This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

EndoGSim: Physics-Aware 4D Dynamic Endoscopic Scene Simulations via MLLM-Guided Gaussian Splatting

Changjing Liu[1,⋆], Yiming Huang[1,⋆], Long Bai[1], Beilei Cui[1], and Hongliang Ren[1,2,⋆⋆]

[1] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
{changjingliu,yhuangdl}@link.cuhk.edu.hk, b.long@ieee.org, beileicui@link.cuhk.edu.hk, hlren@ee.cuhk.edu.hk
[2] Shenzhen Loop Area Institute, Shenzhen, China

Abstract. In robot-assisted minimally invasive surgery, high-fidelity dynamic endoscopic scene reconstruction and simulation are crucial to enhancing downstream tasks and advancing surgical outcomes. However, existing methods primarily focus on visual reconstruction, lacking physics-based descriptions of the scene required for realistic simulation. We propose a unified framework that achieves physics-aware reconstruction and physical simulation of endoscopic scenes through Multimodal Large Language Models (MLLMs)-guided Gaussian Splatting. Our approach utilizes 4D Gaussian Splatting (4DGS) integrated with pre-trained segmentation and depth estimation to represent deformable tissues and tools. To achieve automatic inference of physical properties, we introduce an object-wise material field that initializes material parameters via MLLM and refines them through a differentiable Material Point Method (MPM) under joint supervision from rendered images and optical flow. Validated on both open-source and in-house datasets, our framework achieves superior simulation fidelity and physical accuracy compared to state-of-the-art methods, underscoring its potential to advance robot-assisted surgical applications.

Keywords: 4DGS reconstruction · Physical simulation · Material point method · Material parameter estimation.

1 Introduction

Endoscopic procedures are central to minimally invasive surgery with reduced trauma and faster recovery, motivating the integration of robot-assisted systems to achieve superior precision, consistency, and efficiency [24,8,4]. In this context, accurate dynamic reconstruction and simulation of the endoscopic scene are essential to enable reliable spatial intelligence, support pre-operative planning,

⋆ Co-first authors.
⋆⋆ Corresponding author.

Table 1. Comparison of methods by Video input, Auto Initialization, Object-wise material field, and Physical simulation.

Method	PhysGen	Physics3D	PhysGaussian	GIC	PhysFlow	**Ours**
Video Input	✗	✗	✗	✓	✓	✓
Auto Initialization	✓	✗	✗	✗	✓	✓
Material Field	✗	✗	✗	✗	✗	✓
Physical Simulation	✗	✓	✓	✓	✓	✓

and ensure robust autonomous control [16]. Recent endoscopic scene reconstruction benefits from Neural Radiance Fields (NeRFs) [23,28,32] and 3D Gaussian Splatting (3DGS) [17,31,5], achieving superior reconstruction quality along with real-time rendering performance. Building upon this foundation, 4D Gaussian Splatting (4DGS) [29,16,10] extends static representations to the spatiotemporal domain, making it suitable for deformable surgical scenes reconstruction.

Despite remarkable advances in scene reconstruction, existing methods [17,29] primarily focus on appearance reconstruction and seldom capture the material properties of objects. This limits physically accurate spatial perception, high-fidelity dynamic simulation, and the generation of realistic data essential for surgical automation. To address this challenge, several studies integrate material property estimation and perform simulations on reconstructed deformable scenes [30,3,18,21]. PhysGaussian [30] combines 3DGS with a differentiable Material Point Method (MPM) for physics-based simulation, with material parameters specified through manual initialization. To enable automatic identification of material properties, GIC [3] captures explicit shapes using 3D Gaussian Splatting and infers physical properties through continuum mechanics for simulation. Leveraging rich visual-semantic priors, subsequent works [20,21] further integrate multi-modal large language models (MLLMs) to enhance physics-aware reconstruction and dynamic simulation. PhysGen [20] introduces physics reasoning using large pre-trained visual foundation models, eliminating the need for manual parameter initialization. Physflow [21] leverages multi-modal foundation models and video diffusion to achieve enhanced 4D dynamic scene simulation.

