

基于深度优化NeRF空间失真的三维重建方法研究

俞振宇, 曹春萍

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年3月2日; 录用日期: 2025年3月25日; 发布日期: 2025年4月7日

摘要

在现有神经辐射场隐函数表达能力的限制下, 针对其光线追踪终止深度的不确定性, 深度均方差无法收敛, 进而导致重建结果空间失真、远处渲染效果差的问题, 提出了基于深度信息优化空间失真的神经辐射场。首先使用深度信息引导自适应采样方法: 通过图像分割和Canny算法获取物体深度并解析边缘信息, 据此优化采样点数量和权重分配; 然后融合由深度损失 L_d 上的高斯负对数似然(GNLL)项、轮廓损失和颜色损失作为网络的损失函数, 最小化预测误差; 最后提出了一种高斯平滑的周期编码方式, 降低噪声对位置信息的干扰。实验结果表明, 本文融合的深度信息在三维重建中得以充分利用, 保证渲染质量的同时降低空间失真, 以增强空间逼真感。

关键词

神经辐射场, 空间失真, 深度信息, 自适应采样, 高斯平滑

Research on 3D Reconstruction Method Based on Deep Optimization of NeRF Spatial Distortion

Zhenyu Yu, Chunping Cao

School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai

Received: Mar. 2nd, 2025; accepted: Mar. 25th, 2025; published: Apr. 7th, 2025

Abstract

Under the limitations of the implicit function expression ability of existing Neural Radiance Fields

文章引用: 俞振宇, 曹春萍. 基于深度优化 NeRF 空间失真的三维重建方法研究[J]. 建模与仿真, 2025, 14(4): 61-72.
DOI: 10.12677/mos.2025.144265

(NeRF), the Research on 3D reconstruction method based on deep optimization of NeRF spatial distortion is proposed to address the uncertainty of ray tracing termination depth and the inability of depth mean square error to converge, resulting in spatial distortion of reconstruction results and poor rendering effects at a distance. Firstly, use depth information to guide adaptive sampling methods: obtain object depth and parse edge information through image segmentation and Canny algorithm, and optimize the number of sampling points and weight allocation accordingly; then, the Gaussian Negative Logarithmic Likelihood (GNLL) term that combines the mean square error of light color and depth information is used as the loss function of the network to minimize prediction error; finally, a Gaussian smoothing periodic encoding method was proposed to reduce the interference of noise on position information. The experimental results show that the depth information fused in this paper is fully utilized in 3D reconstruction, ensuring rendering quality while reducing spatial distortion to enhance spatial realism.

Keywords

Neural Radiance Fields (NeRF), Spatial Distortion, Depth Information, Adaptive Sampling, Gaussian Smoothing

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

三维重建技术在虚拟现实产业发展中起着关键作用, 尤其在城市规划、公共场所的精细化管理、文化遗产保护与修复等领域具有重要意义[1]。鉴于神经网络的卓越特征表达性能, 研究者逐渐将深度学习思想引入实景重建任务中, 以追求高质量的三维重建。

当前基于深度学习的三维重建方法大致分为显式方法和隐函数方法[2]。显式方法指的是使用深度学习模型对空间的三维数据结构(体素、点云或网格)进行特征提取和表示学习, 然后通过重建网络生成对应的三维结构和纹理信息。例如: Dai 等人[3]提出了一种可以处理大场景点云数据的三维全卷积神经网络, 采用由粗到精的策略, 逐步对场景进行补全重建。但是, 显式方法三维表现形式的分辨率受限, 对真实世界的复杂拓扑结构重建效果不佳, 它需要大规模地标注数据来训练模型或进行后期处理。然而, 获取大规模的高质量标注数据是一项高成本的任务, 这限制了显式方法在实景数字化领域的发展推广。

隐函数方法则将三维形状表示为隐式曲面, 训练神经网络参数, 学习场景的编码器网络、空间点占据状态或颜色和密度信息, 以适应不同的应用需求和场景, 在提高分辨率上显现出了巨大的潜力。例如: Mescheder 等人[4]提出的占用网络(Occupancy Networks)、Mildenhall 等人[5]提出的神经辐射场模型(Neural Radiance Fields, NeRF)等。这类模型在合成数据集中逐步突破了重建图像分辨率提高的难题, 包括对场景细节和复杂纹理的高度复现。但是, 由于实景通常涉及更大的空间范围、更复杂的几何结构和遮挡关系, 其三维模型要求能反映出更精确的物体位置、更逼真的距离感知, 这极大程度上增加了建模的难度。因此, 针对实景的三维重建不仅需要复现场景的几何、纹理信息, 还需要特别关注三维模型空间真实感的重要性。

综上, 现有方法重建结果空间失真主要体现在两个方面: 整体方面, 缺乏对深度信息的充分考量, 训练过程无法降低深度值损失, 从而导致三维重建结果空间感失真, 物体之间的层次和遮挡关系不明确。细节方面, 无深度监督的局部欠采样导致了远处物体渲染清晰度受损、纹理模糊, 细节无法高度还原。

针对以上问题,本文提出了一种深度引导采样与多元复合损失优化空间失真的NeRF三维重建方法,其主要创新点有:

