

Deform3DGS: Flexible Deformation for Fast Surgical Scene Reconstruction with Gaussian Splatting

Shuojue Yang¹, Qian Li¹, Daiyun Shen², Bingchen Gong³, Qi Dou³, and Yueming Jin¹ (✉)

¹ National University of Singapore, Singapore, Singapore

² Tsinghua University, Beijing, China

³ The Chinese University of Hong Kong, Hong Kong, China
ymjin@nus.edu.sg

Abstract. Tissue deformation poses a key challenge for accurate surgical scene reconstruction. Despite yielding high reconstruction quality, existing methods suffer from slow rendering speeds and long training times, limiting their intraoperative applicability. Motivated by recent progress in 3D Gaussian Splatting, an emerging technology in real-time 3D rendering, this work presents a novel fast reconstruction framework, termed *Deform3DGS*, for deformable tissues during endoscopic surgery. Specifically, we introduce 3D GS into surgical scenes by integrating a point cloud initialization to improve reconstruction. Furthermore, we propose a novel flexible deformation modeling scheme (FDM) to learn tissue deformation dynamics at the level of individual Gaussians. Our FDM can model the surface deformation with efficient representations, allowing for real-time rendering performance. More importantly, FDM significantly accelerates surgical scene reconstruction, demonstrating considerable clinical values, particularly in intraoperative settings where time efficiency is crucial. Experiments on DaVinci robotic surgery videos indicate the efficacy of our approach, showcasing superior reconstruction fidelity PSNR: (37.90) and rendering speed (338.8 FPS) while substantially reducing training time to only 1 minute/scene. Our code is available at <https://github.com/jinlab-imvr/Deform3DGS>.

Keywords: Fast 3D reconstruction · Surgical scene reconstruction · 3D Gaussian Splatting · Deformable scene.

1 Introduction

Three-dimensional (3D) reconstruction of surgical scenes has great potential to facilitate many downstream applications including intraoperative navigation [14], and visualization enhancement [15, 11]. Besides, a high-quality and dynamic 3D scene model has demonstrated potential benefits to surgical training through

D. Shen did this work during his internship at National University of Singapore.

shortening learning curves [1] and remote surgical proctoring [19] by allowing immersive observation of surgical scenes.

Conventional reconstruction methods suffer from a redundant workflow including depth estimation, surface reconstruction, and texture mapping [16]. To enhance the compactness and efficiency, many studies introduce Neural Radiance Field (NeRF) [13], an implicit 3D representation using tiny Multi-Layer Perceptrons (MLPs), to model geometric details and appearance from the captured video. This implicit representation can directly render the photo-realistic novel views, simplifying the bulky conventional workflow. Works [17,22] successfully adapt NeRF into endoscopic scene reconstruction, leading to promising performance in rendering quality and geometry fidelity. However, this line of work suffers from a long training time (hours) and low rendering speed, which significantly impedes their intraoperative applicability. LerPlane [20] encodes the spatial and temporal information using decomposed 4D feature planes to accelerate the reconstruction. This approach inherently models the deformation and directly outputs rendering parameters for deformed tissues, which heavily rely on complex computations performed on feature planes and MLPs, leading to a compromised acceleration.

With the current progress in computer graphics, Gaussian Splatting (GS) [5] emerged as a ground-breaking 3D representation. Driven by its superior performance, many efforts [21,8,6] have been made to adapt this technique to dynamic scene reconstruction through a deformation model to represent motions. 4DGS [18] is one of the pilot studies presenting one efficient solution where decomposed feature planes similar to [2] are used to model time-dependent deformations. Despite the compact feature encoding, 4DGS retains time-consuming feature plane interpolation and decoding with MLPs, which limits the acceleration and intraoperative applicability. Also, 4DGS inadequately utilizes geometric priors, leading to a highly sparse point initialization and a longer time for geometry reconstruction. Therefore, integrating GS into the surgical scene reconstruction framework is still challenging. In addition, although concurrent works extending GS to endoscopic settings show excellent performance in geometry refinement [23], model lightweight [10], and sparse-view synthesis [7], fast reconstruction of deformable surgical scenes is understudied, which inspires our work toward efficiently adapting GS to intraoperative scenes.

