LGS: A Light-weight 4D Gaussian Splatting for Efficient Surgical Scene Reconstruction

Hengyu Liu[1⋆], Yifan Liu[1⋆], Chenxin Li[1⋆], Wuyang Li[1], Yixuan Yuan[1,2(✉)]

[1] Department of Electronic Engineering, The Chinese University of Hong Kong,
[2] CUHK Shenzhen Research Institute

Abstract. The advent of 3D Gaussian Splatting (3D-GS) techniques and their dynamic scene modeling variants, 4D-GS, offers promising prospects for real-time rendering of dynamic surgical scenarios. However, the prerequisite for modeling dynamic scenes by a large number of Gaussian units, the high-dimensional Gaussian attributes and the high-resolution deformation fields, all lead to serve storage issues that hinder real-time rendering in resource-limited surgical equipment. To surmount these limitations, we introduce a Lightweight 4D Gaussian Splatting framework (LGS) that can liberate the efficiency bottlenecks of both rendering and storage for dynamic endoscopic reconstruction. Specifically, to minimize the redundancy of Gaussian quantities, we propose Deformation-Aware Pruning by gauging the impact of each Gaussian on deformation. Concurrently, to reduce the redundancy of Gaussian attributes, we simplify the representation of textures and lighting in non-crucial areas by pruning the dimensions of Gaussian attributes. We further resolve the feature field redundancy caused by the high resolution of 4D neural spatiotemporal encoder for modeling dynamic scenes via a 4D feature field condensation. Experiments on public benchmarks demonstrate the efficacy of LGS in terms of a compression rate exceeding 9× while maintaining the pleasing visual quality and real-time rendering efficiency. LGS confirms a substantial step towards its application in robotic surgical services. Project page: `https://lgs-endo.github.io/`.

Keywords: 3D Reconstruction · Gaussian Splatting · Robotic Surgery.

1 Introduction

Reconstructing dynamic 3D scenes from endoscopic videos holds paramount significance in minimally invasive surgeries [26,31], as it enhances comprehension of the spatial environment surrounding the surgical site, thereby enabling surgeons to conduct more precise and efficient operations [19,11]. Concomitantly, this technology fosters a myriad of subsequent applications, encompassing Virtual Reality (VR) surgeries, medical pedagogy, and the automation of robotic surgeries [17,25]. Subsequent advancements, particularly real-time rendering, have

⋆ Equal contribution

emerged as a cutting-edge methodology conducive to deployment on robotic apparatus, thereby facilitating the progression of robotic surgical automation [16], medical imaging system [12] and aided diagnosis procedure [6].

Previous 3D reconstruction methods encompass various approaches, including those employing depth estimation [18], SLAM-based methods [24,33], and techniques leveraging a sparse warp field [13,5]. With the rising of neural rendering, Neural Radiance Fields (NeRF) [20] is introduced to reconstruct the surgical scenes [26,28,31], unlocking the limited reconstruction quality of dynamic scenes compared with conventional methods. EndoNeRF [26] and its following works [28,31] utilize dynamic neural radiance fields to model deformed surgical scenes and achieve satisfied rendering quality. Recently, 3D Gaussian Splatting (3D-GS) [8], which utilizes an explicit 3D Gaussian representation with specific attributes to model scenes and a differentiated splatting-based rendering technique, has exhibited pleasing reconstruction efficiency from a series of static captures. Further advances of 4D-GS [27] break this *static* limitation by additionally introduce an elaborated high-resolution spatiotemporal feature field to model the time-varying deformation [27] and the relevant progress has been extended to clinical scenarios like reconstructing deformable tissues in dynamic endoscopic scenes in real-time efficiency [15,29,32,34].

