SurgMark: Structure-Aware Watermarking for 4D Gaussian Splatting in Surgical Scenes

Minjae Jeong*, Jaejin Lee*, Yechan Hwang, and Won Hwa Kim

Pohang University of Science and Technology (POSTECH), Pohang, South Korea
{minjaetidtid, jlee4923, yechan99, wonhwa}@postech.ac.kr

Abstract. Watermarking surgical scene representations learned with 4D Gaussian Splatting (4DGS) is challenging due to continuous deformation of anatomical structures and articulated motion of surgical instruments. Although watermarking has been widely studied for images, videos, and static 3D scene representations, these approaches do not directly generalize to dynamic 4D surgical scenes, where naive watermark embedding can disrupt temporal stability or interfere with the analysis of surgical scenes. To address this problem, we propose *SurgMark*, a watermarking framework tailored to 4DGS in surgical scenes. SurgMark consists of: 1) a temporal coherence regularization that preserves the original motion trajectories of Gaussians after watermark embedding; 2) a motion-aware gradient weighting strategy that adaptively scales watermark strength according to the motion magnitude of individual Gaussians; and 3) a segmentation feature consistency loss based on foundation segmentation models to prevent degradation of segmentation performance on anatomical regions and surgical instruments. Extensive experiments on two public surgical datasets demonstrate that SurgMark enables robust and imperceptible watermark embedding while maintaining rendering quality and segmentation performance.

Keywords: 4D Gaussian Splatting · Watermarking · Segmentation

1 Introduction

Recent advances in 4D Gaussian Splatting (4DGS) [21] enable high-fidelity reconstruction of dynamic 3D surgical scenes from minimally invasive surgery videos [7, 14]. These reconstructions are essential for improving spatial understanding and navigation in robotic-assisted surgery [6,8], surgical simulation [16], and AR/VR-based training [22]. As such surgical 4DGS models are extensively adopted within clinical institutions, they naturally become valuable institutional assets derived from sensitive clinical data and expertise. Consequently, safeguarding their authenticity and ownership has become a major concern.

To address this security issue, one potential approach is to embed watermark signals directly into the parameters of a Gaussian Splatting model. Prior works

* These authors contributed equally to this work.

on static 3DGS watermarking [2, 5, 9] leverage differentiable rendering to optimize Gaussian parameters such that rendered views can be decoded as a target watermark. In these approaches, a watermark decoder is first pre-trained and kept frozen, while the Gaussian parameters are fine-tuned to satisfy the decoding objective, thereby embedding the watermark signal into the 3D representation.

In contrast to static 3DGS representations, surgical scenes are inherently dynamic and are reconstructed using 4DGS, where Gaussians continuously deform over time due to anatomical motion and instrument interaction. In such settings, Gaussian parameters are tightly coupled with temporally evolving motion trajectories through a learned deformation mechanism. Directly modifying these time-varying parameters for watermark embedding, without explicitly accounting for motion dynamics, may distort the original geometric trajectories of Gaussians and introduce temporal inconsistencies across rendered frames.

Moreover, surgical 4DGS models are not used solely for visual rendering but are often incorporated into downstream analytical tasks, particularly for anatomical and instrument segmentation [6, 13] in robot-assisted surgery. Such segmentation is typically performed using large-scale foundation models [12, 15, 18] pretrained on diverse visual corpora. These models operate on semantically structured feature representations that capture anatomical structure and instrument context. Consequently, even subtle parameter perturbations introduced during watermark embedding may alter the underlying feature responses, leading to unstable segmentation outputs and potentially undermining safe robotic navigation and context-aware surgical assistance.

In this regard, we propose *SurgMark*, a novel structure-aware watermarking framework tailored to 4DGS models for surgical scenes. To preserve motion and semantic structures, *SurgMark* embeds watermark under explicit structural constraints, which lead to the following **contributions: 1)** We introduce a temporal coherence regularization that constrains watermark embedding to follow the original motion trajectories of Gaussians, which prevents distortion of learned deformation dynamics. **2)** We propose a motion-aware modulation strategy that adaptively adjusts watermark strength according to the motion magnitude of individual Gaussians. **3)** We preserve intermediate features of segmentation models to maintain semantic cues critical for anatomical regions and surgical instruments during watermark embedding. Extensive experiments on two public surgical scene datasets demonstrate that SurgMark achieves robust watermark embedding while preserving rendering quality and segmentation performance.

