基于特征点引导干扰物识别的神经辐射场重建

任皓，李少波，弓茂，王博

(内蒙古科技大学自动化与电气学院，内蒙古 包头 014010)

摘　　要：针对神经辐射场（NeRF）在干扰物体影响下难以实现高质量三维重建的问题，提出一种基于运动恢复结构(SfM)与多视图立体匹配(SAM)模型协同优化的方法。以 SfM 重建过程中的 SIFT 算法为基础，利用动态场景中的几何不一致性进行特征点的识别与匹配，将未匹配的特征点视为动态干扰物，进而引导可以接受点引导分割的 SAM 模型实现动态遮挡物分割，生成静态场景掩膜。基于分割结果，使用掩码感知体积渲染技术预测颜色，并建立四重损失函数：重建损失、结构一致性损失、对抗损失和自监督修补损失。通过联合优化目标的方式约束被修补区域的颜色输出，经多次迭代训练后，实现多视角下被遮挡区域的几何结构与外观的一致性修复，保证辐射场完整性的同时，实现遮挡物的消除。经公开动态场景数据验证表明，利用掩膜体积渲染和联合优化后的重建效果相较于基线模型和主流遮挡物消除方法峰值信噪比(PSNR)平均提升了 5.24 dB，感知图像相似度(LPIPS)降低 35%，该方法为复杂动态环境下的三维重建提供了新范式。

关　键　词：神经辐射场；三维重建；动态场景；遮挡物消除；计算机视觉

中图分类号：TP 391.41　　**DOI**：10.11996/JGj.2095-302X.2026010111

文献标识码：A　　**文 章 编 号**：2095-302X(2026)01-0111-09

Neural radiation field reconstruction based on feature point-guided interference identification

REN Hao, LI Shaobo, GONG Mao, WANG Bo

(School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou Inner Mongolia 014010, China)

Abstract: To address the challenge of achieving high-quality 3D reconstruction with Neural Radiation Fields (NeRF) under the influence of occluding objects, a method based on the collaborative optimization of Structure-from-Motion (SfM) and the Segment Anything Model (SAM) was propose. Building upon the Scale-Invariant Feature Transform (SIFT) algorithm within the SfM reconstruction process, geometric inconsistencies in dynamic scenes were leveraged for feature point identification and matching. Unmatched feature points were treated as dynamic occluders, guiding the SAM model—capable of point-guided segmentation—to perform dynamic occluder segmentation and generate a static scene mask. Based on the segmentation results, mask-aware volumetric rendering was used to predict colors and a quadruple loss function was established: comprising reconstruction loss, structural consistency loss, adversarial loss, and self-supervised patching loss. These objectives were jointly optimized to constrain the color output in patched regions. After iterative training, consistent restoration of geometric structure and appearance in occluded areas across multiple viewpoints was achieved. The radiometric integrity was preserved while occlusions were removed. Validation on public dynamic scene datasets demonstrated that the mask-based volumetric rendering combined with joint optimization produced an average Peak Signal-to-Noise Ratio (PSNR) improvement of 5.24 dB over baseline models and mainstream occlusion removal methods, alongside a 35% reduction in Learned Perceptual Image Patch Similarity (LPIPS). This approach established a new paradigm for 3D reconstruction in complex dynamic environments.

收稿日期：2025-05-30；**定稿日期**：2025-09-08；**通信作者**：李少波，E-mail：12965874@qq.com

Received：30 May，2025；**Finalized**：8 September，2025；**Corresponding author**：LI Shaobo，E-mail：12965874@qq.com

基金项目：内蒙古自然科学基金(2022LHMS06002)

Foundation items：Inner Mongolia Natural Science Foundation (2022LHMS06002)

Keywords: neural radiation field; 3D reconstruction; dynamic scene; occlusion removal; computer vision

基于视觉的三维重建方法是当前计算机图形学和计算机视觉所关注的重点领域[1]。尽管传统的三维重建方法，如运动恢复结构 (Structure-from-Motion, SfM) [2]和多视图立体匹配(Multi-View Stereo, MVS) [3]已经成功实现了稀疏点云重建和稠密几何恢复，但其仍受限于显式表示固有缺陷，如 MVS 在无纹理区域生成的密集点云可因匹配失败产生空洞。伴随着深度学习的迅猛发展，在 2020 年 Mildenhall 团队提出了神经辐射场(Neural Radiance Field，NeRF)[4]，其利用了多层感知机(Multilayer Perceptron, MLP) [5]拟合射线上每个点的密度(σ)以及颜色进而从二维图像中重建对应的三维场景。NeRF 的出现在改变 3D 渲染范式的同时也为今后的三维重建方式提供了新的思路。