Nevertheless, owing to using explicit point-based representations and intricate high-dimensional spatial-temporal fields, the storage burden of 3D representation is significantly enlarged, thereby limiting the practical deployment on resource-limited surgical devices and robots [1,23,10]. In particular, the memory burden by 4D-GS primarily lies in the following folds. (**a**) Redundant Number of Gaussians: Gaussian densification [8] increases the number of Gaussian for the accurate reconstruction of granular details, which results in a significant memory cost [4]. While compared with nature scenes, surgical scenarios require less Gaussians to model the relatively simple environment. (**b**) Redundant Dimension of Gaussian Attributes: Natural scenes require high-dimensional attributes to represent the rich textures and varying illumination. While the intricate environment of surgical scenes such as repetitive textures and view-dependent lighting [3,22] can be represented with fewer parameters, which causes the redundant dimension of Gaussian Attributes. (**c**) High Resolution of Spatial-temporal Fields: A lower-resolution feature field [21] is sufficient to represent deformation, while the high resolution used in spatial-temporal feature fields is redundant to model the dynamics and encode the endoscopic scenario information.

To address the challenges of practical deployment of reconstruction models posed by the aforementioned prohibitively high memory burden. we propose a holistic Lightweight 4D Gaussian Splatting (LGS) framework that allows for achieving satisfactory endoscopic reconstruction with both efficient rendering and storing [7]. In particular, to alleviate the **Quantity** burden of Gaussian representation, we present a Deformation-Aware Pruning (DAP) which identifies the informative Gaussians by a carefully designed deformation score and reduces the redundant Gaussians based on a deformation perception. To address the **High-dimension** burden of Gaussian attributes, we propose a Gaussian-

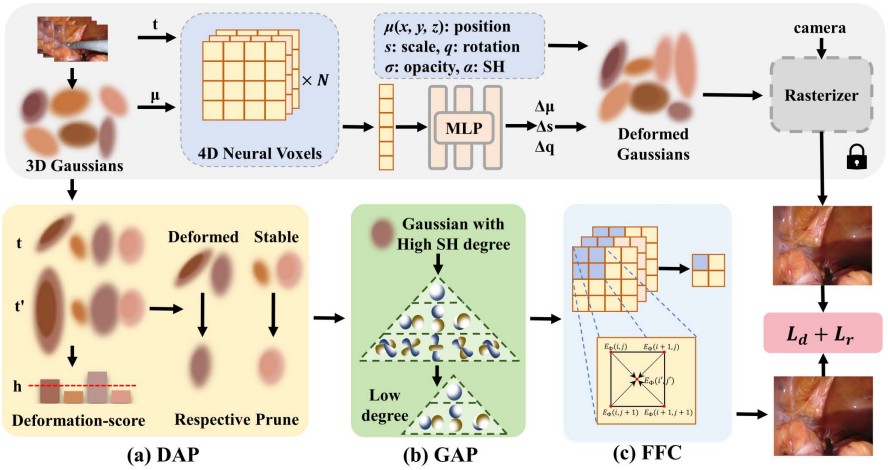

Fig. 1. LGS Overview consists of (a) Deformation-Aware Pruning, (b) Gaussian-Attribute Pruning, (c) Feature Field Condensation, and distillation for optimization.

Attribute Pruning (GAP), which elucidates the uninformative visual patterns modeled by high-frequency components (non-critical textures) in Spherical Harmonic attributes, thereby steering the representation capabilities towards low-frequency attribute characterization. To mitigate the **High-resolution** burden of spatial-temporal deformable fields, we present Feature Field Condensation (FFC) which compactly represents the high-resolution spatial-temporal feature field and performs an adaptive quantized projection of spatial-temporal coordinates. Experimental results show that LGS can achieve higher storage efficiency with an over 9× compression rate, whilst maintaining pleasing reconstruction quality and rendering speed, confirming the efficacy of our framework for practical deployment in clinical robotic surgical equipment.

2 Method

The overview of LGS is shown in Fig. 1. In this section, we first introduce the representation of 3D-GS [8] and its dynamic variant [27] in Sec. 2.1. Then we introduce the DAP and GAP respectively in Sec. 2.2 and Sec . 2.3. In Sec. 2.4 we conduct FFC and finally, we describe the optimization of LGS in Sec. 2.5.

