获得输入视图集合的不同角度摄

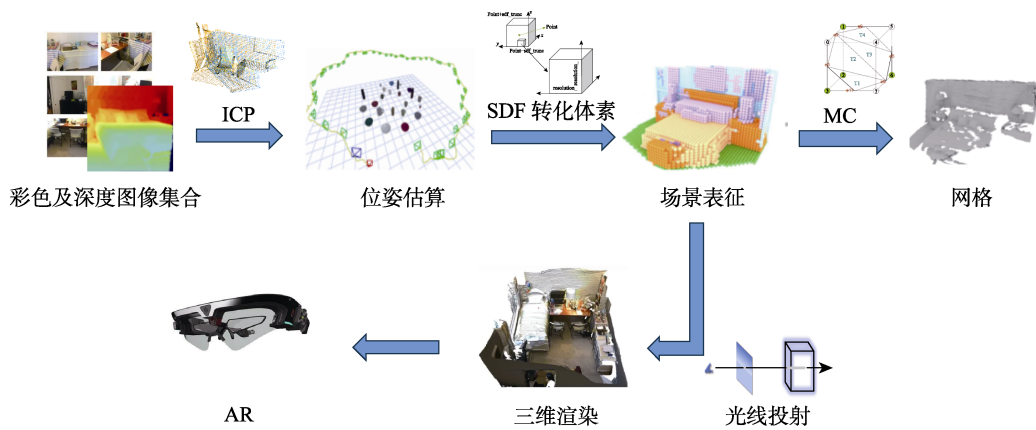


图 1 基于多视角 RGB-D 图像的三维重建

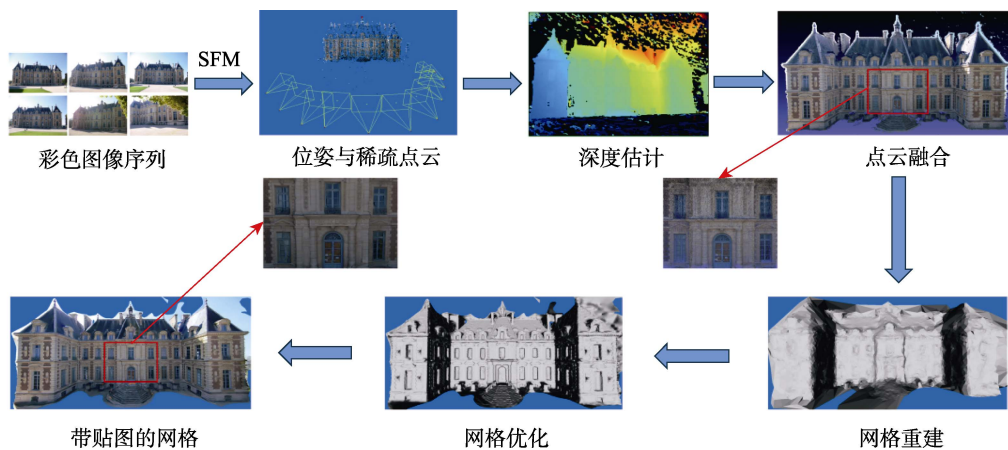


图 2 基于多视角 RGB 图像的三维重建

像机姿态和参数, 建立稀疏点云模型; 然后运用深度估计^[21]和泊松定理^[22]预测像素点颜色等算法, 实现点云模型的稠密重建; 最后对其进行网格优化和纹理贴图, 形成最终的三维模型。

经过多年的发展, 传统的三维重建方式满足了基本三维建模需求, 但仍存在不少缺点, 如在模型的最终重建效果中可能会有孔洞、纹理混叠和体素分辨率限制, 丢失了很多细节。如果实现接近于真实场景的三维表征, 还需要网格优化后进行纹理映射, 存在颜色纹路复杂、高频位置信息锯齿走样、摩尔纹等采样问题。此外, 烦琐的建模过程会降低重建速度, 且在模型可编辑与生成式方面缺乏灵活控制。

1.2 传统三维渲染

三维渲染技术^[23-25]指将三维模型转换成二维图像的过程, 它是计算机图形学中的一个重要分支。三维渲染技术可用于制作电影、游戏、动画等内容, 以及虚拟现实和 AR 应用等。光栅化、光线追踪和光线投射是常用的传统三维渲染技术。

(1) 光栅化^[26-27]指在计算机图形学中将矢量

图形转换为栅格图像的过程。采用局部光照的形式, 在光栅化过程中, 可视化的三维模型是由一个个三角形的网格创建而成的, 计算机会将矢量图形中的每个图形元素转换为位图像素, 并根据像素的颜色、透明度等信息在屏幕上渲染出图像。该技术不如光线追踪逼真, 但是计算量较少。

(2) 光线追踪^[28-29]通过模拟光线在场景中的传播计算每个像素的颜色值。光线从相机位置出发, 经过每个像素并在场景中物体相交。当光线与物体相交时, 根据物体和光线先验知识和属性信息计算在场景物体之间反射、折射、漫反射等, 并继续向前传播, 直至达到光源或被多次反射后被吸收, 最终得到渲染图像。该技术由于能够准确地模拟光线在真实世界中的传播, 因此可产生逼真的图像。但是, 基于全局光照的方法需要对每个像素进行多次光线追踪计算, 因此计算复杂度非常高, 需要高性能计算机或者 GPU 来加速运算。

(3) 光线投射^[30-31]常用于计算场景中的阴影反射和折射等效果。在光线投射中, 射线从相机位

置出发, 贯穿像素平面上的每个点, 并沿着射线方向与场景中的物体进行相交测试. 如果射线与物体相交, 则可以计算出该像素的颜色; 否则, 该像素通常被赋予背景色. NeRF 的体渲染方程便是使用光线投射的原理, 主观认为现实生活中的物体以类似雾、火等类型粒子形态存在, 呈现体密度和纹理信息从而对其进行渲染. 它们的区别是在方程的优化中, NeRF 在射线未与物体相交的点不考虑背景光的影响, 且在光子与粒子发生吸收、放

射、内散射/外散射时认为损失系数一致.

传统三维渲染方式如图 3 所示, 皆基于底层模型实现, 在处理光照、阴影等效果时存在一定的局限性, 难以达到真实场景的效果, 特别是透明、反射等光学成像传播途径中. 该方式对于大规模场景的处理能力有限, 当场景中包含多数物体和复杂几何结构时渲染速度会明显下降, 甚至无法正常进行渲染. 由于计算复杂性的限制, 传统三维渲染方式往往无法达到实时渲染.

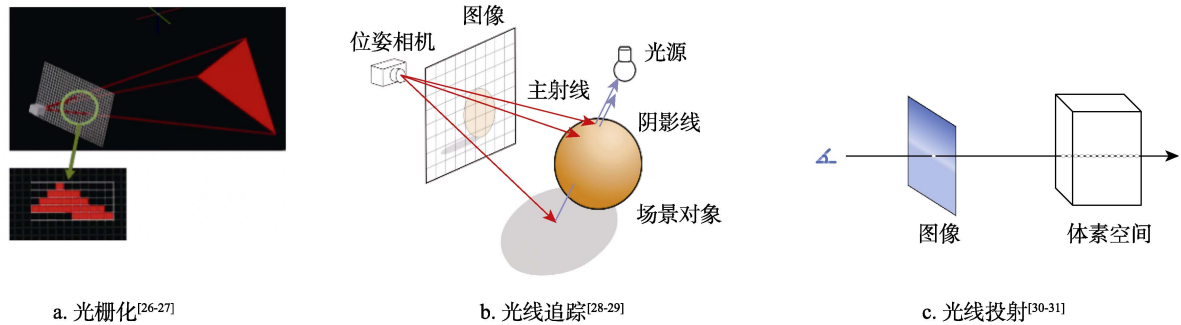


图 3 光栅化、光线追踪和光线投射示意图

2 NeRF 理论

NeRF 作为一种新颖的二维图像生成三维模型技术, 采用体积表征方式, 将场景定义为输入位置(三维)和观察视角(二维)的五维辐射场函数, 并通过多层感知机(multilayer perceptron, MLP)来表示, 输出为场景的体密度和观察视角下的二维 RGB 图像.

NeRF 模型的基本形式如图 4 所示, 将三维场景表示为神经网络函数近似的辐射场, 从 d 方向上接收到的总辐射为

$$L_o(x, d) = L_e(x, d) + \int_{\Omega} f_r(x, d, \omega_i) L_i(x, \omega_i) \cos \theta d\omega_i \quad (1)$$

其中, x 表示当前待分析的三维空间坐标; d 表示光线照射方向; $L_e(x, d)$ 表示 x 为光源点时自身在 d 方向上的辐射量; $\int_{\Omega} f_r(x, d, \omega_i) L_i(x, \omega_i) \cos \theta d\omega_i$ 表示

点光源照射到其他表面后, 折射在 d 方向的辐射, $f_r(x, d, \omega_i)$ 表示散射函数, $L_i(x, \omega_i)$ 表示从 ω_i 向接收到的辐射, θ 表示 ω_i 与 d 的夹角. 光就是电磁辐射, 因此建模辐射光即是建模对应的颜色. NeRF 则是一组可以对渲染方程式(1)近似求解的 MLP, 由辐射场空间内某点的三维坐标 x 和观测方向 d , 通过一个神经隐函数得到该点的对应颜色 c 和密度 σ , 公式为 $F: (x, d) \rightarrow (c, \sigma)$.

根据以上神经网络近似辐射场理论, NeRF 的结构如图 5 所示.

(1) 对五维坐标进行位置编码, 由于 MLP 更倾向于学习映射关系中的低频信息, 因此导致合成新视图中, 颜色纹路复杂部分的高频信息出现纹理混叠、罗马纹等问题. 将低频信息投射到高频空间中实现数据升维, 五维坐标中采样点 $X(x, y, z)$

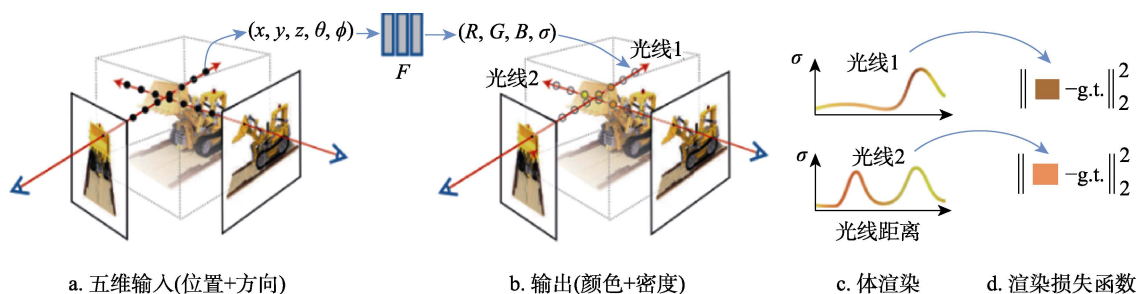


图 4 NeRF 模型运行机制^[6]

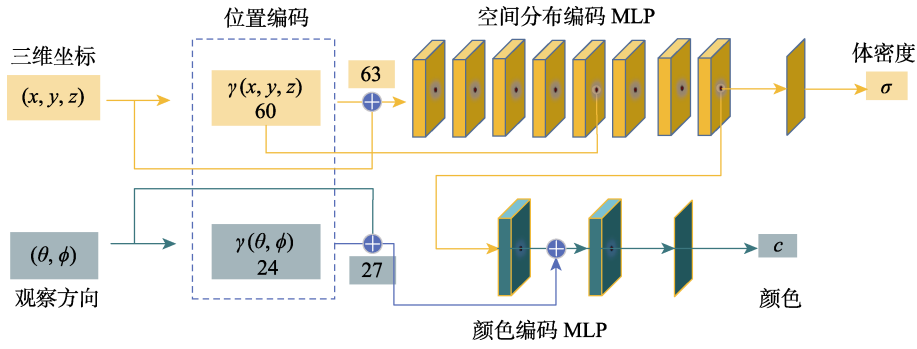


图 5 NeRF 的结构

的三维空间坐标升至 60 维, 二维观察方向 $d(\theta, \phi)$ 升至 24 维, 根据神经网络中输入数据越丰富、输出结果效果越好的理论, 经过原子间位置编码 (interatomic positional encoding, IPE), 增强对多彩复杂纹理等高频特征的渲染质量, 公式为

$$\gamma(p) = (\sin(2^0 \pi p), \cos(2^0 \pi p), \dots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p)).$$

(2) 采用神经网络编码体密度特征, 其架构由 8 层 MLP 构成, 其输入为三维坐标 $X(x, y, z)$ 位置编码后的 63 维向量, 输出为采样点体密度 σ 和 256 维特征向量; 随后该特征向量合并观察方向 $d(\theta, \phi)$ 编码后的 27 维向量, 经过 2 层颜色编码 MLP, 获得颜色值 c . 映射数据基于体渲染方程计算像素点投射射线的合成像素值

$$C(r) = \int_n^t T(t) \sigma(r(t)) c(r(t), d) dt.$$

其中, c 表示采样点颜色; σ 表示体密度; r 和 d 分别表示相机投影射线上的距离和方向; t 表示在相机投影射线上采样点与相机光心的距离, 通过 Coarse-Fine 的采样方式确定三维空间物体所在的位置; $T(t) = \exp\left(-\int_n^t \sigma(r(s)) ds\right)$ 表示采样点粒子的不透明度的总和, 其累积方程的物理意义是人眼看物体时只能观察到表面颜色.

(3) 将输出的合成像素点大小与光线投射的真实像素值进行比较, 确定损失函数 L (光度损失均方误差), 完成 NeRF 模型的收敛 (同时优化 Coarse 和 Fine 双网络结构). 该过程的公式为

$$L = \sum_{r \in R} \left[\|\hat{C}_c(r) - C(r)\|_2^2 + \|\hat{C}_f(r) - C(r)\|_2^2 \right].$$

真实像素点与预测像素点的差分可微是使用神经网络收敛 NeRF 模型的本质; 同时也体现了神经体渲染的另一优点——MLP 是回归模型, 可以实现输入未知的五维坐标, 得到该采样点对应体密

度与颜色, 促使 NeRF 具有强大的新视图合成功能.

