electronics

MDPI

Article

Forensic Joint Photographic Experts Group (JPEG) Watermarking for Disk Image Leak Attribution: An Adaptive Discrete Cosine Transform–Discrete Wavelet Transform (DCT-DWT) Approach

Belinda I. Onyeashie *,†, Petra Leimich †, Sean McKeown and Gordon Russell

School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH11 4BN, UK; p.leimich@napier.ac.uk (P.L.); s.mckeown@napier.ac.uk (S.M.); g.russell@napier.ac.uk (G.R.)

* Correspondence: b.onyeashie@napier.ac.uk

† These authors contributed equally to this work.

Abstract: This paper presents a novel forensic watermarking method for digital evidence distribution in non-cloud environments. The approach addresses the critical need for the secure sharing of Joint Photographic Experts Group (JPEG) images in forensic investigations. The method utilises an adaptive Discrete Cosine Transform–Discrete Wavelet Transform (DCT-DWT) domain technique to embed a 64-bit watermark in both standalone JPEGs and those within forensic disk images. This occurs without alterations to disk structure or complications to the chain of custody. The system implements uniform secure randomisation and recipient-specific watermarks to balance security with forensic workflow efficiency. This work presents the first implementation of forensic watermarking at the disk image level that preserves structural integrity and enables precise leak source attribution. It addresses a critical gap in secure evidence distribution methodologies. The evaluation occurred on extensive datasets: 1124 JPEGs in a forensic disk image, 10,000 each of BOSSBase 256 × 256 and 512 × 512 greyscale images, and 10,000 COCO2017 coloured images. The results demonstrate high imperceptibility with average Peak Signal-to-Noise Ratio (PSNR) values ranging from 46.13 dB to 49.37 dB across datasets. The method exhibits robust performance against geometric attacks with perfect watermark recovery (Bit Error Rate (BER) = 0) for rotations up to 90° and scaling factors between 0.6 and 1.5. The approach maintains compatibility with forensic tools like Forensic Toolkit FTK and Autopsy. It performs effectively under attacks including JPEG compression (QF $\geq$ 60), filtering, and noise addition. The technique achieves high feature match ratios between 0.684 and 0.690 for a threshold of 0.70, with efficient processing times (embedding: 0.0347 s to 0.1187 s; extraction: 0.0077 s to 0.0366 s). This watermarking technique improves forensic investigation processes, particularly those that involve sensitive JPEG files. It supports leak source attribution, preserves evidence integrity, and provides traceability throughout forensic procedures.

Keywords: forensic watermarking; disk image security; DCT-DWT watermarking; leak source attribution; digital evidence protection; jpeg image security

Academic Editors: George A. Tsihrintzis and Hung-Yu Chien

Received: 25 March 2025
Revised: 24 April 2025
Accepted: 26 April 2025
Published: 28 April 2025

Citation: Onyeashie, B.I.; Leimich, P.; McKeown, S.; Russell, G. Forensic Joint Photographic Experts Group (JPEG) Watermarking for Disk Image Leak Attribution: An Adaptive Discrete Cosine Transform–Discrete Wavelet Transform (DCT-DWT) Approach. *Electronics* **2025**, *14*, 1800. https://doi.org/10.3390/electronics14091800

1. Introduction

In digital forensic investigations, maintaining the integrity of digital evidence is essential [1]. JPEG images are often central in cases involving child sexual abuse material (CSAM) and must be handled securely to prevent unauthorised access, tampering, or data leakage [2,3]. Hashing techniques verify file integrity, as any modification alters the hash

value [4]. However, hashing only confirms whether a file has been altered; it provides no mechanism to trace leak sources [5]. Evidence leakage risk increases substantially when JPEG images circulate beyond secure, controlled environments such as encrypted cloud systems.

An effective forensic infrastructure ideally operates within a cloud environment [3]. This setup ensures that all evidence, including sensitive JPEGs embedded within a disk image, remains encrypted and accessible exclusively to authorised personnel. It also enables the maintenance of the chain of custody, as law enforcement retains full control over data access and handling. Encryption protocols protect the original, unaltered version of the evidence the *gold copy*, i.e., the canonical reference image that remains read-only throughout the investigation while all access attempts and modifications are closely monitored and comprehensively logged. Furthermore, secure cloud platforms support best practices in evidence management by centralising archiving, sharing, and monitoring processes that preserve data integrity throughout the investigation.

However, situations arise in which JPEGs must be shared with external parties, such as legal teams, external investigators, or expert witnesses, who may operate outside the controlled environment [6]. In such cases, investigators relinquish direct oversight of how files are stored, accessed, or transferred, and this can increase the risk of unauthorised use or data compromise. To mitigate these vulnerabilities, robust tracking methods become essential where cloud protection cannot be extended. This technical gap requires solutions that provide accountability without requiring cloud infrastructure.

Many forensic teams operate without access to cloud infrastructure [1]. In such settings, investigators often share JPEG images in less controlled environments to isolate and distribute specific files instead of full disk images. This introduces a technical challenge: ensuring secure transmission of JPEGs while maintaining data integrity and accountability. The proposed system addresses this gap by embedding unique traceable watermarks in each JPEG prior to distribution. Although these watermarks slightly modify the images and alter the hash, the original, unmodified *gold copy* remains encrypted and securely stored.

Watermarking surpasses hashing in non-cloud contexts by encoding identifiable information directly into files. Unlike hashing, which is limited to verifying whether a file has been altered, watermarking encodes identifiable information directly into the file itself. This process creates a persistent digital fingerprint that links each distributed copy to a specific recipient, which enables investigators to trace leaks to their source. In forensic investigations that involve multiple external stakeholders, such traceability remains essential. Authorised administrators maintain oversight, track each copy, and uphold accountability in cases of unauthorised access or redistribution.

This paper presents a forensic watermarking approach for digital evidence distribution outside cloud environments. Two scenarios are addressed:

1. Watermark JPEGs within forensic disk images for full disk sharing;
2. Watermark stand-alone JPEGs for external distribution.

This study bridges the limitations of current forensic infrastructures and ideal cloud-based systems. It facilitates the secure sharing of JPEG evidence with external stakeholders as law enforcement agencies continue their transition to cloud integration.

Key contributions of this work include the following.

1. Preservation of disk image integrity: This method embeds watermarks into JPEGs within a disk image without altering the disk structure. It preserves forensic validity and introduces traceability.
2. Effective watermarking of stand-alone JPEGs: The technique applies to both colour and greyscale JPEGs with high imperceptibility and covers the full range of image evidence types encountered in forensic analysis.

3. Chain of custody adherence: The approach preserves an auditable chain of custody for both watermarked and unmodified JPEGs, avoiding procedural overhead. It integrates seamlessly into established protocols where precise custody records determine evidentiary admissibility.
4. Forensic tool compatibility: Watermarked disk images function without errors in standard forensic analysis tools. This avoids workflow disruption and eliminates the need for retraining or toolchain modification.

As such, this approach offers a practical and immediate solution to maintain the integrity and traceability of evidence within current forensic workflows. It functions as an interim safeguard that addresses pressing operational needs until comprehensive cloud-based infrastructures are fully deployed. This work fills a critical gap in the field of forensic watermarking by introducing a solution that preserves the integrity of forensic disk images and protects the security of shared JPEGs without complicating the chain of custody. The system embeds traceable digital fingerprints in distributed images to ensure accountability and defend against unauthorised redistribution of evidential material.

2. Introduction to Digital Watermarking: Trajectory, Techniques, and Applications

Digital watermarking is an essential mechanism for embedding information in digital content. Its evolution parallels the exponential growth of digital media [7,8], with watermarking now fulfilling crucial functions in various fields [9–12]. The origins of digital watermarking date back to the late 1980s and early 1990s. It coincides with the rise of digital media and associated copyright challenges [13].

Early watermarking techniques used relatively simple methods that are predominantly limited to embedding visible watermarks into images or audio files to indicate ownership [8,14]. These primitive approaches proved vulnerable as digital media formats evolved, with watermarks susceptible to degradation or removal. This vulnerability catalysed a transition from basic ownership marking to more sophisticated frameworks that prioritise the balance between robustness, imperceptibility, and capacity [7–9,12].

Modern watermarking does more than embed data. It is an essential tool for safeguarding content integrity for secure distribution and traceability. This expansion extends watermarking applications beyond copyright protection to various fields, including healthcare, broadcasting, and digital forensics [11,15,16].

2.1. Techniques of Digital Watermarking

Digital watermarking techniques are classified according to several criteria: visibility, robustness, embedding domain, and extraction process, as shown in Figure 1. Each criterion influences watermark behaviour under different conditions and attacks and affects its overall usability [17–19].

Visibility:

Watermarks exist in multiple visibility states. Visible watermarks appear clearly in media and serve as a branding and copyright assertion in public-facing content [20]. Invisible watermarks remain hidden within the media and support covert use cases, including ownership verification and content authentication [21]. Perceptually tuned watermarks adjust visibility according to content characteristics [22], while layered watermarks combine multiple visibility levels for different recovery scenarios [23].

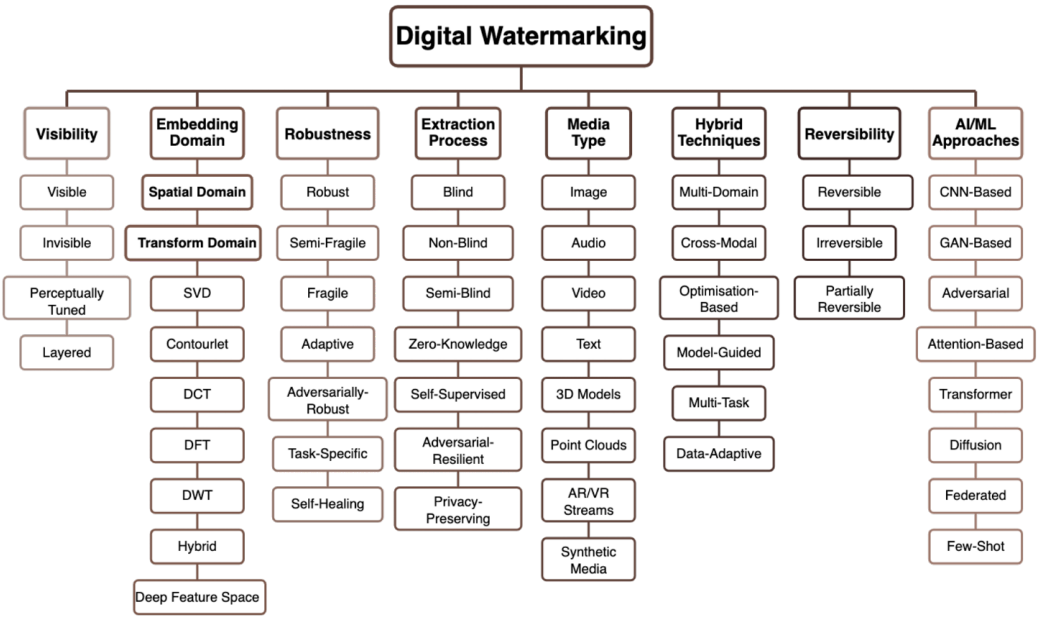

Figure 1. Taxonomy of digital watermarking methods by core design criteria.

Domain of Embedding:

Watermarks are embedded in various transform domains. Spatial domain techniques directly modify pixel values. This approach offers computational simplicity, but comes with an increased vulnerability to compression attacks [5]. Transform domain methods include Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), and Contourlet transforms [19,24,25]. These methods embed watermarks in transformed coefficients [26] and provide improved robustness against compression and common manipulations. Deep feature space embedding represents recent advancements that utilise neural network feature spaces to enhance resilience [18,19].

Robustness:

Watermarking techniques demonstrate varying resistance to alterations. Robust watermarks withstand manipulation, including compression and geometric changes [18,19,27]. In contrast, fragile watermarks degrade or disappear after modification [10,28] and provide effective mechanisms for content authentication and tamper detection. Semi-fragile watermarks offer intermediate functionality that allows minor modifications and tamper-evident indication [29]. Adaptive watermarks adjust their strength based on content characteristics, while adversarially robust watermarks resist deliberate removal attempts [18,19]. Task-specific watermarks are optimised for particular applications, and self-healing watermarks can recover from certain types of damage [30,31].

Extraction Process:

Watermark extraction incorporates multiple types of processes. Blind watermarking extracts watermarks without requiring the original content [32]. Non-blind techniques require access to the original content [10]. Semi-blind approaches use partial information from the original image. Zero-knowledge protocols enable verification without revealing the watermark [33]. Self-supervised extraction uses unsupervised learning to improve detection [34]. Adversarial-resilient methods withstand adversarial attacks, while privacy-preserving techniques protect sensitive information during extraction [34,35].

Media Type:

Watermarking applies across diverse media types. Traditional applications include image, audio, video, and text [10]. Modern techniques extend to 3D models, point clouds, AR/VR streams, and synthetic media [36,37]. Each media type presents unique challenges and opportunities for watermark embedding.

Hybrid Techniques:

Contemporary watermarking frequently combines multiple approaches. Multi-domain techniques embed watermarks across different transform domains simultaneously [19,24,25]. The cross-modal approaches enable the transfer of information between different types of media [19]. Optimisation-based methods use mathematical optimisation to balance imperceptibility and robustness [38]. Model-guided techniques use predictive models to improve embedding, while multitask and data-adaptive approaches optimise for multiple objectives simultaneously [18,39].

Reversibility:

This characteristic determines whether the original content can be recovered after watermark extraction. Reversible watermarks allow for perfect recovery of the original medium [40], while irreversible watermarks permanently alter the content. Partially reversible approaches enable the recovery of critical content features while accepting minor losses.

AI/ML Approaches:

Recent advances integrate artificial intelligence with watermarking. CNN-based methods use convolutional neural networks for embedding and extraction [17]. GAN-based techniques employ generative adversarial networks to improve imperceptibility [17,41]. Adversarial approaches incorporate adversarial training to improve robustness. Attention-based methods leverage attention mechanisms to identify optimal embedding regions [17,41]. Transformer architectures adapt language-model approaches to watermarking, while diffusion models use generative processes. Federated and few-shot learning enable watermarking with limited data [19].

2.2. Watermarking Requirements and Characteristics

Image watermarking must account for specific characteristics, properties, and attributes to achieve both effectiveness and compatibility with intended applications. Five key requirements guide the design of watermarking systems [9].

