PhysioSplat: Physics-Informed Dynamic Gaussian Splatting for Surgical Scene Reconstruction

Hritam Basak[1,2] and Zhaozheng Yin[1]

[1] Dept. of Computer Science, Stony Brook University, New York, USA
[2] Amazon.com
{hbasak,zyin}@cs.stonybrook.edu

Abstract. Despite recent progress in 4D Gaussian Splatting (4DGS), dynamic surgical reconstruction remains challenging due to complex tool-tissue interactions, non-rigid deformations, and wet-surface specularities. Existing methods largely rely on unconstrained visual optimization, ignoring the physical reality of the scene reconstruction. We introduce *PhysioSplat*, a novel framework that advances 4DGS by grounding surgical scene modeling in a unified, three-fold physical model. **First**, we incorporate *kinematic physics* by explicitly disentangling the scene into two disjoint fields: a tool field governed by rigid or articulated motion, and a tissue field driven by a deformation model. Supervised by semantic masks, this separation enables selective tool removal and the temporally consistent recovery of occluded anatomy. **Second**, we enforce *biomechanical physics* through a latent, differentiable regularization model. By predicting per-Gaussian stiffness parameters, it minimizes deformation energy and applies collision-aware repulsion to ensure biologically plausible tissue dynamics during tool interactions. **Third**, we integrate *optical physics* via a specular-aware appearance model. Leveraging the camera-coincident light source common in endoscopy, we mathematically decompose Gaussian radiance into diffuse and view-dependent specular components, successfully stabilizing highlight reconstruction. Together, *PhysioSplat* shifts surgical 4DGS from purely visual reconstruction toward anatomy-aware, physics-regularized modeling. Website.

Keywords: Surgical Reconstruction· Endoscopy · Gaussian Splatting.

1 Introduction

Surgical scene reconstruction, particularly endoscopic reconstruction (ER), is essential for advanced computer-assisted interventions, providing the geometric basis for intraoperative navigation, augmented reality overlays, and robotic automation in minimally invasive surgery. Yet recovering dense, dynamic 3D geometry from monocular endoscopic videos is ill-posed due to ER's unique challenges: highly non-rigid tissue deformation, occlusions by surgical tools, and reflections from fluid-covered anatomical surfaces.

Early reconstruction methodologies relied on discrete representations such as point clouds [18] and surfels [3], yet these struggled to maintain topological coherence. While Neural Radiance Fields (NeRF) [15] introduced continuous high-fidelity modeling, exemplified by EndoSurf's [26] SDF constraints and LerPlane's [24] spatiotemporal encoding [4], their reliance on computationally prohibitive volumetric ray sampling [11] precludes the real-time feedback essential for surgery. Consequently, 3D Gaussian Splatting (3DGS)[8] emerged as a superior alternative, leveraging tile-based rasterization to achieve faster rendering speeds than NeRF [5, 10]. Recent extensions to 4DGS [20] capture temporal dynamics by coupling canonical Gaussians with deformation fields parameterized by voxel or hex-plane grids [23]. In the medical domain, Endo-4DGS [7] and Deform3DGS [25] augment this framework by integrating learnable basis functions to mitigate depth ambiguity and compress temporal features [2, 16, 22].

Despite recent progress, current 4DGS frameworks face key limitations for the heterogeneous physics of ER (kinematic, biomechanical, and optical). **First**, they exhibit a *kinematic bias*: modeling a single deformation field that fails to distinguish the rigid, articulated motion of tools from the elastic, continuum-mechanics-driven deformation of soft tissue. This coupling causes geometric shearing at tool–tissue interfaces and hinders reconstruction of anatomy occluded by tools [14]. **Second**, deformation models lack *biomechanical constraints*, often converging to biologically implausible minima such as volume collapse or interpenetration, especially in sparsely textured regions [21]. **Third**, standard Gaussian models ignore the *optical physics* of the wet, highly specular endoscopic environment, where co-located lighting induces unstable *floater* artifacts and geometric noise in textured areas [17].