2 Method

As shown in Fig. 1, our method embeds watermark directly into dynamic Gaussian parameters while explicitly preserving structural properties of surgical scenes, including motion coherence and segmentation-relevant representations. SurgMark consists of two stages: (1) pre-training a latent-space watermark decoder for reliable message recovery (Sec. 2.3), and (2) fine-tuning a pretrained 4DGS model with learnable time-varying offsets under structural constraints. (Sec. 2.4)

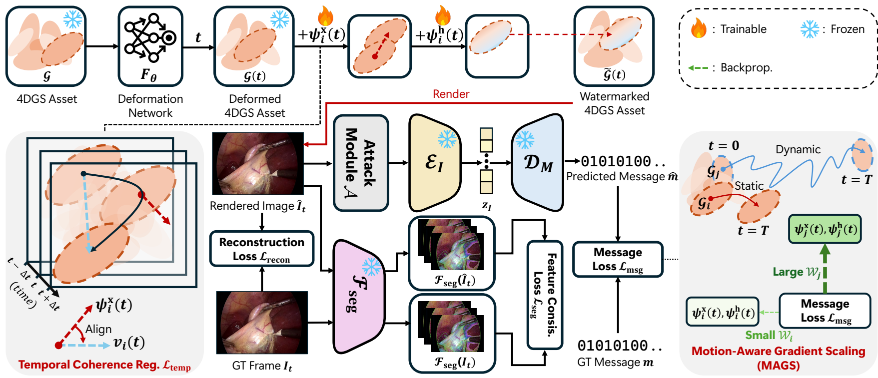

Fig. 1: Overview of *SurgMark*. A pre-trained 4DGS asset $\mathcal{G}$ is fine-tuned by learning watermark offsets $\psi_i^{\mathbf{x}}(t)$ and $\psi_i^{\mathbf{h}}(t)$, yielding $\tilde{\mathcal{G}}(t)$. Attacked rendering $\mathcal{A}(\tilde{I}_t)$ is encoded by CLIP encoder $\mathcal{E}_\mathrm{I}$ into $\mathbf{z}_\mathrm{I}$, which the frozen decoder $\mathcal{D}_\mathrm{M}$ uses to recover the message $\mathbf{m}$. Offsets are optimized with $\mathcal{L}_{\mathrm{msg}}$ for reliable decoding, while $\mathcal{L}_{\mathrm{recon}}$ preserves visual fidelity, $\mathcal{L}_{\mathrm{temp}}$ preserves motion dynamics, MAGS modulates watermark strength, and $\mathcal{L}_{\mathrm{seg}}$ preserves segmentation features.

2.1 Preliminaries: 3D and 4D Gaussian Splatting

3D Gaussian Splatting (3DGS) [10] reconstructs static 3D scenes in real-time by representing a scene as trainable Gaussian primitives $\mathcal{G} = \{(\mathbf{x}_i, \mathbf{h}_i, \alpha_i, \Sigma_i)\}_{i=1}^{N_\mathcal{G}}$, where $\mathbf{x}_i \in \mathbb{R}^3$ denotes the mean position, $\mathbf{h}_i$ the spherical harmonic coefficients, α_i the opacity, Σ_i the covariance, and $N_\mathcal{G}$ the number of 3D Gaussians. A differentiable Gaussian splatting renderer $\mathcal{R}$ [23] produces images $\hat{\mathbf{I}} = \mathcal{R}(\mathbf{v}, \mathcal{G})$ at a given viewpoint $\mathbf{v}$, and $\mathcal{G}$ is optimized by minimizing L1 loss combined with a term based on SSIM [20], which is defined as $\mathcal{L}_{\mathrm{recon}} = \mathcal{L}_1(\mathbf{I}, \hat{\mathbf{I}}) + (1 - \mathcal{L}_{\mathrm{SSIM}}(\mathbf{I}, \hat{\mathbf{I}}))$.
4D Gaussian Splatting (4DGS) [21] extends 3DGS to dynamic scenes by allowing the Gaussians to vary over time t as $\mathcal{G}(t) = \{(\mathbf{x}_i(t), \mathbf{h}_i(t), \alpha_i(t), \Sigma_i(t))\}_{i=1}^{N_\mathcal{G}}$. The temporal dynamics are modeled by a deformation network F_θ that maps the static Gaussians $\mathcal{G} = \mathcal{G}(0)$ to time-varying states via $\mathcal{G}(t) = F_\theta(\mathcal{G}, t)$. By jointly optimizing the deformation network and Gaussian parameters, 4DGS enables real-time reconstruction of dynamic 4D scenes with high visual fidelity.

