FACTA UNIVERSITATIS
Series: **Electronics and Energetics** Vol. 38, Nº 4, December 2025, pp. 581 - 603
https://doi.org/10.2298/FUEE2504581C

Invited paper

OVERVIEW OF DIGITAL IMAGE FORGERY DETECTION

Petar Čisar[1,2]

[1]University of Criminal Investigation and Police Studies, Zemun-Belgrade, Serbia
[2]John von Neumann University, GAMF - Faculty of Engineering and Computer Science, Kecskemét, Hungary

ORCID iD: Petar Čisar https://orcid.org/0000-0001-8009-3347

Abstract. *Digital image forgery detection is a vital area of research in digital forensics, aiming to authenticate visual content in the face of increasingly sophisticated manipulation techniques. This paper presents a comprehensive overview of the field, integrating key concepts in its technical landscape. The categorization of detection methods typically includes active approaches that depend on embedded watermarks or signatures, and passive (blind) techniques that analyze the image content itself without prior information. Underlying these are various detection principles, such as identifying inconsistencies in pixel patterns, compression artifacts, illumination, and sensor noise.*
The paper explores the specific characteristics of detection techniques, analyzing their strengths and limitations. Pixel-based and statistical methods offer efficiency for copy–move and splicing detection but often lack robustness under compression or scaling. Frequency-domain methods and physics-based analysis provide deeper insights, but they can be computationally intensive or sensitive to environmental conditions. The evaluation of detection models is crucial, relying on diverse datasets, realistic manipulation scenarios, and adversarial robustness testing. Effective evaluation metrics include accuracy, precision, recall, F1-score, AUC-ROC and IoU, which collectively assess classification and localization performance.
The deep learning approach in forgery detection has significantly advanced the field, with convolutional neural networks and transformer-based models learning complex tampering artifacts. However, challenges persist in forgery detection, including evolving manipulation methods, dataset limitations, explainability concerns, and vulnerability to adversarial attacks. Finally, the authors discuss trends and future directions, such as self-supervised learning, multimodal forensic integration, domain adaptation, and real-time detection frameworks, paving the way for more resilient and scalable forensic tools.

Key words: *image forgery detection, categorization, detection principles, characteristics, evaluation.*

Received July 11, 2025; accepted July 13, 2025
Corresponding author: Petar Čisar
University of Criminal Investigation and Police Studies. Cara Dušana 196, 11080 Zemun-Belgrade, Serbia
John von Neumann University, GAMF - Faculty of Engineering and Computer Science, Izsáki u. 10, Kecskemét, Hungary
E-mail: petar.cisar@kpu.edu.rs, csiszar.peter@nje.hu

1. INTRODUCTION

In an era characterized by the exponential growth of digital media, images play an integral role in communication, journalism, social media, scientific research, and legal evidence. The ease of image acquisition, manipulation, and dissemination has drastically increased with the availability of powerful editing tools such as Adobe Photoshop, GIMP, and AI-powered generative models like GANs (Generative Adversarial Networks). Although these tools offer creative flexibility and have numerous legitimate applications, they also pose significant risks to the authenticity and reliability of visual information.

Image forgery refers to the intentional alteration of an image to deceive viewers or convey false information. These manipulations can range from simple edits, such as contrast adjustment or object removal, to complex tampering involving seamless integration of objects or synthetic generation of entire scenes. As such forgeries become increasingly convincing, the need for robust and reliable image forgery detection techniques becomes imperative to ensure the integrity of digital images in critical fields such as forensics, journalism, law enforcement, and cybersecurity.

1.1. Challenges in image forgery detection

Detecting image forgery presents several technical and practical challenges. One major challenge is the high visual quality of modern editing tools, which can produce visually seamless forgeries that are nearly impossible to detect with the naked eye. Additionally, the wide variety of manipulations occurring at the pixel, structural, or semantic levels and across different formats and compression types adds to the complexity. The lack of ground truth is another significant hurdle, as there are limited publicly available datasets containing both authentic and manipulated image pairs, making training and evaluation difficult. Furthermore, common image-processing operations, such as compression, resizing, and the addition of noise, can obscure signs of manipulation. Lastly, adversarial techniques, particularly those involving adversarial networks, can generate content specifically designed to deceive detection systems.

1.2. Objectives of the paper

This paper aims to explore and elaborate various methods for image forgery detection through several key objectives. First, it seeks to classify and define different types of image forgeries along with their distinguishing characteristics. It then surveys existing approaches and algorithms used in forgery detection, covering both traditional techniques and those based on deep learning.

The central aim addressed in this paper is to analyze the characteristics of different techniques for detecting image forgeries, with a focus on choosing the most appropriate method for a given case.

The paper also describes the process of evaluating detection models using appropriate datasets and performance metrics. Finally, it discusses the limitations of current techniques and proposes potential directions for future research.

1.3. Structure of the paper

The remainder of this paper is structured to ensure a coherent progression from the foundational concepts to the evaluation and discussion of image-forgery detection methods.

After the introduction, Section 2 presents a review of related work, categorizing existing approaches into traditional signal-based methods, transform-domain techniques, and deep learning-based frameworks, with an emphasis on their development and key contributions.

Section 3 introduces the core concepts of digital image forgery detection, including the taxonomy of manipulation types and an overview of detection strategies. It establishes the theoretical basis for understanding how various techniques operate in practice.

Section 4 provides a detailed analysis of the specific characteristics of different detection methods, examining their strengths, limitations, and applicability to various tampering scenarios. This section also highlights comparative insights across different methodological categories.

Section 5 describes the experimental evaluation framework used to assess the performance of detection models. It outlines the datasets, evaluation metrics, and robustness testing procedures used to ensure the reliability and generalization of the results.

Section 6 focuses on the deep learning approach, elaborating on the architecture and training of neural models, as well as their ability to capture complex forgery patterns. Special attention is given to the challenges of explainability and dataset dependency.

Section 7 discusses the open challenges in the field, including issues of generalization, adversarial robustness, and the detection of subtle or small-scale manipulations.

Section 8 explores emerging research trends and future directions, such as multimodal forensics, universal detection models, and blockchain-based provenance solutions that aim to improve trust and transparency in digital media.

Finally, Section 9 concludes the paper by summarizing the main findings and highlighting the implications of the reviewed approaches for the development of more resilient and interpretable image forgery detection systems.

2. Related Work

Digital forensics and digital image forgery detection are closely interconnected fields that contribute significantly to the identification, analysis, and presentation of digital evidence, particularly in the context of cybercrime. Digital image forgery detection, as a subfield of digital forensics, focuses on identifying manipulations within visual content, which are increasingly exploited in online fraud, misinformation, and other illicit activities. As described in [1], digital forensics provides the tools and techniques to examine digital traces and support legal processes, including image content. Furthermore, the paper [2] emphasizes the importance of structured science-based procedures in forensic investigations, which are also applicable to the detection of inconsistencies in images. Additionally, the study [3] highlights the growing relevance of multimedia analysis as a direction of development in digital forensics. Together, these references demonstrate that image forgery detection is an essential component of modern digital forensics, providing critical insights and evidence in both investigative and judicial contexts.

The field of image forgery detection has evolved significantly over the past two decades, with researchers developing various techniques to identify and localize tampered regions in digital images. The related work can be broadly categorized into three main areas: traditional (signal-based) methods, transform-domain and physics-based approaches, and recent deep learning-based methods.

