MantleMark: Migrating Watermarks from Multi-View Images to Radiance Fields via Frequency Modulation

Ziyuan Luo, *Graduate Student Member*, *IEEE*, Jun Liu, *Senior Member*, *IEEE*, Haoliang Li, *Member*, *IEEE*, Anderson Rocha, *Fellow*, *IEEE*, Renjie Wan, *Member*, *IEEE*

***Abstract*—Multi-view images are essential for modern radiance field reconstruction methods like Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS). While image watermarking is a crucial data protection and ownership verification technique, it faces unprecedented challenges in multi-view scenarios. Traditional 2D watermarking techniques often fail to maintain detectability in rendered views, while existing 3D watermarking methods are typically limited to specific reconstruction methods and require access to the reconstruction process. To address these limitations, we propose MantleMark, a watermarking framework that migrates watermarks from multi-view images to radiance fields via frequency modulation. Our key insight is constructing a mantle-like Frequency-domain Watermarking Representation in 3D frequency space, which can be projected to create view-dependent watermarking patterns. Relying upon the Fourier Projection-Slice Theorem, we embed these patterns through magnitude spectrum modulation in the image frequency domain, enabling watermarks to migrate into 3D representations. This approach ensures watermark detectability in rendered views regardless of the reconstruction methods used by adversaries. Extensive experiments demonstrate that our method achieves robust watermark detection while maintaining high visual quality across various radiance field-based reconstruction methods.**

***Index Terms*—3DGS, Neural Radiance Fields, Watermarking, Fourier Projection-Slice Theorem.**

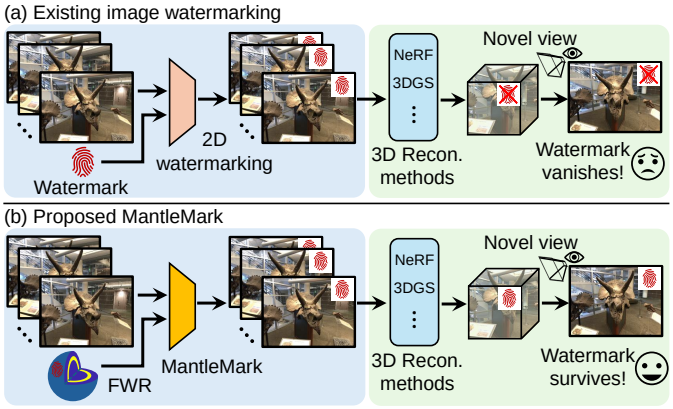

Fig. 1. Challenges in watermark migration from multi-view images to radiance fields. (a) Traditional image watermarks become undetectable in novel views rendered from reconstructed radiance fields. (b) Our MantleMark framework, leveraging a Frequency-domain Watermarking Representation (FWR), successfully preserves watermarks through the reconstruction process while maintaining high visual quality.

I. Introduction

Modern radiance field frameworks such as Neural Radiance Fields (NeRF) [1] and 3D Gaussian Splatting (3DGS) [2] can construct detailed 3D representations from multi-view image collections, enabling novel viewpoint rendering with impressive fidelity. With the widespread availability of these reconstruction tools and the growing prevalence of multi-view imaging datasets, a critical and urgent security threat has emerged. Adversaries can easily collect multi-view images from online sources, reconstruct high-quality 3D representations, and generate unlimited novel views for unauthorized use across diverse applications [3], [4]. This threat is particularly severe as image owners currently lack effective approaches to verify the unauthorized usage of those images rendered from the emerging 3D representations [5]. While traditional image watermarking has been widely adopted for ownership verification in 2D scenarios, directly applying these techniques to protect multi-view image collections is insufficient. The embedded watermarks often become undetectable in rendered views, compromising existing watermarking systems when applied to multi-view imaging scenarios.

This vulnerability arises from a fundamental mismatch between traditional watermarking techniques [6]–[8] and radiance field-based reconstruction [1], [2]. Traditional watermarks are designed to survive 2D image transformations but are not equipped to persist through the dimension transition from multi-view images to 3D representations. During reconstruction, radiance fields optimize for scene consistency across multiple views, which removes any image-level perturbations that do not align with the underlying 3D structure. As shown in Figure 1(a), this optimization process inevitably removes the embedded watermarks, making them undetectable in rendered views and allowing adversaries to obtain watermark-free versions of the protected content. This creates a unique challenge: the 2D watermarks used to protect images should remain intact

This work was carried out at the Renjie Group, Hong Kong Baptist University. Renjie Group is supported by the National Natural Science Foundation of China under Grant No. 62302415, Guangdong Basic and Applied Basic Research Foundation under Grant No. 2024A1515012822, and the HK RGC under a GRF Grant 12203124 and an ECS Grant 22201125. We also thank the support of the São Paulo Research Foundation (Fapesp) through the Horus Project #2023/12865-8 and the Brazilian National Council for Scientific and Technological Research (CNPq) through complementary grants. Haoliang Li is supported by the ECS Grant of HK RGC under Grant 21200522. *(Corresponding author: Renjie Wan.)*

Ziyuan Luo and Renjie Wan are with the Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China (e-mail: ziyuanluo@life.hkbu.edu.hk; renjiewan@hkbu.edu.hk).

Jun Liu is with School of Computing and Communications, Lancaster University, U.K. (e-mail: j.liu81@lancaster.ac.uk)

Haoliang Li is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China (e-mail: haoliang.li@cityu.edu.hk).

Anderson Rocha is with Artificial Intelligence Lab., Recod.ai, Institute of Computing, University of Campinas, Brazil (e-mail: arrocha@unicamp.br).

The source code is available at: https://github.com/luo-ziyuan/MantleMark.

during different radiance field-based reconstructions.

Another key challenge comes from the diversity of radiance field representations. Various methods, like NeRF [1] and 3DGS [2], employ distinct architectures and optimization strategies to represent 3D scenes. This architectural diversity makes it difficult to design a universal protection scheme that works across different radiance field models. While several 3D watermarking techniques have been proposed for radiance fields [9]–[12], they focus on protecting the radiance field models rather than the underlying images. Furthermore, these approaches are inherently model-specific, as watermarking techniques designed for NeRF [1] cannot be applied to 3DGS [2], and vice versa. In addition, these methods require direct access to the reconstruction process, making them impractical for realistic scenarios where content owners distribute their multi-view image collections publicly but cannot control how adversaries might process them [3], [4], [13].

Therefore, protecting multi-view images against radiance field reconstruction poses two fundamental requirements: **cross-dimensional persistence** and **reconstruction independence**. Cross-dimensional persistence requires watermarks to survive different complicated multi-view optimization process, while reconstruction independence demands effectiveness regardless of the underlying radiance field technique employed by adversaries.

To meet the above requirements, our key insight is that watermarks embedded in the frequency domain can achieve superior robustness and cross-dimensional persistence. The frequency domain provides unique advantages that address both fundamental challenges in radiance field-based watermarking. **First**, the frequency domain bridges cross-dimensional persistence through the Fourier Projection-Slice Theorem [14], [15]. Unlike spatial perturbations treated as noise during optimization, frequency-domain patterns are intrinsically linked to the scene's structure, as a 2D projection of a 3D object naturally corresponds to a slice of its 3D frequency spectrum. This fundamental property ensures that our watermarks are embedded as an inherent part of the scene representation rather than removable noise, thereby persisting through the reconstruction process. **Second**, since volumetric rendering in various radiance field methods essentially performs projection and integration operations, frequency-domain patterns maintain their characteristics through these operations. This enables our watermarks to remain effective across different radiance field models, as the frequency response is preserved regardless of the specific radiance field technique used. Moreover, frequency-domain watermarking operates purely in the image space, allowing us to embed watermarks without accessing or modifying reconstruction methods.

Based on these insights, we propose a frequency-domain watermarking framework enabling cross-dimensional migration from multi-view images to radiance fields, as shown in Figure 1(b). As illustrated in Figure 2, we first generate a Frequency-domain Watermarking Representation (FWR) with an annular mask centered at the zero frequency point. View-dependent watermarking patterns are created according to each image's pose and embedded through inverse Fourier transform. Through the Fourier Projection-Slice Theorem, these frequency patterns establish a mapping between 2D image projections and 3D frequency representations, allowing the watermarks to propagate into the reconstructed radiance fields during reconstruction optimization. Our framework determines ownership by detecting the presence of these specific frequency patterns. Since this detection relies on fundamental frequency-domain relationships rather than specific model architectures, our watermarks remain detectable in rendered views from radiance fields reconstructed by different methods.

Our primary contributions can be summarized as follows:

- We identify and formalize the problem of watermark migration from multi-view images to radiance fields, revealing the challenges in preserving watermarks during the reconstruction from 2D images to 3D radiance fields.
- We propose a frequency-domain watermarking framework that operates purely in the image domain while remaining effective across different radiance field models like NeRF [1] and 3DGS [2], without requiring access to the reconstruction process.
- We introduce a Frequency-domain Watermarking Representation (FWR) based on the Fourier Projection-Slice Theorem [14], [15], enabling watermarks to propagate from source images to novel rendered views through view-dependent embedding.

Our Frequency-domain Watermarking Representation forms an annular pattern similar to the Earth's mantle, so we name our framework **MantleMark**. Experiments demonstrate that our method achieves robust watermark detection while maintaining high visual quality across various radiance field-based reconstruction methods.

II. Related work

A. Radiance fields

The introduction of Neural Radiance Field (NeRF) [1] revolutionized 3D scene representation by modeling scenes as continuous volumetric fields through implicit neural networks. This optimization-based approach has inspired numerous improvements in various aspects, including rendering quality [16]–[18], generalization [19], [20], dynamic scene modeling [21], [22], and scene editing [23], [24]. However, NeRF's reliance on Multi-Layer Perceptron (MLP) to implicitly encode scene properties leads to computational inefficiency, as rendering requires multiple network queries along each ray. While various methods have attempted to accelerate optimization and inference [25], [26], they often compromise rendering quality. The emergence of 3D Gaussian Splatting (3DGS) [2] represents a significant advancement in radiance field representation, achieving both rapid optimization and high-quality real-time rendering. Recent works have further enhanced 3DGS in aspects such as sparse view reconstruction [27], [28], quality improvement [29], [30], and model efficiency [31], [32]. As radiance fields become increasingly prevalent in 3D content creation, protecting the copyright of multi-view source images used in their reconstruction has emerged as a critical challenge.

B. 2D digital watermarking

2D digital watermarking plays a crucial role in ensuring image verification, authenticity verification, and content traceability. Early approaches to 2D watermarking focused on embedding information by modifying the least significant bits of pixel values [33], [34]. Subsequently, researchers explored data embedding techniques in various transform domains [35]–[37]. The emergence of deep learning has revolutionized the field of image information hiding [38]–[41]. A pioneering breakthrough came with HiDDeN [38], which leveraged deep encoder-decoder architectures to surpass traditional methods in performance metrics. Building on this foundation, UDH [42] introduced a universal deep hiding framework for cover-independent embedding. The research focus subsequently shifted toward enhancing watermark resilience against various forms of distortion [43]–[46], addressing challenges such as JPEG compression [6], screen recapture scenarios [47]–[52], and multiple combined distortions [53]. Beyond the conventional encoder-decoder paradigm, researchers have explored invertible network architectures [8], [54]–[56], GAN-based methods [57], and approaches leveraging frequency analysis and diffusion models [58], [59]. Recent advances have also focused on high-resolution AI-generated images [60] and localized watermarking through image segmentation [61]. However, these methods are designed solely for the 2D domain and cannot address the migration of watermarks from images to 3D representations, making them inadequate for protecting content against radiance field-based reconstruction.

C. 3D digital watermarking

Traditional 3D watermarking methods primarily focus on mesh representations, where information is embedded by modifying vertex coordinates or mesh properties [62]–[67]. Recent advances in 3D watermarking focus on embedding information directly into 3D radiance fields and extracting them from 2D rendered views [10], [11], [68]. With the emergence of neural radiance fields, CopyRNeRF [10] proposes the first approach to incorporate watermarking into NeRF models. Meanwhile, StegaNeRF [9] presents a novel method to embed steganographic information within neural radiance fields. WateRF [11] and NeRFProtector [69] further advance the field by introducing a fine-tuning strategy for NeRF watermarking. Recently, watermarking techniques for 3D Gaussian Splatting have also emerged [12], [70]–[74]. However, a significant limitation of these approaches is their requirement for direct intervention in the 3D reconstruction process, and they are restricted to specific models and difficult to generalize to others.

