SDFPlane: Explicit Neural Surface Reconstruction of Deformable Tissues

Hao Li[1], Jiwei Shan[1], and Hesheng Wang[1(✉)]

Shanghai Jiaotong University, Shang hai, China
wanghesheng@sjtu.edu.cn

Abstract. Three-dimensional reconstruction of soft tissues from stereoscopic surgical videos is crucial for enhancing various medical applications. Existing methods often struggle to generate accurate soft tissue geometries or suffer from slow network convergence. To address these challenges, we introduce SDFPlane, an innovative method for fast and precise geometric reconstruction of surgical scenes. This approach efficiently captures scene deformation using a spatial-temporal structure encoder and combines an SDF decoder with a color decoder to accurately model the scene's geometry and color. Subsequently, we synthesize color images and depth maps with SDF-based volume rendering. Additionally, we implement an error-guided importance sampling strategy, which directs the network's focus towards areas that are not fully optimized during training. Comparative analysis on multiple public datasets demonstrates that SDFPlane accelerates optimization by over 10× compared to existing SDF-based methods while maintaining state-of-the-art rendering quality. Code is available at https://github.com/IRMVLab/SDFPlane.git

Keywords: 3D Reconstructon · Neural Fields · Robotic Surgery.

1 Introduction

Reconstructing three-dimensional models of deformable soft tissues from stereoscopic endoscopic surgical videos is a critical task. High-quality 3D models significantly enhance various downstream applications[16,12,13,8,3,14,11,23]. For example, in medical education, having detailed surface textures and accurately modeled soft tissue structures is crucial for simulating a virtual surgical environment. This simulation provides a realistic and safe platform for medical personnel to gain essential skills. Additionally, high-quality tissue reconstructions offer detailed geometric information about the target area, helping surgeons understand tissue structures more effectively. This can overcome challenges associated with limited operational fields of view, enhancing the success and safety of surgeries.

Traditional endoscopic 3D reconstruction technologies integrate SLAM for dense or semi-dense mapping of the surgical scene[16,12,13,3,14,11,23]. Nonetheless, these methods often face limitations in dynamic environments or when

Hao Li and Jiwei Shan—Equal contribution.

dealing with topological changes in the scene, thus limiting their practical application. Neural Radiance Fields (NeRF)[15] have emerged as a transformative approach, demonstrating impressive results in various domains, including medical imaging[10]. EndoNeRF[18] pioneers the application of NeRF to the reconstruction of soft tissue deformations. By leveraging both a canonical neural radiance field and a time-dependent neural displacement field, it surpasses traditional constraints, achieving notable reconstructions of deformable tissues. However, the training process of EndoNeRF is computationally demanding, as rendering each pixel requires multiple network queries, leading to slow network convergence.

In recent efforts to advance neural rendering for three-dimensional soft tissue reconstruction, many NeRF-based endoscopic reconstruction methods have been developed[21,20,1]. Among these, Lerplane[20] employs a hybrid approach to model dynamic scenes, which speeds up optimization compared to purely MLP-based methods. However, Lerplane's reliance on volume rendering to depict scene geometry and appearance is primarily tailored for new perspective synthesis, not surface reconstruction[17]. As a result, Lerplane learns only a volume density field, struggling to achieve high-quality geometric reconstruction and thereby limiting its performance in reconstructions. While EndoSurf[21] and LightNeus[1] have explored using signed distance function fields to describe soft tissue geometry, EndoSurf suffers from excessively long training times, often exceeding ten hours, and LightNeus is designed for static scenes. These limitations significantly impede their practical application in surgical settings.

