E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2024)09-2596-14

论文引用格式:Xiao Q, Chen M L, Zhang Y, Huang X H. 2024. Neural radiance field reconstruction for sparse indoor panoramas. Journal of Image and Graphics, 29(09):2596-2609(肖强, 陈铭林, 张晔, 黄小红. 2024. 室内稀疏全景图的神经辐射场重建. 中国图象图形学报, 29(09):2596-2609)[DOI:10.11834/jig.230643]

室内稀疏全景图的神经辐射场重建

肖强,陈铭林,张晔,黄小红*

中山大学深圳校区电子与通信工程学院,深圳　518107

摘　要:**目的**　神经辐射场(NeRF)可以为数字人、交互游戏等虚拟现实应用提供沉浸式环境。然而,现有神经辐射场算法往往依赖大量位置的全景图进行大规模室内场景重建,在稀疏全景图条件下的重建效果不佳。为此,提出了一种用于室内稀疏全景图的神经辐射场重建算法,以实现低开销、高质量的室内新视角合成。**方法**　针对稀疏输入问题,本文首先设计了深度监督策略,以分配更多的采样点在物体表面,从而获取更精细的几何重建结构。然后,引入了未观测视角下的射线形变损失来增强射线约束,从而有效提升了稀疏输入下的室内场景重建质量。**结果**　本文算法在多个室内全景数据集上与较新的神经辐射场重建算法进行了比较。在两幅Replica数据集全景图输入条件下,本文算法在峰值信噪比(peak signal-to-noise ratio,PSNR)指标上比基准算法提升了6%。即使在单幅PNVS数据集全景图输入条件下,本文算法在PSNR指标上也比基准算法提升了11%。**结论**　实验结果表明,本文提出的神经辐射场重建算法能够从稀疏室内全景图中重建高质量场景,实现高真实感的新视角合成。

关键词:神经辐射场(NeRF)重建;稀疏输入;全景图;新视角合成;虚拟现实;数字人

Neural radiance field reconstruction for sparse indoor panoramas

Xiao Qiang, Chen Minglin, Zhang Ye, Huang Xiaohong*

School of Electronics and Communication Engineering, Shenzhen Campus, Sun Yat-Sen University, Shenzhen　518107, China

Abstract: **Objective** Neural radiance field (NeRF) account for a crucial technology used in the creation of immersive environments in various applications, including digital human simulations, interactive gaming, and virtual-reality property tours. These applications benefit considerably from the highly realistic rendering capabilities of NeRF, which can generate detailed and interactive 3D spaces. However, the reconstruction of NeRF typically necessitates to a dense set of multiview images of indoor scenes, which can be difficult to obtain. Current algorithms that address sparse image inputs often perform poorly in the accurate reconstruction of indoor scenes, which leads to less than optimal results. To overcome these challenges, we introduced a novel NeRF reconstruction algorithm specifically designed for sparse indoor panoramas. This algorithm enhances the reconstruction process via the improved allocation of sampling points and refinement of the geometric structure regardless of limited image data. In this manner, high-quality, realistic virtual environments can be synthesized from sparsely available indoor panoramas, which advances the potential applications of NeRF in various fields. **Method** Initially, our algorithm, which is specifically designed to focused on regions of lower latitude, implements a distortion-aware sampling strategy during the ray sampling phase. This strategic approach ensures the sampling of more rays from the central areas of the panorama, which are typically richer in terms of visual information and less distorted compared with the periph-

收稿日期:2023-09-19;**修回日期**:2024-04-10;**预印本日期**:2024-04-17

*通信作者:黄小红　huangxh75@mail.sysu.edu.cn

基金项目:深圳市科技计划资助(KQTD20190929172704911)

Supported by:Shenzhen Science and Technology Program (KQTD20190929172704911)

eral regions. Concentration on these areas we can attain a marked improvement in rendering quality because the algorithm can better capture the essential features and details of the scene. For the further improvement of the reconstruction process, especially in the case sparse image inputs, a panoramic depth estimation network was employed. This network generates a depth map that provides crucial information on the spatial arrangement of objects within a scene. With the estimated depth map, our algorithm incorporates a depth sampling auxiliary strategy and a depth loss supervision strategy. These strategies work in tandem to guide the learning process of the network. The depth sampling strategy allocated a considerable portion of the sampling points in a Gaussian distribution around the estimated depth. This targeted approach enables the network to further comprehension of nuanced understanding of object surfaces, which is essential for accurate scene reconstruction. During the testing phase, our algorithm adopted a coarse-to-fine sampling strategy that aligns with the principles of NeRF. This methodical approach ensures that the network can progressively refine its understanding of the scene, starting with a broad overview and gradually through zooming in on finer details. To maintain the color and depth accuracy throughout the training process, we integrated a depth loss function during the training phase. This function effectively limits the variance of sampling point distribution, which results in a focused and accurate rendering of the scene. In addition, we tackled the issue involving artifacts and improved geometry through the introduction of distortion loss for unobserved viewpoints. This loss function effectively constrains the distribution of unobserved rays in space, which results in realistic and visually pleasing renderings. Moreover, to address the low rendering speed in neural rendering, we developed a real-time neural rendering algorithm with two distinct stages. The first stage involves partitions the bounding box of the scene into a series of octree grids, with each grid's density determined via its spatial location. This process enables the efficient management of the complexity of the scene, which ensures that the rendering process is optimized for speed and quality. Further screening these grids leads to the identification of octree leaf nodes, which are essential for reducing memory consumption and improving the performance. In the second stage, our algorithm leverages the network to predict the color values of leaf nodes from various viewing directions. Spherical harmonics are employed to accurately fit the colors, which guaranteed that the rendered scene is vibrant and true to life. By caching the network model as an octree structure, we enable real-time rendering, which is crucial for applications that demand a seamless and immersive experience. This approach not only substantially improves the rendering speed but also maintains high-quality results, which are essential for the creation of realistic virtual environments. **Result** We evaluated the effectiveness of our proposed algorithm on three panoramic datasets, including two synthetic datasets (i. e. , Replica and PNVS datasets) and one real dataset (i. e. , WanHuaTong dataset). This diverse selection of datasets allowed for a thorough assessment of the algorithm's performance under various conditions and complexities. Our evaluation outcomes illustrate the effectiveness of our algorithm and its superiority over existing reconstruction methods. Specifically, when tested on a Replica dataset with two panoramic images as input, our algorithm exhibited a considerable leap over the current state-of-the-art dense depth priors for NeRF (DDP-NeRF) algorithm. In addition, it achieved a 6% improvement in peak signal-to-noise ratio (PSNR) and an 8% reduction in root mean square error, which reflect improvements in image quality and accuracy. Moreover, our algorithm demonstrated an impressive rendering speed of 70 frames per second on the WanHuaTong dataset, which underscores its capability to handle real depth data with equal proficiency. The algorithm's adaptability is further highlighted in scenarios with challenging panoramic images, such as top cropping and partial depth occlusion. Despite these obstacles, our method effectively recovered complete depth information, which showcases its robustness and reliability in practical applications. **Conclusion** We proposed a NeRF reconstruction algorithm for sparse indoor panoramas, which enables highly realistic rendering from any viewpoints within the scene. Through the implementation of a panoramic-based ray sampling strategy and depth supervision, the algorithm improved the geometric reconstruction quality by focusing on object surfaces. In addition, it incorporated a deformation loss for unobserved viewpoints, which strengthened ray constraints and elevated the reconstruction quality under sparse input conditions. Experimental validation of different panoramic datasets demonstrated that our algorithm outperforms current techniques in terms of color and geometry metrics. This condition leads to the creation of highly realistic novel views and supports of real-time rendering, with potential applications in indoor navigation, virtual reality house viewing, mixed-reality games, and digital human scene synthesis.

