

中图法分类号: TP751.1 文献标识码: A 文章编号: 1006-8961(2025)07-2584-13

论文引用格式: Huang Y, Guo Y, Lu Y, Jiang P and Wang F. 2025. 3D reconstruction of neural radiation fields constrained by the Manhattan structure in urban remote sensing images. *Journal of Image and Graphics*, 30(7):2584-2596(黄洋, 郭宇, 路遥, 姜鹏, 王飞. 2025. Manhattan结构约束神经辐射场在城市遥感图像中的三维重建. *中国图象图形学报*, 30(7):2584-2596)[DOI:10.11834/jig.240544]

Manhattan结构约束神经辐射场在城市遥感图像中的三维重建

黄洋^{1,2,3}, 郭宇^{1,2,4*}, 路遥⁵, 姜鹏³, 王飞^{1,2,4}

1. 人机混合增强智能全国重点实验室, 西安 710049; 2. 视觉信息与应用国家工程研究中心, 西安 710049;
3. 西安交通大学软件学院, 西安 710049; 4. 西安交通大学人工智能与机器人研究所, 西安 710049;
5. 北京市遥感信息研究所, 北京 100192

摘要: 目的 利用卫星遥感影像对地球表面进行三维重建是计算机视觉和遥感领域的研究热点。一些大尺度的三维重建方法都期望在城市地区能有更好的效果。为了使复杂结构渲染更加精细、人造建筑的平面更加平整、训练速度更快, 提出一种快速高精度的城市地区遥感神经辐射场(neural radiance field, NeRF)三维重建算法。方法 根据城市地区相对于其他自然场景呈现的结构性规律, 即城市建筑物通常具有较强的几何特征和规整性, 采用Manhattan坐标系的几何结构约束三维重建过程, 从而确保重建结果的几何准确性和细节保真度。此外, 多分辨率哈希编码模块通过引入哈希表和可学习的位置编码减少原始神经辐射场的多层神经网络, 显著减少了辐射场的训练时间。结果 实验使用DFC2019(data fusion contest)数据集测试, 与当前先进算法进行测试对比。实验结果表明, 本文提出的方法在城市建筑物地区图像上具有高效训练速度的同时取得了更好的渲染质量与高程信息, 有效地实现了复杂建筑群的重建。与现有方法相比, 在图像质量与高程重建误差等指标上均有明显提升。结论 该方法在提高城市复杂建筑群三维重建精度和训练效率方面表现更加优秀, 为城市遥感影像三维重建提供了一种有效的解决方案。

关键词: 遥感影像; 三维重建; 神经辐射场(NeRF); 曼哈顿框架(MF); 哈希编码

3D reconstruction of neural radiation fields constrained by the Manhattan structure in urban remote sensing images

Huang Yang^{1,2,3}, Guo Yu^{1,2,4*}, Lu Yao⁵, Jiang Peng³, Wang Fei^{1,2,4}

1. National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an 710049, China;
2. National Engineering Research Center of Visual Information and Applications, Xi'an 710049, China;
3. School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 4. Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China; 5. Beijing Institute of Remote Sensing, Beijing 100192, China

Abstract: Objective Satellite remote sensing images have high application value in urban planning and construction, natural resource management, surface change monitoring, and scientific research. The 3D reconstruction of Earth's surface by

收稿日期: 2024-09-10; 修回日期: 2024-12-03; 预印本日期: 2024-12-10

* 通信作者: 郭宇 yu.guo@xjtu.edu.cn

基金项目: 国家科技重大专项(2019ZX01008101); 陕西省西安市碑林区科技计划项目(GX2230)

Supported by: National Major Science and Technology Projects of China (2019ZX01008101); Applied Technology R&D Reserve Project of Xi'an Beilin District(GX2230)

using satellite remote sensing images has become a research hot spot in the fields of computer vision and remote sensing. As neural rendering implicit 3D reconstruction becomes more widely applied in the field of remote sensing, research on neural radiance field (NeRF) in remote sensing scenarios has expanded, and various neural rendering algorithms have emerged. Concepts and models, such as urban planning, communication infrastructure construction, digital maps, and smart cities, require large-scale 3D reconstruction methods to perform better in urban areas. Urban buildings frequently exhibit structural regularities, typically presenting geometric structures, such as cuboids, cubes, and prisms. This important prior knowledge has not been effectively utilized in previous remote sensing NeRF algorithms. Therefore, this study aims to introduce geometric structure constraints that are specific to urban areas into remote sensing NeRF 3D reconstruction algorithms. In addition, the rendering method of NeRF and excessive parameters lead to extremely slow training and rendering time. The number of sampling points is determined by the training data, which are responsible for the training and optimization of the entire target space. Reducing the number of sampling points may degrade rendering quality. Therefore, to achieve fast scene reconstruction, optimizing the network structure is necessary, along with reducing its layers and dimensions to accelerate training and rendering.

Method For the 3D reconstruction of complex urban buildings, this study proposes an improved strategy for rendering complex structures more finely and ensuring flatter planes of artificial buildings. During preprocessing, the method adds two adjacent rays to each training ray that do not participate in backpropagation. After obtaining the density information of a scene through a multilayer network, the method inversely solves the depth points of the three rays and determines the surface points of the target. Based on the surface points, a set of surface normals is solved. By using the clustering information of these surface normals, the surface information of the scene and normals perpendicular to the surface is restored and clustered under Manhattan conditions. The three major orthogonal clusters within the Manhattan framework are solved based on the largest cluster, and backpropagation is performed in accordance with constraints, such as orthogonality and normalization, ensuring that the reconstruction target meets Manhattan framework conditions. Training speed is slow due to the extensive computations involved in rendering. To address this issue, the method uses multilayer encoding to reduce the complexity of the neural network. For coarse network sampling points, which must learn the entire scene's 3D information, multi-resolution learnable positional encoding is used to map them into high-dimensional space. For fine network sampling points, which must learn detailed texture information, hash encoding is used. Hash encoding optimizes network input by interpolating sampling points through multi-resolution grids, finding the corresponding hash encoding data, and integrating multiple resolution hash encodings before feeding them into a smaller network for training. This improvement significantly reduces the training time of the radiance field, achieving fast and high-precision reconstruction of complex urban buildings. This method is implemented on Quadro RTX 8000 graphics processing unit by using the PyTorch Lightning framework. The initial learning rate is set to 5×10^{-4} and decreased after each epoch, and the decreasing factor is set to 0.9.

Result Experiments were conducted on four scenes in the 2019 Data Fusion Contest (DFC2019) dataset, compared with open-source projects NeRF, Shadow-NeRF, EO-NeRF, and SaTensorRF. Evaluation metrics included peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and digital surface model (DSM). The results indicated that the proposed method significantly reduced model training time and produced superior DSM and new view renderings for complex urban buildings compared with other methods. In particular, training time under the same number of iterations was only one-third of EO-NeRF and significantly less than those of other remote sensing NeRF algorithms. In datasets with a high proportion of artificial buildings, the method achieved the best metrics among several comparative algorithms. In the 068 dataset, it outperformed other remote sensing NeRF algorithms, with PSNR increasing by 0.67 dB and SSIM by 0.038 7 compared with the best-performing Shadow-NeRF. Similarly, in the 214 dataset, PSNR increased by 0.71 dB and SSIM by 0.052 6 compared with EO-NeRF. In the 260 dataset, PSNR increased by 0.27 dB and SSIM by 0.013 5 compared with EO-NeRF. The generated DSM demonstrated good accuracy, with errors within 3.2 m of the ground truth in the DFC2019 dataset and less than 1.4 m in the 068 dataset.

Conclusion This study proposes an efficient rendering and reconstruction method that combines Manhattan geometric constraints and multi-resolution hash encoding. The Manhattan framework generates surface points during rendering and calculates surface normals for supervision, ensuring that the entire rendering space meets Manhattan framework conditions. The multi-resolution hash encoding module introduces hash tables and learnable positional encoding, reducing the layers and dimen-

sions of the original NeRF and significantly reducing training time. The experiments demonstrate that the algorithm significantly improves modeling accuracy, enhances new view image generation quality, and reduces elevation estimation errors in urban scenes, particularly in artificial building scenarios, while ensuring efficient training performance.

