SurgicalGS: Dynamic 3D Gaussian Splatting for Accurate Robotic-Assisted Surgical Scene Reconstruction

Jialei Chen[1], Xin Zhang[2], Mobarak I. Hoque[3], Francisco Vasconcelos[3], Danail Stoyanov[3], Daniel S. Elson[1], and Baoru Huang[1,3,4]

[1]The Hamlyn Centre for Robotic Surgery, Department of Surgery and Cancer, Imperial College London
[2]Department of Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology
[3]UCL Hawkes Institute, University College London
[4]Department of Computer Science, University of Liverpool

Abstract. Accurate 3D reconstruction of dynamic surgical scenes from endoscopic video is essential for robotic-assisted surgery. While recent 3D Gaussian Splatting methods have shown promise in achieving high-quality reconstructions with fast rendering speeds, their use of inverse depth loss functions compresses depth variations. This can lead to a loss of fine geometric details, limiting their ability to capture precise 3D geometry and effectiveness in intraoperative applications. To address the limitations of existing methods, we developed **SurgicalGS**, a dynamic 3D Gaussian Splatting framework specifically designed for improved geometric accuracy in surgical scene reconstruction. Our approach integrates a temporally coherent multi-frame depth fusion and an adaptive motion mask for Gaussian initialisation. Besides, we represent dynamic scenes using the Flexible Deformation Model and introduce a novel normalized depth regularization loss and an unsupervised depth smoothness constraint to ensure high geometric accuracy in the reconstruction. Extensive experiments on two real surgical datasets demonstrate that SurgicalGS achieves state-of-the-art reconstruction quality, especially in precise geometry, advancing the usability of 3D Gaussian Splatting in robotic-assisted surgery. Our code is available at https://github.com/neneyork/SurgicalGS.

Keywords: 3D Reconstruction · Surgical Scene reconstruction · 3D Gaussian Splatting.

1 Introduction

3D reconstruction of surgical scenes is an important element of robotic-assisted surgery. By creating a 3D model of the observed tissues, it allows for more precise instrument control, enabling a range of downstream applications such as intraoperative navigation [9, 16], robotic surgery automation [8, 10, 17, 26], and

virtual reality simulation [4]. Traditional methods estimate depth with stereo matching [19, 27] or combine simultaneous localization and mapping (SLAM) to fuse depth maps for surgical scene reconstruction [19, 27, 28]. However, these methods either assume that scenes are static or surgical tools are absent, limiting their effectiveness in intraoperative applications.

Neural Radiance Fields (NeRF) [15] show significant progress in scene reconstruction with implicit representations. This technology leverages volume rendering [5] to convert 2D images to 3D scenes. However, the original NeRF method suffers from long training times, low rendering speed, and a lack of explicit geometric representation [3]. Although some methods adapt discrete structures, such as planes [2] and voxel grids [6], to reduce training time from days to minutes, rendering speed remains insufficient for practical use in surgical scenarios.

Recently, 3D Gaussian Splatting (3DGS) [11] can represent scenes as an explicit 3D Gaussian model, significantly improving the rendering speed to real-time. Many methods [14,24] have extended 3DGS to surgical scene reconstruction and demonstrated satisfactory results. However, they fail to capture the precise 3D geometry, which limits their effectiveness in intraoperative use. To address these limitations, we focus on two key questions for reconstructing dynamic surgical scenes: (i) How can depth priors be effectively utilized to initialize Gaussian points? (ii) How can depth priors be leveraged to provide accurate depth constraints in dynamic scenes?

In this paper, we propose a new method for accurately reconstructing the dynamic surgical scene, namely **SurgicalGS**. We first integrate a temporally coherent multi-frame depth fusion framework to reveal tissues, combined with an adaptive motion masking that jointly suppresses depth noise and transient artifacts during the Gaussian initialization. Then, Flexible Deformation Modeling [24] is adopted to represent dynamic scenes. To further enhance depth supervision, we introduce a normalized depth regularization loss and a depth smoothness constraint to ensure accurate geometric reconstruction. We evaluated our method on two public datasets, **EndoNeRF** [22] and **StereoMIS** [7]. Extensive experiments demonstrate the efficacy of our approach, achieving superior geometry accuracy (RMSE: 1.820 mm) and reconstruction quality (PSNR: 38.18).