Key words: neural radiance field (NeRF) reconstruction; sparse input; panorama; novel view synthesis; virtual reality; digital human

0 引 言

新视角合成是计算机视觉和计算机图形学领域中的一个重要研究方向，它的目标是通过多视角二维图片重建三维场景，并在任意新视角下渲染图像。近年来，新视角合成技术和三维重建技术取得了快速发展（龙霄潇 等，2021），并广泛应用到诸多虚拟现实场景中，如室内导航、混合现实游戏和数字人的背景渲染等。

传统的三维重建算法需要进行相机参数估计、点云重建、表面重建和纹理映射等多个步骤，并通过网格、点云和体素等数据形式进行场景表征。但这些方法的重建结果容易出现孔洞和不真实感（杨航等，2023）。基于深度学习的三维场景重建方法受到了广泛关注。其中，神经辐射场（neural radiance field，NeRF）重建算法得到了快速发展（Mildenhall等，2020），在新视角合成方面取得了较好的渲染效果。

神经辐射场利用可微体渲染技术从多幅二维图像重建三维场景，实现新视角合成。目前，大部分稀疏神经辐射场技术都应用于视场角有限的透视图上。相对透视图而言，全景图有360°的广阔视野，在场景表示中具有天然优势。其中，Otonari等人（2022）将神经辐射场技术应用于全景图中，并提出了两种针对全景图的采样策略，以提升神经辐射场重建的质量，从而实现室内大规模场景重建。此外，Hsu等人（2021）提出了OmniNeRF，该工作设计了虚拟点位上的点云反投影策略来解决点云遮挡问题。Li等人（2022a）针对全景图提出了一种轴对齐的位置编码方式，并使用体素网格进行重建，提升重建速度。然而，这些方法均难以解决稀疏全景图输入的问题。

目前，基于神经辐射场技术的稀疏图像重建研究取得了一定进展。这些方法包括使用预训练模型提取先验知识（Chen等，2021a）、引入语义约束或者射线熵正则约束（Kim等，2022）。此外，一些工作引入深度进行监督，其中Deng等人（2022）通过监督运动恢复结构（structure from motion，SfM）获得的关键点深度实现稀疏场景重建。Roessle等人（2022）进一步利用深度补全网络丰富深度信息，这两种方法仍然依赖于SfM的准确性和关键点数量，存在深度不确定性问题。Wang等人（2023）在深度估计的粗深度基础上提出了局部深度排序蒸馏方法，用于缓解深度不准确问题。Yuan等人（2023）通过网格重建和图像微调保证几何的一致性。这两种方法不依赖准确的深度信息，但在输入稀疏的场景下仍然存在监督较弱的问题。

综上所述，用于大规模稀疏全景图的神经辐射场重建仍面临一些挑战：在场景图像稀疏时，射线重叠较少，神经辐射场网络容易过拟合于颜色信息，导致对深度信息估计偏差过大，在稀疏几何监督下问题更为明显。在视图过少时，由于几何监督信息不足，如何结合空间先验知识，引入更多信息进一步约束辐射场重建是稀疏图像重建面临的重要问题。

本文提出了一种用于室内稀疏全景图的神经辐射场重建算法。该算法首先设计了基于全景畸变的射线采样策略，将更多射线采样在有效像素点上并加快收敛。其次设计了基于深度的采样和监督策略，并给予一定自由度用于监督颜色，以缓解深度不准确问题。最后引入未观测视角下的射线约束，算法充分利用空间先验知识，用于缓解稀疏输入下监督信息不足的问题。

本文主要贡献点如下：1）提出了深度采样和监督策略，将更多采样点分配在物体表面，从而获得更精细的场景几何结构，并留有一定自由度去优化颜色信息；2）针对稀疏输入问题，设计了未观测视角的形变损失，利用先验知识约束空间射线分布，进而有效提升了场景重建质量；3）本文算法在多个全景数据集和自渲染的Replica数据集上与其他稀疏辐射场重建方法进行比较。实验结果表明，本文算法相对于对比算法能取得更好的新视角合成效果。

1 相关工作

1.1 神经辐射场

Mildenhall等人（2020）提出神经辐射场，通过使用多层感知机（multilayer perceptron，MLP）来建模三维空间中的隐函数，以获取空间中任意点的密度和与视角相关的颜色信息，然后通过可微的体渲染方程将三维隐函数渲染为二维图像。神经辐射场因其简洁的实现思路和高度逼真的渲染效果而受到广泛关注。

Liu等人(2020)将三维空间分成多个网格,将可训练的特征向量放置在网格顶点上,并使用多层感知机来建模特征向量与空间点密度以及与视角相关的颜色之间的关系。Mip-NeRF(Barron等,2021)提出了使用锥形射线代替单一射线,并通过对位置编码的积分来渲染抗锯齿图像。Mip-NeRF 360(Barron等,2022)提出了一种正则化方式,可以更有效地防止伪影和背景塌陷。Plenoctrees(Yu等,2021a)提出了一种基于Octree的3D表示方法,用于实现神经辐射场实时渲染,极大地提高了渲染速度。Kellnhofer等人(2021)使用符号距离场表示三维空间,并设计了可微分光线追踪器用于渲染二维图像,该方法在三维重建方面优于传统的多视图重建算法。为了更准确地重建几何结构,Wang等人(2021)改进了引起几何偏差的体渲染方程,使其成为无偏估计,从而约束了几何结构的重建,使得重建的几何结构更加合理。Yariv等人(2021)使用符号距离场的拉普拉斯累积分布函数表示体密度函数,以降低重建噪声的影响。

随着后续研究者的不断努力和改进,神经辐射场技术得到了进一步发展。例如,Chen等人(2021a)提出了改进方法以提高神经辐射场的泛化性能,Yu等人(2021a)致力于加快训练速度。这些后续的研究工作使得神经辐射场技术有着广阔的应用前景。

1.2　稀疏输入下的神经辐射场重建

利用稀疏图像重建场景是一个具有挑战性的问题,因为稀疏的图像不能提供足够的三维约束,尤其是在纹理较少的区域。将神经辐射场技术应用于稀疏视角的场景中会导致重建质量严重下降。

针对稀疏输入下神经辐射场重建质量差的问题,现有解决方法可以分为以下3类。第1类是通过预训练的条件模型来提取先验知识。例如,PixelNeRF(Yu等,2021b)从输入图像中提取的局部卷积神经网络(convolutional neural network,CNN)特征,而MVSNeRF(Chen等,2021a)通过3D CNN进行处理。尽管这些方法在小物体上取得了不错的结果,但它们需要许多不同物体的图像数据集进行预训练,在大规模场景中的表现一般。

第2类是增加正则化和语义约束。DietNeRF(Jain等,2021)利用在低分辨率下渲染的未观测视角下的渲染图像编码约束语义一致性,该语义一致性损失可以提供高层次的信息,但在稀疏输入下难以改善场景的几何形状。RegNeRF(Niemeyer等,2022)基于分块渲染来约束场景几何形状和外观,并应用场景空间退火策略。InfoNeRF(Kim等,2022)通过对每条光线的密度熵施加约束,最小化由于视角不足而导致的重建不一致性。

第3类是利用额外的深度进行约束,该方法更适用于场景级别的重建。DS-NeRF(Deng等,2022)通过添加额外的稀疏深度监督来提高重建精度。DDPNeRF(Roessle等,2022)使用深度补全网络将预处理的稀疏深度数据转换为密集深度图用于引导神经辐射场优化。SparseNeRF(Wang等,2023)使用深度排序蒸馏方案进一步缓解深度不准确问题。RGBD-NeRF(Yuan等,2023)通过网格重建保证几何的一致性。