2.1 Preliminaries

3D-GS As introduced in [8], a set of dense Gaussians is utilized to represent the 3D data and achieve real-time rendering of dynamic scenes. Each Gaussian is defined by its center $\boldsymbol{\mu} \in \mathbb{R}^3$ and the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$, which can be decomposed into a scaling factor $\boldsymbol{s} \in \mathbb{R}^3$ and a rotation quaternion $\boldsymbol{q} \in \mathbb{R}^4$

for differentiable optimization. Colors and view-dependent appearance can be represented by opacity $\sigma \in \mathbb{R}$ and SH coefficients $\boldsymbol{\alpha} \in \mathbb{R}^C$ (C devotes the number of SH). After differential splatting [30] and adaptive density control [8], these Gaussians are optimized to achieve real-time rendering for dynamic scenes,

$$\mathrm{G}(\boldsymbol{x}) = e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{1}$$

Dynamic Scene Rendering As the extension of 3D-GS, [27] utilizes a deformation module which involves a spatial-temporal feature field E and a tiny MLP F to calculate deformation. Given a 4D input consisting of the Gaussian center $\boldsymbol{\mu} = (x, y, z)$ and query time t, the feature field E retrieves the latent feature f of the input: $\boldsymbol{f} = E(\boldsymbol{\mu}, t)$. Subsequently, the tiny MLP F calculates the deformation in position, rotation, and scaling: $\Delta\boldsymbol{\mu}, \Delta\boldsymbol{q}, \Delta\boldsymbol{s} = F(\boldsymbol{f})$. Then the i-th Gaussian can be represented as $\boldsymbol{\Theta_i} = \{\boldsymbol{\mu_i}', \boldsymbol{q_i}', \mathbf{s_i}', \sigma_i, \boldsymbol{\alpha_i}\}$, where $\boldsymbol{\mu_i}', \boldsymbol{q_i}', \boldsymbol{s_i}' = \boldsymbol{\mu_i} + \Delta\boldsymbol{\mu_i}, \boldsymbol{q_i} + \Delta\boldsymbol{q_i}, \boldsymbol{s_i} + \Delta\boldsymbol{s_i}$.

2.2 Deformation-Aware Pruning

To distinguish Gaussians that are informative for modeling deformable dynamics, we propose a deformation score to identify the impact of each Gaussian on deformation enabling the follow-up importance-based pruning can be conducted.
Deformation Score Inspired by the color rendering in 3D-GS [8] and the global importance score in [4], the deformation score of each Gaussian is associated with its contribution to all pixels and its deformation in volume across all timestamps. We use the criterion $\mathbb{1}(G(\boldsymbol{X}_i), \boldsymbol{p}_k, t)$ to reflect whether the i-th Gaussian is contributed to the pixel k at timestamp t. The volume deformation of the i-th Gaussian $\Delta V(\boldsymbol{s_i})$ can be calculated with scaling factor $\boldsymbol{s_i}$ and deformation $\Delta\boldsymbol{s_i}$. Consequently, the deformation score is obtained by,

$$d_i = \sum_{t}^{T} \sum_{k}^{HW} \mathbb{1}(G(\boldsymbol{X}_i), \boldsymbol{p}_k, t) \cdot \Delta V(\boldsymbol{s_i}), \Delta V(\boldsymbol{s_i}) = \sum_{t}^{T} \|V(\boldsymbol{s_i}) - V(\boldsymbol{s_i} + \boldsymbol{\Delta s_i})\|_1 \tag{2}$$

where T, H, and W denotes the timestamps, height and weight of the image, $V(\boldsymbol{s}) = 4\pi \boldsymbol{s}_1 \boldsymbol{s}_2 \boldsymbol{s}_3/3$ devotes the volume of Gaussian with $\boldsymbol{s}$.
Deformation-Aware Pruning Based on the deformation score and a threshold h, Gaussians are classified into two categories: stable $SG = \{i | d_i \leq h\}$ and deformed $DG = \{i | d_i > h\}$. Then we conduct DAP respectively on DG and SG: as shown in Eq.3, important score is computed for SG based on opacity and normalized volume, and for DG based on original volume and its deformation,