1) 提出深度引导自适应采样算法,利用图像分割和深度图边缘检测精准指导采样,有助于模型学习更多边缘和远处(即深度值大的)区域特征,改善了远处物体渲染模糊的问题。

2) 计算深度分布的高斯负对数似然(GNLL)估计、轮廓损失,联合颜色损失构成多元复合损失函数优化网络。改进了现有损失函数无法收敛深度误差的问题,进一步降低深度值损失,提升重建结果的空间真实感。

3) 提出了一种高斯平滑的周期编码方式,平滑伴随深度信息导入的噪声,减小数据集中异常像素的不良影响进而提高渲染质量和准确性。

2. 相关工作

基于深度学习的显式三维重建方法通常使用点云、体素和网格去表征三维场景,这种可解释性的数据表征更容易理解,且重建结果的几何精度易控制,因而得以广泛地应用[6]。Charles等人[7]在2017年提出了首个将原始点云作为输入的神经网络——PointNet,用于场景的语义分割,但它并没有充分考虑邻近点之间的相互关系,无法充分学习局部特征。为改进这一问题,Zhao等人[8]在2019年提出PointWeb,通过自适应特征调整模块实现信息交换和点的局部特征学习。体素方面,Dai等人[3]在2018年提出了ScanComplete方法,用于进行场景的补全重建操作。该方法的输入是不完整的TSDF(截断有符号距离函数),并通过使用3D卷积网络以一种从粗到精的策略来进行重建。然而,由于体素的分辨率限制,补全的精度受到了一定的限制[9]。基于此,Coudron等人[10]在2020年简化网络,并提出了基于曼哈顿假设非符号距离空间的数据表示,提取了建筑的永久性结构表示。网格方面,Fang等人[11]在2021年提出了结构感知的方式,借助从三维数据中感知到的空间墙体、地板和天花板等主体构件要素引导重建,显著降低了网格模型的复杂度。然而,这些显式方法重建出的三维模型分辨率有限,且高度依赖大规模、高成本数据,计算开销和成本巨大。

基于隐式函数的三维重建方法则在三维重建分辨率的提升方面有所突破,大幅提升了捕捉真实场景中复杂纹理或细节的能力。这些方法将三维形状表示为隐式曲面,例如,Mescheder等人[4]在2019年提出了Occupancy Network,它通过学习场景中每个空间点的占据状态(occupied/empty),从而实现对三维形状的重建和生成高质量的表面表示。Mildenhall等人[5]在2020年提出了NeRF。该模型用全连接神经网络表达连续场景并编码参数,结合物体渲染技术,从而在视图呈现分辨率上得以明显提升。为实现实景重建的渲染质量提升,学术界很快展开了针对真实场景重建的衍生模型研究[12]。例如,mip-NeRF[13]采用锥追踪完成光线追踪,解决光线混叠问题,最终减少图像锯齿,提升渲染画质。DONeRF[14]提出通过直接智能采样NeRF密度函数的表面来提高渲染速度。Ref-NeRF[15]使用反射亮度取代出射亮度,解决合成中的高光和反光问题,更好地模拟真实场景的反射表面。DS-NeRF[16]利用现成的深度监督学习辐射场的损失,提高稀疏重建质量。NerfingMVS[17]利用基于学习的先验来指导NeRF的优化,解决了重建中形状-辐射模糊性的问题。NeRF-W[18]采用自监督学习方法并将非结构化网络照片集转换为3D场景表示,更好地处理实际场景中的复杂多样的细节问题。

总体看来,NeRF成功突破了图像分辨率提高带来的限制,针对实景重建,现有NeRF的衍生模型大多旨在处理场景细节和纹理以提升渲染画质。但是,其实验结果表明重建结果与真实情况的深度均方根误差(Root Mean Square Error, RMSE)收敛情况不佳,这是由于现有方法没有在整个三维重建过程充分利用深度信息,致使模型渲染时难以捕捉深度特征和远处细节特征,进而导致整体空间失真,远处纹理模糊。在VR数字化场景中,尤其针对日常生活相关场景,其要求三维模型既能展现清晰的画质,又要高

度复现原场景的空间逼真感, 于是深度信息显得至关重要。因此, 针对实景重建, 本文提出了一种融合深度信息的 NeRF 模型, 使用深度信息引导自适应采样, 优化深度一致性约束, 一方面使得模型能够适用于实景建模, 提升图像分辨率; 另一方面输出三维模型对应深度图, 提供更完整的三维场景信息。

3. 方法

3.1. 模型概述

本文方法的整体框架如图 1 所示。首先, 以生活场景数据集 ScanNet (ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes) [19]作为输入, 为改进均匀采样造成的远景(即深度值大的)区域和物体边缘欠采样的不足, 本文使用深度信息监督模型进行自适应采样。依据边缘检测和图像分割结果优化采样, 捕捉更多边缘特征和远处物体细节, 为场景渲染提供更准确的细节特征和对应深度信息, 进而增强渲染结果的空间真实感。在网络训练阶段, 为了使模型的深度估计结果尽可能接近真实深度, 提升三维模型的空间真实感, 本文设计融合深度值的损失函数以促进深度均方根误差收敛, 缩小与原场景深度图的差异。此外, 为了平滑场景图像和深度图的数据噪声, 本文使用高斯编码进一步增强位置信息和深度信息的表示能力。经优化后的网络参数可以计算出预测的 RGB 值、密度 σ 和深度值 d , 最终输出渲染图像和对应深度图像。

