

基于多视图混合一致性约束的神经隐式表面重建方法

朱文倩¹, 宋丽娟^{1,2,3}, 郭新茹¹, 马子睿^{1,2,3}

(1. 宁夏大学信息工程学院, 宁夏 银川 750021; 2. 宁夏“东数西算”人工智能与信息安全重点实验室, 宁夏 银川 750021;
3. 宁夏大数据与人工智能省部共建协同创新中心, 宁夏 银川 750021)

摘要: 在基于神经隐式表面学习的多视图三维重建过程中, 复杂物体的几何形状和外观表示存在潜在的模糊性。因此, 物体的几何细节信息在纹理稀疏区域、边界区域与较大光滑区域中容易丢失, 难以精确恢复。为解决这个问题, 提出一种基于多视图混合一致性约束的神经隐式表面重建方法。该方法采用多视图立体匹配(MVS)、多视图光度一致性与特征一致性、体渲染技术来优化隐式表面表示, 从而重建具有精细几何细节的复杂物体模型。首先, 提出一个基于 MVS 的稠密点生成模块, 通过 MVS 生成稠密点, 来补充物体表面纹理稀疏区域与边界区域的细节信息, 从而实现物体表面的多视图几何优化。其次, 提出多视图混合一致性约束模块, 通过符号距离函数(SDF)定位零水平集, 利用多视图光度一致性约束来对物体光滑区域进行几何约束, 监督所提取的隐式表面, 并对经过线性插值的 SDF 过零处的表面点应用多视图特征一致性约束, 弥补纹理稀疏区域或结构复杂区域像素匹配的误差, 从而优化物体重建模型。最后, 通过应用体渲染技术, 利用隐式的 SDF 得出高质量的图像渲染, 以实现复杂物体的精确表面重建。实验结果表明, 在 DTU 数据集中, 相比于 Colmap 等方法, 所提方法峰值信噪比(PSNR)提升了 40.3% 以上, 实现了物体表面的精确重建。

关键词: 神经隐式表面重建; 多视图立体匹配; 符号距离函数; 多视图混合一致性; 体渲染

源代码链接: <https://github.com/users/zwq1111/projects>

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0252071

Neural Implicit Surface Reconstruction Method Based on Multi-View Mixed Consistency Constraints

ZHU Wenqian¹, SONG Lijuan^{1,2,3}, GUO Xinru¹, MA Zirui^{1,2,3}

(1. School of Information Engineering, Ningxia University, Yinchuan 750021, Ningxia, China;

2. Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Yinchuan 750021, Ningxia, China;

3. Collaborative Innovation Center for Ningxia Big Data and Artificial Intelligence Co-founded by Ningxia Municipality and Ministry of Education, Yinchuan 750021, Ningxia, China)

【Abstract】 Multi-view 3D reconstruction based on neural implicit surface learning includes inherent ambiguities in representing the geometric shape and appearance of complex objects. Consequently, the fine geometric details of an object are prone to being lost in sparse texture areas, boundaries, and large smooth surfaces, making accurate recovery difficult. To address this issue, this study proposes a novel neural implicit surface reconstruction method based on multi-view mixed consistency constraints. This method uses Multi-View Stereo (MVS), multi-view photometric consistency, feature consistency, and volume rendering techniques to optimize the implicit surface representation, enabling the reconstruction of object models with fine geometric details. First, a dense point generation module based on MVS is proposed to supplement detail information in the sparse texture areas and boundaries of the object surface, achieving multi-view geometric optimization of the object surface. Second, a multi-view mixed consistency constraints module is introduced, which uses the Signed Distance Function (SDF) to locate the zero-level set. It applies multi-view photometric consistency constraints to impose geometric constraints on the smooth regions of the object, supervising the extracted implicit surface. Additionally, multi-view feature consistency constraints are applied to surface points at the zero-crossing of the linearly interpolated SDF, compensating for pixel matching errors in texture-sparse or structurally complex regions, thereby optimizing the object reconstruction model. Finally, volume rendering technology is applied to produce high-quality image renderings from the implicit SDF, enabling precise surface reconstruction of objects. Experimental results show that, compared to methods such as Colmap, the proposed method increases the Peak Signal-to-Noise Ratio (PSNR)

基金项目: 宁夏回族自治区重点研发计划(2023BEG02009); 宁夏回族自治区自然科学基金(2024AAC03062)。

作者简介: 朱文倩(CCF 学生会员), 女, 硕士研究生, 主研方向为计算机视觉; 宋丽娟(通信作者), 副教授、博士; 郭新茹, 硕士研究生; 马子睿, 副教授。

收稿日期: 2025-01-20

修回日期: 2025-04-18

E-mail: slj@nxu.edu.cn

by over 40.3% on the DTU dataset and successfully enables accurate surface reconstruction of the objects.

【Key words】 neural implicit surface reconstruction; Multi-View Stereo (MVS); Signed Distance Function (SDF); multi-view mixed consistency; volume rendering

0 引言

三维重建的目的是从不同视角下的多张图像中提取关于物体几何结构的准确信息,这些信息在医学影像分析、数字文物保护等领域具有重要应用价值。与此同时,多视图三维重建技术也逐步拓展至智能导航系统、智慧城市建设等新兴领域。然而,从校准的多视图图像中进行稀疏纹理区域和较大光滑区域的复杂物体的精确重建,仍然是计算机视觉和图形学领域亟待突破的技术难题。传统方法^[1-2]和基于深度学习的多视图立体匹配(MVS)方法^[3-5]通过匹配一组图像中像素或块的光度或特征对应关系来产生深度图。值得注意的是,传统的多视图三维重建方法涉及深度图预测、全局空间融合和表面提取等多个步骤,较为繁琐,不可避免地在重建结果中出现累积误差和伪影,从而影响最终几何结构的重建精度。

受神经辐射场(NeRF)^[6]这一开创性工作的启发,神经隐式表面重建方法近年发展迅速,成为几何三维重建的重要工具。这些方法通常采用符号距离函数(SDF)^[7]进行表面表示,并利用可微分体渲染技术优化隐式 SDF 场,从而实现高保真重建。文献[8]通过结合神经网络和隐式表面表示,提出了一种高效且鲁棒的三维重建方法 NeuS。在 SDF 的一阶近似中,NeuS 采用了一种具有遮挡感知和无偏性的权重函数,能够从稀疏或不完整的数据中重建复杂的几何形状,并在细节捕捉方面表现出色。在 NeuS 的基础上,文献[9]提出了一种基于体渲染的神经隐式表面重建方法 D-NeuS。该方法通过引入两个额外的损失函数,并结合多视图特征一致性约束,显著提高了物体重建的质量。文献[10]则构建了一种融合立体匹配与特征一致性的重建框架 MVSDf,应用 SDF 场与表面光场来分别表示场景的几何与外观,实现物体的高保真重建。文献[11]提出了一种几何一致性神经隐式表面学习方法 Geo-Neus,通过利用稀疏几何与多视图光度一致性显式地执行多视图几何优化,弥补了体渲染积分与基于点的 SDF 建模之间的差距,实现了在复杂结构和光滑区域中的高质量表面重建。

这些方法应用可微渲染,通过最小化渲染图像与输入真实图像之间的差距来联合优化几何形状和

外观表示,但是在绘制新视图时,由于边界区域、纹理稀疏区域与较大光滑区域的结构恢复较为困难,导致重建结果可能出现伪影或缺失等情况,难以恢复物体的高保真几何细节。

为解决以上问题,本文提出一种基于多视图混合一致性的神经隐式表面重建方法。本文的主要贡献包括:

1)提出了基于 MVS 的稠密点生成模块。在输入阶段利用深度 MVS 方法获取高质量稠密点,完善边界区域与纹理稀疏区域的几何信息,对 SDF 网络进行几何监督,避免重建模型出现伪影或缺失的情况。

2)提出了多视图混合一致性约束模块。使用基于 SDF 的插值来定位 SDF 网络的预测表面,使用块匹配来衡量相邻视图之间的光度一致性,监督预测的表面;使用多视图特征一致性对表面复杂结构细节进行细化,并约束体渲染,以实现更准确的表面重建。