1.3　全景图新视角合成

新视角合成是指通过多视角图片重建场景,并合成新视角图像。已有方法主要依靠运动恢复结构来实现(Pittaluga等,2019)。这种方法可以生成稀疏点云和相机参数,然后创建场景的网格表示来渲染新的视角图像。然而,为了精确重建整个场景的细节需要大量图像,因此在现实应用中实用性一般。

另一种新视角合成的方法使用多平面图像(multiplane images,MPI)实现(Zhou等,2018)。MPI提供了前景和背景信息来解决可见性问题,进而合成新视角图像。随后的研究将MPI应用于全景图像,例如Lin等人(2020)和Attal等人(2020)分别提出了多深度全景和多球体图像表示法,用于从360°全景图像渲染6自由度的视角。PNVS(Xu等,2021)通过在源视图中估计房间布局和深度,并将其转换为目标视角,同时构建了单个房间内多个渲染点位的大规模合成数据集。OmniSyn(Li等,2022a)使用球形成本体积算法来预测全景深度,与全景图像合成网格,并使用融合网络生成中间视图。此外,一些工作结合全景深度估计网络(Chen等,2021b)来生成新视角图像,通过深度反投影将三维场景点云渲染为任意视角下的图像,但存在孔洞问题。

一些工作利用神经辐射场技术实现全景图像的新视角合成。OmniNeRF(Hsu等,2021)从单个360°RGBD图像学习整个场景,提出了点云遮挡方案以

扩充训练射线数量，在不同的虚拟相机位置上增强单个全景图像。360FusionNeRF（Kulkarni 等，2023）在 OmniNeRF 的基础上引入预先训练的视觉编码器提取场景的语义特征，对未观察到的区域进行合理的补全。OmniVoxel（Li 等，2022a）提出了一种使用等距全向图像重建神经辐射场的方法，使用轴对齐位置编码方法对颜色特征进行编码。Otonari 等人（2022）提出了两种非均匀射线采样方案，包括失真感知射线采样和内容感知射线采样，这些射线采样方法更适用于全景图。

2 本文方法

本文提出了一种用于室内稀疏全景神经辐射场重建算法，如图 1 所示。算法的主要步骤如下：首先，算法对稀疏输入室内全景图像进行深度估计；然后，使用颜色损失、深度损失和未观测视角下的形变损失监督全景神经辐射场重建；最后，使用八叉树缓存算法对训练好的网络进行网格筛选与转换，实现实时渲染。

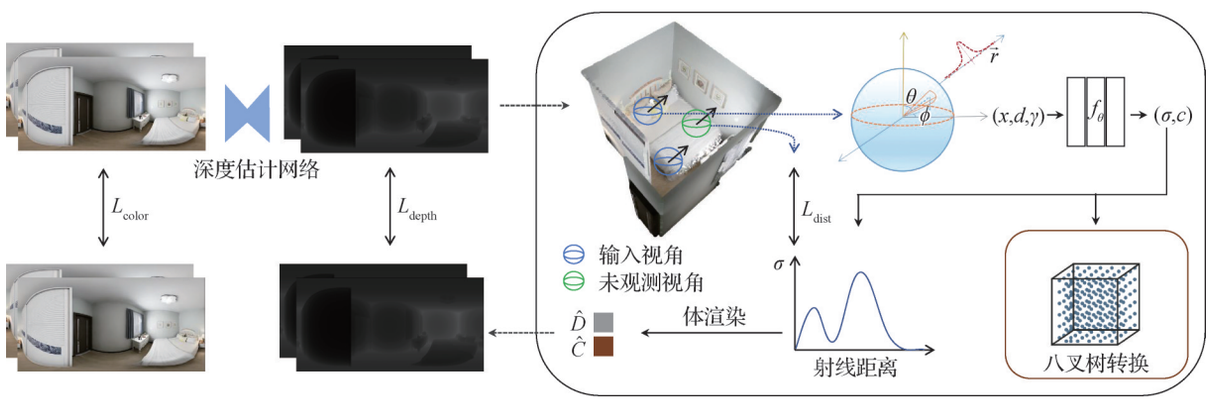


图 1　全景神经辐射场重建算法框架图

Fig. 1　The framework of panoramic neural radiance field reconstruction

2.1 全景神经辐射场

2.1.1 神经辐射场

本文提出了一种适用于全景图的神经辐射场重建算法。对于稀疏输入的全景图，算法首先沿着相机光线采样，并对采样得到的位置 (x,y,z) 和观察方向 (θ,φ) 进行位置编码。编码后，将其与每幅图像优化的隐变量 γ 一起输入神经网络进行预测，得到颜色 $c(r,g,b)$ 和体密度 σ。然后，利用可微体渲染技术得到预测的颜色 $\hat{C}$ 和深度 $\hat{D}$，用于计算颜色和深度损失。在该算法中，主干网络使用 8 层感知器网络，并在视角方向 d 处引入自学习隐变量 γ，以减少不同图像之间由于光照、曝光等差异产生的不一致性，网络模型为

$$(\sigma,c)=MLP(x,d,\gamma) \tag{1}$$

目前，现有的神经辐射场研究大多针对针孔投影模型下的透视图。而全景图采用的是等距投影模型，具有更广阔的视场角。在计算每个像素的射线方程 $r(t)$ 时需要对等距投影模型进行球坐标系转换，转换过程为

$$\begin{cases} x=r\sin\theta\cos\varphi \\ y=r\sin\theta\sin\varphi \\ z=r\cos\theta \end{cases} \tag{2}$$

式中，(x,y,z) 是采样点三维空间坐标，(θ,φ) 是观察方向。全景图上的像素与三维空间坐标的转换如图 2 所示，首先，像素二维坐标可以投影到一个单位球体上。对于全景图上的像素点 $P(i,j)$，其垂直和水平视角可以由 $\theta=\pi i/H$，$\varphi=2\pi j/W$ 计算，其中，H 和 W 分别是全景图的高度和宽度。坐标中心 o 是当前相机中心位置，即光线的起点。光线方向 d 是相机中心到球体表面的单位向量。

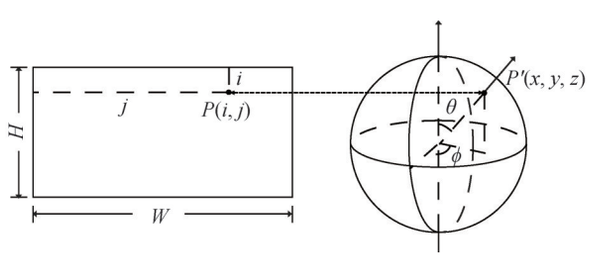


图 2　球坐标系转换

Fig. 2　Spherical coordinate transformation

通过这些转换，全景图中的每个像素位置都能映射到三维空间中的坐标。这样，算法就能计算每个像素对应的射线方程$r(t) = o + t \times d$，进而用于神经辐射场重建。

在射线采样时，算法分别进行粗采样和细采样两个步骤。将得到的采样点进行位置编码，从而捕获高频信息。然后，通过 MLP 网络对每个采样点进行密度σ_i和颜色c_i的预测。网络的输出通过体渲染公式计算射线期望的颜色值$\hat{C}(r)$，具体为

$$\hat{C}(r) = \sum_{i=1}^{N} T_i\left(1 - \exp\left(-\sigma_i\delta_i\right)\right)c_i \tag{3}$$

式中，$T(t)$表示遮挡概率，具体为

$$T_i = \exp\left(-\sum_{j=1}^{i-1}\sigma_j\delta_j\right) \tag{4}$$

式中，$\delta_i = t_{i+1} - t_i$表示两个采样点之间的距离，类似于计算渲染颜色，渲染深度$\hat{D}(r)$可以表示为对射线上所有采样点的距离t_i加权，具体为

$$\hat{D}(r) = \sum_{i=1}^{N} T_i\left(1 - \exp\left(-\sigma_i\delta_i\right)\right)t_i \tag{5}$$