然而，NeRF 依旧存在计算效率低以及遮挡处理欠佳等问题。FastNeRF[6]和 Instant-NGP[7]等针对 NeRF 训练速度慢的问题，利用轻量化 MLP 架构使训练速度提升 60 倍，却忽视了真实环境中人物、车辆的遮挡以及光线变化时产生的阴影等干扰因素。由于重建时默认为静态场景，但在拍摄动态物体时不同视角中的位姿不一致，导致 MLP 拟合时接收到冲突信号，使得目标重建图像存在模糊不清的情况。NeRF-W[8]利用 GLO(Generative Latent Optimization)网络[9]解决了上述问题，通过对每个图片引入外观编码，并在 NeRF 训练时进行共同训练，而面对遮挡物问题时，在 MLP 网络的输出中加入了“不确定因素”的计算。因此，可以使用多个独立的神经网络来分别处理动态遮挡场景和正常无干扰的静态场景。除此之外，Ha-NeRF[10]利用了一个独立的 MLP 来塑造每个图片的可见性，同时使用卷积神经网络(Convolutional Neural Network, CNN)编码器 Encoder 提取输入图片的外观信息，并通过动态-静态解耦与物理引导的遮挡推理，在动态场景重建中实现了高精度与高效率的平衡，但其性能依赖于运动先验的准确性且无法应对复杂动态。而 SF-NeRF[11]对输入的多视角图像，使用预训练的 2D 语义分割模型来提取像素级语义标签，利用多视角几何约束，对分割结果进行跨视角验证，若语义标签出现一致几何冲突(如出现在不同深度)，则标记为潜在的动态遮挡物。在 Block-NeRF[12]中同样利用语义分割的思想，通过空间分块、独立建模、动态融合的三阶段方式，在大规模场景中实现了突破。

尽管 NeRF 的研究取得了突破性进展，但是现有方案(如 Ha-NeRF，SF-NeRF 和 Block-NeRF)定义的动态干扰物均为“行人及车辆”，其识别能力难以泛化至随机物体，从而导致重建效果不佳甚至无法渲染包含动态对象的场景。为此，本文提出一种基于特征点匹配引导 SAM[13] (Segment Anything Model)模型分割生成静态掩码实现自动识别干扰物以及体积渲染联合优化的修复方法，其动态-静态解耦新机制，利用几何不一致性替代语义先验，实现动态干扰物自主检测，并采用掩码渲染公式通过掩膜值解耦动静态场景，最后构建四元损失函数进而实现区域多视角一致性修复。该方式在无需大量先验训练的前提下便可实现动态遮挡物识别与剔除，且大大提升了三维重建的渲染质量。本文算法流图如图 1 所示。

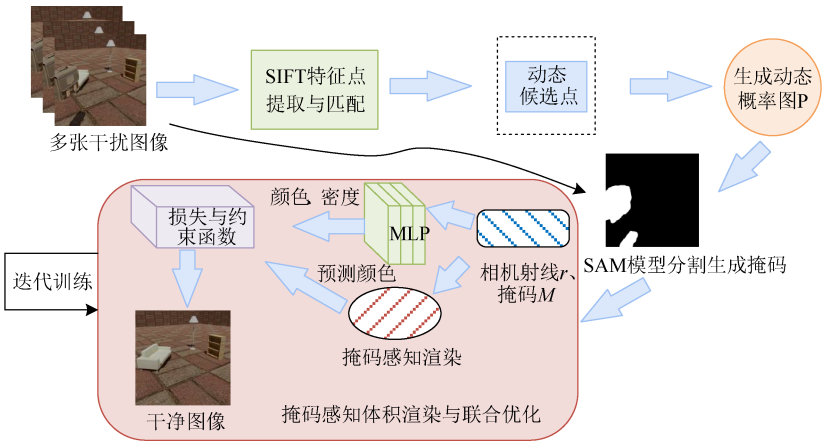


图 1 本文算法流图

Fig. 1 Algorithm flowchart for this paper

1　动态干扰物的检测与识别

在三维重建的过程中，由于动态干扰物的存在。会对重建效果产生影响。因此需要辨别确认该物体是需要被移除的干扰物还是需要重建的目标物。

图 2 展示了干扰物与需要重建场景的关系。识别干扰物最常用的方法是利用语义信息作为先验知识来区分图中的像素是属于遮挡物还是属于目标[14]，其主要特点在于识别物体足够精确但不够通用。Ha-NeRF 和 Up-NeRF[15]通过人工标注方法在 NeRF 训练阶段，分离瞬态干扰物。该方法虽然避免了先验知识依赖，但面临 2 大核心问题：①逐帧手动标注机制导致人力成本高昂，难以扩展至大规模动态场景；②人工标注界限存在主观偏差，易引发分割掩膜与真实辐射场间的对齐误差。基于上述问题，需要找到一个既通用又有足够精度的方法实现瞬态干扰物的识别，提出一种基于 SIFT 算法提取特征点引导分割模型区分动静态物体的方法。

(a)　　(b)

图 2　重建目标与干扰物((a) 无垃圾桶；(b) 有垃圾桶)

Fig. 2　Reconstruction target and disturbance ((a) No trash can; (b) There are trash bins)

1.1　SIFT 特征点提取与匹配

SfM 作为一种从多张二维图像中重建三维场景的技术，其依赖于多视角之间的几何关系，主要流程为：①特征提取；②特征匹配；③基于几何的特征对验证；④特征建树。

在 SfM 算法中，需要在不同图像间建立准确的特征对应关系，其中 SIFT(Scale-Invariant Feature Transform)算法[16]作为特征提取与描述的经典算法，有着很强的尺度不变性以及较高区分度，其在动态干扰物(运动物体)的特征描述与对干扰物的区分上都具有较好的鲁棒性，其区分出的特征点可以为 SAM 模型提供高质量的提示。该算法的核心在于通过高斯金字塔构建多尺度空间 $\boldsymbol{L}$，并筛选得到具有几何不变性的特征点。对于输入图像 $\boldsymbol{I}(x,y)$ 的尺度空间 $\boldsymbol{L}(x,y,\sigma)$ 计算方式为

$$\boldsymbol{L}(x,y,\sigma)=\boldsymbol{G}(x,y,\sigma)*\boldsymbol{I}(x,y) \tag{1}$$

其中，高斯核函数可表示为

$$\boldsymbol{G}(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} \tag{2}$$

