



Efficient EndoNeRF reconstruction and its application for data-driven surgical simulation

Yuehao Wang¹ · Bingchen Gong¹ · Yonghao Long¹ · Siu Hin Fan² · Qi Dou¹

Received: 14 June 2023 / Accepted: 13 March 2024 / Published online: 24 April 2024
© The Author(s) 2024

Abstract

Purpose The healthcare industry has a growing need for realistic modeling and efficient simulation of surgical scenes. With effective models of deformable surgical scenes, clinicians are able to conduct surgical planning and surgery training on scenarios close to real-world cases. However, a significant challenge in achieving such a goal is the scarcity of high-quality soft tissue models with accurate shapes and textures. To address this gap, we present a data-driven framework that leverages emerging neural radiance field technology to enable high-quality surgical reconstruction and explore its application for surgical simulations.

Method We first focus on developing a fast NeRF-based surgical scene 3D reconstruction approach that achieves state-of-the-art performance. This method can significantly outperform traditional 3D reconstruction methods, which have failed to capture large deformations and produce fine-grained shapes and textures. We then propose an automated creation pipeline of interactive surgical simulation environments through a closed mesh extraction algorithm.

Results Our experiments have validated the superior performance and efficiency of our proposed approach in surgical scene 3D reconstruction. We further utilize our reconstructed soft tissues to conduct FEM and MPM simulations, showcasing the practical application of our method in data-driven surgical simulations.

Conclusion We have proposed a novel NeRF-based reconstruction framework with an emphasis on simulation purposes. Our reconstruction framework facilitates the efficient creation of high-quality surgical soft tissue 3D models. With multiple soft tissue simulations demonstrated, we show that our work has the potential to benefit downstream clinical tasks, such as surgical education.

Keywords 3D reconstruction · NeRF · Robotic surgery · Surgery simulation

Introduction

The development of realistic robotic surgery scenes is important for VR-based surgical training. The conventional method for creating these surgery scenes involves manual creation of soft tissue models with in vivo textures by skilled artists. However, this approach is highly time-consuming and restricts the level of detail and variety achievable in surgical simulation. To overcome these limitations, we propose an

automated approach to reconstruct interactive surgical environments using captured real data.

Surgical reconstruction [1–7], as an emerging task, aims to recover the 3D shapes and appearance of soft tissues from in vivo surgery videos. As pointed out by previous literature [6, 7], surgical reconstruction is cursed with three typical challenges over natural scene reconstruction: 1) Soft tissues will undergo large and drastic deformations. Many surgical operations, e.g., cutting and tearing, can even damage the topologies of soft tissues. 2) Surgical tools usually appear on the surgery videos and partially occlude underlying soft tissues from observation. 3) Endoscopic surgery videos are captured in confined in vivo spaces, resulting in limited multi-view geometric clues of the 3D shapes. Our recent work EndoNeRF [7] exploits the strong capacity of NeRF [8] for scene representations and incorporates tailored modules for handling tool occlusion and single-viewpoint

✉ Qi Dou
qidou@cuhk.edu.hk

¹ Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

² Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China

input, achieving significant improvements in surgical reconstruction, particularly for scenes with large deformations. However, EndoNeRF encounters new practical challenges when constructing surgical simulation environments. First, the process of reconstructing a surgical scene from endoscopic videos using EndoNeRF is inefficient, requiring over 10 h for per-scene optimization. Second, the optimized geometry of EndoNeRF is represented in a purely implicit field, i.e., the whole scene is encoded by network parameters. However, many physically-based methods in soft-body simulation [9–12] require explicit geometry model, e.g., meshes, particles, or tetrahedrons, rather than implicit fields. It is also worth noting that the realistic interaction of soft tissues is reliant on the underlying content beneath the tissue surface. While the geometry in the EndoNeRF only represents the surfaces of soft tissues. Hence, apart from surface reconstruction, another significant challenge lies in recovering topologically closed counterparts of soft tissues for simulation purposes.

To fill this gap, this work is the first attempt to create surgical simulation environments with soft tissue surfaces automatically reconstructed from endoscopic surgery videos. Technically, we propose a novel framework for dynamic surgical reconstruction, which can yield realistic and simulator-friendly counterparts of the soft tissues in the input robotic surgery videos. We summarize our main contributions as follows:

- We adopt a novel voxel grids-based scene representation for faster dynamic surgical scene reconstruction.
- We build a pipeline for converting radiance fields into a closed mesh, which enables physically-based simulation of the reconstructed surgical scenes.
- We exhibit multiple robotic surgery simulations with our reconstructed soft tissues on multiple simulation engines, including Taichi MPM [13, 14] and NVIDIA Isaac Sim [15].

This work builds upon a preliminary version presented at MICCAI 2022 [7]. In this paper, we have made significant revisions and extensions to the original conference version. The major improvements include:

- We designed a new deformable scene representation with grid-based radiance fields and 4D tensor-decomposed motion fields for faster training convergence.
- We proposed a novel pipeline for extracting closed meshes from radiance fields, in order to generate simulatable soft tissues.
- We conduct multiple surgical scene simulations with our reconstructed soft tissues (Fig. 1).

Our code is available at <https://github.com/med-air/EndoNeRF>.