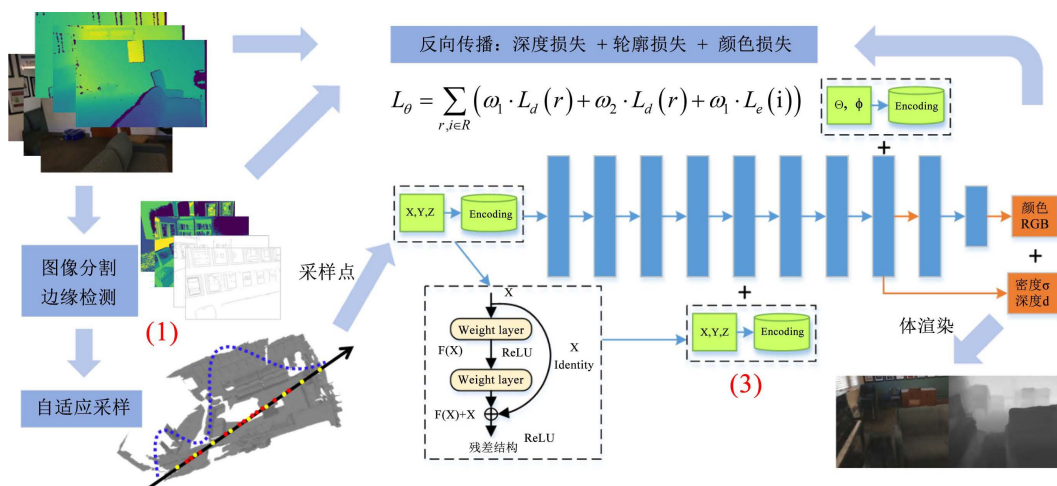


Figure 1. Model framework and operating process description

图 1. 模型框架与运行流程描述

3.2. 深度信息引导自适应采样

对给定实景照片进行三维重建时, 区分场景中不同物体的深度范围和获取边缘信息有助于明确物体定位和遮挡关系, 故应在采样阶段采集到更多相关区域的采样点。然而 NeRF 的均匀采样策略导致物体边缘和远景区域采样不足, 稀疏的采样点并不能帮助模型捕捉更多有效特征, 进而导致渲染出的三维模型空间感失真, 远处物体细节模糊。因而本文提出深度信息引导自适应采样策略, 设计基于连通分量的图像分割算法并统计深度分布, 并结合图像边缘检测结果弥补均匀采样缺陷, 增加边缘和远景区域的采样点密度, 以保留空间真实感、提高远处物体细节的渲染质量。

3.2.1. 基于连通分量的图像分割算法

为了分辨出远景区域的物体, 捕捉更多远处物体细节特征, 需要分割场景中的物体并统计其深度范

围, 实现适应深度变化的采样方案。因此, 本文设计了一种图像连通分量查找算法。相比于其他方法[20], 这种基于连通分量的分割算法在保留空间关系、减少噪声影响具有明显优势。

该方法直接在二维图像中利用明确定义的邻域关系, 其关键组成部分是能够估计任意两束光线的哪些测量点源自同一物体[21]。为了简易高效地找到属于一个对象的组件, 本文首先定义了连通性评价指标 β , 如图 2 所示。

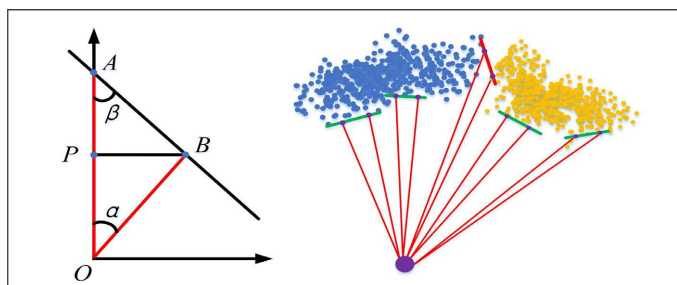


Figure 2. Connectivity definition diagram
图 2. 连通性定义示意图

假设摄像机位于 O 点, 射线 OA 和 OB 代表从摄像机发出的两束光线, 则线段 AB 用于估计物体的表面(如果它们都属于同一物体)。我们首先预定义阈值 θ , 如果 $\beta > \theta$, 则代表这两点属于同一个对象, 如图中绿线上的点被标记为相同的标签; 如果 $\beta < \theta$, 则代表这两点属于不同对象, 如图中红线上的点被标记为不同的标签。

假设线段 OA 和 OB 的长度分别为 d_1 和 d_2 , 由图易得 β 的计算公式为:

$$\beta = \arctan \frac{\|BP\|}{\|PA\|} = \arctan \frac{d_2 \sin \alpha}{d_1 - d_2 \cos \alpha} \quad (1)$$

其中, α 是由相机发射出两条光线的夹角, 初始化光线时通过球面坐标系计算可得。

利用 β 实现物体分割的关键是设计算法查找连通分量。本文使用了一个复杂度为 $O(N)$ 的直通滤波器的变体算法来查找连通分量, 其中 N 是像素数量, 它保证最多访问深度图像中的任何像素两次。从图像左上角开始, 按照从上到下、从左到右的顺序遍历每个像素。遇到未标记的像素时, 从该像素开始进行广度优先搜索(BFS), 直到遍历完图像。