To overcome these challenges, we present SDFPlane, a novel method designed for the efficient and accurate reconstruction of deformable tissues in surgical environments. Our method integrates a spatiotemporal structure encoder, an SDF (Signed Distance Function) decoder, and a color decoder to precisely capture the deformation, geometry, and color of surgical scenes. The encoder employs a multi-resolution HexPlane to capture the deformation of soft tissues across different timeframes. The SDF decoder produces an SDF field, providing an accurate geometric structure of the scene by identifying the surface at a zero-level set. A regularization strategy, informed by SDF properties, ensures the accurate learning of surfaces. The color decoder is tasked with scene color prediction, complemented by SDF-based volume rendering for synthesizing color images and depth maps. Model parameters are optimized by minimizing the discrepancy between actual outcomes and those rendered. Moreover, we introduce an error-guided importance sampling strategy. This approach directs the network's focus towards under-optimized areas during training, thereby enhancing result quality and speeding up the optimization process. Our contributions can be summarized as follows:

1. A novel method for deformable tissue reconstruction that offers superior rendering quality, precise geometry reconstruction, and is 10 times faster in training than previous SDF-based approaches.
2. An innovative error-guided importance sampling strategy to enhance optimization and rendering quality. This method dynamically adjusts the sam-

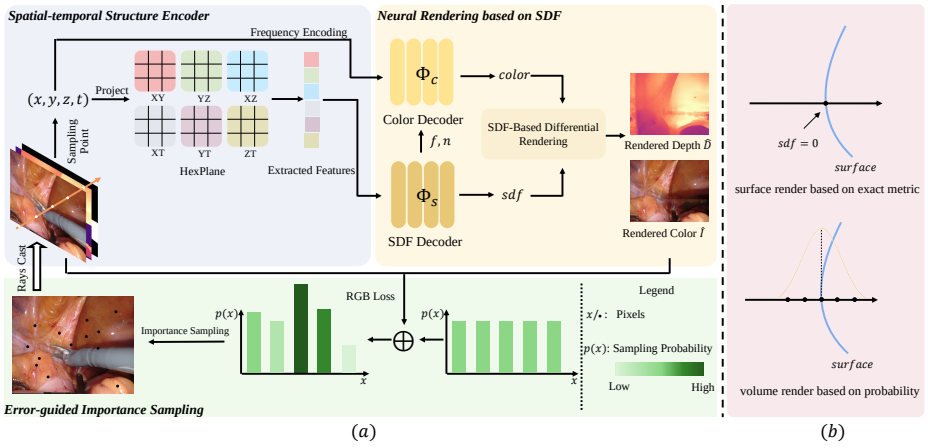

Fig. 1. (a) Illustration of our proposed high-quality and precise reconstruction method, SDFPlane. (b) Difference between SDF-based and density-based structure.

pling probability of pixels and is not restricted by camera movement, making it more broadly applicable.

3. We conducted extensive experiments on the EndoNeRF[18] dataset and StereoMIS[7] dataset, demonstrating that our method surpass all previous endoscopic reconstruction methods in both quantitative and qualitative performance.

2 Method

2.1 Overview

Given a sequence of continuous video from a stereoscopic endoscope, our goal is to reconstruct the shape and texture of deformed soft tissues with high fidelity. Following existing NeRF-based endoscopic reconstruction algorithms[18,20,21], we define our input as a sequence of frame data $\{(I_i, D_i, M_i, P_i, t_i)\}_{i=1}^{T}$. Here, $I_i \in \mathbb{R}^{H\times W\times 3}$ and $D_i \in \mathbb{R}^{H\times W}$ respectively denote the left RGB image and dense depth map for the i-th frame, where H and W are the height and width. The foreground mask $M_i \in \mathbb{R}^{H\times W}$ is utilized to filter out extraneous pixels from surgical tools, blood, and smoke. $P_i \in \mathbb{R}^{4\times 4}$ specifies the image pose for each frame. T signifies the total frame count, with $t_i = i/T$ normalizing each frame's timestamp to the range $[0, 1]$.

Fig. 1 shows the pipeline of our method. SDFPlane is composed of three main components: a spatiotemporal structure encoder, SDF-based differentiable rendering, and error-guided importance sampling. Initially, error-guided importance sampling pinpoints high-priority tissue pixels to formulate corresponding rays (Sec. 2.2). Subsequently, sampling along each ray yields points and their feature vectors, extracted via a spatiotemporal structure encoder (Sec. 2.3). These

feature vectors, combined with encoded coordinate time information, feed into the color and SDF decoders to determine the color and SDF values at each sampling point (Sec. 2.4). Ultimately, the color and depth map for each chosen ray are deduced through SDF-based differentiable rendering, culminating in the formulation of a loss function to refine the entire model (Sec. 2.4).