$$IS_i = \begin{cases} \sum_{t}^{T} \sum_{k}^{HW} \mathbb{1}(G(\boldsymbol{X}_i), \boldsymbol{p}_k, t) \cdot \sigma_i \cdot V_{norm}(\boldsymbol{s_i}), & i \in SG \\ \sum_{t}^{T} \sum_{k}^{HW} \mathbb{1}(G(\boldsymbol{X}_i), \boldsymbol{p}_k, t) \cdot \Delta V(\boldsymbol{s}) \cdot V_{norm}(\boldsymbol{s_i}), & i \in DG \end{cases} \tag{3}$$

where $V_{norm}(\boldsymbol{s}) = (V(\boldsymbol{s})/V_{max90})^{\beta}$ reflects the normalized volume with a norm factor β and V_{max90} reflects the 90% largest volume of all sorted Gaussians. Then we prune the Gaussians with lower important scores in each of the two classes accordingly, which can remove unimportant Gaussians while retaining those crucial for deformation.

2.3 Gaussian-Attribute Pruning

Gaussian Attributes are redundant to represent the intricate environment of surgical scenes caused by repetitive textures and view-dependent lighting. SH coefficients contain 48 floating-point values and represent over 80% of all attributes for each Gaussian [4], which is far more than other attributes that cannot be pruned, such as position, rotation, scaling factor, and opacity. Based on this, GAP reduces the higher degree SH coefficients used to model the view-dependent color and scene reflection. To enhance model memory efficiency and suitability for surgical scenes, we use a threshold to represent the pruned degree of SH coefficients [4]. As described in Eq. 4, GAP adjusts the SH degree from high to low, which effectively reduces the redundancy of each Gaussian's attributes.

$$\alpha_{ic} = \alpha_{ic} * \mathbb{1}(c \leq (h_{sh}+1)^2 + N_{RGB}), c \in C \tag{4}$$

where h_{sh} is the threshold for low SH degree, and N_{RGB} represents the number of SH coefficients to depict RGB colors, which is normally set to 3. To take full advantage of the information contained in the pruned attributes for modeling the surface of the object, we utilize distillation to transfer knowledge from SH coefficients at higher degree, which will be detailed in Sec. 2.5.

2.4 Feature Field Condensation

The deformation module including a spatial-temporal feature field E and a tiny MLP F is used to learn the representation of 4D data and the deformation at different timestamps. Despite the existence of methods such as Gaussian pruning and vector quantization [4,9] that aim to reduce the size of GS-based models, they are primarily designed for static scenes and are unable to address memory issues. To represent the position of the Gaussian and the query time, a spatial-temporal feature field with higher resolution is used for 4D encoding. Though this approach allows the model to capture more details of deformation, it also requires a significant amount of storage space, accounting for over 80% of the required storage, which hinders model deployment on robotic surgical devices.

Motivated by these observations, we propose the FFC to make our model more memory-efficient and easier to deploy on robotic surgical devices. Considering that nearby 3D Gaussians always share similar spatial and temporal information which means adjacent values on the same voxel plane should also be similar. To reduce the size of the voxel plane and preserve every 4D feature as much as possible, we conduct 3D adaptive pooling on each 4D voxel plane,

which is described in Eq. 5. The tiny MLP F remains to preserve the ability of deformation calculation,

$$E'_{\Phi}(i,j) = \frac{1}{r_{\Phi_1} r_{\Phi_2}} \sum_{i'=i \cdot r_{\Phi_1}}^{(i+1)\cdot r_{\Phi_1}} \sum_{j'=j \cdot r_{\Phi_2}}^{(j+1)\cdot r_{\Phi_2}} E_{\Phi}(i',j') \tag{5}$$

where $\Phi = (\Phi_1, \Phi_2) \in \{(x,y),(x,z),(y,z),(x,t),(y,t),(z,t)\}$ devotes the sub-planes of the spatial-temporal feature field, and r_{Φ_i}represents the compression rate of the feature field on the Φ_i axis.