2.1 相参位姿估计

NeRF 模型驱动时, 需要将像素点投射光线相机的姿态, 以及布署三维空间中采样点的世界坐标系位置输入至 MLP 以获取体密度、颜色信息合成模型和新视图. 通常, 基于 COLMAP, Reality-Capture, Agisoft, Metashape 等三维软件, 通过 SFM^[19], 根据输入非结构化视图集合之间的特征点检测匹配, 估算相机参数和姿态提供空间采样点的特征信息, 但估计的过程受视图自身的多因素影响 (模糊、断续视角等), 存在计算误差. 因此, 部分模型选择在训练过程中嵌入学习算法, 反演视图相机参数. 本文对该类 NeRF 变体估测的相机位姿及轨迹准确度进行实验, 包括绝对轨迹误差 (absolute trajectory error, ATE)^[32], 度量相机位置估计的误差 (relative pose error translation, RPEt) 和用于度量相机姿态估计的误差 (relative pose error rotation, RPER), 如表 1 所示. 实验中使用的数据集如下: (1) ScanNet 数据集^[33]. 一个大型三维数据集, 包含 1 513 幅 RGB-D 扫描图像, 涵盖 707 个独特的室内环境、估算的相机参数、表面模型、纹理网格和语义分割实例; 该数据集还为部分扫描提供了 CAD 模型. (2) Tanks and Temples 数据集^[34]. 由工业激光扫描仪采集, 涵盖在真实条件下获取的标准视图序列和高分辨率视频片段; 该数据集还分化了训练和测试样本类, 将测试数据分为中级和高级类别, 以适应不同复杂度的重建任务.

表 1 中, 还对各类优化算法进行定性和定量分析. 其中, 实验平台为 Intel i9-13900K CPU 和 RTX 4090 GPU; 测试和训练源码, 以及数据集的遴选均为原文献作者给出的开源项目. 对比算法如下:

(1) 2021 年, Wang 等^[35]提出了 NeRF-, 在无相机参数位姿的情况下实现 NeRF 渲染. 在模型训练时, 相机参数配置一系列初始值, 并将其设置为可

学习的参数 Π , 嵌入 NeRF 的光度损失中共同收敛优化, 在新视图合成中迭代配准, 公式为

$$\Theta^*, \Pi^* = \arg \min_{\Theta, \Pi} L(\hat{I}, \hat{\Pi} | I).$$

其中, I 和 \hat{I} 分别代表观测与合成视图; Θ 和 Π 表示相机参数(包含参数及各自由度位姿), 通过最小化收敛误差实现初始位姿优化.

(2) 基于相机参数、位姿优化和嵌入 NeRF 收敛方程的思路, 2021 年, Jeong 等^[36]提出 SC-NeRF, 通过学习场景的几何形状, 应用一个新的几何损失函数——投影射线距离损失, 结合非线性相机模型的几何一致性, 可以在没有 COLMAP 估算情况下设定初始值优化位姿. 文献[35-36]皆为正向嵌入优化函数校正位姿信息, 与其相反, 在 NeRF 中可以通过模型梯度下降反演视图位姿参数.

(3) 2021 年, Lin 等^[37]提出 iNeRF, 通过反演物体或场景的 NeRF 进行姿态估计. 对于观测图像 I

的未知相机位姿 \hat{T} , 使用 NeRF 一致的光度损失函数 L , 利用预训练权重 Θ 更新姿态 T 以最小化 L , 公式为 $\hat{T} = \arg \min_{T \in \text{SE}(3)} L(T | I, \Theta)$.

(4) 2021 年, 针对相机位姿存在误差, 缺少(甚至未知)NeRF 隐式表征训练以及多帧同场景配准的联合问题, Lin 等^[38]提出 BARF 模型. 该模型将经典二维图像对齐方法扩展至三维空间, 建立理论支撑, 采用相机位姿与光度损失联合优化方式, 解决了相机姿态失调问题; 但该模型仅适用于拍摄对象在相机前方的场景, 对于大范围的镜头移动场景仍然面临困难.

(5) 2023 年, Bian 等^[39]提出 NoPe-NeRF, 首先从单深度估计网络生成单目深度图重建稠密点云, 然后将未失真的深度图反向投影到点云中; 并提出点云损失和光度损失 2 个约束相对姿态的收敛函数, 共同优化 NeRF 模型, 处理具有挑战性的相机轨迹.

表 1 同场景各类优化算法位姿轨迹准确度对比

数据集	场景	NoPe-NeRF ^[39]			NeRF- ^[35]			SC-NeRF ^[36]			BARF ^[38]			iNeRF ^[37]		
		RPEt ↓	RPEr ↓	ATE ↓	RPEt ↓	RPEr ↓	ATE ↓	RPEt ↓	RPEr ↓	ATE ↓	RPEt ↓	RPEr ↓	ATE ↓	RPEt ↓	RPEr ↓	ATE ↓
ScanNet ^[33]	0079_00	0.752	0.204	0.023	1.706	0.636	0.100	2.064	0.664	0.115	<u>1.110</u>	<u>0.480</u>	0.062	1.263	0.520	<u>0.042</u>
	0418_00	0.455	0.119	0.015	1.402	<u>0.460</u>	0.013	1.528	0.502	<u>0.015</u>	<u>1.398</u>	0.538	0.020	1.438	0.531	0.201
	0301_00	0.399	0.123	0.013	3.097	0.894	0.288	1.133	0.422	0.056	<u>1.316</u>	0.777	0.219	1.329	<u>0.312</u>	<u>0.033</u>
	0431_00	1.625	0.274	0.069	6.799	0.624	0.496	<u>4.110</u>	<u>0.499</u>	0.205	6.024	0.754	<u>0.168</u>	5.312	0.622	0.313
	mean	0.808	0.180	0.030	3.251	0.654	0.224	2.209	0.522	<u>0.098</u>	2.464	0.637	0.117	2.923	<u>0.413</u>	0.219
Tanks and Temples ^[34]	Church	0.034	0.008	0.008	0.626	0.127	0.065	0.836	0.187	0.108	<u>0.114</u>	<u>0.038</u>	<u>0.052</u>	0.623	0.125	0.098
	Barn	0.046	0.032	0.004	1.629	0.494	0.159	8.339	1.491	0.316	<u>0.314</u>	<u>0.265</u>	<u>0.050</u>	1.568	0.535	0.128
	Museum	0.207	0.202	0.020	4.134	1.051	0.346	<u>1.171</u>	<u>0.499</u>	<u>0.142</u>	3.442	1.128	0.263	3.922	0.638	0.211
	Family	0.047	0.015	0.001	2.743	0.537	0.120	<u>1.366</u>	<u>0.438</u>	<u>0.019</u>	1.371	0.591	0.115	3.012	0.507	0.119
	Horse	0.179	0.017	0.003	1.349	0.434	0.018	<u>0.328</u>	<u>0.146</u>	<u>0.012</u>	1.333	0.394	0.014	1.203	0.397	0.018
	Ballroom	0.041	0.018	0.002	0.449	<u>0.177</u>	0.031	<u>0.233</u>	0.483	0.192	0.531	0.228	<u>0.018</u>	0.369	0.210	0.092
	Francis	0.057	0.009	0.005	1.647	0.618	0.207	0.328	<u>0.146</u>	<u>0.012</u>	<u>1.321</u>	0.558	0.082	1.527	0.430	0.153
	Ignatius	0.026	0.005	0.002	1.302	0.379	0.041	<u>0.533</u>	<u>0.240</u>	0.085	0.736	0.324	<u>0.029</u>	1.002	0.315	0.050
	mean	0.080	0.038	0.006	1.735	0.477	0.123	<u>0.890</u>	0.489	0.129	1.046	<u>0.441</u>	0.078	0.998	0.465	<u>0.101</u>

注. 粗体表示最优值, 下划线表示次优值.

2.2 视图集合标准

MLP 体绘制机理揭示了图像的数量(训练光线数量)和质量(差分收敛)会影响渲染的效果. 用少量的图像生成强大的渲染效果是 NeRF 发展追求的一类目标.

2021 年, Yu 等^[40]提出 pixelNeRF, 可从一幅或一组少量图像合成 NeRF. 针对单视图, 由卷积网络提取图像的底层特征进行编码, 并且以单幅图像预测空间参数作为整个场景的采样点的坐标系, 与坐标和观测方向结合获得颜色和体密度进行渲

染重建, 合成某个位置的新视图; 多视图集合由单视图扩展, 面对空间中某点, 提取多视图中存在的多角度投影特征, 通过平均池化算子聚合, 以获得预测的密度和颜色. 与 pixelNeRF 相似, 2021 年, Chen 等^[41]提出 MVSNerf 模型, 首先通过二维卷积神经网络(convolutional neural network, CNN)提取二维图像特征, 再将该特征扭曲到参考视图的平截头体的扫描平面上, 在输入参考视图处构成本体体积; 然后应用三维 CNN 卷积核重建一个具有每体素特征的神经编码体; 最后在空间任意位

置, 使用 MLP 从编码体积内插的特征对体密度和 RGB 辐射度进行回归。

2021 年, Tancik 等^[42]出 Meta-NeRF 模型, 应用标准的元学习算法从训练场景中恢复数据驱动先验, 填补测试场景中的缺失信息, 实现稀疏视图下的 NeRF 合成。与现有模型相比, 2021 年, Kosiorek 等^[43]提出的 NeRF-VAE 模型在多个场景之间学习共享信息, 能够从更少的输入视图中重建新角度场景。NeRF-VAE 是唯一的推销 NeRF 变体, 它以潜在变量的形式使用紧凑的场景表示, 并且可以处理输入中的不确定性。2021 年, Chibane 等^[44]提出立体辐射场(stereo radiance fields, SRF), 其是一种端到端训练的新视图合成技术。受到经典多视图立体方法的启发, SRF 仅需稀疏的视图集合, 通过在立体图像中找到相似的图像区域来估计表面点, 输入图像中立体对应关系的编码, 可以预测每个三维点的颜色和密度, 并且能够泛化到新的场景。

2021 年, Jain 等^[45]提出了 DietNeRF, 其中引入语义一致性损失, 在单个场景上进行训练。语义损失使其能够从任意姿势监督 DietNeRF, 从零开始学习时提高了少量视图合成的感知质量; 在多视图数据集上进行预训练时, 可以用最少一幅观察到的图像呈现新视角图像, 证明了语义损失与深度信息一致时, 可作为 NeRF 渲染进程的监督辅助项。Wang 等^[46]提出了一种名为“图像基渲染(image based rendering, IBR)”的技术, 结合了多视角图像和深度学习。该模型基于 Transformer 架构, 能够处理来自不同视角的输入图像, 并生成高分辨率的光照场, 同时利用插值技术, 提高了低分辨率输入图像的细节质量和视觉保真度。2022 年, Johari 等^[47]提出 GeoNeRF, 该方法包括几何形态推理器和渲染器 2 部分。首先, 几何推理器为源视图序列构建级联的代价体积; 然后, 借助基于 Transformer 的注意力机制和级联的代价体积, 渲染器推断几何形态与外观表征; 最后, 采用体渲染技术渲染精细图像。该架构允许对复杂的遮挡推理, 并扩展了一种适用于 RGB-D 图像的 NeRF 模型, 可以利用深度传感器提供的深度信息优化渲染效果。2023 年, Mu 等^[48]提出一种融入几何先验实现三维隐式重建的 NeRF 方法, 采用 SDF 作为三维表示, 使用相应的深度图构建特征体积, 额外添加深度和表面法线约束, 以恢复细密的几何纹理, 从稀疏视图中学习普适的三维模型。2022 年, Xu 等^[49]提出一个单视图 SinNeRF 模型, 其由特定语义和几何正则化组成, 构建了一个半监督学习过程。通过引入并传播

几何伪标签和语义伪标签指导渐进式训练过程, 呈现出逼真的新视图合成效果。当前, 绝大部分 NeRF 模型渲染需要具有已知相机位姿的图像, 应用 COLMAP 软件估测的不仅有相机参数位姿, 还包含具有深度语义的稀疏点云模型。2022 年, Deng 等^[50]提出了 DS-NeRF, 在训练期间将该信息用作深度监督并添加损失, 促使射线的终止深度分布与点云的三维关键点匹配。该方法可以在训练视图少的情况下渲染效果适宜的图像, 同时提高训练速度; 基于融合深度损失的实验效果表明, 深度是一种具有通用性和实用性的监督信息。

2023 年, Deng 等^[51]提出不同方式的单视图合成模型 NeRDi, 其具有来自二维扩散模型的一般图像先验, 在输入视图约束下, 使用预训练的图像扩散模型最小化其任意视图渲染的扩散损失优化 NeRF, 引入 2 部分语义指导作为扩散模型的条件输入。NeRDi 在本质上有有助于提高多视图内容的连贯性; 此外, 引入基于估计深度图的几何损失, 用于规范 NeRF 的底层三维几何结构。

本文在 NeRF Real(NeRF real-world data)^[52]和 DTU^[53]数据集上, 对面向稀疏视图组合的多类 NeRF 模型的渲染效果进行实验, 结果如表 2 所示。其中, 采用峰值信噪比(peak signal-to-noise ratio, PSNR)^[54]、结构相似性(structural similarity, SSIM)^[55]和学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)^[56]指标进行评估。NeRF Real 数据集作为新颖的视角合成研究工具, 包含使用 SUNCG^[57]和 UnrealCV^[58]创建的人工图像, 以及便携式手机拍摄的 24 幅真实场景照片; 此外, 还提供了一种高效的视角合成算法, 将多平面图像场景描绘与局部光场融合, 拓展了传统的全光采样理论。DTU 数据集采用多视角立体格式, 包含 80 个高度多样化的场景, 每个场景均由 49 或 64 个精确校准的相机位置和结构光标定扫描组成, 获取了分辨率为 1 200×1 600 像素的高质量 RGB 图像。