2 Methodology

2.1 Preliminaries

3D Gaussian Splatting 3DGS [11] uses a differentiable Gaussian representation to model static scenes, allowing for rapid rasterization and fast image rendering. Each 3D Gaussian consists of learnable attributes: position μ, rotation r, scaling s, opacity o, and spherical harmonic (SH) coefficients. For any 3D point x in world coordinates, the impact of a 3D Gaussian on x is defined by the Gaussian distribution:

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), \tag{1}$$

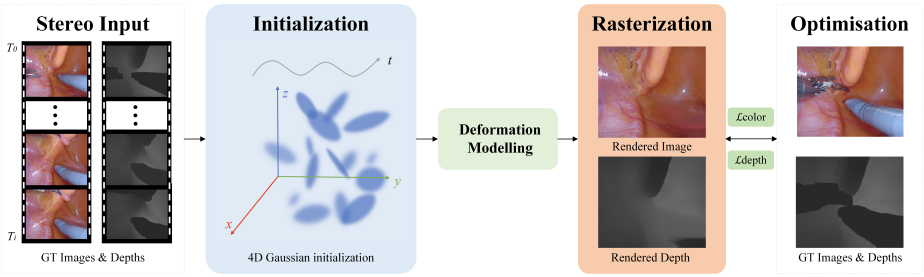

Fig. 1. Proposed framework. Starting from an image and depth sequence, we associate 3D points between frames to initialize the Gaussian point cloud. Then, given a specific time t, the attributes of the Gaussians are modified by the deformation model and rendered as a 2D image and depth map. Finally, color and depth loss are employed to refine results.

$$\Sigma = RSS^TR^T, \tag{2}$$

where Σ is the covariance matrix, R is a rotation matrix, and S is a diagonal scaling matrix. Then, the 3D Gaussian is projected onto the 2D image planes for rendering. The covariance matrix after projection is calculated by $\Sigma' = JW\Sigma W^TJ^T$, where J is the Jacobian of the affine approximation of the projective transformation and W represents the viewing transformation matrix. The final rendering equation for the color $\hat{C}(x)$ and the depth $\hat{D}(x)$ of the pixel x is:

$$\hat{C}(x) = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_i), \hat{D}(x) = \sum_{i \in N} d_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_i), \tag{3}$$

$$\alpha_i = o_i G_i^{2D}(x), \tag{4}$$

where c_i is the color defined from the SH coefficients, d_i is the z-axis in the camera coordinate, and α_i is the density calculated by multiplying the projected 3D Gaussian with the opacity o_i.

Dynamic Representation Following [24], we further represent the dynamic surgical scene using a deformation field, where Fourier and polynomial basis functions $\tilde{b}(t;\theta,\sigma)$ are utilized to learn the motion curve of each Gaussian:

$$\tilde{b}(t;\theta,\sigma) = \exp\left(-\frac{1}{2\sigma^2}(t-\theta)^2\right), \tag{5}$$

where t represents time, and θ and σ denote the learnable center and variance, respectively. Each Gaussian is associated with a set of learnable weights ω and parameters Θ^μ, Θ^r, and Θ^s. These weights and parameters linearly combine the basis functions to represent deformations of position, rotation, and scale, respectively. Taking the position change along the x-axis for an illustration, the

deformation can be expressed with a set of parameters $\Theta_x^\mu = \{\omega_x^\mu, \theta_x^\mu, \sigma_x^\mu\}$ as:

$$\psi^{\mu,x}(t;\Theta^{\mu,x}) = \sum_{j=1}^{B} \omega_j^{\mu,x} \tilde{b}(t;\theta_j^{\mu,x}, \sigma_j^{\mu,x}). \tag{6}$$