The paper [4] provides foundational information on early approaches to image forgery detection, complementing the broader categorization and evaluation strategies discussed in this overview.

Traditional methods operate primarily in the spatial domain and are based on identifying inconsistencies in pixel patterns, noise, and statistical artifacts. Techniques such as block-based and keypoint-based copy-move forgery detection have been explored using DCT [5], PCA [6], Zernike moments [7], SIFT [8], SURF [9] and ORB [10]. Post-processing steps such as RANSAC [11] enhance the robustness of these methods. Splicing detection leverages CFA interpolation inconsistencies [12], JPEG artifacts [13], and PRNU (Photo Response Non-Uniformity) analysis [14]. Error Level Analysis (ELA) also offers intuitive visual cues for JPEG manipulation [15].

Transform-domain methods use frequency analysis to identify subtle manipulations. DCT and DWT have been widely applied to detect double compression and frequency inconsistencies [16], [17]. PRNU-based analysis in the frequency domain, introduced by [14], remains a fundamental method for device identification. Geometric and illumination-based methods provide complementary insights, especially for splicing detection [18], [19]. Techniques for analyzing light source direction [20], shadow geometry [21], and perspective inconsistencies [22] improve explainability and analysis of scene-level consistency.

The introduction of CNNs revolutionized forgery detection. Supervised architectures such as MesoNet [23], ManTra-Net [24], and TamperNet [25] have demonstrated high performance on standard benchmarks. Transformer-based models [26], multi-scale and attention mechanisms [27], and segmentation models such as U-Net [28] further enhance tampering localization. NoisePrint [29] exemplifies noise-based CNNs, while XceptionNet [30] and other architectures have been proven effective in GAN-generated forgery detection. Hybrid models that combine metadata, illumination, and semantic cues offer greater resilience [25], [31].

Standard datasets such as CASIA [32], Columbia [33], CoMoFoD [34], and FaceForensics++ [30] are used for training and benchmarking. Evaluation metrics include F1-score, IoU, and AUC-ROC [35]. Recent work highlights challenges in cross-dataset generalization and robustness against adversarial manipulation [36].

Emerging trends include explainable AI for forensic reasoning [37], domain adaptation [38], zero-shot detection [39], and multimodal analysis [40]. Blockchain-based provenance tracking [41] and universal detection frameworks [42] aim to tackle real-world applicability and content authentication.

Furthermore, the paper [43] on how image enhancement techniques influence facial detection highlights the impact of image pre-processing, an important consideration in the reliability of modern forgery detection processes.

3. IMAGE FORGERY DETECTION

Image forgery detection involves the identification of digital images tampered or manipulated to ensure their authenticity. Tampering can involve adding, removing, or altering content in an image. With the rise of powerful editing tools, detecting forged images has become increasingly important in fields such as journalism, forensics, and legal investigations.

3.1. Categorization of detection methods

In image forgery detection, traditional methods typically involve analyzing visual or statistical inconsistencies in the spatial domain, such as abrupt changes in pixel values,

noise patterns, or texture irregularities. These methods are effective for detecting common manipulations like copy-move or splicing.

On the other hand, examines the image in a frequency or transformed space using techniques like discrete cosine transform (DCT), discrete wavelet transform (DWT), or Fourier transform. This approach can reveal hidden artifacts or anomalies introduced by forgery that are not easily visible in the spatial domain, such as inconsistencies in compression or edge artifacts. Transform-domain methods are particularly useful for detecting subtle manipulations and post-processing operations.

There are two main categories of image forgery detection techniques. Active methods rely on embedded information, such as watermarks or digital signatures added at the time of image capture. Any modification to the image disrupts this information, making it easier to detect tampering. However, this method requires prior preparation and cannot be used for images without embedded data. Passive (blind) methods do not require prior information and are more commonly used. They analyze inconsistencies within the image itself, such as copy-move forgery detection, splicing detection, image retouching detection and compression artifact analysis (detects tampering through inconsistencies in JPEG compression).

Copy-move forgery (CMF) refers to a manipulation technique in which a region within an image is copied and pasted to another location within the same image to conceal or duplicate elements. This technique is often used to hide objects or create false repetition, as shown in Fig. 1.

Fig. 1 Copy-move forgery (a) original image, b) forged image, c) detection of the CMF region) [44]

Splicing, also known as *image composition*, involves combining two or more images to create a single forged image. This method often alters the semantics of the image by introducing external elements, as shown in Fig. 2.

(a) (b) (c)

Fig. 2 Image splicing (a) image used in splicing, and (b) image used in splicing, (c) spliced image [45]

Retouching refers to subtle modifications made to enhance or diminish certain features in an image, often for aesthetic or misleading purposes. This technique is commonly used in advertising and social media platforms.

Additionally, emerging technologies such as GANs have introduced a new class of image forgeries known as deepfakes, which can generate highly realistic yet entirely fabricated images. Modern approaches also use machine learning and deep learning techniques, such as convolutional neural networks (CNNs), to learn features of authentic and tampered images, offering high accuracy and automation.

The following figure shows a general overview of image forgery methods.

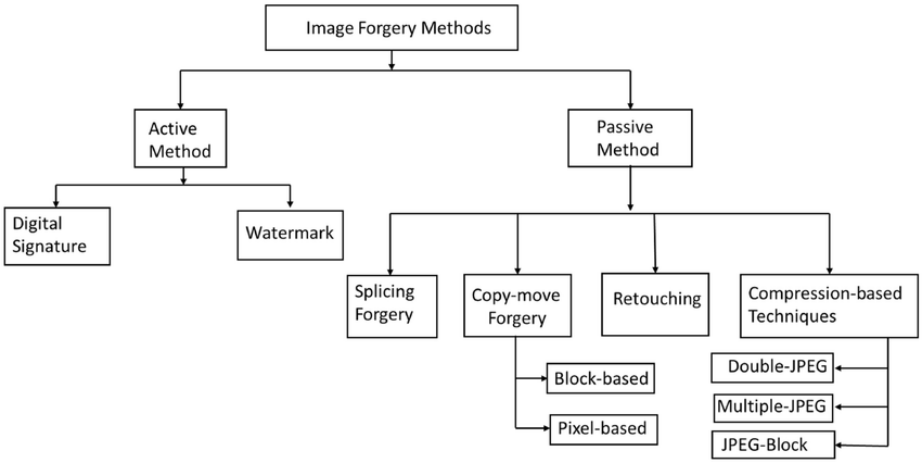

Fig. 3 Overview of image forgery methods [46]

3.1.1. Signal-based (traditional) methods

Taking into account the different detection principles, multiple categories and methods can be distinguished. Signal-based (traditional) methods are oriented towards pixel-level and statistical analysis. Copy-move detection can be realized using different methods. Block-based methods split the image into blocks and extract features (DCT, PCA (Principal Component Analysis), Zernike moments), followed by matching similar blocks. Keypoint-based methods detect keypoints (Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF), Oriented FAST, and Rotated BRIEF (ORB)) and match feature descriptors. Post-processing typically includes the RANSAC (Random Sample Consensus) method for removing geometrically inconsistent matches and filtering.

Splicing detection focuses on identifying signs of double compression, indicating that the region was saved separately before being pasted into the target image. These signs can include inconsistent noise patterns, JPEG blocking artifacts, and different color filter array (CFA) interpolation artifacts.

