Progress in Image Source Identification and Forensic Technology

Zequn Yu *

Mao Yisheng Honors College, Southwest Jiaotong University, Chengdu 611756, Sichuan, China

* Corresponding Author Email: Zaquinn1@outlook.com

Abstract. Image source identification forensics, as a core branch of passive digital image forensics, aims to identify the source device or generation method through the implicit physical or algorithmic characteristics of the image itself, in order to deal with the authenticity threat posed by image forgery technology. This paper systematically sorts out the technical evolution path of image source identification forensics, compares and analyzes the technical characteristics and limitations of traditional model-based methods and deep learning-driven methods, and discusses the four major sub-directions of camera source identification, computer graphics-generated image forensics, AI-synthesized image forensics, and re-acquired image forensics. Based on the shortcomings of existing research, this paper proposes key research directions such as the need to integrate physical prior knowledge with deep learning models, build a cross-modal general framework, and enhance adversarial robustness and unknown source detection capabilities. It also predicts that multimodal feature fusion, self-supervised learning, and explainable modeling will become the focus of future technological breakthroughs.

Keywords: traditional model approach, deep learning method, camera source recognition, AI synthetic image forensics, reacquired image detection.

1. Introduction

The popularity of digital image editing tools (such as Adobe Photoshop) and generative artificial intelligence (such as GANs and diffusion models) has significantly lowered the threshold for image forgery, posing a serious threat to social trust and the authenticity of judicial evidence [1]. Image source identification forensics, as a key means of passive forensics, traces the original source by mining the "source fingerprint" in the image. It has urgent application needs in areas such as news authenticity verification and judicial evidence identification.

However, traditional methods rely on artificially designed physical or algorithmic features. While achieving high accuracy in certain scenarios, they still suffer from poor generalization, computational complexity, and susceptibility to adversarial attacks. While some research has attempted to improve robustness through feature fusion, it remains limited by the inherent bottlenecks of feature engineering. With the rise of deep learning, research has gradually shifted to data-driven automatic feature learning methods.

Based on this, this article aims to continue the work of predecessors, first outlining the existing technical paths, then deeply analyzing the deep-seated problems of current research, and finally putting forward personal opinions and predicting future development trends, providing new methods and directions for subsequent research.

2. Deep Learning-Driven Approaches: Architecture Evolution and Technological Breakthroughs

Deep learning significantly improves the accuracy and generalization of image source identification through end-to-end feature learning, demonstrating significant advantages in practical forensic applications. This section introduces its typical frameworks, principles, and methodological advances in four categories.

2.1. Camera Source Identification

Camera source identification aims to determine which model or camera an image was taken by using device-specific traces embedded in the image (such as sensor noise, color processing pipeline, etc.). Traditional methods mainly rely on hand-designed features, such as sensor pattern noise (PRNU) [2]. PRNU first performs wavelet transform denoising on multiple images to obtain noise residuals; then estimates the camera fingerprint by aligning and averaging multiple noise residuals; finally, correlation detection is used to determine whether the test image comes from the target camera. Although it has high accuracy in some cases, this method is sensitive to image content, compression, and geometric transformations, and has limited generalization ability.

Deep learning methods based on CNN and Transformer architectures have been widely used for camera source recognition [3]. These methods can automatically learn more discriminative features, such as local noise patterns and global color distribution. Multi-branch CNN structures can extract texture and statistical features separately, and then fuse them through an attention mechanism to improve robustness to content changes and compression [4]. Transformer, on the other hand, captures long-range dependencies through a self-attention mechanism, and performs well in cross-brand recognition tasks.