In this paper, we develop a highly efficient framework for deformable surgical scene reconstruction, named Deform3DGS, by introducing the GS technique into surgical scenario with a motion-aware point fusion (MAPF) to initialize the Gaussian point cloud densely. Besides, we propose a novel flexible deformation modeling scheme (FDM) for efficiently representing tissue deformations. FDM models the tissue deformations via an efficient linear combination regression, where learnable basis functions are leveraged to improve both the representation capability and efficiency. Finally, we evaluate our method on EndoNeRF [17] and StereoMIS [3] datasets collected from Da Vinci robotic surgery videos. Experiments indicate the efficacy of our approach, demonstrating superior reconstruc-

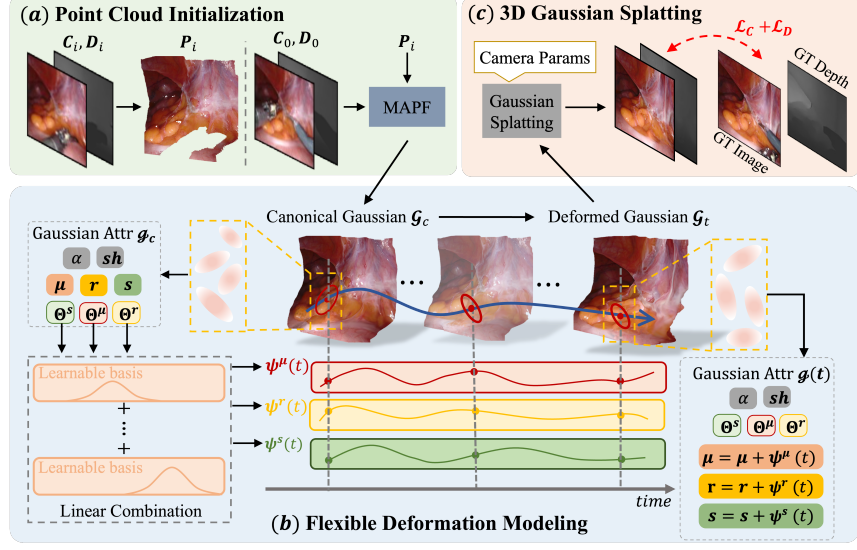


Fig. 1. Illustration of our fast deformable tissue reconstruction framework, Deform3DGS, composed of (a) Point cloud initialization, (b) Flexible Deformation Modeling, and (c) 3D Gaussian Splatting.

tion quality (PSNR: 37.90) and rendering speed (338.8 FPS) while substantially reducing training time to around only 1 minute per scene.

2 Method

Pipeline. As shown in Fig. 1, during the training phase, our framework initializes the endoscopic scene with a Gaussian point cloud $\mathcal{G}$ using the MAPF scheme (Sec. 2.3). The following is the FDM that models deformed tissues by learning the time-dependent changes for each Gaussian point $\mathbf{g}$ (Sec. 2.2). Next, the rendering (Sec. 2.1) is performed to obtain the colored image and depth of the deformed Gaussian point cloud given the camera viewpoint. Finally, the rendered image and depth are supervised by corresponding ground-truth (GT) image and stereo depth to optimize the framework (Sec. 2.5). During rendering (testing), given a query time and camera viewpoint, FDM deforms the learned Gaussian point cloud $\mathcal{G}$ and the view of the deformed Gaussians is rendered.

Problem Setting. Our task is to train a deformable tissue reconstruction model Ψ from an endoscopic surgery video. Given the camera intrinsic matrix $\mathbf{K}$ and extrinsic matrix $\mathbf{T}_i$ recording the camera viewpoint information at the i -th frame, a desired reconstruction model Ψ^* is supposed to render the i -th view $\mathbf{I}_i$ at timestamp t_i as $\mathbf{I}_i = \Psi^*(\mathbf{T}_i, t_i; \mathbf{K})$.