However, there remains a notable gap in adapting these techniques for physics-aware reconstruction and simulation of surgical scenes. In particular, 3DGS-based physical simulation methods [30,18] struggle with dynamic video input, and existing material estimation strategies [30,3,18] often lack informed initialization, potentially leading to physically inconsistent simulations. Furthermore, existing approaches mainly focus on general scenes and rarely consider multi-object material estimation, which is critical in surgical environments involving both instruments and deformable tissues. To overcome these challenges, we propose a unified framework that reconstructs deformable endoscopic scenes from video and simultaneously estimates material parameters to enable realistic physical simulation, as shown in Fig. 1. Specifically, we employ 4DGS [16] with pretrained depth and segmentation models to construct a Gaussian splat representation of the surgical scene. Then, we propose an object-wise material field to

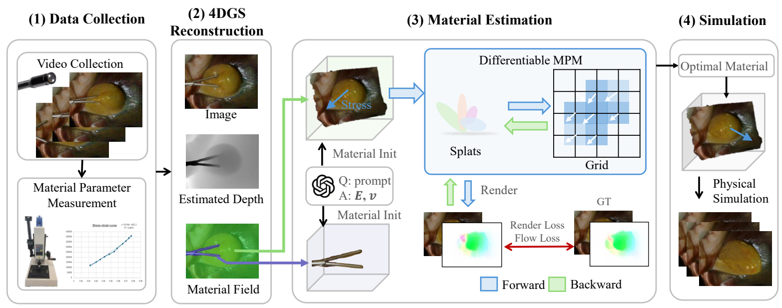

Fig. 1. Overview of our physics-aware framework for surgical scene reconstruction and 4D dynamic simulation with automatic estimation of physical parameters.

estimate the physical properties of the tissues and tools. Material parameters are automatically initialized using large pre-trained multi-modal large language models, and then refined via gradient-based optimization within a differentiable MPM, guided by rendering and optical flow, to ensure both visual realism and physical fidelity. Table 1 summarizes the comparison of our approach with existing methods. Our code is available at https://github.com/ChangjingLiu/EndoGSim. The primary contributions of this work are summarized as follows:

- We introduce a unified framework for the automatic, physics-aware reconstruction and simulation of surgical scenes from endoscopic videos.
- The material field is proposed for material estimation, coarsely initialized via MLLM estimation and then jointly refined with render and optical flow loss in a differentiable MPM, achieving physically plausible simulation.
- Experiments on open-source and in-house datasets demonstrate that our method achieves accurate physics and realistic simulations, offering substantial potential to advance robotic-assisted minimally invasive surgery.

2 Methods

2.1 Preliminaries

Inspired by [16], we reconstruct the surgical scene via 4D Gaussian $\mathcal{G}' = \mathcal{G} + \Delta\mathcal{G}$, including a static 3D Gaussian $\mathcal{G}$ and its deformation $\Delta\mathcal{G} = \mathcal{F}(\mathcal{G}, t)$, where $\mathcal{F}$ denotes the deformation network and t is the time. The spatial-temporal encoder $\mathcal{H}$ is defined to parameterize the 4D trajectory of each Gaussian point.

A multi-head Gaussian deformation decoder $\mathcal{D}$ is designed for decoding the deformation of position μ, rotation $\mathbf{R}$, scaling $\mathbf{S}$, opacity o and spherical harmonics $\mathbf{SH}$ using five efficient and lightweight tiny MLPs. The final 4D Gaussian representation can be formulated as follows:

$$\mathcal{G}' = \{\mu + \Delta\mu, \mathbf{R} + \Delta\mathbf{R}, \mathbf{S} + \Delta\mathbf{S}, o + \Delta o, \mathbf{SH} + \Delta\mathbf{SH}\}. \quad (1)$$

For dynamic scene video $V = \{I_0, ...I_t\}$, we adopt 4D Gaussian Splatting to construct a time-dependent 3D scene representation. Specifically, the π^3 model [27] is integrated to provide depth estimation, and Surgical-SAM-2 [19] is utilized for efficient segmentation of soft tissues and surgical instruments.