We propose ***PhysioSplat***, a physics-informed dynamic Gaussian Splatting framework that links video rendering with physical behavior through: **(1) Physio-Semantic Gaussian Disentanglement:** A dual-branch architecture that separates scene primitives into rigid (tool) and deformable (tissue) fields. This explicit kinematic split enables temporally consistent recovery of occluded anatomy by selectively rendering the tissue field, which monolithic 4DGS cannot provide. **(2) Differentiable Biomechanical Regularization:** A biomechanics-informed loss that penalizes non-physical deformations. Using a latent mass–spring tissue model and collision–repulsion constraints, we enforce quasi-static elasticity and volume preservation, reducing geometric artifacts during tool–tissue interaction. **(3) Specular-Aware Appearance Modeling:** A rendering formulation that decomposes radiance into view-independent diffuse and view-dependent specular components. By explicitly modeling anisotropic reflectance under a co-located light source, we stabilize highlight reconstruction in reflective, fluid-filled scenes.

2 Proposed Method

Given an input endoscopic video sequence $\mathcal{V} = \{\mathbf{I}_t, \mathbf{M}_t, \mathbf{D}_t : t \in [0, T]\}$, where for the t-th RGB frame $\mathbf{I}_t$, $\mathbf{M}_t$ represents given semantic tool mask and $\mathbf{D}_t$ is the given monocular depth map, our objective is to synthesize a physics-informed dynamic Gaussian representation. In Sec. 2.1, we briefly outline the necessary preliminaries of dynamic Gaussian Splatting. Sec. 2.2 details our proposed physio-semantic disentanglement strategy, separating rigid tool kinematics

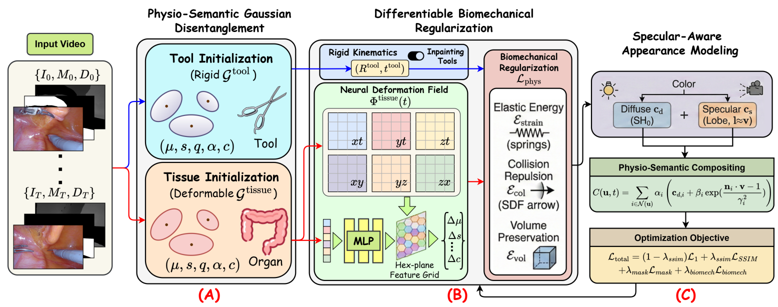

Fig. 1: **(A)** Tool & Tissue Gaussians are disentangled (subsection 2.2), followed by **(B)** deformation modeling through biomechanical regularization (subsection 2.3), **(C)** Gaussians are updated through a Specular-Aware Appearance Modeling (subsection 2.4) using a composite optimization (subsection 2.5).

from soft tissue mechanics. We then formalize the differentiable biomechanical regularization in Sec. 2.3, and present our specular-aware appearance model tailored for co-located endoscopic lighting in Sec. 2.4. An overview of the *PhysioSplat* architecture is illustrated in Figure 1.

2.1 Preliminaries: Dynamic Gaussian Splatting

We represent the surgical scene as a collection of N 3D Gaussians $\mathcal{G} = \{g_i\}_{i=1}^N$. Each primitive is parameterized by a mean position $\boldsymbol{\mu} \in \mathbb{R}^3$, covariance matrix $\boldsymbol{\Sigma}$, opacity $\alpha \in [0, 1]$, and spherical harmonic (SH) coefficients $\mathbf{c}$. To ensure positive semi-definiteness, $\boldsymbol{\Sigma}$ is decomposed as $\boldsymbol{\Sigma} = \mathbf{R}\ \mathbf{S}\ \mathbf{S}^T\mathbf{R}^T$, where $\mathbf{S}$ is a scaling matrix and $\mathbf{R}$ is a rotation matrix, stored as diagonal vector $\mathbf{s}$ and quaternion vector $\mathbf{q}$, respectively. Thus, each Gaussian is represented as $g_i = \{\boldsymbol{\mu}_i, \mathbf{s}_i, \mathbf{q}_i, \alpha_i, \mathbf{c}_i\}$. To model temporal dynamics, following [7], we employ a deformation field $\mathcal{F}_\theta : (g_i, t) \to (\Delta\boldsymbol{\mu}_i, \Delta\mathbf{s}_i, \Delta\mathbf{q}_i, \Delta\alpha_i, \Delta\mathbf{c}_i)$ that maps canonical Gaussians to their deformed state at time t. The rendered color $C(\mathbf{u})$ for a pixel $\mathbf{u}$ is obtained via differentiable α-blending of N ordered primitives:

$$C(\mathbf{u}) = \sum_{i=1}^{N} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \alpha_i = \exp\left(-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu}_i^{2D}(t))^T(\boldsymbol{\Sigma}_i^{2D}(t))^{-1}(\mathbf{u} - \boldsymbol{\mu}_i^{2D}(t))\right)$$

where $\boldsymbol{\mu}_i^{2D}(t)$ and $\boldsymbol{\Sigma}_i^{2D}(t)$ are the 2D mean and covariance matrix obtained after projecting the 3D Gaussian g_i onto the camera image plane. While efficient, this standard formulation lacks physical grounding. Our method injects kinematic, biomechanical, and optical constraints into rendering.