III. Background

Radiance fields represent scenes as continuous functions, mapping 3D coordinates and viewing directions to color and density values. These representations are commonly rendered using volumetric techniques for novel view synthesis. Among various radiance field approaches, Neural Radiance Fields (NeRF) [1] and 3D Gaussian Splatting (3DGS) [2] stand out for their effective use of volumetric rendering principles. The fundamental process involves tracing a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ from a camera position $\mathbf{o} \in \mathbb{R}^3$ through each pixel. The resulting pixel color is computed through volume integration:

$$\hat{\mathbf{C}}(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \tag{1}$$

where the accumulated transmittance is defined as $T(t) = \exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds)$, with t_n and t_f defining the scene's depth range. To numerically evaluate this continuous integral, NeRF adopts a discrete sampling strategy. By dividing the integration interval into N_s segments and applying numerical quadrature, the continuous integral is approximated as:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{i=1}^{N_s} T_i(1 - \exp(-\sigma_i\delta_i))\mathbf{c}_i, \tag{2}$$

where $\delta_i = t_{i+1} - t_i$ is the sampling interval and $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j\delta_j)$ represents the accumulated transmittance up to the i-th sample point. The network F_θ is optimized to minimize a photometric loss:

$$\mathcal{L} = \sum_{k=1}^{N} \sum_{\mathbf{r}\in\mathcal{R}_k} \|\hat{\mathbf{C}}(\mathbf{r}) - \mathbf{C}_k(\mathbf{r})\|_2^2, \tag{3}$$

where N is the number of training images, $\mathcal{R}_k$ denotes the set of rays in the k-th training image, and $\mathbf{C}_k(\mathbf{r})$ is the ground truth color for ray $\mathbf{r}$.

In contrast to NeRF [1], which employs a neural network F_θ to optimize the continuous volumetric representation through photometric reconstruction of training images $\{\mathcal{I}_k\}_{k=1}^N$, 3DGS [2] adopts an explicit approach based on 3D Gaussian primitives. Specifically, the rendering equation in 3DGS [2] is formulated as:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{i=1}^{N_G} \mathbf{c}_i\omega_i \prod_{j=1}^{i-1}(1 - \omega_j), \tag{4}$$

where N_G is the number of sample Gaussians, and $\mathbf{c}_i$ denotes the color of the i-th Gaussian. ω_i represents the 2D Gaussian splat weight computed as:

$$\omega_i = o_i \exp(-\frac{1}{2}(x - \mu_i)^T\Sigma_i^{-1}(x - \mu_i)), \tag{5}$$

where o_i denotes the opacity of the i-th Gaussian, μ_i is the projected 3D Gaussian center, and Σ_i is the covariance matrix in screen space. This formulation enables efficient rendering through the rasterization of 3D Gaussians while maintaining high visual quality.

IV. Proposed method

This section presents MantleMark, a novel watermarking method designed to protect multi-view images against unauthorized radiance field reconstruction. The proposed approach addresses two fundamental challenges: cross-dimensional persistence and independence of the reconstruction method. Cross-dimensional persistence ensures that watermarks survive the transition from 2D images to 3D representations during

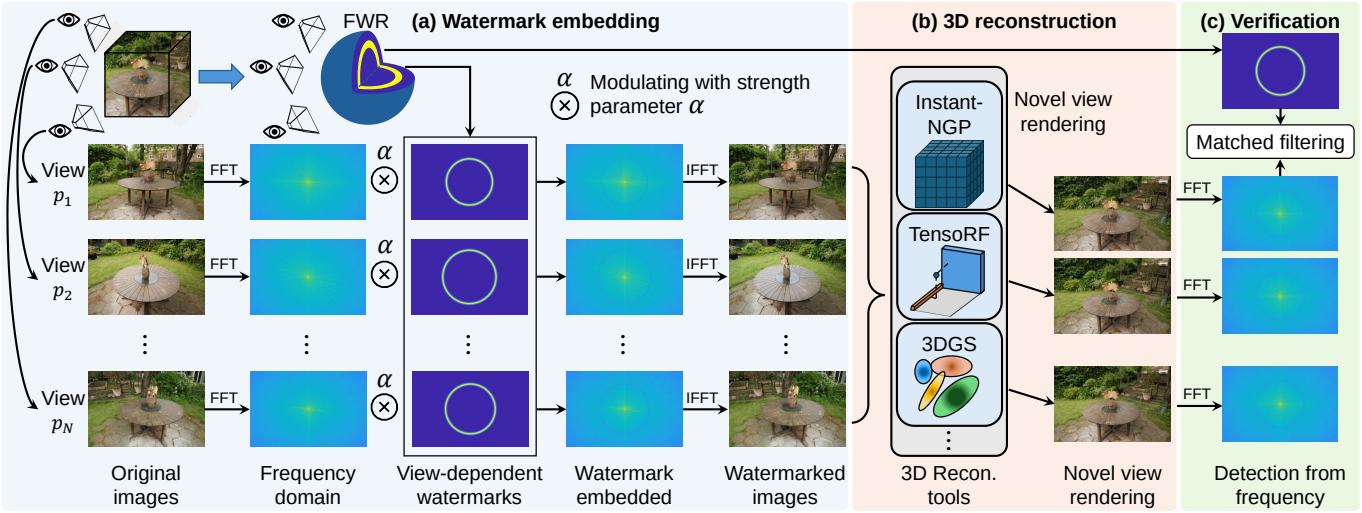

Fig. 2. Overview of the MantleMark framework. **(a)** The process begins with multi-view input images, from which we generate a mantle-like Frequency-domain Watermarking Representation (FWR) utilizing a circular mask in the frequency domain. View-dependent watermarking patterns are created according to camera poses and embedded via frequency modulation. **(b)** The watermarked images retain their visual quality while preserving watermark consistency in 3D space, ensuring robustness even when adversaries attempt novel view synthesis through arbitrary reconstruction methods. **(c)** For watermark verification, we employ a matched filtering approach in the frequency domain to effectively detect the embedded watermarks in synthesized views, providing reliable authentication of content ownership.

multi-view optimization. At the same time, independence allows the watermarks to remain effective regardless of the specific reconstruction methods used by adversaries. We achieve this by embedding watermarks in the frequency domain of the images and ensuring these watermarks propagate to the 3D representation through the Fourier Projection-Slice Theorem [14], [15].

Our framework is illustrated in Figure 2. Given a set of multi-view images to be protected, we first generate a mantle-like Frequency-domain Watermarking Representation (FWR). Then, we create view-dependent watermarking patterns according to each image's pose. These patterns are embedded into the images through a frequency modulation scheme. The embedded watermarks remain detectable in the frequency domain of any synthesized novel views, enabling robust copyright verification regardless of the reconstruction methods employed by adversaries.

A. Problem formalization

Consider a set of N multi-view images $\{\mathcal{I}_i\}_{i=1}^N$ captured from distinct viewpoints $\{p_i\}_{i=1}^N$. We aim to embed detectable watermarks in novel views synthesized from these images.

The watermark embedding process is formalized as a function $E(\mathcal{I}_i, \mathcal{W})$, where $\mathcal{I}_i$ denotes the i-th image and $\mathcal{W}$ represents the watermark information. This function outputs the watermarked image $\mathcal{I}_i^W$:

$$\mathcal{I}_i^W = E(\mathcal{I}_i, \mathcal{W}). \tag{6}$$

This embedding process forms the foundation of our watermarking framework, transforming original images while preserving their visual quality.

Adversaries may employ these watermarked images to reconstruct a 3D scene representation F_θ through a reconstruction function $R(\cdot)$:

$$F_\theta = R(\{\mathcal{I}_i^W\}_{i=1}^N), \tag{7}$$

where θ denotes the parameters of the 3D representation, and $R(\cdot)$ could be any reconstruction method (e.g., NeRF [1] or 3DGS [2]). Given a novel camera pose p^q, the corresponding view is rendered via a function $\Psi(\cdot,\cdot)$:

$$\hat{\mathcal{I}}_{p^q} = \Psi(F_\theta, p^q). \tag{8}$$

This rendering process represents the potential unauthorized generation of novel views from the reconstructed 3D representation.

The watermark detection function $D(\hat{\mathcal{I}}_{p^q})$ is defined as:

$$D(\hat{\mathcal{I}}_{p^q}) = \begin{cases} 1 & \text{if watermark } \mathcal{W} \text{ is detected in } \hat{\mathcal{I}}_{p^q}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This binary detection mechanism enables reliable verification of watermark presence in synthesized views.

We propose MantleMark, a watermarking framework that migrates watermarks from multi-view images to radiance fields while addressing two key challenges. First, for cross-dimensional persistence, the watermark $\mathcal{W}$ enables effective migration from 2D images to 3D radiance fields through frequency modulation, ensuring detectability in novel views rendered from the reconstructed representation. Second, for reconstruction method independence, the embedding process E operates purely in the image domain without requiring access to the reconstruction process R, allowing the watermarks to remain detectable regardless of which radiance field techniques adversaries might employ, such as NeRF [1] or 3DGS [2].

B. Frequency-domain watermarking representations

To enable effective watermark propagation across dimensions, we leverage the Fourier Projection-Slice Theorem [14], [15], which establishes a relationship between the 2D image frequency domain and the 3D frequency domain in volumetric rendering. This allows our watermarks embedded in the 2D frequency domain to effectively propagate to the 3D frequency domain and remain detectable in novel rendered views. The detailed theoretical derivation is provided in Section IV-D.

To implement this concept, we construct the watermark information as Frequency-domain Watermarking Representations (FWR), as illustrated in Figure 3. We choose a mantle-like structure for our FWR because its 2D slices across different viewing angles yield ring patterns with consistent scaling variations, making the implementation of both embedding and detection processes straightforward. The results for other geometric patterns (cube and octahedron) are presented in the ablation study (Section V-H4). We define our mantle-like FWR as a representation $\mathcal{W} : \mathbf{k} \mapsto \mathcal{W}(\mathbf{k})$, where $\mathbf{k}$ represents the 3D coordinates in the Fourier space. As shown in Figure 3(a), We define a spherical shell in the Fourier space to specify the watermarked region between radii $f_{\min}$ and $f_{\max}$. To ensure smooth transitions, the FWR $\mathcal{W}$ is defined using a Gaussian function:

$$\mathcal{W}(\mathbf{k}) = \exp(-\frac{(|\mathbf{k}| - r_c)^2}{2\sigma^2}), \tag{10}$$

where $|\mathbf{k}|$ is the radial distance from the origin in 3D Fourier space, $r_c = (f_{\min} + f_{\max})/2$ is the center radius of the shell, and $\sigma = (f_{\max} - f_{\min})/4$ controls the width of the transition. This formulation creates a smooth, spherically symmetric watermark pattern in the 3D frequency domain, as shown by the radial intensity profile in Figure 3(c).

The spherical design ensures that any 2D slice through the origin of this 3D pattern results in a circular ring with smooth transitions, as shown in Figure 3(b), maintaining rotational invariance while providing consistent watermark characteristics across different viewing angles. When projected onto a 2D image plane, this manifests as an annular region between the scaled radii $f_{\min}$ and $f_{\max}$, preserving the smooth Gaussian transition profile of the original 3D representation. The choice of frequency range $(f_{\min}, f_{\max})$ is crucial: it should be high enough to avoid affecting the main image content but low enough to survive potential rendering and compression artifacts.

This 3D mantle-like FWR design forms the core of our MantleMark framework, inspiring its name while offering several key advantages. First, its rotational symmetry ensures view-independent detection capability. Second, the smooth Gaussian transition helps minimize visual artifacts in the spatial domain. Third, the concentrated frequency band allows efficient watermark embedding and detection while maintaining image quality.

C. Frequency modulation

To enable effective watermark migration through frequency modulation, we propose a view-dependent frequency modulation approach that applies the spherical FWR based on individual viewing directions to modulate the frequency components of multi-view images for watermark embedding. The modulation process can be formulated as $\mathcal{I}_i^W = E(\mathcal{I}_i, \mathcal{W}, p_i)$, where $\mathcal{W}$ is the spherical FWR defined in Section IV-B, and p_i is the viewpoint of image $\mathcal{I}_i$.

