

面向机器人手术模拟器的神经辐射场软组织动态三维重建

陈琪^{1,2}, 秦芝宝^{1,2}, 蔡晓誉^{1,2}, 李世杰^{1,2}, 王梓俊^{1,2}, 石俊生^{1,2*}, 邵永航^{1,2*}

¹云南师范大学物理与电子信息学院, 云南 昆明 650500;

²云南省光电信息技术重点实验室, 云南 昆明 650500

摘要 构建了一种基于自监督的框架,该框架从单目立体内窥镜视频中提取多视图图像,利用图像中的底层三维(3D)信息构建对象的几何约束,实现软组织结构的准确重建。基于分割任意场景模型对内窥镜下的动态手术器械、静态腹腔场景及可形变软组织结构进行分割解耦。该框架利用简单的神经网络多层感知机来表示动态神经辐射场(NeRF)中运动手术器械和形变软组织结构,基于偏斜熵损失对手术场景中的手术器械、腔体场景和软组织结构进行正确分离。在通过使用单目立体内窥镜捕获机器人手术模拟器场景的数据集上,将所提方法的结果与其他方法进行定量定性比较。结果表明本文方法在处理腹腔体场景、软组织结构重建、手术器械的分割解耦,以及来自多视点的3D信息和运动对象的图像分割等方面显著优于当前的方法。

关键词 视觉光学; 神经辐射场; 软组织三维重建; 分割任意场景模型; 分割解耦

中图分类号 O436

文献标志码 A

DOI: 10.3788/AOS231537

1 引言

手术机器人领域随着持续的研究、开发和商业化而迅速发展并日益成熟,人们对手术机器人平台越来越感兴趣,机器人辅助内窥镜场景下的手术可以提高外科医生的表现,使医疗手术更加智能化,有利于更好地理解手术场景的解剖结构。与传统的开放性手术相比,机器人辅助的内窥镜下手术扩大了手术视界并增强了器械的灵巧性。更重要的是,它为人工智能在手术中的应用开辟了新的道路。从单目立体内窥镜视频中重建机器人手术模拟场景是机器视觉和增强现实技术中的一项重要且困难的任務。它通过提供全面的多模态解剖数据和三维(3D)重建环境的智能交互集成,促进改善术前手术可视化,特别有利于机器人手术模拟训练系统^[1]。目前的手术导航系统通常利用分割的术前医学影像如计算机断层扫描/磁共振成像(CT/MRI)提供术中导航。具体来说,基于术前医学影像数据的3D信息和可视化手术中患者的解剖信息,通过立体配准法与当前的机器人手术模拟场景进行匹配。然而,手术中软组织变形和器械移动导致手术场景不断变化、配准不准确,这严重影响手术场景重建的质量与完整性,从而导致手术的安全性下降并限制了现有导航系统在可形变手术场景中的应用。

先前的工作通过计算立体方法在内窥镜场景中实现光学3D重建^[2-3],其中立体视觉是目前较受欢迎的方法。给定预处理后的图像对,通过左右两个图像的像素或使用尺度不变特征变换(SIFT)算法进行图像配准。根据立体配准值得到图像之间的差异,通过计算深度恢复3D点云,实现场景的重建。但在有手术器械遮挡的情况下,寻找左右图像中点之间对应关系(立体匹配)的过程可能会导致错误,特别是当相机视点之间的差异很小或者场景中的对象具有相似的纹理或颜色时,获取的深度信息不准确,无法达到预期的结果。近几年,已有研究重建内窥镜下手术场景时需要使用已知的几何形状作为监督信息,由于光照对角度的变化并不真实,通常假设物体表面是漫反射的,因此物体的3D几何形状的真实值在实际情况下很难获取^[4-5]。在此背景下,Mildenhall等^[6]提出了新的物体表征方法,即神经辐射场(NeRF),使用输入图像作为监督信号,利用图像几何约束条件得到相机参数(相机内外参)及底层3D信息,通过追踪从相机中心到图像像素的光线并沿着光线方向使用由粗到细的策略进行非均匀空间五维(5D)[3D位置坐标 (x, y, z) 和2D视角观察方向坐标 (θ, ϕ)]坐标采样。NeRF采用多层感知机(MLP)将场景中的每一个点的3D坐标 (x, y, z) 和视

收稿日期: 2023-09-08; 修回日期: 2023-12-28; 录用日期: 2024-01-10; 网络首发日期: 2024-02-20

基金项目: 国家自然科学基金(62365017, 62062070, 62005235)、云南省优秀青年基金(202301AW070001)

通信作者: *taiyonghang@126.com

角方向坐标 (θ, ϕ) 作为输入, 产生相应的 RGB 值和体积密度。在整个训练阶段, 利用体渲染技术将所得的 RGB 值和体积密度聚合成 2D 图像。鉴于此过程的可微性, 使用梯度下降法通过最小化观察图像和相应的从辐射场表示中渲染的视图之间的误差, 对该模型进行优化^[7]。渲染期间, 通过光度一致性损失, 对渲染出来的图像与真实图像的均方误差 (MSE) 损失进行进一步优化。此外, NeRF 使用位置编码, 将每个输入的 5D 坐标映射到更高维的空间, 从而能够有效优化 NeRF, 以高频信息描绘复杂的场景细节。

相比之下, 视频分割提供了快速实时提取、准确解剖细节的能力, 这对于持续更新术中导航来说非常重要。研究显示, 将从视频中获得的语义见解与 3D 几何相融合, 可以增强以动态变化为特征的环境中的感知, 例如室内环境^[8]。简而言之, 视频语义分割与 3D 几何相结合, 可以显著提高模型的态势感知和理解能力, 特别是在条件不断变化的环境中, 例如自动驾驶场景、血管内血液流动场景^[9]。从单目立体内窥镜拍摄的视频中重建具有颜色信息的生物组织及遮挡器械的复杂手术场景, 该手术场景通常涉及有纹理的软组织表面和不断变形的组织, 对其使用 2D 先验知识和修复技术则缺乏 3D 理解, 导致得到次优结果。通过引用一种有条不紊的框架 Anything-3D^[10], 巧妙地结合了一系列视觉语言模型^[11]和 Segment Anything 对象分割模型^[12], 将引用的一系列技术集成于一体即分割任意场景模型 (SASM)。SASM 对基本对象概念有广泛的理解, 能够为不同图像或视频中的各种对象创建蒙版, 此功能可以扩展到不属于其训练数据的对象和图像类别中。SASM 的多功能性使其可以满足各种应用程序的需求, 并且可以轻松部署在新的图像领域中, 而无需额外的训练。SASM 能将对象信息提升到 3D, 为多视图条件的 3D 重建任务提供可靠且通用的系统, 以实现重建解耦的 3D 场景表示, 该场景表示允许以自由多视角图和时变方式分别合成动态和静态对象。

本文从单目立体内窥镜拍摄的具有可形变软组织的视频中分割机器人手术模拟场景并对生物软组织结构进行 3D 动态重建, 该方法可以较容易地扩展到多视点视频中。将高度可分割的模型和新兴的神经渲染框架应用于机器人手术模拟场景重建中, 通过对视频中的时变物体和非时变物体利用 SASM 进行分割建模, 可以去除手术场景中特定的动态遮挡物, 并从新的视图中解耦合成干净的背景。综上所述, 本文总结如下: 1) 为适应软组织的几何外观和可形变表示, 利用 MLP 来表示 NeRF 中的机器人手术模拟场景; 2) 构建了基于 NeRF 和 SASM 的混合表示, 来表征内窥镜下机器人手术模拟场景分割; 3) 提出了一个

自监督的方法, 用于解耦动态场景, 并以新的损失促进场景正确分离。为了验证, 使用单目立体内窥镜从不同角度和距离捕获机器人手术模拟场景, 从而建立数据集。在该数据集上进行定量和定性的分析, 结果表明: 与大多数现有方法相比, 本文方法在合成真实机器人手术模拟场景、重建高保真生物软组织结构颜色纹理等细节信息和场景分割方面取得了更好的效果与性能。

2 基于 NeRF 的机器人手术模拟的软组织 3D 动态重建方法

2.1 利用光场复现软组织的颜色信息

医学影像数据^[13-14] (CT、X-ray、MRI 图像 DICOM 格式等) 具有人体内部解剖结构 (例如骨骼、器官和组织) 信息的灰度值图像, 缺失 RGB 信息, 无法用于生物组织纹理、质感和颜色的高保真 3D 结构复现。本文使用内窥镜, 从不同视角和距离捕获手术场景中生物组织的 RGB 图片, 结合经典的体渲染恢复生物组织结构的颜色、纹理等细节信息。人类视觉系统中色觉是通过 NeRF 中的辐射得来的, 人眼能够接收到光, 光从本质上是电磁辐射或振荡的电磁场, 具有特定的属性, 包括波长和频率, 光的视觉感知或颜色与光频率有内在的联系, 其中光的颜色是由其波长决定的^[15-16]。物体产生不同颜色的主要原因是物体表面的发射特性和光谱特性, 可以推断, 建模辐射光即是建模对应的颜色, 本文将从光学层面通过以下步骤推导得到复现软组织的颜色信息。