2.5 Optimization

To achieve the balance between efficient memory and high-quality rendering performance, we use knowledge distillation [7,10] during optimization. We treat the uncompressed and well-trained model as teacher model, and the model processed by DAP, GAP, and FFC as the student model. We minimize the loss $L = L_d + L_r$ to better transfer knowledge from the trained teacher model to the memory-efficient student model and achieve a memory-efficient and high-quality model [7,14]. L_d is the distillation loss between the rendered images of teacher model and student model, and L_r is the rendering loss between the rendered image of student model and the ground truth,

$$L_d = \frac{1}{T}\sum_{t}^{T} \|\hat{\boldsymbol{I}}_{\boldsymbol{tch}}(t) - \hat{\boldsymbol{I}}_{\boldsymbol{stu}}(t)\|_2, \quad L_r = \frac{1}{T}\sum_{t}^{T} \|\boldsymbol{I}_{\boldsymbol{gt}}(t) - \hat{\boldsymbol{I}}_{\boldsymbol{stu}}(t)\|_2 \tag{6}$$

where T is the number of training timestamps, $\hat{\boldsymbol{I}}_{\boldsymbol{tch}}(t)$, $\hat{\boldsymbol{I}}_{\boldsymbol{stu}}(t)$ and $\boldsymbol{I}_{\boldsymbol{gt}}(t)$ are the rendered image of teacher model and student model, and the ground truth image at timestamp t.

3 Experiments

3.1 Experimental Settings

Datasets We conduct experiments on two widely used datasets: ENDONERF [26] and SCARED [2]. ENDONERF [26] consists of two public cases of in-house DaVinci robotic prostatectomy data, each depicting a single-view scene with non-rigid deformation and tool occlusion. Following [31], we use 5 keyframes of SCARED [2] captured by a da Vinci Xi surgical robot. To align with prior work [15,31], we divided each keyframe into 7:1 training and testing sets.
Compared Methods We compare with recent reconstruction methods of dynamic surgical scenes, including NeRF-based methods: EndoNeRF [26], EndoSurf [31], LerPlane [28], and GS-based methods: EndoGaussian [15], EndoGS [34]. We use PSNR, SSIM, and LPIPS as metrics for rendering quality. Following [4], we use the storage size of the model and frames-per-second (FPS)

Table 1. Experimental results on ENDONERF [26] and SCARED [2]

Dataset	Method	Size(↓)	FPS(↑)	SSIM(↑)	PSNR(↑)	LPIPS(↓)
ENDONERF [26]	EndoNeRF [26]	13.00MB	0.035	0.933	36.06	0.089
	EndoSurf [31]	20.00MB	0.040	0.954	36.53	0.074
	LerPlane-9k [28]	274.0MB	0.911	0.926	34.99	0.080
	LerPlane-32k [28]	274.0MB	0.872	0.950	37.38	0.047
	EndoGS [34]	322.7MB	91.75	0.963	37.29	0.045
	EndoGaussian [15]	334.5MB	166.5	0.960	37.78	0.053
	LGS (Ours)	25.00MB	188.3	0.955	37.48	0.068
SCARED [2]	EndoNeRF [26]	6.900MB	0.016	0.768	24.35	0.397
	EndoSurf [31]	14.00MB	0.009	0.802	25.02	0.356
	EndoGaussian [15]	184.0MB	170.56	0.825	26.89	0.272
	LGS (Ours)	20.40MB	194.66	0.826	27.05	0.297

Table 2. Ablation Study on ENDONERF [26] for each component of LGS

Model	Overall Size↓	GS Size↓	Deform Size↓	FPS↑	SSIM↑	PSNR↑	LPIPS↓
w/o DAP	56.70MB	33MB	23.70MB	74.44	0.963	38.36	0.051
w/o GAP	26.30MB	5.0MB	21.30MB	187.9	0.950	37.08	0.096
w/o FFC	329.6MB	3.3MB	326.3MB	177.9	0.945	35.82	0.082
Full model (Ours)	**24.50MB**	**3.3MB**	**21.20MB**	**188.5**	0.957	38.08	0.079

during inference as metrics for memory efficiency and rendering speed respectively.