2.2 Error-Guided Importance Sampling

When sampling pixels from the training dataset, the state of the deformation points varies over time, necessitating an increased sampling frequency. Conversely, the sampling frequency for static points, which remain unchanged across different timestamps, should be reduced to enhance training efficiency. Lerplane[20] addresses this issue by preprocessing the single-view dataset to adjust the sampling weights of the points. Nevertheless, the effectiveness of this method in clinical settings warrants further investigation.

Specifically, the sampling strategy of Lerplane[20] has two main weaknesses. Firstly, Lerplane's sampling strategy uses a prior mask from input images to focus learning on masked areas, but it lacks the ability to adjust based on the model's learning progress. Secondly, the implementation of this sampling involves averaging all images and measuring the differences between each image and the average to find areas with significant deformation. This approach is only suitable for single-viewpoint datasets, as the average image loses its meaning when the camera moves.

To overcome these challenges, we introduce an error-guided importance sampling strategy, which assigns sampling weights based on each pixel's color loss. Specifically, before the initial training iteration, we assign a uniform sampling probability to all pixels (excluding those occluded by the surgical tool) using an initial value β, which is a hyperparameter. Then, in each subsequent iteration, we record the color loss of each sampled pixel to adjust its sampling probability. The higher the loss, the higher the likelihood of being sampled. Formally, at iteration t, the sampling weight W_r^t of the sampled pixel r can be expressed as:

$$W_r^t = \frac{L_r^{t-1}}{\sum_{r \in R} L_r^{t-1}} \tag{1}$$

where L_r^{t-1} represents the rendering loss of pixel/ray r from the most recent sampling. Specifically, if ray/pixel r has been sampled in previous optimization steps, L_r^{t-1} corresponds to its color rendering loss from the most recent iteration; otherwise, if ray/pixel r has never been sampled, L_r^{t-1} is set to the initial value. R denotes all pixels in the training set. Moreover, to mitigate the impact of high loss at noise points, for the first few iterations, we uniformly sample from areas not occluded by the mask. After this warm-up phase, we shift to using error-guided importance sampling.

Therefore, our method can update in real-time based on previous loss data, allowing the sampling weights to be adjusted according to the learning progress. This enables our method to adapt and refine the sampling process throughout the

training period, which minimizes computational resource wastage and enhances model performance. Besides, our proposed method is not restricted by camera movement, making it more broadly applicable.

2.3 Spatial-Temporal Structure Encoder

To process dynamic scenes more effectively, we introduce a spatio-temporal structure encoder, leveraging a multi-resolution six-plane framework[5,2]. This approach decomposes the 4D deformation space into six multi-resolution grid planes, enabling the encoding of spatial points within 2D feature grids. These planes include XY, XZ, YZ for spatial dimensions, and XT, YT, ZT for spatiotemporal dimensions. Specifically, we project the 4D spatiotemporal coordinates $q = (x, y, z, t)$ onto these six planes at a resolution l. On each plane, a projected point is allocated to a grid. Through bilinear interpolation $\mathcal{B}$ performed on the features at the grid's four vertices, we obtain the point's feature on that plane. Formally, this process can be expressed as:

$$f(q)_g^l = \mathcal{B}(F_g^l, \pi_g(q)). \tag{2}$$

where $g \in G = \{XY, XZ, YZ, XT, YT, ZT\}$, π_g projects q onto the g'th plane.