通过不同的 σ 值构建高斯金字塔，式中：σ 表示控制尺度即模糊程度，其大小决定了图像的平滑程度，大尺度对应图像的概貌特征，小尺度对应细节特征；$*$ 表示卷积操作。为提升在尺度空间 $\boldsymbol{L}$ 中关键点的寻找效率，本文提出了高斯差分尺度空间(DoG scale-space)方法，即

$$\boldsymbol{D}(x,y,\sigma)=\boldsymbol{L}(x,y,k\sigma)-\boldsymbol{L}(x,y,\sigma) \tag{3}$$

式中：k 表示尺度倍增因子(通常为 $\sqrt{2}$)。关键点通过检测 $\boldsymbol{D}(x,y,\sigma)$ 的局部极值(与相邻的 26 个点比较)确定。但检测到的局部极值点并非真正的极值点而是离散空间的极值点，因此需要确定精确的关键点的位置和尺度以增强匹配的准确性。利用三维泰勒展开得到极值点的偏移量，即

$$\widehat{\boldsymbol{X}}=-\frac{\partial\boldsymbol{D}^{-1}}{\partial\boldsymbol{X}^2}\frac{\partial\boldsymbol{D}}{\partial\boldsymbol{X}} \tag{4}$$

对应的极值点的方程为

$$\boldsymbol{D}\left(\widehat{\boldsymbol{X}}\right)=\boldsymbol{D}+\frac{1}{2}\frac{\partial\boldsymbol{D}^r}{\partial\boldsymbol{X}}\widehat{\boldsymbol{X}} \tag{5}$$

式中：$\boldsymbol{X}=(x,y,\sigma)^{\mathrm{T}}$ 表示极值点的坐标与尺度，若修正后的偏移量超过 0.5 像素，则调整到临近位置，同时删除对比度 $\left(\boldsymbol{D}\left|\widehat{\boldsymbol{X}}\right|<0.03\right)$ 和边缘相应点 $\boldsymbol{H}$，其计算式为

$$\boldsymbol{H}=\begin{pmatrix}\boldsymbol{D}_{xx} & \boldsymbol{D}_{xy}\\ \boldsymbol{D}_{xy} & \boldsymbol{D}_{yy}\end{pmatrix} \tag{6}$$

设特征值为 α 和 β，r 为阈值，采用 $r=10$，若满足

$$\frac{(\alpha+\beta)}{\alpha\beta}>\frac{(r+1)^2}{r} \tag{7}$$

则判定为边缘响应点并进行去除操作。在关键点邻域内计算梯度幅值 $\boldsymbol{m}(x,y)$ 和方向 $\boldsymbol{\theta}(x,y)$，即

$$\boldsymbol{m}(x,y)=\sqrt{(\boldsymbol{L}(x+1,y)-\boldsymbol{L}(x-1,y))^2+(\boldsymbol{L}(x,y+1)-\boldsymbol{L}(x,y-1))^2}$$

$$\boldsymbol{\theta}(x,y)=\arctan\left(\frac{\boldsymbol{L}(x,y+1)-\boldsymbol{L}(x,y-1)}{\boldsymbol{L}(x+1,y)-\boldsymbol{L}(x-1,y)}\right) \tag{8}$$

式中：$\boldsymbol{L}$ 表示关键点所在的尺度空间值。将关键点周围 16×16 的区域划分为 4×4 的子区域，并计算每个子区域 8 个方向梯度直方图，构成 128 维度的描述子 d_k，为了去除光照变化的影响对其进行归一

化处理，即

$$d_k \leftarrow \frac{d_k}{\| d_k \|_2} \tag{9}$$

对图像 I_i 及 I_j 的描述子集 $\{d_k\}$ 及 $\{d_l\}$ 进行筛选匹配点对 $M_{ij}=\left\{(x_k,y_k)\middle|\|d_k-d_l\|_2<\tau_1,\|d_l-d_k\|_2<\tau_1\right\}$，其中 τ_1 为距离阈值(通常取 0.7 倍最大距离)，至此，完成了 SIFT 算法对图像特征点的提取与匹配。

1.2 干扰物的分割

SAM 分割模型是一种基于 Transformer 的通用图像分割模型，由 3 部分构成：①用于将输入图像转化为高维特征的图像编码器；②将用户所给的提示(如点、框、多点和文本等)编码为向量的提示编码器；③用于生成分割掩码的掩码解释器。在特征点匹配的过程中，干扰物因其具有运动的特性导致其表面无法满足几何一致性的约束，从而导致匹配失败。设未匹配点 $\{x_k\}$ 为动态候选点，动态概率图 $P_{\text{dyn}}(x)$ 通过高斯核扩散建模为

$$P_{\text{dyn}}(x)=\frac{1}{Z}\sum_{x_k\in\text{unmatched}}^{n}\exp\left(-\frac{\|x-x_k\|_2^2}{\sigma^2}\right) \tag{10}$$

式中：Z 表示归一化常数；σ 表示控制动态影像的扩散范围。未匹配点的高斯加权扩散模拟物体的空间连续性，概率越高表示其为动态物体的可能性越大。随后利用动态概率图 $P_{\text{dyn}}(x)$ 作为负向提示优化 SAM 分割的掩码 M，其损失函数为

$$\lambda_{\text{mask}}=-\sum_x\Big[\xi\log p\big(M=1|x\big)\cdot\big(1-P_{\text{dyn}}(x)\big)+(1-\xi)\log p\big(M=0|x\big)\cdot P_{\text{dyn}}(x)\Big] \tag{11}$$