3)应用了经典的体渲染技术,训练神经 SDF 表示并渲染图像,重建出具有完整结构与精细几何细节的高质量三维模型。

1 相关工作

1.1 MVS

多视图三维重建技术作为计算机视觉领域的重要研究方向,目前已经得到了广泛的研究和应用,主要包括从运动中提取结构(SFM)^[12-14]和 MVS 技术^[15]。传统的多视图三维重建利用 SFM 技术从多视角图中提取和匹配相邻视图的特征,并估计相机参数和稀疏三维点。MVS 是一种从重叠图像中恢复密集场景表示的经典方法。根据输出表示,MVS 方法可分为点云^[16]、体积^[17]、深度图^[18-19]3类。基于深度图的方法可以通过预测每个图像的深度信息,并将其融合以得到物体的表面点云,结合不同的表面重建方法来获取物体表面的网格。同时,深度图还能与点云或体积重建进行融合。近年来,深度学习被应用到 MVS 中,文献[20]创新性地提出一种深度 MVS 方法 MVSNet,通过将相邻视图的特征图变形来构建代价体积,并使用三维卷积神经网络对代价体积进行正化。后续研究如文献[21]则提出一种可微 MVS 架构 PatchmatchNet,该架构集

成了迭代的多尺度 PatchMatch 策略,进一步提升了重建性能。

虽然目前使用的方法使整体深度质量有所提高,但在纹理稀疏区域或非朗伯曲面的像素匹配仍然很困难。根据目前的 MVS 研究^[22-23],在点融合和表面重建过程中不可避免地会积累误差,因此,本文提出基于 MVS 的点生成模块生成稠密点来约束物体的几何结构,并对 SDF 网络训练进行监督,更好地改善纹理稀疏区域与非朗伯曲面的重建。

1.2 NeRF

NeRF 是一种用于生成高质量三维重建模型的计算机视觉技术。它利用深度学习技术从多视角图像中提取对象的几何形状和纹理信息,并构建一个连续的三维辐射场,从而实现任意视角和观测距离下的高保真渲染。具体而言,NeRF 以多视角图像和相机参数作为输入,通过一个多层感知机(MLP)模型来表示场景中每个点的颜色和密度。在训练阶段,NeRF 使用图像渲染技术将生成的三维场景渲染为二维图像,并与真实图像进行对比,以优化模型参数,从而提升模型的准确性。

文献[24]提出了一种新方法 NoPe-NeRF,将单目深度整合到无位姿 NeRF 训练中,通过基于深度的表面渲染损失进一步规范相对位姿估计来提高新视角合成与物体重建的质量。文献[25]则提出了一种神经边缘隐式场(NEF),用于表示三维边缘的密度分布,该方法从 NEF 中提取三维边缘点,并使用迭代优化方法重建参数化的三维曲线。MP-NeRF^[26]在处理模糊问题上取得了突破,通过结合多分支融合网络(MBFNet)和基于先验的可学习权重(PLW),增强了对复杂场景细节的捕捉能力,包括纹理和图案信息等。文献[27]则专注于动态三维人体重建场景下 NeRF 训练效率低和体渲染计算复杂度高的问题,提出一种 NeRF 加速方法,引入多分辨率哈希编码作为位置特征编码,提高了 NeRF 对局部细节特征的代表能力,加快了模型收敛。

NeRF 技术在计算机图形学、虚拟现实、增强现实和计算机游戏等领域有着广泛的应用前景。在本文工作中,应用经典的体渲染技术对重建的三维模型进行颜色渲染,获得高质量物体表面。

1.3 神经隐式表面表示与重建

目前,神经隐式表示在三维场景重建中具有良好的发展前景。采用神经隐式函数来表示三维场景的先进方法相继涌现,这些方法可以应用于形状表示^[28]、新视角生成和多视图三维重建。文献[29]提出了一种新的三维重建方法,通过结合卷积神经网络

(CNN)和隐式表示实现从合成数据到真实世界场景的泛化,提高了三维重建的灵活性和准确性。文献[30]针对服装的三维重建技术进行研究,提出基于隐式表达的服装三维重建模型,使用神经网络学习获得的占用函数作为服装三维模型的隐式表达,从而建立三维坐标和模型形状的映射。文献[31]提出基于锥形追踪和网络分解的 NeRF 三维重建方法,使用锥形追踪的方法为每个像素投射一个圆锥体,并将其切割成一系列的圆锥台,沿着该圆锥体进行特征化,通过渲染抗锯齿的圆锥台来降低模糊或伪影效果。

隐式表面重建方法采用表面指示函数,将连续空间坐标映射到占用信息和 SDF,使用行进立方体(MC)算法^[32]来提取任意分辨率下的隐式表面。文献[33]结合表面和体渲染技术提出了一种由粗到精的策略 UNISURF,在由占有场表示的表面周围进行点采样。文献[34]提出仅在物体表面点上渲染光线颜色的策略 IDR,采用人工标记的掩码来保证重建质量。文献[35]提出一种新的模型 VolSDF,通过将体积密度函数定义为应用于 SDF 表示的拉普拉斯累积分布函数来实现更好的几何近似。Neural Warp^[36]通过对 VolSDF 进行扩展,应用光度一致性约束优化几何表面。为了使 SDF 场的几何形状表示更精确, MonoSDF^[37]和 Geo-NeuS 在光度损失的基础上增加了几何损失,减少体渲染过程中可能存在的偏差,保证了物体重建的准确性。

基于上述工作,本文针对复杂物体在较大光滑区域精细节难以精确恢复的问题,提出基于多视图混合一致性的神经隐式表面重建方法,将多视图光度一致性与特征一致性相结合,对 SDF 网络与体渲染进行监督与约束,对复杂物体的结构信息进行细化,提高几何估计的鲁棒性。

2 本文方法

本文所提出的方法,目标是重建由隐式神经网络表示的高保真表面,其整体网络架构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。给定一个物体的具有已知内、外相机参数的多视角图像,通过基于 MVS 的稠密点生成模块生成稠密点,补充纹理稀疏区域与边界区域的几何信息,并与沿着某一视角的一条光线上采样的点共同作为 SDF 网络的输入,提出基于视图方向的 SDF 损失监督网络的训练;使用多视图混合一致性约束模块将基于 MVS 的光度一致性与基于使用多层卷积的 SDF 零水平集的特征一致性相结合,提出混合一致

性损失来约束物体的几何结构;在图像渲染阶段,应用经典的体渲染技术实现图像颜色的无偏渲染,最

终实现复杂物体在纹理稀疏、较大光滑等不同挑战区域的精细重建。

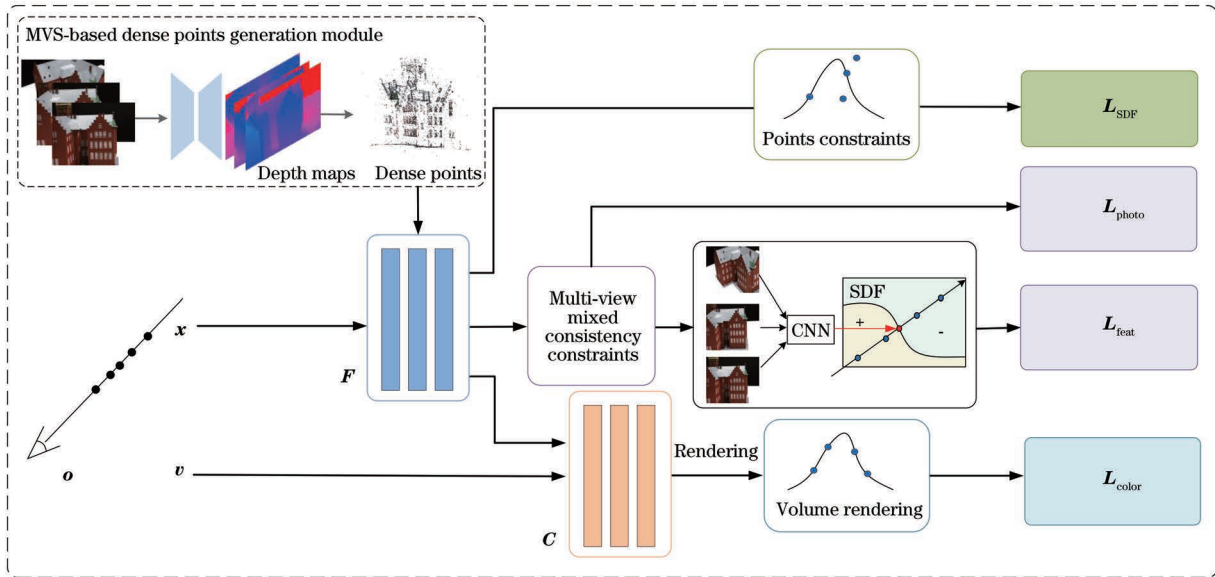


图 1 整体网络结构

Fig.1 Overview of the proposed network architecture

2.1 基于 MVS 的稠密点生成模块

针对纹理稀疏区域与边界区域细节信息难以完整获取的问题,本文提出基于 MVS 的稠密点生成模块,应用深度 MVS 生成稠密点作为整体网络架构的部分输入,以进行网络的几何监督。在以往的数据处理过程中,获取具有已知相机位姿的图像势在必行。利用 SFM 估计这些图像的位置信息并获得重建稀疏的三维点,虽然这些点能够保持一定的精度,但它们不可避免地含有噪声,导致在后续重建复杂对象的几何结构过程中造成误差或伪影。稠密点生成这一部分应用基于 CNN 的深度学习架构,从多视图图像中生成深度图,使用一个三维卷积网络来提取深度视觉图像特征,并对其进行正则化和回归,通过可微单应性变形来构建三维代价体积并经过不断优化来融合不同视角的信息,进行细化以生成最终的三维点位置从而获取重建对象的稠密点,将其作为整体网络架构的一部分输入,以提高纹理稀疏区域与边界区域的重建精度,避免重建模型出现缺失的情况。