光线从眼睛射出, 进入空间并沿着直线方向穿过粒子, 通过计算每根光线从最开始发射到最终打到成像平面上的辐射强度, 可以渲染出投影图像, 这个过程就是体渲染。如图 1 所示, 体渲染把光子与粒子发生作用的过程进一步细化为 3 种类型: 吸收、放射和散射。出射光 L_o 与入射光 L_i 之间的变化量可以表示为这 3 个过程的叠加:

$$L_o - L_i = dL(x, \omega) = R + S - A, \quad (1)$$

式中: $L(x, \omega)$ 为光线, x 为光线上的某个位置点, ω 代表光线发射方向; A 为吸收量; R 为放射量; S 为散射量。

在源自采集内窥镜场景的光学路径的背景下^[17], 图 1 所描绘的区域具有立方体结构 (其特征表示是侧面积为 E , 底面边长为 Δl)。在立方体中, 组成的粒子是半径为 r 的球体, 单个粒子的投影面积 $S_1 = \pi r^2$ (代表阻挡光的面积)。当立方体足够薄 (薄到与粒子一样的厚度, 粒子平铺在立方体一个横截面上) 时, 设 Δl 为立方体的厚度, 立方体的体积为 $E\Delta l$, 粒子数为 $\rho E\Delta l$, 其中, 粒子密度 ρ 为一个关于立方体的底面边长 l 的函数, 每个区域的密度都是不同的, l 为一个关于时间 t 的函数, 在光线传播过程中由于各种光学变化, 光线逐渐

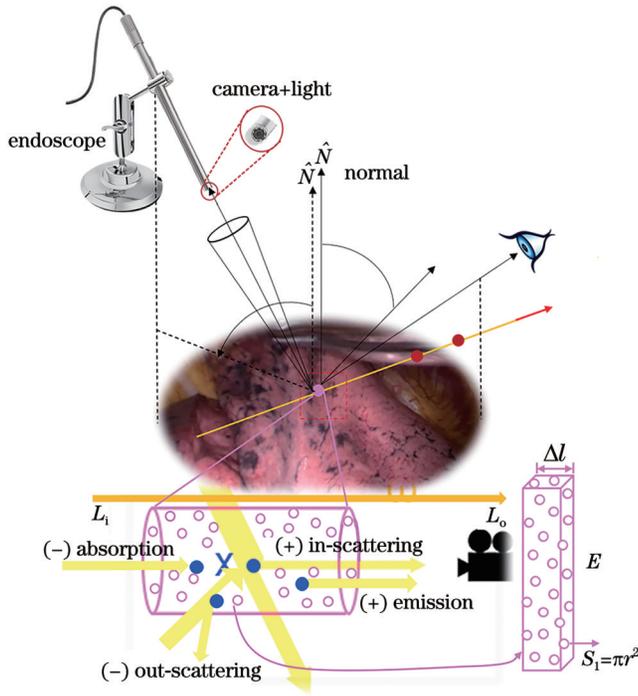


图 1 基于光学的复现软组织的颜色信息示意图

Fig. 1 Schematic of optically based reproduction of color information of soft tissues

衰减。粒子遮挡的面积为 $\rho E \Delta l S_1$ (粒子总数), 占整个底面积的比例为 $\rho E \Delta l S_1 / E = \rho \Delta l S_1$ 。在立方体的一端发射无数条光线 (假设光线朝着相同方向传播), 接收到的光线总强度相比入射光线总强度会有 $\rho \Delta l S_1$ 比例的衰减, 即出射光的强度均值是入射光的 $\rho \Delta l S_1$ 倍。数学上可以表示为

$$L_o - L_i = \Delta L = -\rho S_1 L(l) \Delta l. \quad (2)$$

将式 (2) 转换成常微分方程 $dL/dl = -\rho(l) S_1 L(l) = -\tau_a(l) L(l)$, 即

$$L(l) = L_o \exp \left[-\int_0^l \tau_a(t) dt \right], \quad (3)$$

式中: L_o 为光线的起始点; $\tau_a(t)$ 为光学厚度^[18]。式 (3) 表示吸收过程。

单个粒子发射光的辐射强度为 L_e , 粒子总数为

$$L(l) = \int_0^l \exp \left[-\int_0^t \tau_i(u) du \right] [\tau_a(t) L_e(t) + \tau_s(t) L_s(t)] dt + L_o \exp \left[-\int_0^l \tau_i(t) dt \right] = \int_0^l T(t) \sigma(t) C(t) dt + T(l) L_o, \quad (9)$$

式中: $T(l) = \exp \left[-\int_0^l \sigma(t) dt \right]$ 。

得到穿过场景的任何光线的颜色后, 复现生物软组织结构的颜色信息:

$$C(l) = \int_0^l T(t) \sigma(t) C(t) dt + T(l) L_o, \quad (10)$$

式中: $T(l) = \exp \left[-\int_0^l \sigma(t) dt \right]$ 。

$\rho S_1 E \Delta l$, 则总的发光强度为 $L_e \rho S_1 E \Delta l$ 。在立方体一端接收粒子放射的光线, 能接收到光线的概率为 $\rho S_1 E \Delta l / E = \rho S_1 \Delta l$, 接收到的光线的平均强度为 $L_e \rho S_1 \Delta l$ 。同样地, 可以得到放射光强的常微分方程为

$$\frac{dL}{dl} = L_e(l) \rho(l) S_1 = L_e \tau_a(l), \quad (4)$$

式中: L_e 为一个关于 l 的函数, 立方体内不同位置的粒子所放射的光强有差异。式 (4) 表示放射过程。

粒子除了吸收光子, 也可能会弹射光子, 这个过程称为外散射, 即光子被弹射出原本的光路, 导致光线强度减弱。与吸收一样, 外散射对光线的“削弱”程度, 也与光学厚度相关, 这一过程可以表示为

$$\frac{dL}{dl} = -\tau_s(l) L(l), \quad (5)$$

式中: τ_s 表示外散射对光线的削弱效果, 以区别于 τ_a 。式 (5) 表示外散射过程。

光子可以被弹射走, 就有其他光路的光子被弹射到当前光路, 这一过程就是内散射。内散射的过程比外散射更加复杂, 因为弹射到当前光路的光子可能来自多条不同的光路, 需要综合考虑其他光路的辐射强度以及各种弹射角, 本文不深入探究。内散射的过程可以表示为

$$\frac{dL}{dl} = \tau_s(l) L_s(l), \quad (6)$$

式中: L_s 为其他光路的辐射强度; τ_s 为弹射到当前光路的能量损失比。散射过程可表示为

$$\frac{dL}{dl} = -\tau_s(l) L(l) + \tau_s(l) L_s(l), \quad (7)$$

散射是外散射与内散射的叠加。

由式 (1) 可得

$$\frac{dL}{dl} = -\tau_a(l) L(l) + L_e(l) \tau_a(l) - \tau_s(l) L(l) + \tau_s L(l). \quad (8)$$

假设 $\tau_i = \tau_a + \tau_s$, τ_i, τ_a, τ_s 相等, 统一用 σ 表示, 同时令 $C = L_e + L_s$, 式 (8) 可以进一步简化为

2.2 神经变形场表示

将手术场景表示为时间相关的变形场和规范辐射场, 因此, 机器人手术模拟场景的每一帧图片都可以当作是规范场关于时间的变形。 $\Gamma_\phi(x, y, z; t)$ 表示与时间相关的变形场被编码在一个具有 ReLU 激活的简单 8 层 MLP 中, 并将时空坐标 $(x, y, z; t)$ 作为输入映射为时间 t 处的点 $X(x, y, z)$ 和规范场中的对应点之

间的位移量。 F_θ 表示规范辐射场是另一个具有网络参数 θ 的8层深度全连接神经网络,将连续空间中采样点的位置坐标 $X(x, y, z)$ 和方向坐标 $d(\theta, \phi)$ 映射为RGB颜色值 $c(x, y, z; t)$ 和体积密度 $\sigma(x, y, z)$ 。对于任意时间 t ,任意空间坐标 X 处的RGB颜色值 c 和体积密度 σ 等相关信息可以表示为 $F_\theta[x + \Gamma_\phi(x, y, z; t), y + \Gamma_\phi(x, y, z; t), z + \Gamma_\phi(x, y, z; t); d]$ 。与其他建模方法相比,位移场能够明确地表达所有软组织形变。为了重建手术场景的高频细节、形状、颜色、形变和更

清晰的纹理结构,使用位置编码 $\lambda(\cdot)$ 来表示输入坐标。本文方法中的关键实现步骤概述如图2所示。图2给出了用于机器人手术模拟中手持内窥镜围绕场景旋转一周随意捕获的情况下重建可形变软组织的方法。该方法把时间 t 作为附加维度加入输入中,同时该方法涉及两个部分:一个是变形场 Γ_ϕ ,一个是规范场 F_θ 。变形场 Γ_ϕ 将所有场景变形映射到一个规范场 F_θ ,规范场从每个相机射线输出,得到体积密度 σ 和相关的颜色值 c 。图2中: $r(x, t)$ 表示光线, x 表示位置坐标; \hat{C} 表示颜色损失。

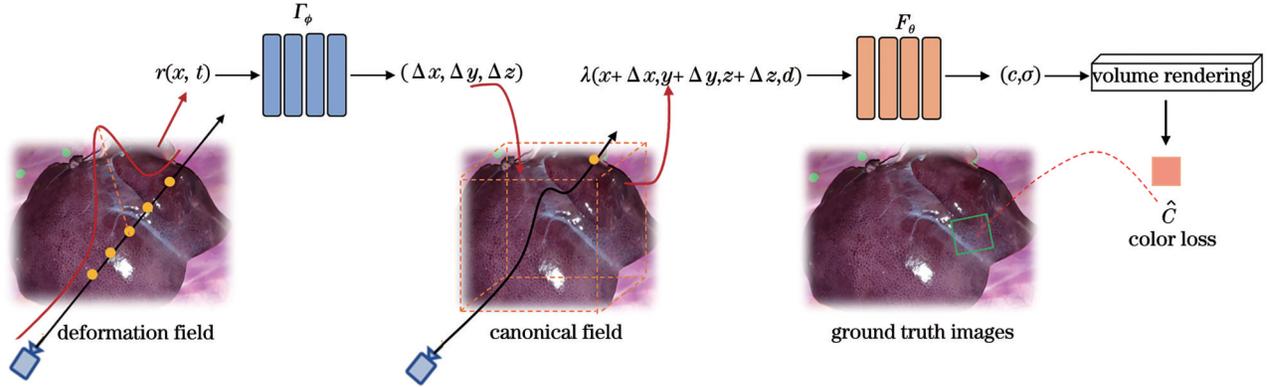


图2 关键实现步骤

Fig. 2 Key implementation steps

2.3 分解混合时空表示

如图3所示,本文将手术场景用SASM分割为3种区域,并用不同的方法对这些区域进行建模。SASM由3个部分组成,分别是图像编码器、提示编码

器和掩码解码器。其中图像编码器是基于稳健且扩展的预训练技术,它的作用是把图像映射到特征空间,使用平均绝对误差(MAE)最小化进行预训练的视觉转换器(ViT)是专为处理高分辨率输入数据而定制的。