2.2 Problem Definition

Given a binary watermark message $\mathbf{m} \in \{0, 1\}^L$, our goal is to embed $\mathbf{m}$ into a pre-trained surgical 4DGS model $\mathcal{G}$ such that the resulting watermarked model $\tilde{\mathcal{G}}$ preserves visual fidelity while enabling reliable message recovery. This is achieved by introducing learnable time-varying offsets $\psi_i^{\mathbf{x}}(t)$ and $\psi_i^{\mathbf{h}}(t)$ to the $\mathbf{x}_i(t)$ and $\mathbf{h}_i(t)$, respectively, where these offsets encode the watermark information:

$$\tilde{\mathcal{G}}(t) = \{\mathbf{x}_i(t) + \psi_i^{\mathbf{x}}(t),\ \mathbf{h}_i(t) + \psi_i^{\mathbf{h}}(t),\ \alpha_i(t),\ \Sigma_i(t)\}_{i=1}^{N_\mathcal{G}}. \tag{1}$$

Here, $\alpha_i(t)$ and $\Sigma_i(t)$ remain unperturbed to avoid direct geometric distortion and visual artifacts. For an arbitrary $\mathbf{v}$ and t, the message is recovered by a general watermark decoding function $\mathcal{D}$ as $\hat{\mathbf{m}} = \mathcal{D}(\tilde{\mathbf{I}}_t)$, where $\tilde{\mathbf{I}}_t = \mathcal{R}(\mathbf{v}, \tilde{\mathcal{G}}(t))$.

2.3 Pre-training the Watermark Decoder

We adopt the watermark decoder introduced in GuardSplat [2], which operates in the latent embedding space of a pretrained CLIP [17] model. Decoding in a latent space decoupled from direct pixel-level supervision enables invisible watermarking, which preserves fine anatomical structures in surgical scenes.

Specifically, the watermark decoder $\mathcal{D}_\mathrm{M}$ is trained to recover an arbitary binary message $\mathbf{m} \in \{0,1\}^L$ from CLIP embeddings. $\mathbf{m}$ is converted into a token sequence processed by the frozen CLIP text encoder $\mathcal{E}_\mathrm{T}$. A fixed bit-to-token mapping table $\Phi[\cdot,\cdot]$ maps each bit value and index to a unique CLIP vocabulary token, yielding the token sequence $\mathbf{w} = \{\Phi(m_i, i)\}_{i=1}^{L}$. $\mathbf{w}$ is zero-padded and encoded by $\mathcal{E}_\mathrm{T}$ to obtain a CLIP text embedding $\mathbf{z}_\mathrm{T} = \mathcal{E}_\mathrm{T}(\mathbf{w})$. The message decoder $\mathcal{D}_\mathrm{M}$, implemented as a multi-layer perceptron, predicts the message logits as $\hat{\mathbf{m}} = \mathcal{D}_\mathrm{M}(\mathbf{z}_\mathrm{T})$ and is trained using $\mathcal{L}_\mathrm{msg} = -\sum_{i=1}^{L}\left[m_i \log \hat{m}_i + (1-m_i)\log(1-\hat{m}_i)\right]$.

During watermark embedding, $\tilde{\mathbf{I}}_t$ is encoded by the CLIP image encoder $\mathcal{E}_\mathrm{I}$ to obtain $\mathbf{z}_\mathrm{I} = \mathcal{E}_\mathrm{I}(\tilde{\mathbf{I}}_t)$, which is passed to frozen $\mathcal{D}_\mathrm{M}$ to recover $\mathbf{m}$.

2.4 Fine-tuning the Surgical 4DGS Model

As formulated in Eq. 1, we fine-tune a pretrained 4DGS model $\mathcal{G}$ by adding learnable time-varying watermark offsets $\boldsymbol{\psi}_i^{\mathbf{x}}$ and $\boldsymbol{\psi}_i^{\mathbf{h}}$ to its Gaussian parameters, which results in a watermarked model $\tilde{\mathcal{G}}$. The optimization is guided by three structural constraints to ensure robust and imperceptible watermark embedding, while preserving important feature representations for surgical segmentation.

