MICCAI

Instrument-Splatting: Controllable Photorealistic Reconstruction of Surgical Instruments Using Gaussian Splatting

Shuojue Yang[1]⋆, Zijian Wu[2]⋆, Mingxuan Hong[1], Qian Li[1], Daiyun Shen[1], Septimiu E. Salcudean[2], and Yueming Jin[1](✉)

[1] National University of Singapore, Singapore, Singapore
ymjin@nus.edu.sg
[2] The University of British Columbia, Vancouver, Canada

Abstract. Real2Sim is becoming increasingly important with the rapid development of surgical artificial intelligence (AI) and autonomy. In this work, we propose a novel Real2Sim methodology, *Instrument-Splatting*, that leverages 3D Gaussian Splatting to provide fully controllable 3D reconstruction of surgical instruments from monocular surgical videos. To maintain both high visual fidelity and manipulability, we introduce a geometry pre-training to bind Gaussian point clouds on part mesh with accurate geometric priors and define a forward kinematics to control the Gaussians as flexible as real instruments. Afterward, to handle unposed videos, we design a novel instrument pose tracking method leveraging semantics-embedded Gaussians to robustly refine per-frame instrument poses and joint states in a render-and-compare manner, which allows our instrument Gaussian to accurately learn textures and reach photorealistic rendering. We validated our method on 2 publicly released surgical videos and 4 videos collected on *ex vivo* tissues and green screens. Quantitative and qualitative evaluations demonstrate the effectiveness and superiority of the proposed method. Our code is available at https://github.com/jinlab-imvr/Instrument-Splatting.

Keywords: Texture Learning · Surgical Instrument · 3D Gaussian Splatting · Controllable Gaussian Splatting.

1 Introduction

In robot-assisted laparoscopic surgery, computer vision-based surgical instrument identification [2,6,7,22,29,27] is of fundamental significance in downstream applications such as augmented reality and autonomous robotic surgery. While deep learning-based methods demonstrated excellent performance in general computer vision domains, adapting these methods to surgical instrument analysis remains challenging due to limited data and annotations.

For robotic surgery automation, existing public datasets either lack instrument pose labels or only provide sole wrist part pose labels [24]. This limits

---
⋆ Equal contribution

imitation learning-based methods where robots directly learn instrument motions from human expert surgical videos [10,19]. To address this issue, a feasible solution is Sim2Real transfer learning, which leverages large amounts of trajectories and videos generated from the simulation environment [17]. Conventionally, most studies [4] adopt CAD mesh models of instruments that exhibit visually unrealistic appearances. This significant sim-to-real gap impairs the performance of models trained on such data. Therefore, a highly realistic and controllable 3D asset of surgical instruments is critical.

With the recent progress in 3D reconstruction, new 3D representations such as Neural Radiance Field (NeRF) [16] and 3D Gaussian Splatting (GS) [9] emerged, which serve as powerful tools that can provide reconstructions with high visual fidelity. Driven by these advances, Zeng et al. [28] utilize 3D GS to reconstruct surgical instruments for novel data synthesis assisting surgical instrument detection. However, this method requires a customized system for data collection and cannot dynamically model surgical instruments with joint changes, which degrades the flexibility and controllability of the reconstructed 3D asset. Similarly, [20] produces instrument reconstruction using large reconstruction model, but lacks controllability. Although previous methods [13,26,12,30,21,25] for dynamic surgical scene reconstruction have demonstrated high visual fidelity on deformable tissue reconstruction, they cannot provide controllable 3D GS since the deformation fields are only conditioned on the timestamps in surgical videos. In addition, the reconstruction quality of these methods is degraded with large and complex instrument motions, which commonly occur in surgical scenarios. Meanwhile, there are emerging works that adopt GS to reconstruct the general robotic arm [14,15] or articulated human body [11], these approaches assume known (or well-estimated) poses and joint states of the articulated objects, which is unpractical in robotic-surgery settings where the kinematics of surgical instruments are unavailable/noisy [2].

