

A Comprehensive Survey on Image Forgery Detection Techniques Using Handcrafted and Deep Features

Shalini Kumari

Computer Science & Engineering

Shri Rawatpura Sarkar University

Abstract

Image forgery has become increasingly sophisticated due to the widespread availability of advanced image editing tools and the rapid growth of generative models, posing serious threats to digital image authenticity and trust. Consequently, image forgery detection has emerged as a critical research area within digital forensics. This paper presents a comprehensive survey of image forgery detection techniques based on handcrafted features, deep learning features, and hybrid approaches that combine both paradigms. Traditional handcrafted methods exploit low-level statistical, texture, frequency-domain, and camera-related artifacts to identify manipulations such as copy-move, splicing, and retouching. While these techniques offer interpretability and robustness to limited data, they often struggle with complex and large-scale forgeries. In contrast, deep learning-based approaches automatically learn high-level semantic representations and demonstrate superior performance on diverse forgery scenarios, albeit at the cost of high data dependency and reduced explainability. To address these limitations, hybrid methods that fuse handcrafted and deep features have recently gained attention, offering improved generalization and robustness. This survey systematically reviews state-of-the-art methods across these three categories, discusses commonly used benchmark datasets and evaluation metrics, and analyzes comparative performance trends. Furthermore, key challenges, including generalization across datasets, robustness to post-processing, and explainability, are highlighted. Finally, potential future research directions are outlined to guide the development of more reliable and scalable image forgery detection systems.

1 Introduction

The rapid growth of digital imaging devices, social media platforms, and online content-sharing services has led to an unprecedented increase in the creation and dissemination of digital images. While this has enabled efficient communication and information exchange, it has also facilitated the widespread manipulation and falsification of visual content. Image forgery, which involves altering or fabricating visual information to mislead viewers, poses serious challenges in critical domains such as digital forensics, journalism, medical imaging, legal investigations, and national security [30, 10].

Image forgeries can broadly be categorized into common manipulation types such as copy-move forgery, splicing, retouching, and object removal or insertion. These manipulations are often visually imperceptible, making manual detection unreliable and necessitating the development of automated image forgery detection (IFD) techniques [7, 5]. Over the past two decades, extensive research efforts have been devoted to designing robust computational methods capable of distinguishing authentic images from forged ones.

Early image forgery detection approaches primarily relied on handcrafted features derived from signal processing and statistical analysis. These methods exploit intrinsic inconsistencies introduced during manipulation, such as abnormal texture patterns, lighting discrepancies, compression artifacts, and noise irregularities. Popular handcrafted techniques include block-based and keypoint-based methods for copy-move detection, frequency-domain representations using Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), local texture descriptors such as Local Binary Patterns (LBP), and moment-based features [2, 1]. Although handcrafted methods are computationally efficient and interpretable, their performance often degrades under post-processing operations such as resizing, rotation, blurring, and compression.

With the advent of deep learning, image forgery detection has witnessed a paradigm shift toward data-driven feature learning. Convolutional Neural Networks (CNNs) have demonstrated remarkable capability in automatically extracting hierarchical and discriminative features from raw pixel data, leading to substantial performance improvements over traditional handcrafted approaches [8, 9]. Deep architectures such as AlexNet, VGG, ResNet, DenseNet, and EfficientNet have been successfully adapted

for various forgery detection tasks, including splicing detection, copy-move localization, and generalized manipulation detection [14, 11].

Despite their superior accuracy, deep learning-based methods face several limitations, including high computational cost, large data requirements, limited generalization across datasets, and vulnerability to adversarial manipulations. To address these challenges, hybrid approaches that combine handcrafted features with deep features have gained increasing attention. Feature fusion strategies aim to leverage the complementary strengths of low-level statistical descriptors and high-level semantic representations, resulting in more robust and generalized detection frameworks [17, 24]. Dimensionality reduction and feature selection techniques such as Principal Component Analysis (PCA), Probabilistic PCA (PPCA), and manifold learning are often employed to mitigate redundancy and improve classifier performance.

In addition to feature extraction strategies, classifier selection plays a crucial role in the effectiveness of image forgery detection systems. Traditional machine learning classifiers such as Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbors (kNN), and ensemble models have been widely used with handcrafted and fused features [12]. More recently, end-to-end deep classifiers and stacking or ensemble-based decision models have been explored to further enhance detection accuracy and robustness [28].