图像质量影响 NeRF 模型的精度, 虽然模型重建后有很多有针对性地提高分辨率和清晰度的算法, 但是它们的合成效果并不贴合现实场景, 且无法还原高频区域纹理细节; 在野外拍摄场景时, 经常会出现散焦或运动引起的图像模糊, 极大地降低了合成的 NeRF 质量。为了解决这个问题, 2022 年, Ma 等^[59]提出 Deblur-NeRF 模型, 其采用一种综合分析方法模拟图像模糊过程, 并重建模糊视图; 其核心思想是通过可变形稀疏核(deformable sparse

表 2 在 2 个数据集下多类 NeRF 模型的渲染效果对比

NeRF 模型	PSNR/dB \uparrow			SSIM \uparrow			LPIPS \downarrow		
	2 视角	5 视角	10 视角	2 视角	5 视角	10 视角	2 视角	5 视角	10 视角
NeRF ^[6]	13.5	18.2	21.5	0.39	0.57	0.67	0.56	0.50	0.52
Pixel-NeRF ^[40]	9.6	9.52	9.7	0.39	0.43	0.40	0.82	0.87	0.81
Meta-NeRF ^[42]	13.1	13.8	14.3	0.43	0.45	0.46	0.89	0.88	0.87
MVSNeRF ^[41]		17.2	17.2		0.61	0.60		0.37	0.36
DS-NeRF ^[50]	20.2	22.6	24.9	0.67	0.69	0.72	0.39	0.33	0.34
NeRF-VAE ^[43]	13.6	18.3	20.8	0.38	0.56	0.63	0.54	0.49	0.48
IBRNet ^[46]	14.4	<u>21.8</u>	<u>24.3</u>	0.52	0.51	0.54	0.53	0.54	0.51
NeRF(DTU) ^[6]	9.91	18.6	22.1	0.37	0.72	0.82	0.62	<u>0.35</u>	<u>0.26</u>
Pixel-NeRF(DTU) ^[40]	19.3	20.4	21.1	0.70	<u>0.73</u>	0.76	0.39	0.36	0.34
Meta-NeRF(DTU) ^[42]	18.2	18.8	20.2	0.61	0.61	0.67	0.40	0.41	0.35
DS-NeRF(DTU) ^[50]	16.9	20.6	22.3	0.57	0.75	<u>0.81</u>	0.45	0.29	0.24
DietNeRF(DTU) ^[45]	17.1			0.62			<u>0.38</u>		
SinNeRF(DTU) ^[49]	<u>19.6</u>			<u>0.69</u>			0.43		
NeRD _i (DTU) ^[51]	17.3			0.61			0.32		

注. 粗体表示最优值, 下划线表示次优值.

kernel, DSK)模块对每个核点的射线原点进行联合优化, 并将其参数化为一个 MLP, 使其能适用于各种模糊类型, 通过共同优化 NeRF 和 DSK 模块以恢复清晰的 NeRF 模型. 基于此方法, 2023 年, Lee 等^[60]提出 DP-NeRF 模型, 通过刚性模糊核增强三维一致性, 利用物理先验和自适应权重细化深度和模糊之间的关系, 降低颜色构成误差, 提高构建的 NeRF 感知质量, 以确保三维几何和外观的模型一致性. Deblur-NeRF^[59]模型基于图像像素上的二维偏移建立模糊核, 由于缺乏三维一致性先验, 破坏了训练 NeRF 的一致性; 然而, 该模型可以根据模糊发生的物理先验渲染 NeRF, 因此, 可保证相机刚性运动的三维一致性, 如图 6 所示.

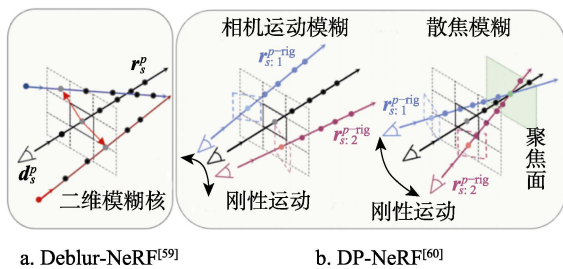


图 6 Deblur-NeRF^[59]与 DP-NeRF^[60]模糊核

现实场景中, 容易发生图像退化(如低光条件下的图像运动模糊), 影响 NeRF 的合成效果. 2023 年, Wang 等^[61]提出一种几何光束调节去模糊 NeRF(BAD-NeRF), 在曝光时间开始和结束时分别用 T_{start} 和 T_{end} 这 2 个姿态表示每幅图像的运动轨

迹, 中间虚拟相机姿势使用 SE3 空间线性插值方式获取. 由生成的虚拟视图合成模糊图像, 最小化合成图像和真实模糊图像之间的光度损失, 对整个网络实现去模糊优化, 虚拟相机在 t 时刻的姿态的公式为

$$T_t = T_{\text{start}} \cdot \exp\left(\frac{t}{\tau} \cdot \lg(T_{\text{start}}^{-1} \cdot T_{\text{end}})\right).$$

其中, τ 表示曝光时间, $t \in [0, \tau]$; T_{start} 和 T_{end} 分别表示曝光开始与结束时刻的相机姿态.

2023 年, Rudnev 等^[62]提出 EventNeRF, 面向运动模糊问题使用以单一颜色事件流作为输入的方法, 其核心是一个完全以自监督的方式从事件中训练的 NeRF, 保留颜色事件通道的原始分辨率, 采样方式根据事件特定, 其合成效果更具视觉吸引力. 从源图像中恢复高频纹理细节, 即使是高质量的训练框架, NeRF 模型生成的新视图仍然受到渲染伪影影响, 如噪声和模糊等. 为了提高基于 NeRF 模型的综合质量, 2023 年, Zhou 等^[63]提出一种通用的 NeRF 不可知的恢复器范式 NeRFLiX, 通过学习退化驱动的视点混合器, 有效地去除现有深度神经网络的 NeRF 原生渲染伪影. NeRF 退化模拟器的公式为 $I^{\text{Dl}} = (I^{\text{gt}} + n) \otimes g$. 其中, n 表示二维高斯噪声图; g 表示各向同性高斯模糊核; \otimes 表示各向同性高斯模糊核对添加高斯噪声的真实图像进行卷积运算. 从原始图像序列 $\{I, I_1^t, I_2^t | I^{\text{gt}}\}$ 构建成对训练数据 $\{I^{\text{gt}}, I_1^t, I_2^t\}$, I_1^t 采用高质量监督视图, 利用该模拟数据上训练的跨视点混合器从

NeRF 渲染图像中恢复增强帧 \hat{I} , 公式为 $L_{\text{Loss}} = |\hat{I} - I^{\text{gt}}|$, where $\hat{I} = f(I, I_1^r, I_2^r; \theta)$. 其中, θ 表示视点混合器可学习参数. 由 NeRF 退化模拟器产生降级视图 I , 从 2 个高质量监督视图 I_1^r 和 I_2^r 中提取有用的信息, 并恢复增强视图 \hat{I} .

面向重建实体进行多视角拍摄时, 相机前方常常存在各类遮挡, 导致部分图像特征受损, 影响 NeRF 的合成效果. 2023 年, Zhu 等^[64]提出了消除遮挡的神经辐射场变体, 由 2 层网络构成, 分别为场景 MLP 和背景 MLP. 首先设计成本体积损失策略, 指导场景 MLP 和相机参数的联合优化; 然后在背景 MLP 中应用深度双向不一致性先验, 通过嵌入深度距离监督还原无遮挡背景. 但是, 当遮挡对象为光损透明玻璃体时, 拍摄图像往往存在反射伪影的噪声污染, 导致该算法不能达到泛化作用. 为解决此问题, 2022 年, Guo 等^[65]提出 NeRFReN 模型, 利用 2 类几何先验设置深度光滑与双向深度一致性损失, 实现反射伪影与建筑实体分离. 该模型对于现实世界中的一般场景可以实现高效收敛, 但当实体与伪影复杂交叉时则无法区别两者的分离.

在夜间环境光照不足、可提取图像信息匮乏的情况下, 2022 年, Mildenhall 等^[66]提出 RawNeRF, 通过监督有噪声的原始输入图像, 在线性 HDR 色彩空间中重建场景. RawNeRF 绕过有损的后处理相机, 应用动态范围压缩和平滑噪声保留原始输入的完整动态范围, 采用图像降噪技术在夜晚的环境下渲染出很好的视图效果.

2.3 过程速度优化

渲染一幅分辨率为 720×1280 像素的图像时, 为了计算出每个像素的最终颜色值, NeRF 需在光线行进的方向上采样 $192(64+128)$ 个点, 并进行 256 次 MLP 训练, 意味着渲染该图像总共需要 235 929 600 次 MLP 训练, 每次推理计算需要超过 100 万次浮点运算, 整个过程需要超过 100T 次浮

点计算. 在源码 NeRF 基于标准自然数据集和常规硬件 RTX2060 驱动支持下, 渲染时间为 6~7 h, 不能满足快速渲染的需求, 庞大的计算量是影响渲染速度的问题之一. 抛开 GPU 读写运转机制的本体因素, 灵活的软件算法可以最大限度地缩减合成时间.

2022 年, Müller 等^[67]提出 Instant-NGP, 采取哈希编码的方式降低计算量, 加快了渲染速度, 其算法的流程如图 7 所示. 哈希编码允许在不牺牲质量的情况下使用更小的 MLP 网络, 大大减少了浮点数和内存访问操作的数量; 多分辨率结构允许网络消除哈希冲突的歧义, 形成一个简单的架构, 最大限度地减少浪费的带宽和计算操作. Instant-NGP 实现了几个数量级的组合加速, 可以在几秒钟内训练高质量的神经图形原语. 2023 年, Yang 等^[68]提出一个基于 Jittor 框架的 NeRF 基准库 JNeRF, 其支持快速训练, 性能和速度可以与 Instant-NGP 相媲美. JNeRF 利用 Jittor 的 JIT(just in time)编译特性, 使得模型能够在运行时被优化, 大幅提高了训练速度; 此外, 通过自动适应不同 GPU 架构, 能够充分地利用硬件资源, 进一步提升了计算效率. JNeRF 广泛应用于三维重建、虚拟现实、游戏开发以及电影特效等领域, 如可以用于快速地构建高逼真的环境模型, 或者在实时交互式应用程序中进行视角变换. 对于计算机视觉与图形图像学研究人员, JNeRF 是一个理想的平台, 可以便捷地实现不同的 NeRF 变体和技术.

神经网络在 NeRF 中并非不可或缺的, 其作用是近似经典渲染方程, 由五维坐标获取颜色和体密度, 并且可以生成新视角的图像. 2022 年, Fridovich-Keil 等^[69]提出一种未使用神经网络的 NeRF 模型 Plenoxels(全光体素), 其中, 场景表示为稀疏三维网格, 通过在光线投射中使用的三线性插值定理计算采样点的颜色和不透明度, 不需要任何神经网络成分. 在标准的基准任务中, Plenoxels 的优化

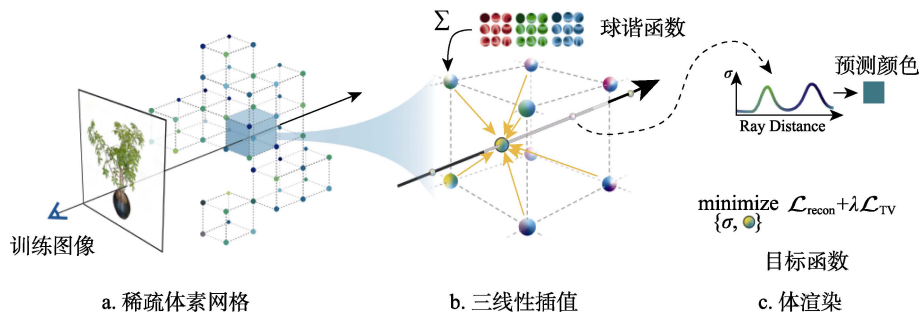


图 7 Instant-NGP 算法流程^[67]

速度比 JNeRF^[68]快 2 个数量级, 且视觉质量没有损失. 2023 年, Yang 等^[70]提出 Recursive-NeRF 模型, 它利用难易层次分化的方式渲染场景, 假设复杂区域由大型神经网络表示, 而简单区域则由小型神经网络编码, 保证了效率和质量之间的平衡. 该模型最终渲染的图像由各层次神经网络组合表征, 提供了一种高效且自适应的 NeRF 渲染和训练方式.

2021 年, Garbin 等^[71]提出的 FastNeRF 是第 1 个基于 NeRF 能够在 GPU 上以 200Hz 的频率渲染高保真的图像的模型. 它通过对 MLP 训练过程的分解, 空间位置和方向分量可以离散化并存储在稀疏的三维网格, 使得缓存网络的输出成为可能, 进而使用光线方向高效地查询渲染图像中的像素

值. FastNeRF 将问题分解为 2 个子网, 它们分别依赖于样本点的三维坐标和摄像机旋转方向, 尽管这种分解减少了缓存大小, 并以超过 200 fps 的速度执行推理, 但内存开销仍然很大. 在此基础上, 2022 年, Wadhvani 等^[72]提出 SqueezeNeRF, 其内存效率是 FastNeRF 的稀疏缓存的 60 倍以上, 并且在推理期间仍然能够在高规格 GPU 上以超过 190 fps 的速度渲染. SqueezeNeRF 将采样点的三维坐标再分解为 (x, y) , (y, z) 和 (z, x) 这 3 类结构, 对不同的结构产生的特征向量与体密度进行缓存离散化、组合计算, 得到体密度和颜色信息, 实现 NeRF 渲染. SqueezeNeRF 与 NeRF 原型网络结构对比如图 8 所示.