在神经网络的训练过程中，算法需要计算渲染的像素颜色与真实颜色之间的差异，并通过反向传播来优化网络参数，颜色损失函数为

$$L_{\text{color}}(r) = \sum_{r\in R_i}\left\|(C(r) - \hat{C}(r))\right\|^2 \tag{6}$$

式中，$\hat{C}(r)$为渲染颜色，$C(r)$为真实颜色，R为空检采样射线。当移动到一个新的视角时，逐像素地生成射线。通过预测射线上采样点的颜色和体密度信息，结合体渲染公式可以渲染出新视角下的全景图。

2.1.2　基于全景畸变的概率射线采样

在神经辐射场的训练过程中，每个训练批次包含多条射线，这些射线采样概率是相等的。然而，由于全景图的畸变特性，穿过每个像素的射线覆盖范围并不相同。在全景图上使用均匀采样策略会导致渲染质量下降。

为了减少球面到平面投影畸变的偏差，本文算法利用每个像素在空间中的覆盖范围来控制采样概率（Otonari 等，2022）。具体而言，算法会计算每个全景图像的像素在单位球面上所占据的面积S_d，并根据该面积分配相应的采样概率，具体为

$$S_d = \int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2}\cos\theta\mathrm{d}\theta\mathrm{d}\phi = \left(\phi_2 - \phi_1\right)\left(\sin\theta_2 - \sin\theta_1\right) \tag{7}$$

式中，$\theta \in [\theta_1, \theta_2]$和$\phi \in [\phi_1, \phi_2]$分别代表球体上的纬度和经度，用于限定每个像素的位置。具体而言，较高纬度区域具有较小的覆盖范围，而较低纬度区域具有较大的覆盖范围。然后，算法计算训练图像中所有像素的S_d值，即像素在单位球面上所占据的面积，并进行归一化处理。这个结果用于训练过程中为每个像素分配采样概率。

对于在顶部和底部存在裁切或残缺的全景图，算法会将这些区域的像素采样概率设置为零，以避免干扰网络训练。

2.2　深度约束

2.2.1　深度估计网络

虽然神经辐射场技术可以在未观测视角下生成高真实感图像，但这依赖于大量的训练图片。当输入视图数量减少时，渲染图像容易过拟合于训练图像，导致在新视角下产生大量伪影，无法学习到场景的几何信息，并且重建效果会显著下降。

为了解决这个问题，本文算法利用深度先验进行约束。通过全景深度估计网络算法 S2Net（accurate panorama depth estimation on spherical surface）（Li 等，2023）估计深度，该算法在各个全景数据集都能取得较好的深度估计效果。

将估计深度与颜色信息反投影到空间点云中，可以在新视角下进行渲染，但会出现大量孔洞。为了在稀疏输入下更好地学习场景的几何信息，本文设计了深度辅助采样策略和深度损失方法。这些深度先验约束能够帮助网络更好地学习场景的几何信息，从而在较少输入全景图的情况下提高渲染的效果，防止过拟合于训练图像。

2.2.2　深度辅助采样策略

在射线采样完成后，本文算法采用深度辅助采样策略在预测深度附近密集选取采样点，从而让网络更多地学习到物体表面细节。这类似于 NeRF 提出的重要性采样，但是本文方法在更短的时间内获得了接近的效果。

具体而言，在训练阶段，算法将一半的采样点均匀分布在近平面和远平面之间，而另一半的采样点则按照高斯分布$N\left(\hat{D}(r), \hat{S}(r)^2\right)$在深度附近进行采样，其中射线权重$w_i$和分布方差$\hat{S}(r)$的计算式为

$$w_i = T_i\left(1 - \exp\left(-\sigma_i\delta_i\right)\right) \tag{8}$$

$$\hat{S}(r)^2 = \sum_{i=1}^{N} w_i \left(t_i - \hat{D}(r) \right)^2 \tag{9}$$

采用这种策略可以在渲染深度不确定的情况下获得更广泛的采样范围。在测试阶段，采取由粗到细采样策略，即先使用较少的采样点进行粗采样并预测结果，然后根据粗采样的结果，进一步分配密集的采样点在体密度更高的区域内。

2.2.3　深度损失函数

为了防止网络过拟合于训练图像而导致重建结果下降，本文算法引入了深度损失函数来约束场景的几何关系。深度损失采用了高斯负对数似然函数（Roessle等，2022），具体为

$$L_{\text{depth}}(r) = \log\left(\hat{S}(r)^2 \right) + \frac{(\hat{D}(r) - D(r))^2}{\hat{S}(r)^2} \tag{10}$$

式中，$\hat{S}(r)$表示采样点距离的方差，$\hat{D}(r)$和$D(r)$分别表示渲染深度和估计深度。该损失一方面约束渲染深度$\hat{D}(r)$尽可能接近实际深度$D(r)$；另一方面约束渲染深度的方差$\hat{S}(r)$尽可能小。此外，为了缓解深度估计网络存在的深度估计不准确问题，只有在满足式（11）的情况下才计算深度损失，具体为

$$\frac{\left| \hat{D}(r) - D(r) \right|}{D(r)} > \beta^{n-1} \tag{11}$$

式中，β为预先设定的深度比例阈值，n是图像的数量。式（11）允许预测深度和估计深度在一定误差范围内时不计算深度损失。

通过设定深度比例阈值，算法可以根据渲染深度和估计深度之间的误差范围来决定是否计算深度损失。当误差在可接受的范围内时，算法将主要关注最小化颜色损失，以保持颜色信息的准确性。这样的设计一定程度上缓解了深度估计的不准确问题，使网络能够在学习过程中具有一定的自适应性和灵活性。

2.3　损失函数设计

2.3.1　未观测视角下的形变损失

受限于稀疏图像输入，神经辐射场重建时容易出现伪影问题。1）稀疏输入下约束较少，网络难以学习到未观测视角下的信息；2）网络过拟合于训练图像，容易学习到错误的场景几何结构。深度辅助采样策略和深度损失函数可以在一定程度上缓解第2类问题，但对于第1类问题，由于缺少真实信息监督，网络难以约束未观测视角下的信息。

考虑到空间中存在一定的先验知识，即空间任意射线会打到物体表面或为空，且该信息无需真实信息进行监督训练，因此可以引入未观测视角下的射线形变损失约束（Barron等，2022）。具体而言，算法使用归一化后的射线距离s来计算形变损失，具体为

$$s = \frac{g(t) - g(t_n)}{g(t_f) - g(t_n)}, \quad g(x) = \frac{1}{x} \tag{12}$$

式中，t_n和t_f分别表示采样点离相机中心射线的最近和最远距离。

为了约束空间中未观测的区域，算法会在场景的包围盒内随机生成相机中心和射线。这些未观测射线可以用于射线分布约束，而不需要真实值作为监督。该方法大大增加了训练信息，从而进一步提升了重建质量。形变损失为

$$L_{\text{dist}}(s, w) = \sum_{i,j} w_i w_j \left| \frac{s_i + s_{i+1}}{2} - \frac{s_j + s_{j+1}}{2} \right| + \frac{1}{3} \sum_i w_i^2 \left(s_{i+1} - s_i \right) \tag{13}$$

式中，s_i表示第i个采样点的归一化距离，w_i表示该采样点对应的权值。该形变损失的目标是鼓励射线w_i在打到物体表面时趋向于单峰分布，或者在未打到物体表面时趋向于零，进而减少辐射场重建过程中产生的伪影，更好地重建场景细节。

2.3.2　总损失函数设计

本文使用的损失函数包括3个部分：颜色损失L_{color}、深度损失L_{depth}和形变损失L_{dist}。颜色损失L_{color}使渲染图像更加接近真实图像；深度损失L_{depth}使渲染深度更加接近估计深度；形变损失L_{dist}用于减弱空间中的伪影。总损失函数L为3个损失函数加权求和得到，具体为

$$L = L_{\text{color}} + \lambda_{\text{depth}} L_{\text{depth}} + \lambda_{\text{dist}} L_{\text{dist}} \tag{14}$$