The x-axis position at time t can be expressed as:

$$\mu_x(t) = \mu_x(t) + \psi^{\mu,x}(t). \tag{7}$$

2.2 Proposed Method: SurgicalGS

Our proposed SurgicalGS integrates geometric information from all frames for dense Gaussian initialization to recover tissues occluded by tools and improve reconstruction quality. Additionally, it employs normalized depth loss and unsupervised depth smoothness to enhance the accuracy of geometric reconstruction.

Dense Initialization with Depth Priors The original 3DGS [11] used the SfM algorithm [18] to generate initialized point clouds for further reconstruction. However, due to limited viewpoints, dynamic lighting conditions, and tool occlusions in surgical environments, SfM cannot generate accurate point clouds in surgical scenes. Thus, we employ the depth map, tissue mask, and camera parameters to extract point clouds of tissues for each frame as:

$$P_i = K_1^{-1} K_2^{-1} D_i (I_i \odot M_i), \tag{8}$$

where P_i denotes the 3D point cloud of the i-th frame, and D_i, I_i, and M_i represent the depth map from stereo-matching, input image, and binary tissue mask for the i-th frame, respectively, K_1 and K_2 refer to the known camera intrinsic and extrinsic matrices, respectively, and $\odot$ denotes element-wise multiplication. However, the point cloud is incomplete due to surgical tool occlusions. We observe that the occluded tissue in the i-th frame could be visible in some of the other frames. Furthermore, a dense initialization of the Gaussian point cloud can more accurately represent the scene [11], leading to improved reconstruction quality for deformable tissues. Based on these observations, we design an adaptive motion mask B_i, that extracts pixels with significant depth variation and occluded tissues, to suppress depth noise and transient artifacts:

$$B_i = \mathbb{I}\left(|D_0 - D_i| > \tau\right) \cup (1 - M_0) \cap M_i, \tag{9}$$

where $\mathbb{I}(\cdot)$ refers to the indicator function, M_0 denotes the tissue mask of the first frame, and τ is an adaptive threshold, that is controlled by the depth map range, to define the significant depth variation. We uniformly downsample the point clouds to reduce the number of points and fuse them to initialize the Gaussian point cloud as follows:

$$P = \{P_0, P_1 \odot B_1, ..., P_T \odot B_T\}, \tag{10}$$

where T refers to the frame length.

Normalized Depth Regularization To ensure accurate alignment between the predicted and actual depth maps, depth regularization is applied to supervise the 3D reconstruction learning process. A common approach is to use $\mathcal{L}_1$ depth loss. However, in dynamic scenes, we observe that directly applying $\mathcal{L}_1$ depth loss results in an overly dense Gaussian point cloud, causing memory overflow and instability in training.

Previous methods [14, 24] incorporate inverse depth maps into the loss computation, effectively stabilizing the optimization process. However, inverse depth maps compress the dynamic range of depth values, reducing the disparity between the binocular and rendered depth maps. This compression minimizes the risk of over-density and enhances the stability of the optimization process. However, there is little variation in the depth map in endoscopic videos. Using inverse depth maps can overly homogenize the depth values, resulting in inaccurate and inconsistent rendered depth maps.

Our observation is that normalization can bring both binocular and rendered depth maps on a consistent scale, ensuring training stability while preserving depth variability. Our normalized depth loss is formulated as:

$$\mathcal{L}_{\hat{\mathrm{D}}} = \|M \odot (\hat{D}_{\mathrm{norm}} - D_{\mathrm{norm}})\|. \tag{11}$$