Given the rapid evolution of forgery techniques and detection methodologies, there is a growing need for comprehensive survey studies that systematically analyze existing approaches, datasets, feature representations, classifiers, and evaluation metrics. While several surveys have addressed specific aspects of image forgery detection, such as copy-move detection or deep learning-based methods, a unified review covering handcrafted, deep, and hybrid feature-based techniques remains essential for guiding future research [25, 21].

This survey aims to provide a comprehensive and structured review of image forgery detection techniques based on handcrafted features, deep learning approaches, and hybrid fusion strategies. The paper systematically examines commonly used datasets, feature extraction methods, classification frameworks, and performance evaluation protocols. Furthermore, it highlights current challenges, emerging trends, and potential research directions to advance the development of robust and generalizable image forgery detection systems.

2 Literature Review

Image forgery detection has become an important research area due to the proliferation of digital media and the potential misuse of manipulated images in critical domains such as journalism, security, and legal investigations [30, 16]. The literature can be broadly categorized into handcrafted feature-based approaches, deep learning-based methods, hybrid techniques, classifier studies, and evaluation protocols.

2.1 Handcrafted Feature-Based Approaches

Early approaches rely on handcrafted features that exploit inconsistencies in the image arising from manipulation. Copy-move forgery detection is one of the most investigated areas. Fridrich et al. [1] proposed a block-based approach using DCT coefficients, whereas Popescu and Farid [2] applied PCA to detect duplicated regions. Amerini et al. [5] introduced a SIFT-based keypoint matching method for robust copy-move detection. Christlein et al. [7] provided a comprehensive evaluation of popular copy-move detection algorithms.

Other approaches focus on splicing, resampling, and photomontage detection. Lukas et al. [3] proposed the use of sensor pattern noise to detect manipulations, and Ng et al. [4] applied higher-order statistics for blind photomontage detection. Local descriptors, frequency-domain transforms, and moment-based features are widely adopted in handcrafted methods [12, 13].

2.2 Deep Learning-Based Approaches

The rise of deep learning enabled automatic extraction of hierarchical features, significantly improving detection performance. CNN-based methods have been widely adopted. Bayar and Stamm [8] introduced a constrained convolutional layer for manipulation detection. Rao and Ni [9] applied deep CNNs for splicing and copy-move detection. Cozzolino et al. [14] proposed Noiseprint CNN to exploit camera model fingerprints. Zhou et al. [11] developed rich feature learning for generalized manipulation detection.

Advanced architectures such as Vision Transformers (ViT) have recently been applied for global feature extraction [18, 23], and contrastive/self-supervised learning has improved robustness under domain shifts [15, 19]. Surveys by Zanardelli et al. [25] highlight the success of deep learning approaches in achieving high accuracy across multiple datasets.

2.3 Hybrid Approaches and Feature Fusion

Hybrid approaches combine handcrafted and deep features to leverage complementary strengths. Jain and Kumar [17] integrated signal processing descriptors with deep CNN features, improving generalization across datasets. Li et al. [24] proposed multi-level feature fusion strategies combining DWT, SPN, and CNN embeddings. Dimensionality reduction methods like PCA and manifold learning are frequently used to remove redundancy and improve classifier efficiency [27, 26].

2.4 Classifier Studies and Ensemble Methods

The choice of classifier significantly affects detection performance. Traditional classifiers such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (kNN) have been widely used with handcrafted or hybrid features [12]. Ensemble and stacking-based classifiers further improve robustness and accuracy. Bogacsovics et al. [28] developed ensemble architectures to combine multiple model predictions. Guillaro et al. [20] introduced TruFor, leveraging all-round clues in an ensemble framework to achieve high detection reliability.

2.5 Datasets and Evaluation Protocols

Robust evaluation requires standardized datasets. CASIA [6], Columbia [4], and NIST datasets [12] provide diverse manipulation types including copy-move, splicing, and retouching. Recent studies emphasize cross-dataset evaluation and benchmarking protocols to assess generalization capability [26, 21]. Surveys [25] underscore the importance of consistent evaluation metrics such as precision, recall, and F1-score for comparing methods.