2.2 Physio-Semantic Gaussian Disentanglement

A single deformation field $\mathcal{F}_\theta$ cannot simultaneously capture the high-frequency articulation of tools and the low-frequency elasticity of tissues without introducing geometric shearing artifacts. To resolve this, we partition the global Gaussian set $\mathcal{G}$ into two disjoint semantic subsets: the *Tool Field* $\mathcal{G}^{\text{tool}}$ and the *Tissue Field* $\mathcal{G}^{\text{tissue}}$, as seen in Figure 1(A). While the ultimate objective is to synthesize tool-free operative views, explicitly reconstructing the tool field $\mathcal{G}^{\text{tool}}$ is imperative as it acts as a necessary volumetric occluder that absorbs photometric gradients, preventing the hallucination of geometric "floaters" in masked regions and provides the requisite 3D boundary for enforcing our biomechanical collision constraints, described in subsection 2.3.

We assign a learnable semantic logit $s_i \in \mathbb{R}$ to each Gaussian, which is optimized alongside geometric parameters using the mask prior $\mathbf{M}_t$. We define the probability of primitive i belonging to the tissue class as $p_i = \sigma(s_i)$, where $\sigma(\cdot)$ is the sigmoid function. The dynamics of the scene are then bifurcated into two distinct motion models using a threshold τ:

$$\boldsymbol{\mu}_i(t) = \begin{cases} \mathbf{R}^{\text{tool}}(t)\boldsymbol{\mu}_i(t-1) + \mathbf{T}^{\text{tool}}(t) & \text{if } p_i < \tau \quad \text{(Rigid)} \\ \boldsymbol{\mu}_i(t-1) + \Delta\boldsymbol{\mu}_i(t) & \text{if } p_i \geq \tau \quad \text{(Deformable)} \end{cases} \tag{1}$$

Here, $\mathbf{R}^{\text{tool}}(t) \in SO(3)$ and $\mathbf{T}^{\text{tool}}(t) \in \mathbb{R}^3$ represent the rotation matrix and translation vector at time t, respectively. For the deformable tissue field ($p_i \geq \tau$), $\Delta\boldsymbol{\mu}_i(t)$ is the positional shift derived from our non-rigid neural deformation field Φ^{tissue}, modeled by a multi-resolution hex-plane encoder. This separation allows us to enforce rigid constraints on the tool while permitting elasticity in the tissue. $R_{\text{tool}}(t)$ and $T_{\text{tool}}(t)$ are jointly optimized via per-frame differentiable rigid registration without external tracking. Notably, $\tau = 0.5$ is empirically robust (PSNR varies <0.3 dB for $\tau \in \{0.3, 0.5, 0.7\}$), as final field assignment is convergence-driven. $R_{\text{tool}}(t)$ and $T_{\text{tool}}(t)$ are jointly optimized via per-frame differentiable rigid registration without external tracking. While Eq. 1 applies a single rigid transform per frame, empirically sufficient for standard laparoscopic instruments, extending to per-segment kinematic trees for articulated tools is future work.

Inpainting via Semantic Visibility A critical challenge in endoscopic reconstruction is the "blind spot" created by surgical tools. To reconstruct occluded tissue, we modify the densification gradients. During training, if a tissue Gaussian $g_i \in \mathcal{G}^{\text{tissue}}$ is occluded by a tool Gaussian $g_j \in \mathcal{G}^{\text{tool}}$ (i.e., depth $z_j < z_i$), we re-weight the gradient flow to g_i using a temporal consistency score $\omega_i = \text{sim}(\hat{\mathbf{c}}_i, \mathbf{c}_i^{\text{adj}})$, computed as the photometric similarity between the occluded Gaussian's predicted appearance and its nearest unoccluded-frame projection. This prevents premature pruning and allows $\mathcal{G}^{\text{tissue}}$ to learn the underlying anatomy even when momentarily hidden. This allows $\mathcal{G}^{\text{tissue}}$ to learn the underlying anatomy even when momentarily hidden. At inference, we can perform *Physio-Inpainting* by selectively rendering only the tissue field:

$$C^{\text{tissue}}(\mathbf{u}) = \sum_{g_i \in \mathcal{G}^{\text{tissue}}} \mathbb{I}(p_i \geq \tau) \cdot \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \mathbb{I}(p_j \geq \tau)\alpha_j) \tag{2}$$

where $\mathbb{I}(\cdot)$ is the indicator function. This effectively "erases" the tool.