2.1 Preliminaries of Gaussian Splatting

3D Gaussian Splatting [5] is a static 3D scene representation that models scenes with the form of a 3D Gaussian point cloud in the world coordinate frame. Each Gaussian point contains learnable attributes: position $\boldsymbol{\mu} \in \mathbb{R}^3$, rotation $\boldsymbol{r} \in \mathbb{R}^4$, scale $\boldsymbol{s} \in \mathbb{R}^3$, opacity α and spherical harmonic (*sh*) coefficients. Given an arbitrary 3D coordinate $\boldsymbol{x} \in \mathbb{R}^3$ in world frame, the spatial impact of a 3D Gaussian point on $\boldsymbol{x}$ is defined by a Gaussian distribution as following:

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right) \quad (1)$$

$$\boldsymbol{\Sigma} = \boldsymbol{R} \boldsymbol{S} \boldsymbol{S}^T \boldsymbol{R}^T, \quad (2)$$

where $\boldsymbol{R}$ is the rotation matrix calculated from $\boldsymbol{r}$, and $\boldsymbol{S}$ is the diagonal matrix of $\boldsymbol{s}$. Next, the 3D Gaussian point is projected to the 2D image plane for rendering. The projected 2D Gaussian with the position $\boldsymbol{\mu}^{2D}$ and covariance matrix $\boldsymbol{\Sigma}^{2D}$ can be analytically computed in pixel coordinate frame given the camera intrinsic and extrinsic parameters ($\boldsymbol{K}$ and $\boldsymbol{T}$). Finally, an α -blending [5] is performed to render the colored images $\hat{\boldsymbol{C}}$ and corresponding depth $\hat{\boldsymbol{D}}$.

2.2 Flexible Deformation Modeling

MLP-based implicit deformation fields represented by Hexplane [2] incur significant computational overhead to training and cannot meet the stringent real-time processing demands of surgical video analytics. To mitigate it, [9] proposed a computationally efficient explicit representation of the deformation field, in which Fourier and polynomial basis functions $b(t)$ parameterized by time are employed to learn the per-Gaussian motion curve. Each Gaussian is endowed with a set of learnable weights $\boldsymbol{\omega}$ that linearly combine the basis functions to generate the motion curve as $\psi(t; \boldsymbol{\omega}) = \sum_{j=1}^B \omega_j b_j(t)$. Despite the adaptability of each Gaussian to deformation through weight adjustments, the temporal deformation is confined to canonical motion, leading to inconsistent deformation representations for different queried times. The model may be compelled to forgo certain specific movements to ensure a coherent overall trajectory, which is detrimental in scenes with intricate and nuanced deformations, such as instrument-tissue interactions. Furthermore, the deformation learned at a particular timestamp globally influences the entire trajectory, resulting in a sub-optimal local deformation representation at different queried times.

To overcome these limitations, we introduce a novel flexible deformation modeling scheme (FDM) which offers flexibility and adaptability to basis functions using learnable parameters. We adopt the Gaussian function with learnable center θ and variance σ .

$$\tilde{b}(t; \theta, \sigma) = \exp\left(-\frac{1}{2\sigma^2}(t - \theta)^2\right), \quad (3)$$

For each Gaussian $\mathbf{g}$ in the point cloud, position $\boldsymbol{\mu}$ and rotation $\mathbf{r}$ are naturally related to tissue motions, and scale $\mathbf{s}$ keeps varying since tissues are prone to elastic deformations during instrument intervention. Thus, an additional set of learnable parameters $\boldsymbol{\Theta}^\mu, \boldsymbol{\Theta}^r, \boldsymbol{\Theta}^s$ are introduced for each Gaussian to describe temporal deformations in their position, rotation, and scale respectively, see Fig. 1. Taking the positional change in x direction as an example, deformation curve can be expressed by a set of parameters $\boldsymbol{\Theta}^{\mu,x} = \{\omega^{\mu,x}, \theta^{\mu,x}, \sigma^{\mu,x}\}$ as:

$$\psi^{\mu,x}(t; \boldsymbol{\Theta}^{\mu,x}) = \sum_{j=1}^B \omega_j^{\mu,x} \tilde{b}(t; \theta_j^{\mu,x}, \sigma_j^{\mu,x}), \quad (4)$$

Due to the locally valuable nature of Gaussian functions, our model ensures that deformations at adjacent moments remain continuous, while those across large time intervals are almost decoupled. Besides, integrating with learnable parameters, this approach yields a deformation model that is not only computationally efficient but also capable of capturing intricate deformation dynamics.