Driven by these limitations, we propose *Instrument-Splatting*, a real-to-sim framework which leverages real monocular surgical videos to create a surgical instrument digital twin represented by controllable 3D GS that can be used for photorealistic simulation. Given untextured instrument CAD models, our pipeline starts with a geometry pretraining that effectively binds the GS to the mesh models with accurate geometry. Next, we propose a novel pose estimation and tracking strategy based on the pretrained GS to estimate the instrument pose and joint states. To handle large inter-frame instrument motions, we first introduce a correspondence matching module to guide the GS to move toward the current frame pose, followed by a structural-prior-based loose regularization to penalize singular pose estimations. Finally, we evaluate our framework on both intraoperative videos selected from EndoVis2017&2018 and in-house dataset collected using da Vinci robotic system. Experiments indicate the efficacy of our approach, demonstrating superior pose estimation accuracy and visual fidelity.
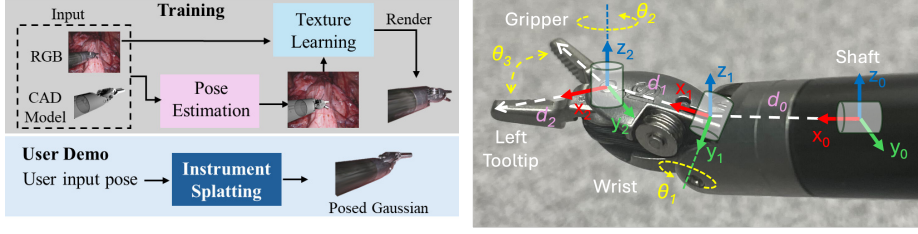
Fig. 1. Instrument-Splatting Pipeline.   **Fig. 2.** Forward kinematics of the LND.

2   Method

Instrument-splatting aims to learn a Gaussian Splatting representation of articulated surgical instruments. The overview pipeline is shown in Fig. 1. During training, we first estimate the instrument pose and joint state given an RGB image and instrument CAD model. Then, images with corresponding poses are fed into the texture learning module to optimize the instrument GS. After training, users can input in-ranged instrument pose with joint variables and get the posed mesh along with rendered photorealistic images. In the following, we present the preliminary knowledge including the instrument forward kinematics and 3D GS (Section 2.1), GS geometry pretraining (Section 2.2), instrument pose estimation & tracking (Section 2.3), and the final texture learning (Section 2.4).

2.1   Preliminary Knowledge

Forward Kinematics. In this study, we adopt the da Vinci EndoWrist Large Needle Driver (LND) as an example surgical instrument. The CAD model of this instrument is open source [8]. Note that the proposed method can be easily adapted to other types of instruments, given their CAD models.

The definition of coordinate systems and forward kinematics diagram are shown in Fig. 2. We regard shaft frame as the base frame and this frame definition consists of two revolute joint variables θ_1 and θ_2, denoting the rotation of the axis z_1 and y_2, and the gripper open angle θ_3. Let $q = \{\theta_1, \theta_2, \theta_3\} \in \mathbb{R}^3$ represent the instrument joint state. Given the translation d_0 and d_1 as 215.9 mm and 9 mm, the forward kinematics chains can be defined as ${}^sT_w(\theta_1) \, {}^wT_g(\theta_2) \, {}^gT_l(\theta_3)$, where ${}^aT_b \in SE(3)$ denotes the rigid transformation from frame a to b, and s, w, g and l denote the frames for shaft, wrist, gripper and left tool tips, respectively. Of note, each of them has only one degree of freedom (DoF) controlled by $\theta_i \in q$. Therefore, we can define a transformation from any joint frame j to the shaft frame as ${}^sT_{j\in\{w, g, l\}}$ given corresponding joint states q. Next, with another transformation from shaft to camera ${}^{cam}T_s$, we define a transform function $\boldsymbol{T_j}(\cdot)$ to transform points $x_j \in \mathbb{R}^3$ in joint frames j to the camera frame as:

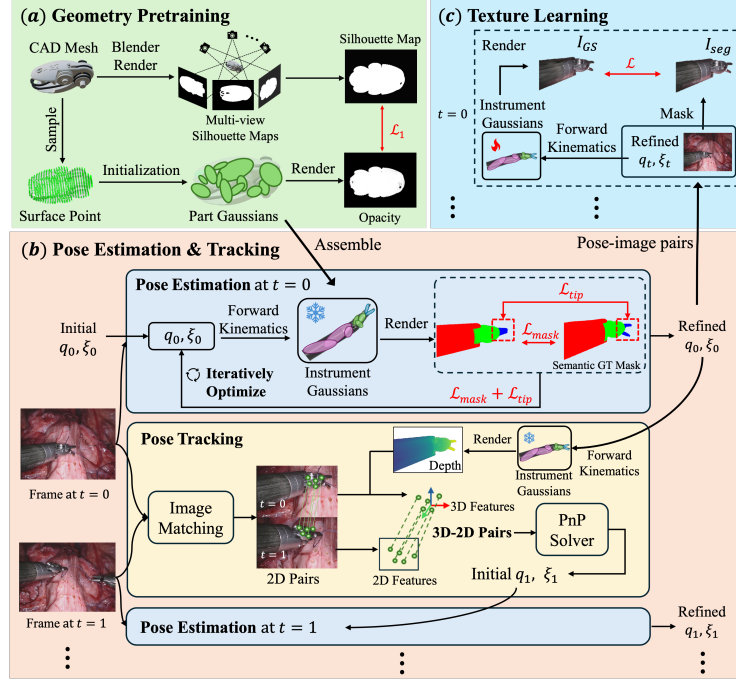$$\boldsymbol{T_j}(x_j; \ q, \xi) = {}^{cam}T_s(\xi) \, {}^sT_j(q) \, x_j, \tag{1}$$

Fig. 3. Overview diagram of Instrument-Splatting methodology.

where we use $\xi \in se(3)$ to parameterize ${}^{cam}T_s$. With Eq. 1, we can define the instrument pose in camera frame given specified (q, ξ).

3D Gaussian Splatting. 3D Gaussian Splatting (GS) [9] is a powerful 3D representation that can reconstruct 3D scenes with a group of Gaussian points. Each Gaussian point contains learnable attributes $\Theta = \{\mu \in \mathbb{R}^3, r \in \mathbb{R}^4, s \in \mathbb{R}^3, \alpha \in \mathbb{R}, sh \in \mathbb{R}^{27}\}$ to define the position, rotation, scale, opacity and colors. With camera parameters, α-blending [9] is performed to render the colored images based on the attributes in 3D GS model.

2.2   Geometry Pretraining

To bind the GS points to each rigid part of the instrument and initialize them with accurate geometric priors, we propose a geometry pretraining strategy, as illustrated in Fig. 3 (a). Specifically, for each part, we densely sample points on mesh surface by ray-tracing, which effectively avoids the sparse point distribution due to oversized triangles. Then, given the CAD model, we adopt Blender [3] to render the ground-truth silhouette maps from multiple viewpoints. Next, the sampled GS points will be trained to learn the geometry information by minimizing the $\mathcal{L}_1$ loss between the rendered opacity and the ground-truth silhou-

---

[3] https://www.blender.org/

ette. Afterward, part Gaussians are combined to generate semantic-embedded instrument Gasussians where one semantic label (i.e., wrist, shaft, and gripper) $f \in \{1, 2, 3\}$ is added to Gaussian parameters Θ to identify the part the Gaussian point belongs to. Therefore, the instrument GS can be controlled by updating μ and r with Eq. 1 as:

$$\mu'_j = \boldsymbol{T_j}(\mu_j; \ q, \xi); \quad r'_j = \boldsymbol{R_j}(r_j; \ q, \xi), \tag{2}$$

where μ'_j and r'_j respectively denote the updated positions and rotations for Gaussian points in part j, $\boldsymbol{R}_j$ means the rotational component derived from rigid-body transformation $\boldsymbol{T_j}$.

2.3 Pose Estimation and Tracking

An accurate alignment between the rendered instrument and the reference image is crucial for appearance learning. Therefore, we design a pose estimation and tracking module to estimate per-frame instrument pose and joint changes, as illustrated in the Fig. 3 (b). Inspired by [2], the pose estimation adopts the render-and-compare framework to align the rendered semantic silhouette with the segmented instrument. For the render-and-compare method, a good pose initialization is crucial. In the initial frame, we manually pick 2D-3D point correspondences of logo landmarks in the wrist part and use the PnP to solve the pose. Only wrist pose is needed for initialization; shaft and gripper poses are fixed relative to it and only 6 3D-2D pairs on wrist are sufficient. We cast the pose estimation task as an optimization problem that can be formulated as:

$$\widehat{\xi}, \ \widehat{q} = argmin \ \mathcal{L}(\xi, \ q) \tag{3}$$

where the loss is $\mathcal{L} = \mathcal{L}_{mask} + \mathcal{L}_{tip}$. The $\mathcal{L}_{mask}$ is $\mathcal{L}_1$ loss. To take advantage of the surgical instrument geometry prior, we propose $\mathcal{L}_{tip} = \mathcal{L}_{dist} + \mathcal{L}_{struct}$ to enforce the tool tip alignment. Given the gripper mask, we use singular value decomposition (SVD) to find the principal axis of the gripper regions, on which the farthest pixel from the wrist part is regarded as the tool tip. We adopt a distance loss $\mathcal{L}_{dist} = ReLU(d - r)$, where d is the Euclidean distance between the rendered tool tip from instrument GS and the estimated tool tip position from segmentation mask, and r is a threshold to determine if the tool tip is aligned well. This loose regularization can alleviate falling into local minima caused by incorrect tool tip detection. In addition, we design a structural loss $\mathcal{L}_{struct} = ReLU(-(\theta_l + \theta_r))$ to constrain the left tip always on the left side of the right tip. θ_l is the left gripper rotation angle, which is a positive value if it rotates leftward and *vice versa* it is a negative value.

To track the instrument pose ξ and joint state q, simply taking the previous frame pose as the initial value struggles to handle the case where the instrument motion between two adjacent frames is large. To address this issue, we propose a novel pose initialization method based on image matching and PnP for pose tracking. We extract features in the wrist and then find the matching correspondences [5] between two adjacent frames. Given the depth rendered from 3D GS,

we can lift the 2D features to 3D. We approximate the frame t's 3D features and frame $t+1$'s 2D features into 2D-3D correspondences and then compute a rough initial pose using a PnP solver for the next frame.

2.4   Texture Learning

After estimating and tracking the pose and joint state over the video, we can control the instrument Gaussian points moving to the target pose in each frame, as shown in Fig. 3 (c). To maintain the geometric details, we freeze μ in this phase, and the other parameters $\{r,\ s,\ \alpha,\ sh\}$ are learnable during this section. Here, we adopt the loss $\mathcal{L} = \mathcal{L}_{mask}+\mathcal{L}_{color}$ to learn the textures, where the $\mathcal{L}_{mask}$ and $\mathcal{L}_{color}$ are the $\mathcal{L}_1$ loss of the semantic silhouette maps and RGB pixels of the instrument, respectively. $\mathcal{L}_{mask}$ is used here to maintain the geometric structure while learning appearances.

3   Experiments and Results

3.1   Datasets and Implementation Details

We establish a new benchmark for the evaluation of surgical instrument reconstruction. It includes two widely used instrument segmentation datasets EndoVis 2017 [3] and 2018 [1]. We select 100 frames containing LND in both datasets. Frames with motion artifacts are excluded to ensure data quality. We split each video into seen and unseen views (7:1) for training and evaluation. Because of the sparse instrument poses in the *in-vivo* datasets, we collect four videos with backgrounds of *ex vivo* tissues and a green screen for fairer validation. Two trajectories with tissue background are to mimic the surgical scenario. The other two trajectories with green screen backgrounds are closer to the camera to capture more texture details. The data collection methodology follows SurgPose [23] using a da Vinci IS1200 system with the da Vinci Research Kit (dVRK) [8]. Regarding annotation, we use SAM 2 [18] to segment the foreground surgical instrument at the part level and obtain corresponding semantic masks. This newly established benchmark, based on our in-house data, will be publicly available.

In this work, we perform geometry pretraining, pose estimation, and texture learning sequentially. We first train the part Gaussian for 10K iterations per part. Then, we adopt 2K iterations to iteratively optimize the pose for each frame, and an early-stop signal will be triggered once the loss remains stable for over 100 iterations. The texture learning lasts for 10K iterations for each video. Only training requires manual initialization; In real inference, we have instrument GS and input user-defined pose for controllable data generation. Therefore, we estimate the instrument poses and use these poses as user input to control the instrument GS for quantitative evaluation. All the experiments are based on the PyTorch framework and conducted with a single NVIDIA RTX A5000 GPU.