Method

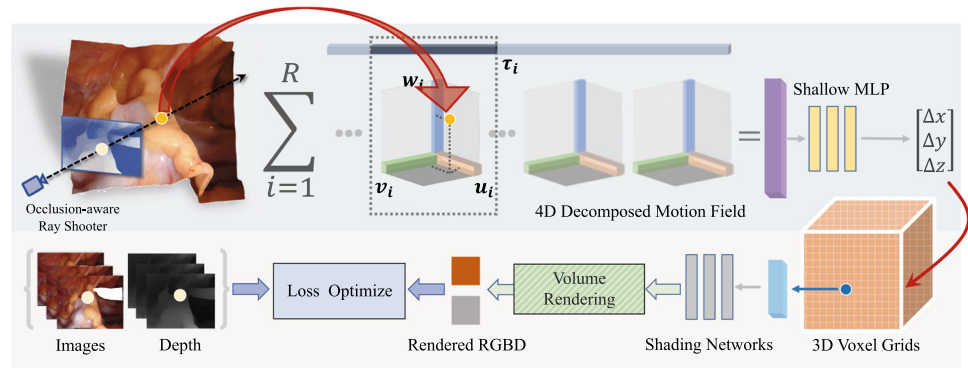
We first aim to propose a dynamic scene representation to model soft tissue's 3D shapes and textures from a stereo video clip of a dynamic surgical scene. Then we devise a particular de-occlusion rendering and stereo depth-supervised loss for optimizing the scene representation. Finally, we fill the reconstructed mesh surfaces into closed meshes and perform soft-body simulations on the filled meshes. The detailed descriptions are as follows.

Efficient EndoNeRF scene representations

In order to enable high-fidelity reconstruction of the surgical simulation environments, we resort to neural radiance fields. The fundamental neural radiance fields [8] for 3D scene representations are modeled in a coordinate-based MLP. Optimizing such scene representation to convergence is slow. Alternatively, we adopt an implicit-explicit voxel grids-based scene representation, which is shown to achieve much faster optimization [16–19]. Specifically, we model the shape and appearance of the scene in density volume grids $\mathbf{V}_\sigma \in \mathbb{R}^{H \times W \times D}$ and feature volume grids $\mathbf{V}_a \in \mathbb{R}^{C \times H \times W \times D}$, where H , W , and D are the resolutions for the x , y and z dimensions and C is the channel number of the appearance features. For the density volume grids $\mathbf{V}_\sigma$, each grid vertex maintains its occupancy probability. For the feature volume grids $\mathbf{V}_a$, each grid vertex holds an appearance code. To map the appearance code into RGB color, we introduce a shallow MLP $S_\Theta : \mathbb{R}^C \rightarrow \mathbb{R}^3$ as a learnable implicit shading module. The geometry and appearance of any point $\mathbf{x}$ in the 3D space can be retrieved via tri-linear interpolation (denoted as $\text{interp}(\cdot)$) of the 8 surrounding vertices' densities and features, i.e., the density $\sigma(\mathbf{x}) = \text{interp}(\mathbf{x}, \mathbf{V}_\sigma)$ and the color $\mathbf{c}(\mathbf{x}) = S_\Theta(\text{interp}(\mathbf{x}, \mathbf{V}_a))$.

Next, we consider surgical scene deformations. A dynamic surgical scene can be decomposed into a canonical radiance field and a time-dependent deformation field [20, 21]. Thereby the dynamic scene at time t can be viewed as the canonical field warped by the deformation field at t . In our proposed method, the canonical radiance field is represented by the $\mathbf{V}_\sigma$ and $\mathbf{V}_a$. To support large and topology-varying deformations, we adopt decomposed 4D motion fields and a 3-layer MLP to model the deformation field, which maps a spatial-temporal coordinate $(\mathbf{x}, t)$ into its corresponding displacement $\Delta\mathbf{x}$. In specific, we define a motion feature field as a $H \times W \times D \times T \times C_T$ tensor $\mathcal{T}$ [22], where T is the resolution of the time dimension and C_T is the temporal feature channel number. Direct dense 5D modeling of the motion feature field is costly in storage and over-high-dimensional

Fig. 1 Pipeline of our proposed FastEndoNeRF framework, consisting of a 4D-decomposed motion field and dense 3D voxel grids



for optimization on sparsely captured frames. Thus, we need to seek another compact representation. Since deformations can be locally continuous and low-rank, as observed in [19, 23], we can decompose this tensor via outer product (Eq. 1):

$$\begin{aligned} \mathcal{T} = & \sum_{r=1}^{R_1} \tau_r^4 \circ \mathcal{V}_r^{1,2,3} \circ b_r^4 + \sum_{r=1}^{R_2} \tau_r^3 \circ \mathcal{V}_r^{1,2,4} \circ b_r^3 \\ & + \sum_{r=1}^{R_3} \tau_r^2 \circ \mathcal{V}_r^{1,3,4} \circ b_r^2 + \sum_{r=1}^{R_4} \tau_r^1 \circ \mathcal{V}_r^{2,3,4} \circ b_r^1, \end{aligned} \quad (1)$$

where R_1, R_2, R_3 and R_4 are expected rank for each dimension, τ_r^l is a 1-D vector of the l -th dimension, b_r^l is a feature basis of the l -th dimension, and $\mathcal{V}_r^{i,j,k}$ is a 3-D volume encompassing i, j, k -th dimensions. For each continuously queried point $(\mathbf{x}, t)$, we tri-linearly interpolate component tensors τ_r^l and $\mathcal{V}_r^{i,j,k}$ to obtain a motion feature vector. Then we feed the motion feature vector into a 3-layer MLP G_ϕ to compute the output displacement vector. In this way, the corresponding coordinates in the canonical field can be obtained by $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}(\mathbf{x}, t)$ with $\Delta\mathbf{x}(\mathbf{x}, t) = G_\phi(\text{interp}(\mathbf{x}, t, \mathcal{T}))$.

