Hard EXIF: Protecting Image Authorship Through Metadata, Hardware, and Content

Yushu Zhang, *Senior Member, IEEE*, Bowen Shi, Shuren Qi*,
Xiangli Xiao, Ping Wang, and Wenying Wen, *Member, IEEE*

***Abstract*—With the rapid proliferation of digital image content and advancements in image editing technologies, the protection of digital image authorship has become an increasingly important issue. Traditional methods for authorship protection include registering authorship through certification organization, utilizing image metadata such as Exchangeable Image File Format (EXIF) data, and employing watermarking techniques to prove ownership. In recent years, blockchain-based technologies have also been introduced to enhance authorship protection further. However, these approaches face challenges in balancing four key attributes: strong legal validity, high security, low cost, and high usability. Authorship registration is often cumbersome, EXIF metadata can be easily extracted and tampered with, watermarking techniques are vulnerable to various forms of attack, and blockchain technology is complex to implement and requires long-term maintenance.**

In response to these challenges, this paper introduces a new framework Hard EXIF, designed to balance these multiple attributes while delivering improved performance. The proposed method integrates metadata with physically unclonable functions (PUFs) for the first time, creating unique device fingerprints and embedding them into images using watermarking techniques. By leveraging the security and simplicity of hash functions and PUFs, this method enhances EXIF security while minimizing costs.

Experimental results demonstrate that the Hard EXIF framework achieves an average peak signal-to-noise ratio (PSNR) of 42.89 dB, with a similarity of 99.46% between the original and watermarked images, and the extraction error rate is only 0.0017. These results show that the Hard EXIF framework balances legal validity, security, cost, and usability, promising authorship protection with great potential for wider application.

***Index Terms*—Authorship protection, metadata, device fingerprint, physical unclonable functions, watermarking.**

Y. Zhang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, also with the School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: yushu@nuaa.edu.cn, zhangyushu@jxufe.edu.cn).

B. Shi is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: shibowen39@nuaa.edu.cn).

S. Qi is with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China (e-mail: shurenqi@cuhk.edu.hk).

X. Xiao and W. Wen are with the School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: wenyingwen@sina.cn; xiaoxiangli@jxufe.edu.cn).

P. Wang is with the School of Engineering, Westlake University, also with Zhejiang University, Hangzhou, Zhejiang 310030, China (e-mail: wangping@westlake.edu.cn).

This work was supported by the Ganpo Talent Program of Jiangxi Province under Grant gpyc20240012 and the Natural Science Foundation of China under Grant 62201233.

(*Corresponding author: Shuren Qi.)

I. Introduction

THE rapid proliferation of digital image content and significant advancements in image editing technologies have markedly transformed the landscape of digital media. As digital imaging devices become increasingly prevalent and accessible, the daily volume of digital images has surged. Concurrently, sophisticated image editing software has empowered users to modify and manipulate digital images easily. While fostering creativity and innovation, this dual evolution has introduced profound challenges in protecting digital image authorship. Impersonation becomes easier and the uniqueness of evidence is threatened, leading to a decrease in the legal effectiveness of common methods of authorship protection. In the event of a dispute over authorship, it becomes more difficult to obtain evidence. Therefore, ensuring the uniqueness and legal interpretability of digital image authorship has become increasingly important.

A core issue in digital image authorship protection is providing indisputable proof of ownership, especially in legal disputes [1]. The initial method of protecting authorship is to register with relevant authoritative third-party organizations and submit relevant materials regarding the author and image information. After the publication period expires and there are no objections, the authorship registration organization will temporarily register the work and issue an authorship registration certificate to the applicant, usually indicating the work name, registration date, author name, and registration number. Finally, the authorship registration agency records and archives the registration information as the legal basis for the authorship of the work, thereby achieving proof of ownership. This type of evidence is highly legally valid and tamper-resistant in a court of law, but it is not a friendly method due to the extremely high time and labor costs required. The fact that it takes months or more from the time a creator submits a registration until they receive proof of registration does not work for creators who need proof in a hurry. For individual creators, this approach is poorly cost-effective and, as the number of images proliferates, it becomes clear that submitting a registration for every single image is inappropriate.

To reduce costs, metadata-based methods for author rights protection have been proposed. Metadata provides a description of an image's content, hardware source, and attributes. Therefore, the shooting information and hardware details embedded in the image metadata can be utilized for author authentication. The main advantages of metadata-based author rights protection methods lie in their simplicity and low cost,

as metadata is inherently stored within the image file. However, with the widespread availability of metadata extraction software, users can easily access plaintext metadata from images, which introduces security vulnerabilities. Metadata can be tampered with freely without leaving any detectable traces, making it susceptible to unauthorized modification. To mitigate these security risks, metadata is often combined with other technologies, such as watermarking [2], which enhances tamper resistance but reduces some of the advantages of metadata, particularly in terms of its simplicity [3] and legal effectiveness.

Digital watermarking [4] is a technology that embeds identification information into multimedia files through algorithms without affecting the original multimedia value and usage. The information embedded in this way is often not directly perceived by users and can withstand certain attacks. Digital watermarking technology can embed author information and other information into original images as evidence in case of copyright disputes. Whether the carrier data is modified or not, watermarking algorithms can be divided into zero watermarking [5] and non-zero watermarking. Zero watermarking refers to the watermark embedding process that does not modify the original image data to ensure the quality of the image is not compromised. Non-zero watermarking corresponds to a watermark embedding algorithm that modifies image data. According to the observability of watermarking, they can be divided into visible [6] and invisible watermarking [7]. Invisible watermarking has good concealment but poor robustness against specific attacks. Watermarking has a certain legal effect when used as evidence and is easy to verify. However, embedding a watermark will impact the original image, and some operations or attacks on the image will also affect the effect of the watermark, which in turn will affect its legal effect. To improve the security of watermarking algorithms, creators choose more complex embedding locations, such as the Speeded-Up Robust Feature (SURF) region, and design more complex processes, but this also increases the overhead. Since the embedding and extraction of watermarking are opposite processes, the corresponding verification process will also be more complicated. If the authorship information or metadata information is directly used as a watermark, there is also the possibility of identity leakage and impersonation. With the advancement of deep learning techniques, a number of deep image watermarking methods [8] [9] have been proposed, offering improved imperceptibility and robustness. However, their practical applicability remains limited due to two key challenges: insufficient resilience against compression-based attacks such as JPEG, and the substantial training and computational resources required for deployment. These constraints, along with the relatively high technical barrier to use, hinder the widespread adoption of deep watermarking in real-world scenarios.

The copyright chain [10]–[12] is a decentralized authorship management system built on blockchain technology, designed to offer a tamper-proof and transparent solution for copyright registration and transactions. Creators can upload their original content (e.g., articles, music, pictures, etc.) in the form of Hash and permanently record the copyright information of the content on the blockchain, including the creator's identity, the time of publication of the work, and the hash value of the work. Once uploaded, this information cannot be tampered with, which ensures that the creator's work can be confirmed and can be used as strong evidence in case of authorship disputes [12]. Furthermore, by integrating smart contracts, the copyright chain automates copyright transactions and authorization processes, reducing associated costs. A core advantage of the copyright chain is its strong resistance to tampering. In traditional centralized copyright registration systems, databases are susceptible to hacking or internal manipulation. However, the blockchain's decentralized node structure and consensus mechanisms ensure that once data is recorded, it cannot be altered; any modification would be detected and rejected by the entire network, significantly enhancing data security. Nonetheless, its implementation poses certain challenges, particularly when handling large volumes of work or frequent transactions, as blockchain networks may incur high storage and computational costs. Additionally, the complexity of blockchain technology may create technical barriers for ordinary users, potentially hindering its widespread adoption.

With the development of authorship protection technology, people have begun to pay attention to the importance of hardware characteristics [13]. Traditional authorship protection technology mainly records hardware information rather than utilizing hardware characteristics. PUF [14]–[16] is a unique hardware-oriented security primitive that does not rely on the complexity of key-based algorithms or intractable mathematical problems as a basis for trust establishment. PUFs utilize subtle mismatches or disturbances in the electrical characteristics of identically designed circuits due to unavoidable and uncontrollable variations in physical parameters during the fabrication of nanoscale devices [17]. PUF can be mathematically modeled as an irreversible mapping of input challenge to output response. Challenge-response pairs (CRPs) are unique for different chips on the same wafer and across wafers, making PUFs an ideal "device fingerprint" [18], which can be used as an important basis for authorship protection.

Overall, metadata technology cannot be directly used as a forensic method due to its inability to resist advanced editing tools and attacks. Although metadata and watermarks are widely used, they may be subject to targeted tampering and deletion, reducing their validity as proof of authorship and leading to reduced legal validity. High-security solutions such as blockchain technology tend to be both expensive and complex, with a high cost of use, while more economical solutions often lack robustness and user-friendliness. When collecting evidence, the evidence chain formed by the above methods did not fully cover the three important pieces of evidence information: metadata, hardware, and image content. The method proposed in this paper achieves authorship protection by utilizing metadata information and hardware characteristics to obtain a unique hardware fingerprint and embedding it into image content. Using the EXIF extension as a PUF challenge, embedding the PUF response as a watermark into the image that needs to be protected, and achieving satisfactory legal validity, security, cost, and usability.

