PHYSMARK: POPULATION-GUIDED CARRIER LEARNING FOR GENERALIZABLE PHYSICAL-CHANNEL WATERMARKING IN 3D GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Medical volumes reconstructed with 3D Gaussian Splatting (3DGS) become digital assets for case discussion, teaching, and online sharing. To support provenance and copyright tracking during both digital and physical dissemination, a verifiable message can be embedded into the 3DGS representation. However, most existing 3DGS watermarking methods design carriers for a single asset and do not explicitly exploit cross-asset regularities in carrier reliability, which limits generalization to unseen assets. We observe that the relative reliability of candidate carriers transfers across 3DGS assets: carriers that are relatively reliable on a finite set of seen assets tend to remain relatively reliable on unseen ones. Based on this observation, we propose PhysMark, a generalizable physical-channel 3DGS watermarking framework. PhysMark converts this regularity into a learnable, structured carrier prior: it initializes structured sparse carriers from population-level responses and then refines the carrier representation through the watermark recovery objective, turning asset-specific carrier search into prior-guided cross-asset carrier learning. In addition, PhysMark integrates the print-and-capture process into training, allowing the carrier to adapt to distortions introduced by printing, halftoning, and photographing. Experiments on medical 3DGS assets across multiple message lengths, unseen assets, digital-domain transformations, and real print-and-capture conditions show that the carrier reliability ranking transfers well across assets, that the prior-guided carrier improves recovery stability on unseen assets compared with single-asset carrier design, and that practical detectability is achieved under real printing and photographing.

1 INTRODUCTION

3D Gaussian Splatting (3DGS) represents three-dimensional content with explicit Gaussian primitives and renders high-quality views through efficient differentiable rasterization, and is an important paradigm for 3D digital assets (Kerbl et al., 2023). In medical scenarios, the 3D volumes of a large number of subjects can be reconstructed into 3DGS assets for case discussion, teaching, and digital sharing. To support provenance verification and copyright tracking during subsequent digital and physical dissemination, a verifiable, asset-associated watermark can be embedded into the 3DGS representation. However, 3DGS consists of a large number of Gaussians with diverse attributes and spatial structures; the same watermark perturbation can produce observably different responses on different assets, which introduces uncertainty when carriers are applied across assets.

Recent 3DGS watermarking research has extended toward model-parameter embedding, rendering robustness, compression and quantization, and cross-content generalization (Huang et al., 2024; Jang et al., 2025; Chen et al., 2025; Huang et al., 2025; In et al., 2026; Li et al., 2026; Cai et al., 2026; Zhao et al., 2026; Wang et al., 2026; Li et al., 2025b). GaussianMarker, 3D-GSW, and GuardSplat design carriers from Gaussian uncertainty, frequency and rendering contribution, and spherical-harmonic attributes respectively (Huang et al., 2024; Jang et al., 2025; Chen et al., 2025); CompMarkGS and Robust3DGSW further address watermark degradation caused by compression and quantization (In et al., 2026; Wang et al., 2026). Meanwhile, MarkSplatter, GS-Marker, and W2M begin to study cross-content generalization and asset-adaptive embedding strategies (Huang et al., 2025; Li et al., 2026; 2025b). Despite this progress, whether a reusable carrier regularity exists

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

across 3DGS assets, and whether such a regularity can support watermark generalization to unseen assets, remains under-studied.

We investigate this question from the perspective of *cross-asset carrier reliability*. For candidate carriers defined in a common canonical space, the absolute response may vary across assets, but what matters more is whether the *relative* reliability among carriers stays consistent. Our analysis shows that the relative reliability of candidate carriers exhibits a transferable ranking across 3DGS assets: a carrier that is relatively reliable on a few seen assets tends to remain relatively reliable on unseen ones. This indicates that carrier construction need not rely on independent search per asset; instead, shared information about carrier reliability can be extracted from several seen assets and used as a prior for carrier learning on unseen ones.

Based on this observation, we propose PhysMark, a generalizable physical-channel 3DGS watermarking framework. PhysMark converts the cross-asset reliability regularity into a learnable structured carrier prior: it first constructs the carrier support and its initialization from population information, and then refines the carrier representation through the watermark recovery objective, so that both cross-asset shared regularities and asset-specific differences are exploited. PhysMark further incorporates the print-and-capture process into training, so that carrier learning adapts to the physical-channel distortions introduced by printing, halftoning, and photographing, and supports reference-free message recovery after dissemination. Our contributions are:

- We systematically analyze the response of candidate carriers across 3DGS assets and find that their relative reliability forms a transferable ranking, providing a population-level prior for carrier construction on unseen assets.
- We convert the cross-asset reliability regularity into a structured carrier prior and refine the carrier representation on top of it, combining shared regularities with asset-specific differences to improve generalization to unseen assets.
- We embed the print-and-capture process into both training and evaluation, establishing a complete pipeline from a digital 3DGS asset to real printing, photographing, and reference-free message recovery.

2 Related Work

Watermarking for 3D Gaussian Splatting 3DGS uses explicit Gaussian primitives as an operable carrier. Medical 3DGS reconstruction has advanced (Yao et al., 2025), and Scaffold-GS provides structured anchor-based representations (Lu et al., 2024). Existing watermarking methods design carriers from Gaussian uncertainty (Huang et al., 2024), frequency and rendering contribution (Jang et al., 2025), or spherical-harmonic attributes (Chen et al., 2025). Compression methods such as HAC, ContextGS, LightGaussian, and PCGS reduce storage but their quantization degrades embedded watermarks (Chen et al., 2024; Wang et al., 2024; Fan et al., 2024; Chen et al., 2026); reviews are provided in (Wu et al., 2024; Youn et al., 2025). Compression and quantization robustness are addressed by CompMarkGS and Robust3DGSW (In et al., 2026; Wang et al., 2026). Steganographic and native extensions are explored in GS-Hider, SecureGS, WateRF, NGS-Marker, Mark4D, InstantSplamp, and StegaNeRF (Zhang et al., 2024; 2025; Jang et al., 2024; Qin et al., 2026; Lee et al., 2026; Li et al., 2025a; 2023), while NeRF Signature and Levenberg–Marquardt watermarking extend to radiance fields and 3D meshes (Luo et al., 2025; Narendra et al., 2024). Attacks and threat models are studied in (Zeng et al., 2026; Deng et al., 2026). Overall, existing work centers on Gaussian-attribute embedding and digital-domain robustness.

