Journal of Network Intelligence ©2024 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **9**, Number **4**, November **2024**

Source Camera Identification Method Based on Multi-Scale Feature Fusion

Li Jiang
Hangzhou Transportation Information Technology Co., Ltd., Hangzhou 310030, China
MerryZH@126.com

Kang-Ming Liang
Polytechnic Institute
Zhejiang University, Hangzhou, 310015, P.R.China
liangkm2021@163.com

Shi-Ying Yu*
Zhejiang Chingo Software Co., Ltd., Hangzhou 310030, China
yusy@chingo.cn

Zhe-Ming Lu
School of Aeronautics and Astronautics
Zhejiang University, Hangzhou 310027, China
zheminglu@zju.edu.cn

*Corresponding author: Shi-Ying Yu

Received August 17, 2023, revised December 20, 2023, accepted June 5, 2024.

Abstract. *Source camera identification has become one of the research hotspots in the field of digital image forensics. Most existing methods are based on deep neural network models. While these methods have improved traceability accuracy compared to traditional methods, their performance in terms of accuracy becomes mediocre when the device categories are expanded. Moreover, they require retraining on a complete dataset or fine-tuning on newly added datasets. To address these challenges, this paper introduces a source camera identification method based on multi-scale feature fusion. Different scales of convolutional kernels are used to sample the input image, and parallel residual networks obtain sensor pattern noises at different granularities. A fusion network layer then inputs the merged features into a Softmax layer for classification results. Furthermore, to avoid repeated training due to class expansion, high-dimensional network features are extracted to construct an index vector database for retrieval classification. Experimental results demonstrate that the multi-scale feature fusion method achieves higher accuracy in camera traceability tasks. Additionally, the proposed retrieval mode effectively addresses the category expansion problem with minimal accuracy loss.*

Keywords: Digital image forensics, Source camera identification, Multi-scale feature fusion, Sensor pattern noise

1. **Introduction.** In the current digital era, due to the rapid advancement of computer technology, the Internet, and artificial intelligence, digital capture devices, especially mobile smart devices, have proliferated to every corner. This means that images can be easily taken, synthesized, edited, tampered with, and quickly shared. Consequently, the reliable determination of an image's origin and integrity has become a focal point for both the

public and experts. The technology of source camera identification is central to the field of digital image forensics [1], allowing for the tracking of the image-producing device and determining whether it has been edited. The practical applications of this technology are quite extensive: from criminal investigations and intellectual property protection to forensic identification and business strategy optimization, its influence is undeniable [2, 3, 4, 5]. On the other hand, deep learning techniques have been extensively applied in various aspects of image and video feature extraction and classification [6, 7, 8, 9].

The generation of images is a complex process, encompassing multiple steps, from the reception of light signals to the final formation of digital images [10, 11]. Each step could potentially embed device-specific characteristics in the image, making it possible to differentiate even between devices of the same model. From the perspective of pattern recognition, the technology of source camera identification is essentially a classification problem, aimed at classifying and recognizing the origin of a device by identifying the unique noise patterns in the image. Existing methods can be primarily categorized into those based on traditional digital image processing algorithms [12, 13, 14, 15, 16, 17] and those rooted in deep learning algorithms [18, 19, 20, 21, 22, 23]. While traditional methods offer acceptable levels of accuracy, deep learning-based approaches typically exhibit superior traceability precision on specific datasets. However, the performance of existing methods on large-scale datasets still requires improvement. Furthermore, when introducing new categories, there is often a significant need for extensive retraining and fine-tuning, undoubtedly resulting in a substantial computational burden.

To further optimize traceability accuracy, this paper introduces a source camera identification method based on multi-scale feature fusion. This method integrates image feature extraction at various granularities and amplifies the features through parallel residual networks, thereby achieving higher traceability precision. In addition, we construct a device index database using high-dimensional feature vectors to efficiently retrieve device fingerprints when new categories are introduced, while maintaining a high level of traceability accuracy. Comprehensive validation on the VISION [22] and Daxing [23] datasets underscores the efficacy of the proposed approach.