Key words: remote sensing images; 3D reconstruction; neural radiance field(NeRF); Manhattan frame(MF); hash encoding

0 引言

随着航天技术的不断发展,高分辨率卫星数量迅速增加,同时遥感卫星上图像采集设备的精度也显著提升,其空间分辨率通常可以达到1 m以内。例如,WorldView-3光学卫星所载设备的空间分辨率可达到0.3 m左右,能够清晰记录地表的空间结构和细节纹理。凭借其精准的数据采集和高效的数据传输能力,这些高分辨率卫星为遥感科学领域提供了大量宝贵的数据资料,并在军事和民用领域发挥着重要作用,其支持着地理信息系统、资源管理和城市规划等多个领域的科学研究(Gwinner等,2016),具有广泛的实际应用价值。

利用遥感影像可为城市规划、灾害监测、通信设施建设和数字城市等提供高精度的三维模型,从而增强对复杂建筑群的空间理解和认知。神经辐射场(neural radiance field, NeRF)(Mildenhall等,2022)技术的发展,为遥感影像三维重建提供了新的可能性。然而,由于模型容量有限,基于NeRF的变体只能在宏观尺度上合理地表示场景,而当导航到更近的地方检查大型户外场景的微观细节时,会出现过度模糊的伪像和不完整的重建。为了应对这些挑战,几种改进方法相继提出。Shadow-NeRF(Derksen和Izzo,2021)和Sat-NeRF(Marí等,2022)已经可以在遥感尺度下完成目标的三维重建,但是仍有一定的缺点,例如,Shadow-NeRF虽然通过辐照度模型处理了多源光源的问题,但在城市地区的优化方面仍显不足,其渲染结果在表面纹理上存在欠缺;在渲染建筑物和街道等非自然物体时,边缘光滑,表面粗糙不平,无法真实反映实际情况,导致数字表面模型(digital surface model, DSM)中物体表面呈现不规则形状,这些问题亟待解决。Trevithick和Yang(2021)提出的GRF(eneral radiance field),虽然能在辐射场中学习一些3D特征,但总体上还是不适用于复杂的城市地区。Sat-NeRF虽然改进了光线抽取模型,在精度上有所提升,但仍存在上述问题。

国内学者在相关领域也取得了诸多进展。例如,李欣等人(2018)提出一种结合深度学习的单幅遥感图像超分辨率重建方法,该方法通过构建基于卷积神经网络的模型,有效地提升了单幅遥感影像的分辨率,为三维重建提供了更加精细的输入数据。而刘晓楠等人(2024)提出一种带深度信息监督的神经辐射场虚拟视点画面合成技术,在深度信息的引导下,显著增强了视点合成的细节表现力,为遥感三维建模和虚拟场景生成提供了新思路。

此外,Gableman和Kak(2024)引入季节性和太阳光照特性,对不同季节和光照条件下的影像进行处理。EO-NeRF(Marí等,2023)增加了几何一致的阴影,改进了阴影模型。Surf_NeRF(Semeraro等,2023)基于Shadow-NeRF对卫星影像的光照条件进行了优化,增强了重建结果的真实感。SpS-NeRF(Zhang和Rupnik,2023)在Sat-NeRF的基础上引入密集深度监督,在数据稀疏的情况下实现了三维重建。rpcPRF(Zhang等,2024)提出一种基于多平面图像的神经辐射场模型,能够适应不同类型的卫星摄像机,提高了模型的适用范围。然而,这些方法在处理大规模城市遥感图像时,仍面临计算复杂度高和训练时间长的问題。通常需要十几个小时以上的训练时间,且模型所占的计算资源庞大,模型训练的成本急剧升高。因此如何进行训练优化和渲染加速也成为一件非常重要的任务。

在此背景下,结合前沿的人工智能(artificial intelligence, AI)的深度学习多模态遥感融合解译技术逐渐引起关注。这些方法通过融合不同类型的遥感数据(如光学、合成孔径雷达(synthetic aperture radar, SAR)和热红外图像),提高了对复杂场景的理解。例如,DeepFusion(Li等,2022)和Multilevel Multimodal Fusion Transformer(Ma等,2024)通过融合不同类型的遥感数据,提高了对复杂场景的理解和分析能力。SpectralGPT(spectral generative pretrained transformer)(Hong等,2024)作为一种光谱遥感大模型将多波段光谱图像建模为3D张量数据,有很强的重建和推理能力。然而,这些技术在处理城市建筑

物的高精度重建方面仍面临挑战,特别是在高层建筑的几何细节与结构规律性的准确刻画上,通常依赖于更细粒度的三维模型。尽管多模态遥感融合技术在某些应用场景中具有优势,但与本文的研究目标相比,它们更多聚焦于数据融合和信息提取,而非针对城市复杂建筑群的高精度重建。

城市地区的建筑往往呈现出一种结构规律性,通常情况下,城市地区重建目标往往都具有几何结构,例如长方体、正方体和棱台等。这种重要先验规律在以往的遥感神经辐射场算法中尚未得到有效利用,本文针对城市复杂建筑群地区高精度高程重建与新视角图像生成问题,提出基于 Manhattan 的几何约束以及多层编码的遥感神经辐射场算法。主要工作如下:1)Manhattan 几何约束框架。通过多层网络获取到场景中的密度信息后,根据密度信息反向求解目标的深度点,并求得目标的表面点,依据表面对目标求解出一组曲面法线,根据此曲面法线的聚类信息还原出场景中的地表信息以及与地表法线垂直的法线信息,并根据此信息对整个渲染过程进行 Manhattan 坐标系的几何约束,使渲染空间中的人造建筑物结构更加平整。2)哈希编码优化。对于粗采样的空间点,算法对其进行多维度的位置编码,以学习空间特征;对于细采样的空间点,算法对其进行多维度的哈希编码,以学习细节纹理特征。通过这些改进,减少了辐射场的训练时间,实现了城市复杂建筑群地区的快速重建目标。

1 基于 Manhattan 几何结构约束的渲染算法

1.1 基于 Manhattan 坐标系的几何约束

神经辐射场在进行渲染的时候是根据已经定义的光线和空间点的密度进行体积渲染的,而在城市地区复杂建筑群的重建是存在一定的先验的。将渲染过程加入 Manhattan 先验,以达到提升渲染质量和数字表面模型精度的结果。在渲染过程中,寻找满足条件的多个正交平面并加以约束 (Straub 等, 2018),使得最终的渲染结果满足 Manhattan 框架 (Manhattan frame, MF) (Coughlan 和 Yuille, 1999)。

设 $\mathcal{E} = \{e_x, e_y, e_z\}$ 表示 MF 中 3 个正交轴,这 3 个正交 $e \in \mathcal{E}$ 轴应该与 MF 场景中的建筑、地表进行对齐,其中 e_z 为垂直于地面的法线。设场景中共计有

m 条法线,则有 $N = \{n_i\}_{i=1}^m$ 表示整个场景中的所有已知法线的集合,算法需要将法线 $n \in N$ 与 MF 框架中正交轴 $e \in \mathcal{E}$ 对齐。在任意一幅给定的视图中,从相机中心发出的光线设为 $r(\varphi)$,得到该光线的终点深度值,具体为

$$\text{depth}(\varphi) = \sum_{i=1}^n T_i \alpha_i t_i \quad (1)$$

$$T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \quad (2)$$

式中, t_i 表示采样点位置分布, T_i 表示沿光线到 t_i 的透过率, σ_i 是光线上第 i 个采样点处评估的密度, $\delta_i = t_{i+1} - t_i$ 是相邻采样点之间的距离, $\alpha_i = 1 - \exp(-\sigma_i \delta_i)$ 为第 i 个采样点处不透明度值。深度值长度的绝对值为光线距重建目标表面点的距离,因此可以根据采样点位置方程计算出目标表面点 x 的三维空间坐标,具体为