Generalizable Watermarking and Carrier Selection For multi-asset deployment, MarkSplatter uses a GaussianBridge for structured representations (Huang et al., 2025), GS-Marker adopts a shared encoder–decoder (Li et al., 2025b), W2M adjusts embedding via policy learning (Li et al., 2026), and Where-What-Why analyzes carrier selection (Cai et al., 2026). Cross-asset generalization relates to medical domain generalization (Yoon et al., 2024) and to rendering generalization in Mip-Splatting, which constrains 3D frequency and replaces 2D dilation with a mip filter (Yu et al., 2024). These works show that generalization depends on carrier selection and per-content response, but carrier reliability transfer across assets remains underexplored.

Print-and-Capture Physical Watermarking Printing, display, and camera recapture cause composite distortions. HiDDeN and StegaStamp demonstrate deep-learning-based robust embedding (Zhu et al., 2018; Tancik et al., 2020). Physical watermarking models color conversion, halftoning, printing degradation, and camera factors (Yu et al., 2005; Bulan et al., 2010). Learnable physical channels (Chen & Li, 2024), deep noise simulation networks (Qin et al., 2023), few-shot cross-channel noise (Fang et al., 2023), lightweight print-camera networks (Guo & Qin, 2024), and unsupervised screen-camera noise layers (Wu et al., 2026) adapt to real dissemination. For 3D printed objects, blind watermarking modifies layer thickness (Delmotte et al., 2020) or uses moment alignment and surface norm distribution (Delmotte et al., 2021). AI-generated image watermarking, active steganography, and medical image watermarking are explored in (Fernandez et al., 2023; Xu et al., 2025; Wen et al., 2023; Yan et al., 2025). Unlike 2D pixel-space methods, 3DGS watermarking modifies Gaussian representations affected by 3D structure and view-dependent rendering, so 2D print-and-capture models do not directly cover Gaussian-level carrier propagation in 3DGS.

3 METHOD

Existing 3DGS watermarking searches carriers per asset, but medical 3DGS varies across subjects in Gaussian visibility, rendering contribution, and propagation response, making single-asset carriers non-transferable. PhysMark observes that absolute carrier responses vary while relative reliability transfers. Carrier learning thus has two stages: population reliability from multi-instance, multi-channel measurements sets structured support and initialization; instance training then refines the carrier under this prior. The watermark is embedded in log-scale Gaussian scale and mapped via instance registration, avoiding per-asset re-search. A differentiable physical channel models printing, halftoning, blur, illumination, JPEG, and sensor noise, with Expectation Over Transformation (EOT) for physical adaptation. The decoder recovers messages from propagated images without clean references or original Gaussians. See Figure 1. Given the i-th medical 3DGS instance G_i, its Gaussians consist of position μ, rotation q, scale s, color c, and opacity α. After canonical registration, the instance has stable Gaussian correspondences. For a watermark message $w \in \{-1, +1\}^K$, where K is the number of watermark bits and w_k is the k-th bit, PhysMark learns a carrier matrix B, produces scale perturbations on the instance to obtain a watermarked model G_i', and then obtains the observation $y_{i,v} = T_\theta(R_v(G_i'))$ through the rendering operator R_v and the propagation channel T_θ, where v is the view index and θ denotes the propagation-channel parameters. The goal is to stably recover the message on unseen instances and under random propagation conditions while preserving rendering quality.

3.1 CROSS-ASSET POPULATION-PRIOR-GUIDED CARRIER LEARNING

PhysMark first constructs a candidate carrier bank in a unified canonical space and learns population-level reliability from the propagation responses of multiple seen instances. Since Gaussians in different instances already have structural correspondences, the same canonical direction can be directly compared across different subjects. We first map the centers of the template Gaussians onto a regular 3D grid, sort them by Morton code, and divide them into R spatially compact local regions $\{\Omega_r\}$, where R is the number of regions and Ω_r is the r-th local region. Within each region, low-order polynomial bases are constructed from local coordinates and orthogonalized by Gram–Schmidt to obtain local candidate directions. The final candidate matrix is $\Phi = [\phi_1, \ldots, \phi_J] \in \mathbb{R}^{N \times J}$, where N is the number of Gaussians, L is the number of local directions per region, $J = RL$ is the total number of candidate directions, and ϕ_j is the j-th candidate direction. Each candidate direction satisfies $\mathrm{supp}(\phi_j) \subseteq \Omega_{r(j)}$ and $\|\phi_j\|_2 = 1$, where $\mathrm{supp}(\cdot)$ is the support set, $r(j)$ is the region index of candidate j, and $\|\cdot\|_2$ is the L2 norm. This local support concentrates the perturbation of a single bit in a local brain region, thereby controlling visual distortion and reducing global interference among different bits. For a candidate direction ϕ, we define its observable response under instance i and propagation channel c as

$$r_{i,c}(\phi) = \|T_c(R_v(G_i + \alpha\phi)) - T_c(R_v(G_i - \alpha\phi))\|_2 \tag{1}$$

where c is the channel index ($c \in \{d, p\}$, corresponding to the digital channel and the print-and-capture channel, respectively), α is the probe perturbation amplitude, and T_c is channel c. Unlike selecting directly based on the magnitude of Gaussian parameters, this definition measures the effective response that a candidate carrier can ultimately produce after rendering and propagation. We