式中，权重系数λ_{depth}和λ_{dist}用于平衡各项损失函数的相对大小。

2.4　八叉树缓存算法

神经辐射场合成新视角图像时需要查询千万次神经网络，这是十分耗时的。为了解决该问题，本文引入了八叉树缓存算法（Yu等，2021a），极大地提升了渲染速度。该算法可以分为以下两个阶段：网格筛选阶段和转换阶段。

1）网格筛选阶段。算法使用八叉树结构对场景进行分割和组织，通过递归将场景划分为8个子区

域。然后，对每个八叉树节点的坐标位置进行编码，并输入到MLP网络，用于预测该位置体密度σ。根据式(15)计算每个体素的的权值 W_i，具体为

$$W_i = 1 - \exp\left(-\sigma_i \delta_i\right) \tag{15}$$

最后，根据设定的阈值过滤掉权值小于阈值的体素，以节省显存消耗。

2)转换阶段。算法利用球谐函数来表示物体在不同观察角度下的颜色信息。首先计算每个八叉树叶子节点在不同视线方向下的颜色值。然后，使用三阶球谐系数进行拟合。在渲染过程中，根据射线经过叶子节点的位置，查询相应的球谐函数值，并计算最终的渲染颜色。

通过以上两个阶段，八叉树缓存算法能将场景从隐式表征转为显式表征，极大提升渲染速度，实现实时渲染。

3　实验结果及分析

3.1　数据集

本文使用以下3个全景图数据集进行实验，这些数据集根据数据来源的不同分为真实场景采集和合成渲染生成的数据集。其中全景图分辨率均为512 × 1 024像素。

1)Replica全景数据集。使用Replica渲染器(Straub等，2019)生成，包括6个房间的高质量合成全景图，每个房间有20～30个点位的全景图，以及对应的位姿和深度图。采样点均匀分布，间隔约1 m。在该数据集上，对比了不同稀疏神经辐射场算法和不同训练图片数量下的结果。

2)PNVS数据集(Xu等，2021)。该数据集为每个源视图渲染了3～6个目标视图，近区域的采样间隔为0.2～0.3 m，远区域为1～2 m，不同点位全景图存在平移而没有旋转。在这个数据集上比较了不同稀疏神经辐射场算法的结果。

3)万花筒数据集。该数据集包含了多个室内场景的实景拍摄全景图、激光雷达扫描的深度信息以及相机平移向量和旋转矩阵。全景图两极部分被裁切，相邻拍摄点之间的间隔约为1 m。在这个数据集上测试了真实深度渲染和实时渲染效果。

3.2　评估指标

本文使用了4个评价指标来衡量网络在新视角合成任务中的性能。包括颜色指标峰值信噪比(peak signal-to-noise ratio，PSNR)、结构相似性(structural similarity，SSIM)(Wang等，2004)、学习感知图像块相似度(learned perceptual image patch similarity，LPIPS)(Zhang等，2018)和深度评价指标均方根误差(root mean squared error，RMSE)。它们用于衡量生成图像与真实图像之间的重建准确度、结构相似性和感知相似度，以及生成深度图与真实深度图之间的准确度。通过评估这些指标，可以更全面地比较不同方法之间的性能差异。

PSNR表示信号最大可能功率和影响它的表示精度的破坏性噪声功率的比值，通过计算均方误差(mean square error，MSE)得到，计算式为

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j) - R(i,j)\right]^2 \tag{16}$$

$$PSNR = 10 \times \lg\left(\frac{MAX_I^2}{MSE}\right) \tag{17}$$

式中，I和R分别表示真实图像和渲染图像的像素颜色。RMSE是深度均方根误差，计算式为

$$RMSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[D(i,j) - K(i,j)\right]^2 \tag{18}$$

式中，D和K分别表示真实图像和渲染图像的像素深度。指标中PSNR和SSIM值越大越好，LPIPS和RMSE值越小越好。

3.3　实验细节

本文使用Pytorch深度学习框架进行模型训练，使用Adam优化器。学习率初始设置为5×10^{-4}，并按指数衰减到5×10^{-5}。模型训练过程中，射线批数量设置为2 048，在未观测视角下的采样射线批数量为256，β值设置为1.25。训练过程在RTX3090显卡上进行，共进行了100 000个epochs。在损失函数中，式(13)中的λ_{depth}设置为0.04，λ_{dist}设置为0.01。八叉树初始网格划分为512^3。

3.4　实验结果分析

3.4.1　对比实验

本文在Replica全景数据集上和PNVS数据集上与现有稀疏图像神经辐射场重建算法进行比较，包括NeRF(Mildenhall等，2020)、InfoNeRF(Kim等，2022)、RegNeRF(regularizing NeRF)(Niemeyer等，2022)、DS-NeRF(depth-supervised NeRF)(Deng等，2022)、DDPNeRF(dense depth priors for NeRF)(Roessle等，2022)、RGBD-NeRF(RGB-depth super-

vised NeRF)(Yuan 等, 2023)和 SparseNeRF(Wang 等,2023)。并给出在3幅及以下训练视图的新视角合成指标。结果如表1和图3所示,本文算法在多项指标上取得了比其他算法更好的结果。

表1展示了不同输入图像情况下的渲染结果。结果显示本文方法在颜色和深度指标上表现最好。在两幅训练图像时,本文方法在PSNR上相比于次优算法DDPNeRF提升了6%。在3幅训练视图时,本文在PSNR和LPIPS两项指标上超越次优算法SparseNeRF,但在SSIM指标上略低于SparseNeRF。

图3展示了两幅训练图像下的渲染图像和渲染深度结果,其中前两行是Replica数据集的结果,后两行是PNVS数据集的结果。可以观察到本文方法在颜色和深度上均取得更好的表现,能够学习到场景的几何信息,并得到更加锐利的边缘。在Replica数据集上,对于墙角的重建,其他方法出现了扭曲,而本文算法可以很好地重建。此外,在椅子的深度图细节上重建效果也更好。在PNVS数据集上,本文算法在桌子细节图和深度图的表现更好,SparseNeRF和其他稀疏NeRF算法不能重建出桌腿,深度图中出现了较多伪影,而本文方法能够很好地重建这些细节,并且减少伪影的出现。

表1　本文算法与其他算法在Replica数据集上的新视角合成实验结果的对比

Table 1　Comparison of experimental results of novel view synthesis on the Replica dataset

算法	1-view			2-view			3-view		
	PNSR/dB	SSIM	LPIPS	PNSR/dB	SSIM	LPIPS	PNSR/dB	SSIM	LPIPS
NeRF(Mildenhall 等,2020)	10.773 6	0.598 6	0.720 1	18.153 3	0.731 5	0.339 2	22.752 4	0.820 4	0.220 6
InfoNeRF(Kim 等,2022)	12.476 9	0.568 8	0.523 2	18.252 1	0.732 7	0.326 3	22.971 2	0.852 6	0.217 7
RegNeRF(Niemeyer 等,2022)	13.900 7	0.614 4	0.505 7	18.623 1	0.748 8	0.317 8	22.322 7	0.835 8	0.225 8
DS-NeRF(Deng 等,2022)	14.452 3	0.643 2	0.452 7	18.983 4	0.768 2	0.293 4	21.456 3	0.812 9	0.231 3
DDPNeRF(Roessle 等,2022)	15.935 3	0.742 4	0.310 4	20.633 4	0.795 6	0.253 6	22.682 9	0.842 4	0.217 1
RGBD-NeRF(Yuan 等,2023)	14.857 9	0.728 9	0.385 6	19.567 3	0.784 2	0.278 3	21.934 5	0.831 4	0.215 6
SparseNeRF(Wang 等,2023)	14.235 6	0.683 4	0.432 3	20.354 6	0.753 4	0.284 1	23.464 2	**0.872 3**	0.203 5
本文	**16.105 3**	**0.760 8**	**0.306 4**	**21.887 1**	**0.835 1**	**0.229 1**	**23.529 4**	0.870 4	**0.201 8**