We combine these features over the six planes using the Hadamard product[5] to produce a final feature vector at this resolution l. Finally, the final feature vector is obtained by concatenating feature vectors of different resolutions:

$$f(q) = \bigcup_l \prod_{g \in G} f(q)_g^l. \tag{3}$$

2.4 SDF-Based Volume Rendering and Optimization

Color and SDF Decoder At a given time t, with the camera origin o and ray direction r, we sample N points $x_i = o + d_i r$, for $i \in \{1, \ldots, N\}$ along the ray. From each sampled point x_i, we derive feature vectors $f(x_i)$ via the spatio-temporal structure encoder. Then the SDF decoder Φ_s predicts the SDF value s_i and a feature vector h_i, and the color decoder Φ_c estimates the RGB value c_i:

$$\Phi_s(f(x_i)) \to (s_i, h_i), \quad \Phi_c(\gamma(x_i), n_i, h_i) \to c_i. \tag{4}$$

Here, $\gamma(*)$ signifies position encoding[15], and the surface normal n_i, calculated from the SDF gradient $n = \nabla_{x_i} s_i$, is incorporated because the proximity of normals at adjacent sampling points generally leads to similar color outputs from the color decoder. This approach effectively regularizes the output SDF, ensuring consistency in the rendering process.

SDF-Based Volume Rendering Following this procedure, we acquire SDF prediction values $\{s_1, s_2, \ldots, s_N\}$ and color prediction values $\{c_1, c_2, \ldots, c_N\}$ for all sampling points along a ray. Following [9,17], depth and color are rendered

by integrating these predicted values along the sampled rays. Specifically, SDF values are first transformed into opacity values[17]:

$$\alpha_i = \max(0, (\phi(s_i) - \phi(s_{i+1}))/\phi(s_i)), \quad \phi(s) = (1 + e^{-s/a})^{-1} \tag{5}$$

where a represents a trainable standard deviation. These opacity values facilitate the rendering of each ray's color and depth:

$$T_i = \prod_{j=1}^{i-1}(1 - \alpha_j), \quad \hat{C}(r) = \sum_{i=1}^{N} T_i \alpha_i c_i, \quad \hat{D}(r) = \sum_{i=1}^{N} T_i \alpha_i d_i \tag{6}$$

Fig. 1(b) illustrates the theoretical differences between SDF-based and density-based rendering approaches. SDF-based methods uniquely identify surface points by leveraging the property $sdf = 0$. In contrast, density-based NeRF networks consider that, along a sampled ray, higher density values indicate a higher probability of the point being on the object surface, lacking an accurate and consistent metric to guide surface generation. Therefore, density-based networks often result in significant noise during 3D reconstruction.

Loss Functions The parameters of the feature plane, color decoder, and SDF decoder are optimized by minimizing the overall loss function, which is defined as follows:

$$L = \lambda_1 L_{color} + \lambda_2 L_{depth} + \lambda_3 L_{eikonal} + \lambda_4 L_{tv} + \lambda_5 L_{sm}. \tag{7}$$

where $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$ are the weighting coefficients. L_{color} and L_{depth} quantify the discrepancies between rendered and actual color and depth values, and we employ color and depth optimization by RMSE loss. We incorporate the Eikonal term[4,6] to regularize SDF values within the 3D space:

$$L_{eikonal} = \sum_{x \in X} (\|\nabla \Phi_s(x)\|_2 - 1)^2. \tag{8}$$

where X represents the set of sampling points on each ray. Moreover, endoscopic soft tissue deformation reconstruction is a serious pathological problem, needing stronger regularizers. Inspired by [5,2], a 2D total variation loss L_{tv} and a smoothness regularizer L_{sm} are used, where L_{tv} is applied to the spatial planes, and L_{sm} is utilized along the temporal dimension of the space-time planes.