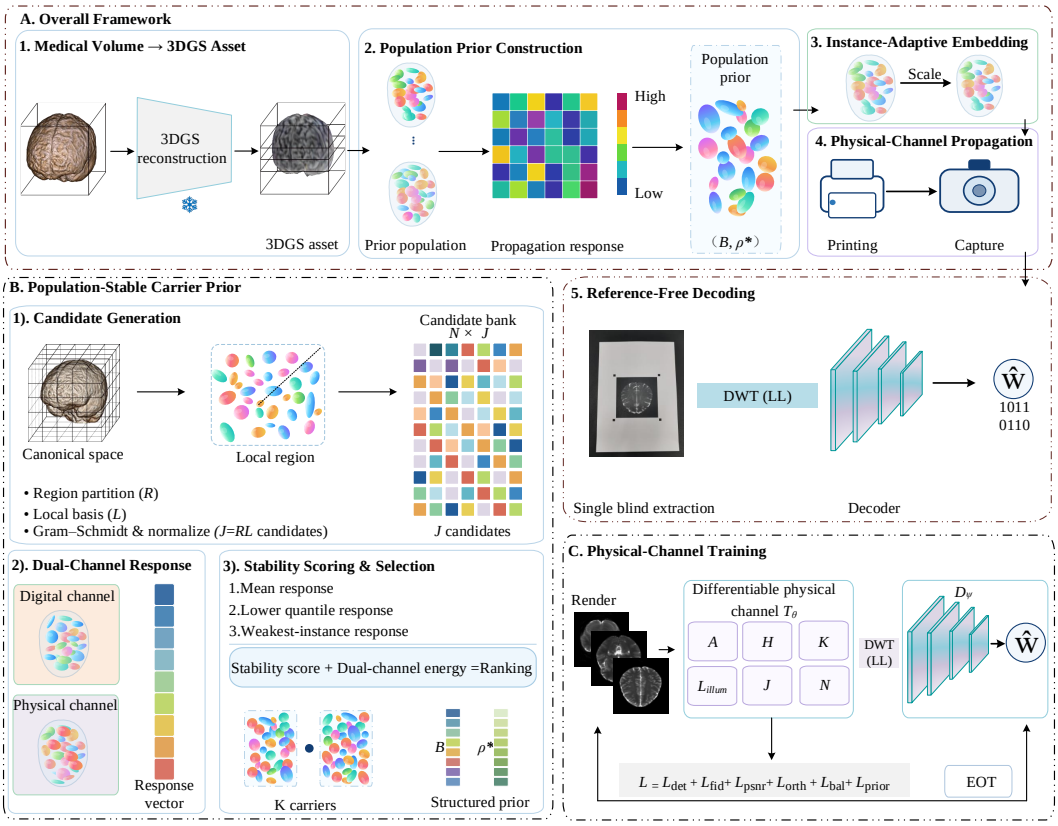

Figure 1: Overview of PhysMark: population-stable carrier prior, instance mapping for log-scale embedding, and end-to-end print-and-capture training with blind decoding.

consider both the digital channel T_d and the print-and-capture channel T_p. For candidate j, the response set $\{u_{i,c,j}\}$ is computed over the prior instance population, from which we further extract the population mean response $\bar{u}_{c,j}$, the lower-quantile response $q_{0.1,c,j}$, and the weakest-instance response $m_{c,j}$, where $u_{i,c,j}$ is the response norm of candidate j under instance i and channel c. The mean response describes the overall detection capability, while the lower quantile and weakest-instance response focus more on cross-subject stability. After normalization, the channel-level stability is fused into the population stability score of the candidate carrier:

$$S_j = \sum_c \omega_c \ \mathrm{norm}(\lambda_1 \bar{u}_{c,j} + \lambda_2 q_{0.1,c,j} + \lambda_3 m_{c,j}) \tag{2}$$

where ω_c is the channel weight, $\lambda_1, \lambda_2, \lambda_3$ are the fusion weights for the three statistics, $\mathrm{norm}(\cdot)$ is a normalization function, and S_j is the population stability score of candidate j. Meanwhile, to emphasize the importance of real physical propagation, we compute the dual-channel response energy $E_j = \omega_d \|u_{d,j}\|_2 + \omega_p \|u_{p,j}\|_2$, where ω_d, ω_p are the energy weights for the digital and print channels with $\omega_p > \omega_d$, and $u_{d,j}, u_{p,j}$ are the response vectors of candidate j under the digital and print channels, respectively. Finally, the response energy and stability are jointly used to construct a normalized reliability ranking of candidate carriers: $s_j = (1 - \lambda_p)\hat{r}(E_j) + \lambda_p \hat{r}(S_j)$, where λ_p is the rank-fusion weight, $\hat{r}(\cdot)$ denotes the descending normalized rank, and s_j is the final reliability score of candidate j.

The key here is not any single statistic itself, but the use of multiple instances to jointly estimate the relative reliability among candidate carriers. If this ranking can be transferred from seen instances to unseen instances, then the population ranking can be further converted into a learnable carrier prior without re-searching independently for each new asset.Based on s_j, we first select highly reliable and low-correlation candidate directions and divide them into K bit groups. For the k-th group G_k,

its initial carrier is

$$b_k^0 = \frac{\sum_{j \in G_k} \phi_j}{\left\| \sum_{j \in G_k} \phi_j \right\|_2} \tag{3}$$

where b_k^0 is the initial carrier of the k-th bit and G_k is the k-th bit group. This yields the initial carrier matrix $B^0 = [b_1^0, \ldots, b_K^0]$. Furthermore, we define the population-stable carrier prior as

$$P = (B, \rho^\star), \quad B = [b_1, \ldots, b_K] \in \mathbb{R}^{N \times K}, \quad \rho^\star = [\rho_1^\star, \ldots, \rho_K^\star]^\top,$$

where B is the canonical carrier matrix (initially B^0), $\rho^\star$ is the bit-strength prior vector, and $\rho_k^\star$ is the strength prior of the k-th bit. However, we do not fix B^0 as the final carrier. Furthermore, based on the selected candidate directions, we define a support matrix $M \in \{0,1\}^{N \times K}$, where $M_{n,k} = 1$ indicates that the n-th Gaussian is allowed to participate in the carrier learning of the k-th bit, and n is the Gaussian index. The carrier parameters are then expressed as $B = M \odot (B^0 + \Delta B)$, where $\odot$ is the element-wise product and ΔB is the carrier residual learned during training. This structure explicitly decouples population knowledge from instance learning: M retains the reliable spatial support already verified at the population level, B^0 provides initialization from reliable directions, and ΔB further corrects the carrier using training instances. In this way, the model neither degenerates back to a completely free random carrier search nor is limited by a fixed prior. To reduce coupling among different bits, we further introduce a carrier orthogonality constraint $L_{\text{orth}} = \|B^\top B - I\|_F^2$, where I is the identity matrix and $\| \cdot \|_F$ is the Frobenius norm.