$$x = o(\varphi) + d(\varphi) \text{depth}(\varphi) \quad (3)$$

式中, $o(\varphi)$ 代表光线 $r(\varphi)$ 的原点, $d(\varphi)$ 代表光线 $r(\varphi)$ 的方向。

将光线 r_1 出射点的上方和右方所发出的光线 r_2 和 r_3 设为光线 r_1 的邻接光线,并组成一个光线三元组 $T = \{r_1, r_2, r_3\}$ 。对于给定的三元组 T ,根据式 (3) 可得一个表面点三元组 $X = \{x_1, x_2, x_3\}$,依据这 3 个非共线的空间点可以粗略地拟合出一个小面积上的切平面以及对应的切平面法线,具体为

$$v = (x_1 - x_2) \times (x_1 - x_3) \quad (4)$$

式中, v 为切平面的法线,为了确保每条经计算后的法线都经过相机中心,还需要对其进行符号矫正和归一化,具体为

$$n = \frac{\text{sign}(o(\varphi)^T \cdot v) \cdot v}{\|v\|} \quad (5)$$

式中, n 为经表面点三元组 X 计算并符号矫正和归一化后的表面法线, sign 为符号函数, v 为未经过校正的切平面法向量。

以上为一条光线的处理流程,在算法中,会选取多个不同相机原点出发的射线得到 m 组光线三元组 $\{T_i\}_{i=1}^m$,根据式 (3) 得到 m 组表面点三元组 $\{X_i\}_{i=1}^m$,并根据式 (5) 得到其对应的法线集合 $N = \{n_i\}_{i=1}^m$,随后通过法线 $\{n_i\}_{i=1}^m$ 从未知的几何结构中恢复 MF。

基于 MF 的假设,算法设集合 N 即代表着 MF 中

不同平面或曲面的表面法线,集合 N 中的大部分向量应该满足 MF 并可以被分配到 3 个相互正交的平面上,少部分被分配到噪声的集合中,如树木或其他不规则物体。算法会自动剔除这些噪声影响并只考虑 MF 结构的目标。算法采用经典的 K-means 算法来处理聚类问题,目的是将集合 N 分解成 k 个向量集合 $C = \{c_i\}_{i=1}^k$,在聚类过程中,为确保表面法线是单位向量,每次迭代时需要对质心进行 L_2 正则化。

由于算法所对应的场景是现实中的城市复杂建筑群,因此向量集合 C 并不会恰好对应 MF 中的 3 个正交轴 $\mathcal{E} = \{e_x, e_y, e_z\}$ 。因此在聚类过程中,选取符合几何结构场景的 3 个聚类质心 $c_{MF} = \{n_1, n_2, n_3\}$,对应的聚类定义为 $C_{MF} = \{N_1, N_2, N_3\}$,将所有的聚类质心与其对应的表面法线集合定义为 $U = \{(C_i, c_i)\}_{i=1}^k$,其中聚类质心 c_i 是由其所对应的集合 C_i 的均值和归一化所确定的,具体为

$$c_i = \left\| \frac{1}{|C_i|} \sum n \right\|, n \in C_i \quad (6)$$

式中, C_i 表示第 i 个聚类, c_i 表示第 i 个聚类的质心。

在经过聚类求得所有对应的法线及其质心的集合 U 后,选择所有聚类中最大的类 $f = \arg\max_i (C_i)$ 的质心向量 $n_1 = c_f$,以表示 MF 坐标系的一个轴。然后通过正交化约束式(7)确定 MF 坐标系中剩下两个正交轴 $n_2 = c_s, n_3 = c_t$,具体为

$$s, t = \arg \min_{ij} (|c_i^T n_1| + |c_j^T n_1| + |c_i^T c_j|) \quad (7)$$

在获得满足 MF 且需要被训练的 3 个聚类 $\{C_1, C_2, C_3\}$ 以及对应的质心 $\{c_1, c_2, c_3\}$ 后,对这 3 个轴进行寻找最满足 MF 框架的算法操作。虽然通过以上的聚类 and 质心就可以从位置环境中恢复 MF,但是算法还是需要考虑 $\{C_1, C_2, C_3\}$ 的排列问题,在理想状态下, $\{c_1, c_2, c_3\}$ 应该是完全两两正交的,但是考虑到算法所应用场景的非理想化,只能尽量地使得 $\{c_1, c_2, c_3\}$ 能够逼近 MF 坐标系。因此算法需要寻找最可满足场景特征的主轴,在城市场景中主轴则代表垂直于地平面的法向量。首先根据基于海拔采样所得到采样点,计算所有的 far 位置的采样点 $\{x_i^{far}\}$,在基于海拔采样的假设中,所有的 far 位置的采样点应该属于一个平面,且此平面为现实意义中的和海拔平行的平面,根据采样点集合 $\{x_i^{far}\}$,可

以根据点的集合计算出海拔平面的法向量 e_z ,具体为

$$e_z = normal(\{x_i^{far}\}) \quad (8)$$

式中, $normal$ 代表对一组属于同一个平面的点求解其所在平面的法线,从而得到 Manhattan 坐标系中的主轴。

1.2 Manhattan 几何结构约束的损失函数

MF 框架的作用是在神经渲染的过程中为其隐式地增加先验条件,使得最终的渲染效果满足基本的 MF 框架。在没有真值做约束的情况下,算法利用 MF 的先验约束(Joo 等, 2016)控制网络的反向传播,对于给定的正交簇 $C_{MF} = \{N_1, N_2, N_3\}$ 和正交质心 $c_{MF} = \{n_1, n_2, n_3\}$ 来说,若要满足 MF,则某个簇 N_i 中的所有表面向量应该满足与其质心 n_i 相同的方向和长度,则 L_{ctr} 的主要目的是使得所有的表面向量向其质心靠拢,在长度和方向上都向其质心逼近,具体损失函数为

$$L_{ctr} = \frac{1}{3} \sum_{i=1}^3 \frac{1}{|N_i|} \sum_{n \in N_i} \|1 - n_i^T n\|_1 + \|n_i - n\|_1 \quad (9)$$

为了增强 3 个正交轴之间的正交性,引入损失函数 L_{ort} ,具体为

$$L_{ort} = \frac{1}{3} (|n_1^T n_2| + |n_1^T n_3| + |n_2^T n_3|) \quad (10)$$

引入 λ_{ctr} 和 λ_{ort} 作为超参数平衡这两个损失函数,再加上神经辐射场的原始的损失函数 L_{nerf} ,得到总损失函数,具体为

$$L = \lambda_{ctr} L_{ctr} + \lambda_{ort} L_{ort} + L_{nerf} \quad (11)$$

有 MF 渲染方法的遥感神经辐射场网络结构如图 1 所示,其输入是抽样光线上的采样点 x ,输出对应的密度值 $\sigma \in \mathbf{R}^+$ 和辐射值 $c \in [0, 1]^3$,随后分别进行深度渲染和颜色渲染,被颜色渲染后的值经过 RGB 的真值监督后输入到 L_{nerf} 中,被深度渲染后的值,根据该点入射射线 r_1 和其下方以及右方发出的邻近光线 r_2 和 r_3 所计算出的深度值 d_1, d_2 和 d_3 计算表面点,根据 3 个不同面的表面点计算法线数据,最终生成该切平面下的法线 n_i 。根据若干不同位置的光线进行求解,最终形成一组表面法线集合 $\{n_i\}_{i=1}^m$,在其中选择出符合 MF 条件的法线聚类 $C_{MF} = \{N_1, N_2, N_3\}$ 以及聚类质心 $c_{MF} = \{n_1, n_2, n_3\}$,最后输入到 MF 框架中,通过式(11)进行相应的优化。