2.3 Differentiable Biomechanical Regularization

The deformation field Φ^{tissue} is mathematically ill-posed (infinitely many vector fields can satisfy the photometric loss). To restrict the solution space to biologically plausible deformations, we introduce a *Differentiable Biomechanical Regularization*, shown in Figure 1(B). We approximate the tissue as a discrete mass-spring system in the canonical space and enforce continuum mechanics constraints via a biomechanical energy functional $\mathcal{L}_{\text{biomech}}$. The loss at time t is:

$$\mathcal{L}_{\text{biomech}} = \lambda_{\text{strain}}\mathcal{E}_{\text{strain}} + \lambda_{\text{col}}\mathcal{E}_{\text{col}} + \lambda_{\text{vol}}\mathcal{E}_{\text{vol}} \tag{3}$$

where $\mathcal{E}_{\text{strain}}$, $\mathcal{E}_{\text{col}}$, and $\mathcal{E}_{\text{vol}}$ are described below, balanced by their respective hyperparameters λ_{strain}, λ_{col}, and λ_{vol}.

Local Elastic Strain Energy: To enforce quasi-static elasticity and enforce local rigidity, we utilize an As-Rigid-As-Possible (ARAP) formulation. For each tissue Gaussian g_i, we construct a local neighborhood $\mathcal{N}(i)$ based on Euclidean distance in the canonical frame. Strain energy penalizes local distance deviations:

$$\mathcal{E}_{\text{strain}} = \sum_{t=1}^{T} \sum_{g_i \in \mathcal{G}^{\text{tissue}}} \sum_{j \in \mathcal{N}(i)} w_{ij} \left(|\boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_j(t)|_2 - |\boldsymbol{\mu}_i(0) - \boldsymbol{\mu}_j(0)|_2 \right)^2 \tag{4}$$

where $w_{ij} = \exp(-|\boldsymbol{\mu}_i(0) - \boldsymbol{\mu}_j(0)|^2/\zeta^2)$ are stiffness weights derived from canonical proximity, and ζ controls the spatial falloff of the neighborhood influence. This preserves local topology and prevents non-physical stretching.

Tool-Tissue Collision Repulsion: Kinematic-only 4DGS often results in volume interpenetration, where the tool visually passes through the tissue. We enforce a hard physical constraint using a collision potential. We approximate the tool volume at time t as a proxy Signed Distance Field (SDF), Ψ^{tool}, computed from the convex hull of $\mathcal{G}^{\text{tool}}$. The collision energy is:

$$\mathcal{E}_{\text{col}} = \sum_{t=1}^{T} \sum_{g_i \in \mathcal{G}^{\text{tissue}}} \text{ReLU}\left(\epsilon - \Psi^{\text{tool}}(\boldsymbol{\mu}_i(t), t)\right)^2 \tag{5}$$

where ϵ is a safety margin (contact thickness). This term exerts a repulsive force on tissue Gaussians that violate the tool's volume, effectively "pushing" the tissue surface to deform realistically around the surgical tool.

Volume Preservation: Finally, to preclude biologically implausible tissue compression or expansion, we enforce quasi-incompressibility by penalizing deviations of the local deformation gradient's determinant from unity:

$$\mathcal{E}_{\text{vol}} = \sum_{t=1}^{T} \sum_{g_i \in \mathcal{G}^{\text{tissue}}} (\det(\nabla\Phi^{\text{tissue}}(\boldsymbol{\mu}_i(t), t)) - 1)^2 \tag{6}$$