3.2 Instance-Adaptive Embedding and Physical-Channel-Aware Training

Given the learned canonical carrier matrix B, we map the carrier to each specific subject through instance registration. Let Π_i denote the row-alignment operator between instance i and the canonical space, with at most one nonzero entry per row; the instance-level carrier is written as $B_i = \Pi_i B$. Thus, different instances share the same spatial support and carrier structure, but the specific Gaussians carrying the watermark vary with the subject content, and new instances do not need to research the carrier. Considering the visibility of different Gaussian attributes during propagation, we embed the watermark only in Gaussian scale. Position, rotation, color, and opacity remain unchanged: $\Delta\mu = \Delta q = \Delta c = \Delta\alpha = 0$. For the k-th bit, let its embedding strength be ρ_k, initialized with $\rho_k^\star$ obtained from the population response. The instance-level scale modulation is defined as

$$\Delta s_i = \Pi_{[-\delta,\delta]} \left(B_i \left(w \odot \rho \right) \right) \tag{4}$$

where Δs_i is the scale perturbation of instance i, $\Pi_{[-\delta,\delta]}(\cdot)$ is the element-wise clipping operator to $[-\delta, \delta]$, δ is the clipping threshold, $\rho = [\rho_1, \ldots, \rho_K]^\top$ is the learnable bit-strength vector, and ρ_k is the embedding strength of the k-th bit. The Gaussian is updated in the log-scale space: $s_i' = \exp(\log s_i + \Delta s_i)$, where $\log$ is the natural logarithm, $\exp$ is the exponential function, and s_i' is the updated scale. The logarithmic parameterization guarantees that the scale is always positive and makes the embedding appear as a relative scale change; the perturbation clipping provides an explicit amplitude constraint, allowing control between watermark quality and detection performance.

After embedding, the watermarked 3DGS is trained through the real propagation process. We model the complete print-and-capture chain as $T_\theta = N \circ J \circ L_{\text{illum}} \circ K \circ H \circ A$, where A denotes capture geometry variation, H denotes printer gamut mapping and halftoning, K denotes camera optical blur, L_{illum} denotes illumination and exposure variation, J denotes JPEG compression, and N denotes sensor noise. For the hard threshold in halftoning, we adopt a straight-through estimator so that the forward pass remains close to discrete printing behavior while allowing gradients to back-propagate.

To cover variations in real capture, propagation conditions are randomly sampled from $\theta \sim p(\theta)$, and the physically robust detection objective is constructed as

$$L_{\text{rob}} = \mathbb{E}_{i,w,v,\theta} \left[\ell_{\text{det}} \left(D_\psi \left(\text{DWT}_{LL} \left(T_\theta \left(R_v(G_i') \right) \right) \right), w \right) \right] \tag{5}$$

where L_{rob} is the physically robust detection objective, $\mathbb{E}$ is the expectation, ℓ_{det} is the detection loss, D_ψ is the decoder, and DWT_{LL} is the single-level Haar wavelet low-frequency subband. Training adopts a progressive curriculum: initially dominated by clean or weakly degraded samples, the intensity of printing and capture degradation is gradually increased, while a certain proportion of

clean samples is always retained, so that the carrier first acquires basic detectability and then gradually adapts to complex physical propagation. The purpose of this process is not simply to add more noise to the decoder, but to make the carrier itself directly constrained by the detection objective after physical propagation. Therefore, carrier learning and physical-channel learning are not two independent modules, but are jointly completed under the same watermark recovery objective.

3.3 Reference-Free Blind Decoding and Joint Optimization

After print-and-capture, the receiver can only obtain the propagated image, without the corresponding clean image or original Gaussian parameters. Therefore, PhysMark adopts reference-free blind decoding. Since the scale modulation mainly appears as local, smooth brightness changes after rendering, while halftone dots and sensor noise contain more high-frequency components, we first apply a single-level Haar wavelet transform to the propagated observation y and keep the low-frequency subband $L = \mathrm{DWT}_{LL}(y)$. It is then fed into a multi-scale residual decoder D_ψ to obtain the logits of K bits: $\ell = D_\psi(L)$, where ℓ is the K-bit logit vector, and the recovered message $\hat{w}$ is obtained by a sign operation. Real phone photos are perspectively rectified using the four corner registration marks on the page before decoding. During training, PhysMark jointly optimizes message detection, visual fidelity, carrier orthogonality, and population prior consistency. For instance i, view v, and bit k, let the decoder output be $\ell_{i,v,k}$, and let $V_{v,k}$ indicate whether the bit has valid visibility in the current view. The detection loss is

$$L_{\mathrm{det}} = \frac{1}{K} \sum_{k=1}^{K} \omega_k \frac{\sum_v V_{v,k}\, \ell_{\mathrm{BCE}}(\ell_{i,v,k}, w_k)}{\sum_v V_{v,k} + \varepsilon} \tag{6}$$

where L_{det} is the detection loss, ω_k is a mild inverse weight set by the sliding statistics of the per-bit response, ℓ_{BCE} is the binary cross-entropy, and ε is a small constant. To preserve visual quality, we compute between the watermarked rendering without the physical channel x' and the original rendering x:

$$L_{\mathrm{fid}} = \frac{1}{2}\,\mathrm{MSE}(x', x) + \frac{1}{2}\,(1 - \mathrm{SSIM}(x', x)) \tag{7}$$

where L_{fid} is the fidelity loss, MSE is the mean squared error, and SSIM is the structural similarity. We further limit visual distortion through $L_{\mathrm{psnr}} = [\mathrm{PSNR}^\star - \mathrm{PSNR}(x', x)]_+^2$, where $\mathrm{PSNR}^\star$ is the target PSNR lower bound and $[\cdot]_+$ denotes the positive part. To avoid systematic bit bias in the decoder, a balance constraint is added: $L_{\mathrm{bal}} = \frac{1}{K}\sum_{k=1}^{K}(\bar{p}_k - 0.5)^2$, where $\bar{p}_k$ is the average probability that the k-th bit is predicted as $+1$. Meanwhile, the bit strengths are initialized from the population prior and kept in a reasonable range by a soft constraint: $L_{\mathrm{prior}} = \|\rho - \rho^\star\|_2^2$. The final joint optimization objective is

$$L = \lambda_{\mathrm{det}} L_{\mathrm{det}} + \lambda_{\mathrm{fid}} L_{\mathrm{fid}} + \lambda_{\mathrm{psnr}} L_{\mathrm{psnr}} + \lambda_{\mathrm{orth}} L_{\mathrm{orth}} + \lambda_{\mathrm{bal}} L_{\mathrm{bal}} + \lambda_{\mathrm{prior}} L_{\mathrm{prior}} \tag{8}$$

where $\lambda_{\mathrm{det}}, \lambda_{\mathrm{fid}}, \lambda_{\mathrm{psnr}}, \lambda_{\mathrm{orth}}, \lambda_{\mathrm{bal}}, \lambda_{\mathrm{prior}}$ are the weights of the corresponding losses. Here, B is initialized from the structured carrier obtained from the population prior and continues to be learned through the detection objective and orthogonality constraint; ρ is initialized from the prior strength obtained from the population response and is prevented from deviating significantly by the soft prior constraint. Therefore, PhysMark does not simply copy a fixed watermark carrier from a population template. Instead, it exploits the transferable reliability patterns in the population to provide a structured search space for carrier learning and allows training to further adapt to instance differences. At the same time, by directly incorporating physical propagation into the detection objective, the learned carrier is oriented not toward an ideal digital image, but toward the final physically observable propagation result.