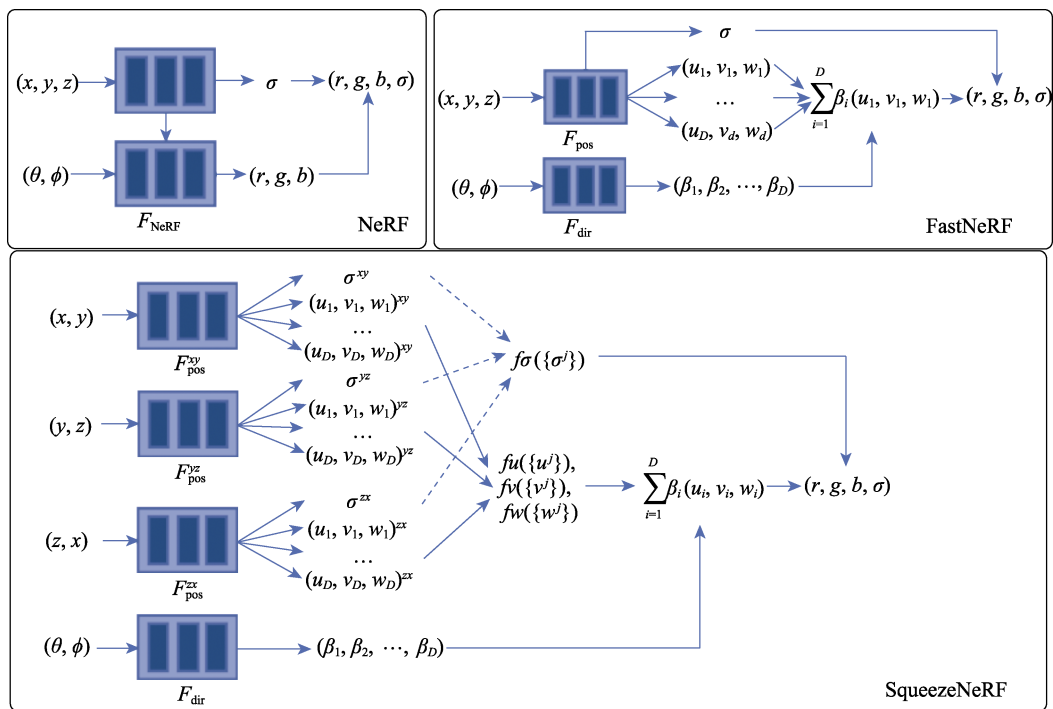


图 8 NeRF^[6], FastNeRF^[71]和 SqueezeNeRF^[72]网络架构

与其他三维技术相比, NeRF 模型在图像合成效果方面有着绝对的优势, 但在硬件运行机制上不能做到适应贴合. 2023 年, Chen 等^[73]提出 MobileNeRF, 将基于纹理多边形的 NeRF 表示方式与光栅化结合, 将重建模型表示为三角网格和纹理映射, 采用两阶段延迟渲染: 第 1 阶段, 不透明度构成的三角网格进行纹理贴图, 即在 GPU 内存中创建延迟渲染缓冲区; 第 2 阶段, 在片段着色器中运行神经延迟渲染器, 使用小型的 MLP 接收特征向量和观察方向, 将特征图像转化为彩色图像. MobileNeRF 加快 NeRF 运行速度, 减小消耗内存, 能够使用传统的多边形

光栅化技术, 提供大规模的像素级并行性.

为了加快体绘制过程, 推动效率和内存的权衡. 2023 年, Li 等^[74]提出 SteerNeRF, 探索了一种新的视角来加速 NeRF 渲染, 利用交互式视点控制中视点变化平滑和连续, 允许使用前面视点的信息减少渲染像素, 并且缩减剩余像素的光线投射采样点的数量. 在整个过程中, 低分辨率的特征图先通过体渲染, 再利用二维神经渲染器生成目标分辨率的输出图像; 通过之前和当前帧的特征对比, 证明该方法不仅可以减少渲染时间, 而且计算量很小, 所需内存也较小.

同时控制摄像机视图和时间变量,从而进行动态渲染.其中, $p(h,t)$ 表示变形后空间采样点位置.通过2个阶段的变形,规范实现在时间变量下的动态 NeRF 渲染.

重建非刚性场景的方法中,通过学习的变形场将每个输入图像中的采样坐标映射到一个规范化的模板坐标空间,但基于变形的的方法难以模拟拓扑变化,因为拓扑变化需要变形场中的不连续性,而这些变形场必须是连续的.为了解决此问题,2021年,Park等^[78]提出 HyperNeRF,将 NeRF 提升至更高维空间,并将每个单独输入图像对应的五维辐射场表示为通过这个空间的一个切片.

动态渲染计算量远远高于静态 NeRF 的合成,渲染速度是面临的一大难点.虽然显式数据结构,如体素特征对于加速训练过程有巨大的潜力,但是体素特征应用于动态场景面临2大挑战——建模时间信息和捕获不同尺度的点运动.2022年,Fang等^[79]使用时间感知的体素特征表示场景,通过引入一个小型坐标变形网络模拟粗糙的运动轨迹,并在辐射网络中进一步增强了时间信息,提出一种多距离插值方法 FDRF 应用于体素特征,以模拟小型和大型运动.该方法在保持高渲染质量的同时,显著地加速了动态辐射场的优化.基于离散体素执行动态非刚性场景的四维变形场建模时,受网络参数制约,模型训练易发生过拟合.为了应对这个挑战,2022年,Liu等^[80]提出 DeVRF,其设计一种静态-动态学习范式,以及实践中易于部署的数据捕捉设置,利用从多视角静态图像中学到的三维体积规范空间,简化只使用少量动态序列学习四维体素变形场的过程,实现了可变形 NeRF 的

高效学习.

动态渲染虽然可以从物体非刚性运动场景中得到逼真的合成新视图,但是从观察空间扭曲到一个通用的规范空间过程中,不会还原透明或者镜面物体引起的光学效应.2023年,Yan等^[81]提出 NeRF-DS,通过重新设定场函数解决上述问题,以观测空间中的表面位置和方向为条件,允许不同姿态的高光表面在映射到公共规范空间时保持不同的反射颜色.本文对非刚性运动场景 NeRF 变种在合成与真实世界数据集下效果进行实验,结果如表3所示,图11所示为其可视化结果对比.

3.2 大范围及无界场景

NeRF 模型针对小范围场景具有出色的重建渲染能力,但在城市建筑群、超分辨率自然环境等大范围场景中,易出现高频信息丢失、训练推理过程缓慢、无法捕捉主体渲染对象的情况.从捕捉城市概况的卫星遥感视图到显示建筑物复杂细节的地面图像,相机距离场景的宽跨度产生具有不同细节和空间覆盖水平的多尺度数据.2022年,Tancik等^[82]为了解决大型城市建筑群复杂的渲染问题,提出 Block-NeRF,采用“总—分—总”的方法将场景分解为区块单独训练的 NeRF;将渲染时间与场景大小解耦,使渲染能够扩展到任意大的环境,并且设计程序对齐相邻的 NeRF 之间的外观,使其无缝结合.2021年,Xiangli等^[83]提出一种渐进式学习范式 CityNeRF,同步增长 NeRF 模型和训练集.首先从浅基础块拟合远处的视图,随着训练的迭代,新的块不断被附加,以适应远-近视图中出现的细节.该方式有效地激活位置编码中的高频通道,并随着训练进程展开更复杂的细节.

表3 非刚性运动场景 NeRF 变种合成效果评价指标对比

NeRF 变种	合成数据集					真实世界数据集				
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	GPU/GB ↓	时间/h ↓	PSNR/dB ↑	SSIM ↑	LPIPS ↓	GPU/GB ↓	时间/h ↓
D-NeRF ^[77]	31.83	0.960	0.0355	10.0	18.40	29.16	0.946	0.0643	12.4	22.10
D-NeRF-dynamic	28.28	0.945	0.0528	9.7	18.50	27.21	0.934	0.0706	13.2	22.30
HyperNeRF ^[78]	<u>33.73</u>	0.965	0.0335	22.5	20.50	28.50	0.944	0.0692	22.0	20.50
HyperNeRF-dynamic	29.16	0.953	0.0555	22.5	19.20	26.53	0.935	0.0802	22.0	19.30
NSFF ^[76]	27.06	0.936	0.0811	21.4	<u>12.80</u>	28.44	0.939	0.0714	22.7	<u>15.30</u>
NSFF-dynamic	18.18	0.858	0.1929	15.0	15.50	19.90	0.909	0.0944	22.7	16.20
DeVRF ^[80]										
DeVRF-dynamic	34.29	0.982	0.0137	4.4	0.13	31.68	<u>0.972</u>	0.0289	6.6	0.18
Nerfies ^[75]	33.09	<u>0.989</u>	0.0432	21.8	18.70	<u>29.58</u>	0.980	<u>0.0576</u>	22.5	19.10
Nerfies-dynamic	32.37	0.991	<u>0.0322</u>	22.0	15.80	23.93	0.920	0.0878	22.0	19.70

注:粗体表示最优值,下划线表示次优值.

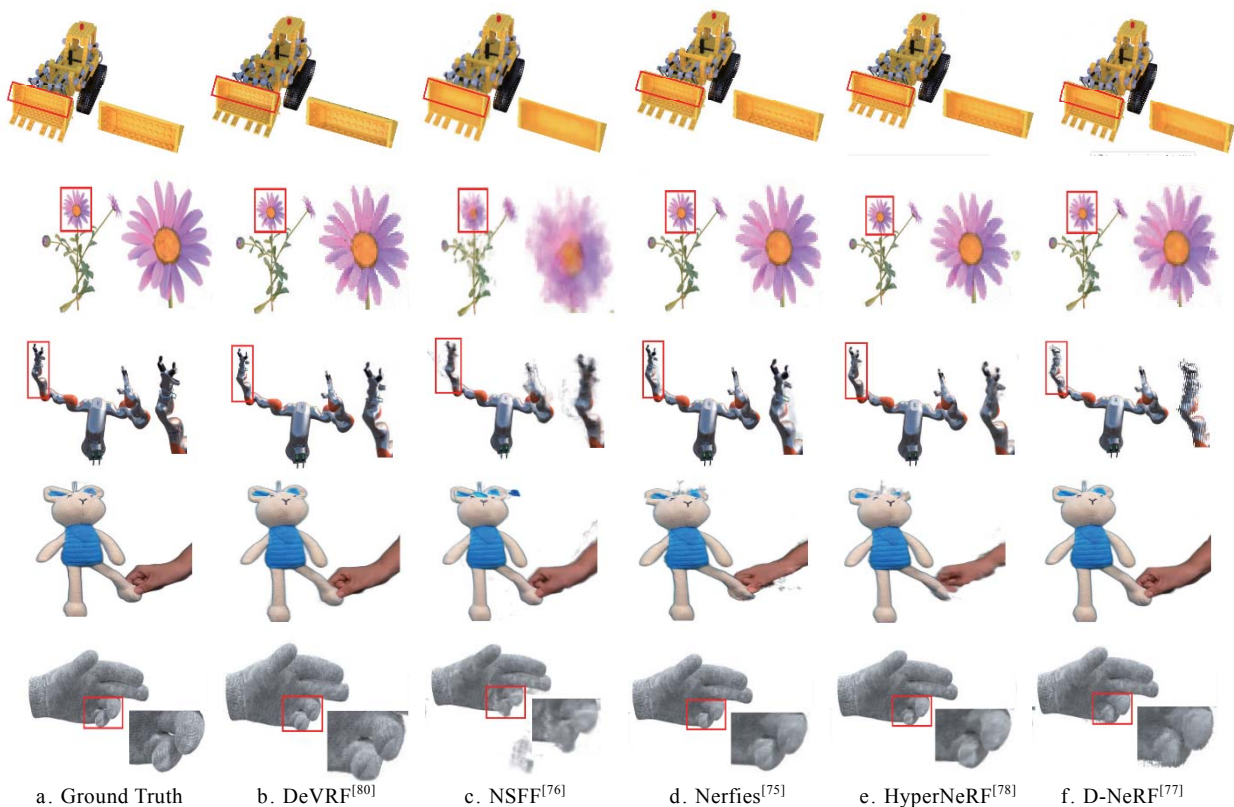


图 11 非刚性运动场景 NeRF 变种合成效果可视化对比

通过 NeRF 构建交互式三维环境, 从大型视觉捕获跨越建筑物、无人机多城市街区航拍等, 与单一对象场景(传统 NeRF)相比, 仍存在多个挑战. 2022 年, Turki 等^[84]提出 Mega-NeRF, 分析大规模训练场景的可见性统计数据, 其参数用于场景的不同区域; 为使数据并行, 提出一种几何聚类算法, 将训练图像划分为可以并行训练的不同 NeRF 子模块; 实验结果表明, 与传统的 NeRF 渲染相比, 该算法实现了 40 倍的加速, 同时 PSNR 质量保持在 0.8 dB 以内.

针对 360°无界场景的渲染问题, 2020 年, Zhang 等^[85]提出的 NeRF++是运用一种反球面参数化的方法, 其将场景空间划分为一个内部单位球体和一个外部反向球体 2 部分, 其中, 内球体包含前景和所有相机体系, 外球体包含场景内的剩余部分. 该方法将 NeRF 网络分为前景和背景网络分别训练渲染, 解决了一定的 360°无界场景模型合成问题. 2023 年, Kulkarni 等^[86]提出的 360FusionNeRF 基于半监督学习框架, 引入几何监督和语义一致性指导渐进式训练, 可以在保留场景特征的同时对未观察到的区域进行合理补全, 实现 360°无界场景渲染.