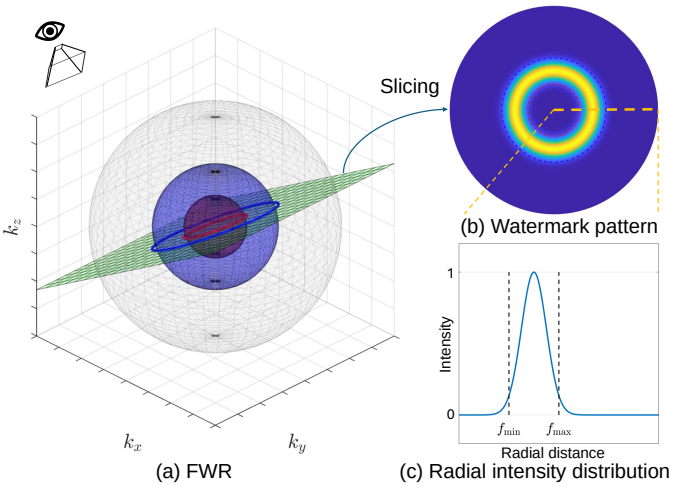

Fig. 3. Visualization of the Frequency-domain Watermarking Representation (FWR). (a) 3D visualization of the spherical shell structure in Fourier space, where a mantle-like watermark pattern is defined between radii $f_{\min}$ and $f_{\max}$ (shown as red and blue spheres), with a slicing plane (green) intersecting the structure. (b) The resulting 2D slice pattern shows the circular ring with smooth Gaussian transitions in the frequency domain. (c) Radial intensity profile of the FWR, illustrating the Gaussian transition between $f_{\min}$ and $f_{\max}$, where the intensity represents $\mathcal{W}(\mathbf{k})$.

According to the Fourier Projection-Slice Theorem [14], [15], we first obtain the 2D watermark pattern $\mathcal{W}_{p_i}$ by slicing the 3D spherical FWR along the viewing plane:

$$\mathcal{W}_{p_i} = S(\mathcal{W}, p_i), \tag{11}$$

where S denotes the slicing operation. Due to the spherical symmetry of our FWR design, this slice naturally results in a circular ring-shaped pattern with a Gaussian transition in the 2D frequency domain.

For frequency modulation, we first transform the input image to the frequency domain:

$$\mathcal{F}_i = \text{FT}\{\mathcal{I}_i\}, \tag{12}$$

where $\text{FT}\{\cdot\}$ denotes the Fourier transform. The watermark is then embedded through magnitude spectrum modulation while preserving the phase information:

$$|\mathcal{F}_i^W| = |\mathcal{F}_i| \cdot (1 - \alpha \mathcal{W}_{p_i}), \tag{13}$$

where α is the watermark strength parameter, and $|\cdot|$ denotes the magnitude spectrum. The phase spectrum remains unchanged:

$$\angle \mathcal{F}_i^W = \angle \mathcal{F}_i, \tag{14}$$

where $\angle$ denotes the phase angle.

Finally, we obtain the watermarked image through Inverse Fourier transform (IFT) and intensity range normalization:

$$\mathcal{I}_i^W = \text{clip}(\text{IFT}\{\mathcal{F}_i^W\}, 0, 255), \tag{15}$$

where clip$(\cdot)$ ensures the pixel values remain within the valid range.

Our frequency modulation approach carefully considers both theoretical and practical aspects. The preservation of phase information ensures that key structural elements of the image remain intact while the inherited smooth transitions from our spherical FWR design effectively suppress unwanted artifacts. By implementing this view-dependent frequency modulation strategy, we achieve robust watermark embedding that adapts naturally to different viewing conditions while maintaining high visual fidelity.

D. Adversarial novel view synthesis

When adversaries obtain the watermarked images $\{\mathcal{I}_i^W\}_{i=1}^N$, they can perform 3D reconstruction using methods like Neural Radiance Fields (NeRF) [1] or 3D Gaussian Splatting (3DGS) [2] to obtain a 3D representation F_θ. NeRF represents scenes using a neural network that maps 3D coordinates to density and color values, while 3DGS uses a set of 3D Gaussians to model the scene geometry and appearance. Both methods can achieve high-quality novel view synthesis through volumetric rendering.

By optimizing the 3D representation F_θ to reconstruct the watermarked training images $\{\mathcal{I}_i^W\}_{i=1}^N$, adversaries can then render novel views through volumetric rendering. As defined in Equation (8), given any novel camera pose p^q, they can generate the corresponding view $\hat{\mathcal{I}}_{p^q}$ using the rendering function $\Psi(F_\theta, p^q)$.

We leverage the frequency-domain properties of volumetric rendering and the Fourier Projection-Slice Theorem to ensure our watermarks remain detectable in these synthesized novel views. The theoretical foundation for this watermark preservation mechanism is detailed in Section IV-F.

E. Watermark detection via matched filtering

Detecting watermarks in novel view images poses unique challenges due to potential geometric distortions and varying frequency responses. To address this, we employ a matched filtering approach, which is optimal for detecting known signals in the presence of additive noise and can adapt to different viewing conditions.

Our watermark detection focuses on binary presence verification [75]–[77]. Given a rendered image $\hat{\mathcal{I}}_{p^q}$ from a novel view p^q, we first compute its Fourier transform FT$\{\hat{\mathcal{I}}_{p^q}\}$ and extract the magnitude spectrum. The detection process utilizes a pose-dependent ring template h_{p^q} that accounts for the geometric transformation of the frequency-domain watermark pattern.

The template is constructed by slicing the 3D FWR as defined in Equation (11), inheriting the smooth Gaussian transition profile from the original spherical design. The template is then normalized to unit energy:

$$\tilde{h}_{p^q} = \frac{\mathcal{W}_{p^q}}{\sqrt{\sum_{\mathbf{u}} \mathcal{W}_{p^q}(\mathbf{u})^2}}, \tag{16}$$

where $\mathbf{u}$ represents the 2D frequency coordinates in the sliced frequency domain.

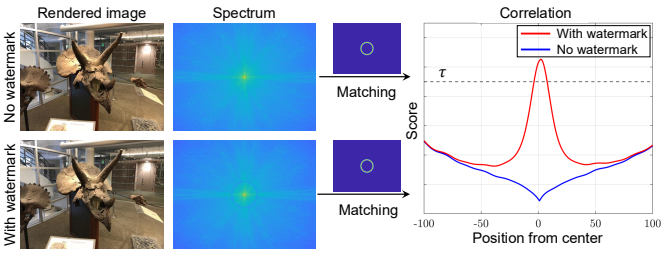

Fig. 4. Illustration of the watermark detection process. Given a pair of rendered images with and without a watermark embedded, we analyze their frequency spectrum, which reveals the embedded ring pattern. The correlation scores clearly distinguish the watermarked image (red) from the non-watermarked one (blue), with the former showing a distinctive peak at the center. The dashed line indicates the detection threshold τ.

For robust detection, we apply logarithmic transformation to enhance contrast in the frequency domain:

$$S_{log}(\mathbf{u}) = \log(1 + |\text{FT}\{\hat{\mathcal{I}}_{p^q}\}(\mathbf{u})|). \tag{17}$$

The detection score is computed using normalized cross-correlation between the log-transformed spectrum and the template:

$$\text{Score} = \text{NCC}(S_{log}, \tilde{h}_{p^q}), \tag{18}$$

where NCC$(\cdot,\cdot)$ denotes the normalized cross-correlation.

A watermark is detected if this score exceeds a predetermined threshold τ, as illustrated in Figure 4. The matched filtering approach provides significant advantages over direct frequency comparison. The logarithmic transformation enhances the visibility of subtle frequency modifications while preserving the relative magnitude relationships in the spectrum. The pose-dependent template design naturally accommodates geometric transformations, and the normalized cross-correlation yields a reliable detection metric that remains consistent across various viewing conditions by inherently normalizing for amplitude variations, making it highly suitable for novel view watermark verification.

F. Theoretical analysis of watermark propagation

This section analyzes how our frequency modulation based watermarks migrate from input multi-view images to novel views through 3D reconstruction and rendering. When adversaries obtain the watermarked images $\{\mathcal{I}_i^W\}_{i=1}^N$, they perform 3D reconstruction to obtain representation F_θ, from which novel views $\hat{\mathcal{I}}_p^R$ can be synthesized. Through the Fourier Projection-Slice Theorem [14], [15], we demonstrate that our frequency watermarks naturally propagate to novel views.

Consider the unified volumetric rendering formulation for both NeRF [1] and 3D Gaussian Splatting [2]. The rendered image can be expressed as an integration along viewing rays:

$$\hat{\mathcal{I}}_p(x,y) = \int_{-\infty}^{\infty} v(x,y,z,p)dz, \tag{19}$$

where $v(x,y,z,p)$ is the volume function at camera pose p.

The Fourier transform of this rendered image satisfies:

$$\begin{aligned} \text{FT}\{\hat{\mathcal{I}}_p(x,y)\} &= \text{FT}\{\int_{-\infty}^{\infty} v(x,y,z,p)\, dz\} \\ &= V(k_x, k_y, 0, p), \end{aligned} \tag{20}$$

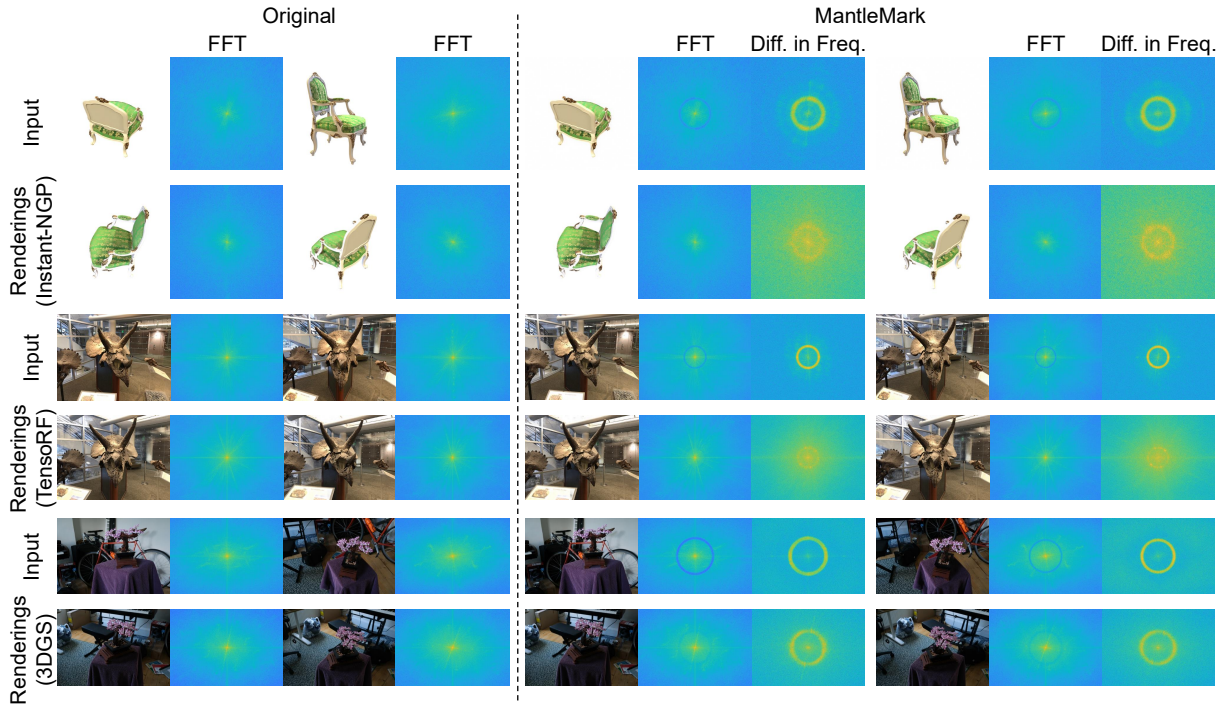

Fig. 5. Qualitative results of our method. We demonstrate the frequency domain analysis of our MantleMark approach on various datasets, including the "chair" scene from the Synthetic NeRF dataset [1], the "horns" scene from the LLFF dataset [79], and the "bonsai" scene from the Mip-NeRF 360 dataset [17]. For each scene, we show both the input images, their renderings, and their corresponding Fourier transforms (FFT). The right columns in the MantleMark display the Difference in the Frequency domain (Diff. in Freq.) between the original and marked images.

where $V(k_x, k_y, k_z, p)$ is the 3D Fourier transform of $v(x, y, z, p)$. This relationship, established by the Fourier Projection-Slice Theorem, shows that the frequency spectrum of any rendered image is a slice of the 3D frequency representation.

During 3D reconstruction, the optimization minimizes the difference between rendered and input images. Using Parseval's theorem [78], this optimization in the spatial domain can be analyzed by examining the frequency components in watermarked and non-watermarked regions separately:

$$\begin{aligned} L_2(\hat{\mathcal{I}}_{p_i}, \mathcal{I}_i^W) &= \beta L_2(V(k_x, k_y, 0, p_i), \mathcal{F}_i^W) \\ &= \beta L_2\big(V(k_x, k_y, 0, p_i)\mathcal{W}_{p_i} + V(k_x, k_y, 0, p_i)(1 - \mathcal{W}_{p_i}), \\ &\qquad (1-\alpha)\mathcal{F}_i\mathcal{W}_{p_i} + \mathcal{F}_i(1 - \mathcal{W}_{p_i})\big), \end{aligned} \tag{21}$$

where β is a normalization factor. In the watermarked regions defined by $\mathcal{W}_{p_i}$, the optimization encourages V to match the attenuated frequency content $(1-\alpha)\mathcal{F}_i$, while in non-watermarked regions defined by $1 - \mathcal{W}_{p_i}$, it preserves the original frequency content $\mathcal{F}_i$.