4 Experiments

We evaluate PhysMark on the IXI dataset (IXI Brain Development Dataset, 2013), reconstructed from T2 MRI via Conditional3DGS. The dataset contains 200 subjects, split by a fixed random seed into an 80/20/100 training/validation/test split. Each subject is an independent Conditional3DGS instance with $N = 100{,}000$ Gaussians. Our method trains end-to-end on an A100 GPU. We adopt PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) as fidelity metrics.

Table 1: Main results under simulated printing for 8/16/32-bit watermarking. A: GaussianMarker; B: 3D-GSW; C: GuardSplat. Values are Bitacc (%).

Method	8-bit		16-bit		32-bit	
	PSNR/SSIM	LPIPS/Bitacc	PSNR/SSIM	LPIPS/Bitacc	PSNR/SSIM	LPIPS/Bitacc
A	38.57/0.988	0.013/88.41	36.30/0.985	0.017/85.31	30.16/0.964	0.034/75.11
B	36.35/0.985	0.015/89.37	33.61/0.981	0.022/87.16	31.95/0.972	0.027/76.22
C	36.39/0.986	0.016/89.50	34.48/0.983	0.020/87.05	32.60/0.976	0.024/76.16
Ours	**42.62/0.999**	**0.001/93.29**	**39.79/0.998**	**0.002/88.54**	**35.95/0.995**	**0.007/77.82**

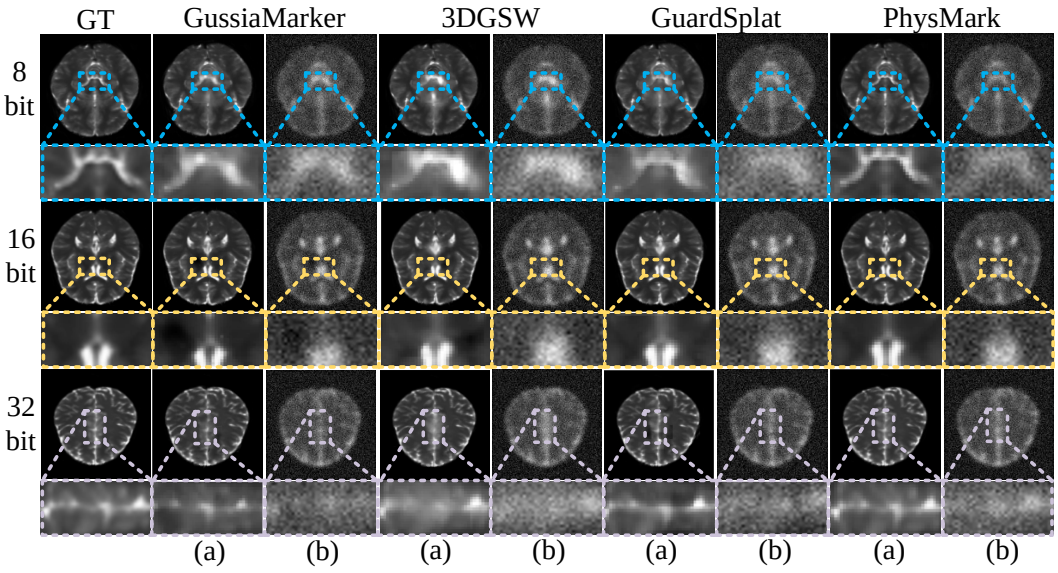

Figure 2: Qualitative comparison under simulated printing for 8-Bit, 16-Bit, and 32-bit watermarking. GT: Ground Truth. (a) Watermarked rendering. (b) Simulated print-and-capture.

4.1 Main Results

The baselines GaussianMarker(Huang et al., 2024), 3D-GSW(Jang et al., 2025), and GuardSplat(Chen et al., 2025) are single-scene methods that cannot generalize across instances. For fair comparison, each baseline is independently trained and tested on each of the 100 test subjects (i.e., 100 separate training runs), using the same data split, rendering resolution, message length, and differentiable physical channel. Figure 2 and Table 1 compare imperceptibility and message recovery. At 8 bits, PhysMark achieves 42.62 dB PSNR and 93.29 Bitacc, outperforming GaussianMarker (38.57 dB/88.41); at 16 bits, PhysMark reaches 39.79 dB/88.54, surpassing GuardSplat (34.48 dB/87.05); at 32 bits, PhysMark attains 35.95 dB/77.82, still outperforming 3D-GSW (31.95 dB/76.22). PhysMark estimates the relative reliability of carriers from the population, selects stable, physically robust, and mutually orthogonal directions, and refines them with a hard-anchor prior, learnable residuals, and orthogonality constraints, thereby yielding lower distortion and more stable recovery on unseen subjects and under print-and-capture conditions.

4.2 Robustness to 2D and 3D Attacks

Table 2 shows that our method achieves relatively high Bitacc under all five attacks and generally outperforms the compared baselines. Under 2D attacks, i.e., JPEG50, Gaussian noise ($\sigma = 0.1$), and Gaussian blur ($\sigma = 0.1$), our method obtains 89.36%, 88.19%, and 89.42%, respectively, while GaussianMarker, 3D-GSW, and GuardSplat reach 80.99%/79.35%/80.99%, 81.54%/79.56%/81.60%, and 81.62%/79.56%/81.62%, respectively. This may be attributed to the fact that the physical channel already includes JPEG compression, additive noise, and blur, allowing

Table 2: Quantitative robustness evaluation of 16-bit watermarking under various attacks.