3 Experiments

3.1 Datasets and Evaluation Metrics

We validated the effectiveness of our proposed method on the EndoNeRF[18] dataset and the StereoMIS dataset[7]. EndoNeRF provides two stereo datasets capturing simple surgical procedures from a fixed intraoperative viewpoint. The

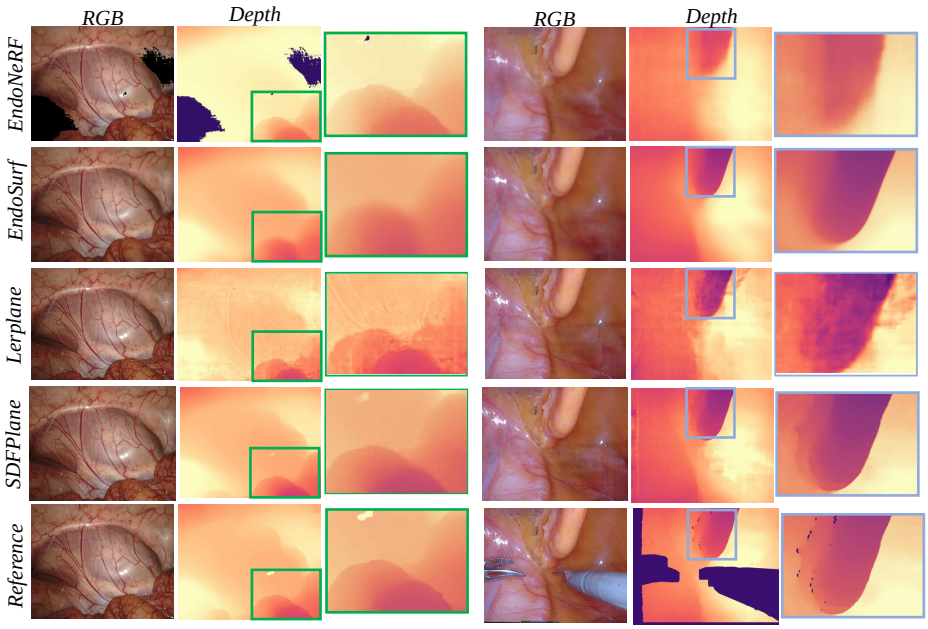

Fig. 2. Qualitative analysis of 2D rendering results in the EndoNeRF's pulling datasets[18] and StereoMIS dataset's P2_8 sequence[7].

dataset includes depth maps corresponding to the images and masks for covering surgical tools. The StereoMIS dataset[7] consists of a series of stereo videos depicting intraoperative surgeries. We extracted 100 frames from two videos (P2_1 and P2_8) with a fixed viewpoint and utilized the built-in depth estimation network of StereoMIS to obtain depth maps. We compared our SDFPlane model with EndoNeRF[18], EndoSurf[21], and Lerplane[20], which are the state-of-the-art NeRF-based methods. Following common practice, we employ PSNR, SSIM[19], and LPIPS[22] as metrics to assess image rendering quality.

3.2 Implementation Details

We implemented our framework using PyTorch, with the multi-resolution six-plane framework selecting resolutions of 64, 128, 256, and 512. The dimension of feature vectors extracted at each resolution is set to 16. The SDF decoder is a tiny MLP with one layer and 64 neurons, while the color decoder is a tiny MLP with two layers and 64 neurons. During sampling, we set $\beta = 0.0003$ and the warm-up rounds to 2000. We assigned $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 0.001, \lambda_4 = 0.0001, \lambda_5 = 0.0001$ to achieve the best performance of the total model. The initial learning rate is set to 0.01, and the initial standard deviation a is set to 0.35. For all experiments, we maintained a uniform batch size of 2048 and conducted training for 9600 epochs. All experiments were performed on a single RTX 4090 GPU.