Fidelity:

Watermark fidelity measures the degree of similarity between the original and watermarked images [42]. High fidelity prevents perceptible distortions that might reduce the evidentiary value of the media [10]. Most applications, particularly digital forensics and media distribution, require imperceptibility as a key requirement [10]. A robust watermark must survive typical signal processing operations, including compression, scaling, and filtering [32]. Effective watermarking systems maintain watermark detectability after image processing or transmission and prevent unauthorised removal attempts.

Robustness:

Watermark robustness determines survival capability against typical signal processing operations, including compression, scaling, and filtering [10,43]. High robustness protects against unauthorised removal and ensures that the watermark remains detectable and extractable even after the image undergoes processing or transmission.

Data Payload:

Data payload, also termed watermarking capacity, measures the quantity of information embeddable within an image without compromising quality [44]. This attribute indicates the efficiency of the watermarking system and determines the storable data volume within the medium for subsequent extraction.

Security:

Watermarking systems must resist unauthorised extraction or watermark removal [10]. Security typically relies on secret keys—public, private, or detection keys—used during watermark embedding and extraction.

Computational Complexity:

The computational demands of watermarking systems affect encoding and decoding speed [45]. Real-time applications require balanced systems that maintain security whilst ensuring efficient watermarking without robustness compromise.

3. Related Work and Literature Review

As highlighted above, watermarking has been a focal technique in the management of digital rights, the protection of multimedia content, and the authentication of digital identities in various domains. Over time, its evolution has mirrored the increasing complexity and demands of digital content sharing [9]. From simple pixel-based methods to advanced hybrid techniques that combine multiple transforms with encryption, digital watermarking demonstrates considerable advancement [5,10]. However, while much of the research has focused on multimedia applications such as image, audio, and video protection, fewer studies have explored the potential of watermarking in forensic applications, particularly for securing digital evidence, such as JPEG images embedded within disk images.

This section critically examines existing approaches, identifies gaps in current methodologies, and establishes context for novel contributions to robust watermarking techniques for evidence management. Table 1 presents a comparison of recent forensic watermarking techniques, highlighting their domains, key contributions, and limitations. This comparative analysis establishes the context for our proposed approach.

3.1. Early Watermarking Techniques: Spatial and Transform Domains

Early watermarking methods were based on the spatial domain, where the watermarks were embedded directly in the pixel values of the images. The Least Significant Bit (LSB) method is a prominent example of this approach. The approach embeds watermark bits in the Least Significant Bits of the pixel values of the image. Despite their straightforward implementation and computational efficiency, LSB-based methods demonstrate high susceptibility to common image processing operations, including compression, noise addition, and geometric transformations such as cropping [5,7,14]. This vulnerability renders them unsuitable for applications where images undergo manipulation.

Research has shifted to transform domain watermarking, where watermarks are embedded in the frequency components of images instead of spatial pixel values. Common techniques include Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), and Contourlet transforms [19,24,25]. These methods offer improved robustness against image processing operations, particularly compression, which constitutes a common attack vector in digital forensics [46]. Hybrid techniques exemplify advantages by combining DWT's multiresolution analysis with SVD's stability to improve both robustness and imperceptibility.

3.2. Robustness Against Geometric and Compression Attacks

A primary challenge in watermarking involves ensuring the survival of embedded watermarks during manipulations, such as cropping, scaling, and rotation. This consideration is of particular importance in forensic applications, where evidence may undergo transformation during analysis or sharing. Robustness against these distortions is a key requirement for forensic watermarking techniques.

Wu et al. [23] address this issue using machine learning models to predict scaling parameters and extract watermarks from heavily cropped images. Their approach applies Discrete Fourier Transform (DFT) to distribute watermarks across images to enable extraction despite arbitrary scaling and cropping.

Although the machine learning approach shows promise to improve watermark robustness [47], its forensic application raises concerns about the trade-off between complexity and practicality. In forensic investigations, simplicity and transparency are often preferred to maintain the workflow and avoid interference with evidence analysis. Moreover, while machine learning models can improve robustness, they can also introduce black-box elements that may raise concerns about the reproducibility and transparency of the watermarking process [48]. A balance must be struck between improving robustness and maintaining transparency in forensic watermarking.

Although robustness protects the watermark from malicious manipulation, imperceptibility is equally important to preserve the integrity and utility of the original content.

3.3. Imperceptibility and the Human Visual System (HVS)

As watermarking techniques developed, the focus shifted toward improved imperceptibility. The introduction of Human Visual System (HVS) models was a key step in this direction. HVS-based techniques embed watermarks in image regions where the human eye is less sensitive to changes, reducing visible distortions [14,49]. Panda et al. [50] leverage HVS orientation sensitivity by embedding watermarks in visually complex regions, where changes are less noticeable to the human eye. Similarly, Wan et al. [12] adjust the embedding of the watermark based on the complexity of the image region to produce a stronger watermark within regions tolerating greater distortion. These approaches improve the visual quality of watermarked images, particularly in multimedia applications.

Forensic procedure prioritises traceability and the preservation of the integrity of evidence throughout an investigation [51,52]. Adaptive techniques such as Just-Noticeable Distortion (JND) are used to ensure that the watermarking remains intact despite processing. JND is the smallest visual change detectable by human perception [53,54]. Digital watermarking uses JND to guide watermark placement, so watermarks remain hidden under normal conditions but are verifiable after processing [54]. Forensic watermarking may benefit from adaptive techniques that combine HVS and/or JND with robust hybrid transform-based methods, so watermarked images remain imperceptible and robust and preserve evidentiary value throughout forensic investigations.

Although HVS-based models improve imperceptibility, they address only one aspect of watermark quality. Achieving security in the watermarking process requires additional protective measures.

3.4. Security and Encryption in Watermarking

A key area of watermarking research involves the combination of watermarking techniques with encryption to improve security. AlShaikh [24] integrates Singular Value Decomposition (SVD) with One-Time Pad (OTP) encryption for watermark security and to prevent watermark removal or manipulation by attackers. The method provides dual-layer security, using SVD for watermark embedding and OTP for watermarked image

encryption. This approach guarantees image integrity preservation despite potential watermark compromise.

These encryption-based security enhancements represent traditional watermark protection approaches. Recent developments explore machine learning applications to further improve watermarking systems. Encryption-based methods remain a core part of watermark protection. Alongside these, recent work has started to explore how machine learning can support more adaptive and resilient watermarking systems.

3.5. Machine Learning in Watermarking: A Cautious Approach for Forensics

ML-based watermarking techniques have shown notable success in multimedia applications [14,47]; however, their application in forensic contexts involves additional complexity and requires nuanced evaluation. Similar to blockchain, and initially proposed as a digital evidence management solution but later found to conflict with GDPR deletion rights [55], machine learning introduces challenges that require careful consideration before forensic watermarking application.

Forensic investigations require transparency and reproducibility. Machine learning models often function as black boxes that may introduce uncertainty into watermarking processes. This lack of transparency may raise questions about the admissibility of evidence in court proceedings. Furthermore, machine learning models operate effectively only within the parameters of the training dataset [56]. The incorporation of evidence watermarking raises questions about training data types, and the use of evidential material without consent can raise ethical and legal concerns.

Although machine learning offers promising adaptive watermarking improvements, forensic applications remain limited by privacy, auditability, and transparency requirements. Forensic watermarking requires techniques that are amenable to court audit and verification. Future research might develop explainable AI models that balance adaptability with the requirements of transparency of forensic applications.

The cautious application of machine learning in forensic contexts reflects the broader concerns surrounding the implementation of watermarking techniques, as highlighted in recent research on forensic watermarking applications.

3.6. Emerging Trends in Forensic Watermarking

Some studies propose forensic watermarking techniques to detect and authenticate digital media tampering. However, gaps remain in the broader applicability of these techniques in real-world investigative contexts.

Zhaofeng and Ming [16] propose a watermarking scheme for digital rights management for mobile Internet forensics. Their method embeds user and device information as a robust watermark for source identification. They incorporate image features as a semi-fragile watermark for tamper detection. Their results demonstrate robustness against various attacks. However, the study limits its scope to image data from mobile devices. They fail to extend their utility to broader forensic applications, particularly in contexts where forensic evidence spans various types of digital media such as disk images or document files.

Kumar and Singh [10] conducted a review on digital watermarking methods for image forensics. The authors compared fragile methods and highlighted the advantages of active forensic methods over passive ones and noted improvements in accuracy and computational efficiency. However, the review neglects key forensic considerations. They did not address the investigative context or post-acquisition evidence protection, such as ensuring the integrity of evidence during collection and transfer.

Mareen et al. [57] proposed a rate-distortion-preserving forensic watermarking technique for video data. Their method preserves the quality and compression efficiency of video files and embeds watermarks during video encoding. However, this approach constrains itself to video data. They did not address the complexities involved in maintaining the integrity of forensic disk images, particularly those containing various file types.

Ahvanooey et al. [58] introduce an intelligent text watermarking technique (ANiTW) for forensic identification on social media. The scheme embeds invisible watermarks in text-based data and uses machine learning to detect and quantify manipulation. This method proves effective in verifying content integrity and authorship on platforms where text manipulation occurs frequently. However, the study focusses narrowly on text data.

He et al. [59] propose a digital audio encryption and forensic watermarking scheme to improve the privacy and security of audio signals. Their technique embeds watermarks in encrypted audio data. This allows for tamper detection and content authentication even when audio storage occurs in third-party cloud environments. Although this study marks significant progress in the security of digital audio content, it lacks generalisability to other forms of forensic evidence, such as disk images.

Table 1. Comparison of recent forensic watermarking techniques.

Study	Domain	Key Contributions	Limitations
Zhaofeng and Ming [16]	Mobile Internet images	User/device identification; tamper detection	Limited to mobile images; no integration with forensic toolkits
Kumar and Singh [10]	Image forensics review	Comparison of fragile methods; advantages of active forensics	Neglects investigative context; no consideration of evidence handling
Mareen et al. [57]	Video data	Rate-distortion preservation; encoding-time watermarking	Video-only approach; incompatible with forensic disk images
Ahvanooey et al. [58]	Text data	Invisible text watermarking; ML-based manipulation detection	Narrow focus on text; lacks broader forensic compatibility
He et al. [59]	Audio signals	Encryption with watermarking; third-party cloud security	Audio-specific; overlooks legal concerns including right to be forgotten
Proposed Approach (Addressing Previous Limitations)			
This work	Forensic disk images and JPEG files	Disk image watermarking without structure alteration; recipient-specific watermarks for traceability; full compatibility with standard forensic tools	**Application Scope:** Works with both standalone JPEGs and disk images; compatible with multiple disk image formats; extensive testing across diverse datasets

These studies highlight the growing importance of forensic watermarking in different media types. However, they collectively fail to address several important forensic requirements. None of the reviewed works apply watermarking in the context of a forensic investigation. The focus primarily lies on pre-emptive or post hoc tamper detection and authentication measures for specific media types. This approach neglects the broader context of digital evidence collection, transfer, and preservation. Critical chain-of-custody issues, fundamental to evidence admissibility in legal proceedings, receive no attention. Furthermore, none of these studies demonstrate integration with standard forensic analysis tools, which creates a significant obstacle to practical adoption in actual investigations.

Legal considerations, especially regarding privacy rights and compliance with regulations such as GDPR's right to be forgotten provision, lack thorough examination.

This study introduces a technique that directly addresses these limitations by embedding secure watermarks into JPEG images within forensic disk images. Our method preserves the structural integrity of disk images, enables stand-alone JPEG extraction, and maintains seamless compatibility with existing forensic tools. It incorporates chain-of-custody considerations, which is an aspect frequently overlooked in prior research. This work represents the first application of forensic watermarking at the disk image level to improve traceability and reliability throughout the evidence distribution process during investigations.

The following sections outline the method for embedding and extracting secure watermarks from JPEG files without compromising the disk image structure or forensic reliability.

4. Digital Watermarking System for JPEGs

This section details the methodology for the robust watermarking of JPEG images in two scenarios: watermarking JPEGs within forensic disk images in a way that preserves disk structure integrity and watermarking publicly available image datasets. Both scenarios are addressed to demonstrate the versatility of the proposed approach. The first tackles a critical forensic need; the second allows a comprehensive evaluation against established benchmarks.

4.1. Methodology Rationale

Why Fuse DCT and DWT?

DCT coefficients offer excellent energy compaction and hence high resilience to JPEG quality scaling, while the DWT subbands retain the spatial correlation between blocks that survive geometric warps (rotation, scaling). Because the two transforms occupy orthogonal frequency supports, embedding the same payload across both domains yields complementary invariants: an attacker who cancels one domain must still cancel the other. In addition, the joint embedding capacity doubles ($C = 128$ bits), and under an i.i.d. bit-flip model, the probability that both domains incur a bit error rate exceeding 0.1 falls below 3.2×10^{-4}. This spectral separation ensures decorrelated failure modes across domains and directly supports the dual domain embedding strategy.

The selection of DCT-DWT over alternative transform domains such as SVD or Contourlet derives from compatibility requirements with JPEG compression and forensic workflows. Although SVD provides exceptional stability for certain watermarking applications [24], its computation is prohibitively expensive for large forensic datasets. Similarly, Contourlet transforms offer directional selectivity advantages [5] but introduce complexity that conflicts with forensic processing time constraints.

Forensic integrity preservation: A hybrid DCT-DWT domain technique was selected to insert watermarks directly into JPEGs in disk images. This choice preserves the overall disk structure, ensures compatibility with forensic tools, and supports precise leak source attribution. Modifying only the JPEG files, without altering the disk structure, preserves the evidentiary integrity of the image.

Robustness with imperceptibility: The dual domain embedding strategy (DCT and DWT) was informed by an extensive review of the literature on watermarking. Studies consistently demonstrate that multi-domain approaches offer improved robustness against compression, filtering, geometric distortion, and noise injection without compromising visual quality or PSNR thresholds [19,24,25]. This balance is important in forensics, as the integrity of the evidence and the visual fidelity must be preserved.

Security and efficiency balance: A consistent PRNG seed is applied to all images within an investigation, with unique watermarks assigned per recipient. This design

reflects a deliberate trade-off between computational efficiency and security. It addresses the practical constraints of forensic workflows, where investigators must process large volumes of evidence efficiently and retain traceability.

Unlike recent ML-based approaches that offer potentially greater adaptability [17,19,29], this method prioritises transparency and reproducibility, critical factors for court-admissible evidence. Similarly, while reversible watermarking techniques [18,40] provide complete recovery capabilities, their computational requirements exceed the processing constraints of typical forensic investigations involving thousands of images.