在本文中,以一幅参考图像和多幅源图像作

为输入,对输入图像进行深度图推断,获取复杂对象的稠密点,获取流程如图 2 所示。首先,采用 8 层的 2D CNN 来提取 N 幅输入图像的深度特征进行密集匹配,减少输入图像的维度,输出 N 幅 32 通道的特征图。与直接对输入图像进行密集匹配相比,所提取的特征图能够显著提高复杂对象的重建质量。其次,从提取的特征图和提供的相机参数中构建三维代价体积,其中的关键部分是可微的单应性变形操作。变形操作是连接二维特征提取与后续三维代价体积正则化的重要步骤。变形操作过程与经典的平面扫描立体视觉类似,它隐式地编码网络中的相机几何,采用可微的双线性插值从通过二维卷积网络获取的特征图中采样像素。在代价体积正则化步骤中,应用多尺度三维 CNN,使用编码器-解码器结构从一个较大的感受野中聚合邻近信息,并采用沿深度方向的 Softmax 操作进行概率归一化,获得初始深度图。最后,应用深度残差学习网络将初始深度图和调整尺寸后的参考图像拼接为 4 通道输入,依次经过 3 个 32 通道的二维卷积层和一个 1 通道的卷

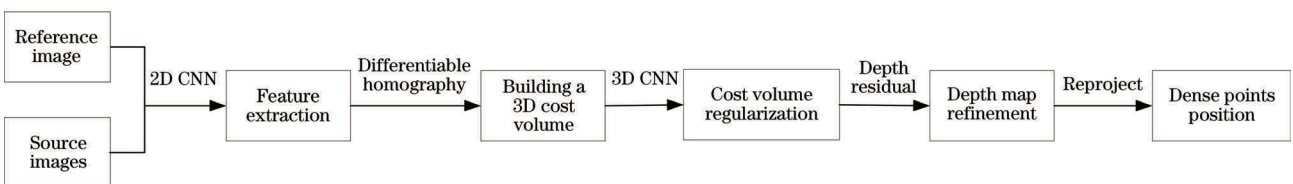


图 2 稠密点获取流程

Fig.2 The acquisition process of dense points

积层学习深度残差,将初始深度图进行回归,并生成精细化的深度图。精细化后的深度图通过按概率加权的每个平面深度值的线性组合来计算,将深度图重投影到三维空间,得到每个视图的稠密点。

由于得到的稠密点可能存在噪声敏感性问题,这种噪声主要来源于低纹理区域的深度估计不准确或遮挡等问题,若直接应用于后续任务,可能会对 SDF 网络训练产生负面影响,导致重建的几何结构出现伪影或缺失,难以恢复物体高保真的几何细节。为确保生成稠密点的质量不会对 SDF 网络训练产生负面影响,采取以下策略:采用深度图融合步骤,

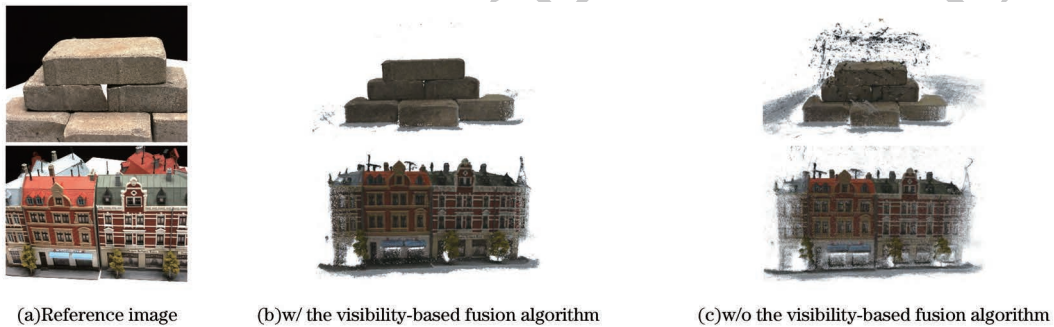


图 3 基于可见性的融合算法降噪效果

Fig. 3 The visibility-based fusion algorithm noise reduction effect

将物体表面 S 用一个隐式场表示,表示方法如下:

$$S = \{x \in \mathbb{R}^3 \mid f(x) = 0\} \quad (1)$$

式中: $f(x)$ 是将点的空间位置映射到 $x \in \mathbb{R}^3$ 的符号距离函数,表示空间点 x 到表面 S 的符号距离, $f(x) = 0$ 则表示空间点 x 在被观测物体的表面上。 $f(x)$ 可以用神经网络 F 来表示,被称为 SDF 网络。它由一个 8 层的 MLP 实现,包含 512 个隐藏单元,中间为跳跃连接,位置编码则被应用于输入位置以捕获高频信息。

当使用相机捕获某种颜色时,这些颜色是沿着光线传输到相机中的光。因此,对于一个特定的相机位置 o 和像素,给出一条由相机发出并穿过一个像素的射线,来自相机光线上的三维点,其中心为 o ,视图方向为 v ,可以表示为:

$$x(t) = o + tv \mid t \geq 0 \quad (2)$$

式中: t 是 x 和 o 之间的距离。

当从视图方向 v 去渲染图像时,依照所提出的网络架构使用 SDF 网络来估计点的 SDF 值并获取点在三维空间中的位置,通过表面上的点来约束物体的整体几何结构,提高物体细节部分重建的准确性。因此,将获取到的点表示出来,提出基于视图方向的 SDF 损失,直接监督 $f(X)$,如式(3)所示:

将来自不同视图的深度图集成到统一的点云表示中,应用基于可见性的融合算法来最小化不同视点之间的深度遮挡;此外,通过确定每个像素的可见视图,并将所有重投影深度的平均值作为最终深度估计,进一步抑制了噪声,有效地降低了生成的稠密点中的噪声对 SDF 网络训练的影响,从而提高了三维重建的精度和鲁棒性。使用基于可见性的融合算法的降噪效果如图 3 所示,其中,“w/”表示使用,“w/o”表示不使用。对照参考图像,应用基于可见性的融合算法生成的点噪声较小,而未应用基于可见性的融合算法生成的点噪声明显,直观地展示了算法的降噪性能。

$$L_{\text{SDF}} = \sum_{x \in X_i} \frac{1}{N} |f(X)| \quad (3)$$

式中: N 为 X 点的个数。用于监督 SDF 网络的损失是不同的,它根据所呈现的视图变化而变化,从而达到引入的 SDF 损失与显色的过程一致的目的。此外,由于具有强纹理的复杂几何结构是获取点的集中分布区域,可以捕获更精细的几何结构。

基于 MVS 生成的稠密点对恢复复杂物体的精细几何结构起到了至关重要的作用。生成的稠密点不仅丰富了物体边界与纹理稀疏区域的几何信息,还补充了物体复杂结构部分的缺少信息,最终实现复杂物体的高保真重建。

2.2 多视图混合一致性约束模块

本文提出多视图混合一致性约束模块来解决目前关于复杂物体重建中结构复杂区域和较大光滑区域细节难以高度恢复的问题。首先,通过获取估计的隐式表面,其几何结构在不同视图之间是一致的。基于这一原理,本文使用 MVS 中的光度一致性约束来对物体光滑区域进行几何约束,监督所提取的隐式表面。其次,使用的多视图特征一致性方法能够匹配特征图之间的单个像素,提高渲染几何形状精细细节的能力,从而获取高