Rendering and optimization

Volume rendering With this scene representation, we can reconstruct the deformable surgical scene by optimizing the loss between rendered color $\hat{\mathbf{C}}$ and ground truth color $\mathbf{C}$. Specifically, the rendered color of the ray $\mathbf{r}(z) = \mathbf{o} + z\mathbf{d}$ at time t can be evaluated by volume rendering as shown in Eq. 2:

$$\begin{aligned} \hat{\mathbf{C}}(\mathbf{r}(z), t) &= \sum_{j=1}^M w_j \mathbf{c}_j, \\ w_j &= \exp\left(-\sum_{i=1}^{j-1} \sigma_i \delta_i\right) \left(1 - \exp(-\sigma_j \delta_j)\right), \end{aligned} \quad (2)$$

where M is the number of sampled points along $\mathbf{r}(z)$, δ_i is the sampling step length, σ_j and $\mathbf{c}_j$ are the density and

color of the j -th sample evaluated by $\sigma(\mathbf{x}_j + \Delta\mathbf{x}(\mathbf{x}_j, t))$ and $\mathbf{c}(\mathbf{x}_j + \Delta\mathbf{x}(\mathbf{x}_j, t))$, respectively. The attenuation term w_j can be regarded as the probability that the ray is transmitted to the j -th sample.

De-occlusion of surgical tools According to the literature [6, 7], soft tissues in surgical videos can often be occluded by surgical tools in the foreground. To address this issue and accurately reconstruct the soft tissues, our approach focuses on training the rays corresponding to tool pixels. Following the methodology proposed in EndoNeRF [7], we generate binary tool masks for the left view of each frame. Instead of the mask-guided ray sampling proposed in EndoNeRF [7], which bypasses rays per training iteration, we pre-compute all possible camera rays and check for intersections between these rays and the tool masks prior to training. This saves the computational costs during the scene optimization procedure, resulting in faster training. Any rays that pass through the tool masks are excluded from the training process. During training, the training batch $\mathcal{R}$ is randomly sampled from the pre-computed rays that have been screened in this manner. By doing so, we ensure that the optimization of the scene representation bypasses the tool pixels. Leveraging the auto-interpolation property of radiance fields, we can patch the occluded soft tissue areas using information from adjacent frames throughout the training procedures.

Distillation of stereo correspondence. To exploit stereo geometry in confined in vivo input, we propose to leverage stereo geometry to enrich 3D clues over the optimization of the scene representation. Very recent work unimatch [24] learns dense correspondence on general vision datasets in a unified formulation for optical flow, stereo matching, and depth estimation tasks. Due to its superior performance over the previous method [25], we propose to distill stereo correspondence learned on general data into the surgical data along with the optimization of the surgical scene. To measure the learned stereo correspondence of the scene representation, we render depth from the radiance fields via $\hat{\mathbf{D}}(\mathbf{r}(z), t) = \sum_{j=1}^M w_j z_j$, where z_j is the distance of the j -th sample along the ray $\mathbf{r}(z)$. The rendered depth is expected to converge to

the estimated stereo depth once well-matched stereo correspondence is attained by optimizing the scene representation. Thus, we estimate stereo depth $\mathbf{D}(\mathbf{r}(z), t)$ by stereo-matching the feature correspondence of the robotic surgery videos from unimatch [24]. Lastly, we add a depth-supervised loss to the objective function, resulting in the final loss function:

$$\mathcal{L} = \sum_{\mathbf{r}(z) \in \mathcal{R}, t \in [0, 1]} \left\| \hat{\mathbf{C}}(\mathbf{r}(z), t) - \mathbf{C}(\mathbf{r}(z), t) \right\|_2^2 + \lambda_d \text{Huber} \left(\hat{\mathbf{D}}(\mathbf{r}(z), t), \mathbf{D}(\mathbf{r}(z), t) \right), \quad (3)$$

where $\mathbf{C}(\mathbf{r}(z), t)$ and $\mathbf{D}(\mathbf{r}(z), t)$ is the corresponding ground truth pixel color and unimatch [24] stereo depth of camera ray $\mathbf{r}(z)$ at the time t . Here we adopt Huber loss [26] which is more stable to outliers. Compared with the stereo depth maps predicted by STTR [25], supervising better depth maps via unimatch can further decrease the training time since the depth refinement module proposed in EndoNeRF [7], which requires depth rendering of all training images, is no longer needed.