Table 1. Quantitative Results of the Key Components

Method	Rendered Semantics Dice↑			Reconstruction Quality		
	Shaft	Wrist	Gripper	PSNR↑	SSIM↑	LPIPS↓
$w/o\ \mathcal{L}_{tip}$	94.625	83.94	63.165	21.77	91.85	0.091
$w/o\ Reg$	95.515	78.4	63.39	22.89	92.63	0.079
w/o GP	-	-	-	14.6	89.54	0.16
Ours	**96.83**	**86.65**	**73.32**	**23.84**	**93.10**	**0.071**

3.2 Quantitative Evaluation of Key Components

First, we verify the effectiveness of the key components of Instrument-Splatting on the EndoVis 2017 and 2018 datasets. As shown in Table 1, we quantify the pose accuracy of our pose estimation and tracking module with 2D projection error similar to [2]. Specifically, we adopt the difference (i.e., Dice score) between rendered semantics maps from posed instrument Gaussians and the corresponding segmentation maps to evaluate the pose accuracy. We test our method by removing the $\mathcal{L}_{tip}$ and the loose regularization strategy, denoted as $w/o\ \mathcal{L}_{tip}$ and $w/o\ Reg$, respectively. As shown in Table 1, both $\mathcal{L}_{tip}$ and the loose regularization lead to significantly improved pose estimation accuracy, particularly for the gripper. We also report the metrics of the reconstruction quality in Table 1, where it is noticeable that an improved pose estimation can enhance the visual fidelity. Table 1 indicates the importance of Geometry Pretraining (GP), which significantly boosts the reconstruction quality. As shown in Fig. 4, six examples are selected from EndoVis18 and in-house dataset for visualization, which indicates that our method can robustly track the instrument pose.
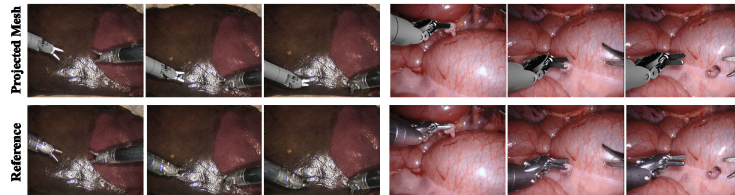
Fig. 4. Visualization of the projected CAD mesh with the estimated poses and corresponding original images.

3.3 Evaluation on Reconstruction Quality

In this work, we are targeting a novel setting that reconstructs articulated surgical instruments with monocular videos and CAD models only. We compare the reconstruction performance of our method with current state-of-the-art reconstruction methods in the surgical domain: EndoGaussian [13] and Deform3DGS [26]. Although both methods cannot provide controllable 3D assets,
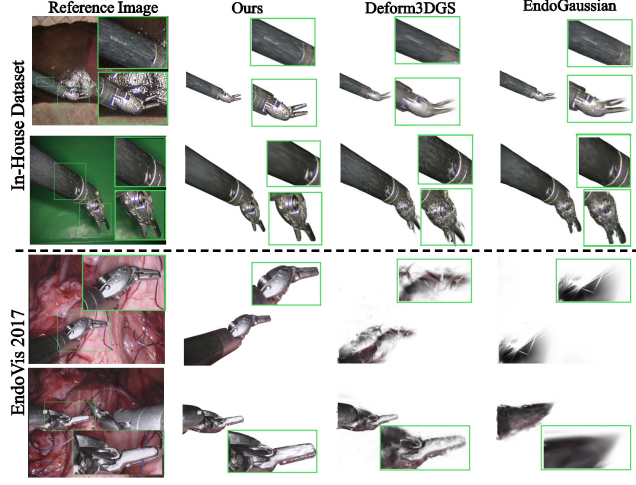
Fig. 5. Visualization of the Novel-view Rendering Results of Different Methods

they demonstrate high visual fidelity in the reconstruction of deformable tissues. As shown in Table 2, we show the results of wrist&gripper, shaft, and the whole

Table 2. Quantitative evaluation of the reconstruction quality on EndoVis2017&2018 and In-House Dataset.