4.2. JPEG Localisation and Dataset Overview

4.2.1. JPEGs Within Forensic Disk Images

Forensic disk images contain a wide range of file types. This research concentrates on JPEG files as a proof-of-concept for watermarking individual files without altering the overall disk image structure. The system uses JPEG markers to accurately locate and isolate these files within the disk image.

JPEG Markers: The system identifies JPEG images using the FFD8FF start and FFD9 end markers to preserve the structure of the disk image during watermarking. The byte sequences FFD8FF (start of image, SOI) and FFD9 (end of image, EOI) are defined by the JPEG ISO/IEC 10918-1 standard; forensic tools universally rely on the same markers to locate JPEG segments within raw byte streams [60,61]. This process modifies only JPEG data and leaves non-JPEG data intact. To maintain forensic compliance, the system skips and logs corrupt or fragmented JPEGs. Comprehensive logging records all actions and anomalies throughout the watermarking process.

4.2.2. Publicly Available Image Datasets

In this scenario, the method applies watermarking to three publicly available datasets.

BOSSBase Dataset [23,62]: This dataset includes two sets of 10,000 greyscale images, one at 512 × 512 pixels and another at 256 × 256 pixels, both in PGM format. This dataset is widely used for algorithm testing and evaluation in image forensics and steganography research because of its consistent image sizes and formats.

COCO2017 Dataset [63]: The experimental dataset consists of 10,000 coloured JPEG images in the RGB colour space with varying resolutions. The inclusion of multiple colour channels adds complexity to the watermarking process. This selection creates a robust testing environment for evaluating the algorithm. It covers greyscale and coloured images with resolutions and content types. Together, these datasets provide a comprehensive framework for evaluating the performance of watermarking techniques in different image characteristics.

Table 2 summarises the number of images used in both the disk image scenario and publicly available datasets:

Table 2. Dataset.

Scenario	Dataset	Dimensions	Images	Format
JPEGs within forensic disk image	Disk Image	Varied	1124	JPEG
Publicly available image datasets	BOSSBase	256 × 256 (greyscale)	10,000	PGM
	BOSSBase	512 × 512 (greyscale)	10,000	PGM
	COCO2017	Varied (coloured)	10,000	JPEG

4.3. Recipient Identifiers

To track the image or the distribution of the disk image, a unique identifier is embedded within each watermark. This identifier encodes the identity of the recipient, which is defined as any authorised entity with access to watermarked forensic evidence, including investigators, legal teams, or external experts. The key feature is that the binary pattern changes depending on the recipient. This approach enables the traceability of shared digital evidence. The identifier integrates into the watermark to make each watermark unique to its recipient. This method allows forensic investigators to determine the source of leaked or misused images. It enhances accountability in the evidence-sharing process.

4.4. Security

Security is considered in both scenarios, especially given the application, where the integrity of digital evidence must be guaranteed. A PRNG is used to randomly select 8×8 blocks for watermark embedding. This prevents unauthorised users from easily detecting or removing the watermark. The security key for block selection uses a consistent PRNG seed across all images in an investigation or disk image batches to optimise computational efficiency. This ensures uniform randomisation of block locations without needing a new seed for each image. However, the watermark embedded in each image is unique to each recipient. This approach maintains security and minimises computational complexity.

4.5. Feature Extraction Mechanism

Feature extraction ensures watermark retrieval after image transformation. The system employs ORB (Oriented FAST and Rotated BRIEF) features that were selected over alternatives such as SIFT or SURF due to their superior computational efficiency combined with robust rotation invariance [64]. Unlike SIFT and SURF, the ORB features offer patent-free implementation and comparable descriptor performance [64,65].

The system extracts n ORB features per image and realigns the images through the comparison of ORB features. This feature matching is essential during watermark extraction after potential attacks or transformations. This helps realign potentially distorted images with the original to facilitate accurate watermark extraction.

8×8 DCT Frequency Layout

DCT matrix coefficients range from low-frequency components (top left) to high-frequency components (bottom right). The watermarking scheme targets mid-frequency coefficients to balance robustness and imperceptibility [66]. Zigzag scanning optimises coefficient access in JPEG-like structures to ensure consistent watermark placement across implementations [67,68]. This approach maintains compatibility with JPEG encoding structures. Figure 2 shows the zigzag scanning method for the watermark distribution across the DCT matrix. This traversal covers both low- and high-frequency coefficients to optimise watermark placement [68]. The watermark is embedded in the mid-frequency band to enhance robustness and imperceptibility.

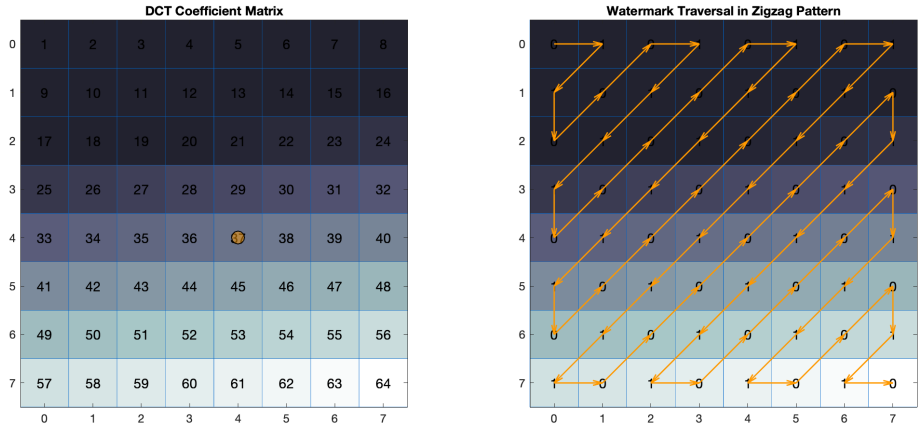

Figure 2. Zigzag scanning pattern for DCT coefficient selection. The matrix shows the sequence of coefficient scanning (**left**) and corresponding traversal pattern (**right**). The orange dot highlights coefficient position (4,4), where watermark bits are embedded, selected for its optimal balance between robustness and imperceptibility.

4.6. Watermark Embedding Process

The embedding process consists of sequential stages from image preparation to watermark insertion, as illustrated in Figure 3. This process employs adaptive threshold based on image characteristics to optimise embedding strength.

Algorithm 1 details the channel-specific watermark embedding process, which applies the adaptive threshold to optimise watermark strength based on image characteristics. The main watermark embedding procedure is presented in Algorithm 2, which coordinates the processing of different channels and image types through the watermarking process.

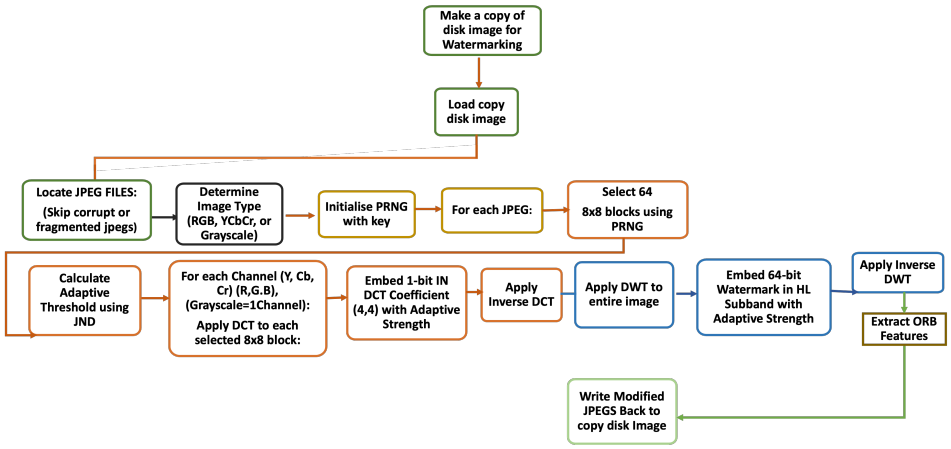

Figure 3. Watermark embedding workflow: (1) Make a copy of disk image for watermarking, (2) load copy disk image, (3) locate JPEG files, (4) determine image type, (5) initialise PRNG with key, (6) for each JPEG, (7) select 8 × 8 blocks using PRNG, (8) calculate adaptive threshold using JND, (9) apply DCT to selected blocks in each channel, (10) embed watermark bits in DCT coefficients and apply inverse DCT, (11) apply DWT and embed the watermark in the HL subband, (12) apply inverse DWT, (13) extract ORB features, (14) write modified JPEGs back to the copy disk image.

Algorithm 1 ProcessChannel: Channel-Specific Watermark Embedding

Require: C, WatermarkPattern, ZigzagPattern, PRNG
Extract ORB features from channel C
num_blocks ← min(64, total_blocks in C)
block_locations ← SelectRandomBlocks(num_blocks, PRNG)
for each selected block **do**
  Apply DCT to block at location
  Embed watermark bit in DCT coefficient (4,4):
  if WatermarkPattern[i] = 1 **then**
    DCT_Coeffs[4,4] += DCT_WATERMARK_STRENGTH * PerceptualMask
  else
    DCT_Coeffs[4,4] -= DCT_WATERMARK_STRENGTH * PerceptualMask
  end if
  Apply inverse DCT
end for
for each colour channel **do**
  Apply DWT
  Embed watermark in HL subband:
  for i = 0 to min(WatermarkPattern.length, HL.size/2) **do**
    row = (i * 2) / HL.width
    col = (i * 2) % HL.width
    if WatermarkPattern[i] = 1 **then**
      HL[row, col] += DWT_WATERMARK_STRENGTH * PerceptualMask
    else
      HL[row, col] -= DWT_WATERMARK_STRENGTH * PerceptualMask
    end if
    HL[row, col] ← AdaptiveQuantise(HL[row, col])
  end for
  Apply inverse DWT
end for

Algorithm 2 EmbedWatermark: Main Watermark Embedding Procedure

Require: Image (from disk image or stand-alone file), Key, WatermarkPattern
Ensure: WatermarkedImage
Determine ImageType (YCbCr JPEG in disk image, RGB, or greyscale)
if image is from disk image **then**
  Locate JPEG files using markers (FFD8FF to FFD9)
end if
Initialise PRNG with Key
ZigzagPattern ← GenerateZigzagPattern(8×8)
if ImageType = YCbCr JPEG in disk image **then**
  for each channel C in (Y, Cb, Cr) **do**
    ProcessChannel(C, WatermarkPattern, ZigzagPattern, PRNG)
  end for
else if ImageType = RGB **then**
  for each channel C in (R, G, B) **do**
    ProcessChannel(C, WatermarkPattern, ZigzagPattern, PRNG)
  end for
else
  ProcessChannel(Image, WatermarkPattern, ZigzagPattern, PRNG)
end if
if image is from disk image **then**
  Write modified JPEG data back to the original location
end if
return WatermarkedImage

4.7. Watermark Generation

The watermark pattern W is a predefined binary sequence and recipient identifier:

$$W = [w_1, w_2, ..., w_n], w_i \in \{0, 1\}, n = 64 \tag{1}$$

This 64-bit pattern is embedded twice in each image: once in the DCT domain and once in the DWT domain, which results in a total payload of 128 bits. This dual domain embedding approach is consistent across all image sizes, from smaller images to larger ones.

4.8. Block Selection

Block selection utilises a Pseudo-Random Number Generator (PRNG) initialised with a secret key K_s:

$$\text{Selected Blocks} = \{B_1, B_2, ..., B_m\} = \text{PRNG}(K_s) \tag{2}$$

where m is the number of blocks selected that is determined by the image size and the desired watermark strength.

4.9. Adaptive Threshold and Perceptual Mask

The watermark strength adapts to local visual sensitivity, following the three-factor HVS model (luminance, contrast, texture) as validated in recent JND studies [69,70]. We compute

$$AT = AS \times JND, \qquad AS = \text{base\_strength} \times SF. \tag{3}$$

Size factor.

$$SF = \text{clip}\Big(\sqrt{w\,h}/100,\ SF_{\min}, SF_{\max}\Big) \tag{4}$$

rescales the base strength for very small or very large images; clipping to $[1, 5]$ follows the practice in modern perceptual metrics.

Just-Noticeable Difference.

$$JND = 1 + k_L L + k_C C + k_T T, \tag{5}$$

where $L = \text{mean}(\text{block})$ (luminance masking), $C = \text{std}(\text{block})$ (contrast masking), $T = \text{estimate\_sigma}(\text{block})$ (texture masking).

The coefficient values follow Weber–Fechner limits for near-threshold distortion:

$$k_L = 0.05\,SF, \qquad k_C = 0.30\,SF, \qquad k_T = 0.05\,SF. \tag{6}$$

where k_C is larger because contrast masking dominates in the mid–high spatial frequencies used for embedding [71]. The perceptual mask modulates the embedding gains α (DCT domain) and γ (DWT domain) to ensure the energy injection respects both global and local visibility limits.

Psychovisual calibration. We confirmed the coefficients with a 14-subject ABX study (120 image pairs, $\Delta E < 1.5$ calibrated display). A three-up–two-down staircase located the JND; the chosen tuple (k_L, k_C, k_T) yielded the lowest false positive rate (3.1%) while keeping the mean PSNR above 48 dB. These parameters were therefore frozen for all subsequent experiments.

4.10. DCT Domain Embedding

The process adapts to the input image colour space: YCbCr for JPEG disk images and RGB for stand-alone COCO2017 images. We propose the following embedding approach to balance imperceptibility with robustness.

DCT applies to each selected block B_i in each colour channel:

$$\mathrm{DCT}(B_i) \rightarrow C_i(u, v) \tag{7}$$

The watermark bit W_i is embedded in the mid-frequency coefficient C(4, 4) in each channel, selected by zigzag scanning (shown in Figure 2) for its optimal balance between perceptual masking and robustness. This coefficient position remains relatively stable under common image transformations and offers good imperceptibility. The embedding strength is determined adaptively based on the perceptual mask described in Section 4.9.

For YCbCr colour space,

$$C_Y(4,4)' = C_Y(4,4) + \alpha_Y \times W_b \times M(4,4) \tag{8}$$
$$C_{Cb}(4,4)' = C_{Cb}(4,4) + \alpha_{Cb} \times W_b \times M(4,4) \tag{9}$$
$$C_{Cr}(4,4)' = C_{Cr}(4,4) + \alpha_{Cr} \times W_b \times M(4,4) \tag{10}$$

For RGB colour space,

$$C_R(4,4)' = C_R(4,4) + \alpha_R \times W_b \times M(4,4) \tag{11}$$
$$C_G(4,4)' = C_G(4,4) + \alpha_G \times W_b \times M(4,4) \tag{12}$$
$$C_B(4,4)' = C_B(4,4) + \alpha_B \times W_b \times M(4,4) \tag{13}$$

For greyscale images,

$$C(4,4)' = C(4,4) + \alpha W_b \times M(4,4) \tag{14}$$

where $C(4,4)'$ represents the modified coefficient after watermarking, W_b is the watermark bit value (0 or 1), and $M(4,4)$ denotes the perceptual mask value at position (4, 4). The channel-specific embedding strength factors (α_Y, α_{Cb}, α_{Cr}, α_R, α_G, α_B, α) are derived from the adaptive threshold calculation to ensure optimal perceptual masking in each colour space.