Extraction of closed meshes for soft-body simulations

After we obtain an optimized dynamic radiance field, we aim to perform physically-based simulations on the reconstruction. Numerically solving physically-based simulation systems requires dividing the object material domain into a number of geometry primitives. Since our reconstructed scene representation only encodes the seen soft tissue surface in an implicit geometry, we need to first obtain its explicit form and convert it into a simulatable object. To do this, we propose the following procedure. We first render the reconstructed canonical radiance fields to color and depth maps. Then, we back-project RGB-D maps into point clouds. Namely, each 3D point (x, y, z) can be computed from a corresponding pixel (P_u, P_v) with depth value $\hat{\mathbf{D}}$ as $(x, y, z) = (\hat{\mathbf{D}}(P_u - C_x)/f, \hat{\mathbf{D}}(P_v - C_y)/f, \hat{\mathbf{D}})$, where (C_x, C_y) is the principal point and f is the focal length. Bilateral filtering is also applied to smooth the point clouds. After conversion to point clouds, we perform Poisson surface reconstruction to extract the mesh surface from the simplified point clouds. Subsequently, we need to construct supporting structures underneath the surface for deformable object simulations. Material Point Method (MPM) and Finite Element Method (FEM) both require a closed mesh surface as the input for discretization. For the MPM solvers, dense particles are sampled to fill the soft tissue surface [27]. As for the FEM solver, robust tetrahedral meshing algorithms [28–30] are proposed to convert surface objects into tetrahedrons. Thus, we tailor an efficient *mesh-open2closed* algorithm that can universally enclose the reconstructed mesh surfaces. The

Algorithm 1: mesh-open2closed (Sec. 2.3)

Data: Open mesh surface $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ which has one hole, soft tissue thickness ζ

Result: Closed mesh surface with updated $\mathcal{V}$ and $\mathcal{F}$

$\mathcal{A} \leftarrow$ non-manifold edges of $\mathcal{F}$;

for each vertex $v \in \mathcal{V}$ **do**

$adj_v \leftarrow \{\}$;

end

for each edge $\{u, v\} \in \mathcal{A}$ **do**

$adj_u \leftarrow adj_u \cup \{v\}$;

$adj_v \leftarrow adj_v \cup \{u\}$;

end

/ Conduct BSF to sort boundary vertices into an ordered list $\mathcal{C}$ */*

$\{u, v\} \leftarrow \text{Head}(\mathcal{A})$;

$\mathcal{C} \leftarrow \{u\}$;

while $v \neq u$ **do**

$\text{Enqueue}(\mathcal{C}, v)$;

$u \leftarrow v$;

for each $k \in adj_v$ **do**

if $k \notin \mathcal{C}$ **then**

$v \leftarrow k$;

break;

end

end

end

$\text{Enqueue}(\mathcal{C}, \text{Head}(\mathcal{C}))$;

$center \leftarrow (\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} X_v, \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} Y_v, \zeta)$;

$\mathcal{V} \leftarrow \mathcal{V} \cup \{center\}$;

$first \leftarrow \text{True}$;

/ Iteratively build a base plane of the soft tissue surface and connect the base with the boundary vertices along $\mathcal{C}$ */*

while $! \text{isEmpty}(\mathcal{C})$ **do**

$v \leftarrow \text{Dequeue}(\mathcal{C})$;

if $!first$ **then**

/ Projection of the last v */*

$p \leftarrow v'$;

end

/ Projection of v in the appended base of the soft tissue */*

$v' \leftarrow (X_v, Y_v, \zeta)$;

$\mathcal{V} \leftarrow \mathcal{V} \cup \{v'\}$;

if $! \text{isEmpty}(\mathcal{C})$ **then**

$u \leftarrow \text{Head}(\mathcal{C})$;

$\mathcal{F} \leftarrow \mathcal{F} \cup \{(v, u, v')\}$;

end

if $!first$ **then**

$\mathcal{F} \leftarrow \mathcal{F} \cup \{(p, v, v')\}$;

$\mathcal{F} \leftarrow \mathcal{F} \cup \{(p, center, v')\}$;

end

$first \leftarrow \text{False}$;

end

pseudocode of the algorithm is given in Algorithm 1, where the input mesh vertices $\mathcal{V}$ and triangles $\mathcal{F}$ are structured in 2D arrays, and X_v, Y_v denote the x and y dimensions of vertex v . The algorithm begins with constructing the boundary edges of the reconstructed surface and organizing them in a list. Those boundary edges can be classified by non-manifold test,

i.e., manifold edges should be simultaneously included in 2 triangles. After finding non-manifold edges, we can conduct a breadth-first search (BFS) to sort boundary vertices into an ordered list. Then, we iteratively build a base plane of the soft tissue surface in the shape of the open mesh boundary and connect the base with the boundary vertices along the ordered list. During this procedure, we loop for each vertex v in the ordered edge list and find the projection of v in the appended base of the soft tissue. Then we connect the projection of the last v , the projection of the current v , and the center of the base plane to create new faces. It is noteworthy that our algorithm is designed for the input mesh with a single “hole”, i.e., there is only one connected edge. This assumption usually holds since the incisions on soft tissues are relatively shallow in in vivo surgical scenes. If there are two disjoint surfaces represented in the reconstructed field, a solution is to run the algorithm separately for each surface.