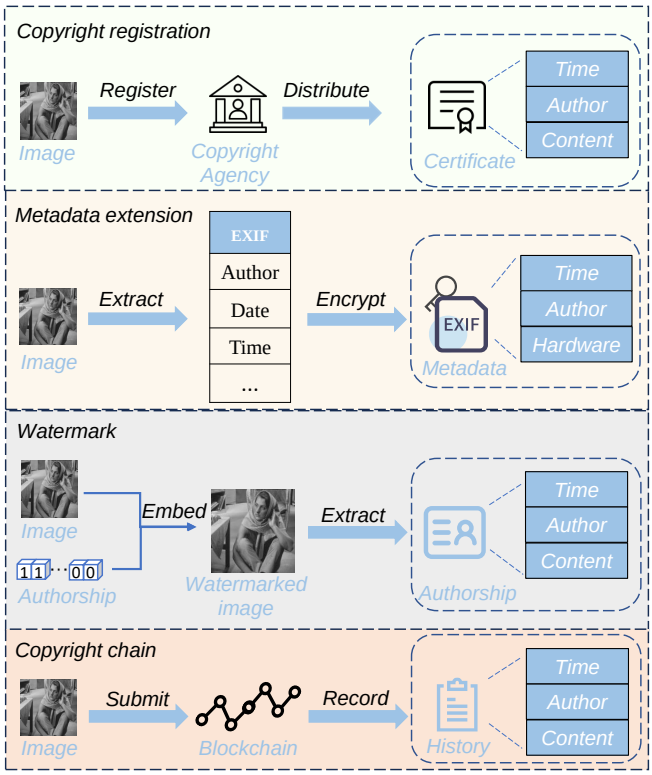


Fig. 1. The four commonly used authorship protection technologies.

We emphasize that the proposed framework is not limited to a specific watermark embedding algorithm or a particular implementation of sensor-based PUF generation. Our objective is to establish a novel and extensible framework that systematically addresses the problem of authorship protection.

Our contributions are:

- We propose a novel and scalable authorship protection framework that, for the first time, integrates metadata, hardware characteristics, and watermarking techniques to generate an accurate and reliable chain of evidence.
- We solve the problem of metadata tampering by using hash functions to extend metadata, providing stable input for subsequent authentication. By using advanced sensor PUFs, utilizing hardware features to provide unique fingerprints, and embedding them into images using watermarking, the security and trustworthiness of the evidence chain are enhanced.
- The proposed framework achieves an average PSNR of 42.89 dB on the dataset, and the average similarity between the original image and the watermark image is 99.46%, with a bit error rate (BER) of only 0.0017. At the same time, it has been verified that it has good robustness and tamper resistance under various attacks, showcasing a complete chain of evidence in case of disputes.

II. Related Work

Firstly, we review the existing authorship protection work and analyze the existing technologies, summarized by Table I and Fig.1.

TABLE I
COMPARISON OF EXISTING AUTHORSHIP TECHNOLOGIES. ↑ INDICATES THE HIGHEST DEGREE, → INDICATES A MODERATE DEGREE, AND ↓ INDICATES THE LOWEST DEGREE.

	Legal effect	Tamper resistant	Cost	Usability
Copyright registration	↑	↑	↑	↑
Metadata extension	→	↓	↓	↑
Digital watermark	→	→	→	→
Copyright chain	→	↑	↑	↓
Hard EXIF	↑	↑	→	↑

A. Copyright Registration

Copyright registration is a formal method for protecting digital image rights by gaining legal recognition through a copyright authority. This process involves submitting detailed information about the work, including creation details, author identification, and ownership declarations. The primary advantage of copyright registration is its strong legal validity, widely accepted across global legal systems, making it a powerful tool in legal disputes. However, the process can be costly and time-consuming, involving application, documentation, and potential legal fees, which may limit accessibility for individual creators and small enterprises. The complexity and time required for registration may also deter creators from using this method, and verifying registration during legal disputes can also be labor-intensive.

B. Metadata Extension

Metadata [19] [20] describes the content, source, and attributes of a file. In digital images, metadata is typically stored in the file's header or footer, such as EXIF data in JPEG formats. EXIF, a widely recognized standard, includes details like camera settings, shooting parameters, and author identification. Modern cameras and smartphones often add geographic location information (GPS coordinates) when capturing images. In digital forensics, EXIF data can verify the time, location, and equipment used to capture an image, aiding in confirming its authenticity and integrity. However, the reliability of EXIF data as legal evidence is limited. The widespread availability of tools to modify EXIF data undermines its credibility, and courts are cautious in accepting it as standalone proof, often requiring additional corroborative evidence. Moreover, most modern devices capture only limited EXIF information, and users frequently overlook its importance. When images are uploaded or processed, EXIF data can be stripped away, reducing its value as evidence. For security, EXIF data is vulnerable to tampering, and while cryptographic digital signatures or blockchain methods [21] have been proposed, they face compatibility issues with editing tools and privacy concerns. Integrating these security measures directly into camera hardware remains a challenge. From a cost perspective, using metadata for authorship protection is relatively economical since it is automatically generated. However, its security vulnerabilities [22] and limited legal acceptance may render it less viable, particularly when considering the costs of verifying tampered metadata during legal disputes. In terms of usability, EXIF metadata is easily accessible

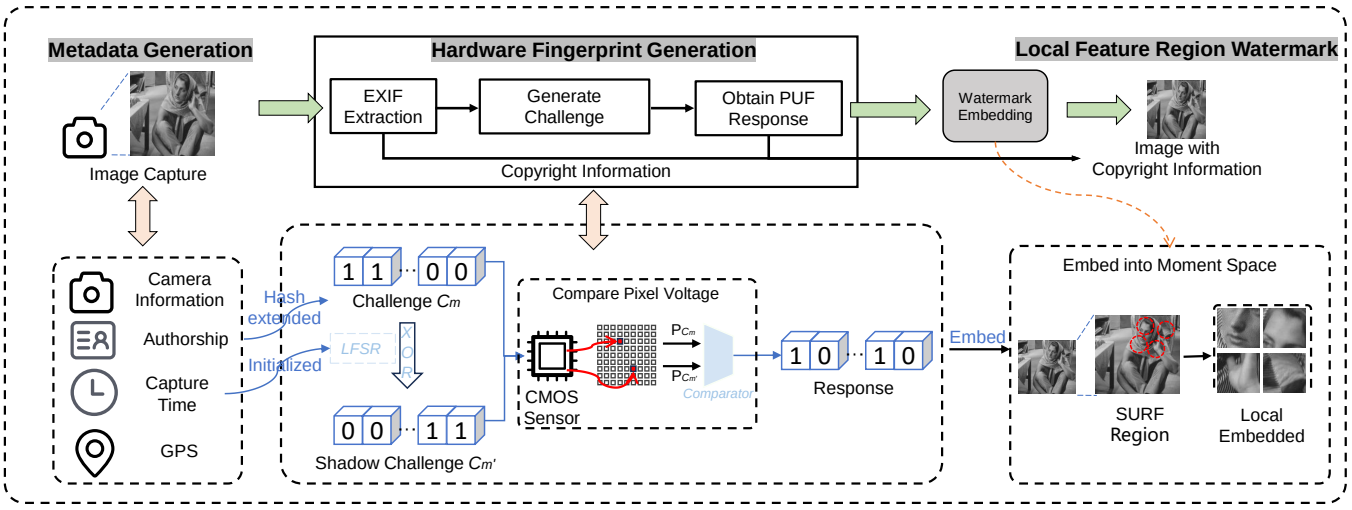


Fig. 2. The framework of Hard EXIF

with standard software but is also easily modified, which compromises its reliability. While a basic understanding of EXIF data is straightforward, verifying its authenticity in legal contexts can be complex and may require forensic expertise, limiting its practicality for robust authorship protection.

C. Digital Watermark

Digital watermarking is a widely adopted technology for embedding hidden information into digital content, serving purposes like authorship protection, information security, and data integrity verification. It is categorized into visible [23] and invisible [24] watermarking, each addressing different needs. Visible watermarks are used for explicit copyright notices by embedding logos or text into content, while invisible watermarks embed imperceptible information to protect content without altering its appearance.

Watermarking techniques are further divided into time-domain and frequency-domain approaches. Time-domain watermarking directly embeds and extracts watermarks from the original pixel data using simple algorithms but suffers from limited robustness. Frequency-domain watermarking, on the other hand, embeds watermark information within frequency-domain coefficients using methods like Discrete Cosine Transform (DCT) [25], [26] and Discrete Wavelet Transform (DWT) [27], offering enhanced resistance to attacks and greater stability. Advanced methods embed watermarks into image moments—mathematical [28] constructs capturing the shape and geometry of images—using techniques like Zernike moments (ZM) [29], Polar Harmonic Transform (PHT) moments [30] [31], which provide higher resilience against various attacks. With the advent of deep learning, invisible watermarking techniques leveraging machine learning have gained prominence. Deep Convolutional Neural Networks [8], [9], [32] enable watermark embedding that maintains visual quality while ensuring detectability by trained networks. Generative Adversarial Networks (GANs) [33] [34] further improve watermark concealment through adversarial training. However, these advanced techniques require significant training resources and are vulnerable to white-box attacks. More recently, watermarking research has extended into AI-generated content (AIGC), such as text-to-image [35], text-to-audio [36], and video synthesis [37]–[39], where signals are embedded during generation to indicate model identity or prompt ownership. These approaches aim to trace content back to the generative model, offering attribution in purely digital pipelines. Future watermarking frameworks may need to support cross-model [40] embedding and verification mechanisms, share attribution information among multiple modalities [41]–[43], and establish trusted anchors at the generation model or device end.

The legal validity of digital watermarking, particularly for invisible watermarks and those generated by deep learning models, is constrained by current legal frameworks. While visible watermarks are more readily accepted as evidence, their visual representation of personal information introduces additional challenges. Images with visible watermarks bring visual modifications to the image itself as well, and visible watermark removal is currently possible using deep learning techniques. Invisible watermarks, on the other hand, typically require supplementary verification to attain legal enforceability. The security of digital watermarks is contingent upon the robustness of embedding techniques against image manipulation and attacks. Matrix-based and deep learning techniques generally offer enhanced security as they integrate the watermark more deeply into the image structure or leverage complex models, making removal more challenging. However, advanced hybrid attacks [44] may still compromise these watermarks.

In terms of cost, time-domain techniques [45] are cost-effective due to their simplicity, whereas deep learning approaches incur higher costs owing to their complexity and resource demands. The availability of watermarking methods also varies: spatial domain-based watermarking techniques, such as the Least Significant Bit (LSB) [46] technique, are easier to implement. The use of deep learning methods, on the other hand, requires specialized knowledge and tools. While

deep watermarking can often maintain visual quality, the embedding and verification process, especially in the context of forensics, can be time-consuming and technically demanding, limiting its usefulness in large-scale author protection.