2.6 Summary and Research Trends

The literature demonstrates a clear evolution from handcrafted feature-based approaches to deep learning and hybrid fusion strategies. Transformer-based models and contrastive learning are emerging as state-of-the-art methods. Ensemble approaches remain critical for robust performance. Despite advancements, challenges such as detection of small manipulations, resilience to adversarial attacks, and cross-domain generalization remain open research directions [30, 16, 29].

longtable

Table 1: Summary of Image Forgery Detection Approaches from Literature

Approach / Dataset	Methodology	Features	Classifier / Performance
CASIA / Copy-Move	Block-based forgery detection using DCT similarity	DCT coefficients	Thresholding / Good for small duplicated regions
CASIA / Copy-Move	Keypoint-based detection using PCA or SIFT descriptors	PCA / SIFT	Matching / Robust to rotation and scaling
Multiple Cameras / Splicing	Sensor Pattern Noise analysis to detect inconsistencies	SPN residuals	Correlation / Effective for splicing detection
Columbia / Photomontage	Higher-order statistical analysis of image blocks	Higher-order statistics	SVM / Detects lighting inconsistencies
CASIA	Benchmark dataset for tampering detection	Standard dataset	Provides standard evaluation protocol

Continued on next page

Table 1 continued from previous page

Dataset	Methodology	Features	Performance
CASIA / Copy-Move	Evaluation of multiple detection algorithms	Various	Comparison study / Evaluates robustness
CASIA / General Manipulation	Deep learning using CNN filters for universal manipulation detection	Learned CNN filters	CNN / High accuracy across manipulations
CASIA / Splicing	Deep CNN features for splicing and copy-move detection	CNN features	CNN / Robust detection of splicing and copy-move
Multiple Datasets	Review of media forensic methods	Review	Comprehensive survey of techniques and datasets
CASIA / Manipulation	Multi-level CNN feature extraction	CNN features	CNN / High accuracy for splicing detection
NIST / Forensics	Evaluation of image forensics detectors	Handcrafted + CNN	Benchmarking / Standard evaluation metrics
CASIA / Hybrid	Fusion of handcrafted and deep features	Handcrafted + CNN	SVM / Improved detection accuracy
Multiple Cameras / Splicing	CNN-based camera model fingerprinting	Camera model residuals	CNN / Generalization across devices
ImageNet / CASIA	Contrastive learning for visual representation	CNN embeddings	Contrastive loss / Robust features
Multiple Datasets / Deepfake	Review of deepfake detection methods	Review	Surveys challenges in deepfake detection
CASIA / Hybrid	Signal processing + deep learning feature fusion	Handcrafted + CNN	SVM / Robust and accurate detection
ImageNet / CASIA	Vision Transformer (ViT) with patch embeddings	Patch embeddings	Transformer / Global manipulation detection
CASIA / Cross-domain	Contrastive learning to improve cross-domain robustness	CNN / Transformer features	Contrastive loss / Cross-domain robustness
CASIA / Ensemble	TruFor method combining multiple clues	CNN + SPN + DWT	Ensemble / High accuracy and robustness
Multiple Datasets / Copy-Move	Review of advances in copy-move detection	Review	Comprehensive survey of methods
CASIA / Few-Shot	Few-shot learning with cross-domain adaptation	CNN features	Few-shot learning / Improved generalization
CASIA / ViT	Vision Transformer for global manipulation detection	Patch embeddings / Attention	Transformer / Accurate detection
CASIA / Fusion	Deep feature fusion using DWT and CNN embeddings	DWT + CNN embeddings	SVM / Improved generalization
Multiple Datasets / Deep Learning	Survey of deep-learning approaches	Review	Comprehensive deep learning survey
CASIA / Benchmarking	Benchmarking protocols and metrics for evaluation	Handcrafted + Deep	Standard evaluation metrics
CASIA / Multi-Domain	Multi-domain feature learning using CNN and handcrafted features	CNN + Handcrafted	CNN / Robust across domains

Continued on next page

Table 1 continued from previous page

Dataset	Methodology	Features	Performance
CASIA / Ensemble	Ensemble architectures combining CNN, SPN, and DWT	CNN + SPN + DWT	Ensemble / High accuracy and robustness
ImageNet / CASIA / CLIP	Feature transfer using CLIP embeddings	CLIP embeddings	Zero-shot transfer / Visual-text supervision
Multiple Datasets / Survey	Review of general image forgery detection techniques	Review	General survey of IFD methods

3 Datasets

Several benchmark datasets have been developed to evaluate the performance of image forgery detection methods. These datasets provide a wide range of manipulated images, including copy-move, splicing, and retouching forgeries. Table 2 summarizes the most commonly used datasets in the literature.