Experiments

Evaluation of efficient EndoNeRF

We conducted an evaluation of our proposed method on a set of typical clips of robotic surgery videos, captured from 10 cases of our in-house DaVinci robotic prostatectomy dataset. In addition to the cases used in EndoNeRF [7], the new cases contain suturing, bleeding, and cutting on soft tissues. Each clip lasted for 4 to 8 s and was sampled into 45 ~ 180 frames. These clips were captured from stereo cameras, and they encompassed challenging scenes with non-rigid deformation and tool occlusion. To establish the effectiveness of our new method, we compared it with two strong baselines: the recent NeRF-based method EndoNeRF [7] and the traditional DynamicFusion-based approach E-DSSR [6]. For qualitative evaluation, we exhibit the reconstruction objects produced by our method, including reconstructed point clouds, surface meshes, and closed meshes. Due to clinical regulations, it is infeasible to collect ground truth depth for numerical evaluation on 3D structures. To perform quantitative comparisons, we instead used photometric errors, such as PSNR, SSIM, LPIPS, and training time, as evaluation metrics. This evaluation methodology is consistent with that used in previous work on surgical scene reconstruction, such as [6, 7], and is widely used in the field of neural rendering.

Figure 2 showcases our reconstruction outcomes, including extracted point clouds, soft tissue mesh surfaces, and closed meshes. Our FastEndoNeRF algorithm excels at reconstructing watertight surfaces of soft tissues from videos, faithfully capturing the intricate in vivo textures. Despite the presence of large deformations, our method tracks the dynamics of the soft tissues using our proposed 4D-decomposed motion field. For tool occlusion in the input

videos, our method manages to patch tool-occluded areas by leveraging information from adjacent frames, ensuring a comprehensive and watertight representation of the dynamic soft tissue. In order to ensure that the reconstructed surface is suitable for simulation purposes in contemporary simulation engines, we have employed a mesh extraction scheme capable of constructing high-resolution meshes with intricate textures and shapes from the reconstructed point clouds. Furthermore, our proposed *mesh-open2closed* algorithm facilitates the creation of a closed structure by appending a base to the mesh surface. This closed structure is essential for enabling accurate simulations in the chosen environment. In Fig. 3, we run our method and the original EndoNeRF [7] on the same NVIDIA RTX 3090 GPU for 3 min and compare their training efficiency. Due to the limited training time, the reconstruction results obtained with EndoNeRF remain noisy and blurry. Conversely, our method demonstrates impressive performance even at an early training stage (i.e., 10 s to 60 s), with the ability to approximate the scene’s appearance and shape accurately. This validates the superior training convergence speed of our proposed scene representation. It is noteworthy that our model employs ~160 M parameters, consuming 4GB GPU memory for training each case. Without factorizing the 4D deformation field, the 4D deformation field would necessitate an allocation of over 12GB of memory during the training procedure, which shows the effectiveness of our compact dynamic scene representations.

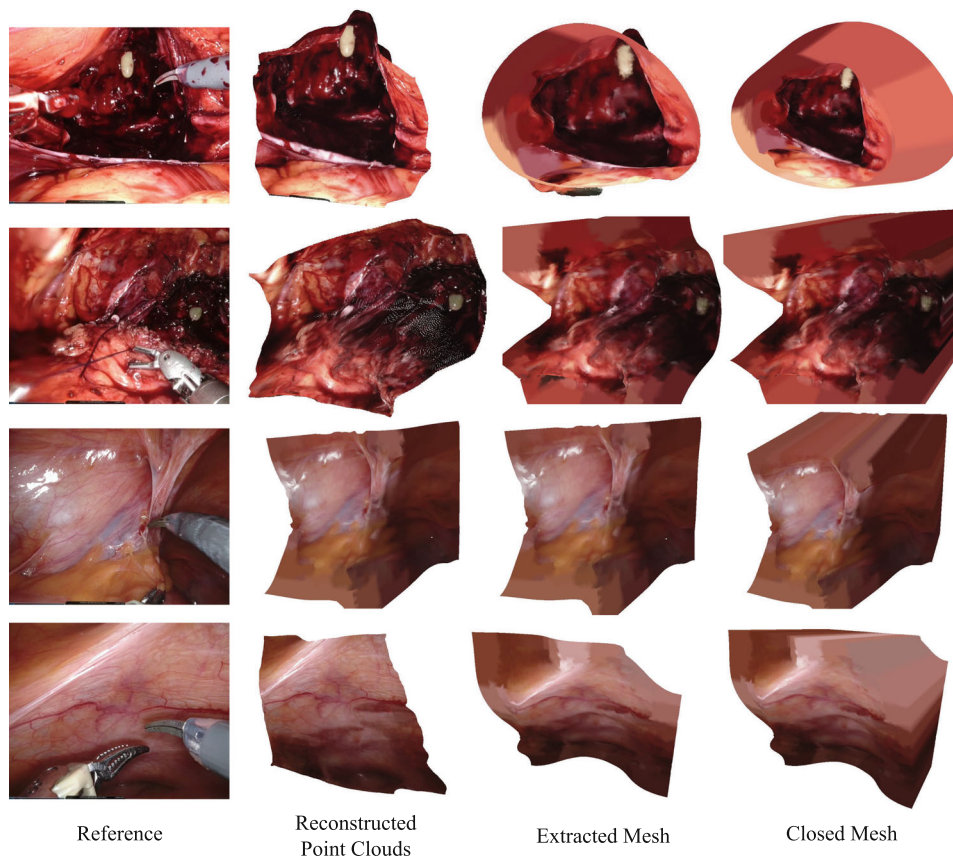
Table 1 displays a quantitative comparison of the metrics PSNR, SSIM, LPIPS, and training time. Both methods exhibit impressive photometric results when compared to the traditional method of E-DSSR [6]. Despite a slight decrease in performance, FastEndoNeRF achieves a remarkable training time improvement of approximately 20 times faster than EndoNeRF. By training FastEndoNeRF for 27 min, we can achieve comparable quality to EndoNeRF trained for over 10 h. This highlights the efficiency and effectiveness of the FastEndoNeRF approach.