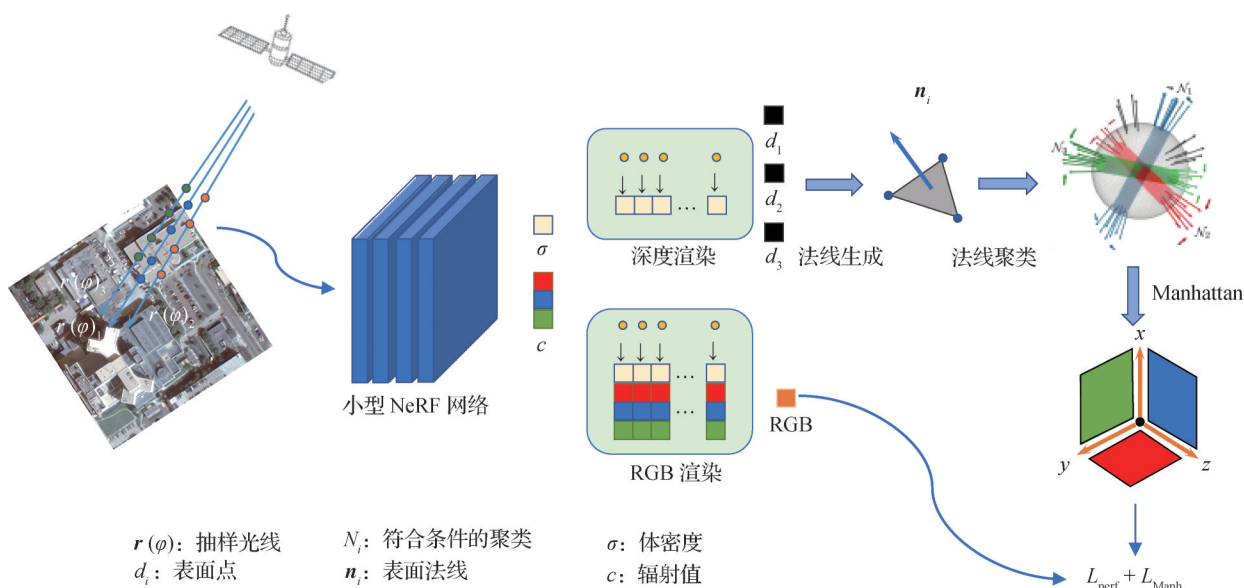


图1 Manhattan几何结构约束的渲染及聚类流程图

Fig. 1 Rendering and clustering flow chart of Manhattan geometric constraint

2 多分辨率哈希编码加速优化

2.1 细网络的哈希编码

细网络需要表达的场景更加细致、表面纹理更加丰富,哈希能够保存更多的高频细节特征,所以将采样点的体素特征保存在哈希表中,并使用此特征的线性插值作为多层神经网络的输入。

细网络下的哈希编码首先定义了场景 M 中的 n 级分辨率,分别为 $L = \{l_1, l_2, \dots, l_n\}$,其中每级的分辨率都是独立的。对于整个3D场景 M 来说,多分辨率中每一级分辨率都将场景 M 分隔为若干个体素集合 $V = \{v_i\}$ 和顶点集合索引 $P^i, i \in L$ 。使用体素的8个顶点的编号可以将 i 级分辨率下的第 p 个体素 v_p 表达成为集合的形式。对于每一级分辨率而言,有与其对应的哈希表 H^i ,每一级分辨率下的哈希表 H^i 的条目数量都为 T ,哈希表 H^i 中每个元素定义长度为 K 的特征向量, i 级分辨率下哈希表 H^i 的特征向量集合为 $F^i = \{f_1, f_2, \dots, f_T\}$,其中 f_j 为 K 维特征向量。

由于空间限制了哈希表的长度,采用哈希表处理顶点和特征向量的映射关系时,可能存在多个顶点映射到同一个特征向量上的情况。没有显式的处理可能存在的哈希冲突,但是算法在多分辨率的条件下依然可以视做是有效的映射关系,因为空间上一点的 x 的最终输入到网络中的特征向量 y 是由多

分辨率下的多个特征向量级联而成,因此可以忽略一定的哈希冲突。即使出现哈希冲突,也可以修改超参数哈希表长度 T 、特征向量长度 F ,使得哈希冲突所造成的影响最小化。算法采用的哈希函数形式,具体为

$$H(x) = \left(\bigoplus_{j=1}^d x_j \pi_j \right) \bmod T \quad (12)$$

式中, $\bigoplus$ 表示按位进行异或运算, π_j 是可选的质数,式(12)本质上是对一维数据进行了线性同余置换,然后对数据进行异或操作,从而减少了维度对哈希效果的影响,为了实现伪随机独立性,需要对 d 个维度中的 $d-1$ 个维度进行相应的置换,且质数 π_j 可以选择 $\pi_j = 1$,以便更好地获得缓存一致性。

对于采样点 x ,在维度 $L = \{l_1, l_2, \dots, l_n\}$ 下都根据点 x 在超立方体中的相对位置对集合中的所有特征向量进行线性插值后,得到 n 个与其相关的特征向量 $F_x = \{f_1, f_2, \dots, f_n\}$,将其级联后组成了该点 x 在所有分辨率下的特征向量 γ_x ,再与辅助输入 $\xi \in \mathbb{R}^E$ 进行级联后组成神经辐射场的网络输入 $y = (\gamma_x, \xi)$ 。

2.2 粗网络的哈希编码

粗网络更需要表达目标的整体特征,其是为细网络的采样做铺垫的,不需要哈希编码的方式来保存其细节纹理上的编码信息,因此算法采用可训练参数的位置编码将低维度的采样点映射到高维空间。

由于在采样过程中,当粗网络的采样点不合理

时,会出现采样混叠的现象,即本该是高频信号的相关特征被采样成了低维信号,因此算法采用锥体光锥代替采样光线(Barron等,2021)。对于原始的采样光线,算法采用多元高斯函数来近似整个圆锥的截面。采样时在圆锥的截面上进行选取点,从而能够考虑到所有的高频信息。对于光线上的采样点 $\mathbf{x}$,经锥体采样后可得采样点,具体为

$$\mathbf{x}^c = f(\mathbf{x}) \quad (13)$$

式中, f 是基于光锥采样的变换函数。在得到对应光锥采样点 $\mathbf{x}^c$ 后,将其在分辨率下可学习的位置编码中进行高维映射,具体为

$$\gamma_p(\mathbf{x}^c) = (\sin(\mathbf{x}^c), \cos(\mathbf{x}^c), \dots, \sin(2^{Q-1}\mathbf{x}^c), \cos(2^{Q-1}\mathbf{x}^c)) \quad (14)$$

式中, $\gamma_p(\mathbf{x}^c)$ 表示多分辨率下的连续映射, Q 表示频率,与当前分辨率的级别相关,被设置在最低分辨率和最高分辨率之间,具体为

$$l_{\min} \leq 2^Q \leq l_{\max} \quad (15)$$

式中, $l_{\max}$ 为最高的分辨率级别, $l_{\min}$ 为最低的分辨率级别。

为了使得多分辨率下编码算法能够在遥感神经辐射场的场景中正确工作,算法需要制定一些采样策略使得其满足本文的框架。由于其采样方式是基于海拔的策略,无法用传统的边界模型对其进行多分辨率采样,因此本文为其设计了一个新的采样策略。对于基于海拔采样的光线可表示为

$$\mathbf{r}(\varphi) = \mathbf{o} + t\mathbf{d} \quad (16)$$

$$\mathbf{d} = \frac{\mathbf{x}_{\text{end}} - \mathbf{x}_{\text{start}}}{\|\mathbf{x}_{\text{end}} - \mathbf{x}_{\text{start}}\|_2} \quad (17)$$

式中, $\mathbf{x}_{\text{start}}$ 表示距离光线原点最近的采样点, $\mathbf{x}_{\text{end}}$ 表示距离光线原点最远的采样点。算法以采样最高点和最低点作为整个采样空间的边界框,以保证多分辨率能够稳定的覆盖在所有的采样空间上。对于所有的采样点 $\{\mathbf{x}_i\}_{i=1}^N$,取其采样点的最高点和最低点的极限值,然后根据此极限值计算新的边界框 $\mathbf{aabb}$,具体为

$$\mathbf{aabb} = ((\min\{\mathbf{x}_i\}_{i=1}^N, \min\{\mathbf{y}_i\}_{i=1}^N, \min\{\mathbf{z}_i\}_{i=1}^N), (\max\{\mathbf{x}_i\}_{i=1}^N, \max\{\mathbf{y}_i\}_{i=1}^N, \max\{\mathbf{z}_i\}_{i=1}^N)) \quad (18)$$