Unsupervised Depth Smoothness In surgical scenes, specular highlights, homogeneous surfaces, and large disparity discontinuities make it difficult for stereo-matching algorithms to establish accurate correspondences, leading to noise in the depth maps [20]. To remove the influence of noise and enforce the smoothness of the rendered depth, we employ a total variation loss. In addition, to preserve depth details, we apply the Canny edge detector [1] as a mask to prevent the regularization of edges with significant depth variations. We regularize the difference between a pixel $\hat{D}_{i,j}$ in depth map and its adjacent pixel as:

$$\mathcal{L}_{\mathrm{smooth}} = \frac{1}{|\hat{D}|} \sum_{i,j} \mathbb{I}_{\mathrm{ne}}(\hat{D}_{i,j}) \cdot \left(|\hat{D}_{i,j} - \hat{D}_{i+1,j}| + |\hat{D}_{i,j} - \hat{D}_{i,j+1}| \right) \tag{12}$$

where $\mathbb{I}_{\mathrm{ne}}(\hat{D}_{i,j})$ is the result of Canny edge detection indicating whether the pixel (i, j) is on the edge.

Total Loss Function The final loss for optimization is defined as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{color}} + \underbrace{\left(\mathcal{L}_{\hat{\mathrm{D}}} + \lambda_{\mathrm{smooth}} \mathcal{L}_{\mathrm{smooth}} \right)}_{\mathcal{L}_{\mathrm{depth}}} \tag{13}$$

where $\mathcal{L}_{\mathrm{color}}$ is the original photometric loss in 3DGS [11].

3 Experiments

3.1 Experiment Setting

Dataset and Evaluation Metrics We evaluate the proposed method and compare it with existing methods on two public datasets: (i) the EndoNeRF

Table 1. Quantitative evaluation of our SurgicalGS method against existing methods in endoscopic scene reconstruction. 'Speed' denotes the rendering speed (FPS). The optimal and suboptimal results are shown in **bold** and underlined respectively. The unit of depth metrics is millimeters.

Data	Method	ABS REL ↓	SQ REL ↓	RMSE ↓	PSNR ↑	SSIM↑	LPIPS ↓	Speed ↑
EndoNeRF	EndoNeRF [22]	0.0232	0.4999	3.229	35.63	0.941	0.153	0.04
	EndoSurf [25]	0.0226	0.5896	3.075	34.91	0.953	0.112	0.04
	LerPlane [23]	0.0351	0.5717	5.734	35.00	0.927	0.099	0.93
	EndoGaussian [14]	0.0340	0.2362	2.926	37.71	0.958	0.062	148.35
	Deform3DGS [24]	0.0324	0.2810	3.275	**38.39**	**0.962**	**0.059**	**332.52**
	SurgicalGS (Ours)	**0.0219**	**0.1115**	**1.820**	38.18	0.960	0.062	194.80
StereoMIS	EndoNeRF [22]	0.0315	0.7183	3.022	28.79	0.809	0.266	0.06
	EndoSurf [25]	0.0189	0.4031	2.457	29.36	0.861	0.211	0.05
	LerPlane [23]	0.0354	1.0906	5.521	29.09	0.789	0.179	0.95
	EndoGaussian [14]	0.0292	0.6057	5.050	31.02	0.878	**0.132**	130.15
	Deform3DGS [24]	0.0330	0.7547	4.888	**31.61**	**0.888**	0.135	**308.66**
	SurgicalGS (Ours)	**0.0082**	**0.0391**	**2.174**	31.54	0.885	0.148	214.06

dataset [22], which contains *in vivo* prostatectomy data captured from stereo cameras at a single viewpoint and provides estimated depth maps with stereo matching [13] and manually labeled tool masks. (ii) the StereoMIS dataset [7], which is a stereo video dataset captured from *in vivo* porcine subjects containing large tissue deformations. We estimate depth maps using the pre-trained RAFT model [21]. Following [25], we divide frames of each scene into training and testing sets with a 7:1 ratio. We use PSNR, SSIM, and LPIPS to evaluate the similarity between actual and rendered images, common depth metrics, similar to [12], to measure the quality of the depth map, and frames per second (FPS) to evaluate reconstruction efficiency.