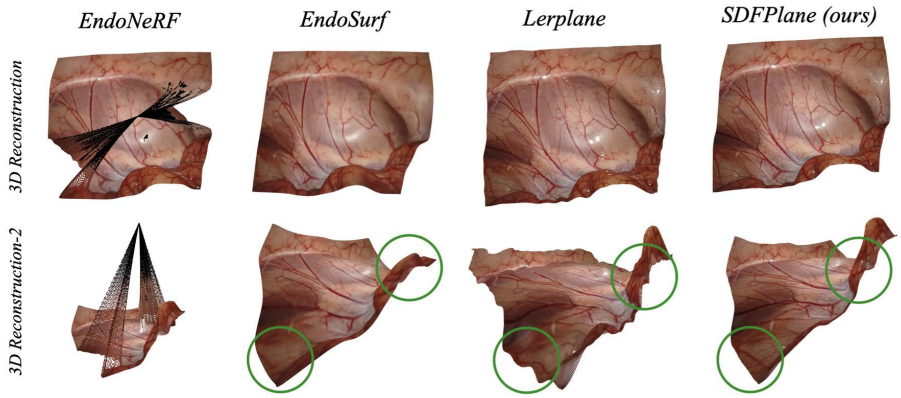

Fig. 3. Qualitative analysis of 3D reconstructing results in the StereoMIS dataset's P2_8 sequence[7] from two distinctive perspectives.

3.3 Qualitative and Quantitative Results

Table 1. Quantitative results of EndoNeRF[18] and StereoMIS[7] dataset on metrics of PSNR, SSIM and LPIPS.

Methods	ENDONERF-cutting			ENDONERF-pulling			StereoMIS-1			StereoMIS-2			
Metrics	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	TIME
EndoNeRF	25.7634	0.8548	0.2225	27.6759	0.8945	0.1717	27.2859	0.5772	0.4440	17.8167	0.6179	0.338	≥5h
EndoSurf	29.8103	0.8958	0.2429	31.4422	0.921	0.2251	25.956	0.5545	0.516	28.2647	0.7449	0.4265	≥5h
LerPlane	35.1085	0.9198	0.0925	37.9586	0.9501	0.0622	31.3784	0.7933	0.1366	34.5649	0.8956	0.0696	10min
SDFPlane w/o sample	35.7176	0.928	0.0767	38.9014	0.9569	0.0524	31.6454	0.8059	0.1257	34.8484	0.9018	0.0675	-
SDFPlane(ours)	**36.2011**	**0.9342**	**0.0606**	**39.7202**	**0.9601**	**0.0379**	**32.2624**	**0.8382**	**0.083**	**35.7656**	**0.9179**	**0.0464**	34min

Table 1 presents the performance of four methods on four datasets, with SDFPlane consistently surpassing the previous three approaches across all quality metrics. Due to the introduction of additional gradient backpropagation, our training time is slower than Lerplane[20], but compared to the SDF-based EndoSurf[21], our training speed has significantly improved.

Fig. 2 displays the RGB and depth images rendered by four methods on the EndoNeRF dataset [18] and the StereoMIS dataset [7]. Fig. 3 illustrates the 3D reconstruction visualization, offering views from two different angles for a more thorough comparison. Notably, EndoNeRF[18] shows black holes on both sides of the central tissue in the P2_8 dataset, highlighting areas of suboptimal reconstruction. EndoSurf[21], with a limited number of training iterations, does not capture sufficient detail in the color map, leading to blurred local details and depth map contours in P2_8. While Lerplane[20] provides detailed color fitting, its depth map introduces excessive noise, causing visible ripples on the tissue surface. In contrast, SDFPlane's visualizations significantly outperform these

methods. SDFPlane not only yields high-quality, detailed color images but also produces depth maps with sharp outlines. Moreover, in 3D reconstruction, SDFPlane addresses Lerplane's issue of multi-wave surfaces, achieving a smoother overall surface reconstruction.

3.4 Ablation Study

To demonstrate the effectiveness of key components, we conducted ablation studies, and Table 1 presents the results. Retaining both the Spatial-Temporal Structure Encoder and the SDF-based neural rendering module (SDFPlane w/o sample), our method outperforms current state-of-the-art algorithms, showing the effectiveness of the proposed key components. Additionally, integrating the error-guided importance sampling module (SDFPlane(ours)) further boosts system performance, showing the efficiency of our proposed pixel sampling approach.