该算法有效地查找图像中的连通分量, 并为属于不同物体的像素分配对应的标签, 实现图像分割。算法的时间复杂度仅为 $O(N)$, 并且最多访问深度图像中的任何像素两次。因此, 该方法只需要最少的参数调整即可实现良好的分割性能。经处理后的结果如图 3 所示。

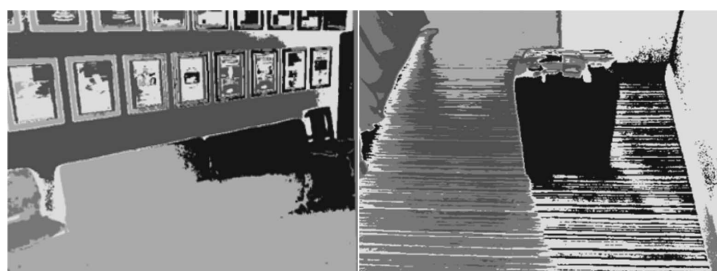


Figure 3. Image segmentation renderings
图 3. 图像分割效果图

3.2.2. 基于 Scharr 算子的自动阈值边缘检测算法

为明确场景中物体的层次结构以进一步增强空间真实感和提升渲染画质, 本文对图像的物体边缘进行解析和提取, 在众多图像边缘检测算法中, Canny 多级检测算法[22]简单直观、性能优越且准确性高。它的实现过程如图 4 所示。

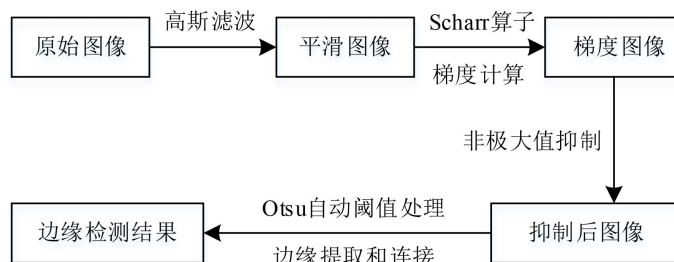


Figure 4. Process description of automatic threshold Canny algorithm
图 4. 自动阈值 Canny 算法流程描述

首先, 输入原始图像, 使用采用高斯滤波器对图像进行平滑处理, 得到平滑图像。

然后, 使用梯度算子计算图像梯度, 进一步增强边缘。梯度算子主要有 Roberts 算子、Sobel 算子、Laplacian 算子、Scharr 算子等。鉴于深度图像不同于常规 RGB 图像, 其边缘相对模糊、特征差异不明显, 所以本文使用对边响应更加强烈的 Scharr 算子计算图像的梯度和方向。首先对于每个像素, 应用 Scharr 算子进行卷积操作, 分别计算其在 x 和 y 方向上的梯度 G_x 、 G_y , 则梯度幅值 G 的计算公式为:

$$G = \sqrt{G_x^2 + G_y^2}. \quad (2)$$

同时, 计算每个像素的梯度方向, 并量化为四个主要的离散方向($0^\circ, 45^\circ, 90^\circ, 135^\circ$), 以便于后续边缘细化。之后, 对得到的梯度图像进行非极大值抑制, 以确保对边缘有且只有一个准确的响应, 经过非极大值抑制后的图像边缘更加清晰。最后, 执行双阈值处理, 以提升边缘的置信度和连续性。由于固定阈值无法提取每个角度场景的边缘信息, 因而使用 Otsu 算法确定自适应阈值。[23]得到阈值之后, 按照双阈值判定法则即可提取强边缘、弱边缘、非边缘, 并标记为不同颜色, 经过连接最终获得边缘检测结果。

经过上述操作, 场景深度图像的边缘信息被有效提取, 这些边缘信息都蕴含着丰富的特征, 是后续重建工作的关键, 提取的边缘效果如图 5。

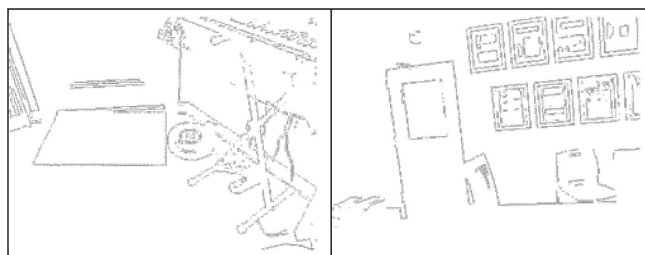


Figure 5. Edge detection renderings
图 5. 边缘检测效果图

3.3. 多元复合损失

经深度信息监督获得的采样点提供了观察方向和颜色信息以及深度信息, 它们被输入到神经网络中,

通过优化迭代网络参数训练 NeRF 模型。由于融合了新的数据——深度信息, 原模型的一致性约束条件无法充分利用深度信息, 导致模型的深度损失难以收敛, 其重建结果与原场景深度差异较大, 空间真实感不佳。为了促进深度误差的收敛, 本文设计深度损失函数并联合颜色与密度损失共同优化网络参数。