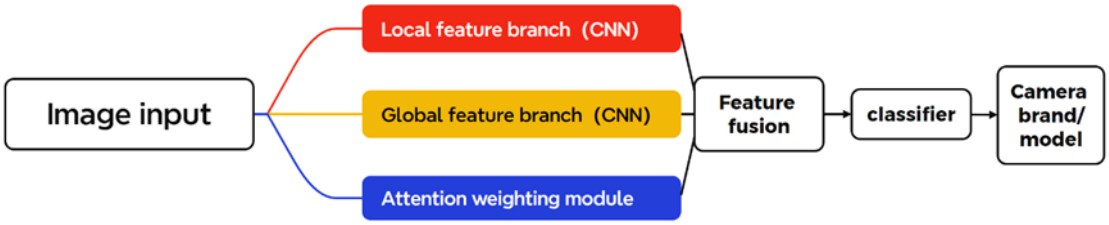

Figure 1. Typical camera source recognition framework using a multi-branch CNN combined with an attention mechanism (Picture credit: Original)

As shown in Figure 1, the local feature branch uses convolutional layers to extract detailed features such as sensor noise and edge response. The global feature branch extracts color distribution and brightness statistics through global average pooling or fully connected layers. The attention weighting module performs a weighted fusion of the features from the two branches to highlight device-specific patterns. The classifier outputs the camera model or device ID. This method effectively leverages differences in camera hardware and image processing pipelines to achieve high-precision source identification. In forensic evidence collection, this method can be used to track the source device of images involved in a case.

2.2. Image Forensics from Computer Graphics

Image forensics in computer graphics aims to distinguish between real-life photographic images and computer-generated (CG) images, typically generated by 3D rendering software such as Blender and Maya. Traditional methods rely on hand-crafted physical consistency features, such as lighting consistency, which examines whether shadow direction and specular reflections conform to physical laws; texture statistics, which uses high-order wavelet statistics to detect overly uniform or repetitive textures; and edge anomalies, where CG image edges are often overly sharp or lack natural transitions. While these methods are somewhat effective, their limited feature representation makes them inadequate for high-quality rendered images.

In recent years, deep learning- based frameworks have made significant progress by automatically learning unnatural patterns in CG images (such as unusual edges, texture transitions, and inconsistent lighting) [1]. Contrastive learning frameworks (such as the Twin Towers network) improve feature discrimination by bringing similar samples closer together and pushing dissimilar samples further away. The Transformer architecture can effectively capture long-range dependencies between image patches and identify unnatural repetitive textures. Inverse rendering techniques attempt to infer

parameters such as lighting and materials from images. If the inference results seriously violate physical laws, the image is judged to be CG.

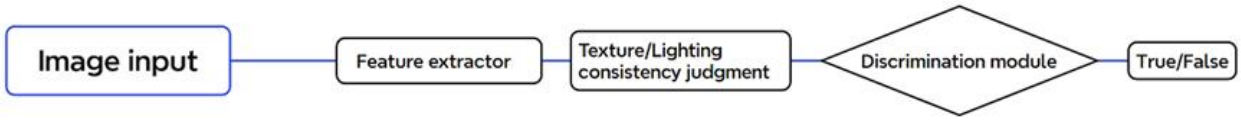

Figure 2. Typical process of CG image forensics (Picture credit: Original)

As shown in Figure 2, the feature extractor uses CNN or ViT to extract multi-scale features, focusing on edges, textures, and illumination responses [5]. The consistency analysis module verifies the physical consistency of the extracted features (e.g., shadow direction, reflection consistency). The discriminant module is used to output whether the image is real or CG. This method distinguishes CG images based on their non-realistic characteristics during the physical rendering process. In media content review, it can automatically identify fake news images generated by CG.

2.3. AI Synthetic Image Forensics

AI-generated image forensics aims to detect images synthesized by generative models (such as GANs and diffusion models). These images often have traces of frequency anomalies and local statistical inconsistencies. Traditional methods mainly include frequency domain analysis, which uses FFT/DCT to detect checkerboard artifacts or spectral anomalies introduced by the generative model [6]; and local statistical detection, which analyzes whether the color distribution, local variance, etc., are natural. These methods are effective for early GAN-generated images, but have limited detection capabilities for new generation technologies such as diffusion models.

Deep learning methods primarily focus on multimodal fusion and latent space analysis [7]. Frequency domain analysis methods are often combined with spatial features to form a multi-stream network structure. Metadata (such as EXIF information) is also used to assist in judgment. Focusing on diffusion models, researchers have begun analyzing the distribution characteristics of their latent spaces to identify synthetic images [8].