Initial application for surgical scene simulation

Virtual surgical training platforms have become increasingly significant in surgery education and training [31, 32]. However, building a surgical education and training platform is associated with several challenges, including limited exposure to real-life surgical cases, and limited access to high-fidelity simulation. Our proposed framework can overcome these challenges by providing a reconstructed realistic environment for surgical trainees to practice and master their skills.

Real-time FEM simulation. Here we first build a real-time virtual surgery simulation in NVIDIA Isaac Sim [15], where FEM is the solver for simulating reconstructed continuum

Fig. 2 Reconsruction results. The first column gives the reference input image, the second column exhibits reconstructed point clouds, the third column shows the meshes obtained by Poisson surface reconstruction on the point clouds, and the last column displays the closed meshes appended with a base



objects. In the first row of Fig. 4, we import a reconstructed closed mesh into NVIDIA Isaac Sim and tune its physical properties to make it behave like soft tissues. Owing to advanced GPU acceleration, NVIDIA Isaac Sim enables real-time FEM simulation and rendering, producing high-fidelity deformations under the dissection interaction. The automatic reconstruction of the simulation environment from real surgical videos ensures that the in vivo textures are accurately preserved, thereby enhancing the visual realism of surgical simulations. The proposed algorithm for closed mesh extraction facilitates material domain discretization for the FEM solver within Isaac Sim. If the imported meshes are not closed in Isaac Sim, the mesh tetrahedralization procedure will fail, resulting in unreasonable simulation effects. Moreover, the creation procedure for this simulation environment is highly scalable, thanks to the efficiency of the surgical reconstruction pipeline.

MPM simulation While the FEM solver in NVIDIA Isaac Sim achieves basic soft-body simulation, it lacks the ability to perform damage operations on continuum objects, which is considered a crucial aspect of simulating soft tissues. In order to address this limitation, we employ the Material Point Method (MPM) [33], a hybrid grid-particle method that combines the strengths of both Eulerian and Lagrangian approaches. This method enables us to handle large deformations and complex material behavior, as demonstrated in recent papers [34, 35]. To specifically support damage deformations on soft bodies and achieve two-way coupling between rigid and non-rigid objects, we implement the state-of-the-art MLS-MPM [14]. In Fig. 4, the second row illustrates an example of soft tissue damage resulting from dissection. It is evident that MLS-MPM is capable of accurately capturing the incision behavior on the soft tissues. While MPM offers soft tissue damaging simulation, it is characterized by high computational costs and falls short

Table 1 Quantitative evaluation and comparison of our method and baselines. We evaluate photometric errors and training time of the dynamic reconstruction

Methods	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Time $\downarrow$
E-DSSR [6]	13.398 $\pm$ 1.387	0.630 $\pm$ 0.057	0.423 $\pm$ 0.047	0
EndoNeRF [7]	29.831 $\pm$ 2.208	0.925 $\pm$ 0.020	0.081 $\pm$ 0.022	> 10h
FastEndoNeRF	29.586 $\pm$ 3.48	0.892 $\pm$ 0.040	0.145 $\pm$ 0.050	27min

Fig. 3 Comparisons of reconstruction quality between EndoNeRF and our method within the first 3 min of training