Method	JPEG50	Noise (0.1)	Blur (0.1)	3D Noise (0.1)	3D Prune (0.2)
GaussianMarker	80.99	79.35	80.99	84.67	81.94
3D-GSW	81.54	79.56	81.60	84.97	82.21
GuardSplat	81.62	79.56	81.62	85.05	82.38
Ours	**89.36**	**88.19**	**89.42**	**89.19**	**90.69**

Table 3: Gaussian attribute ablation under 16-bit watermarking. Values are Bitacc (%).

Carrier attribute	PSNR	SSIM	LPIPS	Bitacc
Position	37.80	0.996	0.005	77.49
Rotation	36.28	0.995	0.007	79.70
Opacity	34.44	0.992	0.011	83.30
Color	35.00	0.993	0.009	77.59
Ours	**39.79**	**0.998**	**0.002**	**88.54**

Table 4: Results of core component ablation under 16-bit watermarking. Values: Bitacc (%).

Configuration	PSNR	SSIM	LPIPS	Bitacc
w/o hard anchor	37.71	0.996	0.004	72.71
w/o physical channel	**41.93**	**0.999**	**0.002**	65.40
Ours	39.79	0.998	**0.002**	**88.54**

print-and-capture training to transfer a certain degree of robustness to common 2D digital attacks without per-attack fine-tuning. Under 3D attacks, i.e., 3D parameter noise ($\sigma = 0.1$) and 3D random pruning (20%), our method achieves 89.19% and 90.69%, respectively, while the best baseline GuardSplat obtains 85.05% and 82.38%. The former may be related to the explicit constraint on scale deviation by log-scale clipping, and the latter may benefit from localized carriers that confine the energy of each bit to a local region. Overall, our method shows relatively better robustness in both the digital domain and the 3D parameter domain.

4.3 Ablation Studies

Anchor Generalization Under the 8-bit protocol, single-subject anchors generalize poorly: the cross-subject rank correlation of stability is only 0.81, and the partial correlation of reliability is 0.68–0.77. Population-aggregated selection reduces the cross-subject standard deviation from 0.033 to 0.028. However, static population anchors are still insufficient: without a prior, physical decoding yields a cross-subject standard deviation of 0.099 and only 32.86 dB PSNR. Treating the population prior as a learnable inductive bias and fine-tuning under orthogonality constraints reduces the standard deviation to 0.055 and raises PSNR to 42.62 dB. Thus, the population prior provides a stable starting point, while learnable carriers compensate for individual differences.

Carrier Attribute and Component Ablation Under a unified perturbation budget and the same 16-bit protocol, Table 3 compare five Gaussian attributes as carriers. Scale (log-scale) achieves the best overall performance (PSNR 39.79 dB, SSIM 0.998, LPIPS 0.002, Bitacc 88.54%). Position and rotation obtain higher PSNR but lower Bitacc (77.49% and 79.70%), while color/SH and opacity are worse. This confirms that scale perturbations appear as local low-frequency brightness changes that are smooth-invariant to blur and halftone dots. For component contribution, Table 4 compares the full method with two variants. The full method achieves 88.54% Bitacc, higher than w/o hard anchor (72.71%) and w/o physical channel (65.40%). Without the hard anchor, fidelity is acceptable but accuracy drops; without the physical channel, fidelity is highest but Bitacc is lowest, showing that digital-only optimization is insufficient for physical propagation. The full method balances fidelity and robustness.

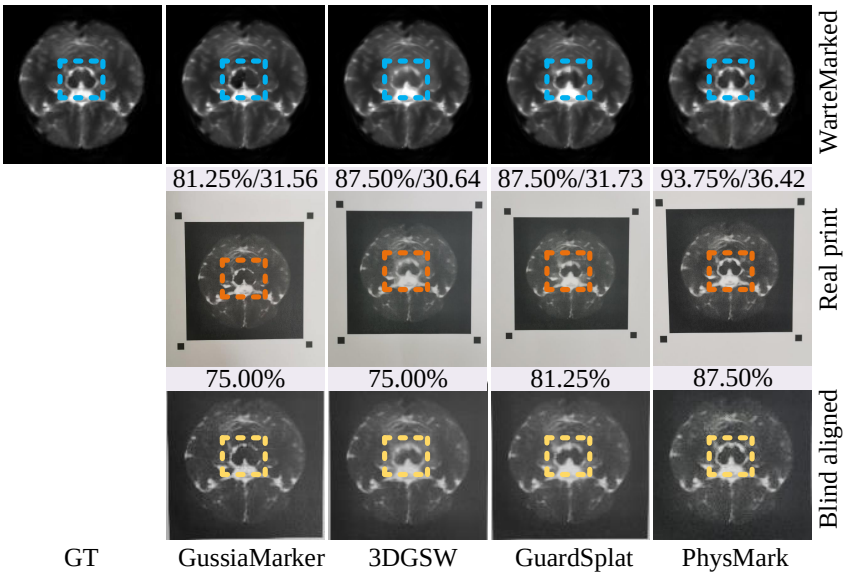

Figure 3: Real print-and-capture comparison. GT: Ground Truth. The three rows show watermarked digital renderings, real printed photos, and blind-aligned decoding results, respectively.

4.4 Real Print-and-Capture Validation

At 16 bits, Figure 3 compares our method with three baselines under real print-and-capture. The three rows show watermarked digital renderings, real printed photos, and blind-aligned decoding results; numbers report Bitacc and PSNR (watermarked row) or Bitacc only (other rows). In watermarked renderings, PhysMark achieves 93.75% Bitacc at 36.42 dB PSNR, outperforming GaussianMarker (81.25%, 31.56 dB), 3D-GSW (87.50%, 30.64 dB), and GuardSplat (87.50%, 31.73 dB). After real printing and blind alignment, PhysMark obtains 87.50% Bitacc (2/16 errors), GaussianMarker 75.00% (4/16), 3D-GSW 75.00% (4/16), and GuardSplat 81.25% (3/16). Notably, our method is evaluated on an unseen subject with a single photo, whereas baselines are trained and tested on this specific scene; despite this, PhysMark still achieves the best performance, demonstrating its generalization. This experiment validates the end-to-end physical pipeline and provides a calibration reference for the simulated channel. Real and simulated results show similar error patterns, suggesting the differentiable channel captures key aspects of real degradation. Large-scale quantitative evaluation is conducted under the simulated channel; extending real capture to more printers, phones, and lighting is left for future work. 3D-GSW and GuardSplat are also competitive but require per-scene training, whereas our method generalizes to unseen subjects without retraining..