Table 2: Summary of commonly used datasets for image forgery detection.

Dataset	Forgery Type	Number of Images
CASIA v1.0 [6]	Splicing	921
CASIA v2.0 [6]	Splicing, Copy-Move	5123
Columbia Image Splicing Dataset [4]	Splicing	180
NIST Nimble Challenge [12]	Multiple	1000+
MICC-F220 [5]	Copy-Move	220
CoMoFoD [7]	Copy-Move	200

These datasets provide standardized benchmarks to evaluate both handcrafted and deep learning-based forgery detection methods. Researchers often apply data augmentation techniques to increase training samples, particularly for deep learning approaches [8, 9].

A wide variety of datasets have been developed to evaluate image forgery detection methods, focusing on different types of manipulations such as splicing, copy-move, and multiple forgery operations. The **CASIA dataset** is among the most widely used, with v1.0 containing 921 spliced images and v2.0 offering 5123 images with both splicing and copy-move manipulations [6]. The **Columbia Image Splicing Dataset** provides 180 spliced images specifically for splicing detection research [4].

For robust evaluation of multiple forgery types, the **NIST Nimble Challenge dataset** includes over 1000 images covering splicing, copy-move, and other manipulations [12]. Smaller, specialized datasets such as **MICC-F220** [5] and **CoMoFoD** [7] are widely used for evaluating copy-move forgery detection methods.

These datasets have become standard benchmarks, enabling researchers to compare handcrafted, deep learning, and hybrid approaches under controlled experimental settings. They also present challenges such as varying image resolutions, diverse content, and post-processing effects, which are critical for assessing the robustness and generalization capability of forgery detection algorithms.

4 Feature Extraction Techniques

Feature extraction plays a pivotal role in image forgery detection, as the effectiveness of any detection system largely depends on the quality and discriminative power of the extracted features. The goal of feature extraction is to represent images in a compact and informative manner, capturing the intrinsic properties of authentic and manipulated regions while suppressing irrelevant information. Effective features should be robust to common post-processing operations, such as scaling, rotation, compression, blurring, and noise addition, which are often applied to conceal traces of forgery [30, ?]. Feature extraction techniques in the context of image forgery detection can be broadly classified into three main categories: handcrafted features, deep learning-based features, and hybrid approaches that combine the strengths of both. Handcrafted features rely on domain knowledge and mathematical formulations to capture statistical, structural, and textural inconsistencies introduced by image manipulation.

These include techniques based on block-matching, keypoint detection, frequency-domain analysis, local texture descriptors, and camera-specific noise patterns [1, 2, 5, 3]. While handcrafted methods are computationally efficient and interpretable, their performance can degrade under sophisticated post-processing or cross-dataset scenarios.

Deep learning-based feature extraction, on the other hand, leverages neural networks to automatically learn hierarchical representations directly from raw pixel data. Convolutional Neural Networks (CNNs), Residual Networks (ResNet), DenseNets, and Vision Transformers (ViTs) have demonstrated significant improvements in forgery detection accuracy due to their ability to capture both low-level texture details and high-level semantic information [8, 11, 14, 18, 23]. Furthermore, contrastive learning and self-supervised learning frameworks enhance the generalization capabilities of deep models, particularly when labeled data is limited [15, 19].

Hybrid feature extraction approaches aim to combine the complementary strengths of handcrafted and deep features. By fusing statistical, textural, and noise-based handcrafted descriptors with deep embeddings, these methods can exploit both domain-specific and data-driven information. Hybrid frameworks often employ dimensionality reduction techniques, such as Principal Component Analysis (PCA) or Probabilistic PCA (PPCA), to eliminate redundancy and improve classifier performance. Moreover, ensemble and transformer-based architectures have been utilized to further enhance detection robustness across diverse datasets and manipulation types [17, 13, 28, 20].