This frequency domain relationship reveals why our watermarks propagate to novel views. The optimization forces the 3D frequency representation V to match the attenuated patterns in watermarked regions while maintaining content fidelity elsewhere. Due to the continuity of the 3D Fourier transform, and our spherical FWR design, the watermark patterns are naturally embedded into the entire 3D frequency space. When rendering novel views, these patterns manifest in the frequency spectra of rendered images through the projection-slice relationship, enabling watermark detection from arbitrary viewpoints.

G. Implementation details

We implement our method using PyTorch and evaluate it on multiple reconstruction methods, including Instant-NGP [26], TensoRF [25], and 3DGS [2]. We follow their widely used implementations and default training configurations. The Frequency-domain Watermarking Representation (FWR) is implemented using a 2D Fast Fourier Transform (FFT). Throughout our experiments, we set the watermark strength parameter α between 0.3 and 1.0. For watermark verification, we randomly sample 600 novel viewpoints around the scene. The detection threshold τ is set by fixing the False Positive Rate (FPR) at 1% using a validation set of non-watermarked images. All experiments are performed on a single NVIDIA Tesla V100 GPU.

V. Experiments and Results

A. Experimental setting

Dataset. We evaluate our methods on three commonly used datasets for novel view synthesis: the LLFF dataset [79], Synthetic NeRF dataset [1], and Mip-NeRF 360 dataset [17]. The Synthetic NeRF dataset contains 8 detailed synthetic objects, each with 100 images captured from virtual cameras arranged on a hemisphere pointing inward. The LLFF dataset comprises 8 real-world scenes with mainly forward-facing images, each containing between 20 to 62 images. The Mip-NeRF 360 dataset contains 9 indoor and outdoor scenes

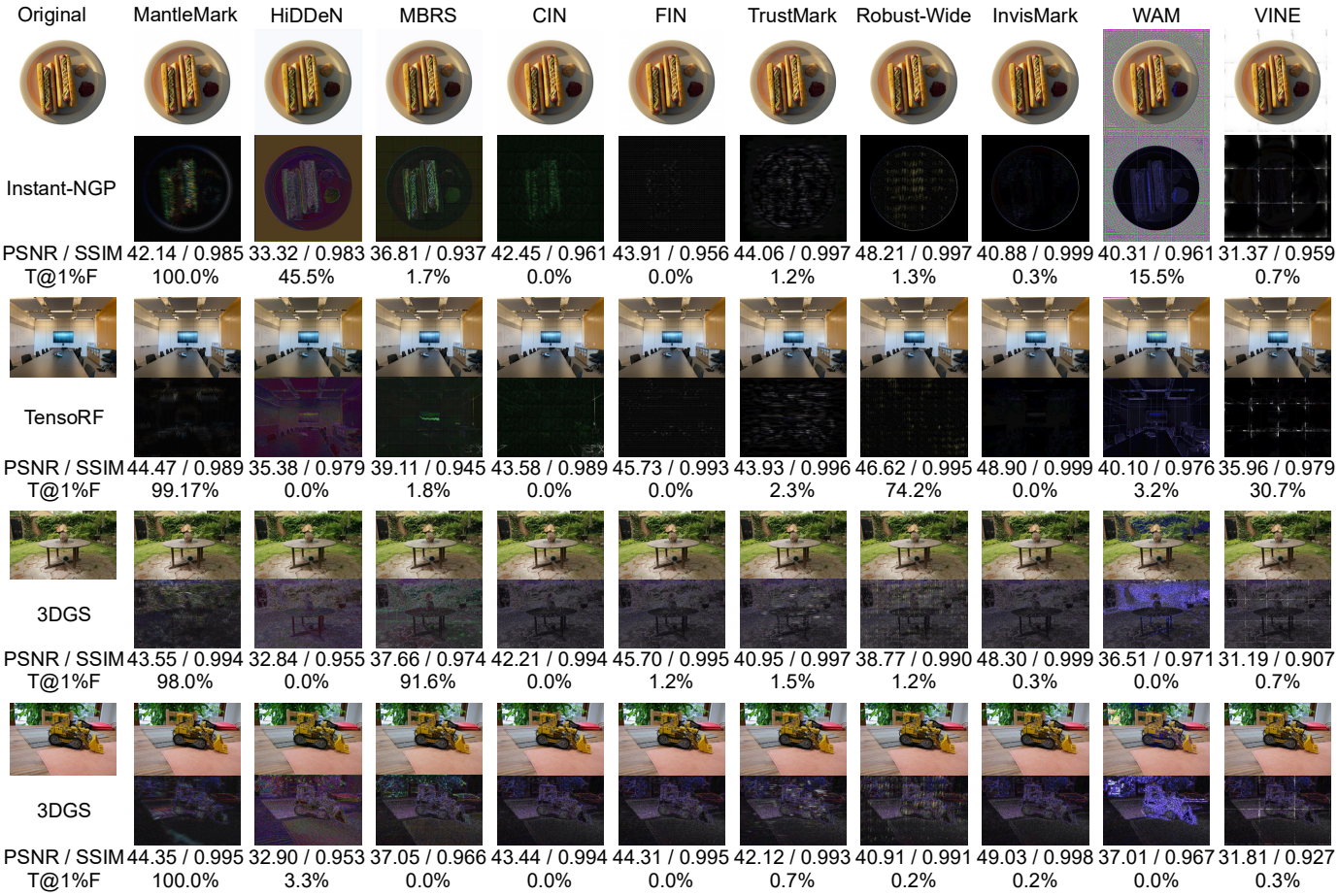

Fig. 6. Qualitative results compared with baselines (HiDDeN [38], MBRS [6], CIN [7], FIN [8], TrustMark [57], Robust-Wide [59], InvisMark [60], WAM [61], and VINE [58]). The comparison demonstrates watermarking effects across different reconstruction methods: results using Instant-NGP [26] are shown for the "hotdog" scene, TensoRF [25] for the "room" scene, and 3DGS [2] for the "garden" and "kitchen" scenes. For each method, the protected images are shown in the top row, with their corresponding differences ($\times 10$) between watermarked and original results displayed in the bottom row. The metrics below each result indicate PSNR/SSIM values, while TPR@1%FPR (T@1%F) shows the watermark detection success rate.

captured with 100-200 images per scene, covering 360-degree views of complex environments.

Baselines. We compare our method with four deep learning-based watermarking methods:

1) **HiDDeN [38]**, which employs a deep encoder-decoder architecture for image watermarking;
2) **MBRS [6]**, an end-to-end method that enhances robustness against JPEG compression through mini-batch real and simulated JPEG compression;
3) **CIN [7]**, which combines invertible and non-invertible mechanisms for high imperceptibility and robustness against strong noise attacks;
4) **FIN [8]**, a flow-based framework utilizing invertible up-down-sampling neural blocks for simultaneous embedding and extraction;
5) **TrustMark [57]**, a GAN-based watermarking method that balances image quality with watermark recovery accuracy for arbitrary resolution images;
6) **Robust-Wide [59]**, a robust watermarking method against instruction-driven image editing using Partial Instruction-driven Denoising Sampling Guidance;
7) **InvisMark [60]**, a watermarking technique for high-resolution AI-generated images that achieves high imperceptibility with extended payload capacity;
8) **WAM [61]**, the Watermark Anything Model that enables localized watermarking by segmenting images and recovering multiple hidden messages;
9) **VINE [58]**, a watermarking method that enhances robustness against image editing techniques by leveraging frequency analysis and diffusion models.

All baselines are implemented using publicly available code.

Evaluation methodology. We evaluate our proposed MantleMark method in three key aspects: detectability, imperceptibility, and robustness. For *detectability*, we adopt three metrics: the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, the True Positive Rate (TPR) at 1% False Positive Rate (FPR), denoted as TPR@1%FPR, and Accuracy as evaluation metrics. Higher values indicate better watermark detection performance. For *imperceptibility*, we employ Peak Signal-to-Noise Ratio (PSNR), Structural Similarity score (SSIM) [80], and Learned Perceptual Image Patch Similarity (LPIPS) [81]. Higher PSNR and SSIM values or lower LPIPS values indicate better imperceptibility. For *robustness*, we assess the watermarking performance under

different types of image-level distortions following Tree-Ring [77], including rotation, cropping, JPEG compression, blurring, noise, and color jitter.

B. Qualitative results

1) Watermark detectability: As shown in Figure 5, we analyze our MantleMark in the frequency domain across different datasets and radiance field models. For each scene, we visualize the input images, their rendered novel views, and their corresponding Fourier transforms. The right columns display the differences between the original and watermarked images. Our embedded watermarks manifest as distinctive ring-like patterns in the frequency domain, which are consistently preserved across different radiance field models, including Instant-NGP [26], TensoRF [25], and 3DGS [2].

These results provide strong empirical validation of our theoretical framework based on the Fourier Projection-Slice Theorem [14], [15]. The theorem predicts that frequency-domain patterns embedded in multi-view images should naturally propagate to the 3D representation and remain detectable in rendered novel views. As evidenced by the consistent ring patterns visible in both the input watermarked images and their corresponding rendered views, our watermarks successfully migrate from the 2D frequency domain to the 3D radiance field representation. This cross-dimensional persistence occurs because the volumetric rendering process, which integrates along viewing rays, corresponds to a projection operation in the frequency domain. Consequently, the watermark patterns embedded in our mantle-like Frequency-domain Watermarking Representation (FWR) are preserved through the reconstruction and rendering pipeline, regardless of the specific radiance field architecture employed. These results demonstrate that our method successfully embeds watermarks that remain imperceptible in the spatial domain while being detectable through frequency analysis of the synthesized novel views, validating the effectiveness of our frequency-domain approach for cross-dimensional watermark migration.

2) Comparison with baselines: Qualitative comparisons between our MantleMark and watermarking methods HiDDeN [38], MBRS [6], CIN [7], FIN [8], TrustMark [57], Robust-Wide [59], InvisMark [60], WAM [61], and VINE [58] are presented in Figure 6 across different datasets and radiance field models. Our method demonstrates two major advantages. First, MantleMark achieves superior imperceptibility, as evidenced by the high PSNR/SSIM scores and minimal differences across all test scenes. Second, our method demonstrates robust watermark detection performance in novel view synthesis across different radiance field models, including Instant-NGP [26], TensoRF [25], and 3DGS [2], as evidenced by the high TPR@1%FPR scores shown in Figure 6.

Existing methods exhibit various limitations. HiDDeN [38] exhibits visible artifacts in the watermarked images, particularly noticeable as color distortions, though it achieves moderate detection rates (45.5% TPR@1%FPR on Instant-NGP). While MBRS [6], CIN [7], and FIN [8] achieve reasonable imperceptibility in input images, their watermarks largely degrade or completely disappear in the synthesized novel views, as shown by their low detection rates (0.0%-1.8%). TrustMark [57] and InvisMark [60], despite maintaining high visual quality with minimal frequency-domain distortions, fail to preserve watermark detectability across novel views, achieving TPR@1%FPR values of only 0.2%-1.3%. Robust-Wide [59] demonstrates improved robustness compared to earlier methods with detection rates reaching 74.2% on TensoRF, but still falls significantly short of our method's performance and shows highly inconsistent behavior across different reconstruction models (1.2%-74.2%). WAM [61] shows limited detection capability (0.0%-15.5% TPR@1%FPR) and exhibits highly localized frequency patterns that do not propagate effectively through volumetric rendering. VINE [58], while designed for robustness, produces noticeable artifacts in the frequency domain and achieves limited detection success (0.3%-30.7%) in novel views. These results validate that MantleMark effectively addresses the challenge of maintaining high imperceptibility and reliable detectability in multi-view imaging scenarios, outperforming existing methods by successfully embedding watermarks that migrate from 2D frequency domain to 3D representations.

C. Quantitative results

We provide comprehensive evaluation results of MantleMark compared with watermarking methods across different radiance field models, including Instant-NGP [26], TensoRF [25], and 3DGS [2]. The results for Instant-NGP [26] and TensoRF [25] are shown in Table I, while the results for 3DGS are presented in Table II. Both tables demonstrate that MantleMark achieves superior watermark detection performance while maintaining image quality across datasets.