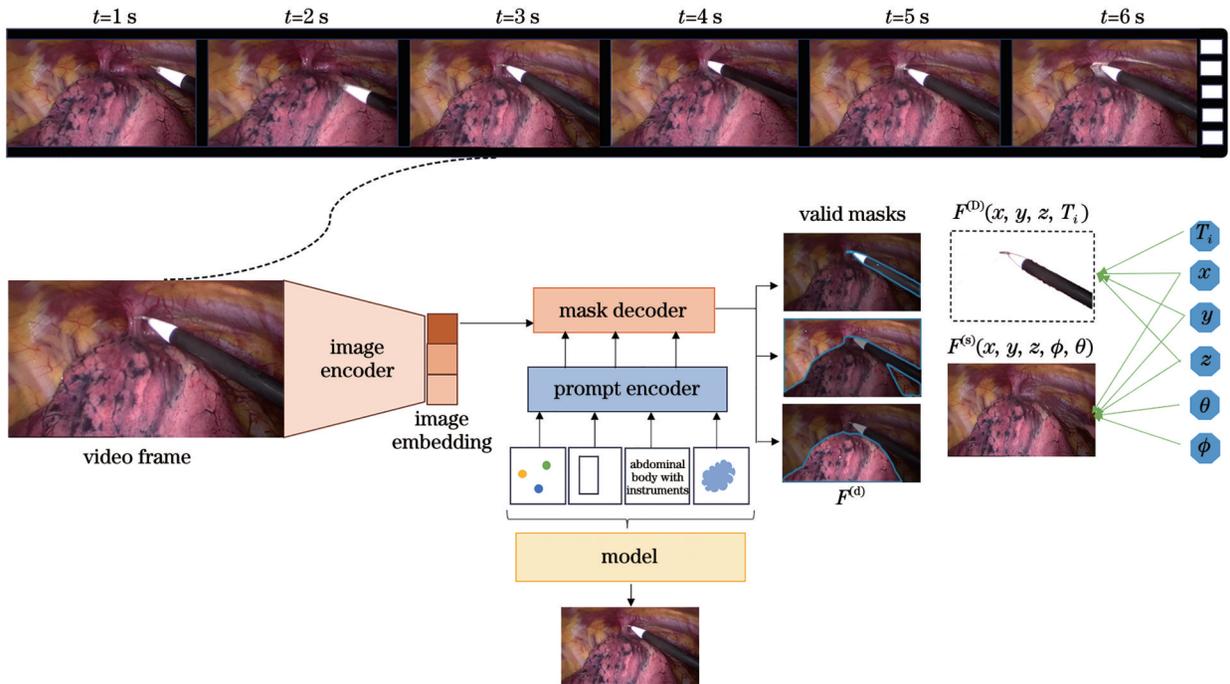


图3 将内窥镜下手术场景重建为复合辐射场,用SASM将内窥镜下手术场景分割为3个区域,分别为动态区域、静态区域、可变形区域
Fig. 3 Reconstruct the endoscopic surgery scene into a composite radiation field, and use the SASM to divide the endoscopic surgery scene into three regions: dynamic region, static region, and deformable region

图像编码器对每张图像运行一次,在提示编码器之前使用。提示编码器将点、框、文本等信息通过位置编码表示,这些编码与每种提示类型的学习嵌入相加,并且使用对比语言-图像预培训(CLIP)的现成文本编码器来表示自由文本。掩码解码器将图像嵌入、提示嵌入和输出标记输出映射到掩码。采用一个修改的Transformer解码器块,修改的解码器块使用提示自注意力和交叉注意力两个方向来更新所有嵌入。在运行两个块之后,对图像嵌入进行上采样,并使用MLP将输出令牌映射到动态线性分类器,然后在每个图像位置计算掩码前景概率,最后完成场景的分割。使用SASM将单目立体内窥镜视频完成分割后,手术场景被分割为3种区域,对这3种区域进行建模,如图4所示(手术场景假设分为3个区域,其中:静态场由用于建模时不变区域的静态特征体的MLP表示;变形场是

由用于建模低频时变外观的具有时间 t 输入的微小MLP组成;分解域中采用了一个小的MLP来预测概率,变形场和辐射场由两个小的MLP表示),分解场 $f(\cdot)$ 是由显式缓存的特征 V_f 和一个小型的MLP解码器 D_f 组成。由于变形稀疏且随时间 t 的频率变化较小,因此变形场 $F^{(d)}$ 使用较小的MLP表示。显式缓冲特征 V_{static} 和一个小型的MLP解码器用来表示静态场 $F^{(s)}$ 。时间 t 和从 V_{static} 中得到的特征作为小型MLP的输入,使用小型MLP是为了对由时变照明引起的时变外观变化进行建模,假设输入数据也是低频信号,动态场 $F^{(D)}$ 是显式保存的特征 $V_{dynamic}$ 。值得注意的是,在以上显式表示中, V_f 和 $V_{dynamic}$ 使用 $4D(x, y, z, t)$ 作为输入,因此使用流式通道。最后,期望特征向量 v 由一个辐射场 $r(\cdot)$ 解码,视角方向 (θ, ϕ) 也被发送到辐射场 $r(\cdot)$ 中。图4中下标 k 代表dynamic、static、deform。

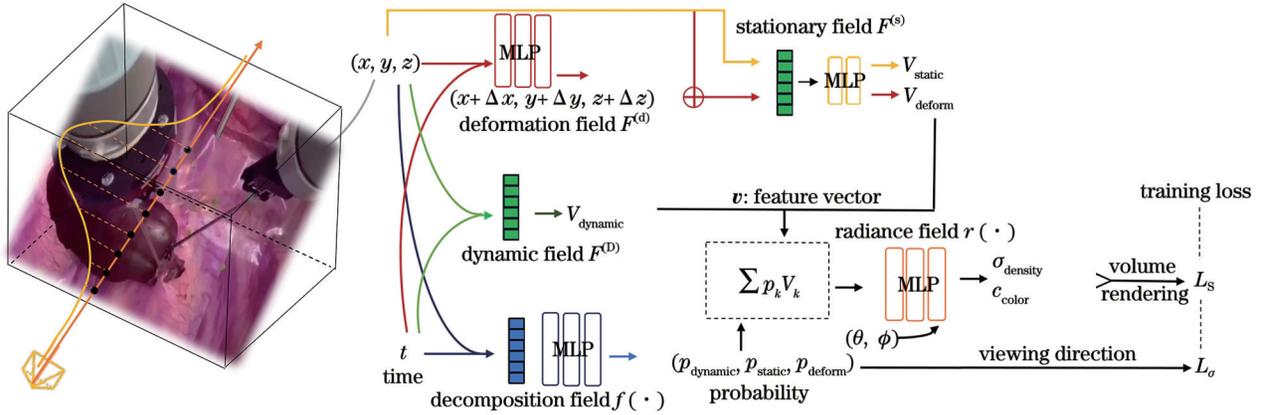


图4 总体框架概述

Fig. 4 Overview of the general framework

2.3.1 静态区域

静态区域在手术场景中具有恒定的几何结构和固定的位置,例如腹腔。此外,本文假设静态区域的形态外观不会随着时间的推移而急剧变化,即在时间上是低频信号。这是因为观察到的几何外观变化是由光照条件引起的,并且反照率是不随时间变化的。因此,将静态场 $F^{(s)}$ 表示为连续的空间相关密度 σ 和空间视图相关辐射 c :

$$F^{(s)}(x, d) = \begin{cases} \sigma^{(s)}(x) \in \mathbf{R} \\ c^{(s)}(x, d) \in \mathbf{R}^3 \end{cases}, \quad (11)$$

式中: $x \in \mathbf{R}^3$ 为空间坐标; $d \in \mathbf{R}^3$ 为视角方向。

2.3.2 变形区域

变形区域是具有变形表面的对象,例如图2中的脏器。变形区域可以分为刚性运动区域和非刚性运动区域。非刚性运动总是有序呈现。变形区域 Γ_d : $(x, y, z, t) \rightarrow (\Delta x, \Delta y, \Delta z)$ 表示变形点,然后将变形点 $(x + \Delta x, y + \Delta y, z + \Delta z)$ 作为查询点发送到预定义的规范空间场景中。

2.3.3 动态区域

对序列中某个点移动到另一个空间位置进行建模,例如将手术场景中的手术器械从中间区域移动到上方区域。采用具有随时间变化的输入 (x, y, z, t) 的特征区域 $n(\cdot)$ 来表示动态区域 $F^{(D)}$:

$$F^{(D)}(x, d, T_i) = \begin{cases} \sigma^{(D)}(x, T_i) \in \mathbf{R} \\ c^{(D)}(x, d, T_i) \in \mathbf{R}^3 \end{cases}, \quad (12)$$

式中: $T_i \in \mathbf{R}^m$ 为每帧时间潜在编码。

2.4 监督损失与优化

本文的训练过程与NeRF的训练过程几乎一致,首先从观测数据集中随机采样一批相机射线 l ,然后对这些射线 l 上的点由粗到细地采样,以进行神经网络的训练。在实际的重建和体渲染任务中,想要获得分解场的精确监督信号很困难。因此,本文提出了一组正则化,以自我监督的方式促进这种分解。