2.3 Point Cloud Initialization

To further boost the reconstruction performance and stabilize the training, we introduce a Gaussian point cloud initialization before the deformation modeling. Specifically, we first employ the camera model and intrinsic matrix to extract the 3D tissue point clouds for each frame as:

$$\mathbf{P}_i = \mathbf{K}^{-1} \mathbf{D}_i (\mathbf{I}_i \odot \mathbf{M}_i), \quad (5)$$

where $\mathbf{P}_i$ and $\mathbf{I}_i$ respectively denote 3D point cloud and 2D pixel coordinates from the i -th frame, $\mathbf{D}_i$, $\mathbf{M}_i$ denote the i -th depth map and valid foreground (i.e., tissue) mask, respectively, $\mathbf{K}$ means the intrinsic matrix, $\odot$ means the element-wise multiplication. By default, the first frame is selected to initiate the Gaussian point cloud as the canonical state, i.e., $\mathbf{P}_c = \mathbf{P}_0$. However, with the presence of tool occlusions in the colored image, some pixels in $\mathbf{I}_0$ are filtered out by $\mathbf{M}_0$, resulting in voids and local sparsity on the initialized point cloud. This unevenly distributed initialization consumes more time densifying the point cloud and leads to a sub-optimal efficiency. Holding an assumption that dense Gaussian point distribution facilitates reconstruction on intensively deformed regions of the surgical scene, we develop a Motion-Aware Point Fusion (MAPF) scheme to selectively fuse points exhibiting intensive motions. Specifically, a motion-aware occlusion mask $\mathbf{F}$ is computed by combining occluded regions and pixels with large color differences from a pixel-wise-averaged image as following:

$$\mathbf{F} = \mathbb{I}(|\mathbf{C}_0 - \sum_j^N \mathbf{C}_j / N| > \tau) \cup (\mathbf{1} - \mathbf{M}_0), \quad (6)$$

where $\mathbb{I}(\cdot)$ refers to the indicator function, $\mathbf{C}_i$ represents the colored image of the i -frame, N is the total number of frames, and τ is the threshold determining pixels with significant motions. $\mathbf{F}$ masks regions with large motions and local sparsity of the initialized Gaussian points $\mathbf{P}_c$ (i.e., $\mathbf{P}_0$). Finally, 3D points of $\mathbf{P}_i$ with 2D projected pixels in mask $\mathbf{F}$ will be fused with the $\mathbf{P}_c$ to initialize the canonical state as shown in Fig. 1.

2.4 Optimization

Our proposed framework jointly optimizes the canonical Gaussians $\mathcal{G}_c$ and the deformation model given by Θ . Given the tissue mask $\mathbf{M}$, we train our deformable tissue reconstruction framework by supervising the rendered images and depths by ground-truth colored images and stereo depth maps as following:

$$\mathcal{L}_C = \|\mathbf{M} \odot (\hat{\mathbf{C}} - \mathbf{C})\|, \quad \mathcal{L}_D = \|\mathbf{M} \odot (\hat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\|, \quad (7)$$

where the $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$, $\mathbf{C}$, and $\mathbf{D}$ denote the rendered image, rendered depth, GT image, and stereo depth, respectively. The overall training loss is summarized $\mathcal{L} = \mathcal{L}_C + \mathcal{L}_D$.