Temporal Coherence Regularization. To preserve the original motion patterns of Gaussians, we enforce directional consistency between the original Gaussian trajectory and the offsets $\boldsymbol{\psi}_i^{\mathbf{x}}$ at each t. This encourages $\boldsymbol{\psi}_i^{\mathbf{x}}$ to align with the underlying motion direction instead of introducing arbitrary perturbations.

We define the local unit velocity of the i-th Gaussian based on its original trajectory as $\mathbf{v}_i(t) = \frac{\mathbf{x}_i(t+\Delta t) - \mathbf{x}_i(t-\Delta t)}{\|\mathbf{x}_i(t+\Delta t) - \mathbf{x}_i(t-\Delta t)\|_2}$, where Δt denotes the time interval between adjacent timesteps. Similarly, the unit direction of $\boldsymbol{\psi}_i^{\mathbf{x}}$ is computed as $\bar{\boldsymbol{\psi}}_i^{\mathbf{x}}(t) = \boldsymbol{\psi}_i^{\mathbf{x}}(t)/\|\boldsymbol{\psi}_i^{\mathbf{x}}(t)\|_2$. The temporal coherence regularization is an average cosine distance of $\mathbf{v}_i(t)$ and $\bar{\boldsymbol{\psi}}_i^{\mathbf{x}}(t)$ over all Gaussians, which is defined as

$$\mathcal{L}_\mathrm{temp} = \frac{1}{N_\mathcal{G}} \sum_{i=1}^{N_\mathcal{G}} \left(1 - \mathbf{v}_i(t)^\top \bar{\boldsymbol{\psi}}_i^{\mathbf{x}}(t)\right). \tag{2}$$

Motion-Aware Gradient Scaling (MAGS). Surgical 4DGS scenes exhibit highly heterogeneous motion patterns across Gaussians. Uniform watermark supervision may cause insufficient watermarking in dynamic regions or unnecessarily perturb static regions. To adapt the embedding strength to the underlying motion characteristics of each Gaussian, we propose an MAGS strategy.

We define a cumulative motion score as $\mathcal{S}_i = \sum_{t=0}^{T-1} \|\mathbf{x}_i(t+\Delta t) - \mathbf{x}_i(t)\|_2$, where T denotes the number of timesteps in the surgical 4DGS scene. Then, the score is normalized using min–max normalization across all Gaussians to obtain motion-aware scales as $\mathcal{W}_i = \frac{\mathcal{S}_i - \min(\mathcal{S})}{\max(\mathcal{S}) - \min(\mathcal{S}) + \epsilon} \in [0,1]$, where $\min(\mathcal{S})$ and $\max(\mathcal{S})$ denote the minimum and maximum motion scores over all Gaussians, respectively, and $\epsilon \in \mathbb{R}^+$ is a small constant for numerical stability.

Rather than reweighting the message loss directly, MAGS modulates the gradients of per-Gaussian watermark offset parameters. Specifically, the gradients of $\boldsymbol{\psi}_i^{\mathbf{x}}$ and $\boldsymbol{\psi}_i^{\mathbf{h}}$ with respect to $\mathcal{L}_{\text{msg}}$ are scaled by $\mathcal{W}_i$ as

$$\nabla_{\boldsymbol{\psi}_i^{\mathbf{x}}} \leftarrow \mathcal{W}_i \cdot \nabla_{\boldsymbol{\psi}_i^{\mathbf{x}}} \mathcal{L}_{\text{msg}}, \qquad \nabla_{\boldsymbol{\psi}_i^{\mathbf{h}}} \leftarrow \mathcal{W}_i \cdot \nabla_{\boldsymbol{\psi}_i^{\mathbf{h}}} \mathcal{L}_{\text{msg}}, \tag{3}$$

which allows dynamic Gaussians to receive stronger watermark supervision while reducing unnecessary perturbations in static regions.