2.4.1 动态因子和静态因子

分解不能共存于同一空间位置的物理对象,合理的解决方案是该空间应具有由静态场景或动态对象占

据的任何位置,但不能同时由两者占据。为了实现这种方案,本文将动态体积密度与静态体积密度的空间比率表示为

$$w(x, T_i) = \frac{\sigma^{(D)}(x, T_i)}{\sigma^{(D)}(x, T_i) + \sigma^{(S)}(x)} \in [0, 1], \quad (13)$$

惩罚其偏离分类 $\{0, 1\}$, 二元熵损失为

$$L_b(l, T_i) = \int_{t_n}^{t_f} H_b\{w[l(t), T_i]\} dt,$$

$$H_b = -[x \log_2 x + (1-x) \log_2 (1-x)], \quad (14)$$

式中: $[t_n, t_f]$ 为光线 l 传播的距离。

然而,由于动态神经网络[式(12)]的强大表现力,不管手术场景是动态的还是静态的,优化损失[式(14)]将导致部分场景建模为动态场景。为了解决这个问题,本文提出了一个偏斜的熵损失,以使损失稍微偏向于对具有偏斜超参数 k 的场景的静态解释,获得了所需的损失:

$$L_s(l, T_i) = \int_{t_n}^{t_f} H_b\{w[l(t), T_i]^k\} dt, \quad (15)$$

式中: $H_b = -[x \log_2 x + (1-x) \log_2 (1-x)]$ 。

2.4.2 静态场景正则化

根据经验,静态分量可能会把相机位姿滥用为当前时间帧的提示,并将动态场景学习为稀疏点云,从而导致外观高频变化,如图 5 所示。这种模糊来自于这种情况,即我们使用的是单目立体内窥镜,在拍摄的过程,相机几乎从未两次访问完全相同的位置。也就是说,相机位姿和时间变量之间存在一对一的映射。这里通过对沿射线的密度分布施加先验来解决这个问题,惩罚会导致云状伪影的密度分布损失:

$$L_{\sigma s}(r) = - \int_{t_n}^{t_f} p(t) \log_2 p(t) dt, \quad (16)$$

式中: $r(t)$ 为动态场景的相机射线; $p(t) =$

$$\frac{\sigma^{(S)}[r(t)]}{\int_{t_n}^{t_f} \sigma^{(S)}[r(s)] ds}, r(s) \text{ 为静态场景的相机射线。图 5 中}$$

$L_{\sigma s}$ 为密度分布损失。

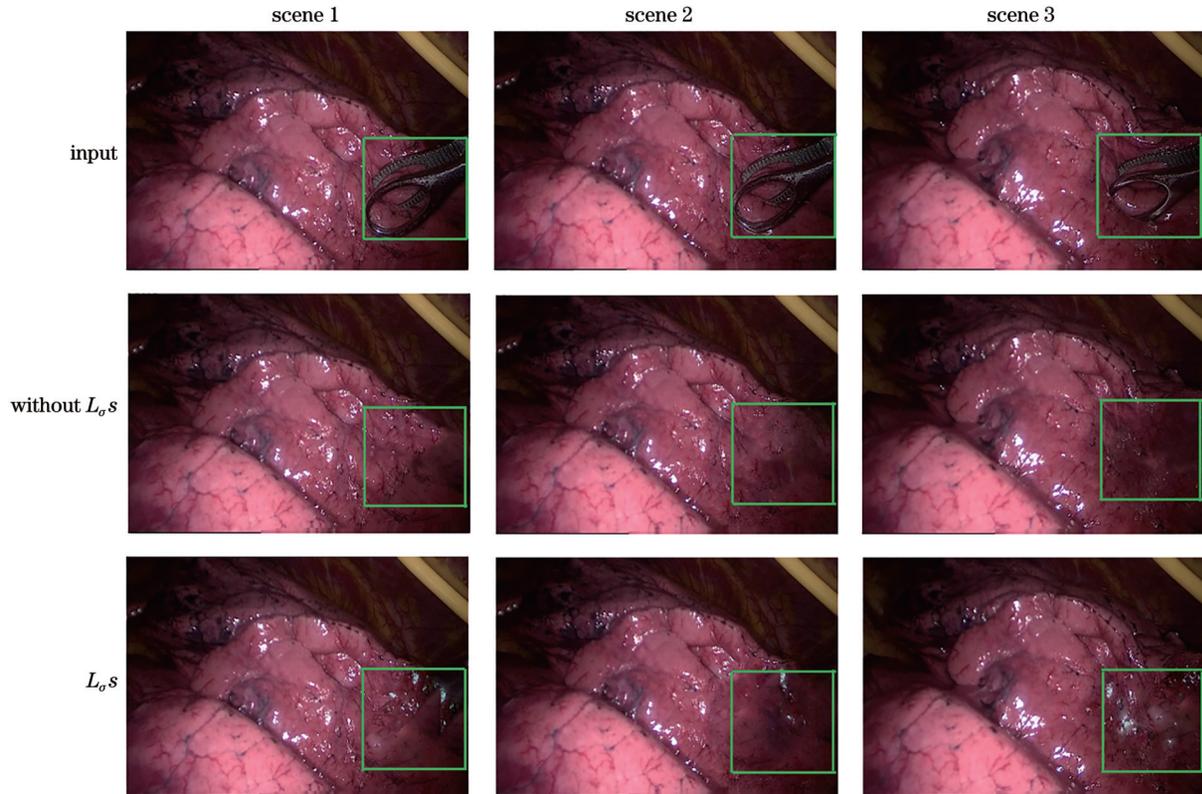


图 5 通过促进沿着每个相机射线在静态分量中更集中的密度分布,使得恢复的背景包含更多的高光,进而增强图像的真实感
Fig. 5 By promoting a more concentrated density distribution along each camera ray in the static component, the recovered background contains more highlights to enhance the realism of the image

3 实验

3.1 实验步骤与评估指标

实验装置被部署在达芬奇研究工具包(dVRK)^[19]上进行评估。如图 6 所示,将一个猪的脏器(肾脏、肝

脏、肺脏)放在模拟人体腹腔体场景的模型上,圆形亮片贴在人体腹腔体模型内用来模拟真实的手术场景,引脚末端执行单目立体内窥镜工作站视频帧。为了在场景中产生变形,使用 dVRK 来控制单个手术机械臂,即患者侧机械手(PSM),以抓取和拉动猪组织。进行

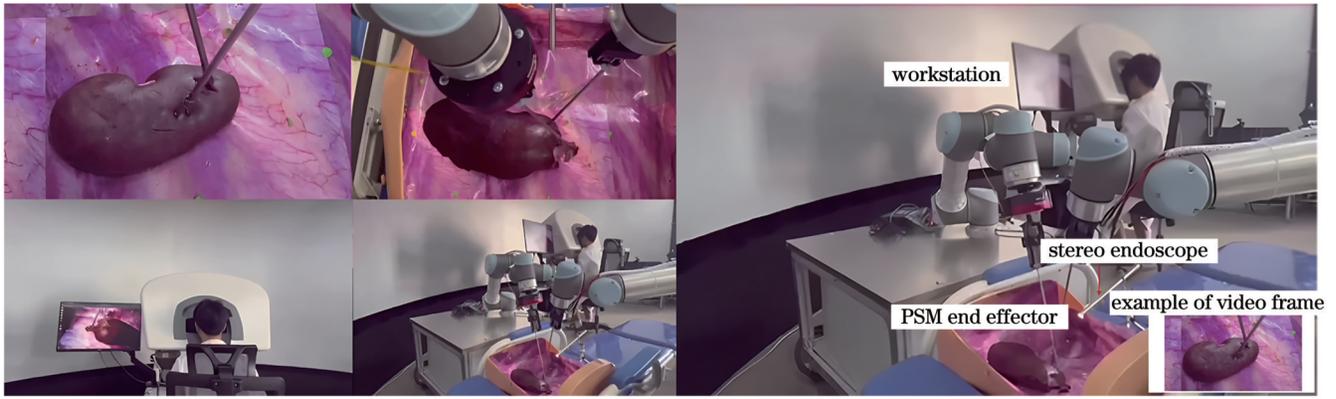


图 6 实验设置与机器人手术模拟系统

Fig. 6 Experimental setup and robotic surgical simulation system

了 6 组实验, 每组实验由 200 帧组成, 原始立体视频的视频分辨率和帧速率分别为 $1920 \text{ pixel} \times 1080 \text{ pixel}$ 和 30 frame/s 。在手术机器人脏器切除术数据集和猪的各个脏器(肺脏、肾脏、肝脏)数据集的单目内窥镜立体视频上评估了所提出的方法, 该数据集是典型的内窥镜下机器人手术模拟的立体视频。在内部达芬奇机器人脏器切除术的实验过程中, 将约 50 个引脚连接到猪组织表面(图 6), 覆盖重要区域, 如变形较大的区域和组织边界, 并在每次实验中提取其在图像平面中的轨迹。通过检测跟踪模式获得引脚轨迹, 其中通过识别图像的色调-饱和度值(HSV)色彩空间中的区域来实现检测。视频是在距离场景不同的位置和视点下拍摄的, 旨在捕捉具有非刚性变形软组织结构、动态手术器械和静态腹腔体的手术场景并对捕获的软组织(肺脏、肾脏、肝脏)结构进行 3D 重建。通过使用标准图像质量指标^[20][包括峰值信噪比(PSNR)、结构相似度(SSIM)、学习感知图像块相似度(LPIPS)]对所提出的方法与现有方法进行比较评估。此外, 为了进行定性评估, 展示了重建的生物软组织结构, 比较不同方法获得的纹理、颜色和几何细节, 并使用色差公式进行定量评估, 发现所提方法在复现软组织结构颜色信息方面具有显著的效果。