In summary, feature extraction remains a cornerstone of image forgery detection systems. The choice of feature type—handcrafted, deep, or hybrid—directly influences the robustness, accuracy, and generalization capability of detection methods. Consequently, recent research has increasingly focused on hybrid and deep feature frameworks that can adapt to the evolving sophistication of image manipulations.

4.1 Handcrafted Feature Extraction

Handcrafted feature extraction methods form the foundation of early image forgery detection research. These approaches rely on manually designed descriptors derived from signal processing, statistical analysis, and texture modeling to capture artifacts introduced during image manipulation. Block-based and keypoint-based techniques, such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Scale-Invariant Feature Transform (SIFT), and Speeded-Up Robust Features (SURF), are extensively used for detecting copy-move and splicing forgeries by identifying duplicated or inconsistent regions within an image [1, 2, 5].

Texture-based descriptors, including Local Binary Patterns (LBP) and Gray Level Co-occurrence Matrix (GLCM), analyze local spatial relationships and have proven effective in identifying subtle texture irregularities caused by tampering. In addition, noise-based features, particularly Sensor Pattern Noise (SPN), exploit camera-specific fingerprints to expose splicing and composite forgeries where regions originate from different imaging devices [3]. Despite their interpretability, low computational complexity, and robustness in constrained environments, handcrafted features often struggle to generalize across datasets and are sensitive to common post-processing operations such as compression, resizing, rotation, and smoothing.

4.2 Deep Learning-Based Feature Extraction

Deep learning-based feature extraction has significantly advanced the field of image forgery detection by enabling automatic learning of discriminative representations directly from raw image data. Convolutional Neural Networks (CNNs) have demonstrated superior performance in capturing hierarchical features, ranging from low-level edges and textures to high-level semantic inconsistencies. Architectures such as AlexNet, VGG, ResNet, DenseNet, and EfficientNet have been widely adopted for forgery detection tasks, including splicing detection, copy-move localization, and generalized manipulation detection [8, 11, 14].

Recent research has also explored Vision Transformers (ViT), which model global contextual dependencies through self-attention mechanisms, making them particularly effective in detecting complex and large-scale manipulations [18, 23]. Furthermore, contrastive learning frameworks enhance representation robustness and cross-dataset generalization by learning invariant features from positive and negative sample pairs [15, 19]. While deep learning approaches achieve state-of-the-art accuracy, they typically require large labeled datasets, high computational resources, and careful regularization to avoid overfitting.

4.3 Hybrid Feature Extraction

Hybrid feature extraction approaches aim to combine the complementary strengths of handcrafted and deep features to improve robustness and generalization. By integrating low-level statistical descriptors with high-level CNN embeddings, hybrid frameworks capture both fine-grained forensic traces and semantic inconsistencies [17, 13]. Feature fusion can be performed at various stages, including early fusion (feature-level concatenation), late fusion (decision-level aggregation), or intermediate fusion through attention mechanisms.

To address feature redundancy and high dimensionality, dimensionality reduction techniques such as PCA and Probabilistic PCA (PPCA) are often applied prior to classification [24]. Recent hybrid models also incorporate ensemble learning and transformer-based architectures to further enhance detection accuracy under diverse attack scenarios [28, 20]. These approaches demonstrate improved robustness against post-processing and cross-dataset variations, making them attractive for real-world deployment.

4.4 Classification Methods

After feature extraction, classification models are employed to distinguish between authentic and forged images. The choice of classifier significantly influences the overall performance, robustness, and inter-pretability of image forgery detection systems.

4.4.1 Traditional Machine Learning Classifiers

Traditional machine learning classifiers remain popular due to their simplicity, efficiency, and effectiveness when combined with handcrafted or fused features. Support Vector Machines (SVMs) are widely used because of their strong generalization capability in high-dimensional feature spaces [2]. Random Forests (RF) and k-Nearest Neighbors (kNN) classifiers are also commonly applied for statistical and texture-based features, offering robustness to noise and ease of implementation. Ensemble learning strategies, which combine multiple classifiers, further enhance detection performance by reducing variance and improving decision stability [28].