2.2 Continuum Mechanics

Continuum mechanics provides the foundational description of the motion of soft tissues during surgical procedures. Specifically, the deformation of a continuum is described via the time-dependent continuous map $\mathbf{x} = \phi(\mathbf{X}, t)$, where $\mathbf{X}$ represents the undeformed material space and $\mathbf{x}$ denotes the deformed space at time t. The deformation gradient $\mathbf{F}(\mathbf{X}, t) = \nabla_{\mathbf{X}} \phi(\mathbf{X}, t)$ captures the local deformation of the material, encoding its rotation, stretch, and shear.

Our work employs a hyperelastic model parameterized by a vector $\theta_p = \{E, \nu\}$, where E is Young's modulus and ν is Poisson's ratio. To accurately represent these properties, the evolution of ϕ must satisfy mass and momentum conservation: mass conservation ensures uniform density during deformation, while momentum conservation is formulated as follows:

$$\rho(\mathbf{x}, t)\dot{\mathbf{v}}(\mathbf{x}, t) = \nabla \cdot \sigma(\mathbf{x}, t) + \mathbf{f}^{ext}, \tag{2}$$

where ρ is the mass density, σ is the Cauchy stress tensor, and $\mathbf{f}^{ext}$ denotes external forces (tool-tissue interactions). The deformation gradient $\mathbf{F} = \mathbf{F}^E \mathbf{F}^P$ decomposes into elastic ($\mathbf{F}^E$) and plastic ($\mathbf{F}^P$) contributions, allowing realistic modeling of tissue elastic recoil and permanent tearing under surgical forces.

2.3 Material Point Method (MPM)

The Material Point Method (MPM) [25,7] provides an efficient solver for the continuum equations in Eq. 2 and naturally extends to point-based representations such as 3DGS, as demonstrated in PhysGaussian [30]. The MPM operates in a particle-to-grid (P2G), grid update, and grid-to-particle (G2P) transfer loop. In P2G process, MPM transfers mass and momentum from particles to grids as:

$$m_i^n = \sum_p w_{ip}^n m_p, \quad m_i^n \mathbf{v_i^n} = \sum_p w_{ip}^n m_p (\mathbf{v_p^n} + C_p^n (\mathbf{x_i} - \mathbf{x_p^n})), \tag{3}$$

where p and i denote the fields on the Lagrangian particles and the Eulerian grid, respectively. Each particle p carries a set of physical properties, including its volume V_p, mass m_p, position $\mathbf{x}_p^n$, velocity $\mathbf{v}_p^n$, deformation gradient $\mathbf{F}_p^n$, and affine momentum C_p^n at time t_n. The term w_{ip}^n represents the B-spline kernel function centered at the i-th grid node and evaluated at the particle position $\mathbf{x}_p^n$.

After the P2G transfer, the velocities on the grid are updated by integrating both internal and external forces:

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n - \frac{\Delta t}{m_i} \sum_p \tau_p^n \nabla w_{ip}^n V_p^0 + \Delta t \frac{\mathbf{f}^{ext}}{m_i}, \tag{4}$$

where τ_p^n is the Cauchy stress tensor, ∇w_{ip}^n is the gradient of the kernel function. In G2P, the velocities are transferred back to particles to update their states:

$$\mathbf{v}_p^{n+1} = \sum_i w_{ip} \mathbf{v}_i^{n+1}. \tag{5}$$

The deformation gradient $\mathbf{F}_p$ is updated to track changes, with adjustments made for plasticity as needed. Specifically, our framework builds upon differentiable MLS-MPM [14], which delivers significantly improved forward simulation stability and efficient back-propagation capabilities.

2.4 Object-wise Material Field Parameter Estimation

Material Field Initialization. Accurate material properties are essential for physically consistent deformations, and their initial values significantly influence the optimization process. In our pipeline, we first apply Surgical-SAM-2 [19] to segment video frame I, obtaining distinct masks for tissues and instruments $M_{tissue}, M_{tools} = SAM(I)$. Subsequently, we infer initial material attributes θ_p^{init} by prompting GPT-4o [1] with the segmented visual inputs: $\theta_p^{init} \leftarrow GPT(prompt, I \odot M_{tissue}, I \odot M_{tools})$. The prompts are detailed in the supplementary material.