ELA is a passive digital image forensics technique used in image forgery detection, particularly effective for identifying manipulated or spliced regions in JPEG images. The process involves recompressing the image and measuring the local error caused by compression. Inconsistent error levels across different areas of the image can indicate tampered regions.

3.1.2. Transform-domain analysis

Transform-domain analysis is dominantly focused on the frequency domain and transformed space. DCT/DWT-based features detect inconsistencies in frequency content: JPEG artifacts, double compression, etc. PRNU extracts the PRNU 'fingerprint' of the sensor; inconsistencies betray tampering. Missing/altered PRNU patterns suggest splicing or replacement. CFA interpolation pattern detection detects inconsistencies in demosaicing artifacts and is useful for camera-origin authentication.

3.1.3. Deep learning-based methods

In image forgery detection, deep learning methods have emerged as powerful tools capable of automatically learning complex features from data without the need for handcrafted rules. Techniques such as CNNs are widely used to detect subtle inconsistencies in texture, edges, or patterns that may indicate tampering. These models can be trained on large datasets of authentic and forged images to distinguish between manipulated and original content. Deep learning approaches often outperform traditional methods, especially in detecting sophisticated forgeries like those created by GANs, though they may require substantial data and computational resources. The goal of deep learning-based methods is to learn spatial and semantic patterns that indicate manipulation without hand-crafted features.

Supervised CNN-based methods involve training CNNs on labeled data that distinguish between forged and authentic regions or images. Notable example architectures include TamperNet, MesoNet, and ManTra-Net, which uses a spatial pyramid structure. These models can also incorporate segmentation networks such as U-Net to accurately localize tampered regions within an image. The features used by these models are typically learned directly from raw pixel data but can also be extracted from frequency or gradient domains to improve detection accuracy.

Multiscale and attention-based networks are designed to combine global context, such as overall scene consistency, with local anomalies like blending artifacts. These networks use attention layers to focus on potentially suspected regions, particularly edges and junctions where tampering is more likely to occur. Recently, transformer-based architectures, such as those built on the Swin Transformer, have been gaining popularity for their effectiveness in capturing both local and global features in image forgery detection.

Noise- and artifact-based detection using CNNs focuses on identifying inconsistencies in camera-specific noise patterns. One prominent approach is NoisePrint, where a convolutional neural network is trained to extract unique noise residuals for each camera model. By analyzing these residuals, the method can detect inconsistencies in noise levels between tampered and genuine regions, revealing potential forgeries.

GAN detection focuses on identifying images that have been altered or entirely generated by generative adversarial networks, such as StyleGAN or DeepFakes. This approach relies on recognizing feature-level inconsistencies, including spectral artifacts, the absence of natural sensor noise, or other statistical anomalies that distinguish generated content from real images. A well-known example is the use of XceptionNet, which has been effectively applied to DeepFake detection tasks.

3.1.4. Hybrid and specialized techniques

In image forgery detection, hybrid and specialized techniques integrate multiple methods to enhance detection accuracy and reliability. Hybrid approaches often combine spatial, transform-domain, and deep learning techniques to capture both low-level artifacts and high-level features. Specialized techniques focus on specific clues that may indicate tampering, for example, metadata analysis inspects embedded file information for inconsistencies, illumination analysis checks for unnatural lighting or shadows, and semantic analysis evaluates the logical coherence of scene elements. These methods are especially useful for detecting sophisticated or subtle forgeries that may evade conventional detection techniques.

There are different hybrid & specialized techniques. Metadata analysis involves examining inconsistencies in EXIF (Exchangeable Image File Format) data, such as the camera model, GPS coordinates, and timestamp. This method also includes comparing the metadata information with the actual visual content of the image to identify potential signs of manipulation or forgery.

Illumination analysis focuses on detecting the direction of light, the behavior of shadows, and inconsistencies in reflections within an image. It is particularly useful for splicing detection, as mismatched lighting between different regions can indicate that elements have been artificially inserted.

Geometric constraints involve analyzing perspective inconsistencies, such as through vanishing point analysis, to identify signs of image manipulation. This approach also includes evaluating the object size in relation to the estimated camera focal length, where mismatches may indicate tampering.

Semantic inconsistencies are detected by applying scene understanding techniques that analyze objects, humans, and the overall context within an image. For example, an object that appears to be floating in the air or that does not cast a shadow can indicate manipulation or forgery.

4. Specific Characteristics of Detection Techniques

This chapter elaborates the strengths and limitations of various techniques typical for certain methods.

4.1. Pixel-level (passive) methods

Pixel-level methods operate directly on the image content, without relying on any embedded data or prior knowledge about the source. As passive techniques, they analyze statistical and structural inconsistencies at the pixel level to reveal signs of tampering. These methods are widely used in digital forensics due to their noninvasive nature and applicability to images from unknown or untrusted sources.

4.1.1. Noise and sensor pattern analysis

Noise and sensor pattern analysis extracts the unique "fingerprint" of the image sensor, known as the photo-response non-uniformity (PRNU), which is specific to each individual camera. Inconsistencies in this pattern can betray tampering and are often used as forensic evidence.

One of the key strengths of PRNU-based analysis is its ability to provide a unique camera fingerprint, capable of identifying the source camera even among devices of the same model. This technique is passive, meaning that it does not require any prior modification or watermarking of the image; analysis can be performed post-capture. It is also robust to common processing, since PRNU patterns typically survive standard operations such as compression, resizing, and mild filtering. The method is particularly effective for tampering detection, since the absence or distortion of the expected PRNU pattern can reveal image forgeries, including splicing or localized edits. Moreover, PRNU analysis is widely validated and has become a well-established and trusted technique in the field of digital forensics.

However, the method also has certain limitations. It can be weak in small regions, as a sufficient number of pixels is needed to estimate a reliable PRNU pattern; therefore, small splices may escape detection. Additionally, PRNU is vulnerable to post-processing; operations such as rescaling, heavy compression, or denoising may degrade or completely remove the signal. The technique performs best when a reference image from the same camera is available, and in its absence, detection performance may degrade. Finally, there is a risk of false positives, especially when similar PRNU patterns appear on different devices of the same model, which can mislead the detection system.

4.1.2. CFA interpolation artifacts

CFA interpolation artifacts detect anomalies in the regular pattern produced by the Bayer filter during the demosaicing process. These patterns are generally stable and device specific, which makes them useful in forensic analysis to identify tampered or manipulated regions within an image.

One of the primary strengths of this method is its ability to assist in the detection of image tampering. Because CFA artifacts follow a predictable pattern, any inconsistencies introduced by editing or splicing are likely to disrupt that regularity. It also supports camera consistency verification, since demosaicing patterns are often unique to specific camera models, allowing investigators to check whether the image is consistent with the claimed capture device.

Like PRNU, CFA-based detection is a passive and non-invasive approach, requiring no prior image modification or embedded data. It is particularly well-suited for forensic scenarios where the original capture context is unknown. Furthermore, it enables forgery localization, as altered regions typically distort the interpolation pattern, making it possible to identify edited areas with relatively fine granularity. An additional advantage is its effectiveness on JPEG images. Despite the lossy nature of JPEG compression, CFA artifacts often remain intact, allowing the analysis of images shared online or stored in compressed formats.