式中：ξ 表示控制静态区域的权重，通过反向传播调整 SAM 的注意力权重，动态概率图 $P_{\text{dyn}}(x)$ 抑制动态区域的注意力权重；$p\big(M=1|x\big)$ 表示 SAM 输出的像素 x 属于静态掩膜的概率。值得注意的是，未匹配的特征点并非完全来自于动态干扰物，低纹理区域和重复纹理区域同样会导致特征点匹配失败，但是动态概率图提供的是一个连续的概率估计，而非二值化的动态点标签，其可更准确地反映了空间位置属于动态物体的可能性，而非绝对判定利用 SAM 本身具有强大的零样本泛化能力和对图像上下文的理解能力重点关注几何冲突区域，并将其划分为动态干扰区域。

SAM 分割的多视角一致性原理：SIFT 未匹配的特征点本质上反映的是三维场景的运动不一致性，其空间分布具有视角不变性的特点。利用高斯扩散生成的动态概率图将离散特征点转化为连续概率场，迫使 SAM 在分割时优先关注几何冲突区域。而在 NeRF 优化阶段，掩膜值被转化为透明度加权的概率混合。即使 SAM 分割存在小范围边界抖动，NeRF 的连续隐式表示会通过密度场平滑性自然吸收误差，从而确保多视角一致性。综上所述利用未匹配点的空间分布，结合 SAM 的分割能力，可有效区分待重建物体与干扰物，实现了减少工作量的同时使其应用范围更加广泛。为了进一步评估分割的有效性，本文在人工合成数据集 Kubric 上统计了特征点分割的准确率，见表 1。实验通过减少初始图像平滑，以保留更多高频细节；并在更多尺度层检测特征点，以提升对物体覆盖。由表 1 可知，大部分未匹配特征点为干扰物，而少部分未匹配特征点为不规则分布，SAM 并未将其划分为干扰区域。

表 1 特征点分割准确率评估

Table 1 Feature point segmentation accuracy evaluation

数据集	特征点总数	未匹配点数量	未匹配点为干扰物数量
Bag	182	46	39
Pillow	157	31	22
Cars.	126	30	22
Chair	243	62	45
平均	174	169	128

表 2 中 4 种场景下的定量实验结果对算法性能进行了综合评估。实验采用准确率(Accuracy, Acc)和交并比(Intersection-over-Union, IoU)2 种评价指标。其中 Acc 是指图像中正确分类的像素百分比，即预测类别正确的像素的比例，Acc 越高，说明生成的掩膜和真实的掩膜重叠区域越大。IoU 是模型对某一类别预测结果和真实值的交集与并集的比值，IoU 越大说明生成的掩膜与真实掩膜越接近。分析表明，本方法在 Acc 和 IoU 指标上均优于 SAM 模型，其生成的掩膜与真实标注具有更高的空间重叠度及轮廓一致性。因人工标注作为基准真值不参

表 2 3 种方法的结果对比/%

Table 2 Comparison of results from three methods/%

数据集	手工标注		SAM		本文	
	Acc	IoU	Acc	IoU	Acc	IoU
Bag	—	—	98.1	92.5	98.4	95.2
Pillow	—	—	97.6	90.6	98.4	95.3
Cars	—	—	97.4	93.1	97.7	94.2
Chair	—	—	97.1	90.0	98.3	94.8

注：—表示手工标注为基准，未参与统计。

与性能指标比较，所以在计算效率方面，远快于手工标注，结果表明，本文提出的分割框架利用特征点引导结合 SAM 模型，可以在保持较高的计算效率的同时保持精度。

2　干扰物排除修补的方法

在复杂开放场景的三维重建过程中，由于动态干扰物的存在，会使多视角图像产生严重的不一致性。这类干扰物的主要特点是不同视角的空间位置会发生显著变化并伴随着不可预见性。传统的基于多视图的几何优化算法在估计场景结构和相机位姿的过程存在误差，导致重建目标物产生表面空洞或伪影等问题，如何从多视角观测数据中实现鲁棒地剔除干扰物，并恢复场景的几何完整性成为了重建时面临的挑战。本文受 SPIn-NeRF[17]的启发，通过掩码感知的体积渲染和联合优化目标的方式，实现了多视角下的被遮挡区域的几何与外观一致性修复。

2.1　掩码感知体积渲染

在传统的 NeRF 中，其体积渲染公式通过沿射线积分颜色(RGB)与密度 σ 重建像素点，即

$$\begin{aligned}\widehat{C}(r)&=\int_{t_n}^{t_f}T(t)\sigma\left(r(t)\right)c\left(r(t),d\right)\mathrm{d}t\\T(t)&=\exp\left(-\int_{t_n}^{t}\sigma\left(r(s)\right)\mathrm{d}s\right)\end{aligned}\tag{12}$$

式中：$\sigma(x)$ 表示在 x 处的体积密度；$c(x,d)$ 表示场景在 x 处沿方向 d 的颜色；$T(t)$ 表示一条射线从近端 t_n 到当前位置 t 的累计透射率。但在真实场景中，由于存在动态遮挡物，导致射线采样不理想，输入的多视角图像包含遮挡区域掩码 $M=\{M(r)\}$，其中若 $M(r)$ 为 1，则表示该区域为待重建区域，若 $M(r)$ 为 0，则表示该区域为干扰区域。为分离待重建区域与干扰区域提出掩码感知体积渲染，即

$$\begin{aligned}\widehat{C}(r)=\int_{t_n}^{t_f}T(t)\sigma\left(r(t)\right)\Big[&M\left(r(t)\right)c\left(r(t),d\right)+\\&\left(1-M\left(r(t)\right)\right)c_{\mathrm{fill}}\left(r(t),d\right)\Big]\mathrm{d}t\end{aligned}\tag{13}$$