Image quality. As shown in Table I, on the Synthetic NeRF dataset [1], our MantleMark with $\alpha = 0.4$ achieves PSNR at 43.847dB, higher than InvisMark [60] (39.796dB) while slightly lower than Robust-Wide [59] (45.845dB). On the LLFF dataset [79], our method maintains high quality at 42.264dB. As shown in Table II, on the Mip-NeRF 360 dataset [17], our method achieves 44.454dB with $\alpha = 0.4$. Notably, even with increased watermark strength at $\alpha = 1.0$, our method maintains PSNR above 34dB across all datasets, demonstrating an effective quality-robustness trade-off.

Watermark detection. Regarding watermark detection, MantleMark demonstrates exceptional performance across all tested models. On the Synthetic NeRF dataset, as shown in Table I, our method with $\alpha = 1.0$ achieves TPR@1%FPR at 0.968 with Instant-NGP [26] and 0.988 with TensoRF [25]. On the LLFF dataset, our method achieves TPR@1%FPR values ranging from 0.962 to 1.000 across different strength settings and rendering models. As shown in Table II, with 3DGS [2], our method achieves strong detection performance on both datasets. At the highest watermark strength ($\alpha = 1.0$), TPR@1%FPR reaches 1.000 on Synthetic NeRF and 0.963 on Mip-NeRF 360.

Existing methods show significantly lower detection rates. On the Synthetic NeRF dataset with Instant-NGP, the majority of baseline methods show limited detection performance, with CIN [7], FIN [8], TrustMark [57], Robust-Wide [59], InvisMark [60], and VINE [58] achieving TPR@1%FPR below

TABLE I

QUANTITATIVE COMPARISON OF VARIOUS METHODS ON SYNTHETIC NERF DATASET [1] AND LLFF [79] DATASETS. WE EVALUATE BOTH IMAGE QUALITY METRICS AND WATERMARK DETECTION METRICS FOR INSTANT-NGP [26] AND TENSORF [25] MODELS. THE BEST RESULTS ARE SHOWN IN **BOLD**, AND THE SECOND BEST RESULTS ARE UNDERLINED. ↑ (↓) INDICATES HIGHER (LOWER) VALUES ARE BETTER. EACH VALUE IS AVERAGED OVER 600 RANDOMLY SAMPLED RENDERED VIEWS, REPEATED 5 TIMES, WITH 95% CONFIDENCE INTERVALS REPORTED IN SUBSCRIPT.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	Instant-NGP [26]			TensoRF [25]		
					AUC↑	TPR@1%FPR↑	Accuracy↑	AUC↑	TPR@1%FPR↑	Accuracy↑
Synthetic [1]	HiDDeN [38]	32.855	0.984	0.014	$0.573_{0.023}$	$0.069_{0.008}$	$0.634_{0.015}$	$0.725_{0.019}$	$0.164_{0.012}$	$0.706_{0.013}$
	MBRS [6]	36.680	0.935	0.024	$0.688_{0.018}$	$0.032_{0.006}$	$0.650_{0.012}$	$0.753_{0.016}$	$0.037_{0.007}$	$0.698_{0.011}$
	CIN [7]	41.950	0.963	0.003	$0.498_{0.028}$	$0.003_{0.002}$	$0.512_{0.019}$	$0.505_{0.027}$	$0.003_{0.002}$	$0.515_{0.018}$
	FIN [8]	43.738	0.960	0.042	$0.492_{0.029}$	$0.004_{0.002}$	$0.511_{0.020}$	$0.507_{0.026}$	$0.003_{0.002}$	$0.518_{0.017}$
	TrustMark [57]	42.683	0.996	**0.002**	$0.500_{0.027}$	$0.007_{0.003}$	$0.503_{0.019}$	$0.500_{0.028}$	$0.014_{0.005}$	$0.504_{0.020}$
	Robust-Wide [59]	**45.845**	**0.997**	**0.002**	$0.516_{0.025}$	$0.012_{0.004}$	$0.522_{0.017}$	$0.517_{0.024}$	$0.006_{0.003}$	$0.522_{0.016}$
	InvisMark [60]	39.796	0.996	0.003	$0.520_{0.024}$	$0.006_{0.003}$	$0.523_{0.016}$	$0.497_{0.029}$	$0.007_{0.003}$	$0.518_{0.019}$
	WAM [61]	39.338	0.960	0.030	$0.541_{0.022}$	$0.135_{0.011}$	$0.568_{0.014}$	$0.543_{0.021}$	$0.153_{0.013}$	$0.577_{0.013}$
	VINE [58]	30.752	0.957	0.177	$0.504_{0.026}$	$0.008_{0.003}$	$0.512_{0.018}$	$0.515_{0.025}$	$0.015_{0.005}$	$0.526_{0.017}$
	MantleMark ($\alpha = 0.4$)	43.847	0.991	0.006	$0.876_{0.009}$	$0.507_{0.016}$	$0.819_{0.008}$	$0.878_{0.008}$	$0.462_{0.015}$	$0.820_{0.007}$
	MantleMark ($\alpha = 0.8$)	38.562	0.976	0.018	$0.990_{0.004}$	$0.912_{0.009}$	$0.964_{0.004}$	$0.991_{0.003}$	$0.902_{0.010}$	$0.971_{0.003}$
	MantleMark ($\alpha = 1.0$)	36.735	0.967	0.026	$\mathbf{0.998}_{0.001}$	$\mathbf{0.968}_{0.007}$	$\mathbf{0.989}_{0.002}$	$\mathbf{0.999}_{0.001}$	$\mathbf{0.988}_{0.006}$	$\mathbf{0.995}_{0.002}$
LLFF [79]	HiDDeN [38]	33.530	0.970	0.028	$0.465_{0.031}$	$0.120_{0.014}$	$0.560_{0.021}$	$0.707_{0.018}$	$0.302_{0.019}$	$0.751_{0.012}$
	MBRS [6]	37.416	0.960	0.018	$0.710_{0.019}$	$0.395_{0.022}$	$0.733_{0.013}$	$0.709_{0.020}$	$0.269_{0.018}$	$0.704_{0.014}$
	CIN [7]	42.127	0.991	0.006	$0.497_{0.029}$	$0.000_{0.001}$	$0.511_{0.020}$	$0.496_{0.030}$	$0.000_{0.001}$	$0.511_{0.021}$
	FIN [8]	44.271	0.995	0.005	$0.499_{0.028}$	$0.000_{0.001}$	$0.513_{0.019}$	$0.493_{0.031}$	$0.001_{0.001}$	$0.513_{0.022}$
	TrustMark [57]	42.422	0.995	**0.002**	$0.631_{0.024}$	$0.272_{0.020}$	$0.636_{0.016}$	$0.553_{0.027}$	$0.123_{0.015}$	$0.561_{0.018}$
	Robust-Wide [59]	41.545	0.992	0.004	$0.786_{0.015}$	$0.505_{0.024}$	$0.772_{0.011}$	$0.777_{0.016}$	$0.481_{0.023}$	$0.768_{0.012}$
	InvisMark [60]	**46.974**	**0.998**	**0.002**	$0.508_{0.027}$	$0.006_{0.003}$	$0.520_{0.018}$	$0.500_{0.029}$	$0.005_{0.003}$	$0.518_{0.020}$
	WAM [61]	37.653	0.973	0.033	$0.503_{0.028}$	$0.005_{0.003}$	$0.505_{0.019}$	$0.510_{0.026}$	$0.018_{0.006}$	$0.510_{0.018}$
	VINE [58]	36.924	0.985	0.028	$0.649_{0.022}$	$0.182_{0.017}$	$0.632_{0.015}$	$0.660_{0.021}$	$0.176_{0.016}$	$0.639_{0.014}$
	MantleMark ($\alpha = 0.4$)	42.264	0.989	0.006	$0.999_{0.001}$	$0.997_{0.002}$	$0.998_{0.001}$	$0.997_{0.002}$	$0.962_{0.008}$	$0.985_{0.003}$
	MantleMark ($\alpha = 0.8$)	36.593	0.963	0.024	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$
	MantleMark ($\alpha = 1.0$)	34.705	0.945	0.037	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$

0.015. WAM [61] achieves the best baseline performance at 0.135, but still far below our method. On the LLFF dataset, baseline methods show varied but generally poor detection performance, with most methods achieving TPR@1%FPR below 0.4, significantly lower than our method's detection rate. Similar patterns are observed across all datasets and rendering models, validating that MantleMark successfully achieves robust watermark detection while preserving high visual quality across various novel view synthesis scenarios.

D. Watermark pattern propagation analysis

To provide deeper insights into MantleMark's superior detection performance, we analyze the propagation of watermark patterns from protected input images to rendered novel views. As illustrated in Figure 7, we visualize the pixel-wise differences (amplified by 10× for visibility) between watermarked and original images for both protected inputs and their corresponding rendered outputs.

MantleMark demonstrates a critical advantage: the watermark patterns embedded in protected images successfully propagate to the rendered novel views. As shown in the left column, the differences between protected and clean images reveal subtle perturbations manifesting as multi-ring circular patterns concentrated along the edges of the circular field of view. Importantly, in the rendered novel views (right column), the difference images still exhibit these multi-ring patterns, demonstrating that the characteristic annular structures are successfully preserved across the novel view synthesis process.

Existing methods like MBRS [6] and Robust-Wide [59] fail to maintain pattern consistency across views. While the differences in MBRS's protected input images (red boxes) reveal regular grid-like patterns, these structured patterns completely disappear in the differences of rendered novel views (blue boxes), with no trace of the grid structure visible. Similarly, the differences in Robust-Wide's protected images show wave-like patterns, but these characteristic wave patterns are entirely absent in the differences of rendered images, indicating complete loss of watermark information during the reconstruction process. This loss of pattern integrity directly explains their poor detection performance in Table I and Table II, where TPR@1%FPR values drop to nearly zero.

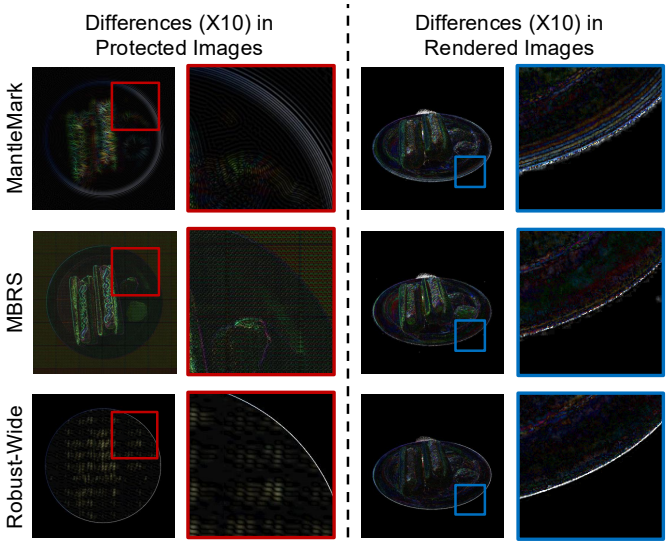

Fig. 7. Visual comparison of watermark pattern propagation from protected images to rendered views. We show pixel-wise differences (× 10) between watermarked and clean images. MantleMark successfully propagates watermark patterns to rendered views, while baseline methods (MBRS [6] and Robust-Wide [59]) lose their patterns in the rendering process.