Implementation Details All the experiments are conducted on the RTX3090 GPU and the PyTorch framework. We employ the Adam optimizer with an initial learning rate of 1.6×10^{-3}. For all scenes, the model is trained for 6K iterations with the same loss function and initialization strategy. We set $\lambda_{\text{smooth}} = 0.0001$ and apply 17 learnable Gaussian basis functions to compose the Flexible Deformation Model in our experiments.

3.2 Quantitative and Qualitative Results

As listed in Table 1, although EndoNeRF [22], EndoSurf [25], and LerPlane [23] effectively reconstruct deformable tissues, they suffer from low rendering speed and struggle with rendering high-quality images, which limits their effectiveness for real-time surgical scene reconstruction. On the other hand, EndoGaussian [14] and Deform3DGS [24] improve image quality and rendering speed based on 3DGS, but they fail to reconstruct accurate depth maps, limiting their intraoperative reliability. Benefiting from normalized depth regularization, our method

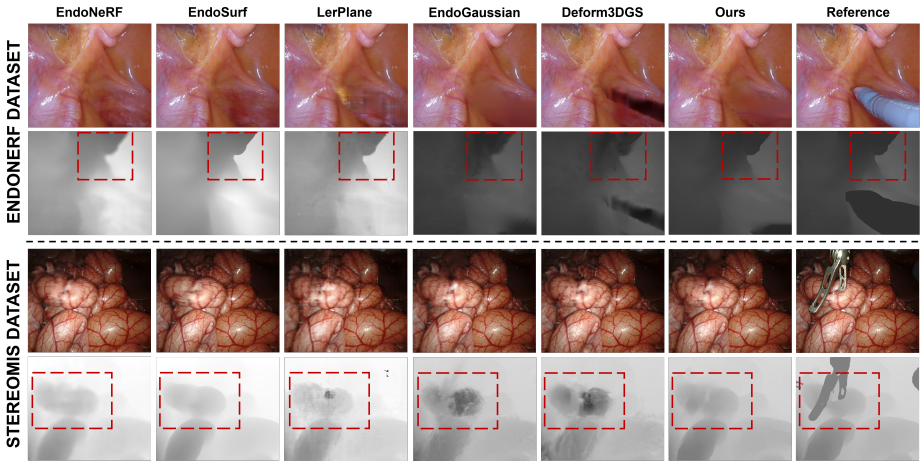

Fig. 2. Visualisation of the 3D reconstruction results. Further details can be seen by zooming in.

reduces geometry error by 37.79% (RMSE: 2.926 $mm \rightarrow 1.820\ mm$) and achieves a suboptimal result in rendering quality. Besides, although the rendering speed of our method is slower than that of Deform3DGS [24], it still achieves real-time performance, as most real-world endoscopes operate at 30-60 FPS.

We also visualize several scenes as shown in Fig. 2. It can be seen that our method effectively models complex tissue motions and preserves texture details. In addition, our method reconstructs an accurate and smooth depth map. In contrast, EndoNeRF [22], EndoSurf [25], and LerPlane [23] fail to adjust scaling to match the real depth distribution, while EndoGaussian [14] and Deform3DGS [24] struggle with capturing precise edges, especially in regions with deformable tissues. These results demonstrate that our method achieves accurate reconstructions of the dynamic surgical scene, highlighting its potential for applications in intraoperative navigation and robotic surgery automation.