3.2 实施细节

模型框架使用 Pytorch 实现, 且框架是普遍适用的, 适用于任何采用显示特征的混合表示。将场景标准化为设备坐标(NDC), 以处理多视图内窥镜视频, 然后在 NDC 空间内投射光线。使用两个主干 InstantNGP^[21]和 TensoRT^[22]来实现该框架。在实现框架中, 变形网络是通道数为 256 的 4 层 MLP, 静态场使用的是通道数为 64 的 2 层 MLP。辐射场 $r(\cdot)$ 是通道数为 64 的 4 层 MLP, 并且具有与主干中的解码器相同的结构。对于基于 InstantNGP 的模型, 其级别数为 8, 每个条目的特征数为 4, 基于 TensoRT 的模型与基于 LLFF 格式数据集的实验模型具有相同的设置。使用具有 64 个粗略样本和 64 个精细样本的分层体积采

样, 对手术场景进行 50000 次迭代训练, 光线批处理大小为 1024, 学习率从 10^{-2} 到 10^{-5} 呈指数衰减, 此训练过程在 Ubuntu20.04 的单个 NVIDIA GeForce RTX 3060 GPU 上约需要 20 min。

4 实验结果

4.1 定量和定性的结果

在合成图像质量方面, 对于定性评估, 针对(顶部)手术真实场景运用高度可分割模型 SASM 进行分割, 将内窥镜下机器人手术模拟场景分割为动态手术器械、腹腔体背景和软组织结构(图 7, 为了更好地说明分割对象, 使用高度可分割模型 SASM。SASM 只适用于解耦动态目标, 因此分割部分不包括在整个捕捉过程中保持静止的对象, 例如手术场景中的腹腔体)。将所提出的方法在软组织三维重建方面的准确性与 NeRF^[21]和 InstantNGP^[23]进行比较, 以衡量其性能。对于 NeRF, 相机本身存在畸变和不合理的信号采样即采样频率低于信号频率, 很难重建手术模拟场景, 重建的生物软组织结构会出现锯齿和伪影(图 8), 呈现较差的重建效果; 对于 InstantNGP, 内存限制问题导致 GPU 内存消耗较高, 这种内存限制会影响图像的分辨率。另外, InstantNGP 中使用的哈希网格可能难以有效地表示和覆盖大空间范围的场景, 这可能会影响模型在高分辨率设置下的性能, 因此重建的内窥镜下机器人手术模拟场景中的软组织结构在高分辨率设置中存在局限性^[24]。对于定量评估, 将本文结果与合成数据的地面真实情况进行了比较, 并将 PSNR(衡量重建的保真度)^[25]、SSIM(衡量重建与真实情况之间的结构相似性)和 LPIPS(度量两图像之间的差异)作为生物软组织结构的新视图合成的度量, 见表 1。本文使用的方法进一步增强了手术重建细节的能力, 研究结果表明, 本文方法在不影响重建质量的情况下实现了较为显著的优化, 在机器人手术模拟器的未来临床应用中显示出巨大的发展潜力。

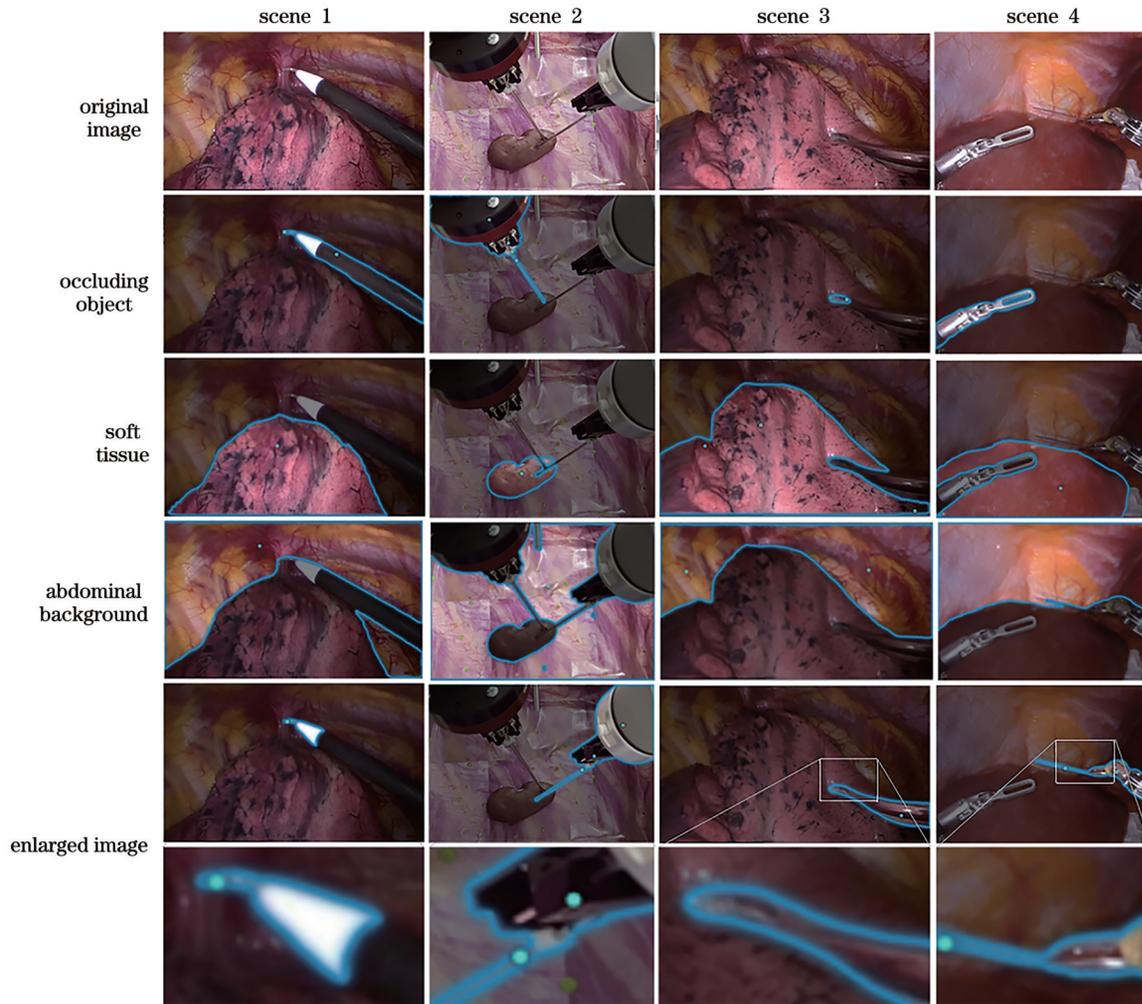


图 7 在真实手术场景上分割的可视化结果

Fig. 7 Visualization results of segmentation in real surgical scenes

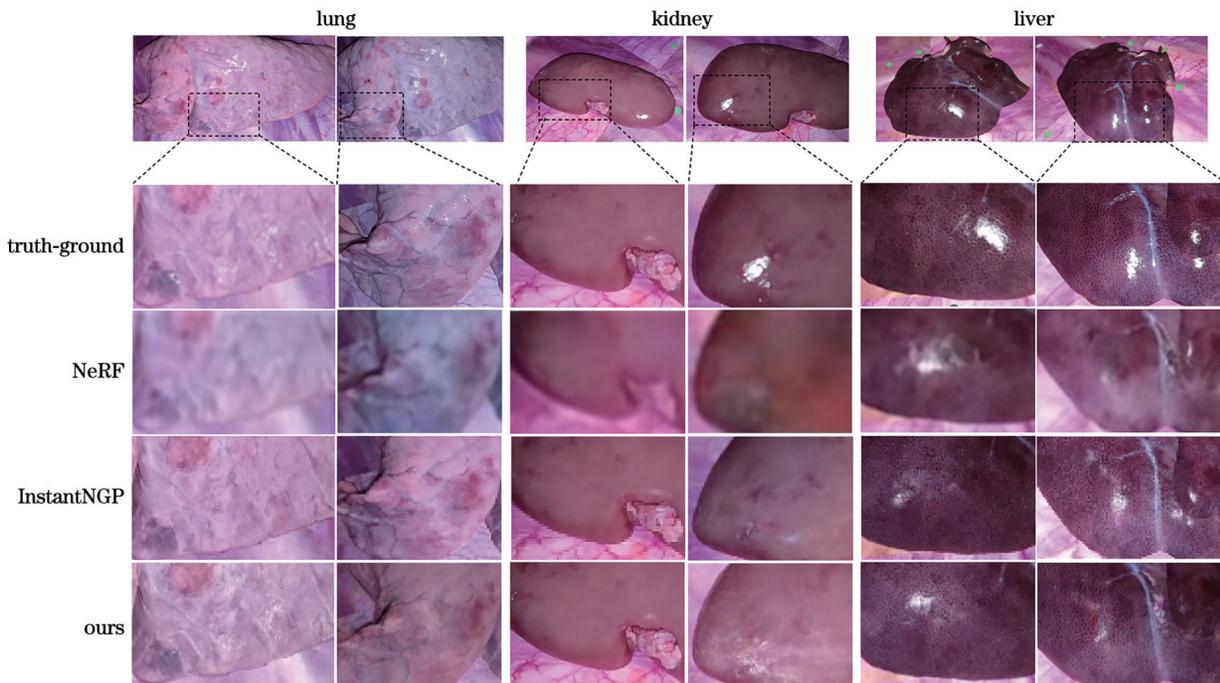


图 8 机器人手术模拟内窥镜场景下的各软组织重建结果(定性)

Fig. 8 Results of each soft tissue reconstruction in the endoscopic scenario of robotic surgery simulation (qualitative)

表1 对每个机器人手术模拟场景下各个软组织结构进行 50000 次训练,引入监督损失后重建的数据集,并与地面真实情况进行比较(定量)