此处加入了修补颜色区域 $c_{\mathrm{fill}}(x,d)$，用于干扰区域的潜在外观建模。利用掩码 $M(r)$ 的引导，待重建区域的颜色由 $c(r,t)$ 直接监督，而在干扰区域中，由 $c_{\mathrm{fill}}(x,d)$ 生成并利用多目标约束以及优化。将射线 $r(t)$ 离散化为 N 个采样点 $\{t_i\}_{i=1}^{N}$，渲染公式为

$$\widehat{C}(r)=\sum_{i=1}^{N}T_i\alpha_i\left[M_ic_i+\left(1-M_i\right)c_{\mathrm{fill},i}\right]\tag{14}$$

式中：$\alpha_i=1-e^{-\sigma_i\delta_i}$ 表示第 i 点的透明度 $(\delta_i=t_{i+1}-t_i)$；$T_i=\prod_{j=1}^{i-1}\left(1-\alpha_j\right)$ 表示累计透射率；$c_i=c(x_i,d)$；$c_{\mathrm{fill},i}=c_{\mathrm{fill}}(x_i,d)$；$M_i=M(x_i)$ 表示采样点 $x_i=r(t_i)$ 处的掩码值。该设计旨在对待重建区域和干扰区域进行解耦，以优化遮挡区域的几何 $(\sigma(x))$ 与外观。

2.2　定义损失函数

为了保证待重建区域的重建精度以及干扰区域中的修复的合理性，本文定义了 4 项损失函数：

1）重建损失 $\Upsilon_{\mathrm{recon}}$，是为了约束待重建区域的渲染颜色与真实像素值一致性，即

$$\Upsilon_{\mathrm{recon}}=\sum_{r\in R}\left\|\sum_{i=1}^{N}T_i\alpha_iM_ic_i-C_{\mathrm{gt}}(r)\right\|_2^2\tag{15}$$

式中：R 表示采样射线的集合；$C_{\mathrm{gt}}(r)$ 表示射线 r 对应的真实像素的颜色。

2）结构一致性损失 $\Upsilon_{\mathrm{struct}}$，是为了避免干扰区出现空洞等非真实因素失，即

$$\Upsilon_{\mathrm{struct}}=\lambda_d\sum_r\left\|\nabla D(r)\right\|_2^2+\lambda_s\sum_x\left\|\nabla\sigma(x)\right\|_2^2\tag{16}$$

该式可分为深度平滑和密度平滑。在深度平滑中，$D(r)=\sum_{i=1}^{N}T_i\alpha_it_i$ 表示射线 r 的期望深度，$\nabla D(r)$ 表示深度值的空间梯度，避免相邻射线的深度突变。在密度平滑中，$\nabla\sigma(x)$ 表示密度场的空间梯度，避免密度发生突变。

3）对抗损失 γ_{adv}，引入判别器 D 提升干扰区域的纹理真实性，即

$$\begin{aligned}\Upsilon_{\mathrm{adv}}=\mathrm{E}_r\Bigg[&\log D\left(\sum_{i=1}^{N}T_i\alpha_i\left(1-M_i\right)c_{\mathrm{fill},i}\right)+\\&\log\left(1-D\left(C_{\mathrm{gt}}(r)\right)\right)\Bigg]\end{aligned}\tag{17}$$

判别器 D 为 2D 卷积网络，输入 64×64 的图像块，输出为真实性概率，输出值域[0，1]。利用对抗训练使得 c_{fill} 生成真实纹理。$\sum_{i=1}^{N}T_i\alpha_i\left(1-M_i\right)c_{\mathrm{fill},i}$ 表示生成输出的遮挡区域渲染结果，$C_{\mathrm{gt}}(r)$ 表示真实像素颜色。在训练中采用交替频率控制：每训练生成器一步，需更新判别器两步，从而避免判别器过强导致生成器梯度消失。并添加 WGAN-gp 作为梯度惩罚，提升收敛性。通过动态训练策略与梯度惩罚，判别器通常在 20 ~ 50 k 迭代后稳定，输出真实值概率 ＞0.9，生成块概率在 0.4 ~ 0.6 区间波动。

4）自监督修补损失 Υ_{self}，旨在训练阶段生成随机的辅助掩码 M' 遮挡部分待重建区域，以加强

模型在上下文中恢复内容的能力，即

$$\Upsilon_{\text{self}}=\sum_{r}\left\|\sum_{i=1}^{N}T_i\alpha_i\left(1-M_i^{'}\right)c_{\text{fill},i}-C_{\text{gt}}\left(r\right)\right\|_1 \quad (18)$$

总损失函数为：$\Upsilon_{\text{all}}=\Upsilon_{\text{recon}}+\Upsilon_{\text{struct}}+\lambda_{\text{adv}}\Upsilon_{\text{adv}}+\lambda_{\text{self}}\Upsilon_{\text{self}}$，其中 λ_{adv} 和 λ_{self} 表示超参数，分别取 0.01 和 0.05。在训练过程中采用分段方式，前期集合初始化阶段，仅优化 Υ_{recon} 及 Υ_{struct}，训练 100 000 次用于初步恢复场景结构。在联合优化阶段引入 Υ_{adv} 和 Υ_{self}，继续训练 100 000 次用于细节恢复。经过 2 个阶段共训练 200 000 次训练操作，可达到良好的修复效果。