The fundamental reason behind this discrepancy lies in the 3D consistency of the embedded watermarks. Traditional

TABLE II

QUANTITATIVE COMPARISON OF VARIOUS METHODS ON SYNTHETIC NERF DATASET [1] AND MIP-NERF 360 [17] DATASETS. WE EVALUATE BOTH IMAGE QUALITY METRICS AND WATERMARK DETECTION METRICS FOR 3DGS [2] MODELS. THE BEST RESULTS ARE SHOWN IN **BOLD**, AND THE SECOND BEST RESULTS ARE UNDERLINED. ↑ (↓) INDICATES HIGHER (LOWER) VALUES ARE BETTER. EACH VALUE IS AVERAGED OVER 600 RANDOMLY SAMPLED RENDERED VIEWS, REPEATED 5 TIMES, WITH 95% CONFIDENCE INTERVALS REPORTED IN SUBSCRIPT.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	3DGS		
					AUC↑	TPR@1%FPR↑	Accuracy↑
Synthetic [1]	HiDDeN [38]	32.855	0.984	0.014	$0.956_{0.012}$	$0.837_{0.015}$	$0.942_{0.010}$
	MBRS [6]	36.680	0.935	0.024	$0.287_{0.035}$	$0.008_{0.004}$	$0.510_{0.022}$
	CIN [7]	41.950	0.963	0.003	$0.513_{0.026}$	$0.004_{0.002}$	$0.517_{0.018}$
	FIN [8]	43.738	0.960	0.042	$0.515_{0.025}$	$0.010_{0.004}$	$0.519_{0.017}$
	TrustMark [57]	42.683	0.996	**0.002**	$0.502_{0.027}$	$0.005_{0.003}$	$0.503_{0.019}$
	Robust-Wide [59]	**45.845**	**0.997**	**0.002**	$0.541_{0.023}$	$0.031_{0.007}$	$0.542_{0.016}$
	InvisMark [60]	39.796	0.996	0.003	$0.492_{0.028}$	$0.009_{0.004}$	$0.511_{0.020}$
	WAM [61]	39.338	0.960	0.030	$0.506_{0.026}$	$0.011_{0.005}$	$0.507_{0.018}$
	VINE [58]	30.752	0.957	0.177	$0.502_{0.027}$	$0.012_{0.005}$	$0.510_{0.019}$
	MantleMark ($\alpha = 0.4$)	43.847	0.991	0.006	$0.990_{0.005}$	$0.941_{0.011}$	$0.958_{0.005}$
	MantleMark ($\alpha = 0.8$)	38.562	0.976	0.018	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$
	MantleMark ($\alpha = 1.0$)	36.735	0.967	0.026	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$	$\mathbf{1.000}_{0.000}$
Mip-NeRF 360 [17]	HiDDeN [38]	34.005	0.950	0.037	$0.701_{0.020}$	$0.075_{0.010}$	$0.688_{0.014}$
	MBRS [6]	38.210	0.957	0.016	$0.800_{0.016}$	$0.405_{0.021}$	$0.777_{0.012}$
	CIN [7]	43.708	0.993	**0.002**	$0.443_{0.032}$	$0.004_{0.002}$	$0.511_{0.022}$
	FIN [8]	44.167	**0.995**	0.005	$0.461_{0.030}$	$0.006_{0.003}$	$0.505_{0.021}$
	TrustMark [57]	42.020	0.991	0.005	$0.499_{0.028}$	$0.011_{0.005}$	$0.503_{0.019}$
	Robust-Wide [59]	40.595	0.990	0.007	$0.497_{0.029}$	$0.007_{0.003}$	$0.510_{0.020}$
	InvisMark [60]	35.238	0.942	0.038	$0.484_{0.031}$	$0.003_{0.002}$	$0.515_{0.021}$
	WAM [61]	36.779	0.952	0.049	$0.502_{0.027}$	$0.000_{0.001}$	$0.502_{0.019}$
	VINE [58]	32.655	0.927	0.067	$0.516_{0.025}$	$0.005_{0.003}$	$0.523_{0.017}$
	MantleMark ($\alpha = 0.4$)	**44.454**	0.993	0.004	$0.932_{0.011}$	$0.742_{0.018}$	$0.882_{0.009}$
	MantleMark ($\alpha = 0.8$)	39.115	0.977	0.016	$0.984_{0.006}$	$0.942_{0.012}$	$0.970_{0.005}$
	MantleMark ($\alpha = 1.0$)	37.299	0.965	0.024	$\mathbf{0.990}_{0.004}$	$\mathbf{0.963}_{0.009}$	$\mathbf{0.980}_{0.004}$

TABLE III

ROBUSTNESS EVALUATION OF VARIOUS METHODS UNDER IMAGE MANIPULATIONS ON THE LLFF DATASET [79]. WE REPORT THE AREA UNDER CURVE (AUC) SCORES UNDER VARIOUS ATTACK SCENARIOS. FOR MANTLEMARK, WE EVALUATE DIFFERENT WATERMARK STRENGTHS (α).

Method	None	Rotation	Cropping	JPEG	Blurring	Noise	Color jitter	Average
HiDDeN [38]	0.302	0.215	0.287	0.294	0.198	0.245	0.268	0.251
MBRS [6]	0.269	0.192	0.258	0.263	0.176	0.221	0.242	0.225
Robust-Wide [59]	0.481	0.368	0.465	0.473	0.325	0.412	0.438	0.414
GaussianMarker [12]	1.000	0.996	1.000	1.000	0.982	0.998	0.999	0.996
GuardSplat [74]	1.000	0.998	1.000	1.000	0.991	0.999	1.000	0.998
3D-GSW [72]	1.000	0.997	1.000	1.000	0.986	0.999	0.999	0.997
MantleMark ($\alpha = 0.4$)	0.997	0.774	0.995	0.996	0.685	0.908	0.854	0.869
MantleMark ($\alpha = 0.8$)	1.000	0.972	0.999	0.999	0.885	0.998	0.994	0.975
MantleMark ($\alpha = 1.0$)	1.000	0.990	0.999	1.000	0.938	0.996	0.987	0.985

image watermarking methods embed signals designed for individual 2D images but lack the multi-view geometric consistency required for radiance field reconstruction. During optimization, these inconsistent patterns across different viewpoints are treated as noise and filtered out by the reconstruction algorithm. Our frequency-domain approach, by contrast, ensures geometric and photometric consistency across multiple views through the Fourier Projection-Slice Theorem [14], [15]. This 3D-consistent embedding allows the watermark patterns to be interpreted as part of the underlying scene structure, thereby propagating successfully to novel views.

E. Robustness against image manipulations

Following the benchmark protocol established [77], we systematically evaluate our watermarking scheme's robustness against six common image manipulations: rotation (75°), JPEG compression (quality factor 25), random cropping (75% of original size with rescaling), Gaussian blur (kernel size 8×8), Gaussian noise ($\sigma = 0.1$), and color jitter (brightness factor uniformly sampled from $[0, 6]$). Since most baseline methods show negligible detection performance as demonstrated in Table I, we select representative methods with relatively better performance for robustness evaluation. As shown in Table III, we report AUC scores on the LLFF dataset [79] under different watermark strengths α.

MantleMark significantly outperforms existing methods across all manipulation scenarios. Even with moderate strength ($\alpha = 0.4$), our method achieves an average AUC of 0.869, substantially higher than the best baseline Robust-Wide [59] at 0.414. As expected, increasing the watermark strength further enhances robustness, with $\alpha = 1.0$ achieving an average AUC of 0.985, demonstrating exceptional resilience to various image manipulations.

F. Comparison with 3D watermarking methods

Recent works [12], [72], [74] have proposed watermarking techniques for 3D Gaussian Splatting representations. However, these methods fundamentally differ from MantleMark

TABLE IV
COMPARISON WITH 3D WATERMARKING METHODS ON SYNTHETIC NERF DATASET [1] UNDER 3DGS RECONSTRUCTION. MANTLEMARK ACHIEVES COMPARABLE DETECTION PERFORMANCE WHILE OPERATING INDEPENDENTLY IN THE IMAGE DOMAIN WITHOUT REQUIRING ACCESS TO THE RECONSTRUCTION PROCESS.

Method	PSNR↑	SSIM↑	AUC↑	TPR@1%FPR↑
GaussianMarker [12]	32.12	0.972	1.000	1.000
GuardSplat [74]	38.90	0.992	1.000	1.000
3D-GSW [72]	34.24	0.977	1.000	1.000
MantleMark ($\alpha = 0.4$)	36.54	0.988	0.990	0.941
MantleMark ($\alpha = 0.8$)	34.52	0.965	1.000	1.000
MantleMark ($\alpha = 1.0$)	33.19	0.953	1.000	1.000

in motivation and approach. They embed watermarks directly into 3DGS model parameters, requiring direct access to and control over the reconstruction process, which is impractical when content owners cannot control how adversaries process their images. Moreover, these methods are model-specific, as techniques for 3DGS [2] cannot protect NeRF [1] reconstructions and vice versa.

MantleMark protects multi-view images before any reconstruction occurs, operating purely in the image domain without requiring access to the reconstruction process. This provides two critical advantages. First, owners can protect images independently of reconstruction methods used by adversaries. Second, watermarks persist across different reconstruction techniques such as NeRF [1] and 3DGS [2].

Nevertheless, we compare MantleMark with 3D watermarking methods under the 3DGS scenario in Tables III and IV. The results show that MantleMark achieves comparable detection performance (AUC=1.000 with $\alpha \geq 0.8$) while maintaining competitive visual quality. The slightly lower PSNR compared to GuardSplat [74] is expected, as GuardSplat directly optimizes 3DGS parameters with reconstruction access, while our method protects images without such access. This validates that MantleMark maintains effectiveness while providing broader applicability across diverse reconstruction pipelines.

G. Computational complexity analysis

We analyze the computational efficiency of MantleMark from both theoretical and empirical perspectives, demonstrating its practical advantages for real-world deployment.

Theoretical complexity analysis. Our MantleMark framework has well-defined computational complexity. For watermark embedding, the main operations include forward FFT transformation with complexity $O(N \log N)$, frequency domain modulation with complexity $O(N)$, and inverse FFT transformation with complexity $O(N \log N)$, where N is the number of pixels. Therefore, the overall embedding complexity is $O(N \log N)$. For watermark detection, the complexity includes forward FFT with $O(N \log N)$ and matched filtering with $O(N)$, resulting in overall detection complexity of $O(N \log N)$. This is significantly more efficient than deep learning-based methods that require neural network inference with multiple layers and parameters.

Empirical time comparison. We measure the actual watermark embedding time on a standard desktop environment with Intel Xeon CPU and a single NVIDIA Tesla V100 GPU, and compare with existing methods in Table V. MantleMark requires only 3.1 seconds per image on average, enabling rapid processing of entire multi-view datasets. 3D watermarking methods like GaussianMarker [12] and GuardSplat [74]+HiDDeN [38] require over 12 hours for embedding watermarks into the entire scene. NeRF-based methods such as CopyRNeRF [10] and WateRF [11] are even more time-consuming, requiring 84 and 14 hours respectively. While GuardSplat [74] achieves faster embedding at 15 minutes, it still requires significantly more time than our per-image approach. This dramatic efficiency advantage makes MantleMark particularly suitable for protecting large-scale multi-view datasets.

TABLE V
WATERMARK EMBEDDING TIME COMPARISON. MANTLEMARK REQUIRES ONLY SECONDS PER IMAGE, WHILE OTHER METHODS NEED MINUTES TO HOURS FOR THE ENTIRE SCENE, DEMONSTRATING SUPERIOR EFFICIENCY FOR PRACTICAL DEPLOYMENT.

Method	Time	Method	Time
MantleMark	3.1s/image	GaussianMarker [12]	12.1h
CopyRNeRF [10]	84h	GuardSplat [74]	15m
WateRF [11]	14h	GuardSplat [74]+HiDDeN [38]	12.5h

TABLE VI
ABLATION STUDY ON FREQUENCY BAND SELECTION ON THE MIP-NERF 360 DATASET [17]. WE EVALUATE BOTH IMAGE QUALITY METRICS AND WATERMARK DETECTION METRICS UNDER DIFFERENT FREQUENCY RANGES. ↑ INDICATES HIGHER VALUES ARE BETTER.

$f_{\min}$	$f_{\max}$	Bandwidth	PSNR↑	AUC↑	TPR@1%FPR↑
80	90	10	40.56	0.927	0.691
80	100	20	38.04	0.987	0.941
100	110	10	41.66	0.906	0.662
100	120	20	39.11	0.983	0.940
120	130	10	42.59	0.930	0.689
120	140	20	40.07	0.982	0.907

Space complexity analysis. Our frequency-domain approach maintains reasonable memory requirements. The FWR representation requires minimal additional storage, as it only needs to store the frequency mask pattern rather than complex model parameters. The embedding process operates in-place on frequency coefficients without requiring additional large memory buffers, unlike some deep learning methods that need to store intermediate network activations. This lightweight design ensures MantleMark can be deployed efficiently across various hardware configurations.

H. Ablation studies

We conduct comprehensive ablation studies to analyze our MantleMark framework. We study the watermark strength parameter α that controls the visibility-robustness trade-off, the frequency band selection ($f_{\min}$, $f_{\max}$) that determines the watermark location in the frequency domain, and evaluate the effectiveness of our Gaussian-based FWR definition.

1) Impact of watermark strength: We analyze the impact of watermark strength α on the trade-off between visual quality and detection performance. As shown in Table I, Table II, and Table III, increasing α leads to better watermark detection performance across all metrics and models. With $\alpha = 1.0$, our method achieves nearly perfect detection performance on both Instant-NGP [26] and TensoRF [25] models, and strong

TABLE VII
COMPARISON BETWEEN TRUNCATION APPROACH AND OUR GAUSSIAN-BASED FWR DEFINITION ON THE LLFF [79] "FERN" SCENE. WE EVALUATE BOTH IMAGE QUALITY METRICS AND WATERMARK DETECTION METRICS. THE BEST RESULTS ARE SHOWN IN **BOLD**. ↑ (↓) INDICATES HIGHER (LOWER) VALUES ARE BETTER.