Segmentation Feature Consistency Loss. Interpretability is critical in surgical scenes, where anatomical segmentation is a key component. To prevent degradation of high-level semantic features critical for anatomical segmentation, we enforce consistency between intermediate feature maps of ground-truth and watermarked renderings using a frozen foundation segmentation model. Let $\mathcal{F}_{\text{seg}}(\cdot)$ denote the feature extractor of the segmentation model. The segmentation feature consistency loss $\mathcal{L}_{\text{seg}}$ is defined as

$$\mathcal{L}_{\text{seg}} = \left\| \mathcal{F}_{\text{seg}}(\tilde{\mathbf{I}}_t) - \mathcal{F}_{\text{seg}}(\mathbf{I}_t) \right\|_2^2. \tag{4}$$

Training Objective. To improve robustness of the watermark against realistic image degradations, we incorporate a differentiable attack module $\mathcal{A}$ during training. Specifically, the message decoded from the attacked rendered image is obtained as $\hat{\mathbf{m}}_{\text{att}} = \mathcal{D}_{\text{M}}(\mathcal{E}_{\text{I}}(\mathcal{A}(\tilde{\mathbf{I}}_t)))$, where $\mathcal{A}$ applies common image-level attacks (e.g., rotation, cropping, etc) on the rendered image. Then, $\mathcal{L}_{\text{msg}}$ in Sec. 2.3 is computed using $\hat{\mathbf{m}}_{\text{att}}$. The final loss is defined as a weighted sum of all losses

$$\mathcal{L}_{\text{final}} = \lambda_{\text{msg}} \mathcal{L}_{\text{msg}} + \lambda_{\text{recon}} \mathcal{L}_{\text{recon}} + \lambda_{\text{temp}} \mathcal{L}_{\text{temp}} + \lambda_{\text{seg}} \mathcal{L}_{\text{seg}}, \tag{5}$$

where λ_{msg}, λ_{recon}, λ_{temp}, and λ_{seg} are hyperparameters.

3 Experiment

3.1 Experimental Settings

Dataset. We conduct experiments on two public surgical video datasets for semantic segmentation: CholecSeg8K [4] and EndoVis18 [1]. These datasets contain 5 laparoscopic cholecystectomy and 2 robot-assisted surgical videos, respectively, both with pixel-level annotations for instruments and anatomical structures. We adopt an 8:1 train–test split per sequence as in SurgTPGS [6].

Implementation Details. The watermark decoder was trained following the setting of [2]. A base 4DGS model is pretrained based on [21], and its rendered

Table 1: Quantitative evaluation of watermarking performance, including multiple attack scenarios, namely None (no attack), Brt (brightness ×2), Crp (40% center cropping), JPG (JPEG compression at 50% quality), Rot (30° rotation), and Res (75% resizing). The best results are shown in **bold**, and the second-best results are <u>underlined</u>.

Method	PSNR ↑	SSIM ↑	**Bit Accuracy (%) ↑**						
			None	Brt	Crp	JPG	Rot	Res	Avg
				CholecSeg8K					
4DGS+HiDDeN [24]	27.83	0.9738	83.08	80.96	43.72	69.90	66.99	79.34	70.67
4DGS+SSL [3]	28.34	0.9747	82.28	77.55	50.40	53.71	60.00	78.19	67.02
4DGS+WAM [19]	27.64	0.9659	85.29	82.13	71.29	61.31	66.98	81.62	74.77
3D-GSW [9]	28.39	0.9752	87.85	84.31	55.36	<u>73.95</u>	<u>71.15</u>	85.46	76.35
GuardSplat [2]	<u>32.32</u>	<u>0.9874</u>	<u>90.42</u>	<u>87.45</u>	<u>71.82</u>	70.36	69.56	<u>88.89</u>	<u>79.75</u>
Ours (SurgMark)	**35.91**	**0.9939**	**98.08**	**96.73**	**90.21**	**90.35**	**88.28**	**97.79**	**93.57**
				EndoVis18					
4DGS+HiDDeN [24]	26.65	0.9812	82.29	78.97	44.50	62.80	66.27	75.95	68.46
4DGS+SSL [3]	26.81	0.9814	80.58	78.61	49.86	56.18	66.33	79.32	68.48
4DGS+WAM [19]	26.25	0.9769	85.27	79.25	<u>65.92</u>	62.95	66.00	78.80	73.03
3D-GSW [9]	27.37	<u>0.9837</u>	<u>88.10</u>	<u>84.62</u>	55.16	<u>72.92</u>	<u>74.01</u>	<u>86.61</u>	<u>76.90</u>
GuardSplat [2]	<u>31.45</u>	0.9833	85.77	82.49	64.04	72.47	68.50	83.83	76.18
Ours (SurgMark)	**37.96**	**0.9939**	**93.56**	**92.37**	**73.76**	**81.35**	**85.82**	**92.46**	**86.55**