注:加粗字体表示各列最优结果。

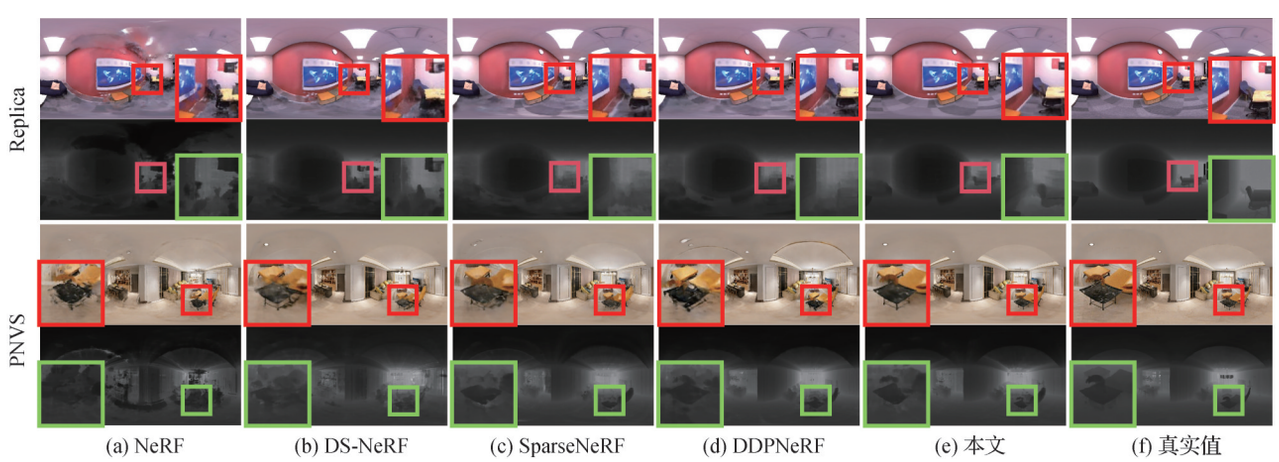


图3　本文算法与其他算法新视角合成实验结果对比

Fig. 3　Comparison of experimental results of novel view synthesis

((a) NeRF; (b) DS-NeRF; (c) SparseNeRF; (d) DDPNeRF; (e) ours; (f) ground truth)

随着输入图像的增加,图像的约束变得更强,此时本文方法与其他辐射场重建算法相比差异变小。然而,在输入图像较少的情况下,本文方法在新视角合成方面相比其他算法具有更好的渲染结果。在处理稀疏输入图像时,基于正则约束的方法(如InfoNeRF,RegNeRF)约束较弱,容易过拟合于训练视图,

渲染效果较差。而基于深度的方法(如DDPNeRF，DS-NeRF，SparseNeRF等)通过深度指导获得更强的约束，结果相对更好。其中DS-NeRF使用SfM获取稀疏点云进行监督，深度监督信息较少。DDPNeRF在此基础上进一步使用深度补全网络补全监督信息，但存在深度不准确的问题。RGBD-NeRF通过重建网格一定程度上缓解了这个问题，但渲染图像时依赖网格重建质量。SparseNeRF提出的深度排序蒸馏方法弥补了深度冲突，在图像较多时有一定优势，但图像较少时监督较弱。本文方法提出了具有鲁棒性的深度监督策略缓解深度不准确的问题，结合未观测视角下的形变损失，算法利用场景先验进一步增加监督信息，因此在细节上表现更好。

3.4.2　消融实验

本文的核心算法设计主要有深度采样策略、深度损失函数、基于全景畸变的概率采样和未观测视角下的形变损失。为了验证各个模块的有效性，本文在Replica全景数据集上进行了5组实验。分别是去除深度辅助采样策略的结果(实验A)、去除gnll损失的结果(实验B)、射线随机筛选的结果(实验C)、去除形变损失的结果(实验D)以及本文完整方法的结果(实验E)。

消融实验的结果如表2和图4所示。从结果中可以观察到，实验A去除深度辅助采样策略，缺少深度采样的引导，在桌子表面上出现了错误的结构。实验B显示引入深度损失后，在渲染图像的PSNR和RMSE上都有显著提升。深度损失能更好地学习到场景几何信息。实验C不使用基于全景畸变的采样概率策略，由于采样概率问题导致图像底部模糊。实验D去除未观测视角下的形变损失时，物体边缘更加模糊，引入未观测视角下的形变损失可以充分利用空间先验知识，提升稀疏输入下的场景重建质量。此外，本文还比较了引入观测视角下的损失，结果显示对于未观测视角约束的结果更好。

表2　消融实验对比结果

Table 2　Comparison results of ablation study

实验	模型	PNSR/dB	SSIM	LPIPS	RMSE
A	w/o sample	21.078 7	0.807 2	0.246 9	0.292 6
B	w/o gnll	19.701 2	0.785 1	0.271 7	0.342 6
C	w/o prob	20.900 1	0.803 5	0.263 8	0.253 5
D	w/o dist	21.212 3	0.819 7	0.237 6	0.235 6
E	本文	**21.887 1**	**0.835 1**	**0.229 1**	**0.223 4**

注：加粗字体表示各列最优结果。w/o X代表缺失X模块。

3.4.3　单幅输入图像实验

本文进一步验证了算法在单幅训练视图下的渲染效果。在合成数据集PNVS中进行了测试，其中训练数据仅包括1幅全景图，首先通过S2Net(Li等，2023)进行深度估计。在距离训练视图附近0.2～0.3 m的范围内进行了新视角的合成，并与现有的单图像全景图新视角合成方法OmniNeRF(Hsu等，2021)、360FusionNeRF(Kulkarni等，2023)进行了比较。实验结果如图5所示。结果显示，本文方法在处理细节方面表现更好。而OmniNeRF和360FusionNeRF方法依赖准确深度信息反投影，导致部分墙面出现空洞。本文方法在估计深度上具有一定的宽容度，并引入了未观测视角射线的形变损失约束，从而更好地处理了细节边缘。与360FusionNeRF相比，本文方法的PSNR从21.15 dB提升至23.52 dB。即使输入单幅图像，RMSE仍可达到0.24 m，依然能得到令人满意的重建结果。

3.4.4　多幅输入图像实验

本文方法在Replica全景数据集上进行了多幅输入图像测试，比较输入图像数量逐步增多时NeRF和本文方法的渲染结果。输入图像数量测试范围为1～12幅，其中图像指标PSNR和深度指标RMSE的结果如图6和图7所示。

从结果可以观察到，在输入图像较为稀疏的情

图4　消融实验结果

Fig. 4　Results of ablation study

(a) OmniNeRF　(b) 360FusionNeRF　(c) 本文

图5　本文算法与其他算法在单幅训练图像下的渲染结果对比

Fig. 5　Comparison of rendering results with a single training image ((a) OmniNeRF; (b) 360FusionNeRF; (c) ours)