3.3.1. 颜色与深度的一致性约束

为了优化辐射场, 对于批处理中的每个像素, 通过评估体积渲染积分的离散形式来计算其颜色。^[24] 具体而言, 像素确定了一条射线, 其起点位于相机中心, 计算机模拟的光线沿此方向通过物体的路径进行采样。对于近平面和远平面内的每个采样位置, NeRF 可以获取该点的颜色和密度信息。

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{k=1}^K w_k \mathbf{c}_k \quad (3)$$

$$w_k = T_k (1 - e^{-\sigma_k \delta_k}) \quad (4)$$

其中,

$$T_k = e^{(-\sum_{k'=1}^k \sigma_{k'} \delta_{k'})} \quad (5)$$

$$\delta_k = t_{k+1} - t_k \quad (6)$$

除了预测光线的颜色之外, 还需要深度先验和方差来监督辐射场。首先, 对训练样本点在不同相机方向上进行蒙特卡洛采样生成多个深度值, 形成一个深度分布。然后, 取各样本点深度均值作为深度先验, 计算该深度分布标准差, 公式如下:

$$z(\mathbf{r}) = \sum_{k=1}^K t_k \quad (7)$$

$$s(\mathbf{r})^2 = \sum_{k=1}^K (t_k - z(\mathbf{r}))^2 \quad (8)$$

其中, t_k 表示当前样本点在第 k 个方向上的深度。

3.3.2. 完整损失函数

经上述推理, 网络的损失函数可描述为:

$$L_\theta = \sum_{r,i \in R} (\omega_1 \cdot L_d(r) + \omega_2 \cdot L_d(r) + \omega_1 \cdot L_e(i)) \quad (9)$$

其中, 颜色输出上的均方误差(MSE)项为:

$$\mathcal{L}_c(\mathbf{r}) = \|\hat{\mathbf{C}}(\mathbf{r}) - \mathbf{C}(\mathbf{r})\|_2^2 \quad (10)$$

深度输出上的高斯负对数似然(GNLL)项为:

$$\mathcal{L}_{\text{depth}}(\mathbf{r}) = \begin{cases} \log(\hat{s}(\mathbf{r})^2) + \frac{(\hat{z}(\mathbf{r}) - z(\mathbf{r}))^2}{\hat{s}(\mathbf{r})^2} & P \text{ 或 } Q \\ 0 & \text{其他} \end{cases} \quad (11)$$

其中, $\hat{z}(\mathbf{r})$ 表示预测深度值, $\hat{s}(\mathbf{r})$ 表示预测深度标准差, 条件 P 、 Q 分别为:

$$P = |\hat{z}(\mathbf{r}) - z(\mathbf{r})| > s(\mathbf{r}) \quad (12)$$

$$Q = \hat{s}(\mathbf{r}) > s(\mathbf{r}) \quad (13)$$

这表明深度损失应用于至少满足以下条件之一的射线: 1) 预测深度和目标深度之间的差异大于先验标准差或 2) 预测标准差大于先验标准差。这样, 损失就会促使 NeRF 在先验深度中最确定的表面观测值

的一个标准偏差内终止光线。此外, 通过解析获得的物体边缘信息可以度量生成边缘差异, 其损失函数可以表示为:

$$\mathcal{L}_e(i) = \frac{1}{N} \sum_{i=1}^N \|E_g(i) - E_r(i)\| \quad (14)$$

其中, N 表示像素总数, $E_g(i)$ 和 $E_r(i)$ 分别是生成图像和真实图像在第 i 个像素位置的边缘像素值。

综上, 该损失函数不仅降低深度损失和轮廓损失, 还保留了一定的密度分配自由度, 以最大程度地减少颜色损失。

3.4. 高斯平滑的周期编码

尽管原有的编码函数更好地拟合包含高频变化的数据, 显著提高了模型的性能, 但由于实景图片及其深度图隐匿了更多噪声。这些噪声可能来自传感器误差或环境干扰, 直接使用原始位置信息会导致模型过拟合噪声, 降低重建质量。为了平滑噪声, 保留场景位置信息的连续性, 本文首先使用高斯编码增强位置信息的表达能力, 然后再使用 \sin 和 \cos 编码保留原有编码的周期性特征。

高斯函数在频域中表现为低通滤波器, 能够有效抑制高频噪声, 同时保留低频信息, 从而平滑数据中的噪声干扰; 而周期编码(如 \sin 和 \cos 函数)则能够捕捉数据中的周期性特征, 在频域中表现为对周期性信号的高效表示。通过将高斯平滑与周期编码结合, 高斯平滑周期编码在时域上实现了对噪声的抑制, 同时在频域上保留并增强了数据的周期性特征。这种结合不仅优化了位置信息的表达能力, 还提高了模型对噪声的鲁棒性。完整编码方案步骤如下:

- 1) 首先对原始位置信息的 3D 坐标归一化至 $[-1, 1]$ 内, 假设得到归一化后的坐标值为 $p(x, y, z)$ 。
- 2) 对每个坐标值进行高斯编码, 高斯函数的形式是:

$$g(p) = ae^{-\frac{(p-\mu)^2}{2\sigma^2}} \quad (15)$$

其中, μ 是样本坐标值高斯分布的均值, σ 是样本坐标值高斯分布的标准差[25]。通过对位置坐标 p 进行加权平滑, 减少了噪声引起的局部波动, 增强了位置信息的鲁棒性。