2.4 Specular-Aware Appearance Modeling

The standard Spherical Harmonic (SH) model in 3DGS assumes a static lighting environment. This assumption is violated in endoscopy, where the light source moves dynamically with the camera. This leads to severe "floater" artifacts in wet, specular regions as the optimization attempts to model moving highlights as shifting geometry. We propose a **Co-located Specular Decomposition** (as seen in Figure 1(C)) that modifies the standard rendering equation. For each Gaussian g_i, we decompose the total emitted radiance $\mathbf{c}_i$ into a view-independent diffuse component $\mathbf{c}_{d,i}$ and a view-dependent specular component $\mathbf{c}_{s,i}$. The diffuse component $\mathbf{c}_{d,i}$ is represented purely by the degree-0 SH coefficient (SH_0). We model the specular component $\mathbf{c}_{s,i}$ using a simplified Gaussian lobe. Under the co-located camera-light assumption, the illumination half-vector perfectly aligns with the view direction vector $\mathbf{v}$. Rendered color $C(\mathbf{u}, t)$ is evaluated as:

$$C(\mathbf{u}, t) = \sum_{i \in \mathcal{N}(\mathbf{u})} \alpha_i \left(\underbrace{\mathbf{c}_{d,i}}_{\text{Diffuse } (\mathrm{SH}_0)} + \underbrace{\beta_i \exp\left(\frac{\mathbf{n}_i \cdot \mathbf{v} - 1}{\gamma_i^2} \right)}_{\text{Specular Lobe } \mathbf{c}_{s,i}} \right) \tag{7}$$

The specular term $\mathbf{c}_{s,i}$ is parameterized by a learnable albedo β_i and roughness γ_i. Crucially, the normal $\mathbf{n}_i$ is not a free parameter but derived on-the-fly from the rendered depth gradient ($\mathbf{n}i \approx abla\mathbf{D}$rend), anchoring specular effects to tissue geometry and stabilizing optimization in wet regions.

2.5 Optimization Objective

The complete training objective $\mathcal{L}_{\text{total}}$ is a weighted combination of photometric reconstruction, semantic supervision, and biomechanical regularization:

$$\mathcal{L}_{\text{total}} = (1 - \lambda_{\text{ssim}})\mathcal{L}_1 + \lambda_{\text{ssim}}\mathcal{L}_{\text{SSIM}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{biomech}}\mathcal{L}_{\text{biomech}} \tag{8}$$

where $\mathcal{L}_1$ and $\mathcal{L}_{\text{SSIM}}$ are the standard L_1 and Structural Similarity Index Measure losses. Crucially, these photometric losses are computed with a pixel-wise mask $\mathbf{M}_t$ that excludes tool-occluded regions, i.e., $\mathcal{L}_1 = \|(\hat{C}(\mathbf{u}, t) - I_t) \odot (1 - \mathbf{M}_t)\|_1$, thereby jointly optimizing the underlying geometry, diffuse colors, and specular parameters without requiring tool-free ground-truth frames, where $\mathcal{L}_{\text{mask}}$ is the pixel-wise binary cross-entropy loss between the ground-truth tool masks $\mathbf{M}_t$ and the 2D semantic probability maps, which are obtained by rendering the 3D semantic logits s_i to the image plane via differentiable α-blending.

3 Experiments and Results

Datasets and Evaluation Metrics: We evaluate our method on three dynamic endoscopic benchmarks: **(1)** EndoNeRF dataset [19] consisting of two prostatectomy procedures accompanied by stereo-derived depth maps. **(2)** StereoMIS dataset [6] comprising 11 minimally invasive surgical sequences acquired using the da Vinci Xi robotic platform on in-vivo porcine models. Following [2], we select two representative segments *P2_1* and *P2_2* sequences for evaluation. **(3)** SCARED dataset [1], which provides diverse endoscopic views of a

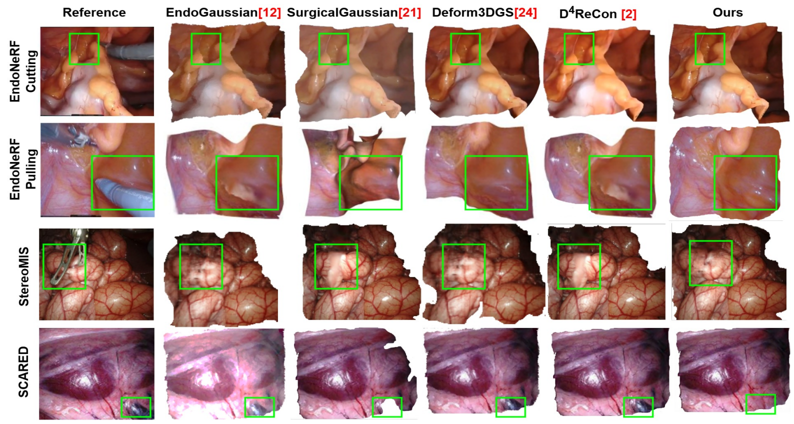

Fig. 2: Qualitative comparison of PhysioSplat. **Row 1,4**: Disentanglement removes kinematic shearing. **Row 2**: Biomechanical regularization prevents volume collapse. **Row 3**: Specular-aware modeling suppresses wet-region artifacts.