4 Conclusion

In this paper, we introduce SDFPlane, an efficient and precise method for reconstructing deformable tissues from endoscopic videos. Utilizing a spatiotemporal structure encoder based on multi-scale planar composition alongside SDF-based volume rendering, our approach achieves higher-quality reconstructions. Notably, our method's speed surpasses that of existing SDF-based methods by more than tenfold. Moreover, the error-guided importance sampling strategy we propose significantly enhances the system's reconstruction accuracy and convergence speed. Our experimental results demonstrate that SDFPlane excels in various intraoperative settings. We believe our work will contribute valuable insights into deformable soft tissue reconstruction and foster advancements in fields like surgical robotics.

Acknowledgments. Research supported by the General Program of National Natural Science Foundation of China (Grant No.62361166632, 62225309, 62073222 and U21A20480).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Batlle, V.M., Montiel, J.M., Fua, P., Tardós, J.D.: Lightneus: Neural surface reconstruction in endoscopy using illumination decline. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 502–512. Springer (2023)
2. Cao, A., Johnson, J.: Hexplane: A fast representation for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 130–141 (2023)

3. Chen, L., Tang, W., John, N.W., Wan, T.R., Zhang, J.J.: Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality. Computer methods and programs in biomedicine **158**, 135–146 (2018)
4. Crandall, M.G., Lions, P.L.: Viscosity solutions of hamilton-jacobi equations. Transactions of the American mathematical society **277**(1), 1–42 (1983)
5. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: K-planes: Explicit radiance fields in space, time, and appearance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12479–12488 (2023)
6. Gropp, A., Yariv, L., Haim, N., Atzmon, M., Lipman, Y.: Implicit geometric regularization for learning shapes. arXiv preprint arXiv:2002.10099 (2020)
7. Hayoz, M., Hahne, C., Gallardo, M., Candinas, D., Kurmann, T., Allan, M., Sznitman, R.: Learning how to robustly estimate camera pose in endoscopic videos. International journal of computer assisted radiology and surgery pp. 1–8 (2023)
8. Huang, Q., Wang, Y., Luo, H., Li, J.: On mimicking human's manipulation for robot-assisted spine ultrasound imaging. Robotic Intelligence and Automation **43**(4), 373–381 (2023)
9. Johari, M.M., Carta, C., Fleuret, F.: Eslam: Efficient dense slam system based on hybrid representation of signed distance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17408–17419 (2023)
10. Khan, M.O., Fang, Y.: Implicit neural representations for medical imaging segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 433–443. Springer (2022)
11. Ma, R., Wang, R., Pizer, S., Rosenman, J., McGill, S.K., Frahm, J.M.: Real-time 3d reconstruction of colonoscopic surfaces for determining missing regions. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. pp. 573–582. Springer (2019)
12. Mahmoud, N., Cirauqui, I., Hostettler, A., Doignon, C., Soler, L., Marescaux, J., Montiel, J.M.M.: Orbslam-based endoscope tracking and 3d reconstruction. In: Computer-Assisted and Robotic Endoscopy: Third International Workshop, CARE 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 3. pp. 72–83. Springer (2017)
13. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE transactions on medical imaging **38**(1), 79–89 (2018)
14. Marmol, A., Banach, A., Peynot, T.: Dense-arthroslam: Dense intra-articular 3-d reconstruction with robust localization prior for arthroscopy. IEEE Robotics and Automation Letters **4**(2), 918–925 (2019)
15. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
16. Schmidt, A., Mohareri, O., DiMaio, S., Yip, M.C., Salcudean, S.E.: Tracking and mapping in medical computer vision: A review. Medical Image Analysis p. 103131 (2024)
17. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. arXiv preprint arXiv:2106.10689 (2021)

18. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 431–441. Springer (2022)
19. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
20. Yang, C., Wang, K., Wang, Y., Yang, X., Shen, W.: Neural lerplane representations for fast 4d reconstruction of deformable tissues. arXiv preprint arXiv:2305.19906 (2023)
21. Zha, R., Cheng, X., Li, H., Harandi, M., Ge, Z.: Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 13–23. Springer (2023)
22. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
23. Zhou, H., Jayender, J.: Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. pp. 331–340. Springer (2021)