3) 对于每个坐标值, 将高斯编码后的结果与 \sin 和 \cos 编码值相乘, 生成多尺度的特征表示, 这种多尺度编码能够同时捕捉场景的低频全局结构和高频局部细节, 从而显著提升 NeRF 对复杂场景的重建能力。映射函数 γ 可以形式化表示为:

$$\gamma(p) = (\sin(2^0 \pi p) \cdot g(p), \cos(2^0 \pi p) \cdot g(p), \dots, \sin(2^{L-1} \pi p) \cdot g(p), \cos(2^{L-1} \pi p) \cdot g(p)) \quad (16)$$

其中, 参数 L 的值决定了输出空间的维度。实验中, 对于位置坐标编码设置 $L=10$, 对于方向向量编码设置 $L=4$ 。将编码后的结果作为输入特征提供给神经网络进行训练和预测, 并通过实验证明了其在实际应用中的有效性。

4. 实验及结果分析

4.1. 评价指标

按照大部分 NeRF 相关文献采用的量化评价标准, 本文计算了测试结果 RGB 图像的 PSNR、SSIM、和学习感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS) [26], 用于度量渲染画质的质量。其中, PSNR 值越大, 表示重建图像和原始图像相似度越高; SSIM 值越大, 表示重建图像和原始图像相似度越高; LPIPS 值越小, 表示重建图像和原始图像之间的感知差异越小, 图像质量越好。

为了证实深度信息的有效性, 本文计算 NeRF 光线追踪终止深度与 ScanNet 数据集真实深度的均方

根误差(Root Mean Square Error, RMSE), 用于度量真实世界三维模型的深度一致性。RMSE 越小, 表示预测深度值与真实深度值差异越小, 三维模型的空间感还原度越高。

4.2. ScanNet 数据集实验结果

测试过程中, 本文使用 ScanNet 中的 4 个不同场景数据集进行实验, 记录相关实验数据并求平均值作为本文模型的参考实验结果, 详细数据见表 1。

Table 1. Experimental results of ScanNet data set

表 1. ScanNet 数据集实验结果

场景编号	PSNR	SSIM	LPIPS	RMSE
Scene708	25.95	0.789	0.302	0.585
Scene738	19.67	0.706	0.248	0.149
Scene758	20.87	0.789	0.259	0.147
Scene781	20.81	0.708	0.298	0.416

从表中可以看出, 不同场景的实验数据略有差异, 由于 Scene708 场景纹理相对简单, 其数据噪声较低, 自适应采样优化后能更好地提升整体渲染画质, 其 PSNR、SSIM 值相对较高。然而由于场景光线偏暗, 场景中深度差异区分度有所降低, 无法提供精准的深度信息监督模型训练, 其深度值 RMSE 为 0.585。因此本文在多个场景中进行了实验, 使得实验结果更具有泛化性, 最后取其平均值作为本文模型的参考实验结果, 用于后续消融实验和对比实验。

4.3. 消融实验

为了验证本文方法的有效性, 针对三个创新点分别设计消融实验, 并记录实验结果见表 2。

Table 2. Results of ablation experiments

表 2. 消融实验结果

模型	PSNR	SSIM	LPIPS	RMSE
本文 - 自适应采样	19.13	0.699	0.386	0.420
本文 - 损失函数	21.04	0.739	0.325	0.461
本文 - 高斯编码	21.98	0.746	0.301	0.362
本文 - 完整模型	21.83	0.748	0.277	0.324

由表中数据易得, 本文提出的优化方法在渲染画质和空间保真两方面均有显著效果。其中, 利用深度信息实现采样的自适应优化, 能对场景的边缘细节和深度分布进行分析, 获取更多有效特征, 从而保留更多的场景细节并还原空间真实感, 其 PSNR、SSIM 分别相对提升了 14.1%、7.0%, LPIPS、深度 RMSE 分别降低了约 0.11、0.10。融合深度信息的损失函数为模型提供了额外的约束和指导, 充分利用监督信号帮助模型更准确地学习场景, 以最直接的方式降低深度损失, 降低三维模型与原场景的深度差异, 它在保证渲染画质的同时, 其 LPIPS 和深度 RMSE 分别降低了约 0.05 和 0.14。高斯平滑的周期编码方案在一定程度上降低了数据的噪声, 进一步增强位置信息的表示。实验结果表明, 相较于仅使用周期编码, 高

斯平滑周期编码能够显著提升模型在含噪声数据上的性能，验证了其在平滑噪声与保留周期性特征之间的平衡能力。虽然在渲染画面的 PSNR 值上产生了略微损失，但是 LPIPS 和深度 RMSE 指标分别降低了约 0.02 和 0.04。上述消融实验数据充分说明了本文提出的方法在针对实景三维重建工作中能发挥优异性能，其在维持渲染画质的基础上，较大降低图像的感知差异和三维模型的深度差异，达到了预期效果。