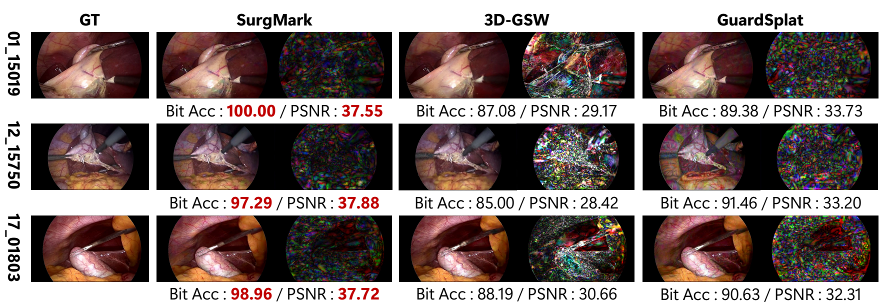

Fig. 2: Qualitative evaluation of watermarked renderings on CholecSeg8K.

frames are used as ground-truth references during watermarking to assess quality preservation and robustness. In Sec. 2.4, $\tilde{G}$ is fine-tuned using the Adam [11] optimizer with learning rates of 2e-4/1e-3 ($\boldsymbol{\psi}_i^{\mathbf{x}}(t)/\boldsymbol{\psi}_i^{\mathbf{h}}(t)$) for 25 epochs. λ_{msg} is set to 0.5, and λ_{recon}, λ_{temp}, λ_{seg} are all set to 1.Following [13], bounding boxes are used as prompts for the segmentation models, where each box is computed from the ground-truth mask of its class.

Baselines & Evaluation. We compare SurgMark with 1) 2D image watermarking decoders (4DGS+HiDDeN [24], SSL [3], WAM [19]) directly adapted to watermarking 4DGS models, where their optimizations are supervised with $\mathcal{L}_{\text{msg}}$ and $\mathcal{L}_{\text{recon}}$, and 2) 3DGS watermarking frameworks (3D-GSW [9], GuardSplat [2]). We report PSNR and SSIM [20] for rendering quality, bit accuracy for robustness, both on clean and attacked scenarios. To assess the impact of watermarking on segmentation performance and its generalizability across foundation

Table 2: Segmentation performance on the CholecSeg8K and EndoVis2018 dataset.

Model	Method	mIOU (%) ↑								
		01_00080	01_00240		01_15019		12_15750	17_01803	seq_5	seq_9
		Liver	Liver	Grasper	Abdominal Wall	Grasper	L-hook Electrocautery	Abdominal Wall	Kidney-Parenchyma	Instrument-Shaft
SAM [12]	w/o watermark	85.11	86.08	77.37	95.75	75.81	81.01	71.44	62.00	69.34
	3D-GSW [9]	77.57	75.48	65.21	91.83	64.78	62.69	63.90	54.57	54.62
	GuardSplat [2]	78.83	78.94	65.68	91.88	64.13	66.14	61.47	56.87	63.16
	Ours (w/o $\mathcal{L}_{seg}$)	78.15	79.19	66.13	92.21	66.47	66.16	68.13	58.44	62.61
	Ours (w/ $\mathcal{L}_{seg}$)	**83.28**	**85.97**	**74.22**	**94.65**	**73.01**	**80.93**	**71.04**	**60.51**	**69.37**
SAM2 [18]	w/o watermark	90.91	90.41	77.74	97.14	74.82	79.18	61.64	67.29	78.30
	3D-GSW [9]	73.91	83.80	56.49	91.75	65.38	62.28	55.64	62.02	56.36
	GuardSplat [2]	85.74	82.35	66.75	92.18	65.18	68.81	56.29	62.10	65.71
	Ours (w/o $\mathcal{L}_{seg}$)	86.85	88.54	68.90	93.83	67.47	70.08	58.57	63.57	69.16
	Ours (w/ $\mathcal{L}_{seg}$)	**89.74**	**90.35**	**75.47**	**95.72**	**72.39**	**77.54**	**60.07**	**65.17**	**76.40**
MedSAM [15]	w/o watermark	96.45	94.71	74.71	98.18	87.83	83.76	62.84	48.35	70.70
	3D-GSW [9]	88.97	88.92	59.23	88.06	74.92	64.55	60.26	38.59	56.63
	GuardSplat [2]	90.50	88.76	68.21	85.35	74.93	68.56	57.39	41.21	61.22
	Ours (w/o $\mathcal{L}_{seg}$)	90.84	92.65	69.57	92.93	76.65	72.39	60.48	44.59	63.52
	Ours (w/ $\mathcal{L}_{seg}$)	**94.57**	**94.29**	**73.14**	**96.21**	**86.11**	**83.17**	**62.72**	**47.18**	**68.63**