2.3 改进多分辨率哈希编码合成算法

本节的多分辨率哈希编码算法主要由多分辨率模型、哈希编码网络、位置编码网络和神经辐射场多层感知机(multilayer perceptron, MLP)组成。如图2所示,算法执行流程如下:对于给定的细网络的采样

空间点 $\mathbf{x}$,首先对空间进行多分辨率分割,确定点 $\mathbf{x}$ 的空间位置后,求得所有分辨率下的顶点集合 $\mathbf{P}^i$,再由不同分辨率下的哈希函数 $H^i(\mathbf{P}^i)$ 求得所有特征向量并进行级联,得到特征向量 γ_x 后再连接上辅助输入 ξ 得到神经辐射场网络输入 $\mathbf{y}$ 。对于粗网络中的每个采样点,首先在多个分辨率层级下进行编码,通过线性插值获取每层对应的特征向量,并赋予不同权重后进行拼接,形成该采样点的最终输入特征向量 $\mathbf{y}$ 。在神经辐射场框架中,该输入 $\mathbf{y}$ 被选入一个全连接的神经网络 $m(\mathbf{y}, \phi)$,其中 ϕ 表示网络中所有可训练的权重参数。

3 实验结果及分析

3.1 实验设置

实验主要采用的是IEEE GRSS Data Fusion Contest的数据集(DFC2019)(Le等,2019)。使用1张Quadro RTX 8000的GPU(graphics processing unit)对模型进行训练,初始学习率设为 5×10^{-4} ,并在每个epoch后进行递减,递减因子设置为0.9。算法在第3次迭代的时候启动Manhattan模块,为保证变量的一致性,多分辨率哈希编码算法中参数的值设置为固定值 $T = 2^{19}$, $K = 2$, $n = 16$,这些超参数调节是在多次实验下确定的最优解。对于神经辐射场中的多层神经网络而言,由于多分辨率哈希编码算法的框架存在,原先的8层256维的神经网络被替换为4层128维度的神经网络,保证了整个模型的训练加速。本文算法使用PyTorch-Lightning框架,该框架是对PyTorch框架的抽象以及包装,复用性和可读性较强,比较适合部署大型的深度学习算法和项目,在实验中均使用了相同算力的GPU以及实验环境。所对比的方法借助了开源项目NeRF、Shadow-NeRF、EO-NeRF、SaTensorRF(Zhang和Li,2023)。

3.2 评价指标

对于算法所生成的新视角的质量,采用比较常用的峰值信噪比(peak signal-to-noise ratio, PSNR),结构相似性(structural similarity, SSIM)和数字表面模型(digital surface model, DSM)作为评估算法质量的指标。PSNR是一种评价图像质量的客观标准,有着广泛的应用场景,当其介于30~50之间时,表示图像与对比的图像仅有非常小的误差,人

眼难以察觉,当介于 20~30 之间时,表示人眼可以看出小的误差,但两者整体差异不大。因此其指标越高越好。SSIM 从亮度、对比度和结构方面量化图像属性,是一种用于衡量两幅图像之间的相似度的指标,相对 PSNR 来说,其在图像之间相似度的度

量更加符合人眼对图像的质量判断。SSIM 范围为 $[0,1]$,数值越高,表示合成的虚拟视图与原视图的相似度越高。DSM 是指包含了地表覆盖物的地面高程模型,真值为机载雷达所获取的地表真实信息。

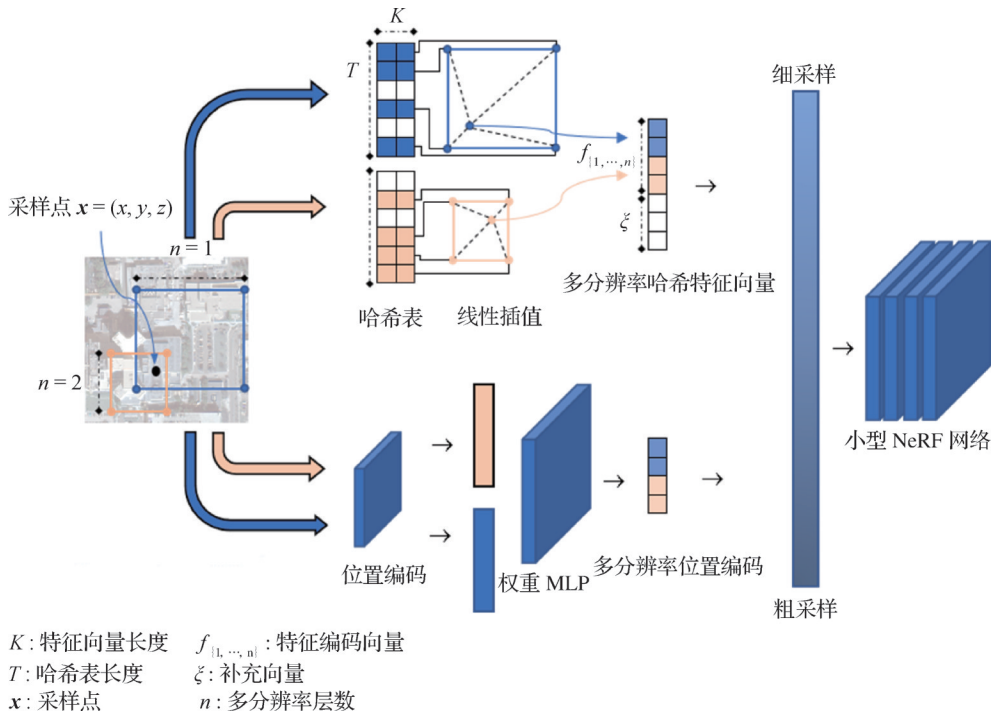


图2 改进多分辨率哈希编码的网络流程图

Fig. 2 Network flow chart of improved multi-resolution hash coding

3.3 实验结果及对比

3.3.1 训练时间分析

本节评估了本文所提出的算法在 DFC2019 (data fusion contest) 数据集里的 4 个场景下的训练运行时间开销情况。相比于其他的遥感神经辐射场算法,本文所提出的算法在训练时间方面表现最好,由于多分辨率哈希编码的作用,模型整体达到了很快的训练时间且能保持很好的质量。具体训练时间如表 1 所示,其所有的模型都是在 10 次 epoch 下完成的训练。对于原 NeRF 算法,其训练时间达到 8 h 以上且渲染效果一般,因为其没有加入遥感场景的相应模型和约束。对于 Shadow-NeERF 算法,由于其加入了辐照度模型,在渲染质量上有了明显的提升,但是其主体网络仍然是原 NeRF 算法的多层网络,所以在渲染时间上大于原 NeRF 算法,需要 11 h 以上。EO-NeRF 改良了光照的一致性以及阴影一致性,并在训练时间上与 Shadow-NeRF 比较接近。

SaTensoRF 算法将采样点的特征用多个维度和平面表示,减少了多层神经网络的数量,其训练时间大大减小。

本文算法由于加入了多分辨率哈希编码模块,将原 NeRF 网络的多层网络规模减小到 3 层 128 的网络,因此在训练速度上加快了很多,训练时间仅为 EO-NeRF 的 1/3,相比于其他遥感神经辐射场算法也大大减少了训练时间,这说明在辐照度模型上增加多分辨率哈希编码模块减少了大量的训练参数,有效地减少了模型的训练时间,最终达到了加速效果。

3.3.2 渲染效果分析

本节评估了本文所提出的算法在 DFC2019 数据集里的 4 个场景下新视角渲染效果情况,评价指标分别为 PSNR 和 SSIM。由表 2 和表 3 可知,相比于其他遥感神经辐射场算法,本文算法在 PSNR 和 SSIM 上相对较高。由于在渲染阶段加入了 Manhattan 框架的约束,在部分数据集的场景中表现较好。

