MICCAI 2025

PAPER HIGHLIGHT

D4Recon: Dual-stage Deformation and Dual-scale Depth Guidance for Endoscopic Reconstruction

Hritam Basak Zhaozheng Yin

Dept. of Computer Science, Stony Brook University, USA

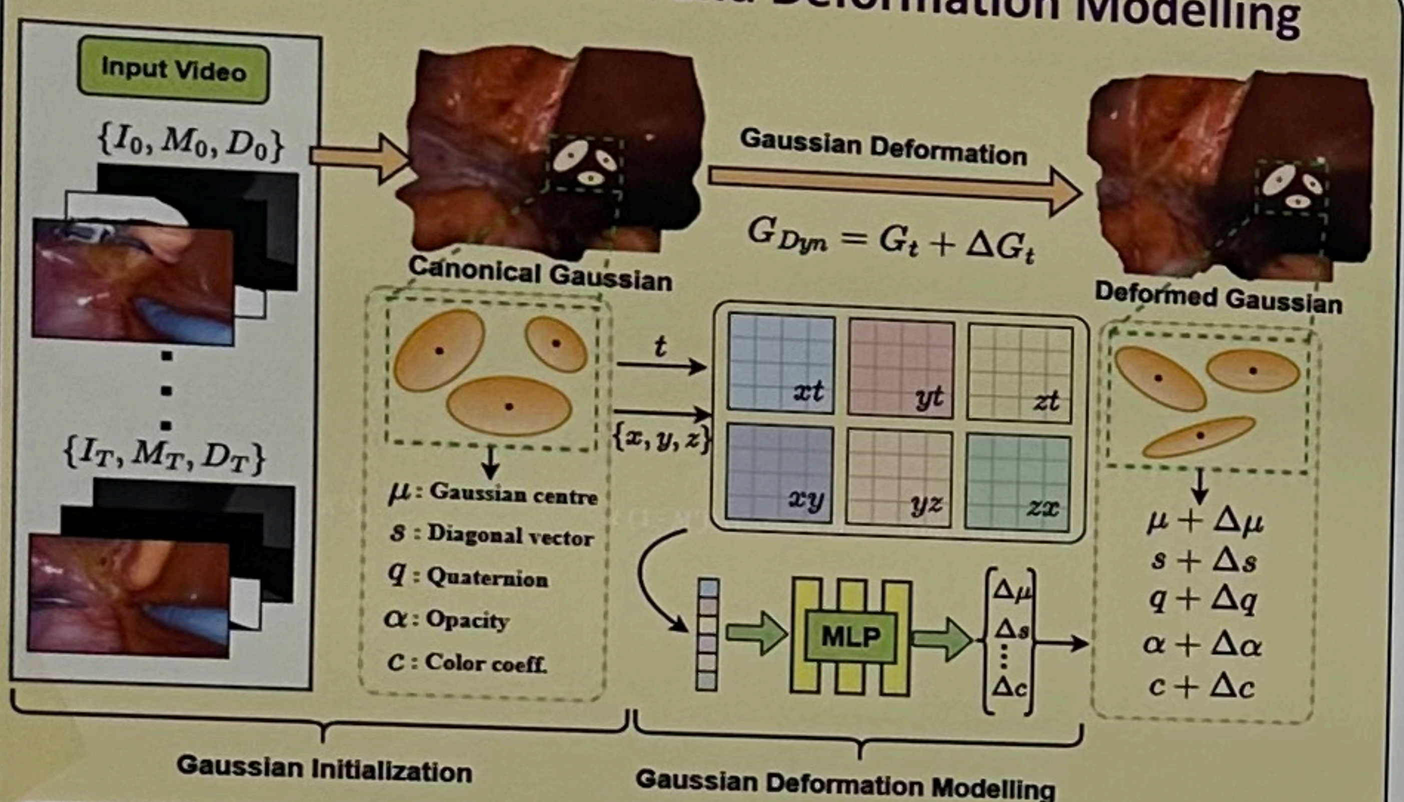
Stony Brook University

Our Proposal and Contribution

- We propose a new **Dual-stage Deformation Modeling** to address **multi-view inconsistency** and **dynamic tissue deformation** via **dual Score Distillation Sampling (SDS)** loss.
- A novel **Dual-scale Hard and Soft Depth Guidance** framework is introduced for **global geometric consistency** and **fine-grained structural details**, respectively.
- We propose **Dynamic3DGS**, by extending 3DGS into a **dynamic, per-frame representation** that incrementally adapts to temporal variations, by **robust modeling of deformable surgical scenes**.
- Demonstrates **state-of-the-art** results on multiple **dynamic** and **static endoscopic** datasets.