Material Field Optimization. Optimization is selectively performed based on the object: For rigid instruments and objects that undergo negligible deformation, the initialized parameters are frozen. For deformable objects, we further optimize these parameters to better match the observed dynamics. Specifically, given the input video sequence $V = \{I_0, \ldots, I_t\}$, where I_t denotes the observed frame at the time step t, we employ a pre-trained RAFT optical flow estimator [26] to compute optical flow fields $U(I_t, I_{t+1})$ for the observed frames and $\hat{U}(\hat{I}_t, \hat{I}_{t+1})$ for the rendered frames $\hat{I}_t$. We supervise appearance (rendering loss between I_t and $\hat{I}_t$) and motion (optical flow differences) using a joint loss:

$$\mathcal{L} = \sum_t \left\| I_t - \hat{I}_t \right\|_2^2 + \lambda \sum_t \left\| U(I_t, I_{t+1}) - \hat{U}(\hat{I}_t, \hat{I}_{t+1}) \right\|_2^2, \tag{6}$$

where λ is adopted to balance the contributions of the loss terms. The loss is then used to update the material parameters θ_p via gradient descent, enhancing the fidelity of the physical simulation.

3 Experiments

3.1 Data Curation

Our dataset comprises two parts: 1. Two open-source surgical datasets. 2. An in-house dataset. We incorporate sequences from public EndoNeRF [28] and CholecSeg8K [13] following [28,15], while assigning ground-truth material parameters to the simulated scenes. Our in-house dataset PorcineEndo was constructed

Table 2. System identification performance (RE ↓ / EPE↓) on the all dataset.

Method	PhysG. [30]	Physics3D [18]	GIC [3]	PhysFlow [21]	**Ours**
v01_080	0.150/0.059	0.230/0.071	0.096/0.066	3.836/6.368	**0.081**/**0.051**
v01_240	0.617/0.408	0.834/0.336	0.129/0.068	3.290/1.601	**0.035**/**0.016**
pulling	1.672/1.650	0.491/1.014	1.048/1.769	4.651/1.683	**0.260**/**0.166**
cutting	2.041/0.592	0.251/0.141	0.235/0.136	0.196/0.142	**0.146**/**0.077**
gallbladder	0.192/0.617	3.417/7.740	0.380/0.700	0.296/0.609	**0.165**/**0.293**
stomach	0.187/0.612	0.964/1.578	0.199/**0.461**	0.668/0.988	**0.147**/0.501
Avg.	0.810/0.657	1.031/1.803	0.348/0.533	2.157/1.898	**0.139**/**0.184**

Table 3. Physical realism scores by surgeons, rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) per video.

Method	PhysG. [30]	Physics3D [18]	GIC [3]	PhyFlow [21]	**Ours**
Physical Realism ↑	3.63	3.33	3.77	3.26	**4.02**

through an ex-vivo experiment involving fresh porcine stomach and gallbladder tissues. Video sequences were recorded using a monocular endoscope, and uniaxial tensile tests were conducted on tissue specimens using a HANDPI HLD digital force gauge. The Young's modulus was computed via a 30-point linear least-squares regression within the elastic region for each specimens, achieving accuracy with an average $R^2 > 0.99$. Additionally, the Poisson's ratio was set to incompressible in this in-house dataset.

3.2 Implementation Details

All experiments are conducted on the A6000 GPU with the Python PyTorch framework. Reconstruction is performed using 4DGS with parameters from [16], while pre-trained π^3 [27] and Surgical-SAM-2 [27] are used with default settings. The simulation is based on the warp [22] implementation of MLS-MPM. The simulation region is discretized onto a structured grid with a typical resolution of 50^3 to enable efficient particle-grid interactions in the framework. Specifically, we employ 25 sub-steps between successive renderings and include the gradient computed at the last sub-step in the optimization. The sub-step duration is set to 2×10^{-4}, ensuring precision and accuracy in the simulation. On our setup, the whole material estimation process requires approximately 15 minutes per scene.