表 1 在 DFC2019 数据集上的训练时间比较结果

Table 1 Comparison results of training time on DFC2019 dataset

方法	004	068	214	260
NeRF	8 h 15 min	8 h 52 min	9 h 43 min	9 h 03 min
Shadow-NeRF	11 h 25 min	12 h 43 min	14 h 05 min	12 h 05 min
EO-NeRF	15 h 43 min	16 h 30 min	18 h 21 min	16 h 07 min
SaTensorRF	5 h 53 min	6 h 12 min	6 h 35 min	5 h 43 min
本文	4 h 15 min	4 h 57 min	5 h 12 min	5 h 28 min

注:加粗字体表示各列最优结果。

本文方法着重处理的是城市地区的复杂建筑群的重建优化,068、214 和 260 场景中的人造建筑的所占空间更大、结构更加复杂,有高楼以及整齐的地面,这都是 Manhattan 框架重点渲染的部分。在 068、214 和 260 场景上,能够获得多个对比算法中最优的指标。在 068 场景上,本文所提出的算法优于其他的遥感神经辐射场算法,比效果最好的 Shadow-NeRF 还要高 0.67 dB 的 PSNR 和 0.038 7 的 SSIM,同样地,在 214 场景上,比 EO-NeRF 高出 0.71 dB 的 PSNR 和 0.052 6 的 SSIM,在 260 场景上,比 EO-NeRF 高出 0.27 dB 的 PSNR 和 0.013 5 的 SSIM。而 004 场景大部分都是小型建筑、林地和草地等,因为这个数据集基本上被大量的植被覆盖,难以找到满足 Manhattan 框架的目标,因此在这个数据集上本文方法的表现相对于其他的几个算法而言效果不佳。但是从总体来看,本文所提出的算法在城市地区复杂人造建筑较多的情况下,可以很好地处理其中的渲染问题。

图 3 是在 DFC2019 数据集上可视化的渲染效果。图中第 1 列指的是不同的场景的代号。图像对应的则是验证集下的新视角图像。可以看出,本文

表 2 在 DFC2019 数据集上的 PSNR 比较结果

Table 2 Comparison results of PSNR on DFC2019 dataset

方法	PSNR/dB ↑			
	004	068	214	260
NeRF	18.07	11.02	18.37	13.80
Shadow-NeRF	25.22	23.59	23.25	21.51
EO-NeRF	25.49	23.17	23.78	22.32
SaTensorRF	24.98	22.06	21.56	21.22
本文	25.29	24.26	24.49	22.59

注:加粗字体表示各列最优结果。“↑”表示值越大越好。

表 3 在 DFC2019 数据集上的 SSIM 比较结果

Table 3 Comparison results of SSIM on DFC2019 dataset

方法	SSIM ↑			
	004	068	214	260
NeRF	0.420 2	0.569 6	0.812 5	0.402 0
Shadow-NeRF	0.810 2	0.865 2	0.906 0	0.790 8
EO-NeRF	0.855 6	0.854 5	0.890 2	0.834 0
SaTensorRF	0.805 6	0.834 4	0.867 1	0.804 9
本文	0.843 5	0.903 9	0.942 8	0.847 5

注:加粗字体表示各列最优结果。“↑”表示值越大越好。

算法可以处理遥感条件下新视图渲染,并在公共数据集上能够取得较好的效果。在 068、214、260 场景上,本文算法能够获得多个对比算法中最优的指标。如图 3 所示,这 3 个数据集的场景处于城区,其内包含了大量的人造建筑物,包含了很多的大楼、平整的地面等,其结构呈现出一种规律性,Manhattan 框架在处理这样的场景时,会将这些符合 Manhattan 特征的目标的法线进行聚类,并将地面、墙壁这些具有正交性的渲染目标变得更加平滑,使得建筑物的表面更加平整、边缘更加锋利。而在 004 场景上,Manhattan 所起到的作用并不明显,因为该场景基本上被大量的植被覆盖,还包含水域、林地等,不满足 Manhattan 框架的前提条件,在渲染过程中模块只能对平整的地面和少部分的建筑物进行建模,其他的区域都会设置为非 Manhattan 结构,并在渲染期间被剔除,只进行普通的渲染,因此只能达到一般的效果,但这也符合预期。

如图 4 所示,以 DFC2019 数据集下的 068 和 214 场景为例,第 1 行数据为 068 场景中的最高楼的右半部分,第 2 行数据为 214 场景的右下角,这两个样例比较满足本文所提出的 Manhattan 假设。从可视化

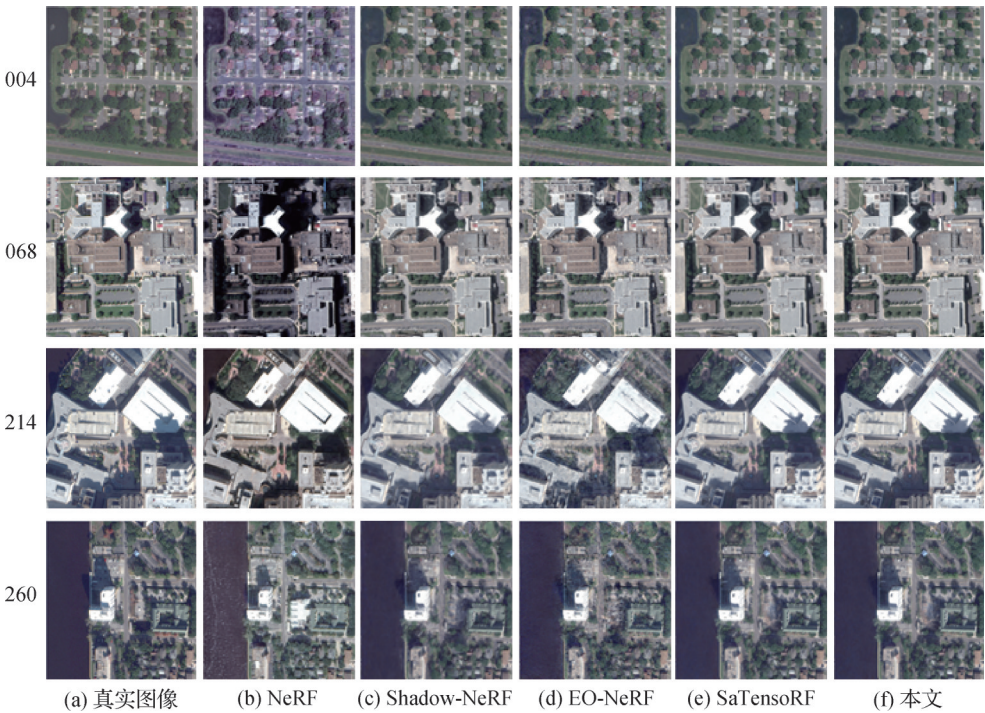


图 3 渲染效果可视化结果

Fig. 3 Rendering effect visualisation results

((a) ground truth; (b) NeRF; (c) Shadow-NeRF; (d) EO-NeRF; (e) SaTensoRF; (f) ours)

的结果可以看出,本文算法在对人造建筑的结构处理方法是最佳的,在 068 场景下,本文算法在处理建筑边缘时效果更好,相对其他算法,能够更有效地渲染边缘和建筑的角,使目标更加满足 Manhattan 条件,同时能够使得建筑平面更加平整。而 NeRF 无法处理光照问题,所以渲染出现了问题,Shadow-NeRF 整

体看来比较模糊,SaTensoRF 的边缘有着更多的锯齿状结构,这是渲染时表面点没有归一到一个平面上的结果。EO-NeRF 在渲染时也出现了墙变成弧形和出现锯齿的问题。而本文算法能有效地处理以上问题,渲染出的结构整体质量较高,整体较为平整。