精度物体表面。

2.2.1 多视图光度一致性约束

在使用基于 MVS 的稠密点生成模块的情况下,本文所提出的网络可以捕获边界区域的几何细节,为其提供显式约束,但较大的光滑区域仍然缺乏显式的几何约束。进一步,在具有 MVS 约束的隐式曲面上设计了几何监督。使用表面的隐式表示,用隐式函数的零水平集来提取表面,所估计的表面表示为 SDF 值是 0 的表面。应用一致的采样策略,从相机位置 o 开始,沿着光线的方向 v 进行移动,对应的三维点为 $x(t_i)$,预测的 SDF 值为 $f(x)$,并进一步将 $f(x_i)$ 表示为关于 t 的函数 $f(x_t)$,确定采样点 t_i ,它的 SDF 值符号与下一个采样点 t_{i+1} 不同。它们的 SDF 值满足下式:

$$f(x(t_i))f(x(t_{i+1})) < 0 \quad (4)$$

由于分层抽样策略重视在表面点附近采样,假设 $x(t_i)$ 和 $x(t_{i+1})$ 之间的距离较小,使用可微的线性插值来逼近 SDF 为 0 的表面点,所获得的交点满足:

$$t^* = \frac{f(x(t_i))t_{i+1} - f(x(t_{i+1}))t_i}{f(x(t_i)) - f(x(t_{i+1}))} \quad (5)$$

由于光线与物体表面可能存在一个或多个交点,考虑到遮挡问题,只使用沿光线的第一个交点,以此来保证隐式表面的采样点对于相应的视图均可见。

对于表面的一部分小区域 s , s 在图像上的投影为一个小像素块 q 。 s 对应的块(patch)在不同视图之间应该是几何一致的,遮挡情况除外。类似于传统的 MVS 方法中的 patch,本文使用中心点及其法线来表示 s 。为了方便起见,将 s 在参考图像 I_{ref} 的相机坐标中表示为以下方程:

$$\mathbf{n}^T \mathbf{x} + d = 0 \quad (6)$$

式中: \mathbf{x} 是通过公式计算得到的交点; \mathbf{n}^T 是在 x 处自动微分得到的法向量。

引入一个单应性矩阵 \mathbf{H} 来定位参考图像中某点 p 的像素值。将参考图像 I_{ref} 的像素块 q_i 中的图像点 p 与源图像 I_{source} 的像素块 q_{is} 中的对应点 p' 相关联:

$$\begin{aligned} p &= H_s p' \\ \mathbf{H} &= \\ \mathbf{K}_{source} \left(\mathbf{R}_{source} \mathbf{R}_{ref}^T - \frac{\mathbf{R}_{source} (\mathbf{R}_{source}^T - \mathbf{R}_{ref}^T \mathbf{t}_{ref}) \mathbf{n}^T}{D} \right) \mathbf{K}_{ref}^{-1} \end{aligned} \quad (7)$$

式中: \mathbf{K} 为内参矩阵; \mathbf{R} 为旋转矩阵; \mathbf{t}_{ref} 为平移向量。为了衡量不同视图之间的光度一致性,引入参

考图像和源图像之间的归一化互相关(NCC)值:

$$N_{NCC} = \frac{\text{Cov}(I_{ref}(q_i), I_{source}(q_{is}))}{\sqrt{\text{Var}(I_{ref}(q_i)) \text{Var}(I_{source}(q_{is}))}} \quad (8)$$

式中: Cov 表示协方差; Var 表示方差。将渲染图像作为参考图像,计算其采样块与其在所有源图像上对应块之间的 NCC 值。为了处理遮挡,取每个采样块计算的 NCC 值中最好的 4 个,使用它们来计算相应视图的光度一致性损失:

$$L_{photo} = \frac{\sum_{i=1}^N \sum_{s=1}^4 1 - N_{NCC}(I_{ref}(q_i), I(q_{is}))}{4N} \quad (9)$$

式中: N 为渲染图像上采样像素的个数。

在应用多视图光度一致性损失的情况下,该方法可实现多视图间隐式表面的几何一致性和物体较大光滑区域的高质量重建,降低重建模型缺失的概率。

2.2.2 多视图特征一致性约束

应用线性插值 SDF 为 0 的表面点多视图特征一致性,可以捕获精细的局部几何细节。通常,光度一致性方法侧重于比较 RGB 图像块之间的光度距离,需要表面法线来计算单应性,而多视图特征一致性方法只需匹配特征图之间的单个像素,能够弥补纹理稀疏区域或结构复杂区域像素匹配的误差。因此,在应用多视图光度一致性约束的基础上进一步引入多视图特征一致性约束来对重建物体表面施加多视图几何约束,减少物体重建伪影的生成,实现物体表面的精确重建。

在神经隐式表面上应用多视图几何约束的一个关键步骤是以可微的方式获得表面点。在基于表面渲染的重建中,可微光线追踪常用于寻找相机光线与物体表面的交点。然而,为了优化基于体渲染的表面重建,应用光线追踪寻找表面点因不需要进行彩色渲染导致额外的计算,可能会出现表面点偏向于真实表面的现象。应用非线性插值方法获得 SDF 为 0 的点则需要较高的计算成本,还会导致几何表面和体渲染表面点之间出现偏差。相比之下,应用线性插值的方法更简单直观,且直接利用了体渲染中已经计算的 SDF 值,不需要额外的计算开销,能够很好地与体渲染框架结合。因此,对线性插值 SDF 为 0 的表面点应用多视图特征一致性,能够获得精确的几何细节并保证重建区域的高精度。线性与非线性插值方法对比如表 1 所示。在表中将线性插值与非线性插值方法的实现复杂度、精度、适用场景与重建计算开销作对比,清晰地展示了线性插值的优势。

表 1 线性插值与非线性插值比较

Table 1 Comparison of linear interpolation and nonlinear interpolation

对比项	线性插值	非线性插值
实现复杂度	简单(仅需相邻点 SDF 的线性组合)	复杂(需迭代或高阶计算)
精度	较高	较低
适用场景	体渲染框架、平滑或中等曲率表面	尖锐表面或透明物体
计算开销	较低(25 h)	较高(33 h)

使用相同的分层采样策略,获取 SDF 值为 0 的表面交点,获得交点后,在多个视图中比较该点的特征。在网络中,使用 CNN 从 RGB 图像中提取特征,然后使用多视图特征一致性损失来约束神经隐式表面重建。

假设一个表面点 x 在视图 0 上通过光线追踪得到,将其在视图 0 上的投影及其相邻视图记为 $\{p_j\}_{j=0}^{N_v}$,其中, N_v 为相邻源视图数。由于这些投影在空间中指的是同一个三维点,因此它们的深层图像特征应该是一致的。多视图特征一致性损失定义为:

$$L_{\text{feat}} = \frac{1}{N_c N_v} \sum_{j=1}^{N_v} |F_0(p_0) - F_j(K_j(R_j x(t^*) + t_j))| \quad (10)$$

式中: N_c 为特征通道数; F 为提取的特征图; p_0 为光线投影的像素; $\{K_j, R_j, T_j\}$ 为第 j 个源视图的相机参数。

因此,多视图混合一致性的损失为多视图光度一致性损失和特征一致性损失之和:

$$L_{\text{mixed}} = L_{\text{photo}} + L_{\text{feat}} \quad (11)$$

通过应用多视图混合一致性约束体渲染,可实现物体在较大光滑区域与结构复杂区域的精确重建,获取精细的几何结构。

2.3 基于体渲染的神经隐式表面重建

目前,基于体渲染的神经隐式表面重建被广泛应用于各个场景领域。在本节中,采用一种使用体渲染的基于 SDF 的神经隐式表面重建方法。通过体渲染网络 C 来估计符号距离函数 $f(x)$ 以及颜色场,渲染网络 C 由一个具有 256 个隐藏单元的 4 层 MLP 进行参数化。由提出的 SDF 表示来渲染图像,沿着光线的颜色通过体渲染进行累积:

$$C(o, v) = \int_t^{t_0} T(t) \rho(x(t)) c(x(t), v) dt \quad (12)$$

式中: $C(o, v)$ 是图像中该像素的输出颜色; $c(x(t), v)$ 为沿视图方向 v 的点处的颜色; $\rho(t)$ 为体渲染中的体积密度。在输出颜色 C 与 $f(x)$ 之间建立适当的联系推导出合适的权重函数 $\omega(t)$ 是从输入图像中学习正确 $f(x)$ 表示的关键。因此,为使权重函数 $\omega(t)$ 具有遮挡感知的能力,使用公式 $\omega(t) = T(t) \rho(t)$ 来定义权重函数,其中 $T(t) = \exp(-\int_0^t \rho(\mu) d\mu)$ 表示沿光线的累积透射率。