D. Copyright Chain

The copyright chain is a blockchain-based system [11], [47]–[49] for digital authorship protection, offering decentralized, tamper-resistant, and transparent management [50], [51]. It uses blockchain's distributed ledger and smart contracts to record copyright information securely, ensuring the authenticity and traceability of records. The process begins with the creator generating a digital fingerprint and relevant authorship details [21], which are then recorded on the blockchain, creating an immutable copyright identifier. Any subsequent actions, such as authorizations or transfers, are also recorded, enabling full traceability. While the decentralized and untamperable nature of blockchain supports legal validity, challenges remain in the legal acceptance of blockchain-based evidence. In terms of security, blockchain provides strong tamper resistance, but concerns remain about scalability and performance. Concerning cost, the copyright chain reduces reliance on third parties, lowering overall costs. However, the initial setup and maintenance of the system may be costly. Usability is also a challenge, as managing blockchain transactions and smart contracts requires specialized knowledge. Despite these challenges, the copyright chain shows promise for the future of authorship management, pending further legal recognition and technical advancements.

III. The Proposed Approach: "Hard EXIF"

The Hard EXIF is composed of three key stages, as illustrated in Fig.2. Firstly, metadata is hashed, which is a challenge putting into the CMOS sensor PUF. Second, the unique response is obtained through the properties of PUF. Finally, the response is embedded into the feature region of the image by using watermark technology, which realizes the triple utilization of the image metadata, content, and hardware.

A. Metadata Processing

When an image is captured, EXIF data is automatically generated by the digital camera or smartphone and embedded into the image file. The process begins when the shutter button is pressed. The camera captures the light signal received by the image sensor and converts it into digital image data. Simultaneously, the camera's firmware collects various details related to the capture, including the make and model of the camera, the date and time of the shot, exposure time, aperture value, ISO sensitivity, focal length, flash status, and geographic location (if GPS is supported). These details are organized into a structured block of metadata, known as EXIF data, which is embedded in the image file's header. Although other metadata formats exist, such as the International Press Telecommunications Council (IPTC) and Extensible Metadata Platform (XMP), our framework primarily utilizes EXIF data but can be extended to incorporate additional formats.

Once the EXIF data is generated, it is processed using a hash function to produce a fixed-length hash value, ensuring its integrity and uniqueness. A hash function converts an input of arbitrary length into a fixed-length output, typically resulting in an n-bit hash value (in this paper, a 256-bit hash). The result is a unique EXIF hash code that is subsequently used for hardware binding. The discussion of hash length will be given in the experimental section.

B. Binding Hardware Information Using Sensor PUFs

To avoid the complexities of traditional image processing techniques, our approach leverages CMOS image sensor-based PUFs that utilize the inherent FPN of the sensor to generate unique digital signatures for each device. FPN arises due to manufacturing variations in the sensor's pixel array and readout circuits. Although FPN is typically divided into PRNU [52] [53] and DSNU, we focus primarily on DSNU [14] [15] [54] for its ability to generate reliable signatures even in low-light conditions.

Our PUF design is inspired by the methodology presented in [15]. Initially, a circuit switch is closed to reset the capacitor and sample the input offset of the operational amplifier. Following this, the reset pixel is activated, and its voltage is sampled. The amplifier's output, the difference between the reset and signal voltages, forms the basis of the PUF's digital signature. During PUF operation, the reset pixel output voltage is read directly. To enhance the precision and reliability of this design, we integrated a bypass transistor in parallel with the correlated double-sampling circuit. As mentioned earlier, this configuration allows PUF mode to bypass CDS, which is crucial for maintaining the randomness and uniqueness of the PUF response. In normal sensing mode, the bypass transistor is off, and the CDS functions by subtracting the reset signal from the capacitor at the column level, thereby reducing FPN. Due to the lossless readout capability of CMOS image sensors, operating the image sensor in PUF mode does not affect its original functionality. Compared to approaches that rely solely on metadata and hash functions, the incorporation of a sensor PUF enables all verification procedures to be executed internally within the sensor hardware. This provides intrinsic tamper resistance without the need for external storage or validation, ensuring a streamlined and secure authorship verification mechanism.

In our implementation, as illustrated in Fig.2, the EXIF data generated during image capture is encoded to form a hash sequence C_m, which serves as the challenge for the PUFs. The address decoder decodes C_m to retrieve pixel voltages, while an additional internal challenge, $C_{m'}$, is generated by an n-bit Linear Feedback Shift Register (LFSR), initialized with a user-selected n-bit seed N $(0 < N < 2^n)$, typically derived from the timestamp T_{stamp} at the time of image capture.

The LFSR generates the internal challenge pairs C_m and $C_{m'}$, which are applied to the CMOS image sensor array to localize a pair of active pixel sensors. The corresponding reset voltages P_{C_m} and $P_{C_{m'}}$ are read by disabling the associated dual sampling circuit and then compared to generate the response bit w (watermark). The output bit w is determined

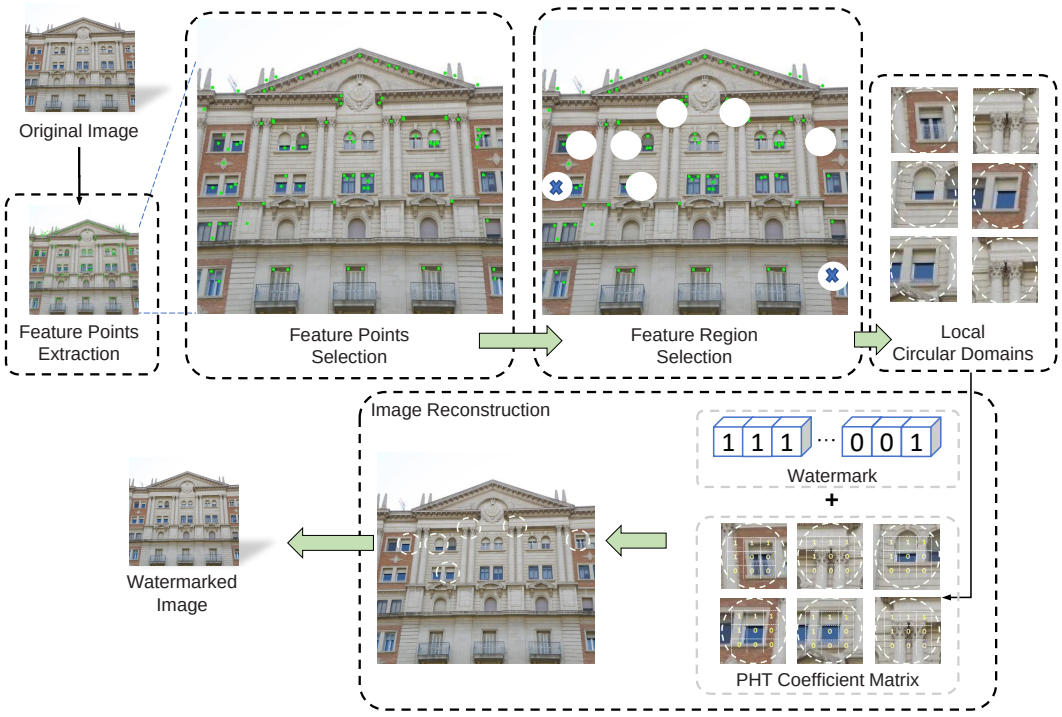


Fig. 3. The proposed watermarking method for binding hardware information (PUF response).

as 0 or 1 based on whether the difference between P_{C_m} and $P_{C_{m'}}$ exceeds the empirically determined threshold P_{th}. If the difference surpasses P_{th}, the bit is considered stable and retained as the response bit. Otherwise, the LFSR generates a new challenge, repeating the process until a stable CRP is identified or the pixel matrix is fully utilized. The threshold P_{th} is calibrated to adjust noise tolerance, ensuring stable response bits despite variations in temperature, voltage, and time. This unique CRP mapping for each PUF instance guarantees a high degree of uniqueness, enabling precise differentiation of individual cameras across different models and brands.

C. Embedding Hardware Fingerprints

Watermark embedding is divided into two steps. Firstly, SURF features are used to select feature regions, and secondly, watermark embedding in PHT moments is performed on these feature regions. The flowchart has been provided by the Fig.3.

1) Feature Region Selection: SURF [55] is a highly efficient local feature detector and descriptor that builds upon the Scale Invariant Feature Transform (SIFT) [56]. SURF offers significant advantages in terms of scale, translation, illumination, contrast, and rotation invariance, outperforming SIFT and other widely used feature extractors. SURF identifies feature points by locating the extrema of the Hessian matrix determinant within the scale space. The resulting feature descriptors are both rotationally and scale invariant. To enhance computational efficiency, the SURF algorithm employs a box filter rather than a Gaussian filter, allowing the matrix determinant to be expressed as

$$\det\left(H_{\text{approx}}\right) = D_{xx}D_{yy} - \left(0.9D_{xy}\right)^2, \tag{1}$$

where D_{xx} is the convolution of the approximate Gaussian second-order derivative with the image I at the point $P = (x, y)$, and $D_{yy}D_{xy}$ are similar. To generate the feature description, the SURF extracts 4 × 4 rectangular blocks of regions around the feature point, and each sub-region counts 25 pixel-points of Haar wavelet features in the horizontal and vertical directions, which are the sum of values in the horizontal direction, the sum of values in the vertical direction, the sum of values in the horizontal direction in absolute terms, and the sum of values in the vertical direction in absolute terms in 4 directions. These 4 values are taken as the feature vectors for each sub-block region. For each pixel in the image, SURF calculates the response value of that pixel at its scale (and adjacent scales). SURF compares the response values of each pixel with its neighboring pixels to determine if it is an extremum point.