However, the method has limitations. It is sensitive to pre-processing; actions such as re-mosaicing, interpolation, or heavy filtering may erase or alter the CFA traces, reducing their reliability. The technique is also model-specific, as different cameras and even different RAW-to-JPEG conversion pipelines use varying demosaicing algorithms, which can limit the general applicability of generic detectors. Lastly, there is a potential for a high false alarm rate, since certain genuine image characteristics, such as high-ISO noise or naturally textured regions, may mimic irregularities typically associated with tampering.

4.1.3. JPEG compression artifacts (double compression)

JPEG compression artifact analysis focuses on identifying signs of double compression, which typically occurs when an image has been edited and subsequently saved again. This process leaves behind quantization inconsistencies and block-level traces, particularly within the 8×8 DCT grid, that may indicate tampering.

One of the main strengths of this method is its ability to support tampering detection, as recompression often introduces subtle artifacts that are not present in original images. Moreover, it allows for forgery localization by detecting anomalies at the level of compression blocks, thereby highlighting regions where editing and re-saving have taken place.

An additional advantage is that it does not require an original image. The method operates passively and works directly on the suspect file, which is especially valuable in cases where the unedited original is unavailable. It is also effective in web and social media images, as compression artifacts are frequently preserved even after online distribution. Furthermore, this approach is model-free, meaning it does not depend on device-specific characteristics such as sensor or demosaicing patterns. This makes it suitable for a broad range of images, even when camera metadata is missing or unreliable.

Despite these strengths, JPEG artifact analysis has certain limitations. It assumes a JPEG processing pipeline, so it is not applicable to PNG, RAW, or other non-JPEG formats. If a forger performs global recompression of the entire image after editing, the distinctive double-compression artifacts may disappear completely. Furthermore, the method often relies on alignment with the original 8×8 block grid; even a slight shift in the recompression grid can obscure evidence of tampering, making detection unreliable.

4.2. Format-based methods

Format-based methods exploit the structure of image encoding and file metadata to detect signs of manipulation. Rather than analyzing visual content directly, these techniques focus on compression artifacts, error patterns, and inconsistencies in embedded metadata that may indicate tampering.

4.2.1. Error Level Analysis (ELA)

Error Level Analysis (ELA) is a passive forensic technique that works by recompressing an image at a known JPEG quality level and visualizing the resulting error values block-by-block. Regions that have been altered typically exhibit different error intensities compared to the rest of the image. These differences can then be highlighted and visually inspected to identify potential tampering.

Among its key strengths, ELA supports tampering detection by exposing areas with an inconsistent compression history, which may indicate editing or splicing. The method is also visual and intuitive, producing a heatmap-like output in which suspicious regions appear visually distinct, allowing even users without deep technical expertise to interpret the results. ELA is fast and easy to apply, requires minimal computational resources and is readily available through common forensic software or online platforms.

Another advantage of ELA is that it operates without needing camera-specific information or embedded metadata, which is useful when such data is unavailable or has been removed. It is also widely adopted as a first-pass analysis tool within forensic workflows, helping analysts identify regions of interest for a more in-depth examination using other techniques.

However, ELA has several limitations. It is fundamentally qualitative rather than quantitative, functioning as a visual aid rather than a precise measurement tool. As such, it can be difficult to automate and may be prone to subjective interpretation. The method is also highly sensitive to parameters, especially the choice of recompression quality, which can significantly affect the results and lacks standardization. Finally, ELA is ineffective against global recompression. If a forger saves the entire image uniformly at the same compression level after manipulation, the distinguishing artifacts that ELA depends on may be completely masked.

4.2.2. Format fingerprint & metadata analysis

Format fingerprint and metadata analysis focuses on examining embedded EXIF/IPTC metadata and comparing the structural 'footprint' of an image file with known patterns associated with specific devices or editing software. Discrepancies or inconsistencies in these data can serve as strong indicators of manipulation.

This method enables the identification of devices and software, as metadata often includes details about the camera model, firmware version, or editing application used to create or modify the image. It can also reveal that tampering clues, missing, inconsistent, or altered metadata fields, may suggest that the image has been changed after capture. In addition, timeline verification is possible by analyzing timestamps to detect suspicious gaps or inconsistencies in the image history.

Another advantage is the applicability to batch analysis. Format fingerprints, such as the default EXIF tags associated with certain applications, can help to quickly identify groups of images processed by the same tool or source. Furthermore, this technique is noninvasive and fast, requiring no modification to the image content and allowing efficient examination of large datasets.

However, metadata analysis has several limitations. It is easily stripped or forged; any user can edit or remove metadata fields without altering pixel-level content. This leads to a high false negative rate, especially when forgers are aware of forensic practices and take steps to preserve plausible metadata. Additionally, metadata analysis is not sufficient on its own to confirm tampering. It may raise suspicion, but it cannot localize specific altered regions or provide visual proof of manipulation.

4.3. Geometry- and physics-based methods

Geometry- and physics-based methods in digital forgery detection rely on understanding the physical and spatial properties of the real world. These methods offer several distinct strengths.

One of the key advantages is their independence from content. Geometry- and physics-based techniques are robust to image semantics; unlike deep learning or statistical methods, they do not rely on the scene content or specific training data. Instead, they use universal physical principles that make them applicable to a wide variety of image types.

Another major strength is explainability. These methods produce easily understandable results, such as inconsistencies in shadows, reflections, or perspective. This interpretability makes them especially valuable in forensic and legal contexts where clear visual evidence is essential.

They are also effective in the detection of sophisticated forgeries. Manipulations that involve object insertion, removal, or relocation often fail to preserve physical consistency,

such as lighting direction, perspective alignment, or depth cues. Geometry and physics-based methods can detect such inconsistencies.

In terms of generalization, these methods do not require large training sets. Since they rely on physical models rather than learned representations, they can function effectively even when data are scarce, unlike many deep learning techniques.

Another benefit is their versatility in all modalities. Geometry-based principles, such as structure-from-motion or shadow geometry, can be applied to both still images and video content, making these methods broadly applicable in various forensic contexts.

Finally, these methods are often complementary to other detection approaches. They can be used synergistically with data-driven or statistical forgery detectors. For example, a neural network might detect suspicious regions, while a geometry-based method can confirm inconsistencies in shadows or vanishing points, strengthening the overall reliability of forensic analysis.

4.3.1. Illumination inconsistency

Illumination analysis checks that lighting direction, shadows, and specular highlights across the scene are mutually consistent. This technique can reveal when objects have been inserted or modified without regard for the existing light conditions.

However, it has certain limitations. It often requires an approximate knowledge of the 3D geometry or light source positions, which may not always be available. The method assumes relatively uniform lighting, so it may not perform well in outdoor scenes or in environments with mixed lighting conditions. Furthermore, if a forger applies global adjustments, such as brightness or contrast corrections, apparent inconsistencies may be masked.

4.3.2. Perspective and shadow geometry

Perspective and shadow geometry analysis ensures that object proportions, vanishing points, and shadow projections follow coherent geometric rules. This is particularly useful in identifying content that has been artificially inserted or altered.

Its limitations include the need for manual intervention, as automatic vanishing-point estimation can be unreliable in cluttered or low-texture regions. This method may struggle with small or subtle splices, where pasted objects produce negligible geometric distortion. Additionally, it typically assumes planar surfaces; irregular or curved shapes can violate the assumptions required for consistent geometric modeling.