为了正确描述体积密度,重新定义 $\rho(t)$,令 $\rho(t) = \phi_s(f(x(t)))$,而 $\phi_s(\mu) = se^{-\mu}/(1+e^{-\mu})$ 被称为 S 密度。尽管权重函数具有一定的遮挡感知能力,但由于其在重建表面中引入了固有误差,导致它是存在偏差的,因此,当 t 满足条件 $t^* = \text{argmin}\{t | o + tv = x, t \in (0, +\infty)\}$ 时,权重函数在光线到达物体表面点 $x(t^*)$ 时能够得到最大值并使得 $f(x(t^*)) = 0$,直接使用归一化的 S 密度作为颜色渲染的权重函数,此时构建的权重函数 $\omega(t)$ 是无偏的,满足:

$$\omega(t) = \frac{\phi_s(f(x(t)))}{\int_0^{+\infty} \phi_s(f(x(\mu))) d\mu} \quad (13)$$

因此,定义一个不透明密度 $\sigma(t)$ 与体渲染中的标准体积密度 $\rho(t)$ 相对应,遵循体渲染的基本原理,以新的方式根据上述定义构造既具有遮挡感知能力又无偏的权重函数: $\omega(t) = T(t) \sigma(t)$,其中, $T(t) = \exp(-\int_0^t \sigma(\mu) d\mu)$ 。

将最后渲染的颜色与输入图像进行比较,用于体渲染网络的监督,实现图像的高质量渲染。

2.4 损失函数

网络的整体损失函数由 4 个子损失构成,各个子损失分别作用于网络的不同阶段。用于训练本文网络的整体损失函数被定义为以下各个项的加权:

$$L = \alpha L_{\text{eik}} + \beta L_{\text{SDF}} + \gamma L_{\text{mixed}} + L_{\text{color}} \quad (14)$$

式中: α 、 β 和 γ 是调整损失函数中不同任务的权重,根据模型在训练集和验证集上的重建性能,将其分别设为 0.1、1.0 和 0.5。 L_{eik} 被本文采用作为 SDF 零水平集的监督。在每次训练迭代中监督三维空间中所有点的 SDF 值:

$$L_{\text{eik}} = \frac{1}{N} \sum_{i=1}^N (|f(x)| - 1)^2 \quad (15)$$

L_{SDF} 表示基于视图方向的 SDF 损失,利用基于 MVS 生成的稠密点对 SDF 网络进行直接监督; L_{mixed} 则表示多视图混合一致性损失,以多视图光度

与特征一致性为主,对体积渲染进行约束,以实现更准确的表面重建。 L_{color} 是渲染像素颜色和真实像素颜色之间的差异:

$$L_{\text{color}} = \frac{1}{N} \sum_{i=1}^N |C_i - \hat{C}_i| \quad (16)$$

3 实验结果与分析

3.1 实验设置与评价指标

整体网络模型在单个 NVIDIA Titan RTX 显卡上进行训练,采用 Adam 优化器,学习率设置为 5×10^{-4} ,算法使用 PyTorch 1.11.0 的深度学习框架,约进行 25 h 的 300 000 次迭代。网络训练完成后,可通过 Marching Cubes 从预定义边界框的 SDF 场中提取一个体积大小为 512^3 的网格。

本文使用的 SDF 网络 F 由一个具有 256 个隐藏单元的 8 层 MLP 建模,中间采用跳跃连接。体渲染网络 C 是由一个包含 256 个隐藏单元的 4 层 MLP 组成的。位置编码应用于 6 个频率的空间位置 x 和 4 个频率的观察方向 v 。体渲染采用分层采样策略为每条射线采样点,同时每次采样 512 条射线。在多视图特征一致性中,使用 $N_c = 32$ 个特征通道将每个参考视图与 $N_v = 2$ 个相邻的源视图进行比较。

本文使用倒角距离(CD)和峰值信噪比(PSNR)对算法重建的表面进行数值评估。CD 是一种用于衡量两个点云或形状之间近似距离的度量方法,它在三维重建中是一个重要的评估指标。它通常用于衡量预测的三维模型与真实模型之间的差异,该值越小,说明重建质量越好,与目标更接近。CD 评价指标被表示为:

$$d_{\text{CD}}(S_1, S_2) = \frac{1}{S_1} \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \frac{1}{S_2} \sum_{y \in S_2} \min_{x \in S_1} \|y - x\|_2^2 \quad (17)$$

PSNR 被用于比较真实图像与处理后图像的相似度,以此来评价图像的渲染质量,其数值越高,表明图像渲染的质量越好。PSNR 评价指标被表示为:

$$R_{\text{PSNR}} = 10 \lg \left(\frac{N_{\text{MAX}}}{E_{\text{MSE}}} \right) \quad (18)$$

3.2 数据集

本文采用 DTU 数据集^[38]和 BlendedMVS 数据集^[39]。DTU 数据集是来自丹麦技术大学的一个用于计算机视觉和三维重建研究的开放数据集。它包含了丰富的室内和室外场景的高质量图像,并且提供了由这些图像生成的精确的三维几何信息,如相机位姿和深度信息等。DTU 数据集由机器人使用 RGB 相机在严格控制的环境中拍摄,共包含 124 个不同场景的实拍数据集,每个场景由 49 或 64 张分辨率为 $1\,200 \times 1\,600$ 像素的图像组成,包括了许多具有不同材质、颜色和几何形状的物体,其中包括精细表面和复杂三维结构。本文采用其中的 15 个不同场景数据集作为测试集与其他方法的重建结果作比较。

BlendedMVS 是一个大规模数据集,为基于学习的 MVS 算法提供了足够的训练样本。BlendedMVS 数据集中包含不同类型的图像,涵盖了各种场景,包括城市、建筑、雕塑和小物体。每个场景由分辨率为 768×576 像素的图像采集,视角范围为 $31^\circ \sim 143^\circ$ 。相关实验表明:和其他数据集相比,使用 BlendedMVS 数据集训练的网络模型具有更好的泛化能力。因此,本文使用 BlendedMVS 数据集中的个别场景来展示本文方法的可视化效果。

3.3 对比实验

将本文方法与当前基于神经隐式学习的表面重建方法和传统重建方法进行比较。对于基于神经隐式学习的方法,将本文方法与 IDR、MVSDf、NeuS 和 D-NeuS 进行比较;对于传统重建方法,将本文方法与 Colmap 进行比较。以 CD 与 PSNR 来展示本文所提方法的有效性。本文使用基于隐式表面表示的方法进行多视图重建。DTU 数据集上 CD 的定量结果如表 2 所示,最优结果均加粗标注。与传统重建方法 Colmap 和基于神经隐式学习的表面重建方法 IDR、NeuS、MVSDf 相比,本文方法的重建性能具有较大优势。

表 2 DTU 数据集上 CD 的定量结果

Table 2 Quantitative results of CD on the DTU dataset

Method	Scan 24	Scan 37	Scan 40	Scan 55	Scan 63	Scan 65	Scan 69	Scan 83	Scan 97	Scan 105	Scan 106	Scan 110	Scan 114	Scan 118	Scan 122	mean
Colmap	0.81	2.05	0.73	1.22	1.79	1.58	1.02	3.05	1.40	2.05	1.00	1.32	0.49	0.78	1.17	1.36
IDR	1.63	1.87	0.63	0.48	1.04	0.79	0.77	1.33	1.16	0.76	0.67	0.90	0.42	0.51	0.53	0.90
NeuS	1.00	1.37	0.93	0.43	1.10	0.65	0.57	1.48	1.09	0.83	0.52	1.20	0.35	0.49	0.54	0.84
MVSDf	0.83	1.76	0.88	0.44	1.11	0.90	0.75	1.26	1.02	1.35	0.87	0.84	0.34	0.47	0.46	0.88
Ours	0.75	1.77	1.08	0.38	1.19	1.09	0.51	1.27	0.88	0.75	0.48	0.86	0.31	0.40	0.45	0.87