In the proposed method, we select feature regions by identifying several optimal feature points based on SURF feature points. These optimal feature points are prioritized by sorting them according to their response values, which are equal to the absolute value of the determinant of the Hessian matrix, ensuring the prominence and stability of these feature points within the image. For each selected feature point, an appropriate radius is specified, centering the circular region on the feature point. The radius is typically chosen based on image resolution and practical requirements, ensuring the circular region encompasses a sufficient number of pixel points to characterize the local features of the feature point effectively.

2) Embedding of Robust Watermarks and Reconstruction of Watermarked Image: In practice, watermarking methods

Fig. 4. Dataset and visualization of experimental results: These figures show the entire process from image capture to embedding of watermarked images in feature regions from top to bottom.

should be able to handle conventional signal processing operations and geometric distortions. Invariant domain-based watermarking is a method that has been proposed in recent years to achieve this goal.

The PHT is a new kind of orthogonal moment defined on the circular domain. The magnitudes of PHT are invariant to image rotation and scaling. Compared to ZM, the computation cost of PHT is extremely low. Besides, the PHT is free of numerical instability issues so that high-order moments can be obtained accurately. As a result, we believe PHT is more suitable for watermarking. PHT is a generalized name for the Polar Complex Exponential Transform (PCET), Polar Cosine Transform (PCT), and Polar Sine Transform (PST). They have been grouped under the name PHT because their kernels are basic waves and trigonometric functions.

We use PCET as an example. First, we determine the maximum order N and repetition M, which are two integers greater than zero. The order n and repetition m of each PCET moment A_{nm} should be satisfied: $-N \leq n \leq N$and $-M \leq m \leq M$. The general orthogonal moments are defined by projecting the image onto the orthogonal kernel function, which is denoted $V_{nm}(\rho, \theta)$ in this paper. In the unit circle domain, the orthogonal kernel consists of a radial component and a circular component:

$$V_{nm}(\rho, \theta) = R_n(\rho)e^{im\theta} = e^{i2\pi n\rho^2}e^{im\theta}, \tag{2}$$

where $R_n(\rho) = e^{i2\pi n\rho^2}$ is the radial part, $e^{im\theta}$ is the circular part. Furthermore, the kernels satisfy the following orthonormal conditions as

$$\int_0^1 R_n(\rho)[R_{n'}(\rho)]^* \rho d\rho = \frac{1}{2}\delta_{nn'}, \tag{3}$$

$$\delta_{nn'} = \begin{cases} 1, & n = n', \\ 0, & n \neq n', \end{cases} \tag{4}$$

where $[R_{n'}(\rho)]^*$ is the conjugate of $R_n(\rho)$.

For image $f(\rho, \theta)$, the PCET moment of order n with repetition m is

$$A_{nm} = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \left[e^{i2\pi n\rho^2}e^{im\theta}\right]^* f(\rho, \theta)\, \rho\, d\rho\, d\theta. \tag{5}$$

PHT is defined for analog images, for digital images, the moments can only be obtained approximately. Therefore, we need to choose the exact moments. Wang et al. [31] showed that the moment of repetition $m = 4j$ is unstable. The orthogonality of the PCET moments is biased when j is an integer and cannot be computed accurately from a discrete image, and for conjugate moment pairs, only moments with an orthogonal order or repetitions are used and the same operation has to be done on the conjugate moments after the watermark embedding. We take $C = \{A_{nm},\ n \leq N, m \geq 0,\ m \neq 4j\}$ as the set of eligible PCET moments for embedding the watermark.

It is important to note that some moments are conjugate. From the definition of PHT, we know that for PCET moments, the conjugate moments are: $A^*_{nm} = A_{-n-m}$. Due to the orthogonal property of PCET, an image can be expressed in terms of the moments, namely image reconstruction from PCET. Based on the principle of amplitude embedding, we can reconstruct the image with the watermark without affecting the quality of the image itself by

$$f(\rho, \theta) = \sum_{n=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} A_{nm}V_{nm}(\rho, \theta). \tag{6}$$

To achieve a better watermark effect, this paper imitates the advanced Quantization Index Modulation (QIM) watermark embedding method in [30], one-bit watermark w is embedded into the magnitude of PCET moment $A_{n,m}$ as

$$\left|A^{Pw}_{n_i,m_i}\right| = \begin{cases} Q(A^P_{n_i,m_i}, s) \times s + \frac{3}{4}s + D_e, & \text{if } w_i = 1 \\ Q(A^P_{n_i,m_i}, s) \times s + \frac{1}{4}s + D_e, & \text{if } w_i = 0 \end{cases} \quad (7)$$

where $A^P_{n_i,m_i} = A_{n_i,m_i} \times T, i = 0, 1, ..., l$. Based on experience, T is generally taken as 100. s represents quantization step length and is a positive even integer greater than 0 and $Q\left(\left|A^{Pw}_{n_i,m_i}\right|, s\right) = \left\lfloor \frac{\left|A^{Pw}_{n_i,m_i}\right|}{s} \right\rfloor$. $D_{e_i} = \left|A^P_{n_i,m_i}\right| - \left\lfloor\left|A^P_{n_i,m_i}\right|\right\rfloor$is the decimal part of $A^P_{n_i,m_i}$, which keeps unchanged in the embedding process. We identify the image reconstructed by the modified PHT moments and their conjugates as I_{rw}, which is computed as

$$I_{rw} = \sum_{i=1}^{L} \Big[(A^w_{n_i,m_i} - A_{n_i,m_i}) V_{n_i,m_i} + \left(A^w_{n_i,m_i} - A_{n_i,m_i}\right) V_{-n_i,m_i} \Big]. \quad (8)$$

Finally, adding this impact on local area I of the original cover image, the watermarked local area I_w can be obtained as $I_w = I + I_{rw}$.

3) Watermark Extraction and Verification: After embedding the watermark, we can get an embedded image with a unique fingerprint, the fingerprint coding itself is meaningless, and the attacker can not infer any information. However, hash and PUFs, which are two strong one-to-one irreversible maps, are bound by the authorship.

Due to both intentional and unintentional attacks, the transmitted information through each channel may be interfered with by different types of transmission noise. During the detection process, we claim the presence of a watermark in image I_w if at least two copies of the embedded watermark are correctly detected. Similarly, the image is first subjected to the same circular region feature extraction as in embedding to obtain the SURF feature points, the response value ranking is performed to generate the circular region, the PCET within the region are extracted, and the claimed moments $A_{n'_i,m'_i}$ are screened with the same criteria as used for screening the moments in embedding as

$$w_i = \begin{cases} 1, & \text{if } G_{n'_i,m'_i} - Q(|A_{n'_i,m'_i}|, s) \times s \geq \frac{1}{2}s, \\ 0, & \text{if } G_{n'_i,m'_i} - Q(|A_{n'_i,m'_i}|, s) \times s < \frac{1}{2}s, \end{cases} \quad (9)$$

where $G_{n'_i,m'_i} = \lfloor ||A_{n'_i,m'_i}| - \alpha| \rfloor + \alpha$ and $\alpha = \text{mod}(s, 4)/4$. Due to the multiple feature regions being embedded, the extracted watermark information may vary. For this reason, the extracted watermark information is corrected by a voting mechanism. This method not only improves the robustness and concealment of the watermark but also ensures the stability and reliability of the watermark information in image processing and geometric transformations, providing an efficient and reliable technical means for image authorship protection. Subsequently, when verifying the authorship identity, the EXIF information of the original image containing the watermark can be outputted as a response through the same process and compared with the extracted watermark to determine the authorship identity.

TABLE II
PSNR (IN DB) OF DIFFERENT QUANTIFY STEP s

Image	Step s					
	s=34	s=36	s=38	s=40	s=42	s=44
Lena	43.1	42.3	42.1	40.8	41.0	40.3
Barbara	42.8	42.1	43.2	41.6	41.0	40.9
Peppers	42.7	42.1	42.1	41.7	41.2	40.8
Car	48.2	48.8	49.6	49.2	49.0	48.8
Building	44.9	45.2	45.8	45.6	44.5	44.2

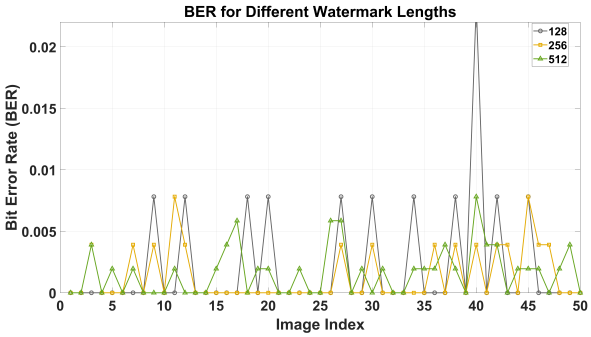


Fig. 5. The impact of different EXIF extension lengths (watermark capacity) on Hard EXIF framework. The average BER of short EXIF lengths is closer to 0, but in some images, the BER is high, while the average BER of long EXIF is high but remains within an acceptable range.

IV. Evaluation and Discussion

A. Experiment Setup

In this section, we comprehensively analyze the Hard EXIF framework from the legal effect, tamper resistance, cost, and usability. Since legal validity is guaranteed by concealment and anti-tamper ability, our experiment focuses more on these two parts. The experimental dataset uses the Google Universal Image Embedding (GUIE) dataset. GUIE has a total of 130000 images, including real pictures of clothing, art, cars, and landmarks, which meet our requirements for authorship protection of real scenes. In this paper, we select 10K images from the dataset for experimentation and use them as the test set D_{test} for testing. Fig.4 shows the test dataset and process images. To demonstrate the generality of the framework, an image is presented for each type of input.