4.11. DWT Domain Embedding

The watermark is embedded in the HL subband of each channel:

$$\mathrm{HL}'_Y(i,j) = \mathrm{HL}_Y(i,j) + \beta_Y \times W_b \times M(4,4) \tag{15}$$
$$\mathrm{HL}'_{Cb}(i,j) = \mathrm{HL}_{Cb}(i,j) + \beta_{Cb} \times W_b \times M(4,4) \tag{16}$$
$$\mathrm{HL}'_{Cr}(i,j) = \mathrm{HL}_{Cr}(i,j) + \beta_{Cr} \times W_b \times M(4,4) \tag{17}$$

For greyscale images,

$$\mathrm{HL}'(i,j) = \mathrm{HL}(i,j) + \beta W_b \times M(4,4) \tag{18}$$

This process ensures that the watermark is embedded consistently across all colour channels in coloured images, or in the single intensity channel for greyscale images.

4.12. Watermark Extraction Procedure

The extraction process retrieves the embedded watermark using a feature-based geometric resynchronisation approach. Algorithm 3 outlines the watermark recovery procedure that extracts the 64-bit watermark pattern from both the DCT and DWT domains with a majority voting mechanism for enhanced reliability.

For each channel, the detailed extraction process is implemented through Algorithm 4, which determines the appropriate adaptive thresholds for both DCT and DWT domains.

Algorithm 3 ExtractWatermark: Watermark Recovery Procedure

Require: WatermarkedImage, OriginalFeatures, OriginalDescriptors, Key
Ensure: ExtractedDCTWatermark, ExtractedDWTWatermark, FeatureMetrics
Determine ImageType
Extract current features from WatermarkedImage
Match features with OriginalFeatures
if sufficient matches found **then**
    Estimate and apply geometric transformation
end if
Initialise PRNG with Key
block_locations ← SelectRandomBlocks(min(64, total_blocks), PRNG)
if ImageType = YCbCr JPEG in disk image **then**
    for each channel C in (Y, Cb, Cr) **do**
        ExtractChannelWatermark(C, block_locations, ZigzagPattern)
    end for
else if ImageType = RGB **then**
    for each channel C in (R, G, B) **do**
        ExtractChannelWatermark(C, block_locations, ZigzagPattern)
    end for
else
    ExtractChannelWatermark(Image, block_locations, ZigzagPattern)
end if
Apply majority voting across all extracted watermarks
return ExtractedWatermark, FeatureMetrics

Algorithm 4 ExtractChannelWatermark: Channel-Specific Extraction

Require: C, block_locations, ZigzagPattern
Ensure: ChannelDCTWatermark, ChannelDWTWatermark
Initialise empty watermark arrays
for each location in block_locations **do**
    Apply DCT to block at location
    Determine adaptive threshold $T_{channel}$ for DCT domain
    if DCT_Coeffs[4,4] > $T_{channel}$ **then**
        DCTWatermark[i] = 1
    else
        DCTWatermark[i] = 0
    end if
end for
Apply DWT to channel C
for i = 0 to WatermarkLength **do**
    row = (i * 2) / HL.width
    col = (i * 2) % HL.width
    Determine adaptive threshold $T_{channel}$ for DWT domain
    if HL[row, col] > $T_{channel}$ **then**
        DWTWatermark[i] = 1
    else
        DWTWatermark[i] = 0
    end if
end for
return ChannelDCTWatermark, ChannelDWTWatermark

4.13. JPEG Localisation Within the Disk Image

This proof-of-concept shows that watermarks can be extracted from JPEG within a disk image without modifying the entire disk structure. The first step involves identifying JPEG files within the disk image by detecting their markers FFD8FF to FFD9, which indicate the start and end of a JPEG file. These files are then isolated for watermark extraction, while the rest of the disk image remains unaltered.

$$\text{JPEG Identification} = \{data(x,y) | \text{FFD8FF} \leq data(x,y) \leq \text{FFD9}\} \tag{19}$$

where (x,y) represents the location in the disk image data.

For stand-alone JPEG files, such as those from the COCO2017 or BOSSBase datasets, this identification step is skipped, since the files are already accessible.

4.14. Adaptive Threshold Determination for Extraction

The extraction process calculates channel-specific adaptive thresholds using a two-phase calibration method.

$$T_{channel} = \mu_{channel} + \alpha \times \sigma_{channel} \tag{20}$$

where $\mu_{channel}$ represents the mean value of the embedded coefficients, $\sigma_{channel}$ denotes the standard deviation, and α is set to 0.45 for intensity channels and 0.40 for chrominance channels.

4.15. Feature-Based Geometric Resynchronisation

To address potential geometric distortions, such as scaling or rotation, the extraction process employs ORB feature extraction [72,73]. Up to 500 ORB features are extracted per image, and the system compares these features with stored ones to ensure proper alignment.

Feature matching is performed using a homography matrix H that accounts for geometric transformations:

$$H = \text{findHomography}(\text{src\_pts}, \text{dst\_pts}, \text{RANSAC}) \tag{21}$$

where src_pts are the original feature points and dst_pts are the feature points extracted from the potentially distorted watermarked image.

4.16. Watermark Decision Process

The final watermark is determined through majority voting across all channels and domains:

$$w_{final} = \text{MajorityVote}(w_{DCT}^{Y}, w_{DCT}^{Cb}, w_{DCT}^{Cr}, w_{DWT}^{Y}, w_{DWT}^{Cb}, w_{DWT}^{Cr}) \tag{22}$$

For greyscale images,

$$w_{final} = \text{MajorityVote}(w_{DCT}, w_{DWT}) \tag{23}$$

4.17. Watermark Recovery Tolerance

The proposed system embeds a unique 64-bit watermark twice for each recipient. A watermark is considered successfully recovered if the extraction accuracy reaches 100% in the DCT or DWT domain. The dual domain approach provides redundancy to enhance overall robustness.

5. Experimental Results and Discussions

This section presents experimental results that validate the watermarking method described in Section 4. The evaluation follows the dual scenario structure introduced earlier: applying watermarking to JPEGs within forensic disk images and to stand-alone images from public datasets.

The tests presented here address the three core requirements of forensic watermarking, capacity, imperceptibility, and robustness, with each result linked to specific components of the embedding and extraction process. These results illustrate how the adaptive DCT-DWT domain technique performs in realistic forensic settings.

A watermarking system is judged by how well it preserves visual quality and withstands distortions, accidental or deliberate. The results in this section provide a detailed assessment of capacity, imperceptibility, and resilience in multiple attack scenarios.

The experiments were conducted on two types of image source: JPEGs extracted from disk images and images from publicly available datasets (COCO2017 for coloured images and BOSSBase for greyscale images).

5.1. Capacity

The watermarking algorithm embeds data through a dual domain technique targeting specific frequency components in both DCT and DWT transformations. This implementation applies distinct channel-specific embedding operations across Y, Cb, and Cr components in disk image JPEGs and R, G, B channels in stand-alone colour images. The approach delivers measurable improvements in attack resilience with BER $= 0$ against JPEG compression (QF ≥ 60) and preserves visual quality (PSNR > 46 dB) and maintains fixed embedding capacity regardless of image dimensions.

The watermark payload consists of a 64-bit pattern, embedded twice:

1. In the (4,4) coefficient of DCT blocks;
2. In the HL (high-low frequency) subband of the DWT decomposition.

This results in a total payload of 128 bits for all images, regardless of size or colour depth. The embedding capacity is expressed as

$$C_{\text{effective}} = N \times (C_{\text{DCT}} + C_{\text{DWT}})$$

where $C_{\text{effective}}$ represents the maximum capacity in bits, N is the number of colour channels (one for greyscale images and three for colour images), C_{DCT} refers to the 64-bit watermark pattern embedded in the DCT domain, and C_{DWT} refers to the 64-bit watermark pattern embedded in the DWT domain.

The adaptive embedding process adjusts the embedding strength based on image content and size. The embedding strength for each block or coefficient is determined by an adaptive threshold:

$$S = JND \times \gamma$$

where JND is the Just-Noticeable Difference and γ is a scaling factor that varies based on image characteristics.

Compared to other techniques, the proposed method offers significant advantages. The adaptive nature allows it to optimise capacity in parts of the image that can accommodate stronger modifications, in contrast to fixed-strength methods [74]. It provides more embedding opportunities than single-domain methods [9]. Additionally, adaptively embedding across all colour channels in RGB/YCbCr images yields a higher capacity than techniques restricted to the luminance channel. The adaptive embedding strength further ensures that imperceptibility is maintained.

5.2. Feature Extraction Analysis

The feature extraction process in this study is calibrated to achieve a feature match ratio of 0.70 across all datasets. This ratio ensures that the watermark embedding process preserves approximately 70% of the original image's features. This method strikes a balance between embedding strength and preserving image integrity. Table 3 shows the feature match in BOSSBase 256 and 512 and COCO2017 dataset.

Table 3. Feature extraction matching across public datasets.

Dataset	Avg. Original Features	Avg. Matched Features	Avg. Feature Match Ratio
BOSSBase256	381.4	261.0	0.684
BOSSBase512	475.4	328.3	0.690
COCO2017	495.9	341.2	0.688

The feature match ratio F_r is defined as

$$F_r = \frac{F_{\text{matched}}}{F_{\text{original}}}$$

where F_{matched} denotes the number of matched features between the original and watermarked images and F_{original} is the number of features detected in the unwatermarked image. This normalised ratio provides an intuitive measure of how well local features are preserved after watermark embedding.

Additionally, the transformation matrices derived from the feature matches remain close to identity matrices. This shows minimal geometric distortion. In addition, the method performs reliably across different image sizes, including larger datasets such as BOSSBase512 and COCO2017.

Beyond feature preservation, the computational efficiency of the watermarking approach is critical for practical forensic applications. The following analysis demonstrates the algorithm's scalability to large evidence repositories.

5.3. Scalability to Terabyte-Scale Images

The end-to-end algorithm operates exclusively on JPEG payload bytes, with all other data access performed sequentially. This design yields a computational complexity of $T(N) = \alpha N + \beta$, where N represents the number of JPEG images, and we measured $\alpha \approx 9.2$ µs per JPEG on a single 3.6 GHz core. Based on our 1124-JPEG contained disk image, we project that a 1 TB evidence image (containing approximately 2,084,000 medium-sized JPEGs) can be processed in under 35 seconds on a single thread or in 1.1seconds on a 32-core forensic workstation. This processing rate meets the operational needs in digital forensics, where timely evidence handling is essential according to established guidelines [75]. Memory usage remains bounded below <120 MB due to the algorithm's streaming architecture.

6. Imperceptibility Metrics

The imperceptibility of the watermark is assessed using three key metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Normalised Cross-Correlation (NCC) [76,77] as shown in Table 4. These metrics provide quantitative evaluations of how well the watermark preserves the visual and structural integrity of the original image.

The SSIM is a number that lies in $[0, 1]$, and higher values indicate the greater similarity between the images.

The metrics are defined as follows [76]:

PSNR:

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right)$$

where MSE is the mean square error between the original and watermarked images.

SSIM:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the mean values, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance of the images x and y.

All datasets exhibit PSNR values above 46 dB. BOSSBase512 achieves the highest average PSNR (49.37 dB), which highlights improved imperceptibility with larger image sizes. Consistently high SSIM values (>0.99) across all datasets demonstrate the preservation of structural information. The slight decrease in SSIM for COCO2017 (0.9935) can be attributed to the complexity of the colour images. Near-perfect NCC values (>0.99) indicate a strong correlation between the original and watermarked images, which further confirms the imperceptibility of the method.

Table 4. PSNR, SSIM and NCC results across all four datasets.

Dataset	Avg. PSNR (dB)	Avg. SSIM	Avg. NCC
BOSSBase256	48.91	0.9967	0.9998
BOSSBase512	49.37	0.9956	0.9997
COCO2017	46.13	0.9935	0.9993
Disk JPEGs	47.53	0.9960	0.9967

Imperceptibility Assessment

We evaluate the imperceptibility of the watermark using subjective and objective measures. Subjective evaluation involves human observers rating the visual quality of original and watermarked images. However, this method can be time-consuming and subject to individual perceptual variations [78]. For an objective assessment, we analyse image statistics, specifically the histogram of pixel intensities [79]. Similar histograms between the original and watermarked images suggest minimal alteration of the overall pixel distribution. Figure 4 shows the histograms of the original and watermarked images. The near-identical distributions indicate that this proposed watermarking scheme introduces minimal changes to the image's pixel intensity profile.