Table 1 Dataset reconstructed by the introduction of supervised loss after each soft tissue structure is trained for 50000 times in each robotic surgery simulation scenario, and the dataset is compared with the ground truth

Method	PSNR of lung	SSIM of lung	LPIPS of lung	PSNR of kidney	SSIM of kidney	LPIPS of kidney	PSNR of liver	SSIM of liver	LPIPS of liver
NeRF	19.9	0.553	0.905	20.1	0.584	0.897	23.5	0.596	0.879
InstantNGP	23.5	0.792	0.816	28.5	0.696	0.871	26.0	0.601	0.792
Ours	28.6	0.860	0.792	29.4	0.888	0.794	28.4	0.899	0.738

在 3D 重建色差方面,本文进行了一项分析,涉及评估不同器官(即肺脏、肾脏和肝脏)的重建图像的最大、平均和最小色差。该分析是在内窥镜机器人手术模拟场景下进行的,重点是将这些重建图像与内窥镜相机捕获的原始图像进行比较。用于此评估的度量是欧几里得距离公式^[26]。图 9 展示了各个器脏的色差折线图,色差随着图像尺寸呈倍数增加,在重建器官(肺脏、肾脏和肝脏)中观察到最大、最小和平均色差。值得注意的是,重建的肺部图像与其对应的真实场景之间的最大色差约为 0.95。根据 Lab 色彩空间中既定的

色差标准^[27],该色差在 0.5~1.0 的范围内,表示中等水平的差异,该色差范围通常被认为对于各种应用来说是可以接受的。此外,重建肺部图像的平均色差始终低于 0.1。参考 Lab 色差标准,色差值低于 0.25,表示几乎可以忽略不计的差异,这表明色差已达到最佳匹配状态,重建肺部颜色与实际肺部颜色之间具有很好的一致性。从图 9(a)~(c)中可以推断出本文提出的方法在解决生物软组织重建中的色差方面取得了良好的结果。结果表明,本文方法可以进行准确的色彩再现,特别是在涉及软组织结构的场景中。

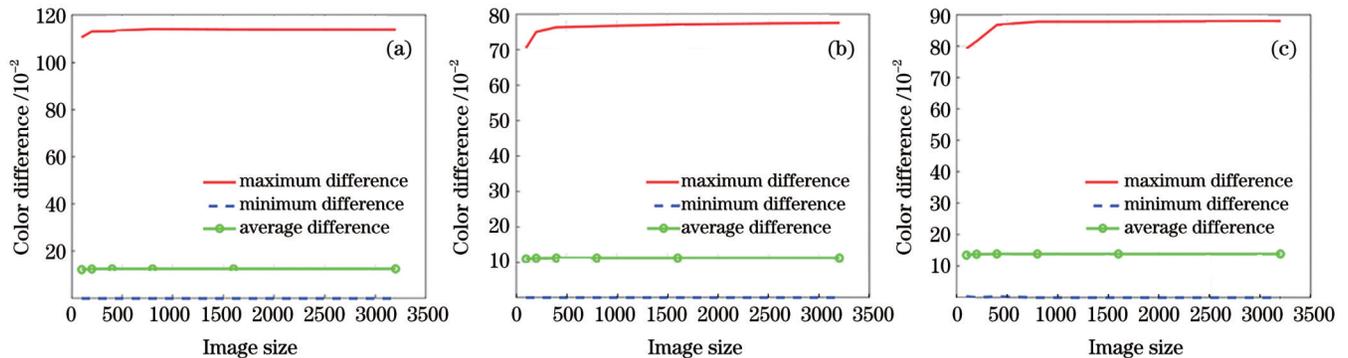


图 9 各器脏的色差折线图。(a)肺脏色差图;(b)肾脏色差图;(c)肝脏色差图

Fig. 9 Line graphs of color difference of each organ. (a) Lung color difference line graph; (b) kidney color difference line graph; (c) liver color difference line graph

最后从位置色差、放大率色差、矢量色差 3 个维度(图 10)进一步说明本文方法重建器脏的颜色接近生物软组织的真实颜色。以最小色差为特征的器脏 3D 重建在机器人手术模拟器领域中具有至关重要的作用,它在手术精度、患者安全和程序创新方面带来了转变。随着这项技术与机器人辅助手术的融合,本文方法会在多个方面产生影响,有助于提高医疗护理标准和改进手术结果。最重要的是,在机器人手术模拟器中部署具有最小色差的 3D 重建重新定义了术前规划的前景。外科医生能够以极高的准确度捕捉到患者器脏解剖结构的精确的可视化图像。有限的色差确保了颜色和纹理的真实保留,使外科医生能够以前所未有的真实感观察生物软组织复杂的轮廓。这种高保真度的表示显著提高了术前评估的精度,允许考虑患者特

定的解剖变化,制定细致的手术策略。

4.2 超参数

本文深入研究了各种参数对重建可变形生物软组织性能的影响。每个实验都涉及单个超参数的修改,在不同的超参数配置下将本文提出的方法与 NeRF 进行比较,评估定量和定性结果。从图 11 可以看出,通过增加迭代次数,重建生物软组织的性能得到改善。图 12 突出显示,PSNR、结构相似性指数(SSIM)和平均结构相似性指数(MSSIM)等指标随着输入图像数量(来自不同观点)的增加而增大。这种改进归因于本文方法可以提供更多的相机位姿,提高重建质量,从而更好地表示底层结构或特征。图像数据量的增加能够在重建过程中捕获更多的细节,减少伪影和噪声,同时提高重建图像的保真度和准确性。

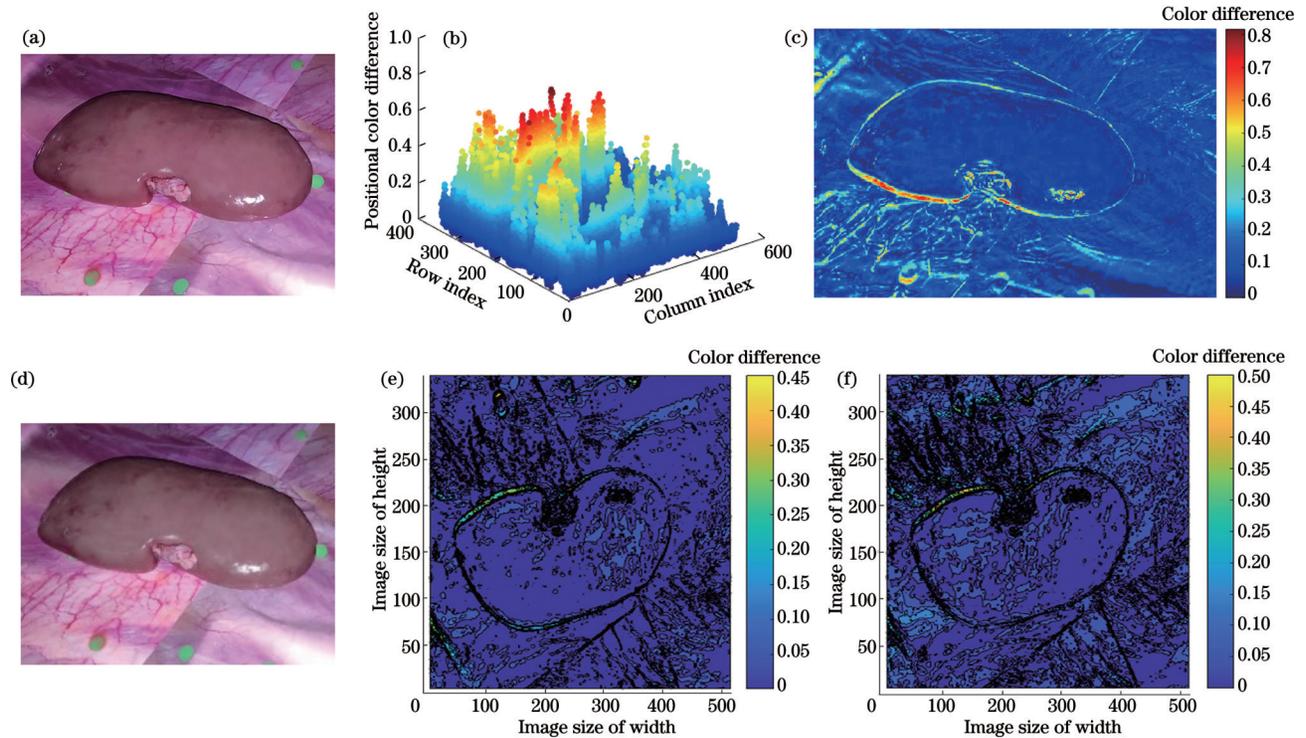


图 10 不同的色差表示说明重建器脏的颜色接近生物软组织的真实颜色。(a)拍摄的原始肾脏图;(b)位置色差 3D 散点图;(c)色差等高线图;(d)3D 重建后的肾脏图;(e)颜色编码地图中 R 通道的矢量色差图;(f)颜色编码地图中 G 通道的矢量色差图
 Fig. 10 Different color difference representations indicate that the color of the reconstructed organ is close to the true color of biological soft tissue. (a) Photographed original kidney map; (b) 3D scatter plot of positional color difference; (c) contour map of magnification chromatic aberration; (d) 3D reconstructed kidney map; (e) vector chromatic aberration map in color coded map of R channel; (f) vector chromatic aberration map in color coded map of G channel