由稀疏视图集合构建大型室内场景辐射场表征时, 存在大型 MLP 超内存训练与高光以及伪影无法恢复问题. 2022 年, Wu 等^[87]提出了 NISR, 基于背景采样方式, 利用表面 MLP 和反射 MLP 这种双 MLP 架构, 分别实现与视图无关、相关的光学效应恢复, 并通过八叉树存储方式加速大规模场景渲染.

NeRF 模型采样基准通过像素点投射光线对场景采样, 在训练或测试数据集以不同分辨率(超大~小范围)观察场景时, 会产生过度模糊或混淆的渲染. 2021 年, Barron 等^[88]提出 Mip-NeRF, 以连续值比例表示场景, 通过高效地渲染抗混叠锥形锥体, 减少混叠伪影, 显著地提高了 NeRF 呈现高频细节的能力. 在处理无界场景时, 由于相机可能面对任何方向, NeRF 和 Mip-NeRF 的场景内容存在随机距离严重影响合成效果的问题. 2022 年, Barron 等^[89]提出 Mip-NeRF 360, 从参数化、效率和模糊 3 个方面, 分别采用非线性场景参数化、在线“蒸馏”和基于失真的优化应对无界场景渲染的挑战; 综合实验数据表明, 该方法能够生成逼真的合成视图和准确的深度, 用于复杂、无界的现实世界场景的重建.

2023 年, Isaac-Medina 等^[90]提出 Exact-NeRF, 解决了较大景深下的远处场景物体的渲染问题. Mip-NeRF^[88]和 Exact-NeRF^[90]的采样机理如图 12 所示. 可以看出, 与采用非线性场景参数化方式的图 12a 相比, 图 12b 改变了采样点编码结构, 基于方形金字塔的形状的采样方式, 获得了精确的 IPE, 可以呈现更好的渲染效果.

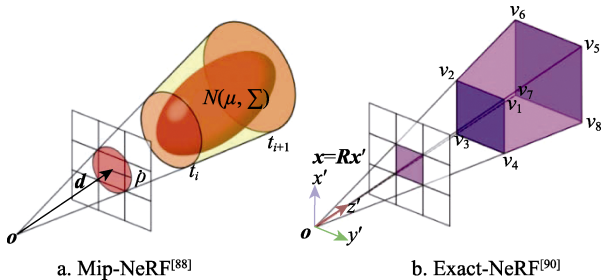


图 12 Mip-NeRF 和 Exact-NeRF 的采样机理

3.3 生成式结合与模型标注

生成式对抗网络(generative adversarial network, GAN)在人工智能领域引起一股热潮, NeRF 与 GAN 的碰撞引领了 NeRF 渲染在生成高分辨率新视图、可控三维模型方面的快速发展. 2020 年, Schwarz 等^[91]提出的 GRAF 奠定了两者结合的基础, 在应用条件辐射场控制 NeRF 模型的形状和外观的情况下, 引入鉴别器获得高分辨率的合成新视图, 但其仅局限于单物体场景. 为此, 2021 年, Niemeyer 等^[92]提出 GIRAFFE, 将场景表示为组合式生成神经特征场, 可以在没有明确监督的情况下, 将单个对象的形状、外观和背景分离; 它与体绘制-神经渲染器结合, 实现快速和可控的图像合成, 且将合成的三维场景结构纳入生成模型, 处理多对象问题. 但是, 以上方法均存在缺乏高质量图像的问题. 2021 年, Chan 等^[93]提出了一种新的生成式辐射场模型 pi-GAN, 称为周期隐式 GAN, 利用具有周期激活函数和体积渲染的神经模型表示场景, 实现视角一致的辐射场, 可用于高质量的三维感知图像合成. 但是, GRAF, GIRAFFE 和 pi-GAN 需要大量的数据, 更适用于简单的几何对象(如人脸, 汽车), 而不是逼真的复杂场景. 2021 年, Meng 等^[94]提出的 GNeRF, 可以在复杂场景中以随机初始化的姿态收敛准确位姿和辐射场. 该算法分为 2 个阶段: 第 1 阶段. 通过对抗性训练, 获取粗略的相机姿态和辐射场; 第 2 阶段. 光度损失精细化参数, 共同优化具有大位移的复杂场景中的三维表示和相机姿态.

基于二维 GAN 图像生成技术生成的图像效果高度逼真, 但忽略了底层三维场景的渲染过程. 2022 年, Sun 等^[95]提出第 1 个基于隐式场景表示的局部可编辑的三维人脸生成模型 FENeRF, 其生成器模块以解耦后形状与外观代码作为输入, 在具有共享几何形状的空间对齐的三维体积中生成相应的语义和纹理; 利用该底层三维表示, FENeRF 可以将边界对齐图像和语义掩码联合渲染, 并利用语义掩码, 通过 GAN 反演对体积场进行编辑, 增强三维 GAN 感知技术的可编辑性. 2022 年, Gu 等^[96]提出一种三维感知生成模型 StyleNeRF, 将三维场景表示与基于样式的生成模型合并, 利用 NeRF 生成低分辨率特征图并采样至高分辨率, 避免超计算量体渲染, 应用正则化项提高三维一致性, 解决了视角条件和噪声注入问题; 但该模型侧重二维图像空间操作, 缺乏三维表示. 2022 年, Chan 等^[97]提出了一种三维感知生成对抗网络, 通过混合显式-隐式三维表示提高三维基础渲染的计算效率, 并提出一种双重判别策略, 以维持神经渲染和最终输出之间的一致性. 该方法将特征生成与神经渲染框架解耦, 利用先进的 CNN 特征生成器 (StyleGAN2) 在三维场景的空间上进行泛化, 生成高质量的三维场景.

2023 年, Kania 等^[98]提出 HyperNeRFGAN 模型, 其流程如图 13 所示. 该模型使用超网络范式生成由 NeRF 表示的三维对象, 其 GAN 架构利用超网络范式将高斯噪声转化为 NeRF 训练权重, 进一步用于渲染二维新视图; 并以传统的二维鉴别器训练基于 GAN 的结构. 该模型的优势是为对象生成专用的 NeRF 表示, 不共享过程的全局参数.

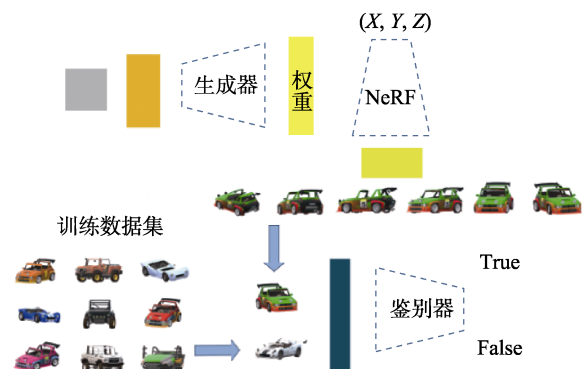


图 13 HyperNeRFGAN 模型^[98]流程

NeRF 显式表征的三维模型可开发诸多功能, 如类似于二维图像的特征语义分割、目标检测等模型标注算法. 2023 年, 为了检测场景中所有物体的

边界框, Hu 等^[99]提出 NeRF-RPN, 利用包含多尺度三维神经体积特征的体素表示, 可以直接在 NeRF 中回归物体的三维边界框, 无需从任何视点渲染 NeRF; 且 NeRF-RPN 是一个通用框架, 可以应用于检测无类标签的物体。

在模型的语义分割方面, 2023 年, Xu 等^[100]提出的 JacobiNeRF 方法不仅对场景的外观进行编码, 同样对场景点、区域或实体之间的语义相关性编码, 旨在捕获它们的相互共变模式; 通过显式规范化学习动态, 对齐高度相关实体的雅可比矩阵, 可在随机场景扰动下最大化它们之间的互信息; 基于该二阶信息合成了一个 NeRF 模型, 表达当网络权重沿着单个实体、区域, 甚至一个点的梯度变化时表现出的语义协同性, 进而实现 NeRF 语义分割。

3.4 引导模型可编辑

NeRF 模型虽然可以高保真地还原实时场景, 但在模型可编辑方面理论先验匮乏, 不能支持外界特征信息指导变形以适应各类需求。2021 年, Liu 等^[101]提出第 1 个在 NeRF 上执行用户控制形变模型 Edit-NeRF, 提出一个模块化网络的条件辐射场, 其中, $z^{(s)}$ 和 $z^{(c)}$ 分别表示融入运行机制的形状和颜色特征。对同一类别的多个实例训练, 该模型在无监督的情况下学习底层部件特征, 允许将用户在二维图像形状、纹理粗略地指引传播至三维空间, 公式为 $\hat{C}(r, z^{(s)}, z^{(c)}) = \sum_{i=1}^{N_c-1} c_i \alpha_i \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$ 。其中, $\alpha_i = 1 - \exp(-\sigma_i \delta_i)$ 与离散化体渲染方程一致。

2021 年, Yang 等^[102]提出的双路径架构

Object-NeRF, 其中, 场景分支编码几何与外观, 对象分支编码每个独立的对象; 条件是可学习的对象激活码, 可以对场景中目标实体进行分类编辑, 但只能支持复制、移动或旋转。Edit-NeRF 利用网络优化实现对整个图像和合成新视图的编辑传输, 但交互场景仅限于二维图像, 不能改变物体的空间位置。为了解决此类问题, 2022 年, Yuan 等^[103]提出 NeRF-Editing, 首先从预训练 NeRF 模型中提取显性三角网格, 直观地编辑变形, 从三角形网格表示中构建一个四面体网格裹着三角形网格; 然后利用三角形网格的变形驱动四面体网格, 将场景几何表面的变化传播到空间离散变形场; 最后利用四面体顶点插值, 完成从离散变形场到连续变形场的传播, 穿过四面体网格的光线会随着连续的变形场发生相应的弯曲, 使最终的渲染结果符合用户的编辑。

Edit-NeRF^[101]在形状操作方面的能力有限, 它只支持添加或删除对象的局部, 编辑过程因其迭代优化的特性变得缓慢。2022 年, Wang 等^[104]提出 CLIP-NeRF, 通过添置 CLIP(一种基于文本-图像对训练的多模态模型), 在形状操作和支持全局变形方面提供了更大的自由度, 还提出一种解纠缠的条件 NeRF 架构, 如图 14 所示。CLIP-NeRF^[104]将学习的变形网络应用于形状调节, 外观代码嵌入体绘制映射颜色输出, 允许对形状和外观进行控制。与此结构类似, 2022 年, Jain 等^[105]提出 Dream Fields, 通过图像-文本模型指导模型生成, 迫使渲染图像符合预训练的超容量 CLIP 模型目标语义; 为了提高保真度和视觉质量, 引入多类几何先验, 如透射率正则化、掩盖密度限制场景边界等。

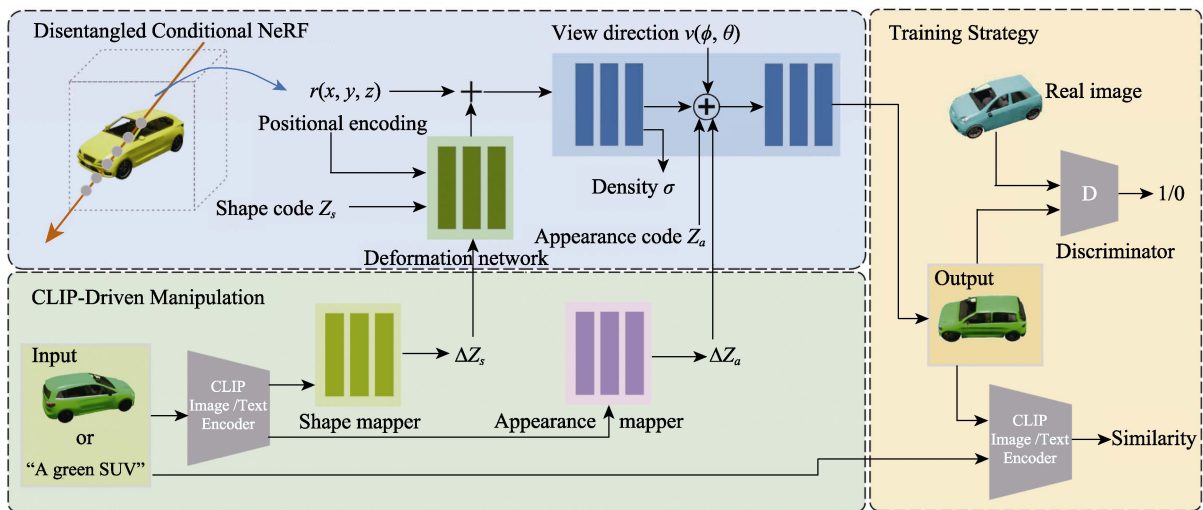


图 14 CLIP-NeRF 架构^[104]

预训练的图像-文本模型可以用于三维合成,但生成的对象缺乏真实感和视觉吸引力. 2023 年, Poole 等^[106]提出 DreamFusion, 其采用类似于 Dream Fields 的方法, 将 CLIP 替换为二维扩散模型的蒸馏损失, 最小化基于扩散前向过程与预训练扩散模型的、共享均值高斯分布族之间 KL 散度; 通过分数蒸馏采样方式进行采样, 并将其与 NeRF 变体相结合, 可以为用户提供的各种文本提示生成高保真、连贯的三维对象和场景. 文本到三维本质上是受约束的, 但可能缺乏引导或执行三维结构的能力. 2023 年, Metzger 等^[107]提出 Latent-NeRF, 应用类似 DreamFusion 的方案训练 Latent-NeRF. 在每次训练迭代中, 从一个随机视点渲染场景, 生成一个特征图 z ; 然后根据随机扩散步骤, z 会被添加噪声, 使用稳定扩散和文本提示对添置噪声的 z 进行去噪; 最后通过稳定扩散将输入噪声从预测噪声中减去, 以近似像素梯度, 反向传播至 NeRF 表示, 对合成体积场具有更优的三维引导效果.