4.4. 对比实验

通过文献查阅和实验操作，本小节评估了现有模型与经改进后的模型实验指标，实验结果对比见表 3。

Table 3. Experimental results comparing different models
表 3. 不同模型对比实验结果

模型	PSNR	SSIM	LPIPS	RMSE
NeRF [5]	19.03	0.670	0.398	1.163
NerfingMVS [17]	16.29	0.626	0.502	0.482
DS-NeRF [16]	20.85	0.713	0.344	0.447
本文	21.83	0.748	0.277	0.324

由表易得，本文方法展现出较好的重建效果。对于 ScanNet 实景数据集进行三维重建，原始 NeRF 在图像 PSNR 和 SSIM 方面有较为理想的结果，但是从深度 RMSE 值为 1.163 可以看出其重建出来的三维模型与原场景深度差异巨大；NerfingMVS 则极大程度上降低了 RMSE，改善了深度失真的问题，但是它损失了较多的渲染画质。DS-NeRF 则在各方面均展现出较优越的性能，是当前性能较好的模型。本文提出的改进方案进一步优化了重建效果，相对于 DS-NeRF，本文方法 PSNR、SSIM 分别提高了 4.7% 和 4.9%，LPIPS、RMSE 分别降低了约 0.07 和 0.12，最终在各项指标上都接近并优于对比实验模型。通过实验对比，验证了本文方法针对实景三维重建的有效性，充分利用了场景的深度信息，增加远景区域采样点数量以获取更多特征，并融合深度损失最小化深度误差，同时平滑数据噪声，为三维重建渲染画质的提升和空间感保真均做出了贡献。

4.5. 渲染细节展示

本文使用测试集进行场景渲染并获得可视化结果，在 4 个不同场景中各选取某个视角的渲染图像和对应深度图像与实况场景进行对比，结果如图 6 所示。

5. 结束语与未来展望

本文针对三维重建空间失真两个具体表现：深度损失明显和远处渲染画质不佳的问题，提出了基于深度信息优化空间失真的神经辐射场。该模型融合深度信息作为监督信号，并有效地利用深度信息引导自适应采样过程，通过图像边缘检测算法和分割算法，更准确地解析场景的复杂细节和深度分布，调整目标区域采样点数量和权重，帮助模型学习到更多场景特征。考虑到光线追踪终止深度，本文设计了新的损失函数优化颜色和深度损失，最小化三维模型的深度与实景深度差异。此外，采用高斯平滑的周期编码方案，降低原始数据的噪声影响，增强模型的表示能力。经实验表明，本文提出的方法能较清晰地重建出实景三维模型，同时较好地复原了场景的空间真实感。

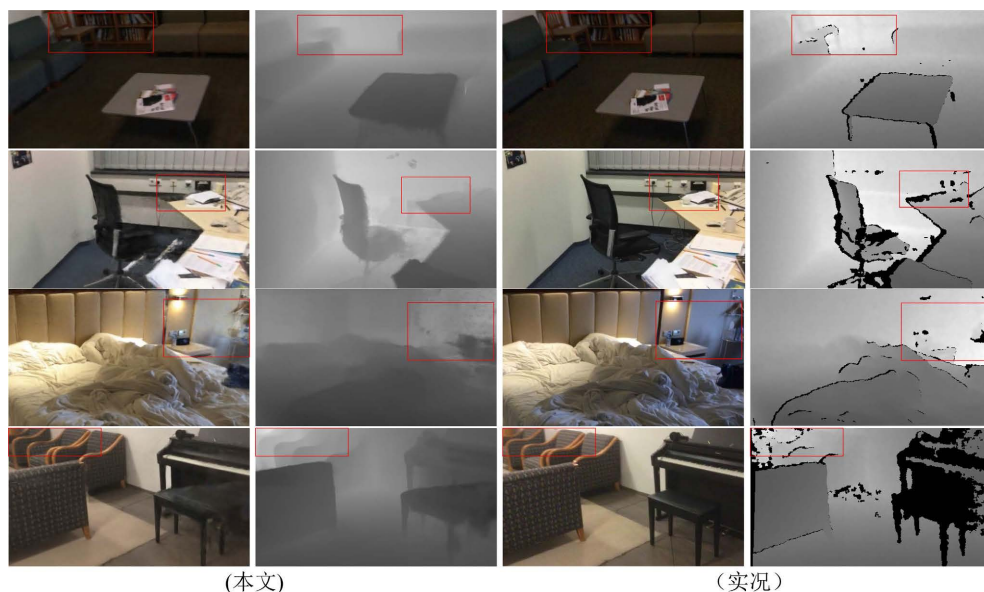


Figure 6. Rendering details comparison chart
图 6. 渲染细节对比图

实景三维重建是一个复杂且充满挑战的研究领域, 未来的研究可以从深度估计、光照建模、伪影消除、大规模场景重建、实时交互等多个方向展开, 同时结合跨领域应用需求, 推动三维重建技术的进一步发展。通过不断优化算法、提升计算效率、扩展应用场景, 三维重建技术将在虚拟现实、自动驾驶、医疗影像等领域发挥更大的作用, 为数字化世界构建更真实、更精细的三维模型。