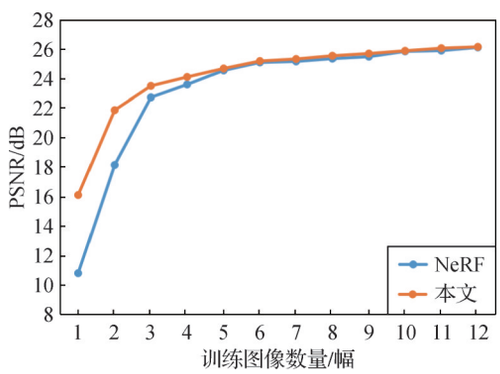


图6　PSNR随训练图像数量的变化

Fig. 6　PNSR with different number of training images

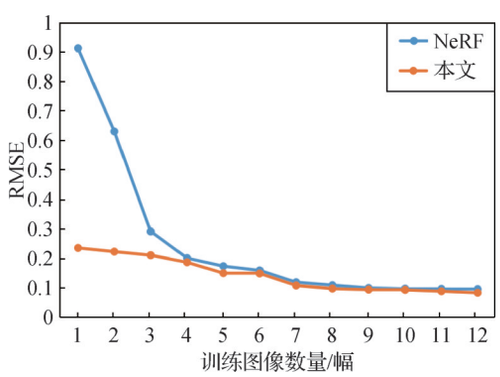


图7　RMSE随训练图像数量的变化

Fig. 7　RMSE with different number of training images

况下，本文方法相对于NeRF在PSNR和RMSE指标上都取得了显著的提升。随着训练图像数量增加，颜色的约束逐渐增强，NeRF的渲染指标也逐渐提升并追平。但是，本文方法引入深度辅助采样策略，使更多的采样点分布在物体表面，并进一步约束未观测视角下的射线分布，这有助于网络更快地收敛，且获得比NeRF更好的细节重建效果。在实际应用场景中，全景图的拍摄点间距平均为1.5 m，在单个房间下拍摄的全景图通常只有3幅左右。在这种情况下，由于输入图像相对较少且分布稀疏，本文方法具有更大的优势。

3.4.5　室外场景实验

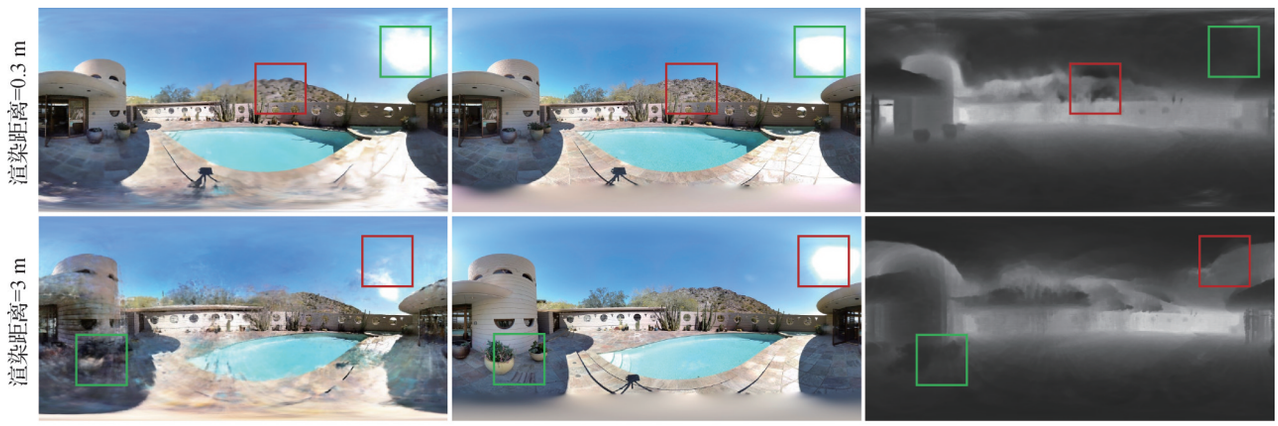
本文也针对室外场景进行了实验。图8(a)(b)展示了在训练视图0.3 m处和3 m处的渲染视图和真实值(Chang等，2017)。可以发现，图中较近处的点位仍能较好地重建出场景，但处于较远点位的太阳和阴影重建效果不太理想。这是由于这些区域的深度估计存在一些问题。如图8(c)所示，天空的深度被预测为靠近相机镜头，因此在室外场景中对远处的重建效果较差。在后续工作中算法拟结合场景分割或其他无界NeRF方法实现更好的室外场景重建效果。

3.4.6　实时渲染实验

本文采用八叉树缓存算法对训练好的网络模型进行网格筛选和转换。算法将隐式的场景表征转换为显式的表征，从而极大提升了渲染的速度。在万花筒数据集的测试实验中，渲染速度从神经辐射场算法的0.2帧/s提升到了70帧/s。

此外，本文方法还支持使用真实深度数据进行监督训练。在万花筒数据集上，本文算法利用了4幅全景图和真实深度进行监督训练，并得到了渲染图像和深度图，如图9(a)(b)所示。需要注意的是，使用深度估计网络获得的结果与真实深度相比，在深度和颜色指标上都有一定程度的降低，这与深度估计网络的性能相关。最后，经过八叉树缓存算法的处理，算法能够在场景内实现自由视角漫游，如图9(c)所示。

从图9(a)可以观察到，尽管训练图像在顶部和



(a) 渲染视图　(b) 真实值　(c) 渲染深度

图 8　室外场景渲染图像

Fig. 8　Results of outdoor rendering images ((a) rendering views; (b) ground truth; (c) rendering depth)

(a) 训练图像　(b) 渲染图像　(c) 实时渲染

图 9　万花筒数据集实验结果

Fig. 9　Results of WanHuaTong dataset ((a) training views; (b) rendering views; (c) real-time rendering)

底部被裁切，并且部分区域的深度信息缺失，但网络训练后仍然能恢复出场景的深度信息细节，并在实时渲染中重建高质量场景。实验结果进一步证明了本文方法在实时渲染方面的有效性。

4　结　论

本文提出了适用于室内稀疏全景图的神经辐射场重建算法。首先，设计了基于全景图的射线采样策略和深度监督策略来为物体表面分配更多的采样点，从而获得了更加精细的几何重建结果。然后，构建了未观测视角下的形变损失来增强射线约束，有效提升了稀疏输入下的场景重建质量。本文算法在多个全景数据集上进行了实验验证。实验结果表明，与对比算法相比，本文算法在室内稀疏全景图输入条件下取得了更好的场景重建效果。在未来工作中，将优化算法在无界场景下的重建，以及加快神经辐射场的训练速度。

参考文献(References)

Attal B, Ling S, Gokaslan A, Richardt C and Tompkin J. 2020. MatryODShka: real-time 6DoF video view synthesis using multi-sphere images//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 441-459 [DOI: 10.1007/978-3-030-58452-8_26]

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: 5835-5844 [DOI: 10.1109/ICCV48922.2021.00580]

Barron J T, Mildenhall B, Verbin D, Srinivasan P P and Hedman P. 2022. Mip-NeRF 360: unbounded anti-aliased neural radiance

fields//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5460-5469 [DOI: 10.1109/CVPR52688.2022.00539]

Chang A, Dai A, Funkhouser T, Halber M, Niebner M, Savva M, Song S R, Zeng A and Zhang Y D. 2017. Matterport3D: learning from RGB-D data in indoor environments//Proceedings of 2017 International Conference on 3D Vision (3DV). Qingdao, China: IEEE: 667-676 [DOI: 10.1109/3dv.2017.00081]

Chen A P, Xu Z X, Zhao F Q, Zhang X S, Xiang F B, Yu J Y and Su H. 2021a. MVSNeRF: fast generalizable radiance field reconstruction from multi-view stereo//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14104-14113 [DOI: 10.1109/ICCV48922.2021.01386]

Chen H X, Li K H, Fu Z H, Liu M Y, Chen Z H and Guo Y L. 2021b. Distortion-aware monocular depth estimation for omnidirectional images. IEEE Signal Processing Letters, 28: 334-338 [DOI: 10.1109/LSP.2021.3050712]

Deng K L, Liu A, Zhu J Y and Ramanan D. 2022. Depth-supervised NeRF: fewer views and faster training for free//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12872-12881 [DOI: 10.1109/CVPR52688.2022.01254]