3.3 Results

We conducted a qualitative and quantitative comparison of our proposed method with state-of-the-art approaches, including PhysGaussian (PhysG.) [30], GIC [3], Physics3D [18], and PhysFlow [21], for surgical scene reconstruction and simulation. Specifically, for a fair comparison under a unified supervision setting, GIC is trained with rendering loss, whereas Physics3D and PhysFlow are trained with direct image supervision instead of generative images. All baselines and

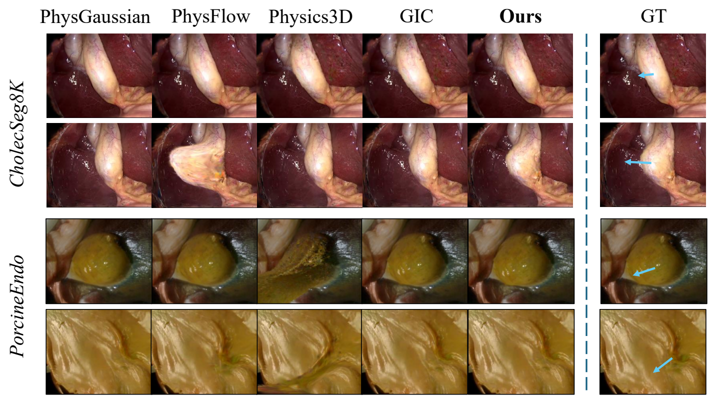

Fig. 2. Qualitative results on the CholecSeg8K and PorcineEndo datasets, shown for representative frames from the simulation sequences.

Table 4. Ablation results (RE ↓ / EPE↓ / FID↓) on the PorcineEndo dataset. The best results are in bold.

Initialization	Manual		Gemini-3-pro [11]		Claude-sonnet-4.5 [2]		GPT-4o [1]	
Optimization	✗	✓	✗	✓	✗	✓	✗	✓
RE ↓	0.309	0.160	0.197	0.064	0.193	0.103	0.144	**0.063**
EPE ↓	0.938	0.514	0.701	0.414	0.679	0.417	0.604	**0.397**
FID ↓	32.28	22.69	22.78	13.43	17.68	13.81	18.27	**13.35**

our method optimize parameters θ_p starting from the same boundary settings per sence, such as including grid resolution, external force.

We evaluate performance using Relative Error (RE) [6] for the estimated material parameters (Young's modulus E, Poisson's ratio ν, shear modulus G, and bulk modulus K; E, G, and K in log-space), End-Point Error (EPE) [9] for optical flow accuracy, and Fréchet Inception Distance (FID) [12] for simulation realism. As shown in Tab. 2, our method consistently achieves lower RE and EPE than Physics3D and PhysFlow in most cases and remains competitive with GIC, demonstrating its effectiveness in estimating material properties and capturing deformation. On the stomach scene, our method is slightly inferior to GIC in EPE. This is mainly due to the uniform color and large deformation in this scenario, which weaken the reliability of local optical flow estimation. Additionally, expert ratings from 33 surgeons indicate that our method obtains the highest realism score (Tab. 3) with the smallest standard deviation, demonstrating superior statistical consistency, highlighting its potential for clinical applications such as haptic feedback training. The qualitative results in Figs. 2 and 3 demonstrate that our approach generates more realistic and stable simulations with superior capture of material deformation than the baselines.

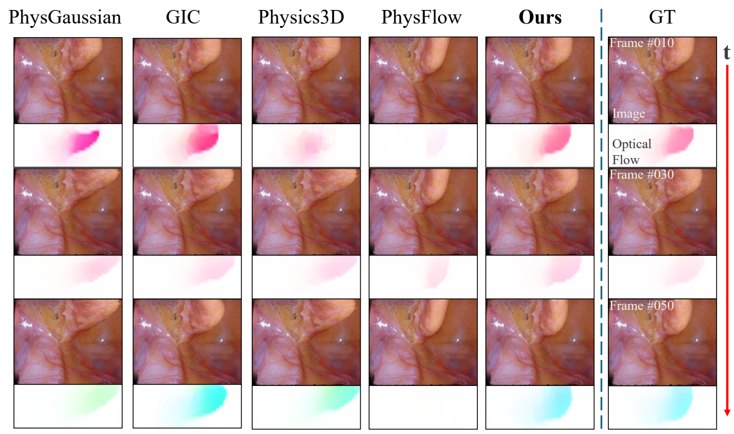

Fig. 3. Qualitative comparison of simulation results from all methods on a sequence of the EndoNeRF dataset, illustrating rendered images and optical flow errors.