porcine abdomen with ground-truth geometry acquired via structured light. Our performance is evaluated in terms of photorealism (PSNR & SSIM) and geometric consistency (LPIPS). These metrics are calculated by comparing predicted color and ground-truth color, excluding surgical tools denoted by masks. Depth maps are obtained via stereo-matching (EndoNeRF, StereoMIS) and structured light (SCARED). Tool masks are generated using SAM [9], following [2, 7]. Depth is used for initializing Gaussian positions and deriving surface normals via $\mathbf{n}_i \approx \nabla \mathbf{D}_{\text{rend}}$.

Implementation Details: We implement *PhysioSplat* in PyTorch on an NVIDIA RTX 4090 (24GB). Training uses Adam for 30K iterations with learning rate $\eta = 1.6 \times 10^{-3}$ and cosine decay. The deformation field Φ^{tissue} uses a multi-resolution hex-plane encoder ($D = 64$) with threshold $\tau = 0.5$. Biomechanical weights are $\lambda_{\text{strain}} = 0.1$, $\lambda_{\text{col}} = 1.0$, $\lambda_{\text{vol}} = 0.05$ (grid search), with collision margin $\epsilon = 2$mm. ζ is computed as the mean distance to $K = 16$ nearest neighbors in canonical space. Specular albedo β_i is initialized to 0.1. Total objective weights: $\lambda_{\text{ssim}} = 0.2$, $\lambda_{\text{mask}} = 0.1$, $\lambda_{\text{biomech}} = 0.01$. Standard 3DGS splitting/pruning is applied, with densification disabled where $\Psi^{\text{tool}} < \epsilon$ to prevent artifacts.

3.1 Comparison with State of the Art

Table 1 summarizes *PhysioSplat*'s performance against SoTA methods across three benchmarks. NeRF-based methods like EndoSurf [26] maintain structural consistency via SDF priors but suffer from high latency and motion blur. Early 4DGS adaptations such as LGS [12] achieve real-time rendering but yield low PSNR from oversimplified deformation fields. Endo-4DGS [7] and EndoGaussian [13] improve depth-aware fidelity but fail to separate rigid tool motion from elastic tissue dynamics. Deform3DGS [25] and EH-SurGS [16] enhance photometric

Table 1: Quantitative comparison of our proposed method, *PhysioSplat* on EndoNeRF [19], StereoMIS [6], and SACRED [1] datasets. The previous best and second-best results are marked in **blue** and **green**.

Method	Category	EndoNeRF-Cutting			EndoNeRF-Pulling			StereoMIS			SACRED			Average	
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	FPS↑	Time(s)↓
EndoSurf [26]	NeRF	34.98	0.953	0.106	35.00	0.956	0.120	30.78	0.856	0.294	25.02	0.802	0.356	0.05	2.5e4
LGS [12]	4DGS	36.21	0.937	0.088	35.89	0.930	0.089	24.47	0.831	0.301	27.05	0.826	0.297	189	134
Endo-4DGS [7]	4DGS	36.56	0.955	**0.032**	37.85	0.959	**0.043**	33.85	0.894	0.165	**26.56**	0.811	0.275	100	330
EndoGaussian [13]	4DGS	38.29	0.962	0.058	37.31	0.958	0.070	34.37	0.899	**0.158**	**26.89**	0.825	**0.272**	191	125
SurgicalGaussian [22]	3DGS	37.51	0.961	0.062	38.78	**0.970**	0.049	30.09	0.845	0.309	22.81	0.802	0.371	86	174
Deform3DGS [25]	3DGS	37.86	0.958	0.059	37.94	0.959	0.061	34.71	0.904	0.163	24.49	**0.830**	0.307	333	**75**
EH-SurGS [16]	3DGS	**39.91**	**0.972**	0.034	**38.72**	0.963	0.062	**34.91**	**0.906**	0.166	23.33	**0.829**	0.311	**374**	**110**
D^{4}Recon [2]	Dyn. GS	**40.13**	**0.978**	**0.029**	**39.98**	**0.986**	**0.049**	**35.03**	**0.910**	**0.155**	26.32	0.819	**0.273**	**335**	120
Ours	**Dyn. GS**	**42.80**	**0.991**	**0.024**	**41.60**	**0.990**	**0.045**	**37.20**	**0.934**	**0.146**	**29.75**	**0.868**	**0.231**	**193**	**145**

Table 2: Ablation results of *PhysioSplat* on surgical reconstruction datasets.