4.4.2 Deep Learning Classifiers

Deep learning classifiers enable end-to-end learning by jointly optimizing feature extraction and classification. CNN-based classifiers have been extensively used for forgery detection due to their ability to model complex spatial dependencies and subtle manipulation artifacts [8, 9, 14]. Vision Transformer-based classifiers further improve performance by leveraging global context modeling and long-range feature interactions [18, 23].

Although deep classifiers consistently outperform traditional models, they require extensive training data and computational resources. To mitigate these challenges, recent studies employ ensemble strategies, transfer learning, and self-supervised pretraining to improve robustness, efficiency, and generalization across datasets.

5 Evaluation Metrics

The performance of the proposed pothole detection framework is quantitatively evaluated using standard classification and object detection metrics that are widely adopted in computer vision and deep learning literature. These metrics provide a comprehensive assessment of detection accuracy, reliability, and robustness under real-world conditions. Specifically, Accuracy, Precision, Recall, F1-score, Average Precision (AP), mean Average Precision (mAP), and the Receiver Operating Characteristic (ROC) curve with Area Under the Curve (AUC) are employed.

5.1 Confusion Matrix

All evaluation metrics are derived from the confusion matrix, which consists of four fundamental components:

- **True Positive (TP):** Correctly detected pothole instances.
- **True Negative (TN):** Correctly identified non-pothole regions.

- **False Positive (FP):** Non-pothole regions incorrectly classified as potholes.
- **False Negative (FN):** Actual potholes that were missed by the detector.

5.2 Accuracy

Accuracy measures the overall correctness of the detection system by computing the ratio of correctly classified instances to the total number of instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Although accuracy provides a general performance indication, it may be misleading in datasets with class imbalance, which is common in pothole detection scenarios.

5.3 Precision

Precision quantifies the reliability of positive detections by measuring the proportion of correctly detected potholes among all predicted potholes:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

High precision indicates a low false alarm rate, which is critical for practical deployment to avoid unnecessary maintenance actions.

5.4 Recall (Sensitivity)

Recall, also known as sensitivity, evaluates the ability of the model to correctly identify actual potholes:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

A higher recall value signifies fewer missed potholes, which is essential for ensuring road safety.

5.5 F1-Score

The F1-score provides a balanced measure by computing the harmonic mean of precision and recall:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

This metric is particularly useful when a trade-off between precision and recall is required.

5.6 Intersection over Union (IoU)

Intersection over Union (IoU) is used to evaluate the spatial overlap between the predicted bounding box and the ground-truth bounding box:

$$IoU = \frac{Area of \text{Overlap}}{Area of \text{Union}} \quad (5)$$

A detection is considered correct if the IoU exceeds a predefined threshold, typically set to 0.5.

5.7 Average Precision (AP)

Average Precision summarizes the precision-recall relationship by computing the area under the precision-recall curve:

$$AP = \int_0^1 P(R) dR \quad (6)$$

where $P(R)$ denotes precision as a function of recall R .

5.8 Mean Average Precision (mAP)

Mean Average Precision is obtained by averaging the AP values across all object classes:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (7)$$

where N represents the total number of classes. Since pothole detection is a single-class problem, mAP is equivalent to AP for the pothole class.

5.9 Receiver Operating Characteristic (ROC) Curve

The ROC curve illustrates the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) at various decision thresholds:

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

$$FPR = \frac{FP}{FP + TN} \quad (9)$$

5.10 Area Under the ROC Curve (AUC)

The Area Under the ROC Curve (AUC) provides a scalar measure of the classifier's discriminative ability:

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (10)$$

A higher AUC value indicates better overall classification performance and robustness against thresh-old variations.