Proposed Method

1. Gaussian Initialization and Deformation Modelling

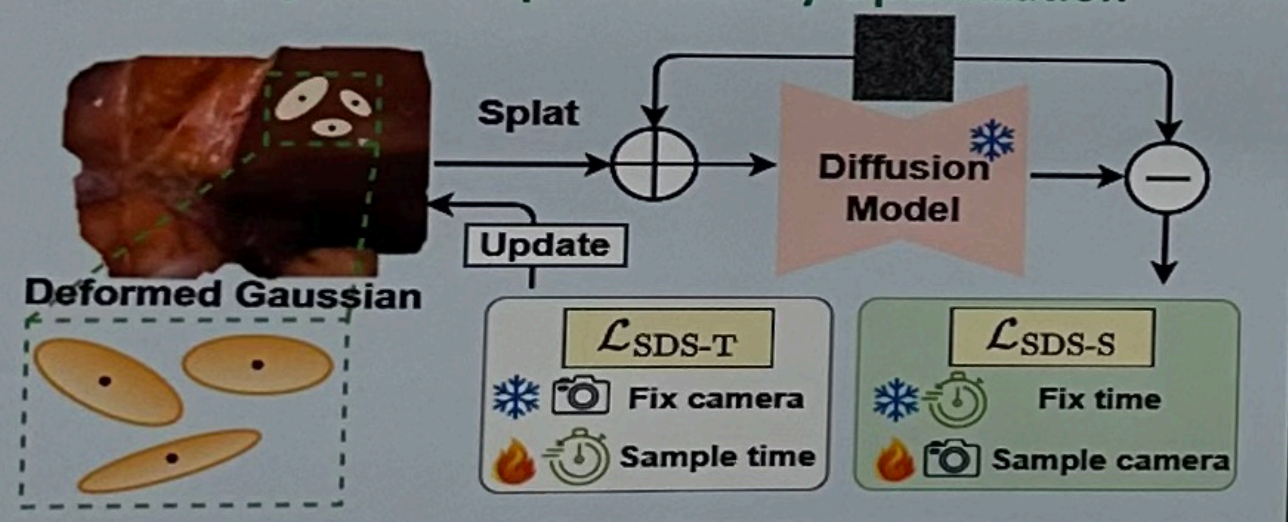


Gaussian Representation $C = \sum_i \left(\alpha_i' \prod_{j=1}^{i-1} (1 - \alpha_j') \right) c_i, \quad D = \sum_i \left(\alpha_i' \prod_{j=1}^{i-1} (1 - \alpha_j') \right) d_i$

Refined PC computation $P' = \{D' K_e^{-1} K_i^{-1} (I' \odot (1 - M'))\}, \quad M' = \cap_{\tau=0}^{t} M_\tau$