Method	PSNR↑	SSIM↑	LPIPS↓	AUC↑	TPR@1%FPR↑	Accuracy↑
Truncation	31.48	0.916	0.067	**1.000**	**100.00%**	**100.00%**
MantleMark ($\alpha = 0.4$)	**44.69**	**0.995**	**0.002**	0.999	**100.00%**	99.50%
MantleMark ($\alpha = 0.8$)	36.85	0.972	0.018	**1.000**	**100.00%**	**100.00%**
MantleMark ($\alpha = 1.0$)	34.96	0.958	0.028	**1.000**	**100.00%**	**100.00%**

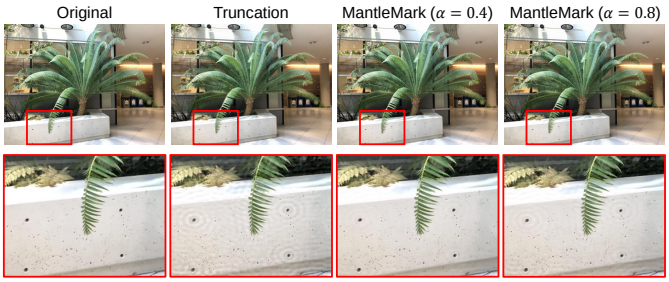

Fig. 8. Visual comparison between truncation approach and our Gaussian-based FWR definition on the LLFF [79] "fern" scene. The truncation approach introduces noticeable ringing artifacts due to sharp frequency cutoff, while our MantleMark with Gaussian transition maintains better visual quality even at higher watermark strengths.

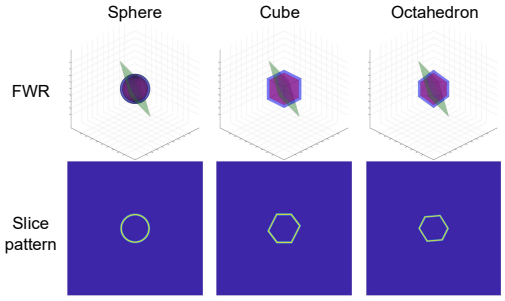

Fig. 9. Visualization of different 3D frequency patterns and their 2D slice projections. Top row shows the 3D geometric structures of sphere, cube, and octahedron patterns in frequency space. Bottom row displays the corresponding 2D slice patterns used for watermark embedding. Different patterns generate distinct geometric shapes.

performance on 3DGS [2]. However, this comes at the cost of visual quality, where PSNR drops from 43.8 to 36.7 on the Synthetic dataset [1] as α increases from 0.4 to 1.0. The robustness evaluation in Table III further confirms this trend, showing that higher α values provide better resistance against various image manipulations, with the average AUC improving from 0.869 to 0.985. This suggests that α effectively controls the visibility-robustness trade-off, allowing users to adjust the watermark strength according to their specific requirements for visual quality and detection reliability.

2) Impact of frequency band selection: The frequency band parameters $f_{\min}$ and $f_{\max}$ determine the location and bandwidth of the watermark pattern in the frequency domain. We conduct experiments with different frequency ranges while fixing the watermark strength at $\alpha = 0.8$. As shown in Table VI, we observe a clear trade-off between image quality and watermark detectability. With a fixed bandwidth of 10, increasing the frequency range from $80-90$ to $120-130$ leads to better image quality as PSNR improves from 40.56 to 42.59, while maintaining similar detection performance with AUC around 0.92. When expanding the bandwidth to 20, we achieve significantly better detection rates with AUC exceeding 0.98, but at the cost of reduced image quality as PSNR drops by approximately 2.5dB. This suggests that wider frequency bands in lower frequency ranges provide better watermark detection reliability, while narrower bands in higher frequency range better preserve image quality.

3) Impact of Gaussian-based FWR definition: We investigate the effectiveness of using a Gaussian function versus a direct truncation approach for defining the Frequency-domain Watermarking Representation (FWR). In the Gaussian approach, the FWR is defined using a smooth Gaussian transition described in Section IV-B. For comparison, we implement a direct truncation approach where the FWR is defined as:

$$\mathcal{W}_{\text{trunc}}(\mathbf{k}) = \begin{cases} 1 & \text{if } f_{\min} \leq |\mathbf{k}| \leq f_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

As shown in Table VII, our Gaussian-based FWR significantly outperforms the direct truncation approach regarding image quality. With $\alpha = 0.4$, our method achieves a PSNR of 44.69, SSIM of 0.995, and LPIPS of 0.002, substantially better than the truncation approach's PSNR of 31.48, SSIM of 0.916, and LPIPS of 0.067. Even at higher watermark strengths ($\alpha = 0.8$ and $\alpha = 1.0$), our method maintains better visual quality while achieving comparable or better watermark detection performance.

Figure 8 visually compares the two approaches. The direct truncation method introduces visible ringing artifacts around edges due to the abrupt frequency cutoff, which is particularly noticeable in the zoom-in view of the concrete wall texture. In contrast, our Gaussian-based approach maintains better visual quality through smooth frequency transitions, effectively suppressing these artifacts while preserving the original image details. These qualitative and quantitative results validate our choice of using a Gaussian function for FWR definition in the MantleMark framework.

4) Impact of different frequency patterns: We evaluate the flexibility of our framework by testing different 3D geometric patterns, including sphere, cube, and octahedron designs. Figure 9 illustrates how these different 3D patterns are sliced to generate 2D frequency masks for watermark embedding. The sphere pattern produces circular slices with consistent shapes across different viewing angles. The cube pattern generates square-shaped slices with varying aspect ratios depending on

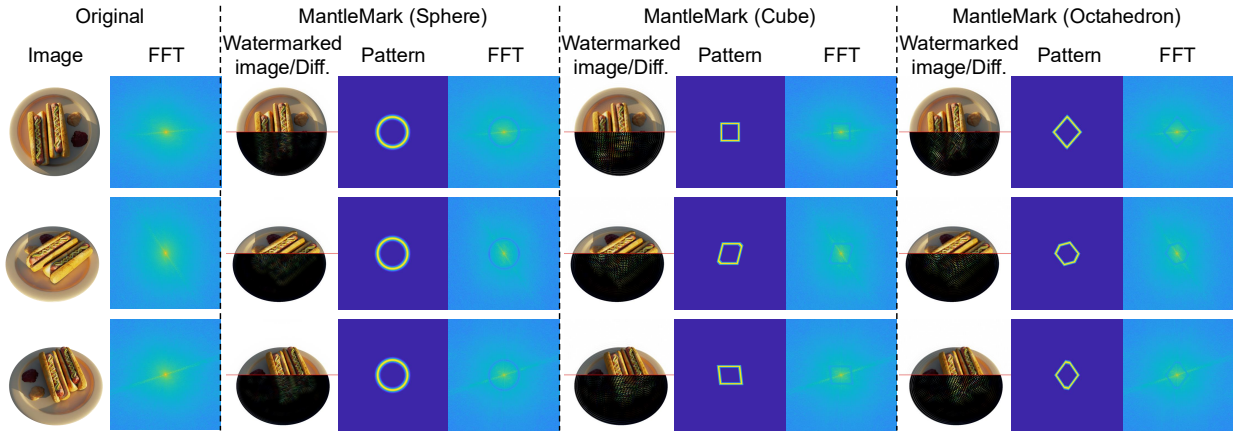

Fig. 10. Visual comparison of watermark embedding with different frequency patterns across multiple views. For each pattern (sphere, cube, and octahedron), we show the original image, its FFT representation, the watermarked image with difference visualization, the 2D slice pattern used for embedding, and the FFT of the watermarked image. Each pattern exhibits unique characteristics in the frequency domain and produces different visual effects, demonstrating the flexibility of our framework to accommodate various geometric pattern designs.

TABLE VIII
COMPARISON OF DIFFERENT FREQUENCY PATTERNS ON SYNTHETIC NERF DATASET [1]. OUR FRAMEWORK SUCCESSFULLY EMBEDS WATERMARKS USING VARIOUS GEOMETRIC PATTERNS, DEMONSTRATING ITS FLEXIBILITY AND GENERALIZATION CAPABILITY.

Pattern	PSNR↑	SSIM↑	LPIPS↓	AUC↑	TPR@1%FPR↑
Sphere	**42.26**	**0.989**	**0.006**	**0.999**	**99.58%**
Cube	39.14	0.975	0.012	0.987	95.32%
Octahedron	38.55	0.969	0.015	0.995	98.76%

the slice orientation. The octahedron pattern creates hexagonal slices with more complex geometric structures.

Figure 10 demonstrates the visual comparison of watermark embedding using these different patterns. For each pattern, we show the watermarked images, the corresponding 2D slice patterns, and their frequency domain representations. The differences in slice patterns directly affect both the watermark visibility and detection reliability. Each pattern exhibits distinct characteristics: the spherical pattern maintains uniform energy distribution across all views, the cube pattern provides structured rectangular frequency regions, and the octahedron pattern offers a unique hexagonal frequency distribution.

Table VIII presents the quantitative comparison between different 3D geometric patterns. Our framework successfully embeds watermarks using all three pattern types while maintaining high detection performance. The spherical pattern achieves a PSNR of 42.26dB and an AUC of 0.999, demonstrating excellent visual quality and detection reliability. The cube pattern maintains strong performance with a PSNR of 39.14dB and an AUC of 0.987, showing the framework's ability to handle structured rectangular frequency regions. The octahedron pattern achieves competitive results with a PSNR of 38.55dB and maintains high detection accuracy with an AUC of 0.995, particularly excelling in the TPR@1%FPR metric at 98.76%. These results demonstrate the versatility of our framework, which can accommodate various 3D geometric patterns while consistently maintaining effective watermark embedding and detection across different viewing angles. This flexibility allows users to select appropriate patterns based on specific application requirements and security considerations.

VI. CONCLUSIONS

In this paper, we present MantleMark, a watermarking framework that enables the effective migration of watermarks from multi-view images to 3D radiance fields. Through frequency-domain embedding and the Fourier Projection-Slice Theorem [14], [15], our approach ensures watermarks propagate into reconstructed 3D representations while operating purely in the image domain. This makes our method immediately applicable to existing image distribution pipelines. Content owners need to process their source images through our frequency-domain embedding before distribution without requiring knowledge or control of downstream reconstruction methods. Extensive experiments demonstrate that MantleMark successfully preserves watermarks across different radiance field models, including NeRF [1] and 3DGS [2], while maintaining high visual quality. Our approach bridges the gap between multi-view watermarking and 3D reconstruction, offering a practical solution for protecting image copyright in the era of radiance fields.

Limitations. While MantleMark demonstrates effective watermark migration across different radiance field representations, several limitations warrant discussion. First, our current implementation focuses on ownership verification through watermark presence detection. While this design ensures robustness and simplicity, incorporating additional metadata such as user identities, timestamps, or usage permissions could further enhance the framework's applicability for comprehensive content management. Second, the watermark strength parameter α presents a trade-off between visual quality and detection robustness. Although our experiments demonstrate that moderate strength values achieve good balance, extreme imaging conditions or aggressive post-processing may require higher strength values that could impact visual quality. Third, while our frequency-domain approach provides cross-dimensional persistence, the detection performance may degrade under certain challenging scenarios, such as extreme viewpoint changes

or significant scene occlusions that were not present in the original training views.

Future work. Several promising directions remain for future exploration. First, increasing watermark capacity could enable content provenance tracking and multi-party authentication through adaptive frequency patterns or layered embedding. Second, developing more sophisticated frequency patterns could enhance robustness against adversarial manipulations by responding to different attack scenarios. Third, extending our approach to dynamic scenes and video sequences presents opportunities for temporal consistency alongside spatial multi-view consistency. Finally, as radiance field technologies evolve, adapting our frequency-domain framework will ensure continued effectiveness. Our work establishes a foundation for cross-dimensional watermarking, opening new possibilities for intellectual property protection in 3D content generation and novel view synthesis.

REFERENCES

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *Proceedings of the European Conference on Computer Vision*, 2020.

[2] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, 2023.

[3] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "NeRF in the Wild: Neural radiance fields for unconstrained photo collections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[4] D. Zhang, C. Wang, W. Wang, P. Li, M. Qin, and H. Wang, "Gaussian in the Wild: 3D Gaussian splatting for unconstrained image collections," in *European Conference on Computer Vision*, 2024.

[5] I. Yoo, H. Chang, X. Luo, O. Stava, C. Liu, P. Milanfar, and F. Yang, "Deep 3D-to-2D watermarking: Embedding messages in 3D meshes and extracting them from 2D renderings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[6] Z. Jia, H. Fang, and W. Zhang, "MBRS: Enhancing robustness of DNN-based watermarking by mini-batch of real and simulated JPEG compression," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021.