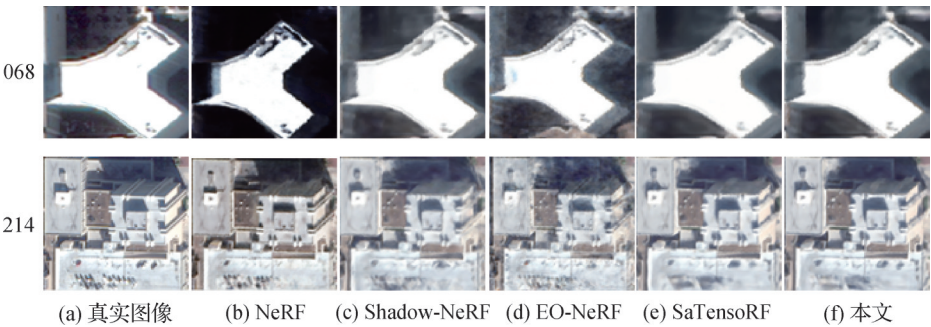


图 4 068 和 214 场景中的细节可视化结果

Fig. 4 Detail visualization results in 068 and 214 scenes

((a) ground truth; (b) NeRF; (c) Shadow-NeRF; (d) EO-NeRF; (e) SaTensoRF; (f) ours)

3.3.3 数字表面模型分析

表 4 是在 DFC2019 数据集中 4 个场景下不同方法生成的数字表面模型与真值的误差值。总体来看,在生成的数字表面模型上精度较好,在 DFC2019

数据集上与真值的误差保持在 3.2 m 以内。在 068、214 和 260 场景中,由于这 3 个场景中符合 Manhattan 框架的人造建筑较多,本文方法效果更加显著,其中 068 场景上的误差小于 1.4 m。图 5 是 DFC2019 数据

集的数字表面模型的可视化结果,其颜色越蓝的区域代表其海拔越低,颜色越红的区域代表其海拔越高。从数据集的不同场景中可以看出,本文算法在人造建筑处理时具有很大的优势,特别是在DFC2019数据集的068、214和260场景中。从可视化的结果可以看出,由于算法对目标的表面上的曲面进行了法线建模,因此所有表面点被约束到了一个整体的曲面上,不会在空中出现游离点,即不会出现类似于Shadow-NeRF算法中出现的空间中的伪影和悬浮物。其次,从真值图像中可以看出,在DFC2019数据集的260场景中最高建筑物的右上方是有两条相互交叉的路面的,多个对比算法都不能很好地渲染出这个目标,而本文算法所生成的DSM可以很明显地看出这两条马路的痕迹,因为这两条马路在模型中被定义为了Manhattan结构。综上所述,本文算法在目标边缘和表面的处理更加优秀,在不同场景下建立的数字表面模型精度都较高,真实地反映出现实世界中的地表模型。

图6是DFC2019数据集的068和214场景的数字表面模型可视化的细节部分。本文算法在对068

表4 在DFC2019数据集上的数字表面模型质量比较结果
Table 4 Comparison results of digital surface model quality on DFC2019 dataset

方法	DSM/m ↓			
	004	068	214	260
NeRF	5.607	7.627	8.035	11.97
Shadow-NeRF	2.031	1.663	4.418	3.449
EO-NeRF	1.456	1.497	3.923	2.235
SaTensorRF	2.125	1.857	4.265	2.954
本文	1.602	1.372	3.190	1.874

注:加粗字体为每列最优值。“↓”表示值越小越好。

场景的大楼表面进行渲染时更加精细,楼顶的整个平面相对于其他对比算法来看也更加平滑。其次,楼与地面之间的落差渲染得很好,从图6中可以看出,颜色从红色能够直接过渡到蓝色,而其他的对比算法都带有少部分的伪影和悬浮物。在214场景上,在对图中左下角的楼房和右上角的楼顶进行渲染时,很好地去除了悬浮物,总体看来表面建模得更加精细,这说明本文算法的Manhattan框架在这个很符合其框架的位置上起到了很大的作用。

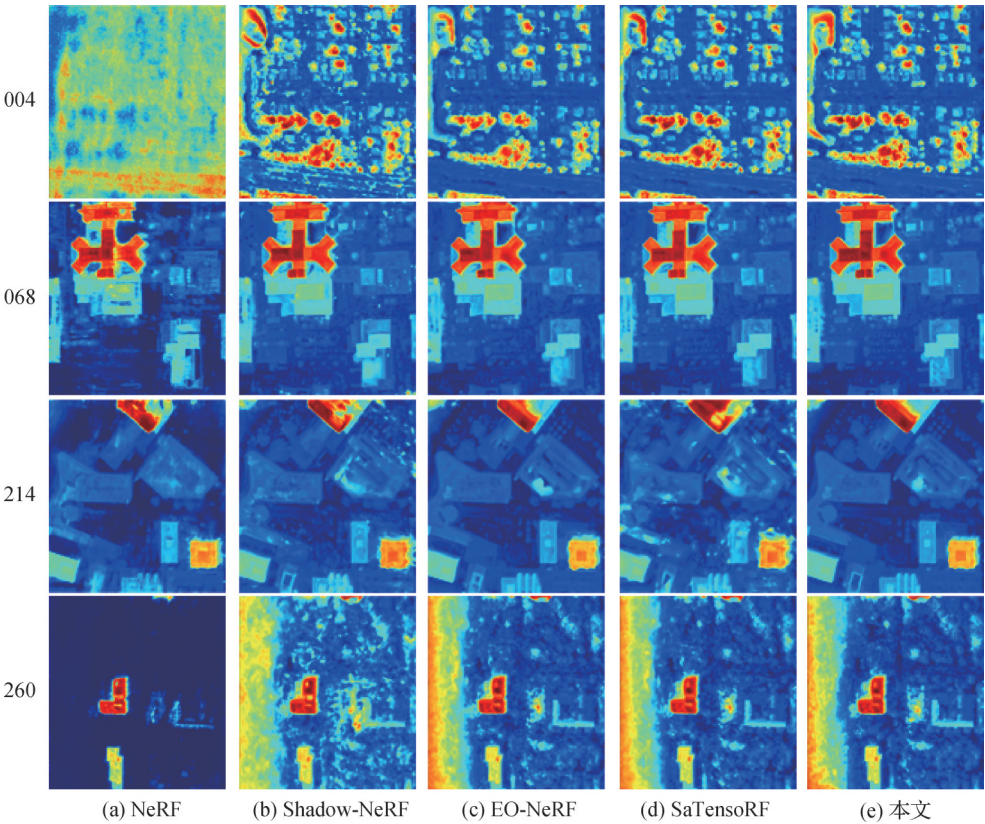


图5 数字表面模型可视化结果

Fig. 5 Digital surface model visualisation results ((a) NeRF; (b) Shadow-NeRF; (c) EO-NeRF; (d) SaTensorRF; (e) ours)

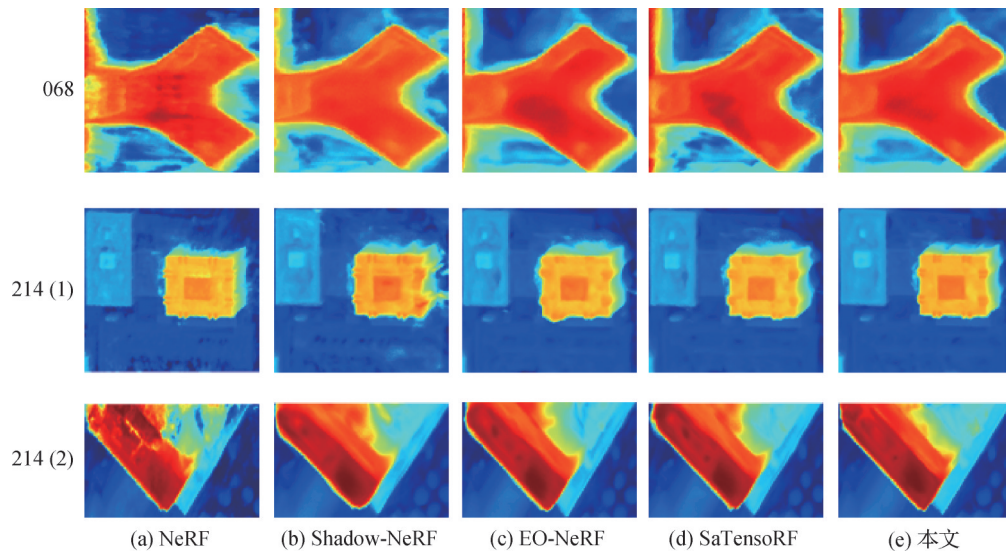


图 6 数字表面模型细节部分比较结果

Fig. 6 Comparison results of detail parts of digital surface model

((a) NeRF; (b) Shadow-NeRF; (c) EO-NeRF; (d) SaTensorRF; (e) ours)