3 实验结果与分析

为验证本算法的有效性与正确性，对 5 种不同场景下的公开数据集进行对比，修复后的结果图像与 3D Gaussian Splatting[18]，NeRF 及 NeRF-on-the-go 进行对比。本文的实验环境基于 Ubuntu20.04 系统，搭载单个 NVIDIA RTX 3080 GPU。每个模型 200 000 次训练时间约为 22~28 h。以蟋蟀为遮挡物的小尺度遮挡测试结果如图 3 所示。

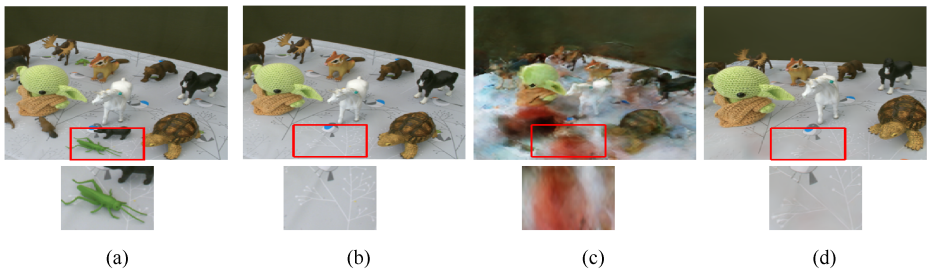


图 3 Yoda 数据集渲染效果((a) 干扰图像；(b) 消除后的图像；(c) NeRF；(d) 3DGS)

Fig. 3 Yoda dataset rendering effect ((a) Interference image; (b) The image after elimination; (c) NeRF; (d) 3DGS)

由图 3 可知，NeRF 无法实现遮挡物的去除，且重建结果存在严重缺陷；由于目标物体表面细节大量丢失，在遮挡区域出现残留伪影。由此可见，NeRF 的隐式表征很难完全解耦动态遮挡物与动态场景的耦合信息。而 3DGS 利用显式点云的策略，使得目标物表面细节的还原优于 NeRF，但其重建结果仍受遮挡物影响，可出现模糊不清的状况。相比之下，本文方法通过动态遮挡掩码的协同优化，不仅实现了伪影与遮挡物的抹除，并可对物体的细节实现高质量的重建。

以木头人为例的大尺度遮挡场景测试如图 4 所示。NeRF 展示了较强的遮挡物去除能力，但其重建质量受限于体渲染的固有特性，且会在木制表面出现马赛克状的模糊伪影，在光照较强的高亮区域出现明显的失真。3DGS 则表现出更强的抗干扰能力，其显示高斯分布模型在物体边缘重建与光照适应方面优于 NeRF，重建结果更加接近真实场景，但在去遮挡方面效果不佳，其重建优势仅表现在无遮挡的静态区域。本方法可以有效抑制了高光区域伪影的生成，使得重建物表面的材质反射特性与原始场景高度一致。

以水瓶为遮挡物的交叉覆盖场景如图 5 所示。NeRF 的重建结果虽然在完整性和去除遮挡物方面有所改善，但仍无法避免重建质量的模糊以及云状伪影，表明其隐式表征对物体建模存在局限性。3DGS 虽然能去除部分遮挡物，并在建模上接近原始图像，但由于删除了几何点云，错误地将罐车当成了与水瓶相似的遮挡物，导致重建物体的完整性

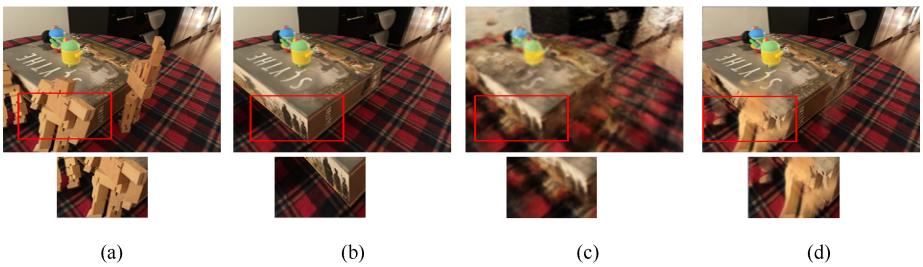


图 4 Android 数据集渲染效果((a) 干扰图像；(b) 消除后的图像；(c) NeRF；(d) 3DGS)

Fig. 4 Android dataset rendering effect ((a) Interference image; (b) The image after elimination; (c) NeRF; (d) 3DGS)

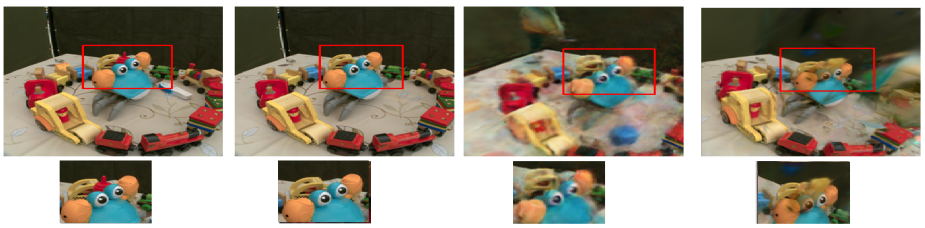


图 5　Carb 数据集渲染效果((a) 干扰图像；(b) 消除后的图像；(c) NeRF；(d) 3DGS)