5 Conclusion

Existing 3DGS watermarking is mostly limited to a single asset, overlooking cross-instance response fluctuations and print-and-capture decay. PhysMark learns a stable carrier prior from population dual-channel responses, selects scale-domain carriers that are strong, stable, and near-orthogonal across instances, and maps them as a learnable inductive bias to each instance. Combined with log-scale embedding and DWT low-frequency blind decoding, it achieves reference-free recovery on unseen subjects and adapts to physical propagation via a differentiable print-and-capture channel and EOT. Experiments show that 3DGS provenance under real-world dissemination depends on digital robustness, cross-instance carrier stability, and physical propagation modeling. Limitations: the differentiable channel cannot cover all variations in paper, printers, illumination, and cameras; under high capacity, a sim-to-real gap and systematic bit errors persist, and fidelity and worst-bit accuracy need improvement. Future work will expand real capture, strengthen weak-bit and worst-case optimization, refine the imaging model, and evaluate stronger attacks.

AI USE STATEMENT

We used a large language model (LLM) for language polishing and manuscript editing, including sentence rewriting, grammar checking, and improving fluency. The LLM did not participate in ideation, research design, experiments, or data analysis, and was not used for other tasks requiring disclosure. All scientific content was independently developed by the authors, who take full responsibility for the paper and ensure that AI-assisted text complies with ethical guidelines.

ETHICS STATEMENT

We have read and acknowledge the ICLR Code of Ethics and comply with its guidelines. This study uses the publicly released, fully de-identified dataset. The work does not involve human-subject intervention or new data collection. We believe the presented work poses no identifiable ethical risk.

REPRODUCIBILITY STATEMENT

We have made every effort to ensure the reproducibility of the results presented in this paper. The experimental settings are described in detail in the paper to facilitate replication.

REFERENCES

Orhan Bulan, Gaurav Sharma, and Vishal Monga. Orientation modulation for data hiding in clustered-dot halftone prints. *IEEE Transactions on Image Processing*, 19(8):2070–2084, 2010. doi: 10.1109/TIP.2010.2046795.

Mingshu Cai, Jiajun Li, Osamu Yoshie, Yuya Ieiri, and Yixuan Li. Where, what, why: Toward explainable 3d-gs watermarking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20701–20710, June 2026.

Changsheng Chen and Xijin Li. Robust image watermarking scheme under halftone distortion with surrogate model. In *Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1–6, 2024.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC: Hash-grid assisted context for 3d gaussian splatting compression. In *European Conference on Computer Vision (ECCV)*, pp. 422–438, 2024.

Yihang Chen, Mengyao Li, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. PCGS: Progressive compression of 3d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 3111–3119, 2026. doi: 10.1609/aaai.v40i4.37304.

Zixuan Chen, Guangcong Wang, Jiahao Zhu, Jianhuang Lai, and Xiaohua Xie. GuardSplat: Efficient and robust watermarking for 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16325–16335, 2025. doi: 10.1109/CVPR52734.2025.01522.

Arnaud Delmotte, Kenichiro Tanaka, Hiroyuki Kubo, Takuya Funatomi, and Yasuhiro Mukaigawa. Blind watermarking for 3-d printed objects by locally modifying layer thickness. *IEEE Transactions on Multimedia*, 22(11):2780–2791, 2020. doi: 10.1109/TMM.2019.2962306.

Arnaud Delmotte, Kenichiro Tanaka, Hiroyuki Kubo, Takuya Funatomi, and Yasuhiro Mukaigawa. Blind 3D-printing watermarking using moment alignment and surface norm distribution. *IEEE Transactions on Multimedia*, 23:3467–3482, 2021. doi: 10.1109/TMM.2020.3025660.

Yangfan Deng, Anirudh Nakra, and Min Wu. Position: 3d gaussian splatting watermarking should be scenario-driven and threat-model explicit. *arXiv preprint arXiv:2602.02602*, 2026.

Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. LightGaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 140138–140158, 2024.

Han Fang, Kejiang Chen, Yupeng Qiu, Jiayang Liu, Ke Xu, Chengfang Fang, Weiming Zhang, and Ee-Chien Chang. DeNoL: A few-shot-sample-based decoupling noise layer for cross-channel watermarking robustness. In *Proceedings of the 31st ACM International Conference on Multimedia (MM)*, pp. 7345–7353, 2023. doi: 10.1145/3581783.3612068.

Pierre Fernandez, Guillaume Couairon, Hervé Jégou, Matthijs Douze, and Teddy Furon. The stable signature: Rooting watermarks in latent diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22409–22420, 2023.

Daidou Guo and Chuan Qin. PCMark-NAS: Lightweight print-camera resilient watermarking networks via neural architecture search. In *ACM Multimedia Asia (MMASIA)*, pp. 1–5, 2024. doi: 10.1145/3696409.3700284.

Xin Huang, Renjie Li, Yiu-Ming Cheung, Ka Chun Cheung, Simon See, and Renjie Wan. GaussianMarker: Uncertainty-aware copyright protection of 3d gaussian splatting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 33037–33060, 2024.

Xiufeng Huang, Ziyuan Luo, Qi Song, Ruofei Wang, and Renjie Wan. MarkSplatter: Generalizable watermarking for 3d gaussian splatting model via splatter image structure. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM)*, pp. 12189–12198, 2025.

Sumin In, Youngdong Jang, Utae Jeong, MinHyuk Jang, Hyeongcheol Park, Eunbyung Park, and Sangpil Kim. CompMarkGS: Robust watermarking for compressed 3d gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2026.

IXI Brain Development Dataset. Information extraction from images (ixi) dataset. *Brain Development*, 2013. URL `https://brain-development.org/ixi-dataset/`.

Youngdong Jang, Dong In Lee, MinHyuk Jang, Jong Wook Kim, Feng Yang, and Sangpil Kim. WateRF: Robust watermarks in radiance fields for protection of copyrights. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12087–12097, 2024.

Youngdong Jang, Hyunje Park, Feng Yang, Heeju Ko, Euijin Choo, and Sangpil Kim. 3D-GSW: 3d gaussian splatting for robust watermarking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5938–5948, 2025. doi: 10.1109/CVPR52734.2025.0055.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023. doi: 10.1145/3592433.