具体来说, IDR、NeuS、MVSDF 均为基于神经隐式学习的表面重建方法。其中, IDR 需要提供人工设计掩码才能实现物体的重建, 较为繁琐, 相比之下, 本文方法更加便捷, 重建精度更高。此外, MVSDF 仅采用单一的多视图特征一致性, 而本文则创新性地提出了多视图混合一致性约束模块。从实现机制来看, MVSDF 需通过计算密集的射线追踪来定位表面点, 这一过程涉及对几何网络的递归查询, 导致较高的计算开销; 相比之下, 本文所使用的多视图特征一致性只需通过局部可微的线性插值从采样点的 SDF 值中确定第一个过零点, 且由于采样点的 SDF 值已经用于体渲染计算从而避免了额外的计算负担, 因此本文

所提出的多视图特征一致性约束在计算效率与实现上均优于 MVSDF 所采用的方案。

图 4 定性比较了本文方法与其他方法在 DTU 数据集上的表面重建效果。Colmap 重建模型表面较为嘈杂, 出现表面细节部分缺失情况, 特别是在较大光滑区域(Scan110 和 Scan114)。而 NeuS 的表面大部分区域重建较好, 但在复杂纹理或细小区域(Scan118)的表面重建较差, 不能良好地处理物体复杂纹理细节部分。相比之下, 本文方法更适用于这些具有挑战性的情况, 能够避免几何细节的缺失与结构复杂区域不能完全恢复的问题, 以更高的精度恢复了物体表面的几何细节。

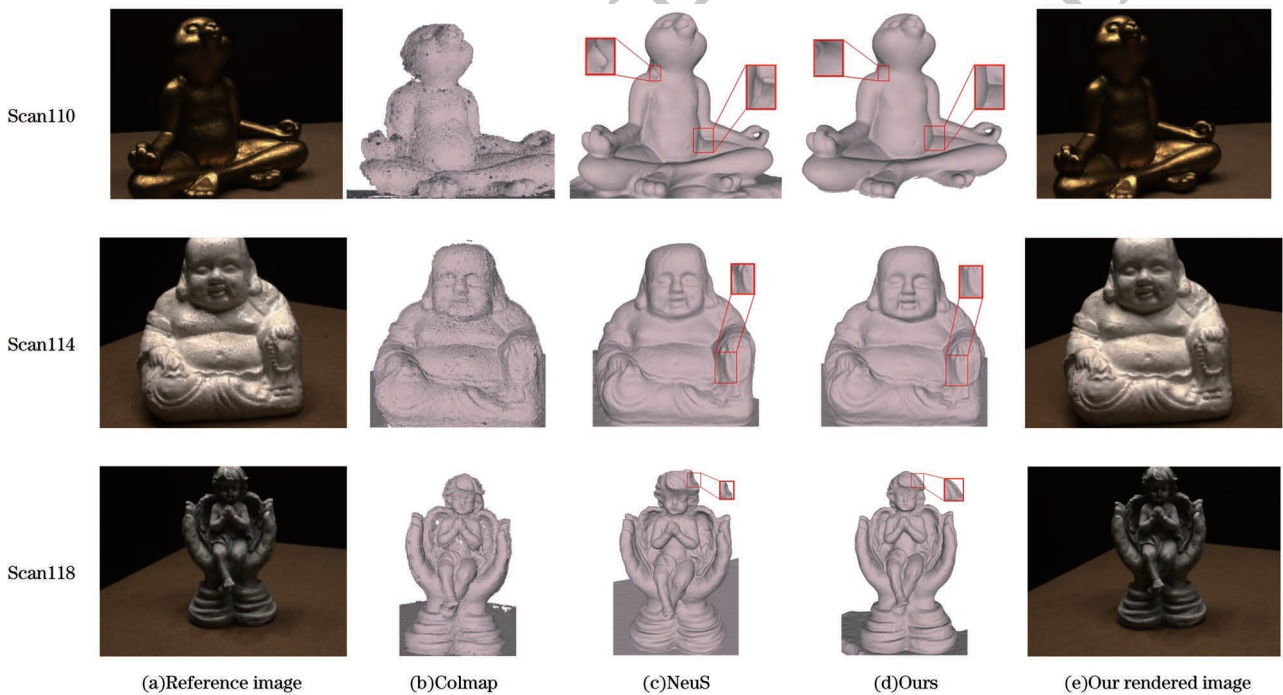


图 4 DTU 数据集上表面重建比较

Fig. 4 Surface reconstruction comparison on the DTU dataset

除表面几何外, 本文方法也实现了真实图像的渲染。DTU 数据集上 PSNR 的定量结果如表 3 所示, 表中定量地评估了不同场景的渲染结果, 最优结果均加粗标注。由表 3 中的结果可以得出, 本文所

提方法的 PSNR 值在多个 Scan 中均高于其他方法, 对照图 4 中的图像渲染结果, 不同场景在物体的结构与颜色恢复上均接近于原参考图像, 实现了场景的高效渲染。

表 3 DTU 数据集上 PSNR 的定量结果

Table 3 Quantitative results of PSNR on the DTU dataset

单位: dB

Method	Scan 24	Scan 37	Scan 40	Scan 55	Scan 63	Scan 65	Scan 69	Scan 83	Scan 97	Scan 105	Scan 106	Scan 110	Scan 114	Scan 118	Scan 122	mean
Colmap	18.44	14.37	19.24	18.27	19.92	13.80	21.23	22.67	18.19	20.43	20.73	17.93	19.08	22.05	22.04	19.23
IDR	23.29	21.36	24.39	22.96	23.22	23.94	20.34	21.87	22.95	22.71	22.81	21.26	25.35	23.54	27.98	23.20
NeuS	28.20	27.10	28.13	28.80	32.02	33.75	30.96	34.47	29.57	32.38	35.07	32.74	31.69	36.97	37.07	31.97
D-NeuS	28.98	27.58	28.40	28.87	33.71	33.94	30.94	34.08	30.75	33.73	34.84	32.41	31.42	36.97	37.17	32.22
Ours	30.40	28.02	28.94	30.37	32.05	32.38	31.68	32.24	29.12	32.45	35.73	33.24	31.98	37.16	37.77	32.25

由表 3 结果得出, 与传统表面重建方法 Colmap 相比, 本文方法的重建性能明显优于 Colmap, 达到

了较大提升; 与基于神经隐式学习的表面重建方法 IDR、NeuS、D-NeuS 相比, 本文方法也表现较好, 在

个别场景中达到最高性能,并实现了最高的平均 PSNR 值。相较于 D-NeuS,本文方法在应用多视图光度一致性的基础上使用多视图特征一致性,能够较好地弥补重建物体在较大光滑区域上的缺乏几何约束的问题。因此,本文方法在整体重建模型效果上具体表现更好。

BlendedMVS 数据集上的定性结果如图 5 所示。本文方法对物体较大光滑区域、边界区域与

结构复杂区域等一些具有挑战性的表面具有鲁棒性,例如室内光照环境下 Sculpture 中的鼻子部分与头部右侧部分,充分体现了多视图混合一致性模块中多视图光度与特征一致性的优势;室外场景 Stone 中的边界与球体等部分能够有效应对光照变化和遮挡问题,准确恢复物体的复杂几何结构。相比之下,Colmap 则难以恢复这些表面的几何细节。