Hsu C Y, Sun C and Chen H T. 2021. Moving in a 360 world: synthesizing panoramic parallaxes from a single panorama [EB/OL]. [2023-09-04]. https://arxiv.org/pdf/2106.10859. 2021.pdf

Jain A, Tancik M and Abbeel P. 2021. Putting NeRF on a diet: semantically consistent few-shot view synthesis//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5865-5874 [DOI: 10.1109/ICCV48922.2021.00583]

Kellnhofer P, Jebe L C, Jones A, Spicer R, Pulli K and Wetzstein G. 2021. Neural lumigraph rendering//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 4285-4295 [DOI: 10.1109/CVPR46437.2021.00427]

Kim M, Seo S and Han B. 2022. InfoNeRF: ray entropy minimization for few-shot neural volume rendering//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12902-12911 [DOI: 10.1109/CVPR52688.2022.01257]

Kulkarni S, Yin P and Scherer S. 2023. 360FusionNeRF: panoramic neural radiance fields with joint guidance//Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Detroit, USA: IEEE: 7202-7209 [DOI: 10.1109/IROS55552.2023.10341346]

Li D, Zhang Y D, Häne C, Tang D H, Varshney A and Du R F. 2022a. OmniSyn: synthesizing 360 videos with wide-baseline panoramas//Proceedings of 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE: 670-671 [DOI: 10.1109/VRW55335.2022.00186]

Li M, Wang S B, Yuan W H, Shen W C, Sheng Z and Dong Z L. 2023. S^2Net: accurate panorama depth estimation on spherical surface. IEEE Robotics and Automation Letters, 8(2): 1053-1060 [DOI: 10.1109/LRA.2023.3234820]

Lin K E, Xu Z X, Mildenhall B, Srinivasan P P, Hold-Geoffroy Y, Diverdi S, Sun Q, Sunkavalli K and Ramamoorthi R. 2020. Deep multi depth panoramas for view synthesis//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 328-344 [DOI: 10.1007/978-3-030-58601-0_20]

Liu LJ, Gu JT, Lin KZ, Chua TS and Theobalt C. 2020. Neural sparse voxel fields//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1313

Long X X, Cheng X J, Zhu H, Zhang P J, Liu H M, Li J, Zheng L T, Hu Q Y, Liu H, Cao X, Yang R G, Wu Y H, Zhang G F, Liu Y B, Xu K, Guo Y L and Chen B Q. 2021. Recent progress in 3D vision. Journal of Image and Graphics, 26(6): 1389-1428 (龙霄潇, 程新景, 朱昊, 张朋举, 刘浩敏, 李俊, 郑林涛, 胡庆拥, 刘浩, 曹汛, 杨睿刚, 吴毅红, 章国锋, 刘烨斌, 徐凯, 郭裕兰, 陈宝权. 2021. 三维视觉前沿进展. 中国图象图形学报, 26(6): 1389-1428) [DOI: 10.11834/jig.210043]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]

Niemeyer M, Barron J T, Mildenhall B, Sajjadi M S M, Geiger A and Radwan N. 2022. RegNeRF: regularizing neural radiance fields for view synthesis from sparse inputs//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5470-5480 [DOI: 10.1109/CVPR52688.2022.00540]

Otonari T, Ikehata S and Aizawa K. 2022. Non-uniform sampling strategies for NeRF on 360° images [EB/OL]. [2023-09-04]. https://arxiv.org/pdf/2212.03635v1.pdf

Pittaluga F, Koppal S J, Kang S B and Sinha S N. 2019. Revealing scenes by inverting structure from motion reconstructions//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 145-154 [DOI: 10.1109/CVPR.2019.00023]

Roessle B, Barron J T, Mildenhall B, Srinivasan P P and Niebner M. 2022. Dense depth priors for neural radiance fields from sparse input views//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12882-12891 [DOI: 10.1109/CVPR52688.2022.01255]

Straub J, Whelan T, Ma LN, Chen YF, Wijmans E, Green S, Engel JJ, Mur-Artal R, Ren C, Verma S, Clarkson A, Yan MF, Budge

B，Yan YJ，Pan XQ，Yon J，Zou YY，Leon K，Carter N，Briales J，Gillingham T，Mueggler E，Pesqueira L，Savva M，Batra D，Strasdat HM，De Nardi R，Goesele M，Lovegrove S and Newcombe R. 2019. The replica dataset：a digital Replica of indoor spaces［EB/OL］.［2023-09-04］. https://arxiv.org/pdf/1906.05797.pdf

Wang G C，Chen Z X，Loy C C and Liu Z W. 2023. SparseNeRF：distilling depth ranking for few-shot novel view synthesis//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Paris，France：IEEE：9031-9042［DOI：10.1109/ICCV51070.2023.00832］

Wang P，Liu L J，Liu Y，Theobalt C，Komura T and Wang W P. 2021. NeuS：learning neural implicit surfaces by volume rendering for multi-view reconstruction//Proceedings of the 35th International Conference on Neural Information Processing Systems.［s.l.］：［s.n.］：27171-27183

Wang Z，Bovik A C，Sheikh H R and Simoncelli E P. 2004. Image quality assessment：from error visibility to structural similarity. IEEE Transactions on Image Processing，13（4）：600-612［DOI：10.1109/TIP.2003.819861］

Xu J L，Zheng J，Xu Y Y，Tang R and Gao S H. 2021. Layout-guided novel view synthesis from a single indoor panorama//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA：IEEE：16433-16442［DOI：10.1109/CVPR46437.2021.01617］

Yang H，Chen R，An S P，Wei H and Zhang H. 2023. The growth of image-related three dimensional reconstruction techniques in deep learning-driven era：a critical summary. Journal of Image and Graphics，28（8）：2396-2409（杨航，陈瑞，安仕鹏，魏豪，张衡. 2023. 深度学习背景下的图像三维重建技术进展综述. 中国图象图形学报，28（8）：2396-2409）［DOI：10.11834/jig.220376］

Yariv L，Gu JT，Kasten Y and Lipman Y. 2021. Volume rendering of neural implicit surfaces//Proceedings of the 35th International Conference on Neural Information Processing Systems.［s.l.］：［s.n.］：4805-4815

Yu A，Li R L，Tancik M，Li H，Ng R and Kanazawa A. 2021a. PlenOctrees for real-time rendering of neural radiance fields//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：5732-5741［DOI：10.1109/ICCV48922.2021.00570］

Yu A，Ye V，Tancik M and Kanazawa A. 2021b. pixelNeRF：neural radiance fields from one or few images//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA：IEEE：4576-4585［DOI：10.1109/CVPR46437.2021.00455］

Yuan Y J，Lai Y K，Huang Y H，Kobbelt L and Gao L. 2023. Neural radiance fields from sparse RGB-D images for high-quality view synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence，45（7）：8713-8728［DOI：10.1109/TPAMI.2022.3232502］

Zhang R，Isola P，Efros A A，Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA：IEEE：586-595［DOI：10.1109/CVPR.2018.00068］

Zhou T H，Tucker R，Flynn J，Fyffe G and Snavely N. 2018. Stereo magnification：learning view synthesis using multiplane images. ACM Transactions on Graphics，37（4）：#65［DOI：10.1145/3197517.3201323］

作者简介

肖强，男，硕士研究生，主要研究方向为神经辐射场和三维生成。E-mail：xiaoq66@mail2.sysu.edu.cn

黄小红，通信作者，男，副教授，主要研究方向为图像解译与信号处理。E-mail：huangxh75@mail.sysu.edu.cn

陈铭林，男，博士研究生，主要研究方向为新视角合成和三维生成。E-mail：chenmlin8@mail2.sysu.edu.cn

张晔，男，博士后，主要研究方向为序列数据分析和三维生成。E-mail：zhangy2658@mail.sysu.edu.cn