还有一些方法不利用上述引入形状外观因子、变形场、扩散等操作, 在保持合成新视图真实感的同时有效地编辑模型的外观. 2023 年, Kuang 等^[108]提出的 PaletteNeRF, 将每个三维点的外观分解为基于调色板的碱基的线性组合, 在场景中共享; 同时, 预测一个依赖于视图的函数, 捕获高光、阴影、镜面反射的颜色残留; 在训练过程中共同优化基函数和调色板. 该方法允许用户通过修改调色板来有效地编辑三维场景的外观, 引导模型可编辑算法分类与关系, 如图 15 所示.

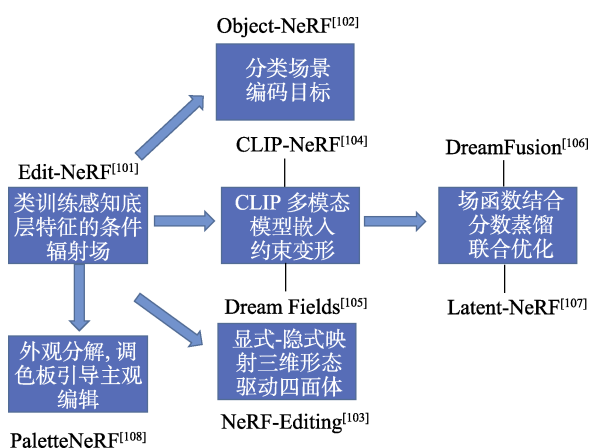


图 15 PaletteNeRF 算法^[108]结构

3.5 人体及细节特征建模

与非刚性物体相比, 人体具有更复杂的三维结构, 在渲染过程中不仅要关注运动状态, 其外部

条件, 如衣服的褶皱、人脸表情等皆增加了重建渲染的难度.

2021 年, Xu 等^[109]提出的 H-NeRF 中引入多个损失函数, 通过 SDF 结构化人体模型限制辐射场, 应用几何约束共同学习观察对象的结构, 包括头发和服装, 用于运动中人体的逼真合成. 2022 年, Weng 等^[110]提出 HumanNeRF, 其以单目视频作为输入, 优化为规范外观和从观察空间到规范空间的运动场; 利用身体姿势估计器纠正初始化的身体姿势, 将运动场分解为骨架刚性运动和非刚性运动, 使用离散网格和连续场表示; 最后叠加 LPIPS 与均方误差(mean square error, MSE)组建损失函数实现收敛, 重建的三维模型具有高保真度. HumanNeRF 从一个移动的单目视频中重建体积神经网络. 2023 年, Weng 等^[111]对 HumanNeRF 进行扩展, 提出 PersonNeRF, 能够从具有不同姿势和外观的非结构化照片集中重建三维模型, 利用 HumanNeRF 一致地映射变形运动场. PersonNeRF 的融合姿态和外观特征与 HumanNeRF 如图 16 所示, 模型可以面向跨时间线的人体服装和姿势图像集合, 渲染新颖的视点、姿态和外观组合图像.

2023 年, Gao 等^[112]提出 SketchFaceNeRF, 其基于粗略结构特征的二维草图引入三平面预测网络, 控制生成外观和纹理嵌入模型; 采用 Mask Fusion 模块将自由视图的二维 Mask 转换为三平面空间的三维 Mask, 引导原始人脸和基于草图生成的人脸融合, 实现高质量人体面部 NeRF. 该方法优于最先进的基于二维草图的图像生成和编辑方法.

SMPL 模型^[113-114]如图 17 所示, 其与 NeRF 的结合推进了神经隐式三维表征在人体建模领域的发展. 2021 年, Peng 等^[123]提出的隐式神经表示方法 Neural body 称为神经体, 用于从稀疏多视角视频中合成动态人物的新视角. 该方法定义一组潜变量, 利用神经网络编码局部几何和外观; 将潜变量锚定到可变形人体模型的顶点上, 表示动态人物. 在表演艺术中, 由于表演者多视角帧间姿态、外观等差距较大, 利用 SMPL 模型对人类表演者创建新颖的自由视点视频是一个高度不适宜问题. 2022 年, Cheng 等^[124]提出的 GNR 基于 SMPL 模型, 利用隐式几何身体嵌入方法和屏幕空间遮挡感知混合技术, 解决了此问题.

为了提高对新姿态和形状的泛化能力, 研究人员引入 SMPL 作为结构化几何体嵌入, 但不准确的 SMPL 估计会对重建结果产生负面影响. 2023 年, Chen 等^[115]基于校正后的多视图图像和配准的

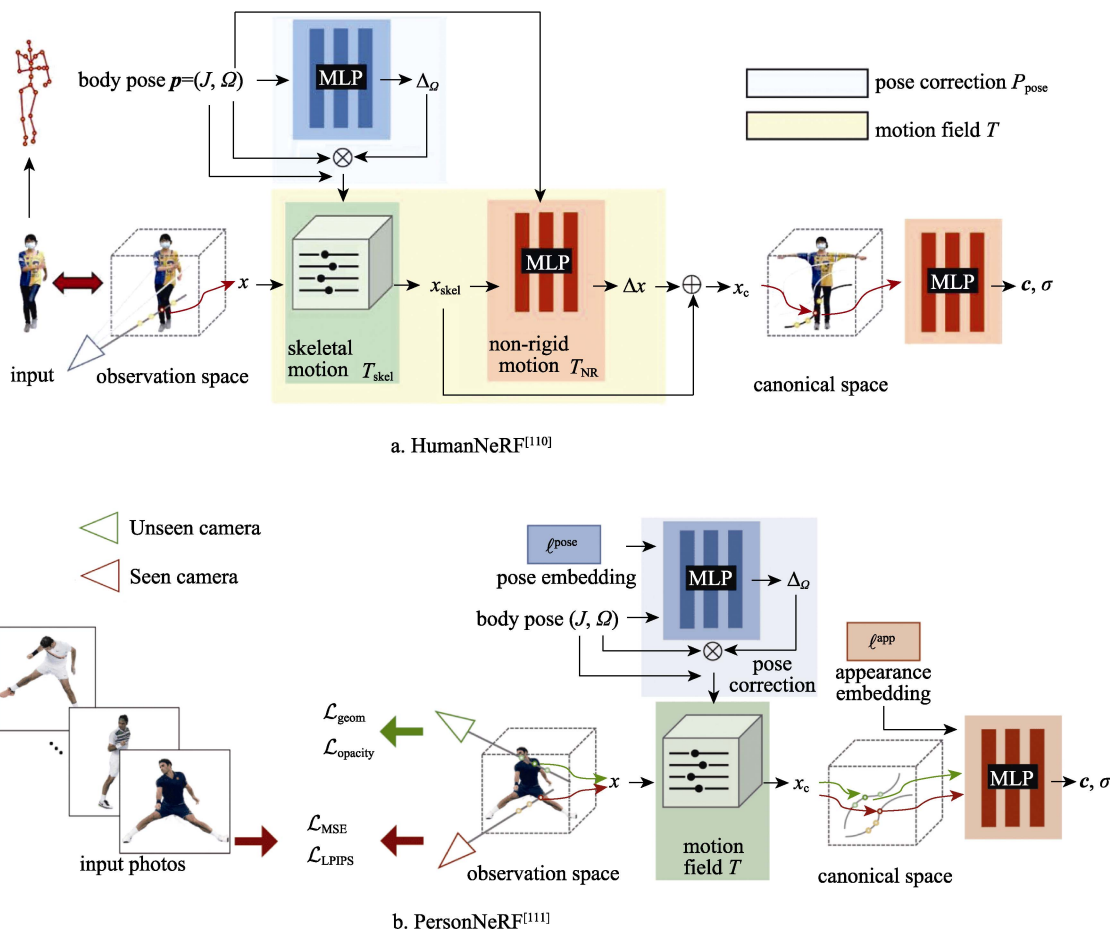


图 16 HumanNeRF^[110]与 PersonNeRF^[111]的差异

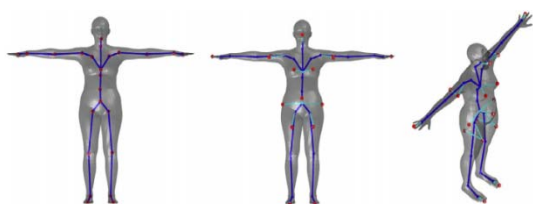


图 17 SMPL 模型^[113-114]

SMPL 提出 GM-NeRF. 首先利用图像编码器提取多视图几何和外观信息, 基于可见性的注意机制构建结构化的几何体嵌入, 并将其进一步扩散形成几何特征体; 对特征体空间中任意点进行三线性插值得到该点几何编码, 并应用几何引导注意力机制直接从多视图几何特征中获取外观代码; 再将几何编码和外观代码输入 MLP 网络中构建神经特征场; 最后采用体绘制与神经渲染生成新视图. 2022 年, Shao 等^[116]提出的 DoubleField 脱离 SMPL 类参数化模型先验知识, 采用双场框架建立表面场和辐射场之间的关联; 引入视图-视图转换器以有效地利用超高分辨率图像输入, 显著地提高了几何和外观的重建质量; 同时, 支持直接推

理、场景特定的高分辨率微调 and 快速渲染.

基于 NeRF 的方法可以合成逼真的人体细节, 但在人体模型存在自遮挡时, 效果不佳. 2021 年, Kwon 等^[117]提出一种基于参数化三维身体模型训练辐射场的方法, 除了利用参数体模型作为几何先验之外, 其核心是将时间和多视图转换器相结合, 有助于聚合时空观测信息, 可以鲁棒性地计算查询点的密度和颜色, 去除遮挡影响. 2022 年, Chen 等^[118]提出一种几何引导的多视图特征融合方法 GP-NeRF, 利用 SMPL 模型估计的几何先验整合输入视图中的不完整信息, 构建目标人体的完整几何体积.

上述方法着重于整体重建, 效果显著, 但在细节处, 如头部、人手等精度不佳. 2022 年, Mihajlovic 等^[119]提出 KeypointNeRF, 与 NHP 等依赖于精确的参数体拟合和时间特征聚合的方法不同, 该方法利用稀疏的三维关键点对相对空间三维信息进行编码, 这一特性使其对姿态的变化具有鲁棒性, 实现了高质量、多视图一致性的头部 NeRF 重建. 2022 年, Hong 等^[120]提出 HeadNeRF, 首先利用基

于 MIL 的隐式函数,并结合多类辅助监督信号(表情、反射率、光亮等)获取三维空间采样点的体密度信息与特征向量;然后渐进式地融合体绘制与二维神经渲染,得到最终合成图像;最后采用高效的收敛函数,提高过程速度与渲染精度.该方法优于最先进的头部重建算法.2023年,Guo等^[121]提出 HandNeRF,使用骨架估计器参数化手的姿态,设计一个姿态驱动的变形场,建立这些不同姿态与共享规范空间的映射关系;在正则空间中,优化一只手的姿态解纠缠 NeRF,增强双手在不易观察区域的几何和纹理,促使高质量地合成模型.

通常,人体 NeRF 合成需要进行数小时的模型训练,难以实际使用.2024年,Peng等^[122]提出 IntrinsicNGP,在哈希编码模块中引入一个连续的、可优化的内在坐标,应用这些坐标,可以利用代理几何形状聚合帧间信息,实现高效的动态对象建模.

结合 NeRF,本文在多个数据集上将人体建模 NeRF 变种进行定量实验,结果如表 4 所示.使用数据集如下:(1) Multi-garment 数据集^[125].该数据集涵盖 356 幅图像,2 078 个真实服装模型,涵盖 10

个类别、563 个服装实例;每件服装均有详细的注解信息,包含三维特征线(如领口、袖口轮廓、下摆等),三维人体姿态以及相应的多视角真实图像.(2) THuman2.0 数据集^[126].该数据集涵盖 500 个高质量人体扫描素材,且每个扫描都配有三维模型和纹理映射,不仅涵盖多样化的动作姿势,还提供了 SMPL-X 参数和对应的网格;借助 Function4D 技术,即便使用稀疏的 RGB-D 传感器,也能实现高精度的人体体积捕捉,并且可达到实时处理的速度;(3) ZJU-MoCap 数据集^[123].该数据集涵盖多个动态人类视频及其相应运动序列,利用包含 23 个同步摄像机的多相机系统进行拍摄,捕捉人类动作,如摇摆、转向、打击等;并生成相应的 SMPL-X 参数,精准地描述了躯干、四肢和表情的变化;(4) GeneBody 数据集^[124].该数据集由 100 名对象和 370 个图像序列构成,通过 48 台空间相机获取多视图数据,涵盖了多种姿势动作、服装、配饰和发型,以及从日常生活到专业场合的行为状态和各具特色的外观造型,每帧都提供了 SMPL-X 估计和前景分割,并且广泛分布在不同人种和年龄的群体中.

表 4 在 4 个数据集上人体建模 NeRF 变种方法的效果对比

NeRF 变种	Multi-garment ^[125]			THuman2.0 ^[126]			ZJU-MoCap ^[123]			GeneBody ^[124]		
	PSNR/ dB ↑	SSIM ↑	LPIPS ↓	PSNR/ dB ↑	SSIM ↑	LPIPS ↓	PSNR/ dB ↑	SSIM ↑	LPIPS ↓	PSNR/ dB ↑	SSIM ↑	LPIPS ↓
IBRNet ^[46]	28.44	0.924	0.0917	25.66	0.916	0.1033	25.25	0.876	0.2323	24.71	0.889	0.1364
NHP ^[117]	26.04	0.925	0.0701	26.99	<u>0.935</u>	0.0734	25.92	0.904	0.1623	22.75	0.872	0.1659
GNR ^[124]	28.61	0.937	0.0511	25.82	0.929	<u>0.0605</u>	25.39	0.903	0.1306	22.21	0.887	0.1254
KeypointNeRF ^[119]	28.36	0.938	0.0471	25.93	0.929	0.0607	25.85	0.911	<u>0.1092</u>	<u>24.34</u>	<u>0.902</u>	<u>0.1236</u>
GM-NeRF ^[115]	30.18	0.947	0.0305	28.88	0.952	0.0335	<u>26.74</u>	<u>0.919</u>	0.0955	23.90	0.906	0.0865
HumanNeRF ^[110]	<u>29.68</u>	<u>0.943</u>					28.18	0.962				
H-NeRF ^[109]	27.15	0.903	<u>0.1007</u>	24.93	0.871	0.1035						
GP-NeRF ^[118]	27.92	0.936	0.0635	<u>27.30</u>	0.932	0.0665	26.36	0.915		23.32	0.883	

注:粗体表示最优值,下划线表示次优值.