4.4. Copy-move and splicing detection

4.4.1. Copy-move (self-similarity) detection

Copy-move detection identifies duplicated patches within the same image using block matching or keypoint clustering. It is particularly effective in detecting cloning operations, in which parts of the image are copied and placed elsewhere to conceal or replicate objects. Since this method relies solely on internal similarities within the image, it does not require any external reference.

This technique enables forgery location to be precise by highlighting regions involved in the copy-move operation. Many algorithms are designed to tolerate slight changes, such as rotation, scaling, or compression between copied regions, making them robust to

minor modifications. Moreover, it is applicable across various image types and can function even when metadata or other forensic cues are missing.

Despite its advantages, this method struggles with non-exact copies. If the forger applies geometric transformations like rotation, scaling, or non-rigid warping, simple block matching may fail to align the altered patches. Furthermore, the exhaustive search required to compare all possible regions can lead to high computational costs, necessitating a trade-off between speed and accuracy. Highly textured regions, such as foliage or brick walls, can also produce a large number of false positives due to natural repetition.

4.4.2. Splicing detection through local features

Splicing detection using local features, such as SIFT, SURF, or other descriptors, focuses on identifying clusters of anomalous keypoints that deviate from the global image statistics. This method is sensitive to local inconsistencies in texture, lighting, noise, or edges, which often indicate that the content has been inserted from another source.

In addition to detecting inconsistencies, the method allows for precise forgery localization by identifying the exact regions where external content was introduced. It is robust to global transformations, such as scaling, rotation, or color adjustments, which enhances its reliability under various post-processing conditions. Since it is a passive technique, splicing detection via local features does not require the original or reference image, nor any embedded metadata. It is also versatile and performs well across different types of visual content, including natural scenes, portraits, and complex backgrounds, especially in cases where traditional forensic signals (such as PRNU) are weak or missing.

However, this approach may be less effective in low-textured areas where few keypoints are generated, making it difficult to locate tampered regions. Post-processing effects, such as blurring or added noise, can further reduce the repeatability of keypoints, allowing splices to go undetected. Additionally, if the inserted content comes from a source image taken with a similar camera model, the feature distributions may overlap, making it hard to distinguish between new and original content.

4.5. Deep-learning-based methods

4.5.1. End-to-end CNN detectors

End-to-end CNN detectors train convolutional networks to spot statistical anomalies, such as tampered edges or inconsistent noise within a supervised setting. One of the key strengths of these methods is automatic feature learning. CNNs can learn complex forensic features directly from the data, without needing hand-crafted rules or manual feature extraction.

These detectors offer high detection accuracy. When trained on large and diverse datasets, CNNs often outperform traditional techniques in identifying a wide range of manipulations, including splicing, copy-move, and deepfakes. Their versatility and scalability allow the same architecture to be adapted for multiple forensic tasks such as image forgery detection, manipulation localization, or even source attribution.

CNN-based systems are also robust to multiple types of manipulation. They are capable of handling subtle or complex changes in geometry, noise patterns, or lighting

inconsistencies. Many of these systems are designed to produce tampering maps, which makes it possible to localize manipulated regions at the pixel or patch level.

However, these methods have limitations. Their performance is highly dependent on the quality and diversity of the training data. If novel or unseen tampering techniques are introduced, the network may not detect them. CNNs are also at risk of overfitting; they may inadvertently learn dataset-specific artifacts, such as particular compression settings, instead of generalizable forgery cues. Another notable challenge is explainability. Due to their black-box nature, CNNs can be difficult to interpret; analysts may struggle to understand why a specific region was flagged, which can be problematic in forensic and legal contexts where clear evidence is required.

4.5.2. GAN-based forgery detectors

GAN-based forgery detectors are specialized networks designed to identify traces of GAN synthesis or deepfake generation. Their primary strength lies in being specifically trained to detect artifacts and patterns that are unique to GAN-generated content, including those produced by models such as StyleGAN or Deepfakes.

These detectors are highly sensitive to subtle artifacts. They can detect minute inconsistencies in texture, facial features, or statistical image properties, artifacts that are often missed by traditional methods. Moreover, with proper retraining, GAN detectors can adapt to evolving GAN architectures, maintaining their effectiveness over time. Some models offer pixel-level localization by generating heatmaps or binary masks that highlight manipulated regions within an image or video.

GAN detectors analyze both visual and statistical cues. They can detect perceptual anomalies such as unnatural eye rendering or inconsistent backgrounds, as well as frequency domain irregularities or missing sensor noise patterns.

Despite their strengths, GAN-based detectors are part of an ongoing arms race. As generative models become more sophisticated and produce fewer artifacts, existing detectors must be continuously re-trained to keep up with them. Furthermore, dataset imbalance presents a challenge, obtaining representative and balanced sets of ‘real’ versus “fake” images can be difficult, and biases in training data may affect performance. GAN detectors also have a limited scope. They are highly effective against synthetic content but may not detect traditional splicing or copy-move operations performed using standard image editing tools.

4.6. Hybrid and fusion approaches

Hybrid and fusion approaches combine multiple cues, pixel artifacts, metadata, geometry, CNN outputs, to improve the robustness and accuracy of image forgery detection systems.

One of the main strengths of these approaches is improved accuracy. By integrating several forensic techniques, such as PRNU, CFA, metadata analysis, and deep learning, overall performance is enhanced, and the rate of false positives is reduced. These systems are also robust to various manipulations, allowing them to detect a wide variety of forgeries, including splicing, copy-move, and GAN-generated images, by leveraging the complementary strengths of individual methods.

Another significant advantage is the availability of complementary evidence. Fusion approaches can integrate multiple types of visual, statistical, and structural information, providing more convincing and reliable forensic conclusions. Their adaptability to complex

cases makes them especially useful in real-world scenarios involving sophisticated tampering techniques, where a single method may fail. Moreover, hybrid systems offer flexibility and scalability. Their modular design allows for easy extension and customization, allowing the incorporation of new detectors or algorithms as manipulation techniques evolve.

However, hybrid systems also present several limitations. Their complexity makes them more difficult to tune and validate. Setting appropriate weights and thresholds for combining modules can be non-trivial. They are also resource-intensive, as running multiple detection pipelines in parallel increases computation time and memory consumption. Additionally, error compounding can occur. Mistakes or uncertainties in one module can influence the results of others, potentially amplifying false positives or false negatives in the overall output.

5. Evaluation of The Detection Model

Evaluating a detection model involves a systematic procedure to measure how effectively the model can identify manipulated (forged) or tampered images. Evaluation is crucial to understand the performance, robustness, and generalizability of the model. The evaluation procedure consists of the following basic steps:

- **Pre-processing and prediction -** These steps begin by preprocessing the test images, which may include steps such as resizing and normalization. These processed images are then passed through the trained model. The model outputs an image-level label (e.g., real or fake) and, for localization models, a forgery mask indicating the manipulated regions.
- **Post-processing** - This phase involves applying thresholds to the prediction scores, such as using 0.5 for binary classification. If necessary, the predicted masks are further refined using techniques such as morphological operations.
- **Metric computation** - Metric computation involves comparing the model's output with ground truth labels or masks. Performance metrics are calculated for each image and then aggregated using statistical measures such as mean or median.

Several approaches are used in the evaluation process.