B. Concealment

For the proposed framework, the key parameters related to concealment include 1) the extended hash length L of EXIF is 256 bits, which is also used as the watermark capacity. The impact of different watermark lengths on the framework is shown in Fig.5, and the impact on security is analyzed later. 2) The feature area radius for watermark embedding is 32. In Fig.6, it can be seen that although the small radius has good concealment, the BER is high, and it does not have an advantage in robustness and capacity. 3) The quantify step s

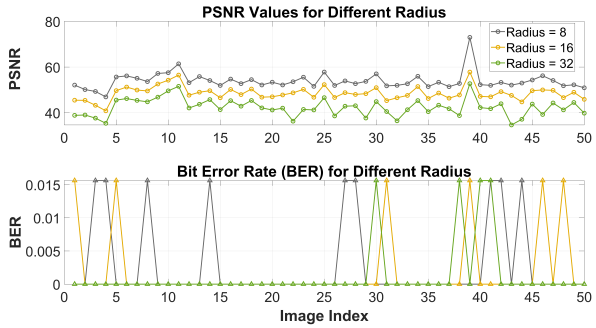


Fig. 6. The impact of different radii on Hard EXIF framework. A smaller radius has better image quality but a higher BER. Watermarked images with a larger radius have a lower BER and better robustness.

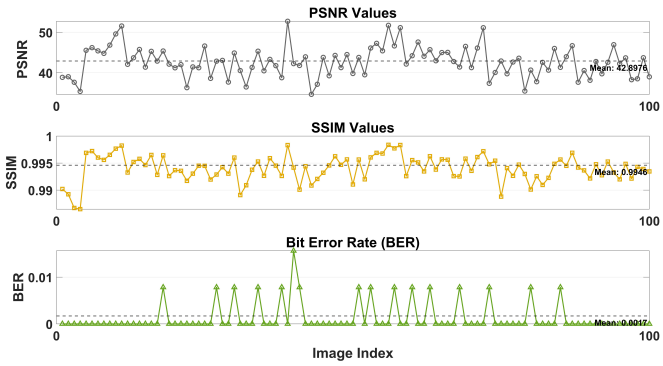


Fig. 7. Concealment assessment: PSNR, SSIM, and BER of the dataset.

is set to 38 as shown in Table II. 4) The maximum order N, which determines the embedding capacity and computation complexity of PCET is 50. 5) For robustness and invisibility, we set the number of embedded circular areas Q to 8.

To evaluate the proposed method, we introduce PSNR and Structural Similarity (SSIM). For the given dataset D_{test}, we obtain an average PSNR of 42.89, SSIM of 99.46%, and BER of 0.0017, as shown in Fig.7, indicating that the proposed method achieves high concealment without affecting image quality.

C. Anti-tamper Analysis

For Hard EXIF, potential threats include tampering and imitation of metadata, PUF attacks, and attacks on the watermark.

1) Metadata Security Analysis: The attacker attempts to obtain the same hash vector by impersonating images and carefully designing EXIF. To evaluate this difficulty, we use information entropy $H(x)$ to measure the randomness and uncertainty of data, which is defined as

$$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \tag{10}$$

where $p(x_i)$ is the probability of the symbol x_i appearing. For EXIF extension data (i.e., hash data) of length L bits, if we assume that each bit is an independent and uniformly distributed binary bit, then the information entropy is $H(x) = L * (-0.5\log_2 0.5 - 0.5\log_2 0.5) = L$. The difficulty of this attack mainly depends on the value of L.

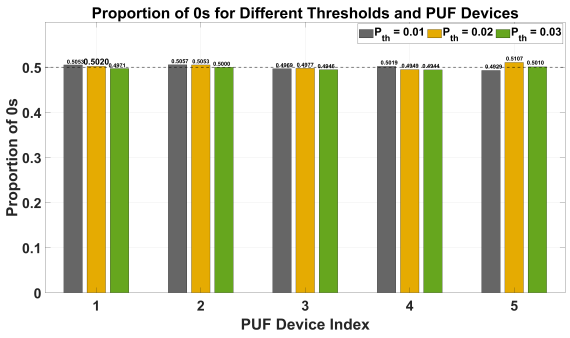


Fig. 8. Uniformity: Measuring the randomness of each device by recording the average of 0 and 1. The closer it is to 50%, the better the randomness.

In practical situations, due to the input space of hash functions being much larger than the output space, there must be situations where different inputs correspond to the same output value, called hash collision. The probability of collision when generating hash values for two different L-bit EXIF data can be estimated through the birthday question. This problem reveals that when a certain number of values are randomly selected, there is a probability that at least two values are the same. In this case, the collision probability G can be approximately expressed as:

$$G \approx 1 - e^{-\frac{L(L-1)}{2\times 2^m}}, \tag{11}$$

where m is the number of bits of the hash output. However, the longer EXIF extension (hash) may not be a better fit for our framework due to the limited watermark capacity. Considering the robustness of the algorithm, we test the different lengths ($L = 128, 256, 512$) of the image shown as Fig.5. It can be intuitively seen from the figure that a BER of 128 bits is higher, while a BER of 512 bits can accommodate more information but is unstable. The length of 256 bits (watermark length) is most suitable for our method.

At the same time, given the hash value, it is not easy to reverse the original input data. This is because the unidirectional nature of the hash function ensures that the attacker can not recover the EXIF data from the hash value, and a small change in the input will result in a significant change in the hash value, ensuring that a small modification to the EXIF data will change the entire hash value.

2) Imitation and Tampering of PUF: The premise of this attack is that the opponent can launch attacks during the PUF stage, imitate and impersonate the hardware used, disrupt the original evidence chain, and gain an advantage in the evidence stage. These attacks are analyzed through the attributes of PUF, the difficulty of the attacks is evaluated, and the tamper resistance of our framework is verified. Our work uses MATLAB to design PUF and generate CRP. Assuming the sensor size is 64×64, 1000 PUFs are simulated, with each PUF generating 240 challenge responses. To better verify the properties, we have set a series of different thresholds $P_{th}(0.01, 0.02, 0.03)$. It should be noted that P_{th} here is used to enhance the security of the framework, meaning that users can regenerate CRPs through P_{th}, which is effective against some malicious attacks.

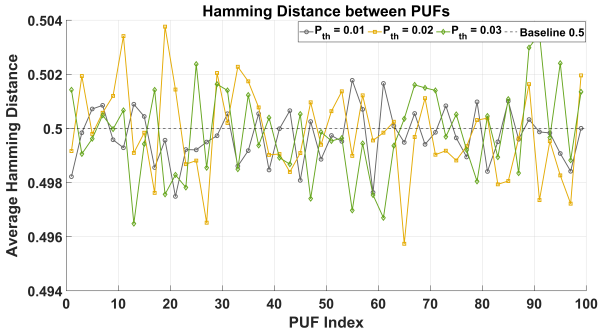


Fig. 9. Uniqueness: Measuring the distance between responses generated by different PUFs under the same challenge.

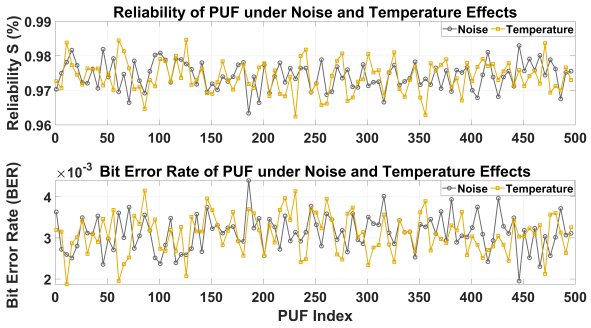


Fig. 10. Reliability: Measuring the distance and error rate of the response to the same challenge under different influences.

Unpredictability. The unpredictability of a PUF assesses the difficulty an attacker faces in predicting its CRP. For a well-designed PUF, the CRP should remain unpredictable for any adversary, even with partial knowledge of existing CRPs. This requires that the correlation between any two CRPs generated by the same PUF is sufficiently low. The unpredictability can be quantified by the entropy of the CRP [57], where the maximum entropy is derived from the number of independent output bits of the PUF, serving as an indicator of its unpredictability. In the case of the proposed sensor PUFs, the sensor comprises $N_{pixel} = H \times V$ pixels, where H and V are the number of rows and columns of the sensor, respectively. Based on the reset voltages, there are $N_{pixel!}$ possible unique orderings of these pixels. Assuming each ordering is equally probable, the number of independent bits can be approximated by $\log_2 N_{pixel!}$.

For the proposed PUFs of a single 64×64 sensor, this yields approximately $\log_2 4096! = 45056$ independent bits, or about 11 bits per pixel, indicating that each pixel contributes roughly 11 bits of entropy. To overcome the PUF by brute-force means, an attacker would theoretically need to test all possible bit combinations, requiring 2^{45056} attempts, an infeasible task. Thus, this analysis demonstrates that predicting the CRP through exhaustive attacks is practically impossible, affirming the resistance of the proposed PUF to cloning and counterfeiting attempts.

Adversaries may also infer attacks from LFSR. The output period of an LFSR depends on the order of its polynomial. An n-bit LFSR has 2^{n-1} states. The adversary needs to try all possible seeds, and this has a complexity of $O(2^n)$ e.g., for a 32-bit LFSR, the complexity is $O(2^{32})$. Second, the complexity is further increased by the variation of timestamps, which are computed when the image is generated, and if the precision is in milliseconds, the adversary needs to try to timestamp every millisecond, which is almost impossible.

Uniformity. In the context of modeling attacks, the assumption is that an adversary can use a given set of CRPs to create a model of the target PUF. With this derived model, other CRPs can be predicted with high accuracy. In the proposed method, it is challenging to derive an additive linear model from the PUF due to the independence of reset voltages among pixels in the pixel array. In PUF outputs with high uniformity, modeling attacks need to deal with more complex patterns because the occurrence of 0s and 1s is unpredictable, and the model needs to learn more complicated features. On the contrary, if the output of PUF is uneven, attackers can capture this bias through simpler models (such as linear models), thereby improving the success rate of modeling.