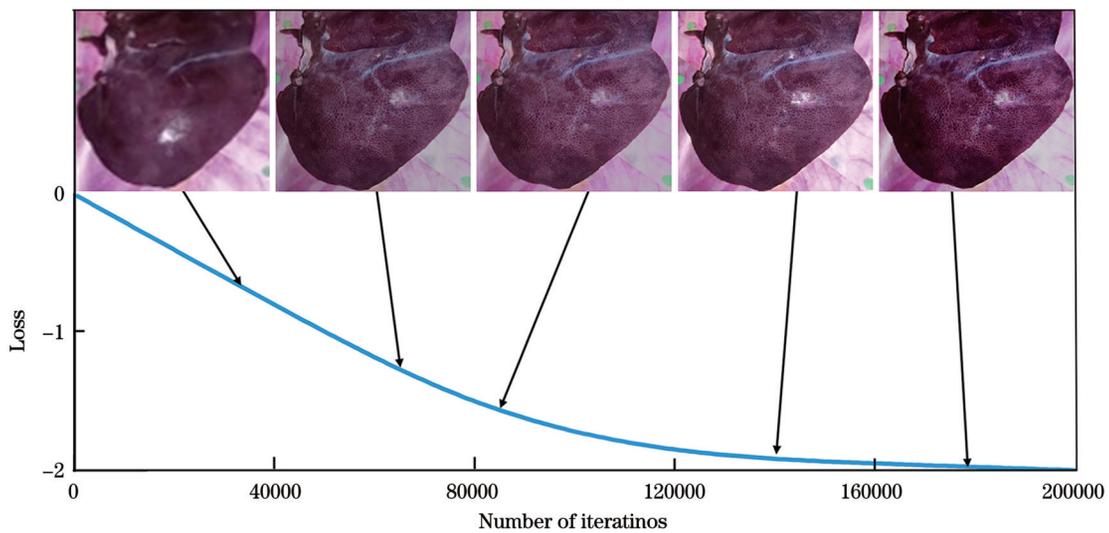


图 11 随着时间的推移(通过增加迭代次数来表示),重建生物软组织的性能得到改善
 Fig. 11 Performance of reconstructed biological soft tissue improves over time (represented by increasing number of iterations)

图 13 表明隐藏层的数量对 PSNR 的影响最小。这一结果与之前的研究^[28]一致,表明哈希网格有效地保留了体积信息。控制每层哈希网格增量大小的比例因子 s 是一个关键参数。过小的 s 可能使网络无法充分编码高频图像细节,而过大的 s 可能会导致网格快速扩展,从而导致严重的哈希冲突。随

着更多特征被编码,附加级别的引入最初提高了可变形软组织重建的准确性。所提出的方法通过采样合并图像正则化。为了验证正则化的有效性,使用不同的正则化权重 γ 进行可变形生物软组织重建。正如预期的那样,较高的 γ 值有助于更精确的重建。

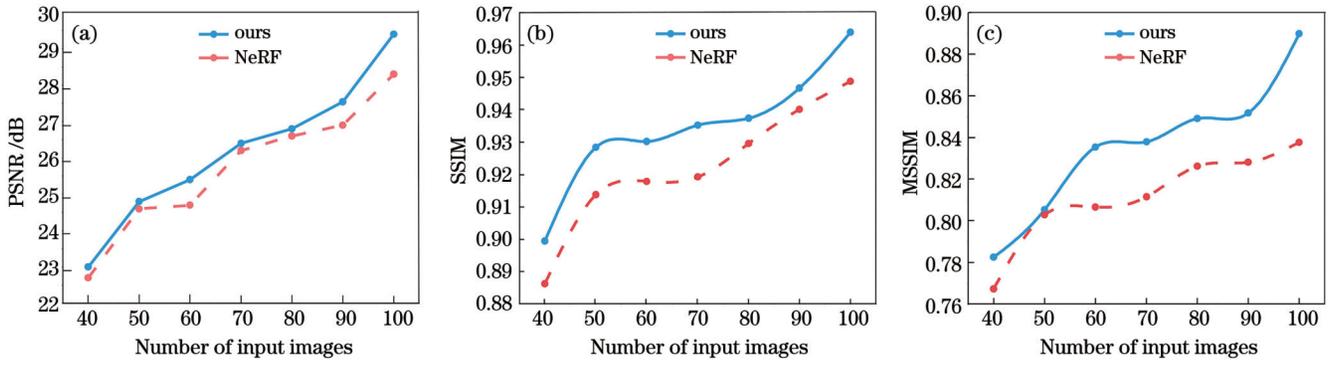


图 12 超参数(输入图片的数量)对重建效果的影响。(a) PSNR;(b) SSIM;(c) MSSIM

Fig. 12 Influence of hyperparameters (the number of input images) on reconstruction performance. (a) PSNR; (b) SSIM; (c) MSSIM

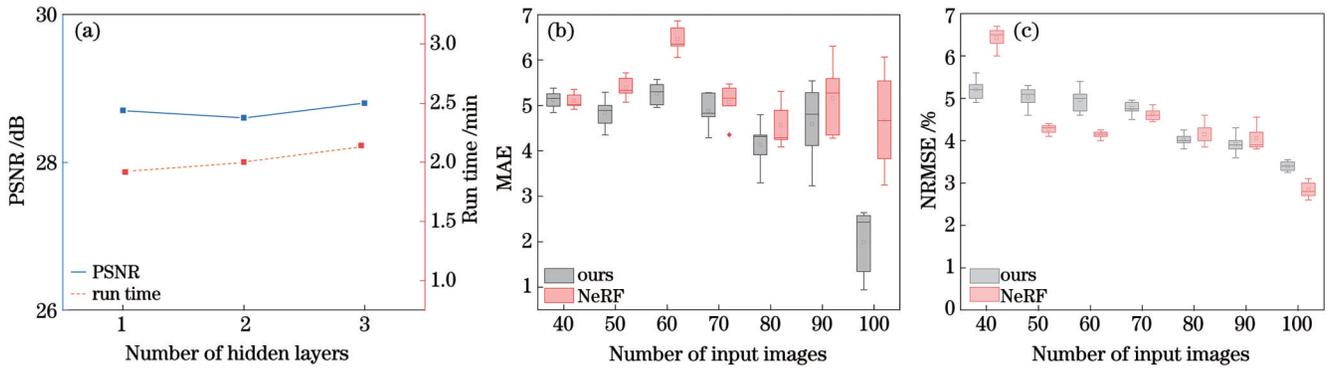


图 13 哈希网格中的超参数(隐藏层的数量、正则化权重)对重建质量的影响。(a) PSNR;(b) MAE;(c) NRMSE

Fig. 13 Influence of hyperparameters (number of hidden layers, regularization weights) in hash grids on reconstruction quality.

(a) PSNR; (b) MAE; (c) NRMSE

4.3 讨论

受采集内窥镜机器人手术场景有限视野的限制,重建复杂机器人手术模拟场景通常需要更多的手术场景信息数据。先前的研究工作主要依赖于明确的场景几何来进行新视点生成。尽管如此,获取实际的手术场景数据通常充满挑战,包括但不限于遮挡、深度变化以及背景元素的干扰。本文提出一种基于 NeRF 和 SASM 的方法,通过 NeRF 来表示光场,通过编辑光场中的相机姿态获得不同视角下的图像,完成角度维度的密集采样。SASM 可分割模块将手术场景分解为动态、静态和可变形区域,有助于生物软组织的重建。本文设计了一种时空混合表示,以便对分解后的各个 NeRF 进行建模。本文方法相对于 NeRF 和 InstantNGP 的主要优势是基于 NeRF 来对手术场景进行隐式表达,为真实的手术场景拟合出准确的隐式函数,准确表达复杂手术场景,能够解决 NeRF 无法重建有遮挡及深度有变化的场景的问题,同时突破了 NeRF 不适合处理复杂的具有可形变动态手术场景的限制。将输入变量通过哈希网格映射到该变量的傅里叶特征,能够有效地解决 InstantNGP 对场景高频纹理信息较难拟合的问题。

5 结论

本文提出了一种新的基于 NeRF 的框架,对随意的多视点单目立体内窥镜视频进行自监督 3D 动态手术场景解耦和生物软组织重建。本文方法对动态手术器械遮挡和可变形软组织结构进行解耦,恢复静态腹腔体背景表示,并实现高质量的新视图合成。本文框架的关键部分是 SASM 和 NeRF, SASM 的高度可分割模块将手术场景分解为动态、静态和可变形区域。本文设计了一种时空混合表示,以便高效地对分解后的各个 NeRF 进行建模。本文方法在机器人手术数据的各种模拟场景中取得了较为优异的性能,例如大幅度移动的手术器械和可形变软组织结构 3D 重建。相信,本文方法可以促进人们对机器人手术模拟器场景的理解,希望新兴的基于 NeRF 的 3D 重建技术可以激发机器人手术模拟器场景理解的新途径,并赋予各种下游临床导向任务。

在未来的工作中,我们的主要任务是在增加本文方法的渲染时间的同时重建被手术器械遮挡的生物软组织结构,以更有效地支持术中手术。我们将致力于降低输入数据的要求,旨在将本文方法的适用性扩展到更广泛的无约束手术场景中。