Exp	4DGS Baseline	PSGD (Sec.2.2)	DBR (Sec.2.3)	SAAM (Sec.2.4)	EndoNeRF-Cutting			EndoNeRF-Pulling			StereoMIS			SACRED		
					PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
(a)	✓	✗	✗	✗	36.2	0.954	0.061	36.7	0.954	0.055	33.5	0.836	0.169	25.10	0.814	0.283
(b)	✗	✓	✗	✗	38.7	0.974	0.035	38.3	0.974	0.049	35.2	0.877	0.162	26.90	0.835	0.269
(c)	✗	✓	✓	✗	40.4	0.983	0.029	40.2	0.988	0.041	36.9	0.921	0.154	27.31	0.842	0.257
(d)	✗	✓	✓	✓	**42.8**	**0.991**	**0.024**	**41.6**	**0.990**	**0.045**	**37.2**	**0.934**	**0.146**	**29.75**	**0.868**	**0.231**

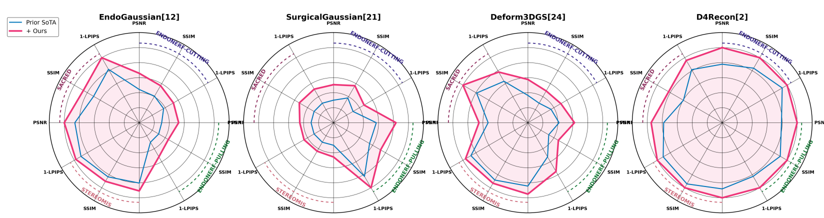

Fig. 3: Radar charts showing the quantitative improvements achieved by integrating our proposed DBR and SAAM modules into recent SoTA methods.

quality but lack physical constraints, while D^{4}Recon [2] offers strong deformation modeling yet exhibits geometric shearing at tool-tissue interfaces. *PhysioSplat* sets a new SoTA, outperforming D^{4}Recon across all metrics with superior geometric accuracy and robustness (Figure 2), while maintaining competitive speed. Furthermore, surface-level geometric evaluation on the SCARED dataset demonstrates that *PhysioSplat* achieves a Chamfer Distance (CD) of 1.83 mm and an F-score@1mm of 0.71. Compared to D^{4}Recon's CD of 2.41 mm and F-score of 0.58, this represents a 24% improvement in geometric reconstruction accuracy. Additional results are in **supplementary material**.

3.2 Ablation Experiments

Our ablation (Table 2) evaluates each module. **(a) Baseline** (monolithic 4DGS [20]) performs worst due to kinematic conflict between rigid tools and elastic tissue. **(b)** Adding PSGD yields +2.5,dB PSNR on Cutting by separating tool/tissue optimization and eliminating boundary shearing. **(c)** DBR further improves geometric fidelity, preventing volume collapse in texture-less regions. **(d)** The full model with SAAM achieves the best perceptual quality (lowest LPIPS), removing floater artifacts in wet environments. Additional results are in the **supplementary material**.

To assess generalizability, Figure 3 shows gains from integrating DBR and SAAM as modular plugins across four baselines [2, 13, 22, 25]. The consistent performance expansion (pink vs. blue) confirms robust, architecture-agnostic improvements in LPIPS and PSNR. Additionally, a mask robustness experiment (±10,px boundary perturbation) yields only 0.2,dB PSNR drop (42.8→42.6), confirming stability under imperfect segmentation.

4 Conclusion

We introduce *PhysioSplat*, a physics-informed dynamic GS framework for high-fidelity surgical scene reconstruction. By disentangling tool-tissue kinematics, enforcing biomechanical regularization, and modeling specular appearance, our method consistently outperforms the state-of-the-art in geometric and photometric accuracy across diverse benchmarks. Future work will explore real-time robotic integration and advanced biomechanical priors for clinical deployment.

Disclosure of Interests. The authors were supported by NSF grants IIS-2331769 and CMMI-2246673.

References

[1] Max Allan et al. "Stereo correspondence and reconstruction of endoscopic data challenge". In: *arXiv preprint arXiv:2101.01133* (2021).