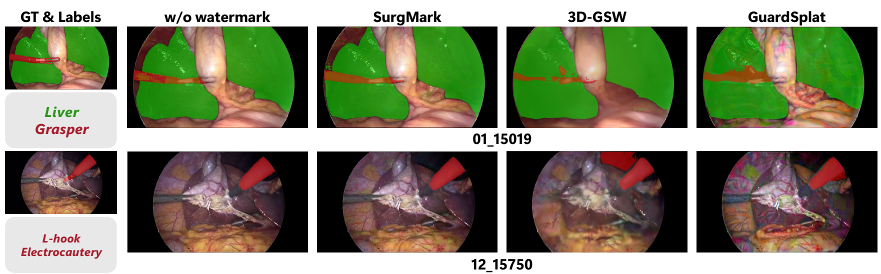

Fig. 3: Qualitative evaluation of segmentation performance on CholecSeg8K.

models, we evaluate SAM [12], SAM2 [18], and MedSAM [15] on watermarked renderings using mIoU. All evaluations are conducted with $L = 48$.

3.2 Quantitative Results

Watermarking performance. As shown in Tab. 1, SurgMark surpasses all baselines, achieving the highest bit accuracy and image fidelity. On CholecSeg8K, our method outperforms bit accuracy over second-best results by 7.66%p, while also increasing PSNR by 3.59dB. Under image-level attacks, SurgMark exhibits only an 8.47%p average drop in bit accuracy, whereas 3D-GSW and GuardSplat experience substantially larger degradations of 21.03%p and 19.84%p, demonstrating stronger robustness against common perturbations.

Segmentation performance. Tab. 2 reports segmentation performance on rendered frames from watermarked 4DGS models. SurgMark consistently outperforms baselines across all labels. Notably, for fine-grained categories such as *Grasper* (*01_00240*), *L-hook Electrocautery* (*12_15750*), and *Instrument_Shaft* (*seq_9*), it achieves substantially higher mIoU than its counterpart without $\mathcal{L}_{\text{seg}}$, demonstrating that $\mathcal{L}_{\text{seg}}$ effectively preserves subtle and detailed structures.

Table 3: Ablation study of *SurgMark*. Results are averaged over videos in the CholecSeg8K dataset.

Use of offsets		Use of components		Bit Acc. (%) ↑	PSNR ↑
$\psi_i^{\mathbf{h}}(t)$	$\psi_i^{\mathbf{x}}(t)$	$\mathcal{L}_{\text{temp}}$	MAGS		
✓	×	×	×	90.42	32.32
×	✓	×	×	88.11	26.63
✓	✓	×	×	93.30	29.24
✓	✓	✓	×	93.97	34.46
✓	✓	✓	✓	**98.08**	**35.91**

Ablation Study. As shown in Tab. 3, using only $\psi_i^{\mathbf{h}}(t)$ yields limited watermark capacity due to restricted parameter freedom, while using only $\psi_i^{\mathbf{x}}(t)$ not only suffers from insufficient capacity but also degrades rendering quality by introducing unconstrained temporal perturbations that disrupt motion consistency. Combining both offsets increases embedding capacity and improves bit accuracy; however, rendering fidelity remains limited. Introducing $\mathcal{L}_{\text{temp}}$ significantly improves PSNR by preserving the original Gaussian trajectories to maintain temporal coherence. Finally, incorporating MAGS achieves the best performance by adaptively scaling watermark gradients according to motion magnitude, which emphasizes the importance of motion-aware supervision in watermarking surgical scenes with diverse motion dynamics.