Robustness and generalization tests - Robustness and generalization tests assess the model's performance under various challenging conditions. In cross-dataset evaluation, the model is tested on a dataset different from the one on which it was trained to evaluate its ability to generalize. Under adversarial conditions, factors such as compression, blur, resizing, and noise are introduced to observe how the model's performance degrades. To test forgery-type generalization, the model is evaluated on manipulation techniques it has not seen during training; for example, a model trained on copy-move forgeries might be tested on splicing.

Qualitative evaluation - Qualitative evaluation involves visually inspecting the predicted masks and overlaying them on the original images to assess their accuracy. This process helps identify false positives and false negatives. Tools such as Grad-CAM or attention maps can also be used to interpret which parts of the image the model focuses on during prediction.

Benchmarking and comparison - A model under examination can be compared against baseline models (traditional or deep learning), state-of-the-art methods, and human performance (in some studies).

Statistical significance - Statistical significance is assessed using techniques such as t-tests, Wilcoxon tests, or bootstrapping to ensure that observed performance improvements are

not due to chance. Additionally, confidence intervals are calculated for performance metrics to provide a measure of reliability.

Example tools for evaluation include Scikit-learn, which is used for classification metrics, OpenCV and NumPy, which are useful for pixel-level comparisons, MATLAB, which supports traditional forensics evaluations, and TensorBoard or Weights & Biases, which are commonly employed for visualizations.

5.5.1. Example usage in experiments

When evaluating an image forgery detection model, the approach depends on the type of model and the aspect being assessed. For classification models, it is important to use image-level metrics such as accuracy, F1-score, and AUC. In the case of segmentation models, metrics such as IoU, Dice coefficient, and pixel-wise precision and recall should be used. To assess robustness, tests should be conducted under various transformations, including resizing, compression, and rotation, followed by a re-evaluation of the relevant metrics.

Choosing the right evaluation metric depends on the task (detection vs. localization) and the dataset characteristics (balanced vs. imbalanced, single vs. multi-type forgeries). Robust evaluation using multiple metrics ensures that an algorithm is not only accurate, but also reliable and generalizable across various conditions.

Table 1 Choosing the appropriate metric depending on the task

Task	Preferred metrics
Image classification	accuracy, precision, recall, F1, AUC
Region localization	IoU, Dice, pixel precision/recall
Boundary analysis	boundary IoU, Hausdorff distance

Summary of common challenges across methods

1. *Post-processing resilience* - Re-compression, filtering, resizing, and color adjustments are the forger's first defense, and they blunt most detectors.
2. *Generalizability* - Many methods are tailored to specific cameras, compression pipelines, or editing tools. Adapting to new scenarios requires retraining or retuning.
3. *Explainability vs. automation* - Highly automated 'black-box' detectors (e.g. deep learning) trade interpretability for throughput. However, forensic experts need clear and visualizable evidence.
4. *Small-scale forgeries* - Tiny splices, subtle object insertions, or minor retouching often fall below the detection threshold of both pixel- and feature-based methods.

In practice, a multicue, analyst-in-the-loop workflow where automated flags are corroborated by human inspection of image regions, metadata, and scene context, remains the most reliable strategy for real-world forensic applications.

5.5.2 Evaluation metrics

In image forgery detection, evaluating the performance of detection algorithms requires the use of appropriate evaluation metrics. These metrics assess how well a method can identify tampered images (image-level detection) or localize manipulated regions (pixel-level localization).

1. Image-level evaluation metrics - These metrics are used when the goal is to classify an entire image as either authentic or forged.
Accuracy **is** the ratio of correctly classified images (both authentic and forged) to the total number of images.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where:

- TP: True Positives (forged images correctly identified)
- TN: True Negatives (authentic images correctly identified)
- FP: False Positives (authentic images wrongly marked as forged)
- FN: False Negatives (forged images wrongly marked as authentic)

Limitation: Can be misleading in imbalanced datasets.

Precision, the proportion of predicted forged images that are truly forged.

$$precision = \frac{TP}{TP + FP} \tag{2}$$

Significance: High precision means fewer false alarms.

Recall (sensitivity or true positive rate) **-** the proportion of actual forged images that were correctly identified.

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Significance: High recall indicates that the method detects most of the forgeries.

F1-score **-** the harmonic mean of precision and recall.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4}$$

Significance: Balances the trade-off between precision and recall.

ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve)

ROC curve - A plot of the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.
AUC - Represents the area under the ROC curve. A higher AUC indicates better model performance.
Significance - AUC close to 1.0 indicates excellent classification ability; 0.5 indicates random guessing.

2. Pixel-level evaluation metrics - Used when the goal is localizing tampered regions within the image, often evaluated using binary masks (ground truth vs. predicted forgery regions).

a. Pixel accuracy, the ratio of correctly classified pixels to the total number of pixels.

$$pixel\ accuracy = \frac{correct\ pixels}{total\ pixels} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Limitation: Like image-level accuracy, can be misleading in cases where the majority of pixels are unaltered.

b. IoU (Intersection over Union) / Jaccard index measures the overlap between the predicted and ground truth tampered regions.

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \tag{6}$$

Range: 0 (no overlap) to 1 (perfect overlap)
Significance: Standard metric for evaluating localization performance.

c. Dice coefficient (a.k.a. Sørensen-Dice index or F1 score for pixels) - a similarity measure between the predicted and actual tampered regions.

$$Dice = \frac{2 \cdot |A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \tag{7}$$

Range: 0 to 1.
Significance: More sensitive to small overlaps than IoU.

d. Boundary-based metrics
Used in fine-grained localization (e.g., detecting edges of tampered regions).

- Boundary IoU - Evaluates the overlap near the boundaries of forged regions.
- Hausdorff distance - Measures the maximum distance between the predicted and ground truth region boundaries.

3. Confusion matrix - A confusion matrix provides a detailed breakdown of classification results. It contains four entries (TP, FP, TN, FN), helps visualize errors and biases in classification, and is useful for both image-level and pixel-level evaluation.

Table 2 Confusion matrix in forgery detection

	Predicted authentic	Predicted forged
Actual authentic	TN	FP
Actual forged	FN	TP

6. Deep Learning Approach in Forgery Detection

The principle of detecting digital image forgery using deep learning involves training a neural network to distinguish between authentic and tampered images based on subtle patterns. This principle consists of the following phases:

1. Input image - A potentially forged digital image is input into the system.

2. Feature extraction - CNN analyzes the image to extract important features such as textures, edges, and inconsistencies. These features can capture noise patterns, compression artifacts, lighting mismatches, and pixel-level anomalies.

3. Classification - The extracted features are fed into a classifier layer (often part of the same deep learning model), which determines whether the image or parts of it are forged or authentic.

4. Output - The system produces a label (e.g. 'forged' or "authentic") and often a forgery map, highlighting damaged regions within the image.

The deep learning process leverages the ability to detect subtle cues that are invisible to the human eye or traditional forensic techniques.

Forgery detection using this approach consists of a data preparation phase and a model training phase:

1. Dataset preparation - Dataset preparation generally consists of the following two phases:

a) Datasets selection – The different datasets can be used to verify forgery detection, such as: CASIA (v1.0, v2.0 – contains tampered and authentic images), Columbia Uncompressed Dataset, Coverage Dataset, NIST Nimble Challenge Dataset, Deepfake/Celeb-DF/FaceForensics++ (for facial manipulations).