4 未来发展方向

综合近年来的文献信息,未来,NeRF 技术将在合成速度、视图数量、不同需求场景通用性、多类生成式算法和模型结构优化与反演 5 个方向上得到重点发展.

(1) 更快的渲染速度.本文综述的各类速度优化算法,如去除神经网络的全光体素和子网分解、离散缓存等技术,其针对小范围场景有一定效果,但对于大范围城市建筑群、超分辨率场景等,仅在位姿估计阶段就会消耗大量时间,后续光线处理

的计算量更加庞大,即使采用嵌入式迭代收敛,省略基于 COLMAP 等软件位姿计算过程,仍无法达到理想效果.因此,渲染速度优化是一个亟需解决的问题.硬件、软件方面需要进步,图形并行计算单元 GPU 基于强大的峰值计算能力和数据并行处理能力,适用于深度学习的训练阶段,形成复杂的神经网络模型;但是在深度学习预测阶段不具优势,平均性能不如专门的 AI 芯片,且功耗很大.软件算法的优化包括缩短相参位姿估算过程、融合辅助监督改进空间点采样编码方式,利用深度监督聚焦高密度采样点,高效的收敛函数降低训练

迭代次数, 将算法处理更贴合 GPU 硬件运行方式, 以减少占用内存量等。

(2) 更少的图像. 由二维图像集合生成三维建模是计算机视觉、计算机图形学发展中面临的巨大难题之一, 伴随 NeRF 技术的发展, 未来 NeRF 单视图的三维重建是主流研究方向. 本文介绍的点云预测、场景信息共享、二维图像卷积三维特征等稀疏视图三维重建算法中, 涉及的图像少但具有代表性, 能够提取大量图像的底层信息用于三维重建; 或者, 利用噪声扩散概率模型和基于分数的生成模型虽然可实现单视图 NeRF 合成, 但缺乏高质量的重建图像. 未来, 可以在场景先验获取方面提取更多辅助信息进行监督, 如几何结构、由局部纹理推测整体结构纹理, 或者改进模型采样编码方式, 充分地利用单视图底层特征的, 促进 NeRF 单视图三维重建的发展。

(3) 适应更多场景. 针对不同的需求场景, 在各种应用的场景中, 通用性 NeRF 模型具有一致性、效果出众的渲染算法, 它并不是目的性的单独场景优化, 如 Block-NeRF 的渲染效果在三维人体重建与城市建筑群达到一致性. 当前, 大范围使用的 Instant-NGP 虽然对于一般场景有泛化性, 但达不到图像的高质量还原, 针对不同的场景使用不同的算法往往费时费力。

(4) 多类生成式算法. 可编辑 NeRF 模型通过文本-图像指导约束形状、外观的改变或者与采样点颜色输出网络, 构建分类网络函数对模型颜色主动调节等. NeRF 可以延伸到同类别三维模型生成, 以现有的同类 NeRF 模型、点云、体素等进行特征学习, 实现端对端同维度变形指引, 与现在研究热潮的生成式 AI 类似. 通过多类 NeRF 之间提取特征, 交互学习生成新颖模型。

(5) NeRF 数据的结构优化和反演. NeRF 模型的文件输出状态有 image, video 和 mesh 这 3 种, 且作为三维显式表达的一种, mesh 格式模型可以实现与其他三维格式交互, 具有广泛的用途; 不考虑 NeRF 模型在图像-文本等形状指导方面的可编辑性时, 对于 mesh 与其他三维格式之间的转化, 现有的三维算法做不到高转化率. 在实际场景中, 对某栋不规则建筑大楼进行三维建模, 将 mesh 格式转向专业建模软件, 如 BIM^[127-128]; 3DMAX 中, 通过算法的转化一定存在数据的损失等问题, 需对转化后的 NeRF 模型进行结构纹理调整、网格优化等. NeRF 将光线投射的体渲染理论贯穿隐式场景中整个物体, 但若仅把重建中心放在物体的表面

将无法得到模型内部信息, 数据的反演算法通过照片信息或者合成模型表面的纹理信息推测物体内部结构模型, 可以与计算机视觉中基于图像信息的各种三维重建方法相结合, 推测内部场景, 并进行算法反演。

5 结 语

自 2021 年 Mildenhall 等^[6]提出 NeRF 至今, 在原始工作的基础上各类优化算法快速演变, 实现原始 NeRF 迭代升级; 广泛应用于三维建模、城市制图、摄影测量、图像编辑和视图合成等多个方向。

本文综合近年来各领域 NeRF 的改进算法, 剖析其关键技术特点, 对比差异与相似性, 量化分析各类 NeRF 优化算法应用效果. NeRF 在二维图像生成三维模型领域具有明显优势: (1) 与传统主动与被动式三维建模相比, 通过使用稀疏的输入视图集合优化底层连续的体积场景函数, 基本上实现了一比一图像与真实场景对照的建模效果, 时效性强. (2) 在不同应用场景中具有多类变体; (3) 与传统三维渲染与经典视图合成技术, 如光场渲染^[129]、局部光场融合^[40]、场景表示网络^[130]、神经体积^[131]等相比, NeRF 只需要图像和位姿学习场景, 在视觉质量方面有更好收敛效果, 并且在后续的优化中添加了各种能动性。

从未来的发展趋势看, 更快的渲染速度、更少的二维图像需求、更多的场景适应性、更强大的 AI 生成式算法, 以及更泛化的反演能力等, 将是 NeRF 技术发展演化的重要方向。

参考文献(References):

- [1] Aldoma A, Marton Z C, Tombari F, *et al.* Tutorial: point cloud library: three-dimensional object recognition and 6 DOF pose estimation[J]. IEEE Robotics & Automation Magazine, 2012, 19(3): 80-91
- [2] Amanatides J, Woo A. A fast voxel traversal algorithm for ray tracing[J]. Eurographics, 1987, 87(3): 3-10
- [3] Laine S, Karras T. Efficient sparse voxel octrees[C] //Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. New York: ACM Press, 2010: 55-63
- [4] Liu L J, Gu J T, Zaw Lin K, *et al.* Neural sparse voxel fields[C] //Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2020: Article No.1313
- [5] Piazza E, Romanoni A, Matteucci M. Real-time CPU-based large-scale three-dimensional mesh reconstruction[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1584-1591

- [6] Mildenhall B, Srinivasan P P, Tancik M, *et al.* NeRF: representing scenes as neural radiance fields for view synthesis[J]. *Communications of the ACM*, 2021, 65(1): 99-106
- [7] Yao M Y, Huo Y K, Ran Y, *et al.* Neural radiance field-based visual rendering: a comprehensive review[OL]. [2023-11-25]. <https://arxiv.org/abs/2404.00714>
- [8] Niu Y D, Liu L M, Huang F Y, *et al.* Overview of image-based 3D reconstruction technology[J]. *Journal of the European Optical Society-Rapid Publications*, 2024, 20(1): Article No.18
- [9] Li Jiyang, Cheng Lechao, He Jingxuan, *et al.* Current status and prospects of research on neural radiance fields[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2024, 36(7): 995-1013(in Chinese)
(李吉洋, 程乐超, 何靖璇, 等. 神经辐射场的研究现状与展望[J]. *计算机辅助设计与图形学学报*, 2024, 36(7): 995-1013)
- [10] He L, Li L H, Sun W C, *et al.* Neural radiance field in autonomous driving: a survey[OL]. [2023-11-25]. <https://arxiv.org/abs/2404.13816>
- [11] Nguyen T A Q, Bourki A, Macudzinski M, *et al.* Semantically-aware neural radiance fields for visual scene understanding: a comprehensive review[OL]. [2023-11-25]. <https://arxiv.org/abs/2402.11141>
- [12] Lin J W. Dynamic NeRF: a review[OL]. [2023-11-25]. <https://arxiv.org/abs/2405.08609>
- [13] Se S, Jasiobedzki P. Photo-realistic 3D model reconstruction[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2006: 3076-3082
- [14] Intwala A M, Magikar A. A review on process of 3D model reconstruction[C] //Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques. Los Alamitos: IEEE Computer Society Press, 2016: 2851-2855
- [15] Ma Z L, Liu S L. A review of 3D reconstruction techniques in civil engineering and their applications[J]. *Advanced Engineering Informatics*, 2018, 37: 163-174
- [16] Aharchi M, Ait Kbir M. A review on 3D reconstruction techniques from 2D images[C] //Proceedings of the 4th International Conference on Smart City Applications on Innovations in Smart Cities Applications Edition 3. Heidelberg: Springer, 2020: 510-522
- [17] Wang F, Zhao Z J. A survey of iterative closest point algorithm[C] //Proceedings of the Chinese Automation Congress. Los Alamitos: IEEE Computer Society Press, 2017: 4395-4399
- [18] Park J J, Florence P, Straub J, *et al.* DeepSDF: learning continuous signed distance functions for shape representation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 165-174
- [19] Vijayanarasimhan S, Ricco S, Schmid C, *et al.* SfM-Net: learning of structure and motion from video[OL]. [2023-11-25]. <https://arxiv.org/abs/1704.07804>
- [20] Jiang S, Jiang C, Jiang W S. Efficient structure from motion for large-scale UAV images: a review and a comparison of SfM tools[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 167: 230-251
- [21] Luo X, Huang J B, Szeliski R, *et al.* Consistent video depth estimation[J]. *ACM Transactions on Graphics*, 2020, 39(4): Article No.71
- [22] Shur M G. The Poisson theorem and Markov chains[J]. *Theory of Probability & Its Applications*, 1985, 29(1): 124-126
- [23] Dennis A R. An overview of rendering techniques[J]. *Computers & Graphics*, 1990, 14(1): 101-115
- [24] Kaufman A, Mueller K. Overview of volume rendering[M] //Hansen C D, Johnson C R. *Visualization Handbook*. Amsterdam: Elsevier, 2005: 127-174
- [25] Takikawa T, Litalien J, Yin K X, *et al.* Neural geometric level of detail: real-time rendering with implicit 3D shapes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 11353-11362
- [26] Laine S, Karras T. High-performance software rasterization on GPUs[C] //Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics. New York: ACM Press, 2011: 79-88
- [27] Li T M, Lukáč M, Gharbi M, *et al.* Differentiable vector graphics rasterization for editing and learning[J]. *ACM Transactions on Graphics*, 2020, 39(6): Article No.193
- [28] Cook R L, Porter T, Carpenter L. Distributed ray tracing[C] //Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1984: 137-145
- [29] Purcell T J, Buck I, Mark W R, *et al.* Ray tracing on programmable graphics hardware[J]. *ACM Transactions on Graphics*, 2002, 21(3): 703-712
- [30] Roth S D. Ray casting for modeling solids[J]. *Computer Graphics and Image Processing*, 1982, 18(2): 109-144
- [31] Dick C, Krüger J, Westermann R. GPU ray-casting for scalable terrain rendering[C] //Proceedings of the 30th Annual Conference of the European Association for Computer Graphics. Aire-la-Ville: Eurographics Association Press, 2009: 43-50
- [32] Sturm J, Engelhard N, Endres F, *et al.* A benchmark for the evaluation of RGB-D SLAM systems[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2012: 573-580
- [33] Dai A, Chang A X, Savva M, *et al.* ScanNet: richly-annotated 3D reconstructions of indoor scenes[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2432-2443
- [34] Knapitsch A, Park J, Zhou Q Y, *et al.* Tanks and temples: benchmarking large-scale scene reconstruction[J]. *ACM Transactions on Graphics*, 2017, 36(4): Article No.78
- [35] Wang Z R, Wu S Z, Xie W D, *et al.* NeRF--: neural radiance fields without known camera parameters[OL]. [2023-11-25]. <https://arxiv.org/abs/2102.07064>
- [36] Jeong Y, Ahn S, Choy C, *et al.* Self-calibrating neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5826-5834
- [37] Lin Y C, Florence P, Barron J T, *et al.* iNeRF: inverting neural radiance fields for pose estimation[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2021: 1323-1330