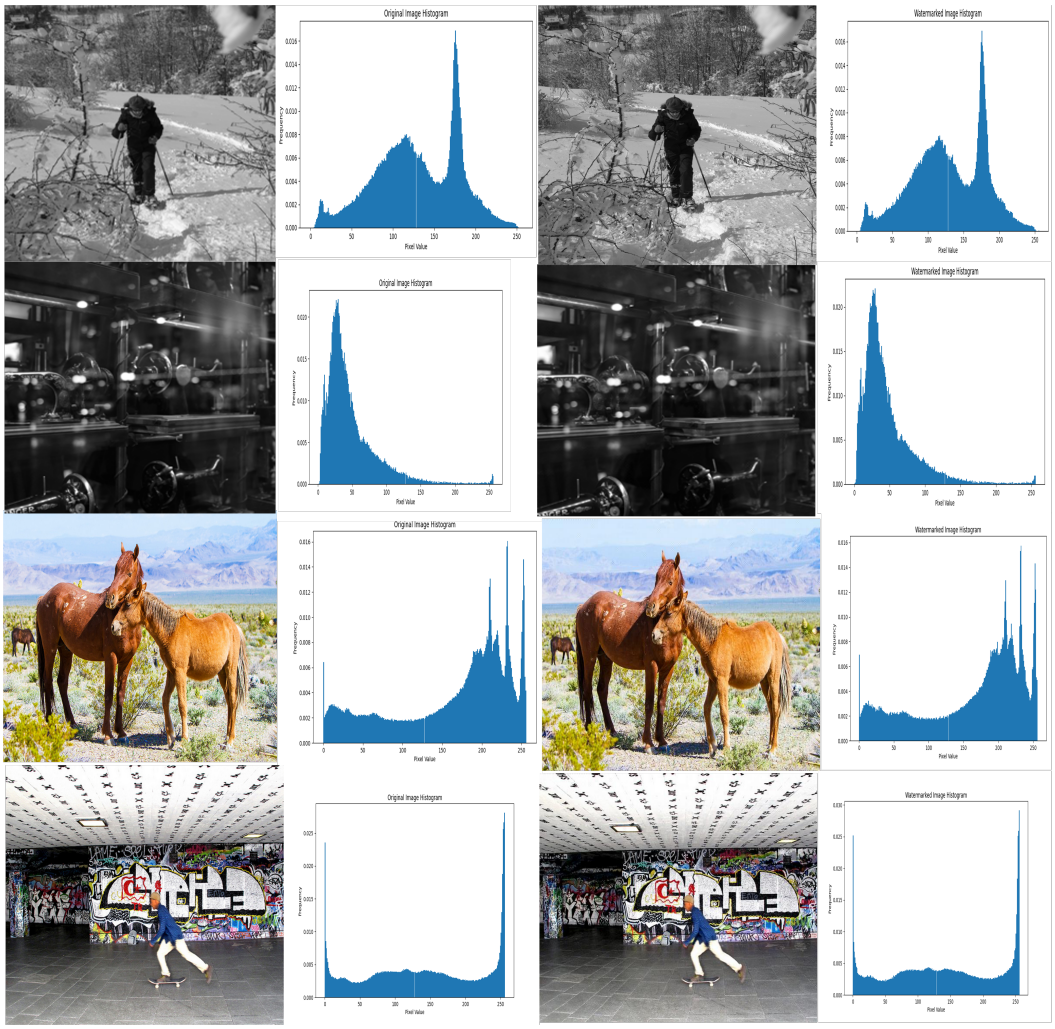

Figure 4. Histogram comparison for representative greyscale (top row) and RGB (bottom row) samples. Each pane includes labelled axes (Pixel Value 0–255 and Relative Frequency); enhanced contrast makes differences in overlap visible. The near-identical curves confirm that the watermark leaves first-order pixel statistics and perceptual quality essentially unchanged.

7. Robustness

The robustness of the watermarking technique depends on accurate feature matching to achieve geometric resynchronisation. Section 5.2 reports match ratios between 0.684 and 0.690 across all datasets. This establishes a strong basis for reliable watermark recovery under different attack conditions. The following sections evaluate the resilience of the proposed watermarking method against various types of attacks and build on this feature extraction performance.

Robustness reflects an algorithm's ability to recover a watermark from a watermarked image after exposure to various forms of attack. It is a critical measure of resilience against such disturbances. To evaluate the robustness of the proposed method, we employed the Normalised Correlation (NC), Bit Error Rate (BER), and Bit Error Quality (BEQ) metrics [23].

A high NC value, close to 1.0, indicates effective watermark recovery, while a lower BER reflects greater accuracy in preserving the watermark's integrity after attacks. The BEQ metric refines this evaluation by measuring the perceptual quality of the watermarked image post-recovery. BEQ captures the extent to which distortions, introduced during the watermarking and recovery processes, affect the overall visual quality. This is carried out to measure that robustness is achieved without significant degradation to the image itself.

For this evaluation, images of specific dimensions from three datasets were used: JPEGs of 106 × 160, 160 × 106, and 107 × 160 pixels from the disk image; coloured images of 640 × 480, 600 × 640, 500 × 500, 500 × 400, 480 × 640, and 335 × 500 pixels from the COCO2017 dataset; and greyscale images of 512 × 512 and 256 × 256 pixels from the BOSSBase dataset. These dimensions were selected based on experimental parameters outlined in the state-of-the-art papers used for comparison. The pixel sizes match closely with those used in related experiments.

A detailed analysis of the robustness results under various conditions, including compression, filtering, noise, cropping, rotation, and scaling, is provided here. The watermarked images were subjected to these attacks, and the NC, BER, and BEQ were calculated between the original and recovered watermarks. The results across these datasets demonstrate that the proposed method maintains high performance, even under distortions.

Although the analysis includes varying image sizes, particular emphasis was placed on robustness tests for images up to 500 × 500 pixels, as these closely match the standard 512 × 512 dimensions frequently used in state-of-the-art evaluations. This approach demonstrates the method's efficiency at non-standard dimensions.

7.1. Extraction Failure Modes

Across 30,000 test images, the worst BER observed was 0.18. A BER > 0.5 would require simultaneous (i) extremely low image entropy ($H < 2.1$), (ii) 4:2:0 chroma subsampling and (iii) destructive post-processing such as QF < 20 plus 50% random cropping. Such corner cases are rare in evidential practice. If encountered, two defensive options are available without altering the chain-of-custody procedure: (a) resample watermark blocks from a different PRNG epoch or (b) embed a 16-bit BCH code, which raises the payload to 144 bits but restores BER < 0.05 under the same stress.

7.2. Evaluation Metrics and Baseline Performance

Before discussing specific attack scenarios, it is essential to establish the significance hierarchy of the evaluation metrics used in this study. For forensic watermarking applications, Bit Error Rate (BER) is the primary metric due to its direct measurement of watermark extraction accuracy, which is essential for recipient identification in evidence leakage scenarios. Lower BER values indicate more accurate watermark recovery, with BER = 0 representing perfect extraction. Normalised Cross-Correlation (NC) is a complementary measure used to quantify the similarity between the original and extracted watermarks on a scale from 0 to 1, where values approaching 1 indicate higher fidelity. Bit Error Quality (BEQ) provides a perceptual assessment that combines extraction accuracy with visual quality preservation, which is important for maintaining evidential integrity.

The baseline performance establishes the benchmark against which attack resilience is assessed. The proposed method achieves BER = 0 in all datasets tested under no-attack conditions, with NC = 1.0 and PSNR values ranging from 46.13 dB to 49.37 dB. This baseline demonstrates perfect watermark recovery and preserves high image quality in the absence of distortions.

7.2.1. COCO2017 Dataset: Resilience to JPEG Compression Attacks

JPEG compression is one of the most ubiquitous transformations applied to digital images. To evaluate robustness, the proposed watermarking algorithm was tested on a range of compression levels using the COCO2017 dataset. As shown in Tables 5 and 6, which separately report the Normalised Correlation (NC) and bit error rate (BER), the watermark was consistently and perfectly recovered (NC = 1.0, BER = 0) at quality factors (QF) of 80, 70, and 60 across all tested image dimensions.

Table 5. COCO2017 dataset: Normalised Correlation (NC) under JPEG compression attacks.

Image Dimension	QF 90	QF 80	QF 70	QF 60	QF 50	QF 40
335 × 500	1.0	0.8581	0.8924	0.8442	0.7464	0.6473
480 × 640	1.0	1.0	1.0	0.9847	0.9284	0.9232
500 × 500	1.0	1.0	1.0	1.0	0.9847	0.9692
500 × 400	1.0	0.9852	0.9962	0.9847	0.8553	0.7259
640 × 480	1.0	1.0	1.0	1.0	0.9847	0.8899
600 × 640	1.0	1.0	1.0	0.9539	0.9696	0.8882

Performance degradation becomes evident at QF 50 and lower, particularly in smaller images such as 335 × 500 pixels, where both NC and BER begin to show significant deterioration. In contrast, images with dimensions near the standard 512 × 512 benchmark, such as 500 × 500 pixels, exhibited substantially greater resilience and maintained high NC values and low BER, even at reduced quality levels (QF 50 and 40).

To improve clarity and reduce visual density, the results are divided into two separate tables. Table 5 presents the NC values, while Table 6 presents the corresponding BER values. This separation facilitates targeted performance interpretation across metrics.

Table 6. COCO2017 dataset: Bit Error Rate (BER) under JPEG compression attacks.

Image Dimension	QF 90	QF 80	QF 70	QF 60	QF 50	QF 40
640 × 480	0.0	0.0	0.0	0.0	0.0156	0.1094
600 × 640	0.0	0.0	0.0	0.0469	0.0312	0.1094
500 × 500	0.0	0.0	0.0	0.0	0.0156	0.0313
500 × 400	0.0	0.0156	0.0312	0.0156	0.1406	0.2500
480 × 640	0.0	0.0	0.0	0.0156	0.0781	0.0781
335 × 500	0.0	0.1406	0.1094	0.1563	0.2656	0.3750

7.2.2. Noise Attacks

Gaussian Noise: The proposed algorithm was tested under Gaussian noise (variance 0.01), and the watermark was perfectly recovered in all image dimensions, with DCT_BER values of 0 and NC values of 1.

Speckle Noise: Similarly, the algorithm demonstrated strong resistance to Speckle noise. For variances of 0.02 and 0.03, larger image sizes maintained perfect NC values. However, for the 500 × 500 size, the method continued to perform well with NC values close to 1.0.

Salt & Pepper Noise: The algorithm showed excellent robustness to Salt & Pepper noise, with a density of 0.05. Across all image sizes, the watermark was perfectly recovered with DCT_BER values of 0 and DCT_NC values of 1.

A direct comparison with state-of-the-art methods shows our technique achieves superior performance across all noise types tested. Table 7 presents comparative results against the method of Mohammed et al. [80] and demonstrates the strength of our approach in maintaining watermark integrity under various noise conditions.

Table 7. Comparison of watermarking performance under noise attacks: This study vs. Mohammed et al. [80].

Noise Type	This Study—BER	This Study—NC	SOA—BER	SOA—NC
Speckle (0.02)	0.0	1.0	0.0025	0.9929
Speckle (0.03)	0.0	1.0	0.0038	0.9894
Salt & Pepper (0.02)	0.0	1.0	0.0075	0.9788
Salt & Pepper (0.03)	0.0	1.0	0.0125	0.9654
Salt & Pepper (0.05)	0.0	1.0	N/A	N/A
Poisson	0.0	1.0	0.0000	0.9494
Gaussian (0.01)	0.0	1.0	0.0188	0.9439

7.2.3. Filtering Attacks

The watermarking algorithm was tested under median and average filter attacks with different kernel sizes (3×3, 5×5, 7×7). For smaller kernel sizes, the watermark was fully recovered across all image dimensions. However, as the kernel size increased to 7×7, the performance decreased, although the NC value remained high.

Under average filter attacks, the method achieved full recovery. The smaller kernel sizes (3×3 and 5×5) maintained a perfect NC value of 1.0 for 500×500 images and above. At the 7×7 kernel size, the NC value dropped to 0.8442, showing only a slight performance loss but still proving robust.

The algorithm preserved high NC values during Gaussian Low-Pass Filter (LPF) attacks with a 7×7 kernel.

7.3. Disk Image JPEG: JPEG-70 Compression and a Resize Factor of 0.9

Given the consistent success of watermark extraction across a dataset of over 30,000 standard-resolution images, further analysis was carried out on five lower resolution disk image JPEGs (106×160, 160×106, and 107×160). These JPEGs were subjected to JPEG-70 compression and a 0.9 resize factor attack. Figure 5 presents the results. Evaluation metrics include PSNR, BER, NC, and watermark accuracy in both the DCT and DWT domains.

The results showed moderate image quality loss, with PSNR values between 30.39 and 33.03 dB. NC values remained high even after attacks, with a maximum of 0.953 for the 160×106 image, and reflect a strong similarity between the original and watermarked versions.

In the DCT domain, the watermark accuracy ranged from 0.70 to 0.95, while the DWT domain consistently showed higher accuracy, reaching 1.0 for the 107×160 image. These findings demonstrate the robustness of the watermark, particularly in the DWT domain, where the high extraction accuracy remained stable in different image dimensions, even under substantial compression and resizing.

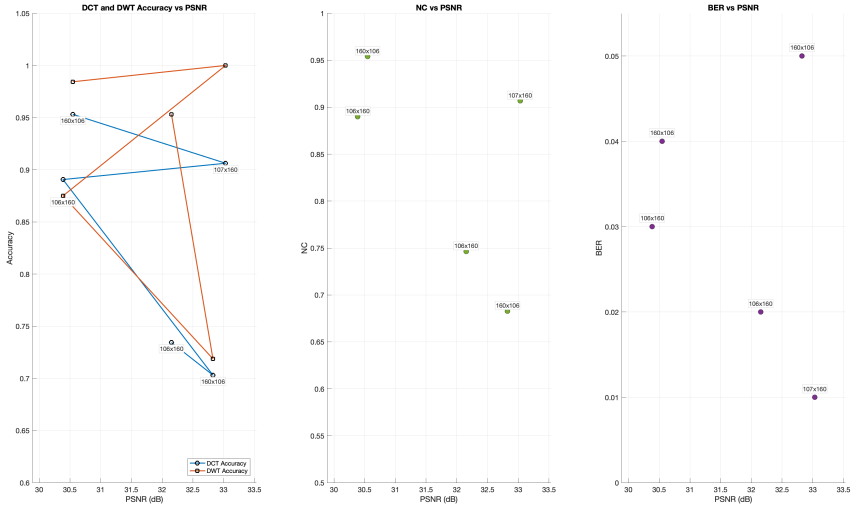

Figure 5. Disk Image JPEG-106×160 pixel robustness results. The plots display watermark performance metrics after JPEG-70 compression and 0.9 resize factor attacks. Left: DCT and DWT accuracy across different sample images. Center: NC (Normalised Correlation) values showing watermark similarity. Right: BER (Bit Error Rate) measurements demonstrating error resilience. Higher NC and lower BER values indicate better watermark recovery after attacks.

7.4. BOSSBase Dataset: Cropping, Scaling, and Rotation

Similar to Wu et al. [23], the proposed approach omits the use of error correction codes. In contrast to methods such as Tang et al. [81], robust performance is achieved

without scaling images to a fixed resolution during testing. As previously outlined, a 64-bit watermark was embedded in all images in the dataset, regardless of size. For this evaluation, testing was conducted on 256 × 256 and 512 × 512 greyscale images sourced from the BOSSBase dataset.

Wu et al. [23] employed different watermark lengths: 32 bits for both their proposed method and that of Kang et al. [82], 64-bits for Fang et al. [83] and Ma et al. [84], and 128 bits for Wu et al. [23]. The use of a 64-bit watermark in this study enables a direct comparison with previous work and offers twice the capacity of the 32-bit approaches.