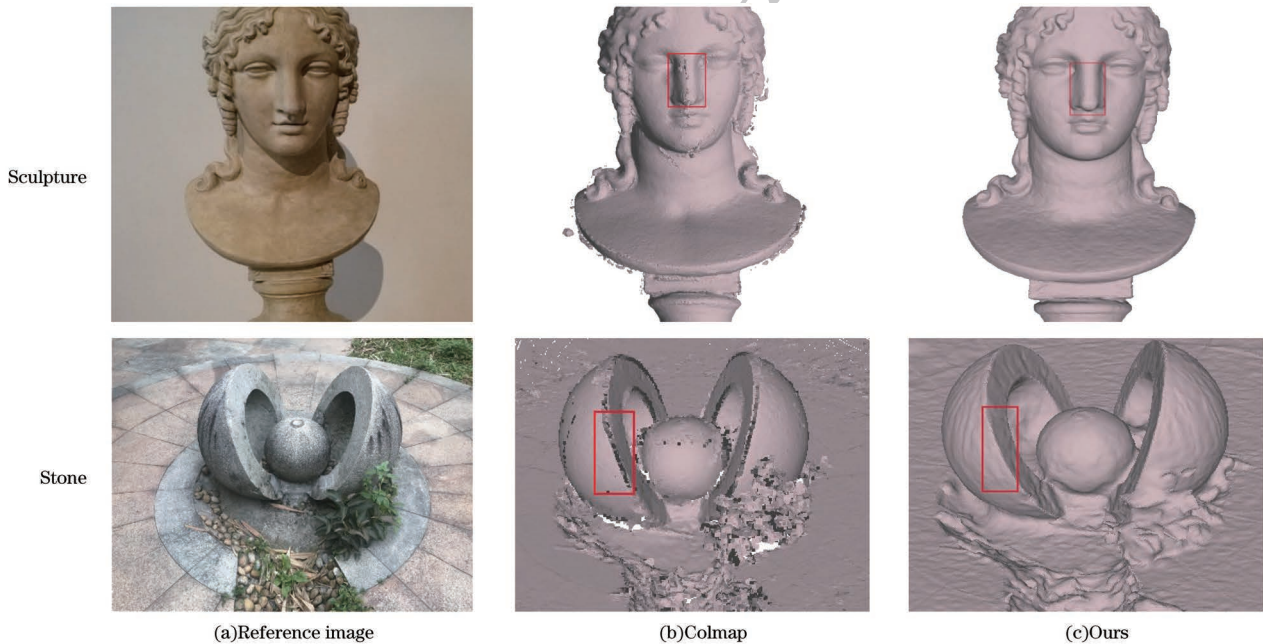


图 5 Blended MVS 数据集上表面重建比较

Fig.5 Surface reconstruction comparison on the BlendedMVS dataset

为了更好地评估本文方法的重建效率,对其在推理阶段的计算开销做对比实验。推理阶段的计算开销如表 4 所示。将本文方法的训练时间与 IDR、NeuS、MVSDF 与 D-NeuS 所花费的训练时间做对比,可以从中看出,各个方法训练时间较为接近,本文方法消耗的训练时间处于中间位置。因此,虽然本文方法在重建三维模型上具有良好的性能,但与其他方法相比仍有提升空间,还需不断地创新优化,在重建效率上有所提高。

表 4 计算开销对比

Table 4 Comparison of calculation overhead 单位:h

Method	Training time(NVIDIA Titan RTX)
IDR	23
NeuS	26
MVSDF	21
D-NeuS	24
Ours	25

3.4 消融实验

通过在 DTU 数据集上的消融实验研究来评估本文方法的不同组成部分。具体来说,本文方法使

用 NeuS 作为构建基线,并逐步合并方法中提出的每一个损失,采取控制变量原则分别对基于 MVS 的稠密点生成模块与多视图混合一致性约束模块两部分进行消融实验。为更好地评估多视图混合一致性约束模块中多视图光度一致性与特征一致性两者的性能,分别对其进行消融实验。在 DTU 数据集内的 Scan24 上做消融实验,定性结果如图 6 所示,将其表面重建效果可视化,图中挑战区域用红色边框标注。与参考图像作对比,本文方法重建模型较为完整,并无缺失,几何结构恢复较好;与只有基于 MVS 的稠密点生成模块相比,模型在物体光滑区域与细节部分未出现偏差;与只有多视图光度一致性或特征一致性相比,在物体较大光滑区域与复杂结构区域表现较好,并未出现伪影或凹凸不平的状况;与只有多视图混合一致性约束模块的重建效果相比,本文方法在左下角光滑区域部分未出现缺失状况,在物体复杂结构区域重建效果较好。因此,本文将基于 MVS 的稠密点生成模块与多视图混合一致性约束模块两者结合应用,吸收各自的优点,提高了方法在不同挑战区域上的重建精度。

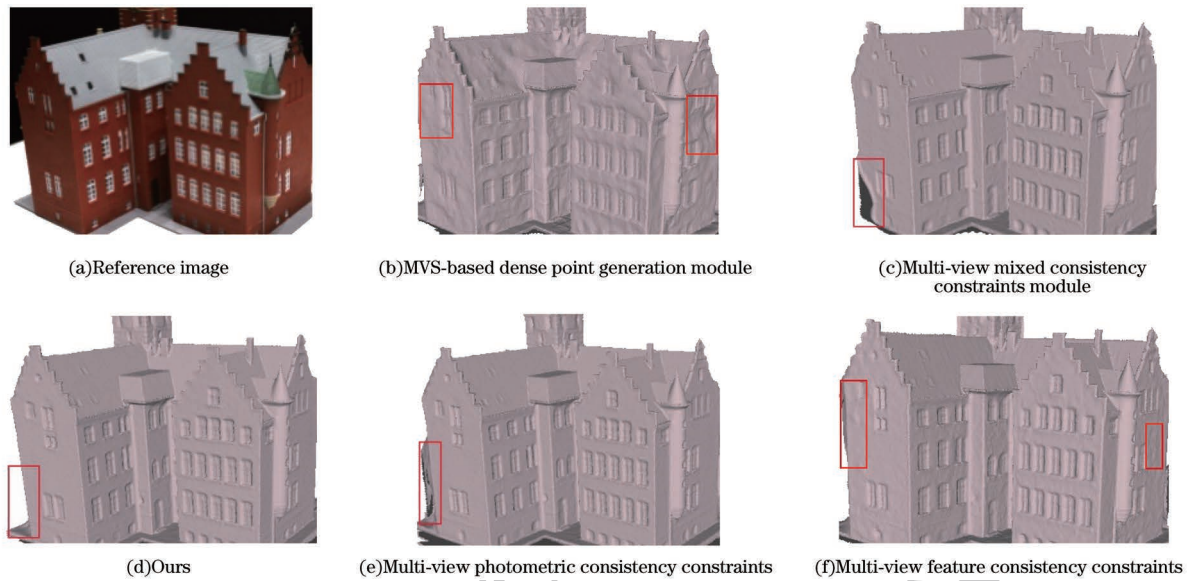


图 6 消融研究定性结果

Fig. 6 Qualitative results of ablation study

消融实验研究的 PSNR 定量结果对比如表 5 所示。本文方法中提出的基于 MVS 的稠密点生成模块和多视图混合一致性约束模块这两项工作均有益于改进几何形状,而在单独使用基于 MVS 的稠

密点生成模块或多视图混合一致性约束模块的情况下,其 PSNR 评估结果均低于两者结合使用的情况。因此,本文方法将两个模块相结合能够更好地提高物体表面重建的效果。

表 5 消融研究定量评估结果对比

Table 5 Comparison of quantitative evaluation results of ablation study

单位: dB

Method	L_{color}	L_{SDF}	L_{photo}	L_{feat}	PSNR
Baseline	✓				28.20
MVS-based dense point generation module	✓	✓			28.91
Multi-view photometric consistency constraints	✓	✓	✓		29.47
Multi-view feature consistency constraints	✓	✓		✓	29.55
Multi-view mixed consistency constraints module	✓		✓	✓	29.64
Ours	✓	✓	✓	✓	30.40

由表 5 消融研究定量评估结果所得, SDF 网络上的三维监督非常稀疏, 本文方法增加基于 MVS 的稠密点生成模块后获取稠密点以增强网络的三维监督, 提高了整体几何结构的约束能力。此外, 通过应用多视图光度一致性约束, 本文方法可以弥补较大光滑区域重建的不足; 通过应用多视图特征一致性约束, 本文方法可以更好地恢复复杂结构区域的细节; 通过应用所提出的多视图混合一致性约束模块, 本文方法实现了更加全面的几何重建。本文方法通过结合基于 MVS 的稠密点生成与多视图混合一致性约束两者的优点, 实现了更加精确的几何结构恢复, 达到了重建性能的提升。

4 结束语

针对物体纹理稀疏、边界、较大光滑区域与复杂结构区域的几何细节难以恢复这一问题, 本文提出

一种基于多视图混合一致性约束的神经隐式表面重建方法。本文方法对复杂物体精细几何细节难以恢复的原因进行了分析, 由于纹理稀疏与边界区域的几何信息不全导致重建物体缺失, 因此提出基于 MVS 的稠密点生成模块生成稠密点作为网络部分输入, 对 SDF 网络进行监督, 约束物体的几何结构并补充纹理稀疏区域与整体轮廓的细节信息。此外, 本文还提出多视图混合一致性约束模块, 使用多视图光度一致性约束对隐式表面进行监督, 优化物体光滑区域的重建; 并将多视图特征一致性约束应用到经线性插值的 SDF 零交叉点处的表面点, 通过这种方式恢复物体的几何细节, 对物体复杂结构进行精确的表面重建。最后, 本文应用体渲染技术对重建模型进行图像渲染, 实现物体的高质量表面重建。在 DTU 和 BlendedMVS 数据集上的实验结果表明, 本文提出的方法能够重建出具有精细几何细