Dataset	Method	Wrist & Gripper			Shaft			Overall Instrument		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
EndoVis17&18	Deform3DGS	19.65	92.55	0.165	22.23	94.33	0.066	17.44	86.49	0.213
	EndoGaussian	18.90	93.45	0.145	20.91	94.30	0.078	16.45	87.75	0.192
	Ours	**25.81**	**96.87**	**0.064**	**31.39**	**96.52**	**0.035**	**23.84**	**93.10**	**0.071**
In-House	Deform3DGS	28.95	97.60	0.046	**34.52**	97.08	**0.023**	**28.39**	94.76	0.048
	EndoGaussian	25.34	96.94	0.059	31.59	**97.23**	0.033	27.47	94.94	0.053
	Ours	**30.44**	**98.27**	**0.038**	32.49	97.14	0.032	27.94	**95.45**	**0.041**

instrument separately. The metrics for the whole instrument are lower than the other two since they are computed based on all the pixels, following [13]. We visualize the rendered images of different methods in Fig. 5. Both EndoGaussian and Deform3DGS fail to reconstruct instruments in real surgical videos (EndoVis2017&2018) due to long-range motions and large time intervals. As a comparison, benefiting from the pose estimation module, our method can effectively capture the significant inter-frame pose change for real surgical videos, leading to a largely improved accuracy. On in-house data with smoother and slower motions, although Deform3DGS/EndoGaussian have comparable or even higher PSNR scores, their reconstruction outputs suffer from significant artifacts on wrist and gripper parts and are far from being photorealistic, as shown in the visualization (Fig. 5). This can be explained by the motion patterns in some

clips of our in-house data where the shaft part only undergoes limited translation, and Deform3DGS/EndoGaussian has a strong capability of capturing the appearance of these regions. Also, Table 2 shows the two methods exhibit an inferior reconstruction quality on wrist and gripper parts, which indicates the existing methods on surgical scene reconstruction cannot handle articulated instrument parts with high dexterity and complex motions.

4   Conclusion

In this paper, we propose Instrument-Splatting, a novel Real2Sim pipeline to reconstruct articulated robotic surgical instruments with high visual fidelity. With only monocular videos and CAD models of the instruments, our Instrument-Splatting can effectively track the per-frame instrument pose and learn the texture through RGB videos. Different from prior works in the surgical domain, our instrument Gaussian is articulated and fully controllable, which allows for manipulation as flexible as real instruments. We believe this digital twin of the instrument can strongly benefit the community working on surgical AI or surgical autonomy. The experimental results demonstrate the superior advances of our proposed method.

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

References

1. Allan, M., Kondo, S., Bodenstedt, S., Leger, S., Kadkhodamohammadi, R., Luengo, I., Fuentes, F., Flouty, E., Mohammed, A., Pedersen, M., et al.: 2018 robotic scene segmentation challenge. arXiv preprint arXiv:2001.11190 (2020)
2. Allan, M., Ourselin, S., Hawkes, D.J., Kelly, J.D., Stoyanov, D.: 3-d pose estimation of articulated instruments in robotic minimally invasive surgery. IEEE transactions on medical imaging **37**(5), 1204–1213 (2018)
3. Allan, M., Shvets, A., Kurmann, T., Zhang, Z., Duggal, R., Su, Y.H., Rieke, N., Laina, I., Kalavakonda, N., Bodenstedt, S., et al.: 2017 robotic instrument segmentation challenge. arXiv preprint arXiv:1902.06426 (2019)
4. Barragan, J.A., Zhang, J., Zhou, H., Munawar, A., Kazanzides, P.: Realistic data generation for 6d pose estimation of surgical instruments. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 13347–13353. IEEE (2024)
5. Chen, H., Luo, Z., Zhou, L., Tian, Y., Zhen, M., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: Aspanformer: Detector-free image matching with adaptive span transformer. European Conference on Computer Vision (ECCV) (2022)