Fig. 5　Carb dataset rendering effect ((a) Interference image; (b) The image after elimination; (c) NeRF; (d) 3DGS)

不强。本文利用特征点准确引导 SAM 分割模型识别出干扰物，并进行重建，最终实现了遮挡物与目标物体的剔除与重建。

从图 6 与图 7 中可以看出，NeRF-on-the-go 在排除干扰物上表现优秀，但在排除干扰后的细节处理上存在问题，如图 6 所示。机器人剔除之后，两侧地板会有明显分界线且地板细节重建也存在分层现象。在图 7 中，墙体依旧会有重建效果不佳的问题。这主要是因为 NeRF-on-the-go 在预测具有强视角依赖效果区域的不确定性方面仍存在问题，如对于高度反射的表面、地板和墙体等，同时需要满足足够的训练视角，当训练视角变得稀疏时，其性能会显著下降。在硬件方面，内存需要 80 G GPU，如果减小批处理大小，则会严重增加训练时间，训练成本远高于当今主流三维重建方法。

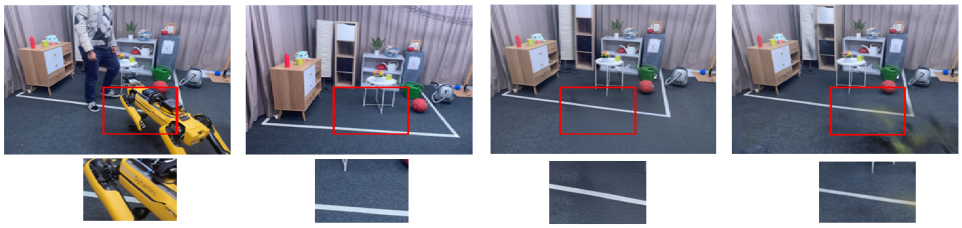


图 6　Corner 数据集渲染效果((a) 干扰图像；(b) 消除后的图像；(c) NeRF；(d) 3DGS)

Fig. 6　Corner dataset rendering effect ((a) Interference image; (b) The image after elimination; (c) NeRF; (d) 3DGS)

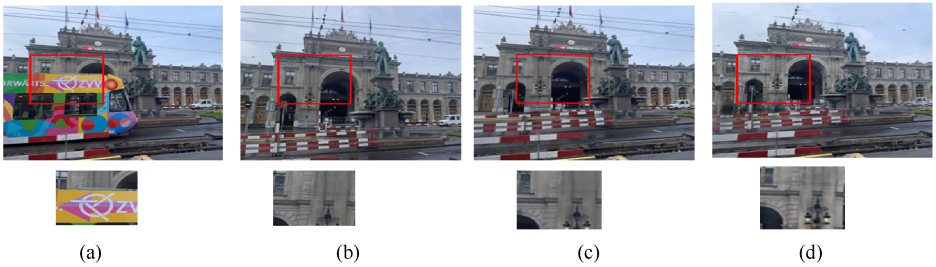


图 7　Station 数据集渲染效果((a) 干扰图像；(b) 消除后的图像；(c) NeRF；(d) 3DGS)

Fig. 7　Station dataset rendering effect ((a) Interference image; (b) The image after elimination; (c) NeRF; (d) 3DGS)

不同数据及重建后的效果评估指标见表 3。其中，峰值信噪比(Peak Signal to Noise Ratio, PSNR)是数字信号处理领域的经典评判指标，以评价重建模型与真实场景之间的一致性，表征重建结果的保真度。学习感知图像块相似性(Learned Perceptual Image Patch Similarity, LPIPS)量化重建结果与真实场景的感知相似性，是常见的评估指标。Loss 在本文中指的是颜色重建误差，由于本实验是剔除图片中存在的干扰物，会引入颜色重建误差，其误差值取决于干扰物与重建物体的颜色接近程度。由表 3 可知，本文方法仅在 Yoda 数据集上略差于 NeRF-on-the-go，在 Android 数据集中，由于存在大遮挡物，导致 Loss 值相对较高。综上，说明本文方法无论是细节保存上还是视觉效果上都极大地还原了原本的视图效果。

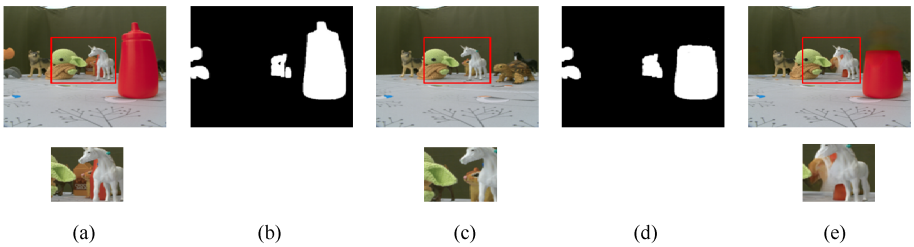
图 8 为分割失败案例。在 Yoda 数据集中：蟹

表 3 重建后的效果评估

Table 3 Post-reconstruction effect evaluation

方法	Yoda			Android			Crab		
	LPIPS ↓	Loss ↓	PSNR ↑	LPIPS ↓	Loss ↓	PSNR ↑	LPISP ↓	Loss ↓	PSNR ↑
NeRF	0.36	0.07	20.46	0.29	0.03	22.03	0.18	0.06	22.65
3DGS	0.14	0.02	27.08	0.16	0.04	23.88	0.09	0.02	27.75
NeRF-W	0.16	0.03	26.64	0.25	0.03	20.62	0.15	0.03	26.91
MipNerf360	0.23	0.02	23.75	0.18	0.02	21.81	0.09	0.02	26.25
On-the-go	0.19	0.03	28.96	0.13	0.04	23.10	0.11	0.03	27.55
本文方法	0.11	0.01	27.55	0.13	0.02	25.32	0.09	0.01	28.01