节的高保真物体表面,并在定量与定性结果分析中都较大程度上优于现有的一些方法。值得注意的是,虽然本文提出的方法能够使物体重建质量得到一定的提高,但其效率仍然有限,不能够实现快速的重建过程。因此,在未来的研究中,将更加专注于模型性能的优化,进一步提高模型重建的效率。

参考文献

- [1] KAZHDAN M, HOPPE H. Screened Poisson surface reconstruction[J]. *ACM Transactions on Graphics*, 2013, 32(3): 1-13.
- [2] XU Q S, TAO W B. Multi-scale geometric consistency guided multi-view stereo[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE Press, 2019: 5478-5487.
- [3] CAO C J, REN X L, FU Y W. MVFormer: learning robust image representations via transformers and temperature-based depth for multi-view stereo[EB/OL]. [2025-01-02]. <https://arxiv.org/pdf/2208.02541>.
- [4] GOESELE M, CURLESS B, SEITZ S M. Multi-view stereo revisited[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2006: 2402-2409.
- [5] FURUKAWA Y, HERNÁNDEZ C. Multi-view stereo: a tutorial[J]. *Foundations and Trends® in Computer Graphics and Vision*, 2015, 9(1/2): 1-148.
- [6] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: representing scenes as neural radiance fields for view synthesis[C]//*Proceedings of European Conference on Computer Vision*. Berlin, Germany: Springer, 2020: 405-421.
- [7] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: learning continuous signed distance functions for shape representation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE Press, 2019: 165-174.
- [8] WANG P, LIU L J, LIU Y, et al. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction[EB/OL]. [2025-01-02]. <https://arxiv.org/pdf/2106.10689>.
- [9] CHEN D C, ZHANG P, FELDMANN I, et al. Recovering fine details for neural implicit surface reconstruction[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE Press, 2023: 4319-4328.
- [10] ZHANG J Y, YAO Y, QUAN L. Learning signed distance field for multi-view surface reconstruction[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE Press, 2021: 6505-6514.
- [11] FU Q, XU Q, ONG Y S, et al. Geo-Neus: geometry-consistent neural implicit surfaces learning for multi-view reconstruction [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 3403-3416.
- [12] 蒲建鑫, 宋方伟, 冷齐齐. 基于 SFM 算法的三维重建关键技术研究[J]. *电子技术*, 2021, 50(6): 36-37.
PU J X, SONG F W, LENG Q Q. Study on key technologies of 3D reconstruction based on SFM algorithm[J]. *Electronic Technology*, 2021, 50(6): 36-37. (in Chinese)
- [13] HUANG H, WU Y L, ZHOU J S, et al. NeuSurf: on-surface priors for neural surface reconstruction from sparse input views[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(3): 2312-2320.
- [14] WANG W, GAO F, SHEN Y. Res-Neus: deep residuals and neural implicit surface learning for multi-view reconstruction[J]. *Sensors (Basel)*, 2024, 24(3): 881.
- [15] SEITZ S M, CURLESS B, DIEBEL J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York, USA: IEEE Press, 2006: 519-528.
- [16] JIANG Y Y, SONG L J. 3D surface reconstruction based on dynamic graph convolutional occupancy network [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2023, 37(14): 2354022.
- [17] JI M Q, GALL J, ZHENG H T, et al. SurfaceNet: an end-to-end 3D neural network for multiview stereopsis[C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy: IEEE Press, 2017: 2326-2334.
- [18] CAMPBELL N D F, VOGIATZIS G, HERNÁNDEZ C, et al. Using multiple hypotheses to improve depth-maps for multi-view stereo [C]//*Proceedings of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2008: 766-779.
- [19] TOLA E, STRECHA C, FUA P. Efficient large-scale multi-view stereo for ultra high-resolution image sets[J]. *Machine Vision and Applications*, 2012, 23(5): 903-920.
- [20] YAO Y, LUO Z X, LI S W, et al. MVNet: depth inference for unstructured multi-view stereo[C]//*Proceedings of European Conference on Computer Vision*. Berlin, Germany: Springer, 2018: 767-783.
- [21] WANG F, GALLIANI S, VOGEL C, et al. PatchmatchNet: learned multi-view patchmatch stereo[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE Press, 2021: 14189-14198.
- [22] XU Q G, XU Z X, PHILIP J, et al. Point-NeRF: point-based neural radiance fields[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE Press, 2022: 5428-5438.
- [23] 王思成, 江浩, 陈晓. 具有跨尺度 Transformer 的高效多视图立体网络[J]. *计算机工程*, 2024, 50(11): 266-275.
WANG S C, JIANG H, CHEN X. Efficient multi-view stereo network with cross-scale Transformer[J]. *Computer Engineering*, 2024, 50(11): 266-275. (in Chinese)
- [24] BIAN W J, WANG Z R, LI K J, et al. NoPe-NeRF: optimising neural radiance field with No pose prior[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE Press, 2023: 4160-4169.
- [25] YE Y F, YI R J, GAO Z R, et al. NEF: neural edge fields for 3D parametric curve reconstruction from multi-view images[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE Press, 2023: 8486-8495.
- [26] WANG X H, YIN Z Y, ZHANG F Q, et al. MP-NeRF: More refined deblurred neural radiance field for 3D reconstruction of blurred images [J]. *Knowledge-Based Systems*, 2024, 290: 111571.
- [27] 肖祎龙, 邓伊琴, 陈志刚. 面向动态三维人体重建的神经辐射场加速方法[J]. *计算机工程*, 2025, 51(8): 95-106.
XIAO Y L, DENG Y Q, CHEN Z G. Acceleration approach for neural radiance field in dynamic 3D human reconstruction [J]. *Computer Engineering*, 2025, 51(8): 95-106. (in Chinese)
- [28] ZHOU J S, WEN X, MA B R, et al. 3D-OAE: occlusion auto-encoders for self-supervised learning on point clouds[EB/OL]. [2025-01-02]. <https://arxiv.org/pdf/2203.14084>.

- [29] PENG S, NIEMEYER M, MESCHEDER L, et al. Convolutional occupancy network [C] // Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 523-540.
- [30] 费煜哲, 蔡欣, 赵鸣博, 等. 基于隐式表达的服装三维重建 [J]. 计算机工程, 2024, 50(5): 220-228.
FEI Y Z, CAI X, ZHAO M B, et al. 3D reconstruction of clothing based on implicit expression [J]. Computer Engineering, 2024, 50(5): 220-228. (in Chinese)
- [31] 景维鹏, 王源锋, 李超. 基于锥形追踪和网络分解的 NeRF 三维重建方法 [J]. 计算机工程, 2024, 50(10): 334-341.
JING W P, WANG Y F, LI C. NeRF 3D reconstruction method based on cone tracking and network decomposition [J]. Computer Engineering, 2024, 50(10): 334-341. (in Chinese)
- [32] LORENSEN W E, CLINE H E. Marching cubes: a high resolution 3D surface construction algorithm [J]. ACM SIGGRAPH Computer Graphics, 1987, 21(4): 163-169.
- [33] OECHSLE M, PENG S Y, GEIGER A. UNISURF: unifying neural implicit surfaces and radiance fields for multi-view reconstruction [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE Press, 2021: 5569-5579.
- [34] YARIV L, KASTEN Y, MORAN D, et al. Multiview neural surface reconstruction by disentangling geometry and appearance [J]. Advances in Neural Information Processing Systems, 2020, 33: 2492-2502.
- [35] YARIV L, GU J, KASTEN Y, et al. Volume rendering of neural implicit surfaces [J]. Advances in Neural Information Processing Systems, 2021, 34: 4805-4815.
- [36] DARMON F, BASCLE B, DEVAUX J C, et al. Improving neural implicit surfaces geometry with patch warping [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE Press, 2022: 6250-6259.
- [37] YU Z, PENG S, NIEMEYER M, et al. MonoSDF: exploring monocular geometric cues for neural implicit surface reconstruction [J]. Advances in Neural Information Processing Systems, 2022, 35: 25018-25032.
- [38] JENSEN R, DAHL A, VOGIATZIS G, et al. Large scale multi-view stereopsis evaluation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE Press, 2014: 406-413.
- [39] YAO Y, LUO Z X, LI S W, et al. BlendedMVS: a large-scale dataset for generalized multi-view stereo networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE Press, 2020: 1787-1796.

文字编辑 金胡考
栏目编辑 宋 圆