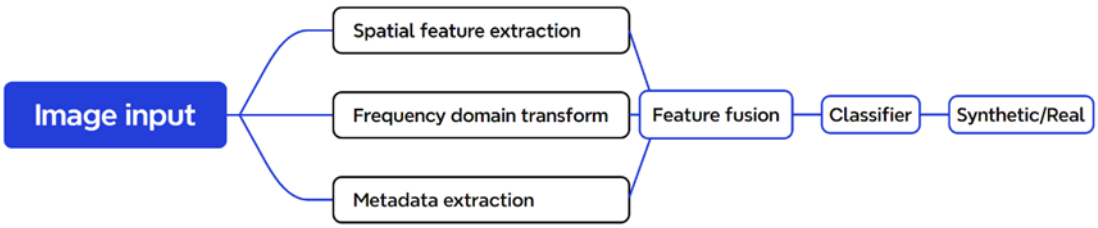

Figure 3. A multimodal fusion AI synthetic image detection framework (Picture credit: Original)

Figure 3, spatial feature extraction is used to extract pixel-level features such as texture and color; frequency domain transformation performs DCT/FFT transformation on the image to extract spectral features; metadata extraction refers to parsing information such as EXIF to determine whether the creation tool is consistent with the visual features [9]; the fusion and classification module is used to integrate multimodal features for final judgment. This method fully utilizes the inconsistency of generated images in multiple dimensions. On social media platforms, this type of technology can effectively detect malicious synthetic content such as Deepfake.

2.4. Re-acquisition of image evidence

Recaptured image forensics aims to detect whether an image has been screen-reproduced or re-photographed. Such images often contain traces of moiré, color distortion, and mixed compression artifacts. Traditional methods include moiré detection, which identifies periodic patterns in the frequency domain; color shift analysis, which examines color differences for anomalies; and compression artifact analysis, which detects the effects of multiple compression artifacts. These methods are effective in simple scenarios, but their effectiveness is limited for complex backgrounds and high-definition screen-reproduced images.

Deep learning methods, combined with frequency domain preprocessing and CNN feature learning, significantly improve detection capabilities [10]. Light field consistency analysis models the interaction between the screen pixel arrangement and the optical properties of the camera, and achieves high-precision detection by analyzing the light field inconsistencies in the retaken images.

Figure 4. Typical process for re-acquiring image forensics (Picture credit: Original)

As shown in Figure 4, frequency domain preprocessing performs DCT/FFT transforms on the image to enhance moiré and periodic patterns. Light field feature extraction uses a CNN to model light field inconsistencies caused by lens distortion and focal length differences. Consistency checking determines whether the image meets the light field constraints of natural photography. The discriminant output determines whether the image is a recaptured image. This method makes a comprehensive judgment based on multiple artifacts introduced during the recapture process. In financial document verification, fraudulent images generated by screen re-photographing can be detected.

3. Evaluation Metrics and Datasets

To fairly evaluate algorithm performance, multiple standard datasets are widely used. These datasets provide a foundation for model training and validation, but they still need to be expanded in diversity and challenge to address emerging forgery techniques. Table 1 lists representative datasets for each sub-discipline and their characteristics.

For camera source recognition, the Dresden database has become a mainstream benchmark because it contains RAW and JPEG images from multiple camera brands, along with physical annotations such as sensor noise. It is typically evaluated using accuracy and F1-score to account for class imbalance.

In the field of computer graphics forensics, Columbia University Library provides comparison samples of real-life images and 3D rendered images, covering a variety of rendering engines. Commonly used indicators such as precision, recall, and AUC are used to evaluate the model's distinguishing ability.

For AI-generated image forensics. It contains facial images synthesized by various generative models and introduces adversarial samples to test the generalization and robustness of the models. Accuracy, F1-score, and EER (equal error rate) are commonly used for comprehensive evaluation.

In the reacquired image forensics task, the Reacquired Image Database provides reacquired images for different screen types and camera device combinations, and annotates the secondary acquisition parameters. It is suitable for detecting fraudulent behaviors such as screen re-photography. Precision and recall rates are commonly used to evaluate model performance.