6 Observations from Literature

A thorough review of existing image forgery detection techniques reveals several important trends and insights. Deep learning approaches, particularly CNN- and transformer-based models, generally outperform traditional handcrafted methods in tasks such as splicing and copy-move detection due to their ability to learn hierarchical and discriminative features directly from raw pixels [8, 9, 14, 23]. However, hybrid methods that combine handcrafted and deep features often provide enhanced robustness, particularly under post-processing operations such as compression, rotation, and scaling [17, 13]. Ensemble and transformer-based architectures further improve performance by capturing both local and global inconsistencies [20, 28, 18]. While deep models dominate recent research, handcrafted methods remain relevant for lightweight, interpretable, and low-data scenarios [1, 2, 5, 3]. The availability of standardized datasets, including CASIA, Columbia, CoMoFoD, and NIST Nimble Challenge, is critical for fair benchmarking and reproducibility [6, 12, 26]. Despite impressive results on individual datasets, cross-dataset generalization remains a challenge due to variations in imaging devices, forgery types, and post-processing operations [15, 19]. Emerging trends such as self-supervised and contrastive learning, few-shot detection, and multi-domain feature learning aim to address data scarcity and improve robustness [22, 24, 27]. Evaluation metrics commonly include accuracy, precision, recall, F1-score, and AUC, highlighting the need for standardized evaluation protocols [26, 25]. Additionally, robustness against adversarial manipulations and the development of interpretable models are important research directions [30]. Overall, these observations illustrate the evolution from handcrafted to deep and hybrid methods, emphasizing the critical roles of feature quality, classifier design, dataset diversity, and evaluation strategies in developing reliable and generalizable image forgery detection systems.

6.1 Future Work and Scope

Despite significant advancements in image forgery detection, several open challenges and research opportunities remain, providing ample scope for future work. One of the primary directions is the development of **generalizable detection frameworks** capable of maintaining robust performance across diverse datasets, manipulation types, and post-processing operations. Many existing methods are trained and

evaluated on specific datasets, leading to performance degradation when applied to unseen data distributions. Future research should emphasize domain adaptation, cross-dataset evaluation, and self-supervised learning to improve generalization.

Another important research avenue lies in **robust detection under advanced and emerging forgery techniques**. With the rapid evolution of generative models, particularly Generative Adversarial Networks (GANs) and diffusion-based image synthesis, forgery artifacts are becoming increasingly subtle and difficult to detect. Future detection systems must incorporate multi-scale analysis, frequency-domain cues, and semantic consistency modeling to effectively counter sophisticated deepfake and AI-generated content.

The integration of **hybrid and multimodal feature learning** also presents a promising direction. Combining spatial, frequency, noise, and semantic features with deep representations can further enhance robustness against complex manipulation pipelines. Additionally, leveraging multimodal data such as metadata, camera sensor information, and temporal cues from image sequences may significantly improve detection reliability in real-world forensic applications.

Lightweight and computationally efficient models are essential for deploying image forgery detection systems on resource-constrained platforms, such as mobile devices and edge computing environments. Future work should explore model compression, knowledge distillation, and efficient transformer architectures to balance detection accuracy with computational efficiency.

Another critical aspect is the need for **standardized datasets, benchmarks, and evaluation protocols**. Although several datasets exist, they vary widely in manipulation types, quality, and annotation standards. Future efforts should focus on creating large-scale, diverse, and continuously updated benchmark datasets that reflect real-world forgery scenarios. Moreover, establishing unified evaluation metrics and testing protocols will enable fair and reproducible comparisons across methods.

Finally, improving the **interpretability and explainability** of image forgery detection models is crucial, especially for applications in legal, forensic, and journalistic domains. Future research should emphasize explainable AI (XAI) techniques that provide visual and quantitative evidence supporting model decisions, thereby increasing user trust and practical usability.

In summary, future research in image forgery detection should aim to develop robust, explainable, efficient, and generalizable frameworks capable of addressing the growing challenges posed by advanced image manipulation techniques. These advancements will play a vital role in ensuring the integrity and trustworthiness of visual media in an increasingly digital world.