2. Optimization

(A) Spatio-Temporal Fidelity Optimization



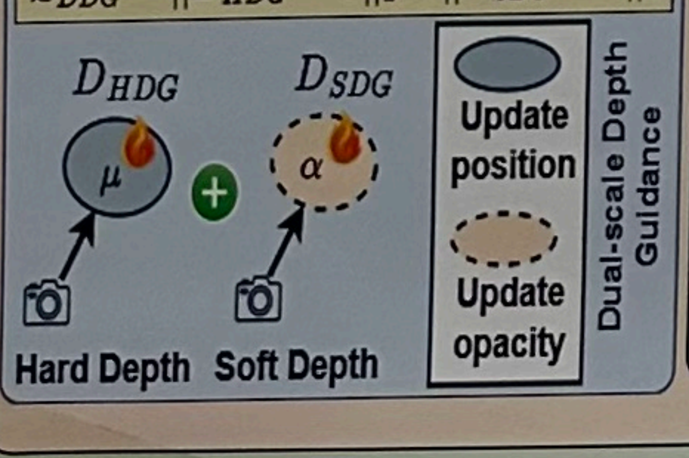
Dynamic Gaussians
$$G_{Dyn} = \{\mu + \psi_\mu(f_t), s + \psi_s(f_t), q + \psi_q(f_t), \alpha + \psi_\alpha(f_t), c + \psi_c(f_t)\} = \{\mu + \Delta\mu, s + \Delta s, q + \Delta q, \alpha + \Delta\alpha, c + \Delta c\}$$

Spatial Deformation Modeling
$$\nabla_{\theta_s} \mathcal{L}_{\text{SDS-S}} = E_{\widehat{P_i}, \epsilon, \sigma} \left[w(\sigma)(\epsilon_\phi - \epsilon) \frac{\partial \mathcal{I}_{\widehat{\mathcal{P}_i}}}{\partial \theta_s} \right]; \quad \epsilon_\phi = \text{Diffusion}\left(\mathcal{I}^{\epsilon}_{\widehat{\mathcal{P}_i}}, \widehat{P_i}, \sigma \right))$$

Temporal Deformation Modeling
$$\nabla_{\theta_T} \mathcal{L}_{\text{SDS-T}} = E_{\widehat{P_i}, \epsilon, \sigma} \left[w(\sigma)(\epsilon_\phi - \epsilon) \frac{\partial \mathcal{I}_{\widehat{\mathcal{P}_i}}}{\partial \theta_T} \right]; \quad \epsilon_\phi = \text{Diffusion}\left(\mathcal{I}^{\epsilon}_{\widehat{\mathcal{P}_i}}, \widehat{P_i}, \sigma \right))$$

(B) Dual-scale Depth Guidance

$$\mathcal{L}_{DDG} = ||D_{HDG} - D||_2 + ||D_{SDG} - D||_2$$



Hard Depth Guidance
$$D_{HDG}(p) = \sum_i \beta (1 - \beta)^{i-1} |\rho - \mu_i|_2$$

Soft Depth Guidance
$$D_{SDG}(p) = \sum_i \alpha_i' \prod_{j=1}^{i-1} (1 - \alpha_j') |\rho - \mu_i|_2$$

Dataset Description

- **Dynamic Surgical Scene Datasets**: **StereoMIS** includes 11 surgical sequences recorded on in-vivo porcine subjects using the da Vinci Xi system, with challenges such as tissue deformation and tool occlusions. Following prior work, two segments (P2_1 and P2_2) are used. **EndoNeRF** consists of stereo-matched prostatectomy sequences that capture complex non-rigid deformations and tool-tissue interactions. We adopt 7:1 training-validation splits and utilize monocular depth maps as priors for both datasets.
- **Colonoscopy Datasets**: **Simulation** sequences are Unity-rendered colonoscopy videos with depth and pose estimated via RNNSLAM. The **In-Vivo** dataset contains real colonoscopy recordings at 270×216 resolution, while **Phantom** is derived from the C3VD collection, featuring high-resolution colonoscopy sequences such as "cecum_t4_b," "desc_t4_a," and "transt_t1_a."

Experimental Results

(I) Comparison with SoTA

Table 1: Quantitative evaluation of D⁴Recon on EndoNeRF and StereoMIS datasets. The best & second-best performances are highlighted in red & blue.