注：↓表示数值越低越好；↑表示数值越高越好。



(a) (b) (c) (d) (e)

图 8 分割失败案例((a) 干扰图像；(b) 成功分割；(c) 成功重建；(d) 失败分割；(e) 失败重建)

Fig. 8 Failed partitioning cases ((a) Interference image; (b) Successful segmentation; (c) Successful reconstruction; (d) Failure segmentation; (e) Failure reconstruction)

爪、白马和红色水瓶分别代表正常分割、过分割和欠分割。在过分割时，掩膜边界与辐射场优化发生冲突，重叠区域同时受静态重建损失与干扰区对抗损失梯度拉扯，导致重建既无法有效消除干扰物体，也难以清晰呈现物体边缘，产边界粘连现象。而在欠分割时，会直接导致干扰物体消除不彻底。造成过分割主要是因为在低纹理区域 SIFT 匹配的失败率提升，导致动态概率图在静态区域扩散。而以红色水瓶为代表的欠分割，在多视角满足几何一致性，且未匹配点低于检测阈值，导致无法引导 SAM 分割。

4 结束语

针对动态场景的三维重建存在的动态遮挡问题，本文提出了一种具有广泛应用性的识别以及修复方法，解决了现阶段大多数的识别需要人工标注以及适配性低的问题。其核心贡献在于：①以引入了动态特征点引导分割模型的方式，代替了传统的语义先验识别干扰物，形成高精度的掩膜；②设计掩码感知的联合优化目标函数，在辐射场训练中融合掩膜与遮挡区域外观一致性损失，实现了多视角下的被遮挡区域的几何与外观一致性修复。通过不同数据集以及不同方法的比较结果表明，本方法能够很好地消除由于动态干扰物的视图不一致带来的伪影，并且保持修复后的图像质量，为后续的三维重建工作提供了扎实的基础。但本研究存在以下局限，当场景长时间保持静止且在输入视图中完全遮挡同一区域的动态物体时，本方法将无法准确识别，而因这类物体可满足几何一致性而被误判为静态场景的一部分，导致其被错误重建，且由于被其完全遮挡的区域因缺乏有效观测数据而难以获得高质量修复。未来工作将探索结合时序运动分析(如光流)和弱语义先验信息，以提升对“伪静态”干扰物的识别与处理能力。

参考文献 (References)

[1] 王稚儒, 常远, 鲁鹏, 等. 神经辐射场加速算法综述[J]. 图学学报, 2024, 45(1): 1-13.
WANG Z R, CHANG Y, LU P, et al. A review on neural radiance fields acceleration[J]. Journal of Graphics, 2024, 45(1): 1-13 (in Chinese).

[2] SCHÖNBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 4104-4113.

[3] SEITZ S M, CURLESS B, DIEBEL J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2006: 519-528.

[4] 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.

[5] PINKUS A. Approximation theory of the MLP model in neural

networks[J]. Acta Numerica, 1999, 8: 143-195.

[6] GARBIN S J, KOWALSKI M, JOHNSON M, et al. FastNeRF: high-fidelity neural rendering at 200FPS[C]//2021 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2021: 14326-14335.

[7] MÜLLER T, EVANS A, SCHIED C, et al. Instant neural graphics primitives with a multiresolution hash encoding[EB/OL]. [2025-04-01]. https://doi.org/10.1145/3528223.3530127.

[8] MARTIN-BRUALLA R, RADWAN N, SAJJADI M S M, et al. NeRF in the wild: neural radiance fields for unconstrained photo collections[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 7206-7215.

[9] BOJANOWSKI P, JOULIN A, LOPEZ-PAZ D, et al. Optimizing the latent space of generative networks[EB/OL]. [2025-03-30]. https://dblp.uni-trier.de/db/conf/icml/icml2018.html#BojanowskiJLS18.

[10] CHEN X Y, ZHANG Q, LI X Y, et al. Hallucinated neural radiance fields in the wild[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 12933-12942.

[11] LEE J, KIM I, HEO H, et al. Semantic-aware occlusion filtering neural radiance fields in the wild[EB/OL]. [2025-04-01]. https://arxiv.org/abs/2303.03966.

[12] TANCIK M, CASSER V, YAN X C, et al. Block-NeRF: scalable large scene neural view synthesis[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 8238-8248.

[13] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//2023 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2023: 3992-4003.

[14] 何高湘，朱斌，解博，等. 基于神经辐射场的新视角合成研究进展[J]. 激光与光电子学进展, 2024, 61(12): 1200005.
HE G X, ZHU B, XIE B, et al. Progress in novel view synthesis using neural radiance fields[J]. Laser & Optoelectronics Progress, 2024, 61(12): 1200005 (in Chinese).

[15] KIM I, CHOI M, KIM H J. UP-NeRF: unconstrained pose-prior-free neural radiance fields[EB/OL]. [2025-04-01]. https://arxiv.org/abs/2311.03784.

[16] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[17] MIRZAEI A, AUMENTADO-ARMSTRONG T, DERPANIS K G, et al. SPIn-NeRF: multiview segmentation and perceptual inpainting with neural radiance fields[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2023: 20669-20679.

[18] KERBL B, KOPANAS G, LEIMKUEHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 139.