4 结 论

为了解决在城市复杂建筑区域的高精度高程重建与新视角图像生成问题,本文提出基于 Manhattan 几何约束与多分辨率哈希编码结合的高效渲染、重建方法。Manhattan 渲染框架的作用是优化神经辐射场中的建模渲染框架,使之适用于本文所提出问题的场景。Manhattan 距离约束在渲染过程能够反向生成表面点、计算曲面法线并加以监督,使得整个场景建模空间符合 Manhattan 条件的约束,增强人造建筑物边缘的建模精度。多分辨率哈希编码模块通过引入哈希表和可学习的位置编码减少辐射场的多层神经网络规模,进而显著降低训练时间,提升建模效率。本文方法在多个公开遥感数据集与当前先进算法上进行对比,实验结果表明,本文算法在保证训练高效性的同时,在城市场景尤其是人造建筑物场景中,能够显著提升建模精度,改善新视角图像生成质量,降低高程信息估计误差。在未来的工作中,将会在神经辐射场增加多光谱渲染的模块和约束,使算法能够输出新视角下的多光谱数据;其次通过多视图关联算法建立训练图像之间的关联性,利用关联性来隐式监督神经辐射场的深度约束。

参考文献(References)

- 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 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5855-5864 [DOI: 10.1109/ICCV48922.2021.00580]
- Coughlan J M and Yuille A L. 1999. Manhattan world: compass direction from a single image by Bayesian inference//Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE: 941-947 [DOI: 10.1109/ICCV.1999.790349]
- Derksen D and Izzo D. 2021. Shadow neural radiance fields for multi-view satellite photogrammetry//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA: IEEE: 1152-1161 [DOI: 10.1109/CVPRW53098.2021.00126]
- Gableman M and Kak A. 2024. Incorporating season and solar specificity into renderings made by a NeRF architecture using satellite images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(6): 4348-4365 [DOI: 10.1109/TPAMI.2024.3355069]
- Gwinner K, Jaumann R, Hauber E, Hoffmann H, Heipke C, Oberst J, Neukum G, Ansan V, Bostelmann J, Dumke A, Elgner S, Erkeling G, Fueten F, Hiesinger H, Hoekzema N M, Kersten E, Loizeau D, Matz K D, McGuire P C, Mertens V, Michael G, Pasewaldt A, Pinet P, Preusser F, Reiss D, Roatsch T, Schmidt R, Scholten F, Spiegel M, Stesky R, Tirsch D, van Gasselt S, Walter S, Wählisch M and Willner K. 2016. The high resolution stereo

- camera (HRSC) of Mars express and its approach to science analysis and mapping for Mars and its satellites. *Planetary and Space Science*, 126: 93-138 [DOI: 10.1016/J.PSS.2016.02.014]
- Hong D F, Zhang B, Li X Y, Li Y X, Li C Y, Yao J, Yokoya N, Li H, Ghamisi P, Jia X P, Plaza A, Gamba P, Benediktsson J A and Chanussot J. 2024. SpectralGPT: spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46 (8): 5227-5244 [DOI: 10.1109/TPAMI.2024.3362475]
- Joo K, Oh T H, Kim J and Kweon I S. 2016. Globally optimal manhattan frame estimation in real-time//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 1763-1771 [DOI: 10.1109/CVPR.2016.195]
- Le Saux B, Yokoya N, Hansch R, Brown M and Hager G. 2019. 2019 Data fusion contest [technical committees]. *IEEE Geoscience and Remote Sensing Magazine*, 7(1): 103-105 [DOI: 10.1109/MGRS.2019.2893783]
- Li X, Wei H W and Zhang H Q. 2018. Super-resolution reconstruction of single remote sensing image combined with deep learning. *Journal of Image and Graphics*, 23(2): 209-218 (李欣, 韦宏卫, 张洪群. 2018. 结合深度学习的单幅遥感图像超分辨率重建. *中国图象图形学报*, 23(2): 209-218) [DOI: 10.11834/jig.170194]
- Li Y W, Yu A W, Meng T J, Caine B, Ngiam J, Peng D Y, Shen J Y, Lu Y F, Zhou D, Le Q V, Yuille A and Tan M X. 2022. DeepFusion: lidar-camera deep fusion for multi-modal 3D object detection//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 17182-17191 [DOI: 10.1109/CVPR52688.2022.01667]
- Liu X N, Chen C Y, Hu X J and Yu H Y. 2024. Virtual viewpoint image synthesis using neural radiance fields with depth information supervision. *Journal of Image and Graphics*, 29(7): 2035-2045 (刘晓楠, 陈纯毅, 胡小娟, 于海洋. 2024. 带深度信息监督的神经辐射场虚拟视点画面合成. *中国图象图形学报*, 29(7): 2035-2045) [DOI: 10.11834/jig.221188]
- Ma X P, Zhang X K, Pun M O and Liu M. 2024. A multilevel multimodal fusion transformer for remote sensing semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5403215 [DOI: 10.1109/TGRS.2024.3373033]
- Marf 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 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. New Orleans, USA: IEEE: 1310-1320 [DOI: 10.1109/CVPRW56347.2022.00137]
- Marf R, Facciolo G and Ehret T. 2023. Multi-date earth observation NeRF: the detail is in the shadows//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Vancouver, Canada: IEEE: 2034-2044 [DOI: 10.1109/CVPRW59228.2023.00197]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Semeraro F, Zhang Y, Wu W Y and Carroll P. 2023. NeRF applied to satellite imagery for surface reconstruction [EB/OL]. [2024-09-10]. <https://arxiv.org/pdf/2304.04133.pdf>
- Straub J, Freifeld O, Rosman G, Leonard J J and Fisher J W. 2018. The Manhattan frame model — Manhattan world inference in the space of surface normals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (1): 235-249 [DOI: 10.1109/TPAMI.2017.2662686]
- Trevithick A and learning Yang B. 2021. GRF: a general radiance field for 3D representation and rendering//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 15162-15172 [DOI: 10.1109/ICCV48922.2021.01490]
- Zhang L L and Rupnik E. 2023. SparseSat-NeRF: dense depth supervised neural radiance fields for sparse satellite images [EB/OL]. [2024-09-10]. <https://arxiv.org/pdf/2309.00277.pdf>
- Zhang T T and Li Y X. 2023. Fast satellite tensorial radiance field for multi-date satellite imagery of large size [EB/OL]. [2024-09-10]. <https://arxiv.org/pdf/2309.11767.pdf>
- Zhang T T, Li Y X and Wei X. 2024. rpcPRF: generalizable MPI neural radiance field for satellite camera with single and sparse views. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5618715 [DOI: 10.1109/TGRS.2024.3382537]

作者简介

黄洋,女,硕士研究生,主要研究方向为遥感影像的三维重建。E-mail:15802997731@163.com

郭宇,通信作者,男,副教授,主要研究方向为场景重建与感知、数字人生成与驱动、三维点云处理。

E-mail:yu.guo@xjtu.edu.cn

路遥,男,副研究员,主要研究方向为遥感图像智能处理。

E-mail:yaolu@bjirs.org.cn

姜鹏,男,硕士研究生,主要研究方向为大规模场景神经渲染。E-mail:jiangpeng@stu.xjtu.edu.cn

王飞,男,教授,主要研究方向为视觉智能处理器设计、计算机视觉、模式识别与智能系统。E-mail:wfx@mail.xjtu.edu.cn