6. Chen, Y., Wu, Z., Schmidt, A., Salcudean, S.E.: A-mfst: Adaptive multi-flow sparse tracker for real-time tissue tracking under occlusion. arXiv preprint arXiv:2410.19996 (2024)
7. Du, X., Kurmann, T., Chang, P.L., Allan, M., Ourselin, S., Sznitman, R., Kelly, J.D., Stoyanov, D.: Articulated multi-instrument 2-d pose estimation using fully convolutional networks. IEEE transactions on medical imaging **37**(5), 1276–1287 (2018)
8. Kazanzides, P., Chen, Z., Deguet, A., Fischer, G.S., Taylor, R.H., DiMaio, S.P.: An open-source research kit for the da vinci® surgical system. In: 2014 IEEE international conference on robotics and automation (ICRA). pp. 6434–6439. IEEE (2014)
9. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)
10. Kim, J.W., Zhao, T.Z., Schmidgall, S., Deguet, A., Kobilarov, M., Finn, C., Krieger, A.: Surgical robot transformer (srt): Imitation learning for surgical tasks. arXiv preprint arXiv:2407.12998 (2024)
11. Kocabas, M., Chang, J.H.R., Gabriel, J., Tuzel, O., Ranjan, A.: Hugs: Human gaussian splats. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 505–515 (June 2024)
12. 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)
13. Liu, Y., Li, C., Yang, C., Yixuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. Arxiv (2024)
14. Lou, H., Liu, Y., Pan, Y., Geng, Y., Chen, J., Ma, W., Li, C., Wang, L., Feng, H., Shi, L., et al.: Robo-gs: A physics consistent spatial-temporal model for robotic arm with hybrid representation. arXiv preprint arXiv:2408.14873 (2024)
15. Lu, G., Zhang, S., Wang, Z., Liu, C., Lu, J., Tang, Y.: Manigaussian: Dynamic gaussian splatting for multi-task robotic manipulation. In: European Conference on Computer Vision. pp. 349–366. Springer (2024)
16. 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)
17. Munawar, A., Wu, J.Y., Fischer, G.S., Taylor, R.H., Kazanzides, P.: Open simulation environment for learning and practice of robot-assisted surgical suturing. IEEE Robotics and Automation Letters **7**(2), 3843–3850 (2022)
18. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
19. Reiley, C.E., Plaku, E., Hager, G.D.: Motion generation of robotic surgical tasks: Learning from expert demonstrations. In: 2010 Annual international conference of the IEEE engineering in medicine and biology. pp. 967–970. IEEE (2010)
20. Wang, X., Zhang, S., Huang, B., Stoyanov, D., Mazomenos, E.B.: Endolrmgs: Complete endoscopic scene reconstruction combining large reconstruction modelling and gaussian splatting
21. 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)
22. Wu, Z., Schmidt, A., Kazanzides, P., Salcudean, S.E.: Augmenting efficient real-time surgical instrument segmentation in video with point tracking and segment anything. Healthcare Technology Letters (2024)

23. Wu, Z., Schmidt, A., Moore, R., Zhou, H., Banks, A., Kazanzides, P., Salcudean, S.E.: Surgpose: a dataset for articulated robotic surgical tool pose estimation and tracking. arXiv preprint arXiv:2502.11534 (2025)
24. Xu, H., Weld, A., Xu, C., Roddan, A., Cartucho, J., Karaoglu, M.A., Ladikos, A., Li, Y., Li, Y., Shen, D., et al.: Surgripe challenge: Benchmark of surgical robot instrument pose estimation. arXiv preprint arXiv:2501.02990 (2025)
25. 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)
26. Yang, S., Li, Q., Shen, D., Gong, B., Dou, Q., Jin, Y.: Deform3dgs: Flexible deformation for fast surgical scene reconstruction with gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 132–142. Springer (2024)
27. Ye, M., Zhang, L., Giannarou, S., Yang, G.Z.: Real-time 3d tracking of articulated tools for robotic surgery. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part I 19. pp. 386–394. Springer (2016)
28. Zeng, T., Loza Galindo, G., Hu, J., Valdastri, P., Jones, D.: Realistic surgical image dataset generation based on 3d gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 510–519. Springer (2024)
29. Zhan, B., Zhao, W., Fang, Y., Du, B., Vasconcelos, F., Stoyanov, D., Elson, D.S., Huang, B.: Tracking everything in robotic-assisted surgery. arXiv preprint arXiv:2409.19821 (2024)
30. Zhu, L., Wang, Z., Jin, Z., Lin, G., Yu, L.: Deformable endoscopic tissues reconstruction with gaussian splatting. arXiv preprint arXiv:2401.11535 (2024)