参 考 文 献

- [1] Widya A R, Monno Y, Okutomi M, et al. Whole stomach 3D reconstruction and frame localization from monocular endoscope video[J]. *IEEE Journal of Translational Engineering in Health and Medicine*, 2019, 7: 3300310.
- [2] Agarwal S, Snively N, Simon I, et al. Building Rome in a day [C]//2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE Press, 2009: 72-79.
- [3] Sun H Y, Wang S L, Bai J X, et al. Confocal laser scanning and 3D reconstruction methods for the subsurface damage of polished optics[J]. *Optics and Lasers in Engineering*, 2021, 136: 106315.
- [4] Chen L, Tang W, John N W, et al. SLAM-based dense surface reconstruction in monocular Minimally Invasive Surgery and its application to Augmented Reality[J]. *Computer Methods and Programs in Biomedicine*, 2018, 158: 135-146.
- [5] Cox I J, Hingorani S L, Rao S B, et al. A maximum likelihood stereo algorithm[J]. *Computer Vision and Image Understanding*, 1996, 63(3): 542-567.
- [6] 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.
- [7] Tong Z, Gao J, Zhang H T. Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks[J]. *Construction and Building Materials*, 2017, 146: 775-787.
- [8] Hoang D C, Lilienthal A J, Stoyanov T. Object-RPE: dense 3D reconstruction and pose estimation with convolutional neural networks[J]. *Robotics and Autonomous Systems*, 2020, 133: 103632.
- [9] Sanz-Requena R, Moratal D, García-Sánchez D R, et al. Automatic segmentation and 3D reconstruction of intravascular ultrasound images for a fast preliminar evaluation of vessel pathologies[J]. *Computerized Medical Imaging and Graphics*, 2007, 31(2): 71-80.
- [10] Shen Q H, Yang X Y, Wang X C. Anything-3D: towards single-view anything reconstruction in the wild[EB/OL]. (2023-04-19)[2023-11-09]. <http://arxiv.org/abs/2304.10261>.
- [11] Li J N, Li D X, Xiong C M, et al. BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation[EB/OL]. (2022-01-28)[2023-11-09]. <https://arxiv.org/abs/2201.12086>.
- [12] Mazurowski M A, Dong H Y, Gu H X, et al. Segment anything model for medical image analysis: an experimental study [J]. *Medical Image Analysis*, 2023, 89: 102918.
- [13] Wiesent K, Barth K, Navab N, et al. Enhanced 3-D-reconstruction algorithm for C-arm systems suitable for interventional procedures[J]. *IEEE Transactions on Medical Imaging*, 2000, 19(5): 391-403.
- [14] González Izard S, Sánchez Torres R, Alonso Plaza Ó, et al. Nextmed: automatic imaging segmentation, 3D reconstruction, and 3D model visualization platform using augmented and virtual reality[J]. *Sensors*, 2020, 20(10): 2962.
- [15] 张璇晔, 朱疆. 内窥光学相干层析成像的研究进展与应用[J]. *中国激光*, 2023, 50(21): 2107103.
Zhang X Y, Zhu J. Research progress and applications of endoscopic optical coherence tomography[J]. *Chinese Journal of Lasers*, 2023, 50(21): 2107103.
- [16] 董晓文, 徐杨, 高程, 等. 基于视锥感光空间的光源相关色温计算[J]. *光学学报*, 2022, 42(2): 0233001.
Dong X W, Xu Y, Gao C, et al. Calculation of light source corrected color temperature based on cone fundamental space[J]. *Acta Optica Sinica*, 2022, 42(2): 0233001.
- [17] Pertsch T, Xiao S M, Majumdar A, et al. Optical metasurfaces: fundamentals and applications[J]. *Photonics Research*, 2023, 11(5): OMFA1-OMFA3.
- [18] 殷永凯, 于锴, 于春展, 等. 几何光场三维成像综述[J]. *中国激光*, 2021, 48(12): 1209001.
Yin Y K, Yu K, Yu C Z, et al. 3D imaging using geometric light field: a review[J]. *Chinese Journal of Lasers*, 2021, 48(12): 1209001.
- [19] Saracino A, Deguet A, Staderini F, et al. Haptic feedback in the da Vinci Research Kit (dVRK): a user study based on grasping, palpation, and incision tasks[J]. *The International Journal of Medical Robotics + Computer Assisted Surgery: MRCAS*, 2019, 15(4): e1999.
- [20] Wang Y H, Long Y H, Fan S H, et al. Neural rendering for stereo 3D reconstruction of deformable tissues in robotic surgery [M]//Wang L W, Dou Q, Fletcher P T, et al. Medical image computing and computer-assisted intervention-MICCAI 2022. Lecture notes in computer science. Cham: Springer, 2022, 13437: 431-441.
- [21] Müller T, Evans A, Schied C, et al. Instant neural graphics primitives with a multiresolution hash encoding[J]. *ACM Transactions on Graphics*, 2022, 41(4): 1-15.
- [22] Chen A P, Xu Z X, Geiger A, et al. TensorRF: tensorial radiance fields[M]//Avidan S, Brostow G, Cissé M, et al. Computer vision-ECCV 2022. Lecture notes in computer science. Cham: Springer, 2022, 13692: 333-350.
- [23] Pupezescu M C, Pupezescu V. Instant neural graphics primitives framework and its emerging applications in digitalization of dental care[C]//2022 E-Health and Bioengineering Conference (EHB), November 17-18, 2022, Iasi, Romania. New York: IEEE Press, 2022.
- [24] Zhao B Y, Mertz J. Resolution enhancement with deblurring by pixel reassignment[J]. *Advanced Photonics*, 2023, 5(6): 066004.
- [25] Jain A, Tancik M, Abbeel P. Putting NeRF on a Diet: semantically consistent few-shot view synthesis[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 5865-5874.
- [26] 刘嘉慧, 杨燕青, 马睿, 等. 有机近红外二区荧光探针研究进展[J]. *中国激光*, 2023, 50(21): 2107101.
Liu J H, Yang Y Q, Ma R, et al. Research progress of organic NIR- II fluorescent probes[J]. *Chinese Journal of Lasers*, 2023, 50(21): 2107101.
- [27] 李铁妮, 刘东远, 张鹏睿, 等. 基于深度卷积编解码网络的 fNIRS 成像技术研究[J]. *中国激光*, 2023, 50(21): 2107107.
Li T N, Liu D Y, Zhang P R, et al. Deep convolutional encoder-decoder neural network approach for functional near infrared spectroscopic imaging[J]. *Chinese Journal of Lasers*, 2023, 50(21): 2107107.
- [28] Tu R C, Mao X L, Ma B et al. Deep cross-modal hashing with hashing functions and unified hash codes jointly learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(2):560-572

Dynamic Three-Dimensional Reconstruction of Soft Tissue in Neural Radiation Field for Robotic Surgery Simulators

Chen Qi^{1,2}, Qin Zhibao^{1,2}, Cai Xiaoyu^{1,2}, Li Shijie^{1,2}, Wang Zijun^{1,2}, Shi Junsheng^{1,2*}, Tai Yonghang^{1,2*}

¹*School of Physics and Electronic Information, Yunnan Normal University, Kunming 650500, Yunnan, China;*

²*Yunnan Key Laboratory of Optoelectronic Information Technology, Kunming 650500, Yunnan, China*

Abstract

Objective Reconstructing soft tissue structures based on the endoscope position with robotic surgery simulators plays an important role in robotic surgery simulator training. Traditional soft tissue structure reconstruction is mainly achieved through surface reconstruction algorithms using medical imaging data sets such as computed tomography and magnetic resonance imaging. These methods fail to reconstruct the color information of soft tissue models and are not suitable for complex surgical scenes. Therefore, we proposed a method based on neural radiation fields, combined it with classic volume rendering to segment robotic surgery simulator scenes from videos with deformable soft tissue captured by a monocular stereoscopic endoscope, and performed three-dimensional reconstruction of biological soft tissue structures to restore soft tissue. By using segmented arbitrary scene model (SASM) for segmentation modeling of time-varying objects and time-invariant objects in videos, specific dynamic occlusions in surgical scenes can be removed.

Methods Inspired by recent advances in neural radiation fields, we first constructed a self-supervision-based framework that extracted multi-view images from monocular stereoscopic endoscopic videos and used the underlying 3D information in the images to construct geometric constraints of objects, so as to accurately reconstruct soft tissue structures. Then, the SASM was used to segment and decouple the dynamic surgical instruments, static abdominal scenes, and deformable soft tissue structures under the endoscope. In addition, this framework used a simple neural network multilayer perceptron (MLP) to represent moving surgical instruments and deformed soft tissue structures in dynamic neural radiation fields and proposed skew entropy loss to correctly predict surgical instruments, cavity scenes, and soft tissue structures in surgical scenes.

Results and Discussions We employ MLP to represent robotic surgery simulator scenes in the neural radiation field to accommodate the inherent geometric complexity and deformable soft tissue. Furthermore, we establish a hybrid framework of the neural radiation field and SASM for efficient characterization and segmentation of endoscopic surgical scenes in an endoscopic robotic surgery simulator. To address the dynamic nature of scenes and facilitate accurate scene separation, we propose a self-supervised approach incorporating a novel loss function. For validation, we perform a comprehensive quantitative and qualitative evaluation of a data set captured using a stereoendoscope, including simulated robotic surgery scenes from different angles and distances. The results show that our method performs well in synthesizing realistic robotic surgery simulator scenes compared with existing methods, with an average improvement of 12.5% in peak signal-to-noise ratio (PSNR) and an average improvement of 8.43% in structural similarity (Table 1). It shows excellent results and performance in simulating scenes and achieving high-fidelity reconstruction of biological soft tissue structures, color, textures, and other details. Furthermore, our method shows significant efficacy in scene segmentation, enhancing overall scene understanding and accuracy.

Conclusions We propose a novel NeRF-based framework for self-supervised 3D dynamic surgical scene decoupling and biological soft tissue reconstruction from arbitrary multi-viewpoint monocular stereoscopic endoscopic videos. Our method decouples dynamic surgical instrument occlusion and deformable soft tissue structures, recovers a static abdominal volume background representation, and enables high-quality new view synthesis. The key parts of our framework are the SASM and the neural radiation field. The highly segmentable module of SASM decomposes the surgical scene into dynamic, static, and deformable regions. A spatiotemporal hybrid representation is then designed to facilitate and efficiently model the decomposed neural radiation fields. Our method achieves excellent performance in various simulation scenes of robotic surgery data, such as large-scale moving surgical instruments and 3D reconstruction of deformable soft tissue structures. We believe that our method can facilitate robotic surgery simulator scene understanding and hope that emerging NeRF-based 3D reconstruction technology can provide inspiration for robotic surgery simulator scene understanding and empower various downstream clinically oriented tasks.

Key words visual optics; neural radiation field; three-dimensional reconstruction of soft tissue; segmented arbitrary scene model; segmentation decoupling