Table 1. Main datasets for image source identification and forensics

Sub-direction	Dataset name	Data scale	Data content	Common evaluation indicators	Features	References
Camera Source Identification	Dresden Image Database	15,000+	RAW and JPEG images captured by various camera brands	Accuracy, F1-score	Include sensor noise annotation	[1][2]
Computer Graphics Forensics	Columbia University Library	5,000	Comparison between real and rendering images	Precision, recall, AUC	Covering software such as Blender/Maya	[1]
AI synthetic image forensics	Deepfake Detection Challenge	128,000	Facial images synthesized by multiple generative models	Accuracy, F1-score, EER (equal error rate)	Contains adversarial examples	[6]
Re-acquired image forensics	Reacquired Image Database	10,000	Combinations of different screen types and shooting equipment	Precision and recall	Marking secondary acquisition parameters	[5]

4. Conclusion

Despite significant progress in deep learning, image source identification forensics still faces three major challenges: First, the model is not interpretable enough and the decision-making process lacks theoretical support, making it difficult to achieve judicially recognized reliability in actual deployment. Second, the robustness to adversarial forces is weak, and even slight perturbations can lead to significant performance degradation, which seriously affects the stability of practical applications. Finally, the detection ability of unknown sources or newly generated models is limited. For example, the false detection rate of diffuse images is high, which cannot effectively cope with the rapidly evolving generation technology. To address the above challenges, the following directions should be focused on in the future: First, the integration of physical priors and deep learning, such as introducing sensor models through noise consistency loss, can improve the reliability of the model in complex real-world environments. Second, the construction of a cross-modal general framework, the use of meta-learning to achieve feature sharing, and reduce the dependence on large amounts of labeled data in actual deployment. Third, the enhancement of adversarial defense mechanisms, such as dynamic adversarial training and feature masking, can improve the stability of the system in adversarial environments. Fourth, the development of zero-shot detection technology, the use of self-supervised learning to achieve unknown source identification, and expand the scope of practical applications. Looking ahead, multimodal feature fusion is expected to improve cross-task detection accuracy and significantly enhance the efficiency of actual forensics. Interpretability tools (such as Grad-CAM) can enhance decision transparency and meet the interpretability requirements of forensic evidence. Self-supervised pre-training will reduce reliance on annotations and lower practical application costs. Light field physics modeling is expected to push re-acquisition detection to even higher accuracy, providing more reliable technical support for key sectors such as finance and justice. These advances will collectively propel image source identification forensics from laboratory research to large-scale practical application.

References

[1] Chen Yifang, He Ziqiang, Wen Guanchen, et al. A review of image source identification and forensics research. Signal Processing, 2021, 37 (12): 21.

[2] Jia R, Wang X. Research on super-resolution reconstruction algorithm of image based on generative adversarial network. Journal of Physics Conference Series, 2021, 1944 (1): 012014.

[3] Banerjee C, Doppalapudi TK, Pasiliao E, Mukherjee T. Deep feature learning for intrinsic signature-based camera discrimination. Big Data Mining and Analytics, 2022, 5 (3): 206 - 227.

[4] Su Shuzhi, Xie Jun, Ping Xinrui, et al. Graph-enhanced canonical correlation analysis and its application in image recognition. Journal of Electronics & Information Technology, 2021, 43 (11): 8.

[5] Zhang Minhai, Wu Xinkai, Zhang Tingting. Digital image processing system based on JPEG compression coding algorithm. Computer Systems and Applications, 2022 (10).

[6] Liu Yuqing, Feng Junkai, Xing Bowen, et al. Intelligent ship target detection based on machine vision. Chinese Ship Research, 2021.

[7] Huang HH. Robust texture classification via group-collaboratively representation-based strategy. Journal of Electronic Science and Technology: English Edition, 2022 (4).

[8] Liu Z, et al. Exploring simple and transferable recognition-aware image processing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (3): 3032 - 3046.

[9] Guo M-H, et al. Attention mechanisms in computer vision: A survey. Computational Visual Media, 2022, 8 (3): 331 - 368.

[10] Hegde C, et al. Indoor group identification and localization using privacy-preserving edge computing distributed camera network. IEEE Journal of Indoor and Seamless Positioning and Navigation, 2024, 2: 51 - 60.