These datasets typically contain authentic images (unaltered), forged images (with splicing, copy-move, inpainting, deepfakes, etc.), ground truth masks (in pixel-level detection datasets, indicating manipulated regions).

b) Data splitting – Data splitting involves dividing the dataset into training, validation, and test sets, commonly using ratios such as 70/15/15 or 80/10/10. To prevent data leakage, it is important to ensure that forgeries and their corresponding original images are kept within the same split.

2. Model training - Model training involves using training data to fit the model. To enhance robustness, data augmentation techniques such as flipping, scaling, and adding noise are applied. Depending on the specific task, whether it requires localization or classification, the model is trained using either pixel-level or image-level labels.

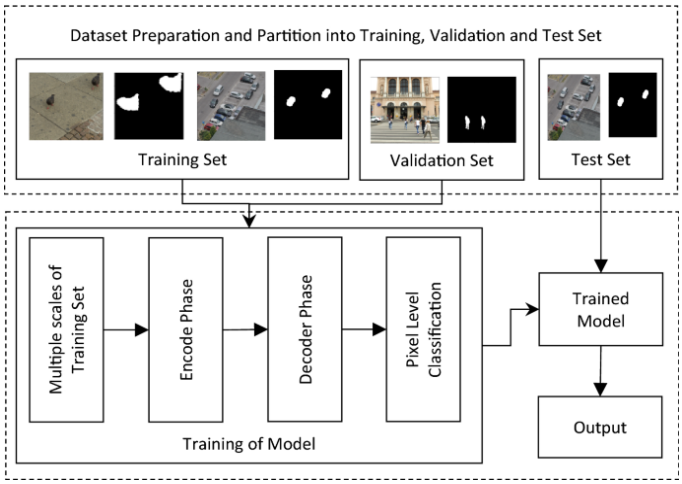

Fig. 3 Detection of copy-move forgery – deep learning model [47]

7. Challenges in Forgery Detection

One of the primary challenges is the emergence of high-quality forgeries, particularly those generated by GANs, which can often evade traditional pixel-level or statistical detection methods. Another difficulty lies in the balance between realism and subtlety: retouching and minor edits may be imperceptible without contextual clues, making them hard to identify.

Generalization remains a significant issue, as models trained on specific datasets often struggle to perform effectively in unseen scenarios. Furthermore, post-processing techniques such as compression, resizing, and blurring can degrade forensic signals, further complicating detection.

The rise of multi-modal fakes, which combine image, audio, and video elements as seen in deepfake pipelines, introduces added complexity to the detection task. Lastly, adversarial examples pose a growing threat, as some manipulations are deliberately crafted to bypass forensic algorithms and avoid detection.

8. Trends & Future Directions

One major trend in image forensics is the development of universal forensic models. The goal is to create a single model capable of detecting multiple types of forgeries, such as splicing, GAN-generated content, and image retouching, through unified representation learning.

Another emerging area is zero-shot and few-shot detection, which leverages techniques like contrastive learning or domain adaptation to identify manipulations the model has not seen during training.

Multimodal forensics is also gaining prominence. This approach combines consistency checks across different media types, such as image, video, audio, and text. Deepfake detection, in particular, increasingly relies on multimodal evidence to identify manipulation.

Explainable forensics is becoming essential, with tools being developed to justify their predictions, for example, indicating why a certain region of an image was flagged as fake. This improves trust and transparency, especially in high-stakes fields like journalism and law enforcement.

Finally, blockchain and provenance tracking are being explored as solutions to ensure the authenticity of content. These involve embedding verifiable provenance chains into digital media. Notable examples include initiatives like the Coalition for Content Provenance and Authenticity (C2PA) and Project Origin.

9. Conclusions

Digital image forgery detection is a rapidly evolving field driven by the increasing sophistication and accessibility of manipulation tools. This paper provides a comprehensive overview of the current landscape that includes traditional, transform-domain, deep learning, and hybrid approaches to detect image forgeries. Through categorization and analysis of detection principles, it is demonstrated how different techniques target various manipulation types, including copy-move, splicing, retouching, and GAN-generated content, each with distinct strengths and limitations.

Traditional signal-based methods remain valuable for their interpretability and efficiency, especially in scenarios with known forensic signatures, such as compression artifacts or sensor noise patterns. However, they often struggle with subtle or heavily post-processed forgeries. Transform-domain methods add robustness against some manipulations but are sensitive to environmental and image-specific factors.

Deep learning has made substantial progress, particularly in detecting complex forgeries and automatically learning high-level patterns. However, its effectiveness depends heavily on the availability and quality of training data, and its black-box nature raises challenges in forensic explainability. Hybrid approaches, which integrate cues from multiple forensic domains, appear especially promising for improving robustness and generalization.

Evaluation of forgery detection methods remains critical and multifaceted. Performance must be assessed not only through standard classification and localization metrics but also through rigorous testing under real-world conditions and adversarial scenarios. This ensures that detection systems maintain reliability beyond the controlled datasets.

Despite the advances, several challenges persist: the arms race with increasingly realistic GAN forgeries, the need for models to generalize across domains, and the importance of explainability in forensic settings. Future work should prioritize universal and explainable forensic models, improved dataset diversity, and real-time or multimodal detection systems. Integrating blockchain-based provenance tracking and advancing forensic transparency will also be key to maintaining trust in digital visual content.

In summary, image forgery detection is moving toward more resilient, scalable, and explainable systems, an imperative step in safeguarding the integrity of digital imagery in an era dominated by AI-generated and edited media.

REFERENCES

[1] P. Čisar, S. Maravić Čisar and S. Bošnjak, *Cybercrime and digital forensics – technologies and approaches*, DAAAM International Scientific Book, Vienna, Austria, 2014, Chapter 42, pp. 525-542.

[2] P. Čisar and S. Maravić Čisar, "Methodological Frameworks of Digital Forensics", In Proceedings of the 9th IEEE International Symposium on Intelligent Systems and Informatics SISY 2011, 2011, pp. 343-347.

[3] P. Čisar and S. Maravić Čisar, "General Directions of Development in Digital Forensics", *Acta Technica Corviniensis*, vol. 5, no. 2, pp. 87-91, 2012.

[4] P. Čisar and J. Fodor, "Digital Image Forgery Identification", In Proceedings of International Scientific Conference "Archibald Reiss Days" 2015, vol. I, 2015, pp. 93-99.

[5] J. Fridrich, D. Soukal and J. Lukáš, "Detection of Copy-Move Forgery in Digital Images," *International J.*, vol. 3, pp. 652-663, 2003.

[6] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions", Technical Report TR2004-515, Dartmouth College, 2004.

[7] S. J. Ryu, M. J. Lee and H. K. Lee, "Detection of Copy-rotate-move Forgery Using Zernike Noments", in Proceedings of Information Hiding (IH 2010), R. Böhme, P. W. L. Fong and R. Safavi-Naini, Eds., Berlin, Germany: Springer, 2010, Lect. Notes Comput. Sci., vol. 6387, pp. 51-65.

[8] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints", *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, Nov. 2004.

[9] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded up Robust Features (SURF)", *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346-359, 2008.

[10] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," In Proceedings of International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2564-2571.

[11] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Commun. ACM*, vol. 24, no. 6, pp. 381-395, June 1981.