Method	Category	EndoNeRF-Cutting PSNR↑	SSIM↑	LPIPS↓	EndoNeRF-Pulling PSNR↑	SSIM↑	LPIPS↓	Average FPS↑	Time(s)↓	StereoMIS PSNR↑	SSIM↑	LPIPS↓	Average FPS↑	Time(s)↓
LerPlane-32K [28]	NeuralPlane	34.66	0.923	0.071	31.77	0.910	0.071	100	240	24.12	0.814	0.327	100	255
EndoSurf [31]	NeRF	34.98	0.953	0.106	35.00	0.956	0.120	0.04	2.5e4	30.78	0.856	0.294	0.05	2.5e4
LGS [14]	4DGS	36.21	0.937	0.088	35.89	0.930	0.089	188	122	24.47	0.831	0.301	190	145
Endo-4DGS [10]	4DGS	36.56	0.955	0.032	37.85	0.959	0.043	100	240	33.85	0.894	0.165	100	420
EndoGaussian [15]	4DGS	38.29	0.962	0.058	37.31	0.958	0.070	193	120	34.37	0.899	0.158	190	130
SurgicalGaussian [26]	3DGS	37.51	0.961	0.062	38.78	0.970	0.049	82	165	30.09	0.845	0.309	86	182
Deform3DGS [30]	3DGS	37.86	0.958	0.059	37.94	0.959	0.061	335	71	34.71	0.904	0.163	332	79
EH-SurGS [22]	3DGS	39.91	0.972	0.034	38.72	0.963	0.062	383	101	34.91	0.906	0.166	365	120
Ours	Dynamic3DGS	40.13	0.978	0.029	39.98	0.986	0.049	336	122	35.03	0.910	0.155	335	120

Table 2: Quantitative evaluation of D⁴Recon on three static datasets.

Method	Category	Simulation PSNR ↑	SSIM ↑	LPIPS ↓	In-Vivo PSNR ↑	SSIM ↑	LPIPS ↓	Phantom PSNR ↑	SSIM ↑	LPIPS ↓
NeRF [18]	NeRF	35.29	0.92	0.14	18.93	0.67	0.43	32.10	0.81	0.39
REIM-NeRF [20]	NeRF	32.22	0.82	0.33	18.94	0.65	0.45	31.66	0.78	0.22
Nice-SLAM [34]	SLAM	35.61	0.84	0.31	20.37	0.77	0.32	28.08	0.88	0.29
Endo-Depth [21]	DepthCNN	38.88	0.93	0.13	23.51	0.79	0.25	30.18	0.86	0.26
Endo2DTAM [9]	SLAM+3DGS	35.62	0.85	0.22	23.19	0.76	0.28	29.93	0.81	0.28
EndoGSLAM [23]	SLAM+3DGS	39.48	0.92	0.10	25.59	0.81	0.19	32.63	0.89	0.21
GPancake [3]	RNNSLAM+3DGS	40.34	0.97	0.05	26.25	0.83	0.21	32.31	0.90	0.20
Ours	Dynamic3DGS	46.79	0.99	0.02	30.63	0.92	0.14	37.82	0.94	0.15

(II) Ablation Experiments

Table 3: Ablation experiment of D⁴Recon on surgical reconstruction datasets. Mean values of cutting and pulling scenes are reported for EndoNeRF dataset.

Exp#	D_{HDG}	D_{SDG}	D_{Any}	$\mathcal{L}_{\text{SDS-S}}$	$\mathcal{L}_{\text{SDS-T}}$	$\mathcal{L}_{\text{SDS}}$	EndoNeRF PSNR ↑	SSIM ↑	LPIPS ↓	StereoMIS PSNR ↑	SSIM ↑	LPIPS ↓
(a)	✗	✗	✗	✗	✗	✗	31.03	0.886	0.129	25.63	0.766	0.331
(b)	✗	✗	✗	✗	✗	✓	32.19	0.903	0.853	26.71	0.781	0.305
(c)	✗	✗	✗	✗	✓	✗	36.67	0.945	0.077	29.27	0.820	0.214
(d)	✗	✗	✗	✓	✓	✗	38.42	0.968	0.052	31.38	0.886	0.176
(e)	✗	✗	✓	✓	✓	✗	38.89	0.970	0.049	32.13	0.890	0.171
(f)	✗	✓	✗	✓	✓	✗	39.69	0.973	0.045	33.91	0.902	0.168
(g)	✓	✓	✗	✓	✓	✗	**40.06**	**0.982**	**0.039**	**35.03**	**0.910**	**0.155**

(III) Qualitative Results