Metric	w/o MF	w/ MF
RE ↓	0.433	**0.178**
EPE ↓	0.366	**0.132**
FID ↓	8.81	**4.01**

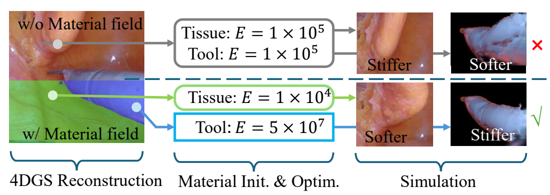

Fig. 4. Ablation on Material Field (MF): quantitative results on EndoNeRF and CholecSeg8K datasets (left), and qualitative comparison with vs. without MF (right).

We perform ablation experiments on material property initialization (manual vs. MLLMs), both with and without subsequent optimization. As shown in Tab. 4, manual initialization yields partial realism and the lowest performance, and subsequent optimization brings moderate improvement. MLLM-inferred properties provide a better starting point without optimization, and MLLM initialization followed by optimization delivers the highest material consistency and visual realism. We further ablate the object-wise material field to evaluate its effectiveness. As shown in Fig. 4, the material field allows object-wise parameter assignment, resulting in more accurate material estimation and physically realistic simulations. The ablation results demonstrate that each component contributes to improved accuracy, physical realism, and overall performance.

4 Conclusion

In this paper, we propose a physics-aware framework for reconstruction and simulation of surgical scenes from endoscopic videos. We first integrate 4DGS with pre-trained depth estimation and segmentation to generate the Gaussian splats representation for the surgical scene. The material field is proposed for material estimation, coarsely initialized via MLLMs-guide estimation and then jointly refined with render and optical flow loss in a differentiable MLS-MPM. The optimized material properties are then incorporated into the simulation pipeline to enable realistic modeling of dynamic surgical scenes. Qualitative and quantitative experiments on open-source and in-house datasets demonstrate that the proposed approach consistently outperforms existing baselines on accuracy and realism, highlighting its potential to advance robotic-assisted minimally invasive surgery with comprehensive perception capabilities.

Acknowledgments. This work was supported by Ministry of Science and Technology (MOST) of China Key Project 2025YFE0122500, Shenzhen-Hong Kong-Macau Technology Research Programme (Type C) STIC Grant SGCX20250526153900001, and Hong Kong RGC CRF C4026-21G, RIF R4020-22, GRF 14211420, 14216020 &14203323.

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

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Anthropic: Introducing claude sonnet 4.5. https://www.anthropic.com/news/claude-sonnet-4-5 (2025)
3. Cai, J., Yang, Y., Yuan, W., He, Y., Dong, Z., Bo, L., Cheng, H., Chen, Q.: Gic: Gaussian-informed continuum for physical property identification and simulation. Advances in Neural Information Processing Systems **37**, 75035–75063 (2024)
4. Chen, E., Chen, L., Zhang, W.: Robotic-assisted colorectal surgery in colorectal cancer management: A narrative review of clinical efficacy and multidisciplinary integration. Frontiers in Oncology **15**, 1502014 (2025)
5. Chen, Y., Xu, H., Zheng, C., Zhuang, B., Pollefeys, M., Geiger, A., Cham, T.J., Cai, J.: Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 370–386. Springer (2024)
6. Dagli, R., Xiang, D., Modi, V., Loop, C., Tsang, C.F., Chen, A.H., Hu, A., State, G., Levin, D.I., Shugrina, M.: Vomp: Predicting volumetric mechanical property fields. In: Proceedings of International Conference on Learning Representations (ICLR) (2026)
7. De Vaucorbeil, A., Nguyen, V.P., Sinaie, S., Wu, J.Y.: Material point method after 25 years: Theory, implementation, and applications. Advances in applied mechanics **53**, 185–398 (2020)
8. Ding, Y., Wang, S., Lan, R., Lin, W., Liu, X., He, W.: Telerobotic surgery: a comprehensive two-decade evolution and the integration of emerging technologies. International Journal of Surgery **112**(1), 1652–1672 (2026)

9. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision (ICCV). pp. 2758–2766 (2015)
10. Gao, B., Zhou, J., Zou, J., Qin, J.: Endord-gs: Robust deformable endoscopic scene reconstruction via gaussian splatting. IEEE Transactions on Medical Imaging **45**(2), 528–541 (2025)
11. Google: A new era of intelligence with gemini 3. https://blog.google/products/gemini/gemini-3 (2025)
12. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in Neural Information Processing Systems **30** (2017)
13. Hong, W.Y., Kao, C.L., Kuo, Y.H., Wang, J.R., Chang, W.L., Shih, C.S.: Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80. arXiv preprint arXiv:2012.12453 (2020)
14. Hu, Y., Fang, Y., Ge, Z., Qu, Z., Zhu, Y., Pradhana, A., Jiang, C.: A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. ACM Transactions on Graphics **37**(4), 1–14 (2018)
15. Huang, Y., Bai, L., Cui, B., Yuan, K., Wang, G., Hoque, M.I., Padoy, N., Navab, N., Ren, H.: Surgtpgs: Semantic 3d surgical scene understanding with text promptable gaussian splatting. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 584–594. Springer (2026)
16. Huang, Y., Cui, B., Bai, L., Guo, Z., Xu, M., Islam, M., Ren, H.: Endo-4dgs: Endoscopic monocular scene reconstruction with 4d gaussian splatting. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 197–207. Springer (2024)
17. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4), 139:1–139:14 (2023)
18. Liu, F., Wang, H., Yao, S., Zhang, S., Zhou, J., Duan, Y.: Physics3d: Learning physical properties of 3d gaussians via video diffusion. arXiv preprint arXiv:2406.04338 (2024)
19. Liu, H., Zhang, E., Wu, J., Hong, M., Jin, Y.: Surgical sam 2: Real-time segment anything in surgical video by efficient frame pruning. arXiv preprint arXiv:2408.07931 (2024)
20. Liu, S., Ren, Z., Gupta, S., Wang, S.: Physgen: Rigid-body physics-grounded image-to-video generation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 360–378. Springer (2024)
21. Liu, Z., Ye, W., Luximon, Y., Wan, P., Zhang, D.: Unleashing the potential of multi-modal foundation models and video diffusion for 4d dynamic physical scene simulation. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR). pp. 11016–11025 (2025)
22. Macklin, M.: Warp: A high-performance python framework for gpu simulation and graphics. In: NVIDIA GPU Technology Conference (GTC). vol. 3 (2022)
23. 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)
24. Raptis, S.P., Theocharopoulos, A., Theocharopoulos, C., Papadakos, S.P., Levantis, G., Kontis, E., Vrahatis, A.G.: Artificial intelligence analysis of minimally invasive surgery data. Journal of Robotic Surgery **20**(1), 186 (2026)

25. Stomakhin, A., Schroeder, C., Chai, L., Teran, J., Selle, A.: A material point method for snow simulation. ACM Transactions on Graphics **32**(4), 1–10 (2013)
26. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 402–419. Springer (2020)
27. Wang, Y., Zhou, J., Zhu, H., Chang, W., Zhou, Y., Li, Z., Chen, J., Pang, J., Shen, C., He, T.: π^3: Permutation-equivariant visual geometry learning. arXiv preprint arXiv:2507.13347 (2025)
28. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 431–441. Springer (2022)
29. 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. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR). pp. 20310–20320 (2024)
30. Xie, T., Zong, Z., Qiu, Y., Li, X., Feng, Y., Yang, Y., Jiang, C.: Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR). pp. 4389–4398 (2024)
31. Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR). pp. 20331–20341 (2024)
32. Zha, R., Cheng, X., Li, H., Harandi, M., Ge, Z.: Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 13–23. Springer (2023)