References

- [1] J. Fridrich, D. Soukal, and J. Lukas, "Detection of Copy-Move Forgery in Digital Images," in *Proc. Digital Forensic Research Workshop*, 2003.
- [2] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 758–767, 2004.
- [3] J. Lukas, J. Fridrich, and M. Goljan, "Detecting Digital Image Forgeries Using Sensor Pattern Noise," in *Proc. SPIE*, 2006.
- [4] T. T. Ng, S. Chang, and Q. Sun, "Blind Detection of Photomontage Using Higher Order Statistics," in *Proc. IEEE ISCAS*, 2009.
- [5] I. Amerini, L. Ballan, R. Caldelli, and A. Del Bimbo, "A SIFT-Based Forensic Method for Copy-Move Attack Detection," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1099–1110, 2011.
- [6] J. Dong, W. Wang, and T. Tan, "CASIA Image Tampering Detection Evaluation Database," in *Proc. IEEE ChinaSIP*, 2011.
- [7] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An Evaluation of Popular Copy-Move Forgery Detection Approaches," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1841–1854, 2012.
- [8] B. Bayar and M. C. Stamm, "A Deep Learning Approach to Universal Image Manipulation Detection," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2491–2503, 2017.
- [9] Y. Rao and J. Ni, "A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries," in *Proc. IEEE WIFS*, 2017.

- [10] L. Verdoliva, "Media Forensics: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 16–28, 2018.
- [11] P. Zhou, X. Han, V. Morariu, and L. Davis, "Learning Rich Features for Image Manipulation Detection," in *Proc. IEEE CVPR*, 2018.
- [12] P. Korus and J. Huang, "Evaluation of Image Forensics Detectors: NIST Case Study," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 3211–3224, 2018.
- [13] C. Hsu, Y. Chen, and Y. Chang, "Hybrid Feature Fusion for Enhanced Image Forgery Detection," *J. Vis. Commun. Image Represent.*, vol. 61, pp. 234–245, 2019.
- [14] D. Cozzolino, G. Poggi, and L. Verdoliva, "Noiseprint: A CNN-Based Camera Model Fingerprint," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 144–159, 2020.
- [15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in *Proc. ICML*, 2020.
- [16] L. Verdoliva, "Media Forensics and Deepfakes: An Overview," *IEEE JSTSP*, vol. 14, no. 5, pp. 910–932, 2020.
- [17] R. Jain and A. Kumar, "Signal Processing and Deep Learning Feature Integration for Forgery Detection," *IEEE Access*, vol. 9, pp. 109827–109840, 2021.
- [18] A. Dosovitskiy et al., "An Image Is Worth 16x16 Words: Transformers for Image Recognition," in *Proc. ICLR*, 2021.
- [19] W. Li, L. Zhang, and N. Liu, "Contrastive Learning for Robust Image Forgery Detection," *Pattern Recognition*, vol. 127, 2022.
- [20] F. Guillaro, D. Cozzolino, and L. Verdoliva, "TruFor: Leveraging All-Round Clues for Trustworthy Image Forgery Detection," in *Proc. ECCV*, 2022.
- [21] R. Keshari, R. Singh, and A. Chaudhary, "Advances in Copy-Move Forgery Detection: A Review," *Multimedia Tools Appl.*, vol. 81, 2022.
- [22] C. Qian, J. Dong, and J. Huang, "Few-Shot Forgery Detection via Cross-Domain Feature Adaptation," *Pattern Recognition*, vol. 134, 2023.
- [23] H. Li, J. Chen, and Y. Zhang, "Vision Transformer for Image Forgery Detection," *IEEE Trans. Multimedia*, vol. 25, pp. 987–999, 2023.
- [24] Q. Li, Y. Xu, and J. Chen, "Deep Feature Fusion for Image Forgery Detection," in *Proc. IEEE ICIP*, 2023.
- [25] M. Zanardelli, L. Bianchi, and G. Poggi, "Image Forgery Detection: A Survey of Deep-Learning Approaches," *Multimedia Tools Appl.*, vol. 82, 2023.
- [26] J. Huang, P. Liu, and H. Chen, "Benchmarking Image Forgery Detection: Protocols and Metrics," *IEEE Trans. Inf. Forensics Security*, vol. 18, 2023.
- [27] L. Wang, P. Liu, and H. Chen, "Multi-Domain Feature Learning for Robust Image Forgery Detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, 2024.
- [28] G. Bogacsovics, B. Harangi, and A. Hajdu, "Developing Ensemble Architectures for Image Forgery Detection," *Multimedia Tools Appl.*, 2024.
- [29] A. Radford, J. Kim, and C. Hallacy, "Learning Transferable Visual Models from Natural Language Supervision," in *Proc. ICML*, 2021.
- [30] H. Farid, "Image Forgery Detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, 2009.