We present an ablation study to evaluate the effect of the proposed methods. We first analyze the influence of our dense initialization strategy in Table 2. The 'w/o init' denotes using point clouds of the first frame to initialize the Gaussian points, which results in a performance decrease, showing the significance of the proposed dense initialization. Furthermore, we compare our $\mathcal{L}_{\hat{\mathrm{D}}}$ with common depth losses, $\mathcal{L}_1$ $\mathcal{L}_{\mathrm{LogL1}}$, and $\mathcal{L}_{\hat{\mathrm{D}}}^{-1}$. As is shown in Table 2, $\mathcal{L}_{\hat{\mathrm{D}}}$ generally outperforms other depth losses and the unsupervised depth smoothness has a slight improvement in depth metrics. Additionally, we find that although $\mathcal{L}_{\hat{\mathrm{D}}}^{-1}$ underperforms on depth metrics, it significantly improves image quality. This may indicate that strict constraints on the positions of Gaussians may limit their flexibility in reconstructing, which might reason $\mathcal{L}_{\hat{\mathrm{D}}}^{-1}$ causes suboptimal rendering results. Fig. 3 demonstrates that our $\mathcal{L}_{\hat{\mathrm{D}}}$ produces a more accurate depth

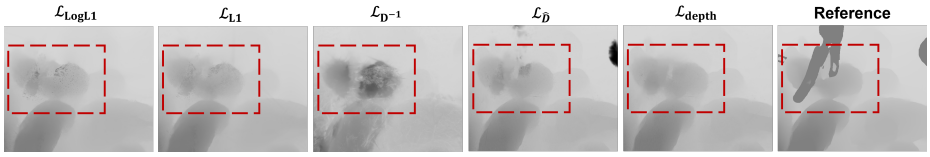

Fig. 3. Visualization of ablation study on depth loss using the StereoMIS dataset. Further details can be seen by zooming in.

Table 2. Ablation of proposed methods. $\mathcal{L}_{\hat{\text{D}}}$ is our normalized depth loss, and $\mathcal{L}_{\text{depth}}$ is normalized depth loss along with unsupervised depth smoothness.

Component	Method	RMSE ↓	PSNR ↑	SSIM ↑
Initialization	w/o Init	2.362	31.43	0.881
	w/ Init	**2.174**	**31.54**	**0.885**
Depth loss	ours-$\mathcal{L}_{\text{LogL1}}$	2.375	23.44	0.643
	ours-$\mathcal{L}_1$	2.436	22.91	0.630
	ours-$\mathcal{L}_{\hat{\text{D}}}^{-1}$	4.956	**32.29**	**0.903**
	ours-$\mathcal{L}_{\hat{\text{D}}}$	<u>2.256</u>	31.54	0.885
	ours-$\mathcal{L}_{\text{depth}}$	**2.174**	<u>31.54</u>	<u>0.885</u>

map, and the unsupervised depth smoothness significantly enhances the overall visual quality.

4 Conclusions

In this paper, we present a novel 3DGS-based approach for accurate surgical scene reconstruction. Different from previous methods, we proposed a normalized depth regularization and unsupervised depth smoother to ensure accurate geometry. Additionally, the dense initialization strategy is introduced to remove artifacts in depth maps and improve reconstruction quality. From the ablation study, we validate the effects of our proposed methods. We also observe that L1 and LogL1 depth losses introduce noise and significantly degrade reconstruction quality. Furthermore, we find that inverse depth loss struggles to capture accurate geometry, but it improves image quality. This indicates that imposing less constraint on the position of the Gaussians allows greater freedom to represent the scene. Besides, we notice that incomplete and sparse initialization leads to a noticeable decline in reconstruction performance. Extensive experiments on two real surgical datasets show that our method achieves state-of-the-art reconstruction quality, particularly in terms of geometric accuracy.

Future work will focus on evaluating the robustness of SurgicalGS under a variety of real-world and synthetic surgical scenarios, including complex and challenging conditions such as blood, smoke, and blurring. Additionally, further exploration is needed to assess the model's ability to maintain high reconstruction quality and geometric accuracy in these adverse conditions, ensuring its

adaptability and reliability in more dynamic and unpredictable surgical environments.

Acknowledgments. This paper is independent research partially funded by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC), the Cancer Research UK (CRUK) Imperial Centre.