3 Experiment

3.1 Experiment Setting

Datasets and Evaluation. We evaluate the proposed method and compare it with existing works on two datasets: 1) EndoNeRF dataset [17] is a collection of stereo endoscopic videos including 6 clips extracted from Da Vinci robotic prostatectomy data. Each clip is captured from a single camera viewpoint with complex surgical instrument occlusion and tissue deformations. 2) StereoMIS dataset [3] is a stereo endoscopic video dataset captured from in-vivo porcine subjects containing diverse anatomical structures and challenging scenes with large tissue deformations. Specifically, we use all 6 scenes of EndoNeRF and select 3 clips from video P2_7 and P3 in StereoMIS datasets with more diverse anatomical structures compared to EndoNeRF dataset. Each selected clip lasts for 4 ~ 5s with 30 *fps*. Following [22], we split frames of each scene into training and testing sets with a ratio of 7:1. We use PSNR, SSIM, and LPIPS to evaluate the reconstruction performance. Also, training time and rendering speed are calculated to evaluate the efficiency.

Implementation Details. For each scene, we normalize the video duration into [0, 1] and empirically apply 17 learnable Gaussian basis functions to compose FDM. The training lasts for 3000 iterations, with an initial learning rate of 1.6×10^{-3} . To stabilize the training, we freeze the densification on Gaussian points number at the initial 600 iterations. All the experiments are based on the PyTorch framework [4] and conducted with a single NVIDIA RTX A5000 GPU.

3.2 Comparison with State-of-the-art Methods

We evaluate our proposed framework by comparing its performance with EndoNeRF [17] and two existing SOTAs on fast reconstruction: LerPlane [20], and EndoGaussian[12]. EndoGaussian is a concurrent work using GS to accelerate the endoscopic reconstruction. As listed in Table 1, despite the effectiveness of reconstructing deformable tissues, EndoNeRF takes a long training time (hours) to reconstruct a scene in seconds, which compromises their intraoperative usability.

Dataset	Method	PSNR $\uparrow$	SSIM($\%$) $\uparrow$	LPIPS $\downarrow$	Time (sec) $\downarrow$	Speed (<i>fps</i>) $\uparrow$
EndoNeRF	EndoNeRF	35.55	93.02	0.09	$\sim$ 21600	0.03
	LerPlane	36.56	94.36	0.07	$\sim$ 600	1.45
	EndoGaussian	<u>37.66</u>	95.89	<u>0.06</u>	<u>138</u>	<u>128.13</u>
	Ours	37.90	<u>95.84</u>	0.06	64	338.80
StereoMIS	EndoNeRF	28.86	74.15	0.27	$\sim$ 21600	0.03
	LerPlane	29.46	77.73	0.20	$\sim$ 600	1.52
	EndoGaussian	<u>30.25</u>	82.75	0.21	<u>151</u>	<u>134.50</u>
	Ours	30.48	<u>82.74</u>	<u>0.21</u>	66	330.37

Table 1. Quantitative evaluation of our proposed framework against existing methods on endoscopic scene reconstruction. ‘Time’ and ‘Speed’ denote the training time and rendering speed (*fps*), respectively. The optimal and suboptimal results are shown in **bold** and underlined respectively.

Method	PSNR $\uparrow$	SSIM($\%$) $\uparrow$	Time(<i>min</i>) $\downarrow$
LerPlane	22.75	82.13	$\sim$ 1
EndoGaussian	35.81	94.67	$\sim$ 1
Ours	37.90	95.84	$\sim$ 1

Table 2. Quantitative comparison with SOTAs given the limited training time around 1 *min*.

Method	PSNR $\uparrow$	SSIM($\%$) $\uparrow$	Time(<i>sec</i>) $\downarrow$
Ours-HexPlane	37.13	95.51	108
Ours-PFS	37.08	95.29	60
w/o MAPF	37.47	95.62	61
Ours	37.90	95.84	64

Table 3. Quantitative analysis of the key components on EndoNeRF dataset. ‘Time’ denotes the training time.

As a comparison, the fast reconstruction method, LerPlane, effectively accelerates the training phase to the minute level and leads to a superior reconstruction quality. Our framework yields noticeable performance gains over LerPlane across all evaluation metrics on reconstruction quality, while substantially improving training efficiency by 10 times to around 1 minute. Furthermore, we compare our method against EndoGaussian [12]. Similar to [18], this method relies on decomposed feature planes to model dynamic Gaussian Splatting, presenting remarkable performance against NeRF-based methods and reaching comparable reconstruction quality to our framework. However, benefitting from the highly efficient FDM, our method leads to over $2\times$ acceleration in both training ($138\text{ sec} \rightarrow 64\text{ sec}$) and rendering ($128\text{ fps} \rightarrow 338\text{ fps}$). Note that our measured training and rendering speeds for EndoGaussian mismatch with the reported values in [12] due to different hardware used.