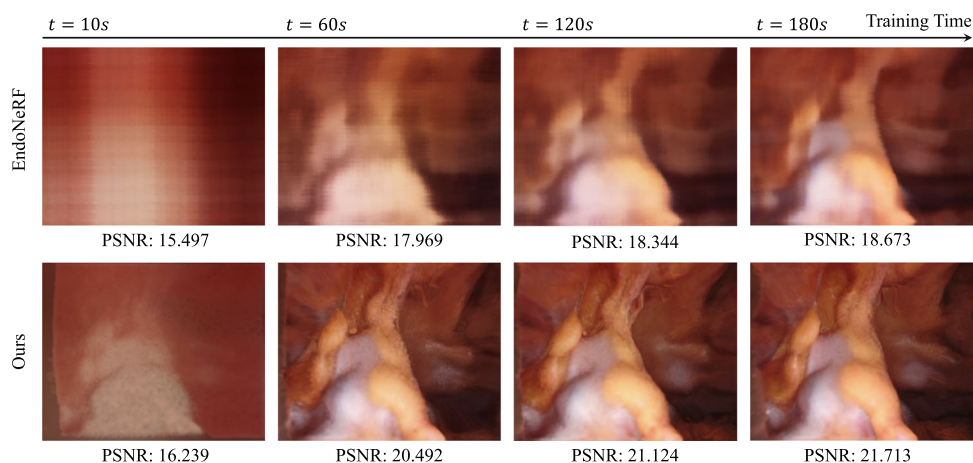
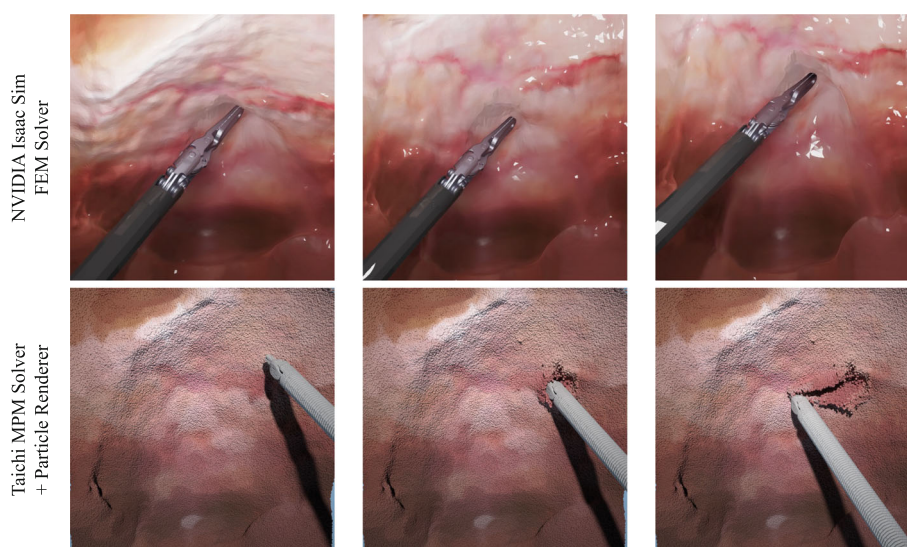


Fig. 4 Soft tissue simulation results. The first row exhibits real-time interaction between surgical tools and reconstructed soft tissues in NVIDIA Isaac Sim [15]. The second presents a simulation example of soft tissue incision with MLS-MPM algorithm [14] implemented in Taichi [13]



of achieving real-time simulations. In the simulation stage, $\sim 5M$ particles are generated for simulation, resulting in a memory consumption of around 5GB.

and is poised to open up new possibilities for next-generation surgical training and surgical robot learning.

Conclusion

We present an innovative and data-driven framework for constructing surgical simulation environments using endoscopic videos. Our approach introduces a new fast dynamic scene representation based on NeRF, which significantly accelerates the 3D reconstruction process of surgical scenes. Additionally, we propose a closed mesh extraction algorithm that converts reconstructed soft tissue surfaces into simulation objects. To demonstrate the versatility and applicability of our framework, we showcase multiple simulations of reconstructed surgical environments for diverse clinical applications. Our proposed methodology aims to inspire a significant advancement in the field of surgical simulation

Limitations and future work There are still some under-explored problems with our current methods. First, the de-occlusion of surgical tools relies on the interpolation of radiance fields, which will cause artifacts in the textures of occluded soft tissues. This could be solved by incorporating generative models to inpaint the textures. Second, as an initial trial, our simulation is based on the naive versions of FEM and MPM. In future, we aim to test more simulation algorithms on our reconstructed soft tissues, e.g., XFEM and XMPM, to achieve more realistic simulation effects.

Acknowledgements This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 24209223), in part by the Hong Kong Innovation and Technology Fund (Project No. ITS/223/22), in part by the Science, Technology and Innovation Commission of Shenzhen Munic-

ipality Project No. SGDX20220530111201008, and in part by InnoHK Multi-Scale Medical Robotics Center.