Implementation Details We use our checkpoints with the training details described in EndoGaussian [15], where the initial number of Gaussians is 3000. For ablation experiments, 30000 Gaussians are used for initialization For DAP, we set the h to 0.5 and β to 0.1. For GAP, we set h_{sh} to 2. For FFC, we set the resolution of the feature field to $[16, 16, 16, 25]$. We implement our framework with Pytorch and use the differential rasterization [4] as the render engine. The training of LGS needs about 2GB GPU memory and 3 minutes.

3.2 Experimental Results

The experimental results on ENDONERF [26] and SCARED [2] are presented in Table 1. It can be observed that the model size of LGS achieved a compression rate of 15× and 9× compared to EndoGaussian [15] on ENDONERF and SCARED, respectively. Moreover, LGS exhibits a model size on the same order of magnitude as EndoNeRF [26] and EndoSurf [31], inferring that LGS is as memory-efficient as NeRF-based methods. The rendering quality difference between LGS and GS-based methods can be ignored with 0.3 lower in PSNR and 0.05 lower in SSIM on ENDONERF, while LGS outperforms EndoGaussian on SCARED in SSIM and PSNR. Moreover, the visual results in Fig. 2 show that LGS can render the details in surgical scenes as well as the GS-based

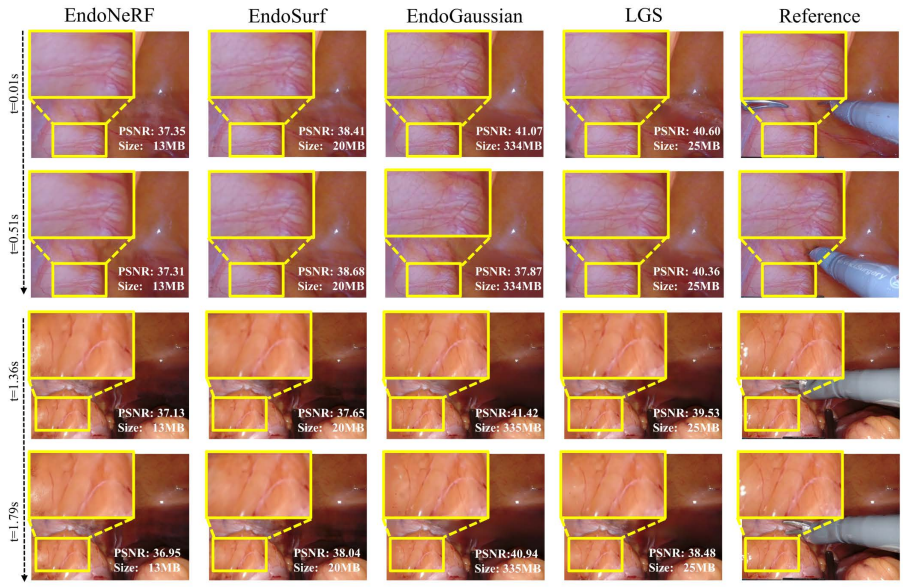

Fig. 2. Rendered images of previous methods and ours: PSNR reflects the quality of the shown image and Size reflects the used memory to store model.

method. LGS has a higher rendering speed than EndoGaussian and EndoGS, 21.8 and 87.9 higher on FPS respectively. However, there is a clear gap between GS-based methods and NeRF-based methods in terms of rendering speed and quality. Therefore, LGS can achieve high-quality and real-time rendering for dynamic surgical scenes with memory efficiency similar to NeRF-based methods.