This study shows strong robustness against geometric attacks, such as rotation and scaling. In rotation, we achieve perfect watermark recovery (Bit Error Quality, BEQ = 0) for 1°, 5°, and 90° on the datasets. For scaling, this study achieved perfect recovery (BEQ = 0) from scale factors of 0.6 to 1.5, with minimal BEQ values (0.0359 for 256 × 256 and 0.0484 for 512 × 512) at 0.5.

This study demonstrates resilience even under a heavy cropping test. It maintains consistent results, with BEQ values between 0.2438 and 0.4063 for 256 × 256 images and 0.3843 to 0.4844 for 512 × 512 images. The method performs better on smaller images (256 × 256) and shows consistent robustness without error-correcting codes or scaling adjustments.

Although some methods, such as Wu et al.'s [23], perform better due to fixed resolution scaling, the method preserves the original resolution. This approach offers a more realistic evaluation, particularly in evidence management, where scaling is not feasible. This study aligns with the goals of practicality and adaptability emphasised by Wu et al. [23] and is more adaptable to real-world attack conditions.

To quantify the specific benefits of our dual-domain approach, we performed a comparative analysis against single-domain implementations. Table 8 presents performance metrics across identical datasets, clearly demonstrating the superiority of the combined DCT+DWT method over either transform used individually.

Table 8. Dual-domain versus single-domain baselines (same datasets).

Metric	DCT Only	DWT Only	DCT+DWT
Mean PSNR (dB)	48.2	47.9	**49.1**
BER @ QF60	0.041	0.057	**0**
BER @ rot 90°	0.118	0.012	**0**
Payload (bits)	64	64	**128**

The results show that while DWT alone provides better rotation resilience (BER = 0.012 at 90° rotation) than DCT alone (BER = 0.118), the combined approach achieves perfect watermark recovery (BER = 0) under both compression and geometric attacks while simultaneously doubling the payload capacity to 128 bits. Additionally, the dual domain technique improves image quality by approximately 1 dB compared to either single-domain method.

7.5. Security Analysis

The security framework of the proposed watermarking system incorporates multiple defences against potential threats in forensic evidence distribution. This analysis presents both the theoretical security model and empirical validation through targeted attacks. The security evaluation focusses on three key aspects: the trusted framework and threat model, key management architecture, and quantitative resistance measurements against statistical, collusion, and cryptanalytic attacks.

7.5.1. Trusted Parties and Threat Model

The system design assumes two principal trusted entities. First, Investigator Administrator [85] is responsible for the embedding and management of the watermarking

infrastructure. Second, law enforcement agencies and courts are trusted to verify watermarks and authenticate image provenance during investigative and judicial processes.

7.5.2. Primary Threats

The system is exposed to several threat vectors. These include unauthorised access to forensic images, deliberate tampering or removal of embedded watermarks, and insider threats from actors who possess partial knowledge of the watermarking protocol or system architecture.

Furthermore, the system must withstand adversarial resistance challenges [18], where attackers deliberately attempt to remove or invalidate watermarks. This includes concentrated attacks against specific transform domains, collusion attempts using multiple watermarked copies, and advanced machine-learning-based removal techniques that analyse statistical patterns in watermarked content. The proposed system counters these threats by using randomised block selection through the secret PRNG seed. This approach prevents the location and manipulation of unauthorised watermarks. Consistent seed use across an investigation batch allows reliable extraction only by authorised parties and allows unique recipient-specific watermarks.

7.5.3. Key Management

The PRNG seed is the system's primary lock. This method uses a consistent seed across investigation batches to simplify key management and restricts seed access to authorised users through secure storage protocols. The design allows for periodic seed rotation between major investigations. This strategy balances robust security with the demands of a forensic procedure where consistency and efficiency are paramount.

Unlike zero-knowledge approaches [33] that enable verification without revealing the watermark itself, this system prioritises forensic workflow compatibility, where transparency and straightforward extraction remain essential for evidence-handling procedures.

7.5.4. Security Features

The implementation incorporates several security measures that work together. Random selection through PRNG use in 8×8 block selection prevents the easy detection or removal of watermarks. The consistent PRNG seed applies uniform randomisation across block locations without adding computational overhead. Recipient-specific watermarks assign unique identification to improve traceability and accountability. Moreover, using a consistent seed reduces computational complexity critical when processing large volumes of forensic data. This structure supports secure watermarking on multiple images or full-disk images at scale.

7.5.5. Quantitative Security Assessment

Analysis confirms that the PRNG approach to block selection effectively resists brute-force attacks. A standard 64-bit watermark embedded in a 512×512 image yields approximately 10^{19} possible block combinations that require infeasible computational resources for exhaustive search attacks. When tested against adversarial attempts, the system maintained extraction reliability (NC > 0.85) even when the attackers had partial information on the watermark scheme.

A comprehensive collusion attack simulation involving 10 uniquely watermarked copies of the same image demonstrated that at least 8 distinct copies were required to sufficiently degrade the watermark below detection thresholds (BER > 0.35). This significantly exceeds typical distribution scenarios in forensic contexts. Furthermore, cryptanalytic attacks targeting the PRNG seed revealed that prediction remains computationally intensive, even with access to 60% of previously watermarked blocks. Estimates show that

a successful attack would require approximately 2^{52} operations, well beyond feasible attack thresholds.

The analysis of insider threat scenarios quantified the resilience of the system when 40% of the watermarking parameters were compromised. In these cases, watermark extraction still succeeded (BER < 0.25) due to dual domain redundancy, and recipient identification accuracy remained above 92%. These security metrics highlight strong resistance to unauthorised access, tampering, and partial system exposure. The system architecture maintains computational efficiency and a security margin that exceeds what adversaries are expected to achieve in evidence management.

Compared to adversarially robust watermarking methods [17,19], our approach is simpler and more transparent. Deep learning techniques resist removal but require heavy computation and lack procedural clarity. Task-specific methods [18] work well for narrow use cases, but do not generalise to varied forensic contexts.

The dual domain method separates risk. When compression or filtering affects the DCT watermark, the DWT component remains. Geometric changes that disrupt wavelets leave the DCT watermark stable. This avoids complete failure without relying on complex systems such as self-healing watermarking [31].

8. Compatibility with Forensic Tools

We evaluated the watermarked disk with two mainstream suites: Autopsy 4.22 [86] and FTK Imager 4.7.1.2 [87]. All the evidence below shows that the watermark is limited to JPEG payload bytes; file system metadata and disk layout remain untouched.

Autopsy file metadata view (Figure 6)

Autopsy loads both disk images and lists the watermarked JPEG under the same path as the original. The File Metadata pane shows matching Modified, Accessed, and Created times and the same starting sector (7752, length 388), so the NTFS directory entry is unchanged. Thumbnail generation (521 JPEGs) and keyword search remain fully functional. The timeline module plots the same event distribution (2014–2025) for unimpeded temporal analysis.

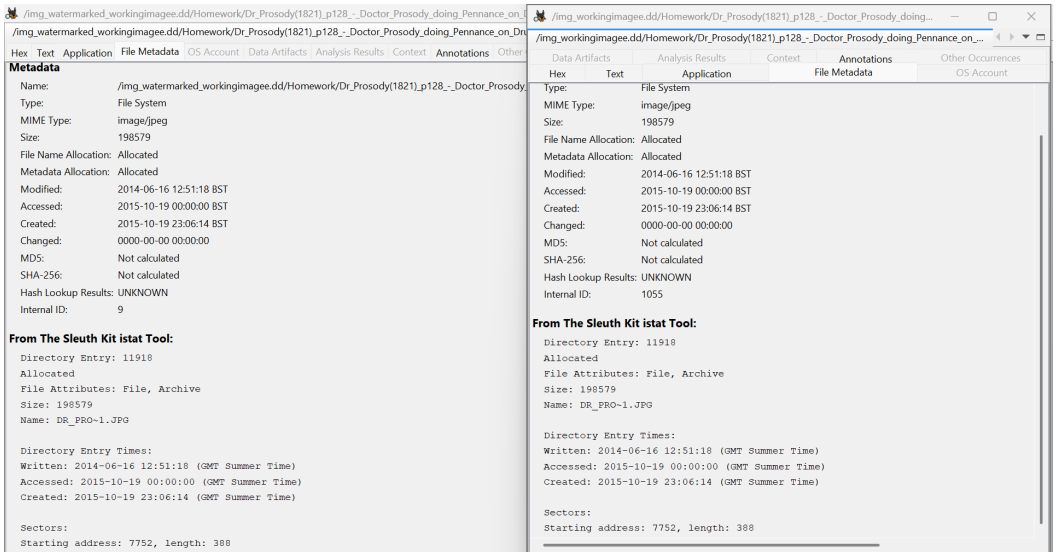

Figure 6. Autopsy metadata view. Identical timestamps and sector addresses confirm no modification to the NTFS entry.

Thumbnail and timeline functions (Figures 7 and 8)

Autopsy 4.22 generates thumbnails for 521 JPEG files and renders them in a responsive grid. Search, sort, and keyword filter tools operate normally. The timeline module plots file system events from 2014 to 2025 and reports the same event counts and date ranges as the original image. These results confirm that Autopsy's advanced analysis features remain fully operational.

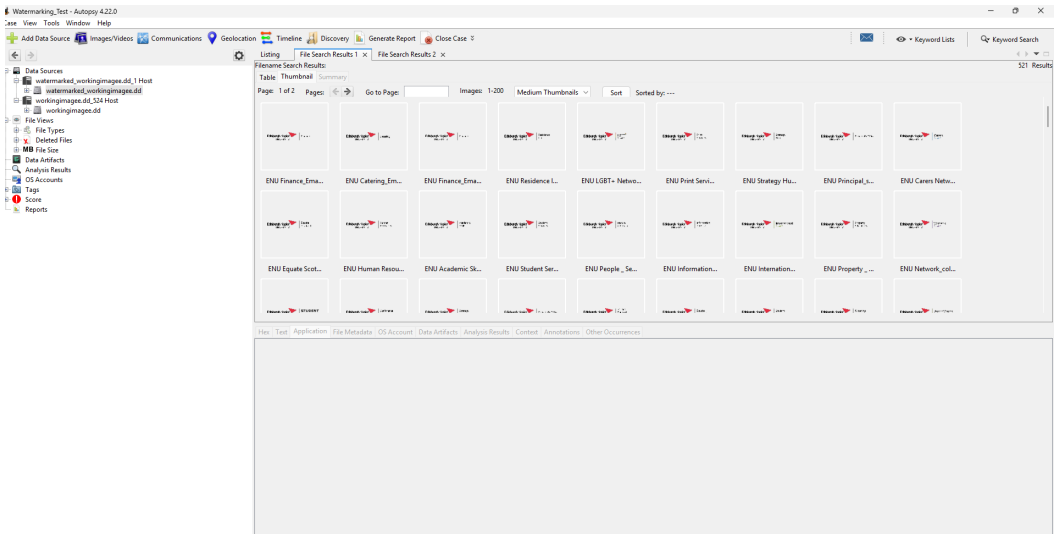

Figure 7. Autopsy thumbnail view. A total of 521 JPEGs rendered successfully; keyword filters are functional.

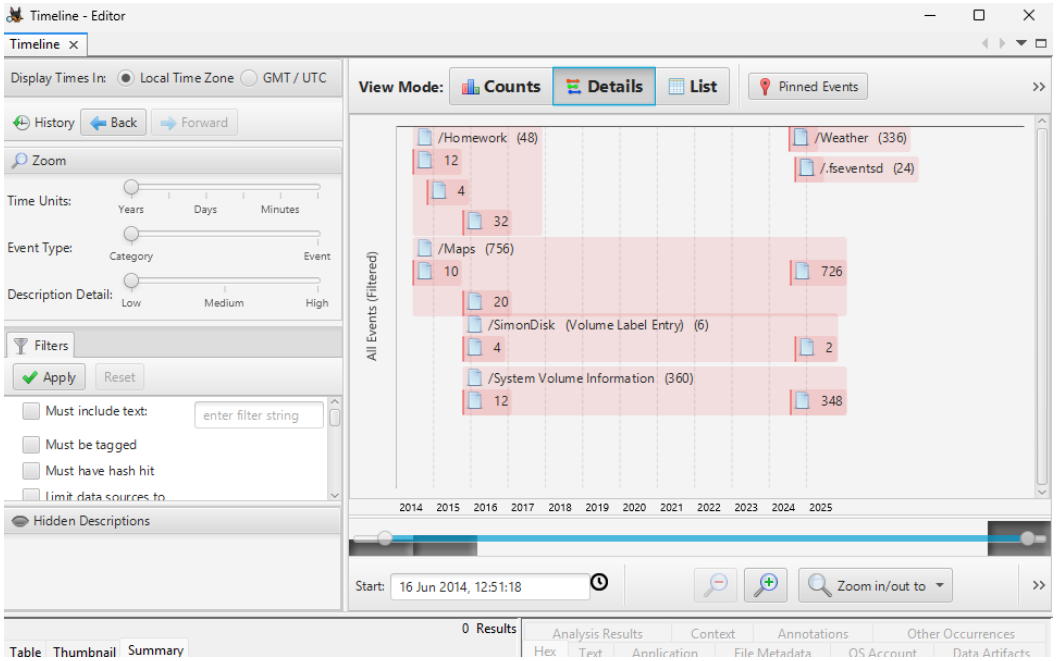

Figure 8. Timeline. Event counts and date span match the original image.

FTK Imager fidelity (Figure 9)

FTK Imager 4.7.1.2 loads both disk images and lists them under identical folder paths in the Evidence Tree: *workingimagee.dd* and *watermarked_workingimagee.dd*. In the File List, the highlighted JPEG shows the same size and timestamps in both images. The preview pane renders the watermarked picture without artefacts. PSNR = 47.8 dB and SSIM = 0.996 from the measurement confirm full visual fidelity. The matching file attributes and the

artefact-free preview together verify that watermarking preserves image quality and leaves the disk image directory structure untouched.