Declarations

Conflicts of interest The authors declare that they have no Conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Liu X, Stiber M, Huang J, Ishii M, Hager GD, Taylor RH, Unberath M (2020) Reconstructing sinus anatomy from endoscopic video—towards a radiation-free approach for quantitative longitudinal assessment. In: MICCAI, pp 3–13
- Chen W, Liao X, Sun Y, Wang Q (2020) Improved orb-slam based 3d dense reconstruction for monocular endoscopic image. In: ICVRV, pp 101–106
- Recasens D, Lamarca J, Fàcil JM, Montiel J, Civera J (2021) Endo-depth-and-motion: reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. *IEEE Robot Automat Lett* 6(4):7225–7232
- Wei G, Yang H, Shi W, Jiang Z, Chen T, Wang Y (2021) Laparoscopic scene reconstruction based on multiscale feature patch tracking method. In: EIECS, pp 588–592. IEEE
- Wei R, Li B, Mo H, Lu B, Long Y, Yang B, Dou Q, Liu Y, Sun D (2022) Stereo dense scene reconstruction and accurate localization for learning-based navigation of laparoscope in minimally invasive surgery. *IEEE Trans Biomed Eng* 70(2):488–500
- Long Y, Li Z, Yee CH, Ng CF, Taylor RH, Unberath M, Dou Q (2021) E-dssr: efficient dynamic surgical scene reconstruction with transformer-based stereoscopic depth perception. In: MICCAI, pp 415–425
- Wang Y, Long Y, Fan SH, Dou Q (2022) Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. MICCAI
- Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R (2020) Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV, pp 405–421
- Müller M, Heidelberg B, Hennix M, Ratcliff J (2007) Position based dynamics. *J Vis Commun Image Represent* 18(2):109–118
- Sifakis E, Barbic J (2012) Fem simulation of 3d deformable solids: a practitioner's guide to theory, discretization and model reduction. In: *Acm Siggraph 2012 Courses*, pp 1–50
- Qian K, Bai J, Yang X, Pan J, Zhang J (2017) Essential techniques for laparoscopic surgery simulation. *Comput Animat Virtual Worlds* 28(2):1724
- Qian K, Jiang T, Wang M, Yang X, Zhang J (2016) Energized soft tissue dissection in surgery simulation. *Comput Animat Virtual Worlds* 27(3–4):280–289
- Hu Y, Li T-M, Anderson L, Ragan-Kelley J, Durand F (2019) Taichi: a language for high-performance computation on spatially sparse data structures. *ACM Trans Gr(TOG)* 38(6):1–16
- Hu Y, Fang Y, Ge Z, Qu Z, Zhu Y, Pradhana A, Jiang C (2018) A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. *ACM Trans Graph(TOG)* 37(4):1–14
- Liang J, Makovychuk V, Handa A, Chentanez N, Macklin M, Fox D (2018) Gpu-accelerated robotic simulation for distributed reinforcement learning. In: *CoRL*, pp. 270–282. PMLR
- Müller T, Evans A, Schied C, Keller A (2022) Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans Graphics (ToG)* 41(4):1–15
- Sun C, Sun M, Chen H-T (2022) Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In: *CVPR*
- Fridovich-Keil S, Yu A, Tancik M, Chen Q, Recht B, Kanazawa A (2022) Plenoxels: Radiance fields without neural networks. In: *CVPR*, pp. 5501–5510
- Fridovich-Keil S, Meanti G, Warburg FR, Recht B, Kanazawa A (2023) K-planes: Explicit radiance fields in space, time, and appearance. In: *CVPR*, pp. 12479–12488
- Pumarola A, Corona E, Pons-Moll G, Moreno-Noguer F (2021) D-nerf: Neural radiance fields for dynamic scenes. In: *CVPR*, pp. 10318–10327
- Park K, Sinha U, Barron JT, Bouaziz S, Goldman DB, Seitz SM, Martin-Brualla R (2021) Nerfies: Deformable neural radiance fields. In: *ICCV*, pp. 5865–5874
- Chen A, Xu Z, Geiger A, Yu J, Su H (2022) Tensorf: Tensorial radiance fields. In: *ECCV*, pp. 333–350
- Cao A, Johnson J (2023) Hexplane: A fast representation for dynamic scenes. In: *CVPR*, pp. 130–141
- Xu H, Zhang J, Cai J, Rezatofighi H, Yu F, Tao D, Geiger A (2023) Unifying flow, stereo and depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*
- Li Z, Liu X, Drenkow N, Ding A, Creighton FX, Taylor RH, Unberath M (2021) Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In: *ICCV*, pp. 6197–6206
- Huber PJ (1992) Robust estimation of a location parameter. In: *Breakthroughs in Statistics: Methodology and Distribution*, pp. 492–518
- Wang X, Qiu Y, Slattery SR, Fang Y, Li M, Zhu S-C, Zhu Y, Tang M, Manocha D, Jiang C (2020) A massively parallel and scalable multi-GPU material point method. *ACM Trans Graph(TOG)* 39(4):30–1
- Hu Y, Zhou Q, Gao X, Jacobson A, Zorin D, Panozzo D (2018) Tetrahedral meshing in the wild. *ACM Trans Graph(TOG)* 37(4):60–1
- Hu Y, Schneider T, Wang B, Zorin D, Panozzo D (2020) Fast tetrahedral meshing in the wild. *ACM Trans Graph(TOG)* 39(4):117
- Si H (2015) Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Trans Math Softw* 41(2):11. <https://doi.org/10.1145/2629697>
- Long Y, Li C, Dou Q (2022) Robotic surgery remote mentoring via AR with 3D scene streaming and hand interaction. *Comput Methods Biomech Biomed Eng: Imag Vis* 11(4):1027–1032

32. Long Y, Wei W, Huang T, Wang Y, Dou Q (2023) Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning. *IEEE Robotics and Automation Letters*
33. Sulsky D, Zhou S-J, Schreyer HL (1995) Application of a particle-in-cell method to solid mechanics. *Comput Phys Commun.* 87(1–2):236–252
34. Wolper J, Fang Y, Li M, Lu J, Gao M, Jiang C (2019) Cd-MPM: continuum damage material point methods for dynamic fracture animation. *ACM Trans Graph (TOG)* 38(4):1–15
35. Wolper J, Chen Y, Li M, Fang Y, Qu Z, Lu J, Cheng M, Jiang C (2020) AnisoMPM: animating anisotropic damage mechanics. *ACM Trans Graph (TOG)* 39(4):37

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.