参考文献

- [1] 宁小娟, 巩亮, 韩怡, 等. 结合语义分割与模型匹配的室内场景重建方法[J]. 中国图象图形学报, 2023, 28(10): 3149-3162.
- [2] 贾鑫. 基于深度学习的双目三维物体稀疏与稠密点云重建[D]: [博士学位论文]. 天津: 天津理工大学, 2022.
- [3] Dai, A., Ritchie, D., Bokeloh, M., Reed, S., Sturm, J. and Niessner, M. (2018) ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4578-4587. <https://doi.org/10.1109/cvpr.2018.00481>
- [4] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. and Geiger, A. (2019) Occupancy Networks: Learning 3D Reconstruction in Function Space. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 4455-4465. <https://doi.org/10.1109/cvpr.2019.00459>
- [5] 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. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer, 405-421. https://doi.org/10.1007/978-3-030-58452-8_24
- [6] 龚靖渝, 楼雨京, 柳奉奇, 等. 三维场景点云理解与重建技术[J]. 中国图象图形学报, 2023, 28(6): 1741-1766.
- [7] Charles, R.Q., Su, H., Kaichun, M. and Guibas, L.J. (2017) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 77-85. <https://doi.org/10.1109/cvpr.2017.16>
- [8] Zhao, H., Jiang, L., Fu, C. and Jia, J. (2019) PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 5560-5568. <https://doi.org/10.1109/cvpr.2019.00571>
- [9] 任飞, 常青玲, 刘兴林, 等. 基于点云的室内结构三维重建综述[J]. 计算机科学, 2022, 49(S2): 351-361.
- [10] Coudron, I., Puttemans, S., Goedemé, T. and Vandewalle, P. (2020) Semantic Extraction of Permanent Structures for the Reconstruction of Building Interiors from Point Clouds. *Sensors*, **20**, Article 6916. <https://doi.org/10.3390/s20236916>

- [11] Fang, H., Pan, C. and Huang, H. (2021) Structure-aware Indoor Scene Reconstruction via Two Levels of Abstraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 155-170. <https://doi.org/10.1016/j.isprsjprs.2021.06.007>
- [12] 韩开, 徐娟. 3D 场景渲染技术-神经辐射场的研究综述[J]. 计算机应用研究, 2024, 41(8): 2252-2260.
- [13] 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. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5835-5844. <https://doi.org/10.1109/iccv48922.2021.00580>
- [14] Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J.H., Chaitanya, C.R.A., *et al.* (2021) DONeRF: Towards Real-time Rendering of Compact Neural Radiance Fields Using Depth Oracle Networks. *Computer Graphics Forum*, **40**, 45-59. <https://doi.org/10.1111/cgf.14340>
- [15] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T. and Srinivasan, P.P. (2022) Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5481-5490. <https://doi.org/10.1109/cvpr52688.2022.00541>
- [16] Deng, K., Liu, A., Zhu, J. and Ramanan, D. (2022) Depth-Supervised NeRF: Fewer Views and Faster Training for Free. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 12872-12881. <https://doi.org/10.1109/cvpr52688.2022.01254>
- [17] Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J. and Zhou, J. (2021) NerfingMVS: Guided Optimization of Neural Radiance Fields for Indoor Multi-View Stereo. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5590-5599. <https://doi.org/10.1109/iccv48922.2021.00556>
- [18] Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A. and Duckworth, D. (2021) NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 7206-7215. <https://doi.org/10.1109/cvpr46437.2021.00713>
- [19] Tubiana, J., Schneidman-Duhovny, D. and Wolfson, H.J. (2022) Scannet: An Interpretable Geometric Deep Learning Model for Structure-Based Protein Binding Site Prediction. *Nature Methods*, **19**, 730-739. <https://doi.org/10.1038/s41592-022-01490-7>
- [20] Cabaret, L., Lacassagne, L. and Oudni, L. (2014) A Review of World's Fastest Connected Component Labeling Algorithms: Speed and Energy Estimation. *Proceedings of the 2014 Conference on Design and Architectures for Signal and Image Processing*, Madrid, 8-10 October 2014, 1-6. <https://doi.org/10.1109/dasip.2014.7115641>
- [21] Bogoslavskyi, I. and Stachniss, C. (2016) Fast Range Image-Based Segmentation of Sparse 3D Laser Scans for Online Operation. 2016 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, 9-14 October 2016, 163-169. <https://doi.org/10.1109/iros.2016.7759050>
- [22] Choi, K. and Ha, J. (2023) An Adaptive Threshold for the Canny Edge with Actor-Critic Algorithm. *IEEE Access*, **11**, 67058-67069. <https://doi.org/10.1109/access.2023.3291593>
- [23] 周钰聪, 叶超, 林子涵, 等. 复杂场景下偏振导航自适应图像分割算法研究[J]. 光学学报, 2024, 44(19): 150-160.
- [24] 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. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 12882-12891. <https://doi.org/10.1109/cvpr52688.2022.01255>
- [25] Dang, W., Liao, S., Yang, B., Yin, Z., Liu, M., Yin, L., *et al.* (2023) An Encoder-Decoder Fusion Battery Life Prediction Method Based on Gaussian Process Regression and Improvement. *Journal of Energy Storage*, **59**, Article ID: 106469. <https://doi.org/10.1016/j.est.2022.106469>
- [26] Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O. (2018) The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 586-595. <https://doi.org/10.1109/cvpr.2018.00068>