To ensure the accuracy of the experiment, we test the uniformity F which measures the randomness of each device. It is calculated as the average of all responses from each device as

$$F = \frac{1}{C}\sum_{c=0}^{C} crp_c, \tag{12}$$

where C denotes the number of bits in the response. Ideally, 0 and 1 should be evenly distributed to resist common machine learning modeling attacks. After our test, as shown in Fig.8, the average percentage of 0 is 50.12%, and the percentage of 1 is 49.87% with $P_{th} = 0.01$, which is very close to the ideal value of 50. This high degree of randomness means that PUFs can effectively resist modeling attacks.

In the proposed method, the LFSR is integrated with the image sensor core to encrypt the input challenges and restrict direct access to the CRPs. The input challenges are encrypted using an LFSR-based stream cipher (XOR) to determine the address of the shadow pixel in the selected region. Different seed values in the LFSR generate distinct random numbers, rendering the original CRPs collected by an attacker invalid after a seed change. Additionally, by reconfiguring the properties of the LFSR, such as modifying the characteristic polynomial or EXIF data, adversaries find it significantly more difficult to predict the proposed PUF output using current modeling attack methods.

Uniqueness. Uniqueness measures the difference in response between devices, and Hard EXIF requires sufficient randomness between devices. It is proposed to measure the difficulty of an attack when an adversary has the same hardware but no matching hash vector, i.e. This property can be measured by calculating U as

$$U = \frac{1}{P}\sum_{i=0}^{D}\sum_{j=i+1}^{D}\left(1 - \mathrm{HD}(C_i, C_j)\right), P = \binom{D}{2}. \tag{13}$$

The HD function refers to the normalized hamming distance, C refers to all responses of the devices, and D is the number

TABLE III
ROBUSTNESS COMPARISON OF HARD EXIF VARIANTS USING DIFFERENT WATERMARKING METHODS

Attacks/Variants	Hard EXIF (Ours)	Hard EXIF + Method in [58]	Hard EXIF + Method in [59]
Identity	**0.0017**	0.0128	0.0125
Rotation 1°	**0.0028**	0.0134	0.0050
Rotation 5°	**0.0059**	0.0076	0.0072
Rotation 10°	**0.0067**	0.0098	0.0088
Median Filter	0.0588	0.0731	**0.0551**
JPEG (q=50)	0.0256	0.0120	**0.0085**
JPEG (q=60)	**0.0168**	0.0185	0.0179
JPEG (q=70)	**0.0095**	0.0165	0.0121
Gaussian Noise (σ=0.1)	0.0184	**0.0145**	0.0167
Gaussian Noise (σ=0.15)	**0.0194**	0.0205	0.0225
Gaussian Noise (σ=0.2)	**0.0335**	0.0432	0.0455
Scaling 50%	0.0132	0.0046	**0.0042**
Scaling 90%	0.0073	**0.0011**	0.0020

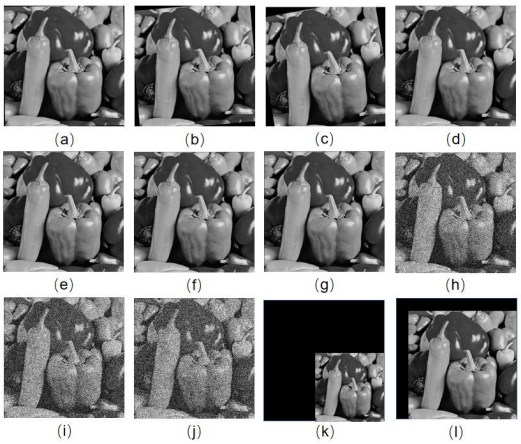


Fig. 11. Pepper images under different attacks : (a) Rotation 1° (b) Rotation 5° (c) Rotation 10° (d) Median filter (e) JPEG compression (q = 50) (f) JPEG compression (q = 60) (g) JPEG compression (q = 70) (h) Gaussian noise (σ = 0.1) (i) Gaussian noise (σ = 0.15) (j) Gaussian noise (σ = 0.2) (k) Scaling 50% (l) Scaling 90%

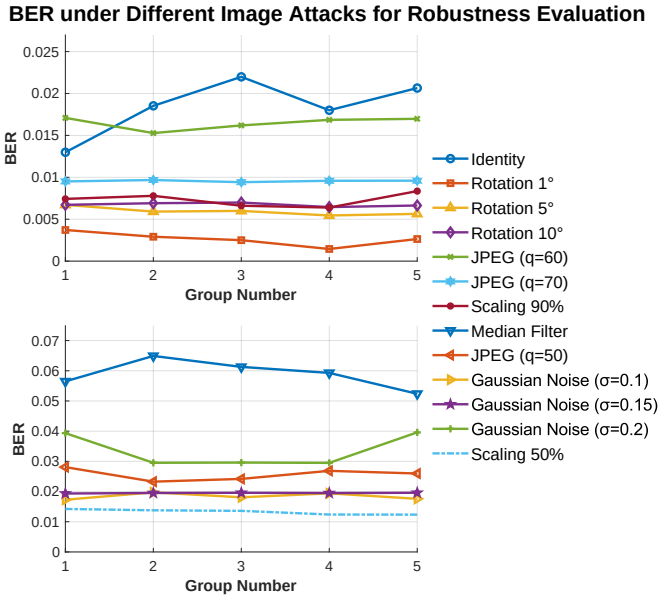


Fig. 12. The robustness performance of the 5 test image groups under Rotation, Median filter, JPEG, Gaussian noise, and Scaling attacks.

of devices. To achieve a balance between robustness and imperceptibility, the ideal value of uniqueness should be close to 0.5, which means that half of the response bit sequences are different between devices.

Through the Monte Carlo simulation, generating CRP from a large number of PUF instances, it is possible to estimate the uniqueness of the PUFs, where each iteration applies a set of PUFs with a unique pixel-voltage matrix to the image sensor. For PUFs with different thresholds, we use the same challenge to obtain different responses for testing. We obtain a uniqueness U of 0.5694 with $P_{th} = 0.01$ in Fig.9, which indicates that the Hamming distance of the response bit sequences between the different devices is slightly higher than the length of the bit sequences by half. It means that the response bit sequences between devices are more distinct, and 56 % of the bits in the response bit sequences of each pair of devices are different. The high Hamming distance indicates that each device's PUF response bit sequence has a high degree of randomness. As can be seen from Fig.9, even the worst distances are close enough to 50 to show that our method is feasible. An attacker cannot infer the responses of other devices from known device responses, much less fully recover the responses.

For brute force reverse cracking, the probability of an attacker succeeding in mimicking the response of an instance is extremely low because the response of each instance is statistically unique. The expected value of the Hamming distance is $64\times0.5694 \approx 36.44$, i.e., the responses of different instances are statistically different by 36.44 bits, which increases the difficulty of the attack.

Reliability. It should be noted that since PUFs realize their non-clonable nature through some physical features, they may also receive physical attacks with effects that can change the generation of CRP, so we introduce reliability to measure the difficulty of this attack. S' measures the reproducibility or stability of a PUF's CRP under different operating conditions. The reliability of a PUF can be measured by its distance, which can be characterized by comparing responses taken at different

TABLE IV
CORRECT EXTRACTION RATES (1-BER) OF DIFFERENT WATERMARKING VARIANTS ON THE COCO DATASET UNDER COMMON DISTORTIONS (MESSAGE LENGTH = 64 BITS).

Distortion Type	**Deep Hard EXIF***	**Hard EXIF + Method in [8]**	**Hard EXIF + Method in [9]**
Identity	99.28%	99.97%	**99.99%**
Cropout (p=30%)	92.82%	**94.78%**	91.96%
Crop (p=30%)	91.37%	**95.25%**	94.81%
Dropout (p=30%)	94.53%	92.92%	**97.99%**
Resize (p=50%)	85.96%	89.41%	**95.53%**
JPEG (q=50)	95.51%	**96.69%**	92.46%
Gaussian Noise (σ=0.1)	**97.98%**	97.12%	96.66%

times to a reference response to the same challenge:

$$S' = 1 - \left(1 - \frac{1}{k}\sum_{j=1}^{k}\frac{\mathrm{HD}(R_i, R_{i,j})}{n}\right) \times 100\%, \qquad (14)$$

where R_i is the standard response bit of the i-th PUF, we apply the same set of challenges k times to the same PUFs, varying the environmental conditions to obtain an n-bit response $R_{i,j}$.

The experimental results indicate an average BER of 0.00155 under noise-only conditions and 0.00154 under temperature-only conditions, as shown in Fig.10. These findings validate the robustness of our differential readout approach against temperature and noise interference, suggesting that physical environmental attacks have minimal impact on the framework. In the reliability assessment, the framework achieved a reliability rate of 97%, further demonstrating its strong tolerance to both noise and temperature variations. This indicates that potential physical attacks are unlikely to compromise the PUF or significantly alter the CRP.

3) Watermark Attack: Due to the irreversibility of PUFs and hash, it is not possible to infer EXIF metadata and PUF responses through watermarking alone. This has been analyzed in the previous section of the attack, and the focus of the attacks in this section is to disrupt the chain of evidence through common image-processing techniques rather than direct forgery or impersonation. We test common image manipulation attacks, including rotation, filtering, noise (σ denotes the standard deviation of the noise), JPEG compression (q denotes the quality factor used in JPEG compression), and scaling. The attacked images are illustrated in Fig.11.

We employ the BER of the extracted watermark after each attack as the robustness evaluation metric. As shown in Fig. 12, the framework achieves a low BER under mild perturbations including Identity (mean: 0.0184, std: 0.0034) and Rotation 1° (mean: 0.0026, std: 0.0008), demonstrating its stability under minor geometric transformations. For Rotation 5° and Rotation 10°, the BERs remain acceptably low (0.0059 and 0.0067, respectively) with minimal variance, indicating resilience against moderate rotation distortions. In contrast, Median filtering, which alters local structures and edge information, results in a higher BER (mean: 0.0589, std: 0.0048), reflecting its stronger destructive impact on feature-region-based watermark embedding. These results confirm both the statistical significance and the stability of the method under diverse attack conditions.