3.3 Ablation Studies

To prove the effectiveness of each component of LGS, we conduct ablation experiments on the ENDONERF [26] dataset. and experimental results are presented in Tab. 2. We can observe that DAP prunes the least unimportant Gaussians, GAP reduces the number of parameters of each Gaussian by 40%, and FFC is also significant for memory efficiency, as it reduces the resolution of spatial-temporal feature field. The use of GAP and FFC can also improve the performance of the model, which respectively solve the repetitive textures mentioned in Sec. 2.3 and utilize the spatial-temporal similarity of nearby Gaussians mentioned in Sec. 2.4. Moreover, Tab. 2 suggests that the three proposed methods all contribute positively to the enhancement of FPS. Overall, the three parts of LGS are all important to achieve high-quality and real-time rendering of dynamic surgical scenes with memory efficiency.

4 Conclusion

This paper presents Light-weight 4D Gaussian Splatting (LGS), a breakthrough method for endoscopic reconstruction that addresses the challenges of rendering efficiency and memory efficiency. Specifically, we employ Deformation-Aware Pruning to minimize the redundancy of unnecessary Gaussian quantities and conduct Gaussian-Attribute Pruning to prune the dimensions of Gaussian attributes which simplifies the representation of textures and lighting in non-crucial areas. Additionally, we use Feature Field Condensation to tackle high-resolution redundancy in the spatial-temporal feature field for dynamic scenes. Our experiments on public datasets demonstrate that LGS achieves a compression rate exceeding 9 $\times$ while maintaining high-quality and real-time rendering. With the efficient memory usage and real-time rendering capabilities of LGS, we take a significant step towards the practical deployment of 4D Gaussian Splatting in robotic surgical services and point-of-care scenarios.

Acknowledgement. This work was supported by Hong Kong Innovation and Technology Commission Innovation and Technology Fund ITS/229/22 and National Natural Science Foundation of China under Grant 62001410.

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

References

1. Ahmed, M.S., Giordano, S.: Pre-trained lightweight deep learning models for surgical instrument detection: Performance evaluation for edge inference. In: GLOBECOM 2023-2023. pp. 3873–3878. IEEE (2023)
2. Allan, M., Mcleod, J., Wang, C., Rosenthal, J.C., Hu, Z., Gard, N., Eisert, P., Fu, K.X., Zeffiro, T., Xia, W., et al.: Stereo correspondence and reconstruction of endoscopic data challenge. arXiv preprint arXiv:2101.01133 (2021)
3. Batlle, V.M., Montiel, J.M., Fua, P., Tardós, J.D.: Lightneus: Neural surface reconstruction in endoscopy using illumination decline. In: MICCAI. pp. 502–512. Springer (2023)
4. Fan, Z., Wang, K., Wen, K., Zhu, Z., Xu, D., Wang, Z.: Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. arXiv preprint arXiv:2311.17245 (2023)
5. Gao, W., Tedrake, R.: Surfelwarp: Efficient non-volumetric single view dynamic reconstruction. In: RSS XIV (2019)
6. He, Z., Li, W., Zhang, T., Yuan, Y.: H 2 gm: A hierarchical hypergraph matching framework for brain landmark alignment. In: MICCAI. pp. 548–558. Springer (2023)
7. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
8. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)