[7] R. Ma, M. Guo, Y. Hou, F. Yang, Y. Li, H. Jia, and X. Xie, "Towards blind watermarking: Combining invertible and non-invertible mechanisms," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[8] H. Fang, Y. Qiu, K. Chen, J. Zhang, W. Zhang, and E.-C. Chang, "Flow-based robust watermarking with invertible noise layer for black-box distortions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[9] C. Li, B. Y. Feng, Z. Fan, P. Pan, and Z. Wang, "StegaNeRF: Embedding invisible information within neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[10] Z. Luo, Q. Guo, K. C. Cheung, S. See, and R. Wan, "CopyRNeRF: Protecting the CopyRight of neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[11] Y. Jang, D. I. Lee, M. Jang, J. W. Kim, F. Yang, and S. Kim, "WateRF: Robust watermarks in radiance fields for protection of copyrights," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[12] X. Huang, R. Li, Y. ming Cheung, K. C. Cheung, S. See, and R. Wan, "GaussianMarker: Uncertainty-aware copyright protection of 3D Gaussian splatting," in *Advances in Neural Information Processing Systems*, 2024.

[13] T. Wu, J. Zhang, X. Fu, Y. Wang, J. Ren, L. Pan, W. Wu, L. Yang, J. Wang, C. Qian *et al.*, "Omniobject3D: Large-vocabulary 3D object dataset for realistic perception, reconstruction and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[14] T. Totsuka and M. Levoy, "Frequency domain volume rendering," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, 1993.

[15] T. Malzbender, "Fourier volume rendering," *ACM Transactions on Graphics (ToG)*, 1993.

[16] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[17] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-NeRF 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[18] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-NeRF: Structured view-dependent appearance for neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[19] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[20] Q. Wang, Z. Wang, K. Genova, P. P. Srinivasan, H. Zhou, J. T. Barron, R. Martin-Brualla, N. Snavely, and T. Funkhouser, "IBRNet: Learning multi-view image-based rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[21] C. Gao, A. Saraf, J. Kopf, and J.-B. Huang, "Dynamic view synthesis from dynamic monocular video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[22] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-NeRF: Neural radiance fields for dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[23] J. Cen, Z. Zhou, J. Fang, W. Shen, L. Xie, D. Jiang, X. Zhang, Q. Tian *et al.*, "Segment anything in 3D with NeRFs," in *Advances in Neural Information Processing Systems*, 2023.

[24] C. Wang, M. Chai, M. He, D. Chen, and J. Liao, "Clip-NeRF: Text-and-image driven manipulation of neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[25] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "TensoRF: Tensorial radiance fields," in *European Conference on Computer Vision*, 2022.

[26] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics*, 2022.

[27] J. Li, J. Zhang, X. Bai, J. Zheng, X. Ning, J. Zhou, and L. Gu, "DNGaussian: Optimizing sparse-view 3D Gaussian radiance fields with global-local depth normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[28] D. Charatan, S. L. Li, A. Tagliasacchi, and V. Sitzmann, "pixelSplat: 3D Gaussian splats from image pairs for scalable generalizable 3D reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[29] A. Guédon and V. Lepetit, "SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[30] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma, "GaussianShader: 3D Gaussian splatting with shading functions for reflective surfaces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[31] S. Niedermayr, J. Stumpfegger, and R. Westermann, "Compressed 3D Gaussian splatting for accelerated novel view synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[32] J. C. Lee, D. Rho, X. Sun, J. H. Ko, and E. Park, "Compact 3D Gaussian representation for radiance field," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[33] R. van Schyndel, A. Tirkel, and C. Osborne, "A digital watermark," in *Proceedings of 1st International Conference on Image Processing*, 1994.

[34] S. Haddad, G. Coatrieux, A. Moreau-Gaudry, and M. Cozic, "Joint watermarking-encryption-JPEG-LS for medical image reliability control in encrypted and compressed domains," *IEEE Transactions on Information Forensics and Security*, 2020.

[35] C. Wang, J. Ni, and J. Huang, "An informed watermarking scheme using hidden markov model in the wavelet domain," *IEEE Transactions on Information Forensics and Security*, 2012.

[36] X. Kang, J. Huang, and W. Zeng, "Efficient general print-scanning resilient data hiding based on uniform log-polar mapping," *IEEE Transactions on Information Forensics and Security*, 2010.

[37] Y. Wang, Q. Ying, Y. Sun, Z. Qian, and X. Zhang, "A DTCWT-SVD based video watermarking resistant to frame rate conversion," in *International Conference on Culture-Oriented Science and Technology*, 2022.

[38] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "HiDDeN: Hiding data with deep networks," in *Proceedings of the European Conference on Computer Vision*, 2018.

[39] S. Baluja, "Hiding images within images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[40] Q. Ying, X. Hu, X. Zhang, Z. Qian, S. Li, and X. Zhang, "RWN: Robust watermarking network for image cropping localization," in *IEEE International Conference on Image Processing*, 2022.

[41] Y. Zeng, J. Tan, Z. You, Z. Qian, and X. Zhang, "Watermarks for generative adversarial network based on steganographic invisible backdoor," in *IEEE International Conference on Multimedia and Expo*, 2023.

[42] C. Zhang, P. Benz, A. Karjauv, G. Sun, and I. S. Kweon, "UDH: Universal deep hiding for steganography, watermarking, and light field messaging," in *Advances in Neural Information Processing Systems*, 2020.

[43] Y. Zhou, Q. Ying, Y. Wang, X. Zhang, Z. Qian, and X. Zhang, "Robust watermarking for video forgery detection with improved imperceptibility and robustness," in *IEEE 24th International Workshop on Multimedia Signal Processing*, 2022.

[44] X. Chen, J. Deng, Y. Chen, C. Li, X. Fang, C. Liu, and W. Xu, "IMPRINTS: Mitigating watermark removal attacks with defensive watermarks," *IEEE Transactions on Information Forensics and Security*, 2025.

[45] Y. Tang, C. Wang, S. Xiang, and Y.-m. Cheung, "A robust reversible watermarking scheme using attack-simulation-based adaptive normalization and embedding," *IEEE Transactions on Information Forensics and Security*, 2024.

[46] Y. Zhang, D. Ye, C. Xie, L. Tang, X. Liao, Z. Liu, C. Chen, and J. Deng, "Dual defense: Adversarial, traceable, and invisible robust watermarking against face swapping," *IEEE Transactions on Information Forensics and Security*, 2024.

[47] H. Fang, W. Zhang, H. Zhou, H. Cui, and N. Yu, "Screen-shooting resilient watermarking," *IEEE Transactions on Information Forensics and Security*, 2018.

[48] H. Fang, Z. Jia, Z. Ma, E.-C. Chang, and W. Zhang, "Pimog: An effective screen-shooting noise-layer simulation for deep-learning-based watermarking network," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[49] H. Fang, K. Chen, Y. Qiu, J. Liu, K. Xu, C. Fang, W. Zhang, and E.-C. 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*, 2023.

[50] G. Liu, Y. Si, Z. Qian, X. Zhang, S. Li, and W. Peng, "WRAP: Watermarking approach robust against film-coating upon printed photographs," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023.

[51] X. Xiao, Y. Zhang, Z. Hua, Z. Xia, and J. Weng, "Client-side embedding of screen-shooting resilient image watermarking," *IEEE Transactions on Information Forensics and Security*, 2024.

[52] G. Gao, X. Chen, L. Li, Z. Xia, J. Fei, and Y.-Q. Shi, "Screen-shooting robust watermark based on style transfer and structural re-parameterization," *IEEE Transactions on Information Forensics and Security*, 2025.

[53] X. Luo, R. Zhan, H. Chang, F. Yang, and P. Milanfar, "Distortion agnostic deep watermarking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[54] J. Jing, X. Deng, M. Xu, J. Wang, and Z. Guan, "HiNet: Deep image hiding by invertible network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[55] Z. Guan, J. Jing, X. Deng, M. Xu, L. Jiang, Z. Zhang, and Y. Li, "DeepMIH: Deep invertible network for multiple image hiding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[56] F. Li, Y. Sheng, K. Wu, C. Qin, and X. Zhang, "LiDiNet: A lightweight deep invertible network for image-in-image steganography," *IEEE Transactions on Information Forensics and Security*, 2024.

[57] T. Bui, S. Agarwal, and J. Collomosse, "TrustMark: Universal watermarking for arbitrary resolution images," *arXiv preprint arXiv:2311.18297*, 2023.

[58] S. Lu, Z. Zhou, J. Lu, Y. Zhu, and A. W.-K. Kong, "Robust watermarking using generative priors against image editing: From benchmarking to advances," in *The Thirteenth International Conference on Learning Representations*, 2025.

[59] R. Hu, J. Zhang, T. Xu, J. Li, and T. Zhang, "Robust-Wide: Robust watermarking against instruction-driven image editing," in *European Conference on Computer Vision*, 2025.

[60] R. Xu, M. Hu, D. Lei, Y. Li, D. Lowe, A. Gorevski, M. Wang, E. Ching, and A. Deng, "InvisMark: Invisible and robust watermarking for AI-generated image provenance," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025.

[61] T. Sander, P. Fernandez, A. O. Durmus, T. Furon, and M. Douze, "Watermark anything with localized messages," in *The Thirteenth International Conference on Learning Representations*, 2025.

[62] H. Zhou, K. Chen, W. Zhang, Y. Yao, and N. Yu, "Distortion design for secure adaptive 3-D mesh steganography," *IEEE Transactions on Multimedia*, 2018.

[63] R. Jiang, H. Zhou, W. Zhang, and N. Yu, "Reversible data hiding in encrypted three-dimensional mesh models," *IEEE Transactions on Multimedia*, 2017.

[64] F. Wang, H. Zhou, H. Fang, W. Zhang, and N. Yu, "Deep 3D mesh watermarking with self-adaptive robustness," *Cybersecurity*, 2022.

[65] F. Wang, H. Zhou, W. Zhang, and N. Yu, "Neural watermarking for 3D morphable models," in *International Conference on Artificial Intelligence and Security*, 2022.

[66] X. Zhu, G. Ye, X. Luo, and X. Wei, "Rethinking Mesh Watermark: Towards highly robust and adaptable deep 3D mesh watermarking," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[67] I. Yoo, H. Chang, X. Luo, O. Stava, C. Liu, P. Milanfar, and F. Yang, "Deep 3D-to-2D watermarking: Embedding messages in 3D meshes and extracting them from 2D renderings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[68] Z. Luo, A. Rocha, B. Shi, Q. Guo, H. Li, and R. Wan, "The NeRF signature: Codebook-aided watermarking for neural radiance fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[69] Q. Song, Z. Luo, K. C. Cheung, S. See, and R. Wan, "Protecting NeRFs' copyright via plug-and-play watermarking base model," in *European Conference on Computer Vision*, 2024.

[70] X. Zhang, J. Meng, R. Li, Z. Xu, J. Zhang *et al.*, "GS-Hider: Hiding messages into 3D Gaussian splatting," in *Advances in Neural Information Processing Systems*, 2024.

[71] X. Zhang, J. Meng, Z. Xu, S. Yang, Y. Wu, R. Wang, and J. Zhang, "SecureGS: Boosting the security and fidelity of 3D Gaussian splatting steganography," in *The Thirteenth International Conference on Learning Representations*, 2025.

[72] Y. Jang, H. Park, F. Yang, H. Ko, E. Choo, and S. Kim, "3D-GSW: 3D Gaussian splatting for robust watermarking," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025.

[73] R. Li, X. Zhang, C. Tong, Z. Xu, and J. Zhang, "GaussianSeal: Rooting adaptive watermarks for 3D Gaussian generation model," *arXiv preprint arXiv:2503.00531*, 2025.

[74] Z. Chen, G. Wang, J. Zhu, J. Lai, and X. Xie, "GuardSplat: Efficient and robust watermarking for 3D Gaussian splatting," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025.

[75] X. Xian, G. Wang, X. Bi, J. Srinivasa, A. Kundu, M. Hong, and J. Ding, "RAW: A robust and agile plug-and-play watermark framework for AI-generated images with provable guarantees," in *Advances in Neural Information Processing Systems*, 2024.

[76] J. Kirchenbauer, J. Geiping, Y. Wen, J. Katz, I. Miers, and T. Goldstein, "A watermark for large language models," in *International Conference on Machine Learning*, 2023.

[77] Y. Wen, J. Kirchenbauer, J. Geiping, and T. Goldstein, "Tree-rings watermarks: Invisible fingerprints for diffusion images," in *Advances in Neural Information Processing Systems*, 2023.

[78] G. B. Folland, *Fourier Analysis and Its Applications*. American Mathematical Soc., 2009.

[79] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics*, 2019.

[80] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, 2004.

[81] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.