Despite the comparable reconstruction quality, as shown in Table 2, our proposed method achieves significantly superior performance within a limited training time, which demonstrates the superiority in intraoperative scenarios.

We also visualize several rendered scenes as shown in Fig .2 for better qualitative evaluation. It can be observed that our method has an enhanced capability of preserving the appearance details and modeling complex tissue motions. Additionally, rendered views given by EndoGaussian indicate comparable rendering quality to ours without a visually perceivable difference. According to these results, our proposed method achieves state-of-the-art (SOTA) performance on endoscopic scene reconstruction, especially its outstanding progress in fast training

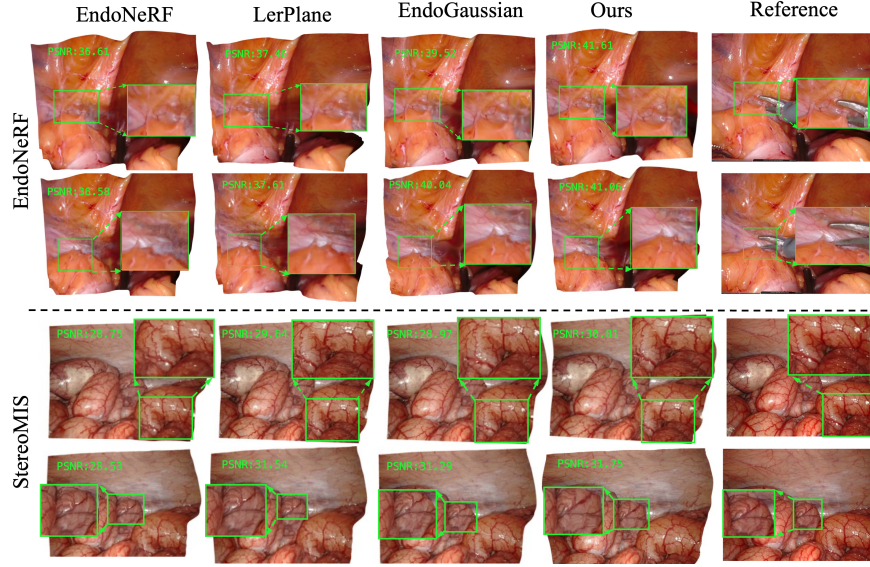


Fig. 2. Visualization of the 3D reconstruction results.

and real-time-level rendering, which indicates great clinical values in intraoperative applications.

3.3 Quantitative Evaluation of Key Components

We first investigate the effectiveness of the proposed FDM by comparing it with existing deformation modeling techniques on the EndoNeRF dataset. With the identical workflow shown in Fig. 1, we replace the FDM with other modeling methods including a combination of Fourier and Polynomial series (Ours-FPS), and a HexPlane-based decomposed feature plane [2] following [18] denoted as Ours-HexPlane. As illustrated in Table 3, despite the acceleration, Ours-FPS shows limited capability of representing complex deformations. On the other hand, using HexPlane to encode spatial and temporal information significantly enhances the representative capability, however, leading to a relatively longer training time. Our proposed FDM achieves the best deformation representing performance without impairing time efficiency. Furthermore, ‘w/o MAPF’ refers to initializing the Gaussian points with only the first frame point cloud, which exhibits a performance drop and thus demonstrates the important role of the proposed MAPF scheme.

4 Conclusion

In this paper, we work toward intraoperative surgical scene reconstruction by proposing a fast and accurate deformable scene reconstruction framework. With

the utilization of Gaussian Splatting, our framework can achieve high-quality rendering at a real-time level. To further accelerate the tissue motion modeling, we introduce an efficient flexible deformation modeling scheme composed of learnable Gaussian basis functions to maintain a strong motion-representative capability. Besides, combined with a motion-aware point fusion scheme for initialization, our framework leads to a SOTA reconstruction quality while significantly minimizing the training time to only 1 *min*/scene, showing the possibility of reconstructing longer sequences with more challenging surgical scenes. Therefore, we believe that our work achieves significant progress in bridging the gap between high-quality rendering and intraoperative applications.