[2] Hritam Basak and Zhaozheng Yin. "D4Recon: Dual-Stage Deformation and Dual-Scale Depth Guidance for Endoscopic Reconstruction". In: *MICCAI*. 2025, pp. 159–169.

[3] Jiadi Cui et al. "Neural Surfel Reconstruction: Addressing Loop Closure Challenges in Large-Scale 3D Neural Scene Mapping". In: *Sensors (Basel, Switzerland)* 24.21 (2024), p. 6919.

[4] Sara Fridovich-Keil et al. "K-planes: Explicit radiance fields in space, time, and appearance". In: *CVPR*. 2023, pp. 12479–12488.

[5] Jiaxin Guo et al. "Free-surgs: Sfm-free 3d gaussian splatting for surgical scene reconstruction". In: *MICCAI*. 2024, pp. 350–360.

[6] Michel Hayoz et al. "Learning how to robustly estimate camera pose in endoscopic videos". In: *International journal of computer assisted radiology and surgery* 18.7 (2023), pp. 1185–1192.

[7] Yiming Huang et al. "Endo-4dgs: Endoscopic monocular scene reconstruction with 4d gaussian splatting". In: *MICCAI*. 2024, pp. 197–207.

[8] Bernhard Kerbl et al. "3D Gaussian splatting for real-time radiance field rendering." In: *ACM Trans. Graph.* 42.4 (2023), pp. 139–1.

[9] Alexander Kirillov et al. "Segment anything". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 4015–4026.

[10] Chenxin Li et al. “Endosparse: Real-time sparse view synthesis of endoscopic scenes using gaussian splatting”. In: *MICCAI*. 2024, pp. 252–262.

[11] Ruilong Li et al. “Nerfacc: Efficient sampling accelerates nerfs”. In: *ICCV*. 2023, pp. 18537–18546.

[12] Hengyu Liu et al. “Lgs: A light-weight 4d gaussian splatting for efficient surgical scene reconstruction”. In: *MICCAI*. 2024, pp. 660–670.

[13] Yifan Liu et al. “Endogaussian: Gaussian splatting for deformable surgical scene reconstruction”. In: *arXiv preprint arXiv:2401.12561* (2024).

[14] Huoling Luo et al. “Digs: diffusion-guided Gaussian Splatting for dynamic occlusion surgical scene reconstruction”. In: *International Journal of Computer Assisted Radiology and Surgery* (2026), pp. 1–12.

[15] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[16] Jiwei Shan et al. “Deformable Gaussian splatting for efficient and high-fidelity reconstruction of surgical scenes”. In: *ICRA*. 2025, pp. 10545–10551.

[17] Idris O Sunmola et al. “Surgical Gaussian Surfels: Highly Accurate Real-time Surgical Scene Rendering”. In: *arXiv preprint arXiv:2503.04079* (2025).

[18] Jiawei Tian et al. “Three-Dimensional point cloud reconstruction method of cardiac soft tissue based on binocular endoscopic images”. In: *Electronics* 12.18 (2023), p. 3799.

[19] Yuehao Wang et al. “Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2022, pp. 431–441.

[20] Guanjun Wu et al. “4d gaussian splatting for real-time dynamic scene rendering”. In: *CVPR*. 2024, pp. 20310–20320.

[21] Weixing Xie et al. “Endo-HDR: Dynamic Endoscopic Reconstruction with Deformable 3D Gaussians and Hierarchical Depth Regularization”. In: *Knowledge-Based Systems* (2025), p. 114914.

[22] Weixing Xie et al. “Surgicalgaussian: Deformable 3d gaussians for high-fidelity surgical scene reconstruction”. In: *MICCAI*. 2024, pp. 617–627.

[23] Chen Yang et al. “Efficient deformable tissue reconstruction via orthogonal neural plane”. In: *IEEE Transactions on Medical Imaging* 43.9 (2024), pp. 3211–3223.

[24] Chen Yang et al. “Neural lerplane representations for fast 4d reconstruction of deformable tissues”. In: *MICCAI*. 2023, pp. 46–56.

[25] Shuojue Yang et al. “Deform3dgs: Flexible deformation for fast surgical scene reconstruction with gaussian splatting”. In: *MICCAI*. 2024, pp. 132–142.

[26] Ruyi Zha et al. “Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos”. In: *MICCAI*. 2023, pp. 13–23.