3.3 Qualitative Results

Fig. 2 presents the rendered watermarked images along with their corresponding difference maps (amplified by $\times 10$) with respect to the ground-truth image. 3D-GSW noticeably destroys or blurs fine anatomical structures and instruments, leading to large visual discrepancies as shown in the difference maps. Also, despite these visible artifacts, it still yields relatively low bit accuracy. GuardSplat exhibits noticeable color distortions in the watermarked renderings, and similarly achieves limited bit accuracy. Conversely, SurgMark produces watermarked images that are nearly indistinguishable from the ground truth, with the smallest difference maps among all methods. Segmentation results in Fig. 3 show that baselines yield degraded masks in fine-grained regions. In contrast, SurgMark preserves masks in thin instruments and anatomical boundaries.

4 Conclusion

In this work, we propose *SurgMark*, a structure-aware watermarking framework for surgical 4DGS models. SurgMark preserves motion dynamics and adaptively scales watermark strength to balance robustness and visual fidelity, while retaining segmentation-relevant representations. Experiments validate reliable watermark recovery and stable segmentation performance of our method, demonstrating practical copyright protection for surgical 4D scene representations.

Acknowledgments. This research was supported by RS-2026-25494850 (60%), RS-2022-II220290 (30%), RS-2019-II191906 (AI Graduate Program at POSTECH, 5%). This work was also partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - ITRC (Information Technology Research Center) grant funded by the Korea government (MSIT) (IITP-2026-RS-2024-00437866, 5%).

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., et al.: 2018 robotic scene segmentation challenge. arXiv preprint arXiv:2001.11190 (2020)
2. Chen, Z., Wang, G., et al.: Guardsplat: Efficient and robust watermarking for 3d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 16325–16335 (2025)
3. Fernandez, P., Sablayrolles, A., et al.: Watermarking images in self-supervised latent spaces. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3054–3058. IEEE (2022)
4. Hong, W.Y., Kao, C.L., et al.: Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80. arXiv preprint arXiv:2012.12453 (2020)
5. Huang, X., Li, R., et al.: Gaussianmarker: Uncertainty-aware copyright protection of 3d gaussian splatting. Advances in Neural Information Processing Systems **37**, 33037–33060 (2024)
6. Huang, Y., Bai, L., et al.: Surgtpgs: Semantic 3d surgical scene understanding with text promptable gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 584–594. Springer (2025)
7. Huang, Y., Cui, B., et al.: Endo-4dgs: Endoscopic monocular scene reconstruction with 4d gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 197–207. Springer (2024)
8. Huang, Y., Cui, B., et al.: Advancing dense endoscopic reconstruction with gaussian splatting-driven surface normal-aware tracking and mapping. In: 2025 IEEE International Conference on Robotics and Automation (ICRA). pp. 4084–4091. IEEE (2025)
9. Jang, Y., Park, H., et al.: 3d-gsw: 3d gaussian splatting for robust watermarking. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5938–5948 (2025)
10. Kerbl, B., Kopanas, G., et al.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
11. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Kirillov, A., Mintun, E., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
13. Li, K., Wang, J., et al.: Feature-endogaussian: Feature distilled gaussian splatting in surgical deformable scene reconstruction. arXiv preprint arXiv:2503.06161 (2025)
14. Liu, Y., Li, C., et al.: Foundation model-guided gaussian splatting for 4d reconstruction of deformable tissues. IEEE Transactions on Medical Imaging (2025)

15. Ma, J., He, Y., et al.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
16. Martyniak, S., Kaleta, J., et al.: Simuscope: Realistic endoscopic synthetic dataset generation through surgical simulation and diffusion models. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4268–4278. IEEE (2025)
17. Radford, A., Kim, J.W., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
18. Ravi, N., Gabeur, V., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
19. Sander, T., Fernandez, P., et al.: Watermark anything with localized messages. In: International Conference on Learning Representations (ICLR) (2025)
20. Wang, Z., Bovik, A.C., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (2004)
21. Wu, G., Yi, T., et al.: 4d gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20310–20320 (June 2024)
22. Yang, Z.: Efficient 3d scene reconstruction and simulation from sparse endoscopic views. arXiv preprint arXiv:2509.17027 (2025)
23. Yifan, W., Serena, F., et al.: Differentiable surface splatting for point-based geometry processing. ACM Transactions On Graphics (TOG) **38**(6), 1–14 (2019)
24. Zhu, J., Kaplan, R., et al.: Hidden: Hiding data with deep networks. In: Proceedings of the European conference on computer vision (ECCV). pp. 657–672 (2018)