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

References

1. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (6), 679–698 (1986)
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, G., Wang, W.: A survey on 3d gaussian splatting. arXiv preprint arXiv:2401.03890 (2024)
4. Chong, N., Si, Y., Zhao, W., Zhang, Q., Yin, B., Zhao, Y.: Virtual reality application for laparoscope in clinical surgery based on siamese network and census transformation. In: Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis. pp. 59–70. Springer (2022)
5. Drebin, R.A., Carpenter, L., Hanrahan, P.: Volume rendering. ACM Siggraph Computer Graphics **22**(4), 65–74 (1988)
6. Fang, J., Yi, T., Wang, X., Xie, L., Zhang, X., Liu, W., Nießner, M., Tian, Q.: Fast dynamic radiance fields with time-aware neural voxels. In: SIGGRAPH Asia 2022 Conference Papers. pp. 1–9 (2022)
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 **18**(7), 1185–1192 (2023)
8. Huang, B., Nguyen, A., Wang, S., Wang, Z., Mayer, E., Tuch, D., Vyas, K., Giannarou, S., Elson, D.S.: Simultaneous depth estimation and surgical tool segmentation in laparoscopic images. IEEE transactions on medical robotics and bionics **4**(2), 335–338 (2022)
9. Huang, B., Tsai, Y.Y., Cartucho, J., Vyas, K., Tuch, D., Giannarou, S., Elson, D.S.: Tracking and visualization of the sensing area for a tethered laparoscopic gamma probe. International Journal of Computer Assisted Radiology and Surgery **15**(8), 1389–1397 (2020)
10. Huang, B., Vo, T., Kongtongvattana, C., Dagnino, G., Kundrat, D., Chi, W., Abdelaziz, M., Kwok, T., Jianu, T., Do, T., et al.: Cathaction: A benchmark for endovascular intervention understanding. arXiv preprint arXiv:2408.13126 (2024)
11. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
12. Kusupati, U., Cheng, S., Chen, R., Su, H.: Normal assisted stereo depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2189–2199 (2020)

13. Li, Z., Liu, X., Drenkow, N., Ding, A., Creighton, F.X., Taylor, R.H., Unberath, M.: Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6197–6206 (2021)
14. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv preprint arXiv:2401.12561 (2024)
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. Overley, S.C., Cho, S.K., Mehta, A.I., Arnold, P.M.: Navigation and robotics in spinal surgery: where are we now? Neurosurgery **80**(3S), S86–S99 (2017)
17. Penza, V., De Momi, E., Enayati, N., Chupin, T., Ortiz, J., Mattos, L.S.: Envisors: Enhanced vision system for robotic surgery. a user-defined safety volume tracking to minimize the risk of intraoperative bleeding. Frontiers in Robotics and AI **4**, 15 (2017)
18. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4104–4113 (2016)
19. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. IEEE Robotics and Automation Letters **3**(1), 155–162 (2017)
20. Stoyanov, D.: Surgical vision. Annals of biomedical engineering **40**, 332–345 (2012)
21. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: ECCV. pp. 402–419. Springer (2020)
22. 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)
23. Yang, C., Wang, K., Wang, Y., Yang, X., Shen, W.: Neural lerplane representations for fast 4d reconstruction of deformable tissues. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 46–56. Springer (2023)
24. Yang, S., Li, Q., Shen, D., Gong, B., Dou, Q., Jin, Y.: Deform3dgs: Flexible deformation for fast surgical scene reconstruction with gaussian splatting. arXiv preprint arXiv:2405.17835 (2024)
25. 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)
26. Zhang, D., Xiao, B., Huang, B., Zhang, L., Liu, J., Yang, G.Z.: A self-adaptive motion scaling framework for surgical robot remote control. IEEE Robotics and Automation Letters **4**(2), 359–366 (2018)
27. Zhou, H., Jagadeesan, J.: Real-time dense reconstruction of tissue surface from stereo optical video. IEEE transactions on medical imaging **39**(2), 400–412 (2019)
28. 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)