- [38] Lin C H, Ma W C, Torralba A, *et al.* BARF: bundle-adjusting neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5721-5731
- [39] Bian W J, Wang Z R, Li K J, *et al.* NoPe-NeRF: optimising neural radiance field with no pose prior[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 4160-4169
- [40] Yu A, Ye V, Tancik M, *et al.* pixelNeRF: neural radiance fields from one or few images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 4576-4585
- [41] Chen A P, Xu Z X, Zhao F Q, *et al.* MVNeRF: fast generalizable radiance field reconstruction from multi-view stereo[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 14104-14113
- [42] Tancik M, Mildenhall B, Wang T, *et al.* Learned initializations for optimizing coordinate-based neural representations[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 2845-2854
- [43] Kosiorek A R, Strathmann H, Zoran D, *et al.* NeRF-VAE: a geometry aware 3d scene generative model[OL]. [2023-11-25]. <https://arxiv.org/abs/2104.00587>
- [44] Chibane J, Bansal A, Lazova V, *et al.* Stereo radiance fields (SRF): learning view synthesis for sparse views of novel scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 7907-7916
- [45] Jain A, Tancik M, Abbeel P. Putting NeRF on a diet: semantically consistent few-shot view synthesis[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5865-5874
- [46] Wang Q Q, Wang Z C, Genova Kyle, *et al.* IBRNet: learning multi-view image-based rendering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Los Alamitos: IEEE Computer Society Press, 2021: 4688-4697
- [47] Johari M M, Lepoittevin Y, Fleuret F. GeoNeRF: generalizing NeRF with geometry priors[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 18344-18347
- [48] Mu T J, Chen H X, Cai J X, *et al.* Neural 3D reconstruction from sparse views using geometric priors[J]. *Computational Visual Media*, 2023, 9(4): 687-697
- [49] Xu D J, Jiang Y F, Wang P H, *et al.* SinNeRF: training neural radiance fields on complex scenes from a single image[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 736-753
- [50] Deng K L, Liu A, Zhu J Y, *et al.* Depth-supervised NeRF: fewer views and faster training for free[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12872-12881
- [51] Deng C Y, Jiang C M, Qi C R, *et al.* NeRDi: single-view NeRF synthesis with language-guided diffusion as general image priors[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20637-20647
- [52] Mildenhall B, Srinivasan P P, Ortiz-Cayon R, *et al.* Local light field fusion: practical view synthesis with prescriptive sampling guidelines[J]. *ACM Transactions on Graphics*, 2019, 38(4): Article No.29
- [53] Jensen R, Dahl A, Vogiatzis G, *et al.* Large scale multi-view stereopsis evaluation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 406-413
- [54] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM[C] //Proceedings of the 20th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2010: 2366-2369
- [55] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. *Journal of Computer and Communications*, 2019, 7(3): 8-18
- [56] Zhang R, Isola P, Efros A A, *et al.* The unreasonable effectiveness of deep features as a perceptual metric[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 586-595
- [57] Song S R, Yu F, Zeng A, *et al.* Semantic scene completion from a single depth image[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 190-198
- [58] Qiu W C, Zhong F W, Zhang Y, *et al.* UnrealCV: virtual worlds for computer vision[C] //Proceedings of the 25th ACM International Conference on Multimedia. New York: ACM Press, 2017: 1221-1224
- [59] Ma L, Li X Y, Liao J, *et al.* Deblur-NeRF: neural radiance fields from blurry images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12851-12860
- [60] Lee D, Lee M, Shin C, *et al.* DP-NeRF: deblurred neural radiance field with physical scene priors[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 12386-12396
- [61] Wang P, Zhao L Z, Ma R J, *et al.* BAD-NeRF: bundle adjusted deblur neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 4170-4179
- [62] Rudnev V, Elgharib M, Theobalt C, *et al.* EventNeRF: neural radiance fields from a single colour event camera[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 4992-5002
- [63] Zhou K, Li W B, Wang Y, *et al.* NeRFLiX: high-quality neural view synthesis by learning a degradation-driven inter-viewpoint MiXer[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 12363-12374
- [64] Zhu C X, Wan R J, Tang Y K, *et al.* Occlusion-free scene recovery via neural radiance fields[C] //Proceedings of the

- IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20722-20731
- [65] Guo Y C, Kang D, Bao L C, *et al.* NeRFReN: neural radiance fields with reflections[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 18388-18397
- [66] Mildenhall B, Hedman P, Martin-Brualla R, *et al.* NeRF in the dark: high dynamic range view synthesis from noisy raw images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 16169-16178
- [67] Müller T, Evans A, Schied C, *et al.* Instant neural graphics primitives with a multiresolution hash encoding[J]. ACM Transactions on Graphics, 2022, 41(4): Article No.102
- [68] Yang G W, Liu Z N, Li D Y, *et al.* JNeRF: an efficient heterogeneous NeRF model zoo based on Jitter[J]. Computational Visual Media, 2023, 9(2): 401-404
- [69] Fridovich-Keil S, Yu A, Tancik M, *et al.* Plenoxels: radiance fields without neural networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5491-5500
- [70] Yang G W, Zhou W Y, Peng H Y, *et al.* Recursive-NeRF: an efficient and dynamically growing NeRF[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(12): 5124-5136
- [71] Garbin S J, Kowalski M, Johnson M, *et al.* FastNeRF: high-fidelity neural rendering at 200FPS[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 14326-14335
- [72] Wadhvani K, Kojima T. SqueezeNeRF: further factorized FastNeRF for memory-efficient inference[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2022: 2716-2724
- [73] Chen Z Q, Funkhouser T, Hedman P, *et al.* MobileNeRF: exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 16569-16578
- [74] Li S C, Li H, Wang Y, *et al.* SteerNeRF: accelerating NeRF rendering via smooth viewpoint trajectory[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20701-20711
- [75] Park K, Sinha U, Barron J T, *et al.* Nerfies: deformable neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5845-5854
- [76] Li Z Q, Niklaus S, Snavely N, *et al.* Neural scene flow fields for space-time view synthesis of dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 6494-6504
- [77] Pumarola A, Corona E, Pons-Moll G, *et al.* D-NeRF: neural radiance fields for dynamic scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 10313-10322
- [78] Park K, Sinha U, Hedman P, *et al.* HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields[J]. ACM Transactions on Graphics, 2021, 40(6): Article No.238
- [79] Fang J M, Yi T R, Wang X G, *et al.* Fast dynamic radiance fields with time-aware neural voxels[C] //Proceedings of the SIGGRAPH Asia 2022 Conference Papers. New York: ACM Press, 2022: Article No.11
- [80] Liu J W, Cao Y P, Mao W J, *et al.* DeVRf: fast deformable voxel radiance fields for dynamic scenes[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2022: Article No.2664
- [81] Yan Z W, Li C, Lee G H. NeRF-DS: neural radiance fields for dynamic specular objects[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 8285-8295
- [82] Tancik M, Casser V, Yan X C, *et al.* Block-NeRF: scalable large scene neural view synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 8238-8248
- [83] Xiangli Y B, Xu L N, Pan X G, *et al.* CityNeRF: building NeRF at city scale[OL]. [2023-11-25]. <https://arxiv.org/abs/2112.05504v1>
- [84] Turki H, Ramanan D, Satyanarayanan M. Mega-NeRF: scalable construction of large-scale NeRFs for virtual fly-throughs[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12912-12921
- [85] Zhang K, Riegler G, Snavely N, *et al.* NeRF++: analyzing and improving neural radiance fields[OL]. [2023-11-25]. <https://arxiv.org/abs/2010.07492>
- [86] Kulkarni S, Yin P, Scherer S. 360FusionNeRF: panoramic neural radiance fields with joint guidance[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2023: 7202-7209
- [87] Wu X C, Xu J M, Zhu Z H, *et al.* Scalable neural indoor scene rendering[J]. ACM Transactions on Graphics, 2022, 41(4): Article No.98
- [88] Barron J T, Mildenhall B, Tancik M, *et al.* Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5835-5844
- [89] Barron J T, Mildenhall B, Verbin D, *et al.* Mip-NeRF 360: unbounded anti-aliased neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. E/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2022: 5460-5469
- [90] Isaac-Medina B K S, Willcocks C G, Breckon T P. Exact-NeRF: an exploration of a precise volumetric parameterization

- for neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 66-75
- [91] Schwarz K, Liao Y Y, Niemeyer M, *et al.* GRAF: generative radiance fields for 3D-aware image synthesis[C] //Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2020: Article No.1692
- [92] Niemeyer M, Geiger A. GIRAFFE: representing scenes as compositional generative neural feature fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 11448-11459
- [93] Chan E R, Monteiro M, Kellnhofer P, *et al.* pi-GAN: periodic implicit generative adversarial networks for 3D-aware image synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 5795-5805
- [94] Meng Q, Chen A P, Luo H M, *et al.* GNeRF: GAN-based neural radiance field without posed camera[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 6331-6341
- [95] Sun J X, Wang X, Zhang Y, *et al.* FENeRF: face editing in neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 7662-7672
- [96] Gu J T, Liu L J, Wang P, *et al.* StyleNeRF: a style-based 3D-aware generator for high-resolution image synthesis[OL]. [2023-11-25]. <https://arxiv.org/abs/2110.08985>
- [97] Chan E R, Lin C Z, Chan M A, *et al.* Efficient geometry-aware 3D generative adversarial networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 16102-16112
- [98] Kania A, Kasymov A, Zięba M, *et al.* HyperNeRFGAN: hypernetwork approach to 3D NeRF GAN[OL]. [2023-11-25]. <https://arxiv.org/abs/2301.11631>
- [99] Hu B R, Huang J K, Liu Y C, *et al.* NeRF-RPN: a general framework for object detection in NeRFs[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 23528-23538
- [100] Xu X M, Yang Y C, Mo K C, *et al.* JacobiNeRF: NeRF shaping with mutual information gradients[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 16498-16507
- [101] Liu S, Zhang X M, Zhang Z T, *et al.* Editing conditional radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5753-5763
- [102] Yang B B, Zhang Y D, Xu Y H, *et al.* Learning object-compositional neural radiance field for editable scene rendering[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 13759-13768
- [103] Yuan Y J, Sun Y T, Lai Y K, *et al.* NeRF-editing: geometry editing of neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 18332-18343
- [104] Wang C, Chai M L, He M M, *et al.* CLIP-NeRF: text-and-image driven manipulation of neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 3825-3834
- [105] Jain A, Mildenhall B, Barron J T, *et al.* Zero-shot text-guided object generation with dream fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 857-866
- [106] Poole B, Jain A, Barron J T, *et al.* DreamFusion: text-to-3D using 2D diffusion[OL]. [2023-11-25]. <https://arxiv.org/abs/2209.14988>
- [107] Metzger G, Richardson E, Patashnik O, *et al.* Latent-NeRF for shape-guided generation of 3D shapes and textures[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 12663-12673
- [108] Kuang Z F, Luan F J, Bi S, *et al.* PaletteNeRF: palette-based appearance editing of neural radiance fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20691-20700
- [109] Xu H Y, Alldieck T, Sminchisescu C. H-NeRF: neural radiance fields for rendering and temporal reconstruction of humans in motion[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.1146
- [110] Weng C Y, Curless B, Srinivasan P P, *et al.* HumanNeRF: free-viewpoint rendering of moving people from monocular video[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 16189-16199
- [111] Weng C Y, Srinivasan P P, Curless B, *et al.* PersonNeRF: personalized reconstruction from photo collections[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 524-533
- [112] Gao L, Liu F L, Chen S Y, *et al.* SketchFaceNeRF: sketch-based facial generation and editing in neural radiance fields[J]. ACM Transactions on Graphics, 2023, 42(4): Article No.159
- [113] Loper M, Mahmood N, Romero J, *et al.* SMPL: a skinned multi-person linear model[J]. Seminal Graphics Papers: Pushing the Boundaries, 2023, 2: Article No.88
- [114] Chen D Y, Song Y Y, Liang F Z, *et al.* 3D human body reconstruction based on SMPL model[J]. The Visual Computer, 2023, 39(5): 1893-1906
- [115] Chen J C, Yi W T, Ma L Q, *et al.* GM-NeRF: learning generalizable model-based neural radiance fields from multi-view images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20648-20658
- [116] Shao R Z, Zhang H W, Zhang H, *et al.* DoubleField: bridging

- the neural surface and radiance fields for high-fidelity human reconstruction and rendering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 15851-15861
- [117] Kwon Y, Kim D, Ceylan D, *et al.* Neural human performer: learning generalizable radiance fields for human performance rendering[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.1895
- [118] Chen M F, Zhang J F, Xu X Y, *et al.* Geometry-guided progressive NeRF for generalizable and efficient neural human rendering[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 222-239
- [119] Mihajlovic M, Bansal A, Zollhöfer M, *et al.* KeypointNeRF: generalizing image-based volumetric avatars using relative spatial encoding of keypoints[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 179-197
- [120] Hong Y, Peng B, Xiao H Y, *et al.* HeadNeRF: a realtime NeRF-based parametric head model[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 20342-20352
- [121] Guo Z Y, Zhou W G, Wang M, *et al.* HandNeRF: neural radiance fields for animatable interacting hands[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 21078-21087
- [122] Peng B, Hu J, Zhou J T, *et al.* IntrinsicNGP: intrinsic coordinate based hash encoding for human NeRF[J]. IEEE Transactions on Visualization and Computer Graphics, 2024, 30(8): 5679-5692
- [123] Peng S D, Zhang Y Q, Xu Y H, *et al.* Neural body: implicit neural representations with structured latent codes for novel view synthesis of dynamic humans[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 9050-9059
- [124] Cheng W, Xu S, Piao J T, *et al.* Generalizable neural performer: learning robust radiance fields for human novel view synthesis[OL]. [2023-11-25]. <https://arxiv.org/abs/2204.11798>
- [125] Bhatnagar B, Tiwari G, Theobalt C, *et al.* Multi-garment net: learning to dress 3D people from images[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 5419-5429
- [126] Yu T, Zheng Z R, Guo K W, *et al.* Function4D: real-time human volumetric capture from very sparse consumer RGBD sensors[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 5742-5752
- [127] Wang B Y, Wang Q, Cheng J C P, *et al.* Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes[J]. Automation in Construction, 2022, 133: Article No.103997
- [128] Zhao Y F, Deng X Y, Lai H H. Reconstructing BIM from 2D structural drawings for existing buildings[J]. Automation in Construction, 2021, 128: Article No.103750
- [129] Levoy M, Hanrahan P. Light field rendering[C] //Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1996: 31-42
- [130] Sitzmann V, Zollhöfer M, Wetzstein G. Scene representation networks: continuous 3D-structure-aware neural scene representations[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2019: Article No.101
- [131] Lombardi S, Simon T, Saragih J, *et al.* Neural volumes: learning dynamic renderable volumes from images[J]. ACM Transactions on Graphics, 2019, 38(4): Article No.65