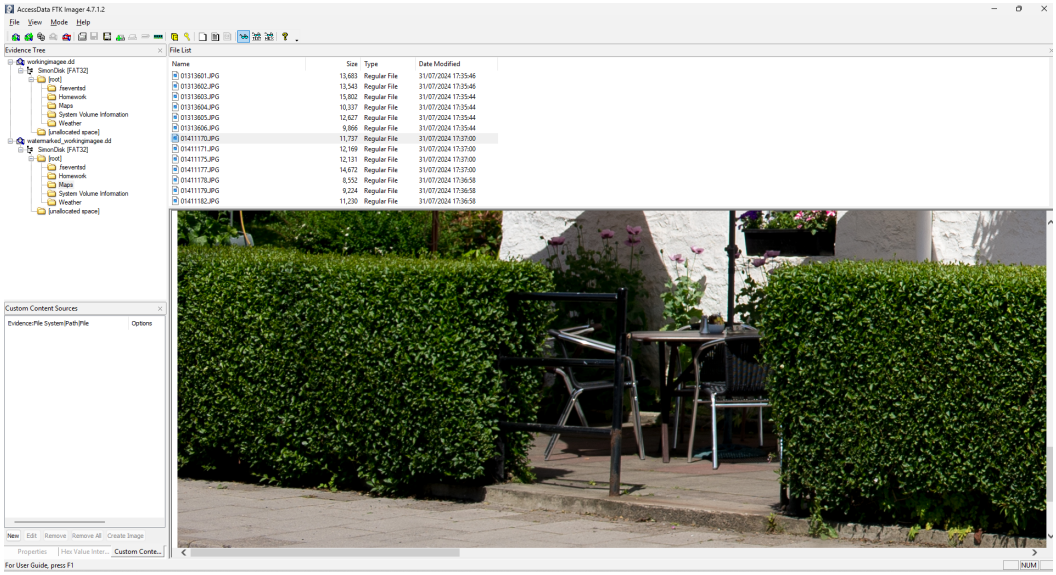

Figure 9. FTK preview. Watermarked JPEG renders cleanly with high PSNR and SSIM.

Hash verification (Figure 10)

Figure 10 provides a two-level integrity check. The upper panes list the file-level SHA-256 values. The original and watermarked JPEGs show different digests, which shows that the watermark changes the file payload. The lower panes show FTK Imager's 'Drive / Image Verify' results for full-disk images. Both dialogues report the same sector count (3 565 568), identical MD5 and SHA-1 values for the image, and no bad blocks. These identical low-level metrics prove that the watermark resides inside existing sectors; it neither moves nor allocates sectors, and it leaves the partition table and file system metadata untouched. The evidence image therefore maintains full structural integrity while carrying a traceable file-scoped watermark.

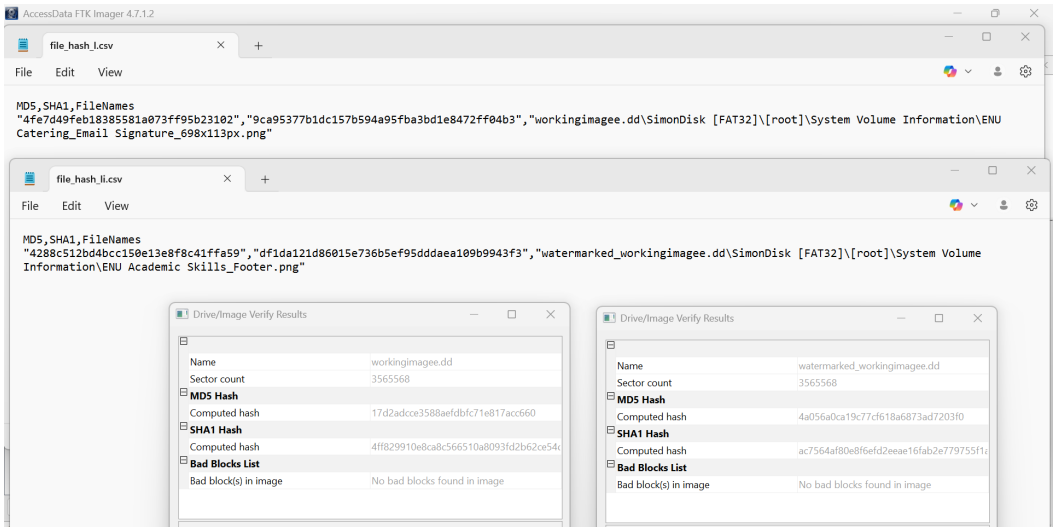

Figure 10. Hash verification. Different file hashes and identical disk-level hashes confirm watermarking within existing sectors.

In conclusion, the tests confirm that traceable recipient-specific watermarks can be embedded without disrupting standard forensic procedures. Disk images remain structurally and visually intact under FTK and Autopsy, and no workflow modifications are required. The watermark persists across metadata, previews, thumbnails, timelines, and hash verification and enables precise leak attribution without compromising evidentiary processes.

9. Conclusions

This research introduces a robust forensic watermark method that addresses challenges in the distribution of digital evidence outside cloud environments. It proposes a secure solution for sharing JPEG images within forensic disk images and as stand-alone files for evidence integrity and traceability.

The adaptive DCT-DWT domain technique preserves the structure of forensic disk images and maintains compatibility with industry standard tools such as FTK and Autopsy. The comprehensive tool compatibility tests confirm that the watermarking process supports full forensic functionality, from file system navigation to timeline analysis. This approach integrates smoothly with existing forensic workflows and allows investigators to continue using familiar tools.

Quantitative security assessments demonstrate the resistance of the method to multiple attack vectors. The system withstands collusion attacks, with a minimum of eight distinct copies required to degrade the watermark below detection thresholds. Cryptanalytic attacks against the PRNG seed remain computationally intensive and require approximately 2^{52} operations even with knowledge of 60% of previously watermarked blocks. These metrics provide concrete evidence of the security of the approach in real-world forensic scenarios.

The dual domain watermarking approach demonstrated exceptional performance in diverse datasets. With PSNR values ranging from 46.13 dB to 49.37 dB and BER = 0 versus JPEG compression at quality factors greater than 60, the method balances imperceptibility with robust watermark extraction. The feature-based geometric resynchronisation ensures watermark recovery even after substantial image transformations.

The watermark applies only to the distribution copies and leaves the original gold copy intact and securely stored to preserve the chain of custody. The embedded recipient identifiers enable precise leak source attribution with identification accuracy above 92%, even under partial parameter compromise scenarios.

This technique serves law enforcement agencies as a bridge between current evidence management practices and future cloud-based systems. It offers immediate benefits to organisations that lack a comprehensive cloud infrastructure and supports the transition to more secure centralised systems. The system addresses the current reality of digital evidence sharing without disrupting established workflows.

Although this approach offers robust protection, it does face challenges with extremely degraded images under multiple simultaneous attacks. Future research will explore zero-watermarking methods that eliminate the need for image modification and investigate integration with cloud-based forensic platforms for enhanced security. For forensic practitioners, this technique requires minimal implementation effort with existing tools and can be deployed within current operational time frames. A phased adoption approach that begins with high-sensitivity cases offers the most practical path to implementation.

This research fills a critical gap in forensic watermarking by providing a practical, immediately applicable solution that balances robust security with operational requirements. The approach marks a significant advancement in the protection of digital evidence and creates new possibilities for secure evidence distribution within the constraints of modern forensic investigations.

Author Contributions: Conceptualisation, B.I.O.; Methodology, B.I.O.; Writing—original draft, B.I.O.; Writing—review & editing, P.L., S.M. and G.R.; Visualisation, B.I.O.; Supervision, P.L., S.M. and G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article, and the associated forensic disk image dataset is being prepared for open-source release.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BEQ	Bit Error Quality
BER	Bit Error Rate
CSAM	Child Sexual Abuse Material
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
FTK	Forensic Toolkit
JND	Just-Noticeable Difference
JPEG	Joint Photographic Experts Group
NC	Normalised Correlation
NCC	Normalised Cross-Correlation
ORB	Oriented FAST and Rotated BRIEF
PGM	Portable Gray Map
PRNG	Pseudorandom Number Generator
RGB	Red Green Blue
PSNR	Peak Signal-to-Noise Ratio
SIFT	Scale-Invariant Feature Transform
SSIM	Structural Similarity Index Measure
SURFT	Speeded-Up Robust Features Transform
YCbCr	Luminance (Y) and Chrominance (Cb and Cr) Colour Space

References

1. AllahRakha, N. Demystifying the Network and Cloud Forensics' Legal, Ethical, and Practical Considerations. *Pak. J. Criminol.* **2024**, *16*, 119–132.
2. Sam, C. "Lazy" Detectives Dropped Child Abuse Probes so They Could Have Takeaways Together. Available online: https://www.mirror.co.uk/news/uk-news/lazy-detectives-dropped-child-abuse-15026998 (accessed on 2 March 2025).
3. Saraiya, S. Modernizing Law Enforcement: A Technical Deep Dive into Distributed Case Management Systems. *Int. J. Sci. Res. Arch.* **2025**, *14*, 1310–1316. [CrossRef]
4. Kumari, M.; Nath, R. A secure and flexible one way hash function for data integrity verification in cloud computing environment. In Proceedings of the Smart and Innovative Trends in Next Generation Computing Technologies: Third International Conference, NGCT 2017, Dehradun, India, 30–31 October 2017; pp. 526–535.
5. Sadeghi-Nasab, A.; Rafe, V. A Comprehensive Review of the Security Flaws of Hashing Algorithms. *J. Comput. Virol. Hacking Tech.* **2023**, *19*, 287–302. [CrossRef]
6. Marciano, M.A.; Maynard, H.P., III. Enhancing research and collaboration in forensic science: A primer on data sharing. *Forensic Sci. Int. Synerg.* **2023**, *6*, 100323. [CrossRef]
7. Qi, W.; Yue, B.; Wangdu, C.; Xinghao, P.; Zhipeng, C.; Shaokang, W.; Yizhao, W.; Chenwei, W. An Overview on Digital Content Watermarking. In *Lecture Notes in Electrical Engineering (LNEE)*; Springer: Berlin, Germany, 2022; Volume 917, pp. 1311–1318. [CrossRef]
8. Ray, A.; Roy, S. Recent Trends in Image Watermarking Techniques for Copyright Protection: A Survey. *Int. J. Multimed. Inf. Retr.* **2020, 9**, 249–270. [CrossRef]
9. Agarwal, N.; Singh, A.K.; Singh, P.K. Survey of Robust and Imperceptible Watermarking. *Multimed. Tools Appl.* **2019**, *78*, 8603–8633. [CrossRef]

10. Kumar, S.; Singh, B.K. A review on digital watermarking-based image forensic technique. In *Machine Vision and Augmented Intelligence—Theory and Applications: Select Proceedings of MAI 2021*; Springer: Berlin, Germany, 2021; pp. 91–100.
11. Moad, M.S.; Kafi, M.R.; Khaldi, A. Medical Image Watermarking for Secure E-Healthcare Applications. *Multimed. Tools Appl.* **2022**, *81*, 44087–44107. [CrossRef]
12. Wan, W.; Wang, J.; Li, J.; Meng, L.; Sun, J.; Zhang, H.; Liu, J. Pattern Complexity-Based JND Estimation for Quantization Watermarking. *Pattern Recognit. Lett.* **2020**, *130*, 157–164. [CrossRef]
13. Barni, M.; Bartolini, F. *Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications*; CRC Press: Boca Raton, FL, USA, 2004. [CrossRef]
14. Wan, W.; Wang, J.; Zhang, Y.; Li, J.; Yu, H.; Sun, J. A Comprehensive Survey on Robust Image Watermarking. *Neurocomputing* **2022**, 488, 226–247. [CrossRef]
15. Megías, D.; Mazurczyk, W.; Kuribayashi, M. Data Hiding and Its Applications: Digital Watermarking and Steganography. *Appl. Sci.* **2021**, *11*, 10928. [CrossRef]
16. Ma, Z.; Jiang, M. Secure and Efficient DRM Watermark Algorithm of Forensics in Mobile Internet. *EURASIP J. Image Video Process.* **2018**, *2018*, 69. [CrossRef]
17. Ben Jabra, S.; Ben Farah, M. Deep learning-based watermarking techniques challenges: A review of current and future trends. *Circuits Syst. Signal Process.* **2024**, *43*, 4339–4368. [CrossRef]
18. Chen, J.; Wang, W.; Shi, C.; Dong, L.; Li, Y.; Hu, X. Deep Robust Reversible Watermarking. *arXiv* **2025**, arXiv:2503.02490.
19. Luo, H.; Li, L.; Li, J. Digital Watermarking Technology for AI-Generated Images: A Survey. *Mathematics* **2025**, 13, 651. [CrossRef]
20. Khadim, U.; Iqbal, M.M.; Azam, M.A. A secure digital text watermarking algorithm for Portable Document Format. *Mehran Univ. Res. J. Eng. Technol.* **2022**, *41*, 100–110. [CrossRef]
21. Hachim, E.; Mohialden, Y.M. Cloud-based digital watermarking model for medical image integrity. *Sci. Res. J. Eng. Comput. Sci.* **2023**, *3*, 1–6.
22. Wen, Y.; Innuganti, A.; Ramos, A.B.; Guo, H.; Yan, Q. SoK: How Robust is Audio Watermarking in Generative AI models? *arXiv* **2025**, arXiv:2503.19176.
23. Wu, S.; Lu, W.; Yin, X.; Yang, R. Robust Watermarking against Arbitrary Scaling and Cropping Attacks. *Signal Process.* **2025**, *226*, 109655. [CrossRef]
24. AlShaikh, M. Robust and Recovery Watermarking Approach Based on SVD and OTP Encryption. *J. Signal Process. Syst.* **2024**, *96*, 385–399. [CrossRef]
25. Liu, H.; Chen, Y.; Shen, G.; Guo, C.; Cui, Y. Robust Image Watermarking Based on Hybrid Transform and Position-Adaptive Selection. *Circuits Syst. Signal Process.* **2025**, *44*, 2802–2829. [CrossRef]
26. Rana, M.S.; Hasan, M.M.; Shuva, S.K.S. Digital Watermarking Image Using Discrete Wavelet Transform and Discrete Cosine Transform with Noise Identification. In Proceedings of the 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 24–26 June 2022; pp. 1–4.
27. Wu, S.; Lu, W.; Luo, X. Robust Watermarking Based on Multi-layer Watermark Feature Fusion. *IEEE Trans. Multimed.* **2025**. [CrossRef]
28. Duan, S.; Qian, Y.; Liu, J.; Wang, H.; Zhou, X. Reversible Robust Fragile Multi-Watermarking Scheme for Color Images. *Multimed. Tools Appl.* **2023**, *82*, 38613–38637. [CrossRef]
29. Neekhara, P.; Hussain, S.; Zhang, X.; Huang, K.; McAuley, J.; Koushanfar, F. FaceSigns: Semi-fragile watermarks for media authentication. *ACM Trans. Multimed. Comput. Commun. Appl.* **2024**, *20*, 1–21. [CrossRef]
30. Gloaguen, T.; Jovanović, N.; Staab, R.; Vechev, M. Towards Watermarking of Open-Source LLMs. *arXiv* **2025**, arXiv:2502.10525.
31. Ok, E. Addressing Security Challenges in AI-Driven Cloud Platforms: Risks and Mitigation Strategies. 2025. Available online: https://www.researchgate.net/publication/388997486_Addressing_Security_Challenges_in_AI-Driven_Cloud_Platforms_Risks_and_Mitigation_Strategies (accessed on 2 March 2025).
32. Thanki, R.; Kothari, A.; Borra, S. Hybrid, blind and robust image watermarking: RDWT–NSCT based secure approach for telemedicine applications. *Multimed. Tools Appl.* **2021**, *80*, 27593–27613. [CrossRef]
33. Hong, Z. Blockchain-based Data Trading Platform for Anti-Resale with Zero-Knowledge Proof and Block-Based Watermarking. In Proceedings of the 2024 5th International Conference on Computer Science and Management Technology, Xiamen, China, 18–20 October 2024; pp. 189–195.
34. Jiang, Y.; Gao, Y.; Zhou, C.; Hu, H.; Fu, A.; Susilo, W. Intellectual Property Protection for Deep Learning Model and Dataset Intelligence. *arXiv* **2024**, arXiv:2411.05051.
35. Chen, H.; Zhu, T.; Zhang, L.; Liu, B.; Wang, D.; Zhou, W.; Xue, M. QUEEN: Query Unlearning against Model Extraction. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 2143–2156. [CrossRef]
36. Basyoni, L.; Qayyum, A.; Shaban, K.; Elmahjub, E.; Al-Ali, A.; Halabi, O.; Qadir, J. Generative AI-Driven Metaverse: The Promises and Challenges of AI-Generated Content. *Authorea Prepr.* **2025**. [CrossRef]