[12] H. Farid, "Detecting Digital Forgeries Using Demosaicing Artifacts", *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 611-615, Sep. 2006.
[13] T. Bianchi and A. Piva, "Image Forgery Localization via Block-grained Analysis of JPEG Artifacts", *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 3, pp. 1003-1017, Sep. 2012.
[14] J. Lukás, J. Fridrich and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise", *IEEE Trans. Inf. Forensics Secur.*, vol. 1, no. 2, pp. 205-214, June 2006.
[15] H. Farid, "Exposing Digital Forgeries from JPEG Ghosts", *IEEE Trans. Inf. Forensics Secur.*, vol. 4, no. 1, pp. 154-160, Mar. 2009.
[16] B. Mahdian and S. Saic, "Blind Authentication Using Periodic Properties of Interpolation", *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 3, pp. 529-538, Sept. 2008.
[17] P. Yang, R. Ni and Y. Zhao, "Double JPEG Compression Detection by Exploring the Correlations in DCT Domain", arXiv preprint arXiv:1806.01571, June 2018.
[18] M. K. Johnson and H. Farid, "Exposing Digital Forgeries Through Chromatic Aberration", In Proceedings of the 8th Workshop on Multimedia and Security (MM&Sec '06), New York, NY, USA: Association for Computing Machinery, 2006, pp. 48-55.
[19] M. K. Johnson and H. Farid, "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting", In Proceedings of the 7th Workshop on Multimedia and Security (MM&Sec '05), New York, NY, USA: Association for Computing Machinery, 2005, pp. 1-10.
[20] C. Riess and E. Angelopoulou, "Physics-based Illuminant Color Estimation as an Image Semantics Clue", In Proceedings of the International Conference on Image Processing (ICIP), Cairo, Egypt, 2009, pp. 689-692.
[21] E. Kee, J. F. O'Brien and H. Farid, "Exposing Photo Manipulation with Inconsistent Shadows", *ACM Trans. Graph.*, vol. 32, no. 3, p. 28, June 2013.
[22] M. Iuliani, G. Fabbri and A. Piva, "Image Splicing Detection Based on General Perspective Constraints", In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Rome, Italy, 2015, pp. 1-6.
[23] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network", In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 2018, pp. 1-7.
[24] Y. Wu, W. AbdAlmageed and P. Natarajan, "ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries with Anomalous Features", In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 9535-9544.
[25] R. Salloum, Y. Ren and C.-C. J. Kuo, "Image Splicing Localization Using a Multi-task Fully Convolutional Network (MFCN)", *J. Vis. Commun. Image Represent.*, vol. 51, pp. 201-209, 2018.
[26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei and Z. Zhang, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", In Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002.
[27] X. Hu, Z. Zhang, Z. Jiang, S. Chaudhuri, Z. Yang and R. Nevatia, "SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization", in Computer Vision – ECCV 2020, *Lecture Notes in Computer Science*, vol. 12366, Cham, Switzerland: Springer International Publishing, 2020, pp. 312-328.
[28] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", in *Medical Image Computing and Computer-Assisted Intervention* – MICCAI 2015, N. Navab, J. Hornegger, W. Wells and A. Frangi, Eds., Lecture Notes in Computer Science, vol. 9351, Cham, Switzerland: Springer, 2015, pp. 234-241.
[29] D. Cozzolino and L. Verdoliva, "Noiseprint: A CNN-based Camera Model Fingerprint", *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 144-159, 2020.
[30] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images", In Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1-11.
[31] M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi and S. Tubaro, "Aligned and Non-Aligned Double JPEG Detection Using Convolutional Neural Networks", *J. Vis. Commun. Image Represent.*, vol. 49, pp. 153-163, 2017.
[32] CASIA Image Tampering Detection Evaluation Database. [Online]. Available: http://forensics.idealtest.org/
[33] Columbia Image Splicing Detection Evaluation Dataset. [Online]. Available: https://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/
[34] D. Tralić, I. Župančić, S. Grgić and M. Grgić, "CoMoFoD – New Database for Copy-Move Forgery Detection", In Proceedings of the 55th International Symposium ELMAR2013, Zadar, Croatia, Sep. 2013, pp. 49-54.

[35] J. Guo, J. Zhang, T. Chen and H. Liu, "Hierarchical Fine-Grained Image Forgery Detection and Localization", In Proceedings of IEEE/CVPR 2023, pp. 9457–9466.
[36] Y. Liu, X. Zhu, X. Zhao and Y. Cao, "Adversarial Learning for Constrained Image Splicing Detection and Localization Based on Atrous Convolution," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 10, pp. 2551-2566, Oct. 2019.
[37] A. A. Solanke, "Explainable Digital Forensics AI: Towards Mitigating Distrust in AI-based Digital Forensics Analysis Using Interpretable Models", *Forensic Sci. Int.: Digit. Investig.*, vol. 42, p. 301403, 2022.
[38] H. Cheng, L. Niu, Z. Zhang and L. Ye, "Generalization Enhancement Strategy Based on Ensemble Learning for Open Domain Image Manipulation Detection", *J. Vis. Commun. Image Represent.*, vol. 107, p. 104396, 2025.
[39] W. Zheng, X. Ke and W. Guo, "Zero-shot 3D Anomaly Detection via Online Voter Mechanism", *Neural Netw.*, vol. 187, p. 107398, 2025.
[40] S. Usmani, S. Kumar and D. Sadhya, "Spatio-temporal Knowledge Distilled Video Vision Transformer (STKD-VViT) for Multimodal Deepfake Detection", *Neurocomputing*, vol. 620, p. 129256, 2025.
[41] N. Xiao, Z. Wang, X. Sun and J. Miao, "A Novel Blockchain-based Digital Forensics Framework for Preserving Evidence and Enabling Investigation in Industrial Internet of Things", *Alexandria Eng. J.*, vol. 86, pp. 631-643, 2024.
[42] I. Amerini, M. Barni, S. Battiato, P. Bestagini, G. Boato, T. S. Bonaventura et al., "Deepfake Media Forensics: State of the Art and Challenges Ahead", arXiv preprint arXiv:2408.00388, Aug. 2024.
[43] I. Vuković, P. Čisar, K. Kuk, M. Banđur and B. Popović, "Influence of Image Enhancement Techniques on Effectiveness of Unconstrained Face Detection and Identification", *Elektronika ir Elektrotechnika*, vol. 27, no. 5, pp. 49-58, 2021.
[44] U. Samariya, S. D. Kamble, S. Singh et al., "A Survey on Copy-Move Image Forgery Detection Based on Deep-learning Techniques", *Multimed. Tools Appl.*, vol. 84, pp. 30603-30662, 2024.
[45] M. D. Ansari, S. P. Ghrera and V. Tyagi, "Pixel-based image forgery detection: A review", *IETE J. Education*, vol. 55, no. 1, pp. 40-46, 2024.
[46] S. Singh and R. Kumar, "Image Forgery Detection: Comprehensive Review of Digital Forensics Approaches", *J. Computat. Soc. Sci.*, vol. 7, pp. 1-39, 2024.
[47] A. K. Jaiswal and R. Srivastava, "Detection of Copy-Move Forgery in Digital Image Using Multi-scale, Multi-stage Deep Learning Model", *Neural Process. Lett.*, vol. 54, pp. 75-100, 2022.