To highlight the adaptability of our framework while evaluating robustness, we conduct a comparative analysis incorporating both traditional watermarking algorithms and deep learning-based watermarking schemes. The distinguishing factor between these variants and the proposed Hard EXIF framework lies exclusively in the watermark embedding and extraction components. Specifically, Method [58] integrates PCET with a logical mapping strategy, whereas Method [59] adopts Fast Quaternion Generic Polar Complex Exponential Transform (FQGPCET) approach. It is worth noting that both methods derive their watermark inputs from response signals generated via hash functions. Experimental results in Table III demonstrate that the proposed strategy achieves higher extraction accuracy than the two traditional variants in the absence of attacks. This improvement is attributed to our selective choice of feature regions and the inherent rotation invariance of the PCET representation. However, the proposed method demonstrates limited robustness against Gaussian noise and median filtering compared to the two variants. This is primarily because such perturbations degrade edge and gradient information in the image, thereby impairing the extraction of SURF feature points and the consistency of feature region matching.

For the results presented in the Table IV, it is important to emphasize that the Deep Hard EXIF variant serves solely as a conceptual demonstration. The Deep Hard EXIF variant adopts an end-to-end deep watermarking framework based on an encoder–decoder architecture with an integrated noise layer. In this framework, the original watermark input is replaced by a 64-bit CRP response, aligning in length with the configurations used in MBRS [8], and CIN [9]. We conduct a comparative evaluation of advanced watermarking techniques MBRS and CIN on the COCO dataset. The noise pool includes Identity, Cropout, Crop, Dropout, Resize, and JPEG compression. Here, p denotes the intensity level of the applied noise.

The tabulated results demonstrate robustness comparable to SOTA methods, along with strong adaptability of the proposed framework. However, as noted in the Introduction, deep learning models inherently require substantial training and computational resources, posing significant challenges for deployment at the sensor level. The objective of our experiment is to illustrate the extensibility of the Hard EXIF variant within such constraints.

TABLE V
TIME COST OF THE PROPOSED SENSOR PUF

Operation	Hardware Module	Latency
Row/Column addressing (select C_m and C'_m)	Row/Column Decoders	< 0.2 μs
Pixel activation and reset voltage reading (P_{C_m}, $P_{C'_m}$)	Sample-and-Hold (S/H) Circuit	< 0.5 μs
Voltage comparison: $\lvert P_{C_m} - P_{C'_m} \rvert$ vs P_{th}	Analog Comparator	< 0.2 μs
Response bit output	Logic gates (XOR + Threshold unit)	< 0.1 μs
Retry if unstable (generate new C'_m)	LFSR + Addressing Logic	< 0.3 μs (if triggered)

D. Cost

From a cost perspective, EXIF metadata management is relatively economical, as it primarily involves modifying or appending data to an image file without requiring additional hardware or complex algorithms. In contrast, digital watermarking incurs costs associated with specialized software for watermark embedding and detection, along with the overhead of preserving watermark integrity. Copyright registration services, particularly those offering formal legal protection, are notably expensive. Copyright hash chains, especially blockchain-based implementations, entail significant upfront costs due to the need to develop and maintain the necessary technical infrastructure. However, their long-term return on investment can be substantial, given the security and immutability they provide.

Our framework capitalizes on the advantages of these existing technologies. EXIF metadata and CMOS-based solutions can be implemented cost-effectively, leveraging existing hardware, while PUF designs can be integrated at the circuit level without the need for additional components. Therefore, the proposed PUF can be easily implemented without affecting or impairing the original function and performance of CMOS image sensors. Device authentication does not require additional expensive security EEPROM/RAM, dedicated encryption modules, or other auxiliary PUF modules. As referenced in [15] [60], the area overhead of a CMOS image sensor-based PUF mainly stems from the inclusion of switch transistors and the CRP generator. A single switch transistor is added per column to bypass the column-level CDS circuitry. The total number of required LUTs and registers [15] are 196 and 27, respectively.

Based on the timing characteristics of each operation, the total latency for generating a single CRP is approximately 1–1.2 μs in Table V. Consequently, a complete 256-bit PUF response—sufficient for robust device-level authentication—can be generated in under 1 ms. This ultra-low latency makes the proposed method well-suited for real-time or resource-constrained environments.

The proposed method employs PCET moments for watermark embedding and extraction, requiring only 1.2 seconds per image. The overall framework demonstrates high computational efficiency and simplicity across both hardware and software layers.

E. Usability

For usability, compared to copyright chain and copyright registration methods, our framework does not have a high threshold for use because metadata and PUF are bound in the hardware stage, and the watermark part is also easy to use due to the blind watermark strategy. At the hardware level, these internal software functionalities can be implemented by hardware manufacturers, eliminating the need for end-users to acquire additional knowledge or skills. Moreover, PUF has been designed as a circuit that can be repeatedly verified during the design process [60], and there are no verification difficulties compared to complex watermark technologies. So Hard EXIF has better usability.

V. CONCLUSION AND FURTHER WORKS

This paper explores the advantages and disadvantages of current authorship protection technologies and analyzes four important attributes: legal effectiveness, tamper resistance, cost, and availability. We find that existing technologies can not balance these characteristics well. In view of this, we propose Hard EXIF, a new authorship protection framework that utilizes metadata, hardware features, and image content to achieve high legal effectiveness, strong resistance to tampering, low usage costs, and good usability, providing a new approach to authorship protection.

With the advent of the AIGC era, achieving identity consistency and ownership traceability across different modalities will emerge as a new challenge. Our proposed framework has the potential to serve as a promising solution to this problem. Therefore, as part of our future work, we will continue to extend the framework and explore its applicability to broader and more diverse scenarios.

REFERENCES

[1] C. Joyce, T. T. Ochoa, M. W. Carroll, M. A. Leaffer, and P. Jaszi, *Copyright law*. Carolina Academic Press Durham, NC, 2016, vol. 85.

[2] H.-C. Huang and W.-C. Fang, "Metadata-based image watermarking for copyright protection," *Simul. Model. Pract. Theory*, vol. 18, no. 4, pp. 436–445, 2010.

[3] S.-C. Pei and Y.-Y. Wang, "Auxiliary metadata delivery in view synthesis using depth no synthesis error model," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 128–133, 2014.

[4] I. Cox, M. Miller, J. Bloom, and C. Honsinger, "Digital watermarking," *J. Electron. Imaging*, vol. 11, no. 3, pp. 414–414, 2002.

[5] C. Wang, X. Wang, Z. Xia, and C. Zhang, "Ternary radial harmonic fourier moments based robust stereo image zero-watermarking algorithm," *Inf. Sci.*, vol. 470, pp. 109–120, 2019.

[6] M. Holliman and N. Memon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 432–441, 2000.

[7] Y. Hu, S. Kwong, and J. Huang, "An algorithm for removable visible watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 129–133, 2005.

[8] Z. Jia, H. Fang, and W. Zhang, "Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 41–49.

[9] 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 *Proc. ACM Int. Conf. Multimedia*, 2022, pp. 1532–1542.

[10] B. Wang, S. Jiawei, W. Wang, and P. Zhao, "Image copyright protection based on blockchain and zero-watermark," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2188–2199, 2022.

[11] L. Xiao, W. Huang, Y. Xie, W. Xiao, and K.-C. Li, "A blockchain-based traceable ip copyright protection algorithm," *IEEE Access*, vol. 8, pp. 49 532–49 542, 2020.

[12] X. Xiao, Y. Zhang, Y. Zhu, P. Hu, and X. Cao, "Fingerchain: Copyrighted multi-owner media sharing by introducing asymmetric fingerprinting into blockchain," *IEEE Trans. Netw. Serv. Manag.*, vol. 20, no. 3, pp. 2869–2885, 2023.

[13] N. N. Rao, P. Thrimurthy, and B. R. Babu, "An efficient copyright protection scheme for digital images using biometrics and watermarking," in *Proc. Int. Sci. Tech. Conf. Computer. Sci. Inf. Tech (CSIT)*. IEEE, 2009, pp. 69–74.

[14] Y. Cao, L. Zhang, and C.-H. Chang, "Using image sensor puf as root of trust for birthmarking of perceptual image hash," in *Proc. IEEE Asian Hardw. Oriented Secur. Trust Symp. (AsianHOST)*. IEEE, 2016, pp. 1–6.

[15] Y. Cao, L. Zhang, S. S. Zalivaka, C.-H. Chang, and S. Chen, "Cmos image sensor based physical unclonable function for coherent sensor-level authentication," *IEEE Trans. Circuits Syst. I-Regul. Pap.*, vol. 62, no. 11, pp. 2629–2640, 2015.

[16] R. Maes and R. Maes, *Physically unclonable functions: Concept and constructions*. Springer, 2013.

[17] Y. Wang, G. Zhang, X. Mei, and C. Gu, "A high-reliability, non-crp-discard arbiter puf based on delay difference quantization," *IEEE Trans. Circuits Syst. I: Regul. Pap.*, 2024.

[18] A. Rullo, C. Felicetti, M. Vatalaro, R. De Rose, M. Lanuzza, F. Crupi, and D. Sacca, "Puf-based authentication-oriented architecture for identification tags," *IEEE Trans. Dependable Secure Comput.*, 2024.

[19] P. Alvarez, "Using extended file information (exif) file headers in digital evidence analysis," *Int. J. Digit. Evid.*, vol. 2, no. 3, pp. 1–5, 2004.

[20] C. Zheng, A. Shrivastava, and A. Owens, "Exif as language: Learning cross-modal associations between images and camera metadata," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. (CVPR)*, 2023, pp. 6945–6956.

[21] R. A. Dobre, R. O. Preda, C. C. Oprea, and I. Pirnog, "Authentication of jpeg images on the blockchain," in *Proc. - Int. Conf. Control, Artif. Intell., Robot. Optim. (ICCAIRO)*. IEEE, 2018, pp. 211–215.

[22] G. Tsudik, "Authorship integrity and attacks," *IEEE Secur. Privacy*, vol. 14, no. 4, pp. 3–5, 2017.