Jaejin Lee, Minjae Jeong, Joonhyuk Park, Yechan Hwang, Seunghun Baek, and Won Hwa Kim. Mark4D: Temporally-consistent watermarking for 4d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 39393–39402, 2026.

Chenxin Li, Brandon Y. Feng, Zhiwen Fan, Panwang Pan, and Zhangyang Wang. StegaNeRF: Embedding invisible information within neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 441–453, 2023.

Chenxin Li, Hengyu Liu, Zhiwen Fan, Wuyang Li, Yifan Liu, Panwang Pan, and Yixuan Yuan. InstantSplamp: Fast and generalizable steganography framework for generative gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2025a.

Lijian Li, Jinglu Wang, Xiang Ming, and Yan Lu. GS-Marker: Generalizable and robust watermarking for 3d gaussian splatting. *arXiv preprint arXiv:2503.18718*, 2025b.

Nan Li, Yike Zeng, Qian Zhang, Qi Zhang, and Zhiyi Pan. Write where it matters: Policy-guided watermarks for 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6580–6590, 2026.

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20654–20664, 2024.

Ziyuan Luo, Anderson Rocha, Boxin Shi, Qing Guo, Haoliang Li, and Renjie Wan. The NeRF signature: Codebook-aided watermarking for neural radiance fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4652–4667, 2025. doi: 10.1109/TPAMI.2025.3550166.

M. Narendra, M. L. Valarmathi, L. Jani Anbarasi, and Amir H. Gandomi. Levenberg–marquardt deep neural watermarking for 3D mesh using nearest centroid salient point learning. *Scientific Reports*, 14(1):6942, 2024. doi: 10.1038/s41598-024-57360-z.

Chuan Qin, Xiaomeng Li, Zhenyi Zhang, Fengyong Li, Xinpeng Zhang, and Guorui Feng. Print-camera resistant image watermarking with deep noise simulation and constrained learning. *IEEE Transactions on Multimedia*, 26:2164–2177, 2023. doi: 10.1109/TMM.2023.3293272.

Hao Qin, Yukai Sun, Luyuan Chen, Mengxu Lu, Feng Zhang, Ming Kong, Zhenhong Du, and Qiang Zhu. NGS-Marker: Robust native watermarking for 3d gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2026.

Matthew Tancik, Ben Mildenhall, and Ren Ng. StegaStamp: Invisible hyperlinks in physical photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2114–2123, 2020.

Boyu Wang, Jun Xia, and Mingsong Chen. Robust3DGSW: Toward robust watermarking for quantization-aware 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19076–19084, 2026.

Yufei Wang, Zhihao Li, Lanqing Guo, Wenhan Yang, Alex C. Kot, and Bihan Wen. ContextGS: Compact 3d gaussian splatting with anchor level context model. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 51532–51551, 2024.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Yuxin Wen, John Kirchenbauer, Jonas Geiping, and Tom Goldstein. Tree-Rings Watermarks: Invisible fingerprints for diffusion images. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 58047–58063, 2023.

Xiao Wu, Hu Guan, Ying Huang, Chunhua Song, Baoning Niu, Shuwu Zhang, and Jie Liu. A comprehensive review of three-dimensional watermarking algorithms. In *Proceedings of the 2024 10th International Conference on Communication and Information Processing (ICCIP)*, pp. 469–475, 2024. doi: 10.1145/3708657.3708734.

Yufeng Wu, Xin Liao, Baowei Wang, Han Fang, Xiaoshuai Wu, Mingyue Chen, and Guiling Wang. Sim-to-real: An unsupervised noise layer for screen-camera watermarking robustness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1303–1310, 2026. doi: 10.1609/aaai.v40i2.37103.

Rui Xu, Mengya Hu, Deren Lei, Yaxi Li, David Lowe, Alex Gorevski, Mingyu Wang, Emily Ching, and Alex Deng. InvisMark: Invisible and robust watermarking for AI-generated image provenance. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 909–918, 2025.

Fei Yan, Zeqian Wang, and Kaoru Hirota. Dual medical image watermarking using SRU-enhanced network and EICC chaotic map. *Complex & Intelligent Systems*, 11(1):101, 2025. doi: 10.1007/s40747-024-01723-6.

Yao Yao, Bo Ouyang, and Cancan Zhao. CG-3DGS: Complexity-guided 3d gaussian splatting for high-fidelity surgical scene reconstruction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 17541–17547, 2025. doi: 10.1109/IROS60139.2025.11246213.

Jee Seok Yoon, Kwanseok Oh, Yooseung Shin, Maciej A. Mazurowski, and Heung-Il Suk. Domain generalization for medical image analysis: A review. *Proceedings of the IEEE*, 112(10):1583–1609, 2024. doi: 10.1109/JPROC.2024.3486204.

Seokhyun Youn, Soohyun Lee, Geonho Kim, et al. SUCCESS-GS: Survey of compactness and compression for efficient static and dynamic gaussian splatting. *arXiv preprint arXiv:2512.07197*, 2025.

Long Yu, Xiamu Niu, and Shenghe Sun. Print-and-scan model and the watermarking countermeasure. *Image and Vision Computing*, 23(9):807–814, 2005.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-Splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456, 2024.

Qingyuan Zeng, Shu Jiang, Jie Lin, Zhen Wang, Kay Chen Tan, and Min Jiang. Fading the digital ink: A universal black-box attack framework for 3dgs watermarking systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 38084–38092, 2026. doi: 10.1609/aaai.v40i44.41146.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.

Xuanyu Zhang, Jiarui Meng, Runyi Li, Zhipei Xu, Yongbing Zhang, and Jian Zhang. GS-Hider: Hiding messages into 3d gaussian splatting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 49780–49805, 2024.

Xuanyu Zhang, Jiarui Meng, Zhipei Xu, Shuzhou Yang, Yanmin Wu, Ronggang Wang, and Jian Zhang. SecureGS: Boosting the security and fidelity of 3d gaussian splatting steganography. In *International Conference on Learning Representations (ICLR)*, 2025.

Longjie Zhao, Ziming Hong, Jiaxin Huang, Runnan Chen, Mingming Gong, and Tongliang Liu. Intellectual property protection for 3d gaussian splatting assets: A survey. *arXiv preprint arXiv:2602.03878*, 2026.

Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei. HiDDeN: Hiding data with deep networks. In *European Conference on Computer Vision (ECCV)*, pp. 657–672, 2018.