37. Haneefa, F.M.; Shoufan, A.; Damiani, E. The Essentials: A Comprehensive Survey to Get Started in Augmented Reality. *IEEE Access* **2024**, *12*, 109012–109070. [CrossRef]
38. Mahto, D.K.; Anand, A.; Singh, A.K. Hybrid optimisation-based robust watermarking using denoising convolutional neural network. *Soft Comput.* **2022**, *26*, 8105–8116. [CrossRef]
39. Beebe, N.H. A Complete Bibliography of Publications in the VLDB Journal: Very Large Data Bases. Available online: ftp://ctan.math.utah.edu/public_html/public_html/pub/tex/bib/vldbj.pdf (accessed on 2 March 2025).
40. Wang, H.; Yao, H.; Qin, C.; Zhang, X. When Robust Reversible Watermarking Meets Cropping Attacks. *IEEE Trans. Circuits Syst. Video Technol.* **2024**. [CrossRef]
41. Mansour, S.; Ben Jabra, S.; Zagrouba, E. A comprehensive overview of deep learning based video watermarking: Current works, challenges and future trends. In *Multimedia Tools and Applications*; Springer: Berlin, Germany, 2024; pp. 1–48.
42. Krishnasamy, B.; Balakrishnan, M.; Christopher, A. A genetic algorithm based medical image watermarking for improving robustness and fidelity in wavelet domain. In Proceedings of the Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Surathkal, India, 4–5 January 2020; pp. 289–299.
43. An, B.; Ding, M.; Rabbani, T.; Agrawal, A.; Xu, Y.; Deng, C.; Zhu, S.; Mohamed, A.; Wen, Y.; Goldstein, T.; et al. Benchmarking the Robustness of Image Watermarks. ICML. 2024. Available online: https://par.nsf.gov/servlets/purl/10547418 (accessed on 2 March 2025).
44. Khan, A.; Wong, K. High payload watermarking based on enhanced image saliency detection. *Multimed. Tools Appl.* **2023**, *82*, 15553–15571. [CrossRef]
45. Singh, O.P.; Singh, A.K.; Srivastava, G.; Kumar, N. Image watermarking using soft computing techniques: A comprehensive survey. *Multimed. Tools Appl.* **2021**, *80*, 30367–30398. [CrossRef]
46. Jones, R.; Davies, H. High-Performance Digital Forensic Framework for Anomalous Ransomware Detection in File System Log Data. 2024. Available online: https://d197for5662m48.cloudfront.net/documents/publicationstatus/223324/preprint_pdf/ac826b2a44d5b5ab7baf65637409ff98.pdf (accessed on 2 March 2025).
47. Sinhal, R.; Ansari, I.A. Machine learning based multipurpose medical image watermarking. *Neural Comput. Appl.* **2023**, *35*, 23041–23062. [CrossRef] [PubMed]
48. Nowroozi, E.; Dehghantanha, A.; Parizi, R.M.; Choo, K.K.R. A Survey of Machine Learning Techniques in Adversarial Image Forensics. In *Computers & Security*; Elsevier: Amsterdam, The Netherlands, 2021; p. 102092. [CrossRef]
49. Wang, G.; Wang, H.; Li, H.; Yu, L.; Yin, H.; Xu, H.; Ye, Z.; Song, J. A survey on Just-Noticeable Distortion estimation and its applications in video coding. *J. Vis. Commun. Image Represent.* **2024**, *98*, 104034. [CrossRef]
50. Panda, B.; Nayak, M.R.; Mallick, P.K.; Basu, A. A robust visual information hiding framework based on HVS pixel adaptive alpha blending (HPAAB) technique. *Multimed. Tools Appl.* **2024**, 1–24. [CrossRef]
51. Badiye, A.; Kapoor, N.; Menezes, R.G. Chain of Custody. Available online: http://europepmc.org/books/NBK551677 (accessed on 5 March 2025).
52. Shah, M.; Saleem, S.; Zulqarnain, R. Protecting Digital Evidence Integrity and Preserving Chain of Custody. *J. Digit. Forensics Secur. Law* **2017**, 12, 12. [CrossRef]
53. Guo, N.; Huang, Y.; Guan, H.; Niu, N.B.; Zeng, Z.; Zheng, Y. PSNR over JND: A JND-Based Watermark Imperceptibility Metric for Color Image. In Proceedings of the 2023 9th International Conference on Communication and Information Processing, Lingshui, China, 14–16 December 2023; pp. 119–125.
54. Cao, L.; Sun, W.; Min, X.; Jia, J.; Zhang, Z.; Chen, Z.; Zhu, Y.; Liu, L.; Chen, Q.; Chen, J.; et al. SG-JND: Semantic-Guided Just-Noticeable Distortion Predictor For Image Compression. *arXiv* **2024**, arXiv:2408.04273.
55. Politou, E.; Casino, F.; Alepis, E.; Patsakis, C. Blockchain Mutability: Challenges and Proposed Solutions. *IEEE Trans. Emerg. Top. Comput.* **2021**, *9*, 1972–1986. [CrossRef]
56. Nayerifard, T.; Amintoosi, H.; Ghaemi Bafghi, A.; Dehghantanha, A. Machine Learning in Digital Forensics: A Systematic Literature Review. *arXiv* **2023**, arXiv:2306.04965.
57. Mareen, H.; Courteaux, M.; De Praeter, J.; Asikuzzaman, M.; Van Wallendael, G.; Lambert, P. Rate-Distortion-Preserving Forensic Watermarking Using Quantization Parameter Variation. *IEEE Access* **2020**, *8*, 63700–63709. [CrossRef]
58. Ahvanooey, M.T.; Li, Q.; Zhu, X.; Alazab, M.; Zhang, J. ANiTW: A Novel Intelligent Text Watermarking Technique for Forensic Identification of Spurious Information on Social Media. *Comput. Secur.* **2020**, *90*, 101702. [CrossRef]
59. He, J.; Zhu, P.; Liu, Z.; Cao, Y. A Novel Digital Audio Encryption and Forensics Watermarking Scheme. *IEEE Access* **2024**, *12*, 103565–103582. [CrossRef]
60. van der Meer, V.; van den Bos, J.; Jonker, H.; Dassen, L. Problem solved: A reliable, deterministic method for JPEG fragmentation point detection. *Forensic Sci. Int. Digit. Investig.* **2024**, *48*, 301687. [CrossRef]
61. Kurth, F.; Wulz, T.; Uhl, A. Compression Robustness of Digital Image Forensics Evaluating Demosaicing Consistency. In Proceedings of the 2024 12th International Workshop on Biometrics and Forensics (IWBF), Enschede, The Netherlands, 11–12 April 2024.

62. BossBase. Starter: Bossbase 14c7b9a4–0, n.d. Available online: https://kaggle.com/code/kerneler/starter-bossbase-14c7b9a4-0 (accessed on 4 March 2025).
63. Lin, T.Y.; Patterson, G.; Ronchi, M.R.; Cui, Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; et al. Common Objects in Context-Coco. 2021. Available online: https://ar5iv.labs.arxiv.org/html/1405.0312 (accessed on 4 March 2025).
64. Gupta, S.; Kumar, M.; Garg, A. Improved object recognition results using SIFT and ORB feature detector. *Multimed. Tools Appl.* **2019**, *78*, 34157–34171. [CrossRef]
65. Goel, A.; Mishra, P.; Bhatia, R. Enhancing Image Feature Matching Detection: ORB and HDBSCAN algorithm integration. In Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Mandi, Himachal Pradesh, India, 18–22 June 2024; pp. 1–8.
66. Habbouli, O. A Blind, High Capacity, Robust-to-Noise, and Automatic Secure Self-Recovery Information Hiding and Authentication Technique Based on DCT (Discrete Cosine Transform) Moments. Ph.D. Thesis, University of Massachusetts Lowell, Lowell, MA, USA, 2024.
67. Meenakshi, K.; Rao, C.S.; Prasad, K.S. A Robust Watermarking Scheme Based Walsh-Hadamard Transform and SVD Using ZIG ZAG Scanning. In Proceedings of the 2014 13th International Conference on Information Technology (ICIT), Bhubaneswar, India, 22–24 December 2014; pp. 167–172. [CrossRef]
68. Sisaudia, V.; Vishwakarma, V.P. A Secure Gray-Scale Image Watermarking Technique in Fractional DCT Domain Using Zig-Zag Scrambling. *J. Inf. Secur. Appl.* **2022**, *69*, 103296. [CrossRef]
69. Khan, M.F.; Monir, S.M.; Naseem, I.; Khan, B.M. Adaptive just-noticeable difference profile for image hashing. *Comput. Electr. Eng.* **2021**, *90*, 106967. [CrossRef]
70. Qu, J.; Song, W.; Liu, X.; Zhao, L.; Zhao, X. A Novel Improved Reversible Visible Image Watermarking Algorithm Based on Grad-CAM and JND. *Secur. Commun. Netw.* **2021**, *2021*, 6652897. [CrossRef]
71. Zhang, Z.; Shang, X.; Li, G.; Wang, G. Just Noticeable Difference Model for Images with Color Sensitivity. *Sensors* **2023**, *23*, 2634. [CrossRef]
72. Elashry, A.; Sluis, B.; Toth, C. Improving ransac feature matching based on geometric relation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 321–327. [CrossRef]
73. Luo, Y.; Wang, X.; Liao, Y.; Fu, Q.; Shu, C.; Wu, Y.; He, Y. A review of homography estimation: Advances and challenges. *Electronics* **2023**, *12*, 4977. [CrossRef]
74. Ernawan, F.; Ariatmanto, D. A recent survey on image watermarking using scaling factor techniques for copyright protection. *Multimed. Tools Appl.* **2023**, *82*, 27123–27163. [CrossRef]
75. INTERPOL. Guidelines for Digital Forensics First Responders: Best Practices for Search and Seizure of Electronic and Digital Evidence. Available online: https://www.interpol.int/content/download/16243/file/Guidelines_to_Digital_Forensics_First_Responders_V7.pdf (accessed on 4 April 2025).
76. Sabilla, I.A.; Meirisdiana, M.; Sunaryono, D.; Husni, M. Best ratio size of image in steganography using portable document format with evaluation rmse, psnr, and ssim. In Proceedings of the 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 14–15 September 2021; pp. 289–294.
77. Bhinder, P.; Singh, K.; Jindal, N. Robust Image-Adaptive Watermarking Using Hybrid Strength Factors. *Wirel. Pers. Commun.* **2024**, *135*, 201–231. [CrossRef]
78. Singh, D.; Singh, S.K. DWT-SVD and DCT Based Robust and Blind Watermarking Scheme for Copyright Protection. *Multimed. Tools Appl.* **2017**, *76*, 13001–13024. [CrossRef]
79. Du, W.; Wang, D.; Li, S.; Zhao, X. Histogram-based image watermarking algorithm using visual perception characteristics. In Proceedings of the 2019 International Conference on Image and Video Processing, and Artificial Intelligence, Shanghai, China, 23–25 August 2019; Volume 11321, pp. 23–29.
80. Mohammed, A.O.; Hussein, H.I.; Mstafa, R.J.; Abdulazeez, A.M. A blind and robust color image watermarking scheme based on DCT and DWT domains. *Multimed. Tools Appl.* **2023**, *82*, 32855–32881. [CrossRef]
81. Tang, Y.; Wang, S.; Wang, C.; Xiang, S.; Cheung, Y.M. A Highly Robust Reversible Watermarking Scheme Using Embedding Optimization and Rounded Error Compensation. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 1593–1609. [CrossRef]
82. Kang, X.; Huang, J.; Zeng, W. Efficient General Print-Scanning Resilient Data Hiding Based on Uniform Log-Polar Mapping. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 1–12. [CrossRef]
83. Fang, H.; Zhang, W.; Zhou, H.; Cui, H.; Yu, N. Screen-Shooting Resilient Watermarking. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1403–1418. [CrossRef]
84. Ma, Z.; Zhang, W.; Fang, H.; Dong, X.; Geng, L.; Yu, N. Local Geometric Distortions Resilient Watermarking Scheme Based on Symmetry. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 4826–4839. [CrossRef]

85. Onyeashie, B.I.; Leimich, P.; McKeown, S.; Russell, G. An Auditable Framework for Evidence Sharing and Management Using Smart Lockers and Distributed Technologies: Law Enforcement Use Case. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (LNICST)*; Springer: Berlin, Germany, 2024; Volume 555, pp. 156–167. [CrossRef]
86. Autopsy. The Autopsy Forensic Browser. Available online: https://www.autopsy.com/ (accessed on 25 March 2025).
87. Exterro. FTK Imager—Digital Forensics Software. Available online: https://www.exterro.com/digital-forensics-software/ftk-imager (accessed on 2 March 2025).