[23] M. S. Kankanhalli, K. Ramakrishnan *et al.*, "Adaptive visible watermarking of images," in *Proc. Int. Conf. Multimedia Comput. Syst.*, vol. 1. IEEE, 1999, pp. 568–573.

[24] M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *Proc. IEEE Int. Conf. Image Process*, vol. 2. IEEE, 1997, pp. 680–683.

[25] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A dct-domain system for robust image watermarking," *Signal Process*, vol. 66, no. 3, pp. 357–372, 1998.

[26] M. Jiansheng, L. Sukang, and T. Xiaomei, "A digital watermarking algorithm based on dct and dwt," in *Proc. - Web Inf. Syst. Appl. Conf., WISA*. Citeseer, 2009, p. 104.

[27] O. Jane, E. Elbaşi *et al.*, "Hybrid non-blind watermarking based on dwt and svd," *J. Appl. Res. Technol.*, vol. 12, no. 4, pp. 750–761, 2014.

[28] S. Qi, Y. Zhang, C. Wang, J. Zhou, and X. Cao, "A survey of orthogonal moments for image representation: Theory, implementation, and evaluation," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–35, 2021.

[29] X.-C. Yuan, C.-M. Pun, and C.-L. P. Chen, "Geometric invariant watermarking by local zernike moments of binary image patches," *Signal Process.*, vol. 93, no. 7, pp. 2087–2095, 2013.

[30] R. Hu and S. Xiang, "Cover-lossless robust image watermarking against geometric deformations," *IEEE Trans. Image Process.*, vol. 30, pp. 318–331, 2020.

[31] X.-y. Wang, Y.-n. Liu, S. Li, H.-y. Yang, P.-p. Niu, and Y. Zhang, "A new robust digital watermarking using local polar harmonic transform," *Comput. Electr. Eng.*, vol. 46, pp. 403–418, 2015.

[32] D. Li, L. Deng, B. B. Gupta, H. Wang, and C. Choi, "A novel cnn based security guaranteed image watermarking generation scenario for smart city applications," *Inf. Sci.*, vol. 479, pp. 432–447, 2019.

[33] J. Fei, Z. Xia, B. Tondi, and M. Barni, "Supervised gan watermarking for intellectual property protection," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.(WIFS)*. IEEE, 2022, pp. 1–6.

[34] J. Huang, T. Luo, L. Li, G. Yang, H. Xu, and C.-C. Chang, "Arwgan: Attention-guided robust image watermarking model based on gan," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–17, 2023.

[35] J. Wang, H. Duan, G. Zhai, and X. Min, "Quality assessment for ai generated images with instruction tuning," 2025. [Online]. Available: https://arxiv.org/abs/2405.07346

[36] J. Wang, H. Duan, G. Zhai, J. Wang, and X. Min, "Aigv-assessor: Benchmarking and evaluating the perceptual quality of text-to-video generation with lmm," 2024. [Online]. Available: https://arxiv.org/abs/2411.17221

[37] Y. Cao, X. Min, W. Sun, and G. Zhai, "Attention-guided neural networks for full-reference and no-reference audio-visual quality assessment," *IEEE Trans. Image Process.*, vol. 32, pp. 1882–1896, 2023.

[38] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Trans. Image Process.*, vol. 29, pp. 6054–6068, 2020.

[39] W. Sun, X. Min, W. Lu, and G. Zhai, "A deep learning based no-reference quality assessment model for ugc videos," in *Proc. of the 30th ACM Int. Conf. Multimedia*, 2022, pp. 856–865.

[40] C. Li, Z. Zhang, H. Wu, W. Sun, X. Min, X. Liu, G. Zhai, and W. Lin, "Agiqa-3k: An open database for ai-generated image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 6833–6846, 2023.

[41] X. Min, G. Zhai, K. Gu, and X. Yang, "Fixation prediction through multimodal analysis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 1, Oct. 2016. [Online]. Available: https://doi.org/10.1145/2996463

[42] X. Min, G. Zhai, J. Zhou, X.-P. Zhang, X. Yang, and X. Guan, "A multimodal saliency model for videos with high audio-visual correspondence," *IEEE Trans. Image Process.*, vol. 29, pp. 3805–3819, 2020.

[43] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Sci. China Inf. Sci.*, vol. 63, pp. 1–52, 2020.

[44] C. Deng, J. Li, and X. Gao, "Geometric attacks resistant image watermarking in affine covariant regions," *Acta Automatic Sinica*, vol. 26, no. 2, pp. 221–228, 2010.

[45] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-lsb data embedding," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 253–266, 2005.

[46] C.-K. Chan and L.-M. Cheng, "Hiding data in images by simple lsb substitution," *Pattern Recognit.*, vol. 37, no. 3, pp. 469–474, 2004.

[47] D. Bhowmik and T. Feng, "The multimedia blockchain: A distributed and tamper-proof media transaction framework," in *Proc. Int. Conf. Dig Signal Process (DSP)*. IEEE, 2017, pp. 1–5.

[48] M. Holland, J. Stjepandić, and C. Nigischer, "Intellectual property protection of 3d print supply chain with blockchain technology," in *Proc. IEEE Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*. IEEE, 2018, pp. 1–8.

[49] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *Int. J. Web Grid Serv.*, vol. 14, no. 4, pp. 352–375, 2018.

[50] B. Wang, S. Jiawei, W. Wang, and P. Zhao, "Image copyright protection based on blockchain and zero-watermark," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2188–2199, 2022.

[51] Z. Meng, T. Morizumi, S. Miyata, and H. Kinoshita, "Design scheme of copyright management system based on digital watermarking and blockchain," in *Proc. IEEE Comput. Softw. Appl. Conf. (COMPSAC)*, vol. 2. IEEE, 2018, pp. 359–364.

[52] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A bayesian-mrf approach for prnu-based image forgery detection," *IEEE Trans. Inf. Forensic Secur.*, vol. 9, no. 4, pp. 554–567, 2014.

[53] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Blind prnu-based image clustering for source identification," *IEEE Trans. Inf. Forensic Secur.*, vol. 12, no. 9, pp. 2197–2211, 2017.

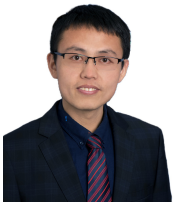
[54] Y. Zheng, Y. Cao, and C.-H. Chang, "A puf-based data-device hash for tampered image detection and source camera identification," *IEEE Trans. Inf. Forensic Secur.*, vol. 15, pp. 620–634, 2019.

[55] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[56] J. Wu, Z. Cui, V. S. Sheng, P. Zhao, D. Su, and S. Gong, "A comparative study of sift and its variants," *Meas. Sci. Rev.*, vol. 13, no. 3, pp. 122–131, 2013.

[57] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc Des. Autom. Conf. (DAC)*, 2007, pp. 9–14.

[58] C.-p. Wang, X.-y. Wang, X.-j. Chen, and C. Zhang, "Robust zero-watermarking algorithm based on polar complex exponential transform and logistic mapping," *Multimedia Tools Appl.*, vol. 76, pp. 26 355–26 376, 2017.
[59] H.-y. Yang, S.-r. Qi, P.-p. Niu, and X.-y. Wang, "Color image zero-watermarking based on fast quaternion generic polar complex exponential transform," *Signal Process. Image Commun.*, vol. 82, p. 115747, 2020.
[60] Y. Gao, S. F. Al-Sarawi, and D. Abbott, "Physical unclonable functions," *Nat. Electron.*, vol. 3, no. 2, pp. 81–91, 2020.



Yushu Zhang (Senior Member, IEEE) received the B.S. degree from the School of Science, North University of China, Taiyuan, China, in 2010, and the Ph.D. degree from the College of Computer Science, Chongqing University, Chongqing, China, in 2014. He held various research positions with Nanjing University of Aeronautics and Astronautics, City University of Hong Kong, Southwest University, University of Macau, and Deakin University. He is currently a Professor with the School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang, China. He is an Associate Editor of Information Sciences, Journal of King Saud University-Computer and Information Sciences, IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Network and Service Management, and Signal Processing. His research interests include multimedia security, blockchain, and artificial intelligence. He has co-authored more than 300 refereed journal articles and conference papers in these areas.

Bowen Shi received the B.S. degree in Information Security from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2023 and is currently pursuing the M.S. degree in Cyberspace Security from Nanjing University of Aeronautics and Astronautics. His research interests include digital watermarking, multimedia security, and copyright protection of digital images.

Shuren Qi is currently a Postdoctoral Fellow with Department of Mathematics, The Chinese University of Hong Kong. His research focuses on Geometric Deep Learning, with applications in Trustworthy AI and Science AI. His research has appeared in several top-tier journals and conferences, such as TPAMI, ICCV, and USENIX Security. His works offer some new designs of invariant representations – from global to local and hierarchical assumptions. More information is available at https://shurenqi.github.io/.

Xiangli Xiao received the B.E. degree in communication engineering from the College of Electronic and Information Engineering, Southwest University, Chongqing, China, in Jun. 2020, and the Ph.D. degree in cyberspace security from the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in Oct. 2024. He is currently a Lecturer with the School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang, China. His current research interests include multimedia security, digital watermarking, blockchain, and cloud computing security.

Ping Wang (Student Member, IEEE) received the B.S. degree from Wuhan Institute of Technology, Wuhan, China, in 2017 and the M.E. degree from Southwest University, Chongqing, China, in 2021. He is currently working toward the Ph.D. degree with Westlake University and Zhejiang University, Hangzhou, China. His research interests include computational imaging and computer vision.

Wenying Wen (Member, IEEE) received the M.S. degree in computational mathematics from the Inner Mongolia University of Technology, Hohhot, China, in 2010, and the Ph.D. degree in computational mathematics from Chongqing University, Chongqing, China, in 2013. She is currently a Professor with the School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang, China. Her research interests include image processing, multimedia security, compressive sensing security, and blockchain.