9. Lee, J.C., Rho, D., Sun, X., Ko, J.H., Park, E.: Compact 3d gaussian representation for radiance field. arXiv preprint arXiv:2311.13681 (2023)
10. Li, C., Lin, M., Ding, Z., Lin, N., Zhuang, Y., Huang, Y., Ding, X., Cao, L.: Knowledge condensation distillation. In: ECCV. pp. 19–35. Springer Nature Switzerland Cham (2022)
11. Li, C., Liu, H., Liu, Y., Feng, B.Y., Li, W., Liu, X., Chen, Z., Shao, J., Yuan, Y.: Endora: Video generation models as endoscopy simulators. arXiv preprint arXiv:2403.11050 (2024)
12. Li, C., Zhang, Y., Li, J., Huang, Y., Ding, X.: Unsupervised anomaly segmentation using image-semantic cycle translation. arXiv preprint arXiv:2103.09094 (2021)
13. Li, Y., Richter, F., Lu, J., Funk, E.K., Orosco, R.K., Zhu, J., Yip, M.C.: Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics. RA-L **5**(2), 2294–2301 (2020)
14. Liang, Z., Rong, Y., Li, C., Zhang, Y., Huang, Y., Xu, T., Ding, X., Huang, J.: Unsupervised large-scale social network alignment via cross network embedding. In: CIKM. pp. 1008–1017 (2021)
15. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv preprint arXiv:2401.12561 (2024)
16. Liu, Y., Li, W., Liu, J., Chen, H., Yuan, Y.: Grab-net: Graph-based boundary-aware network for medical point cloud segmentation. TMI (2023)
17. Lu, J., Jayakumari, A., Richter, F., Li, Y., Yip, M.C.: Super deep: A surgical perception framework for robotic tissue manipulation using deep learning for feature extraction. In: ICRA. pp. 4783–4789. IEEE (2021)
18. Luo, H., Wang, C., Duan, X., Liu, H., Wang, P., Hu, Q., Jia, F.: Unsupervised learning of depth estimation from imperfect rectified stereo laparoscopic images. COMPUT BIOL MED **140**, 105109 (2022)
19. 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)
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Commun. ACM **65**(1), 99–106 (2021)
21. Peng, J., Sun, W., Li, H.C., Li, W., Meng, X., Ge, C., Du, Q.: Low-rank and sparse representation for hyperspectral image processing: A review. IEEE Geosci **10**(1), 10–43 (2021)
22. Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.: Tracking monocular camera pose and deformation for slam inside the human body. In: IROS. pp. 5278–5285. IEEE (2022)
23. Sherif, Y.A., Adam, M.A., Imana, A., Erdene, S., Davis, R.W.: Remote robotic surgery and virtual education platforms: How advanced surgical technologies can increase access to surgical care in resource-limited settings. In: SEMIN PLAST SURG. Thieme Medical Publishers, Inc. (2023)
24. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. RA-L **3**(1), 155–162 (2017)
25. Tang, R., Ma, L.F., Rong, Z.X., Li, M.D., Zeng, J.P., Wang, X.D., Liao, H.E., Dong, J.H.: Augmented reality technology for preoperative planning and intraoperative navigation during hepatobiliary surgery: a review of current methods. HBPD INT **17**(2), 101–112 (2018)

26. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In: MICCAI. pp. 431–441. Springer (2022)
27. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Xinggang, W.: 4d gaussian splatting for real-time dynamic scene rendering. arXiv preprint arXiv:2310.08528 (2023)
28. 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)
29. 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)
30. Yifan, W., Serena, F., Wu, S., Öztireli, C., Sorkine-Hornung, O.: Differentiable surface splatting for point-based geometry processing. TOG **38**(6), 1–14 (2019)
31. Zha, R., Cheng, X., Li, H., Harandi, M., Ge, Z.: Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In: MICCAI. pp. 13–23. Springer (2023)
32. Zhao, H., Zhao, X., Zhu, L., Zheng, W., Xu, Y.: Hfgs: 4d gaussian splatting with emphasis on spatial and temporal high-frequency components for endoscopic scene reconstruction. arXiv preprint arXiv:2405.17872 (2024)
33. Zhou, H., Jayender, J.: Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos. In: MICCAI. pp. 331–340. Springer (2021)
34. Zhu, L., Wang, Z., Jin, Z., Lin, G., Yu, L.: Deformable endoscopic tissues reconstruction with gaussian splatting. arXiv preprint arXiv:2401.11535 (2024)