Acknowledgement. This work was supported by Ministry of Education Tier 1 Start up grant, NUS, Singapore (A-8001267-01-00); Ministry of Education Tier 1 grant, NUS, Singapore (A-8001946-00-00). This work was also supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 24209223)

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

References

1. Boedecker, C., Huettl, F., Saalfeld, P., Paschold, M., Kneist, W., Baumgart, J., Preim, B., Hansen, C., Lang, H., Huber, T.: Using virtual 3d-models in surgical planning: workflow of an immersive virtual reality application in liver surgery. *Langenbeck’s archives of surgery* **406**, 911–915 (2021)
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. 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)
4. Imambi, S., Prakash, K.B., Kanagachidambaresan, G.: *Pytorch. Programming with TensorFlow: Solution for Edge Computing Applications* pp. 87–104 (2021)
5. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4) (2023)
6. Kratimenos, A., Lei, J., Daniilidis, K.: Dynmf: Neural motion factorization for real-time dynamic view synthesis with 3d gaussian splatting. *arXiv preprint arXiv:2312.00112* (2023)
7. Li, C., Feng, B.Y., Liu, Y., Liu, H., Wang, C., Yu, W., Yuan, Y.: Endospase: Real-time sparse view synthesis of endoscopic scenes using gaussian splatting. *arXiv preprint arXiv:2407.01029* (2024)
8. Li, Z., Chen, Z., Li, Z., Xu, Y.: Spacetime gaussian feature splatting for real-time dynamic view synthesis. *arXiv preprint arXiv:2312.16812* (2023)
9. Lin, Y., Dai, Z., Zhu, S., Yao, Y.: Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. *arXiv preprint arXiv:2312.03431* (2023)

10. Liu, H., Liu, Y., Li, C., Li, W., Yuan, Y.: Lgs: A light-weight 4d gaussian splatting for efficient surgical scene reconstruction. *arXiv preprint arXiv:2406.16073* (2024)
11. Liu, X., Stiber, M., Huang, J., Ishii, M., Hager, G.D., Taylor, R.H., Unberath, M.: Reconstructing sinus anatomy from endoscopic video—towards a radiation-free approach for quantitative longitudinal assessment. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III* 23. pp. 3–13. Springer (2020)
12. Liu, Y., Li, C., Yang, C., Yixuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. *Arxiv* (2024)
13. 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)
14. Pelanis, E., Teatini, A., Eigl, B., Regensburger, A., Alzaga, A., Kumar, R.P., Rudolph, T., Aghayan, D.L., Riediger, C., Kvarnström, N., et al.: Evaluation of a novel navigation platform for laparoscopic liver surgery with organ deformation compensation using injected fiducials. *Medical image analysis* **69**, 101946 (2021)
15. Rodby, K.A., Turin, S., Jacobs, R.J., Cruz, J.F., Hassid, V.J., Kolokythas, A., Antony, A.K.: Advances in oncologic head and neck reconstruction: systematic review and future considerations of virtual surgical planning and computer aided design/computer aided modeling. *Journal of Plastic, Reconstructive & Aesthetic Surgery* **67**(9), 1171–1185 (2014)
16. 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)
17. 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)
18. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. *arXiv preprint arXiv:2310.08528* (2023)
19. Wu, T.Y., Meng, Q., Yang, L., Kumari, S., Pirouz, M.: Amassing the security: An enhanced authentication and key agreement protocol for remote surgery in healthcare environment. *CMES-Computer Modeling in Engineering & Sciences* **134**(1) (2023)
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. Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. *arXiv preprint arXiv:2309.13101* (2023)
22. 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)
23. Zhu, L., Wang, Z., Jin, Z., Lin, G., Yu, L.: Deformable endoscopic tissues reconstruction with gaussian splatting. *arXiv preprint arXiv:2401.11535* (2024)