2. **Related Work.** Regarding the source camera identification technology based on digital image processing algorithms, Choi et al. [12] proposed that by measuring the radial distortion of the lens and pixel intensity, it can be considered as a "fingerprint" of the camera, achieving an identification accuracy rate of 90%. Dirik et al. [13] observed that dust on the camera lens leaves distinctive patterns on images, which can be used for identification. Lukás [14] and Goljan [15] were the first to explore the identification of camera sensors through PRNU (Photo-Response Non-Uniformity). They extracted noise fingerprints using wavelet transform and matched them using peak correlation energy. The process includes: preprocessing images, removing noise, calculating sensor pattern noise, and building a camera fingerprint database. Quan et al. [16] found that ISO (camera sensitivity) affects the correlation of PRNU and suggested compensatory processing. Popescu et al. [17] determined whether an image had been tampered with by examining color interpolation traces. Overall, source camera identification techniques based on traditional digital image processing have shown mediocre performance in terms of accuracy.

Source camera identification techniques based on deep learning can be divided into two categories: partially based on deep learning and fully based on deep learning. The former adopts the traditional digital image process framework but replaces traditional low-pass filters with denoising neural networks like DnCNN and FFDNet. References [18, 19, 20, 21, 22, 23] demonstrate that source camera identification based on convolutional neural networks (CNNs) exhibits advantages over previous methods. For instance, Gao [18] employs a multi-stage progressive neural network as a noise extractor, achieving superior performance on the dataset. On the other hand, approaches based on fully deep learning utilize supervised learning and neural network downsampling to transform images into high-dimensional vector representations. The category probabilities are obtained through a Softmax layer, and the output result is determined by the highest probability. This method exhibits higher traceability efficiency and accuracy, but the retraining or fine-tuning is required when expanding sample categories, resulting in significant time overhead. Marra et al. [21] propose a deep learning framework for iris photos and personal cameras, capable of simultaneously authenticating user identity and camera sensors. Chen et al. [22] apply residual networks to source camera identification tasks, achieving commendable results on multiple classic datasets. David [23], through a comprehensive analysis of convolutional networks with different depths, provides an end-to-end traceability implementation using shallow neural networks. While the aforementioned methods are effective for specific device types, they necessitate retraining or fine-tuning when expanding categories. In contrast, our proposed method exhibits stronger generalization capabilities.

3. **Method Elaboration.** This paper presents an image source camera identification method based on multi-scale feature fusion. Given an input image for testing, the model extracts its feature vector, matches it with fingerprints in the device fingerprint database, and ultimately outputs the retrieval results, as illustrated in Figure 1. The upper half of the figure delineates the training process of the neural network model based on multi-scale feature fusion: through supervised learning based on cross-entropy loss on a designated dataset, the model gains the ability for feature extraction. The lower half of the figure represents the model deployment and usage process: input a test image, obtain the feature vector of the image's pattern noise through the model, and then perform retrieval-based source camera identification. In comparison to the upper half of the figure, the model part in the lower half reduces the gray structure, indicating the direct extraction of the model's high-dimensional vector as the feature vector for the image's pattern noise. The overall process is as follows:

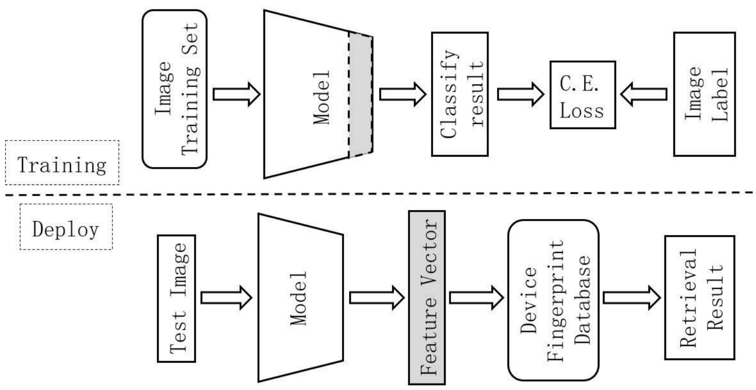

Figure 1. Overview of the Method

3.1. **Neural Network Based on Multi-scale Feature Fusion.** The Multi-scale Fusion Convolutional Neural Network (MFCNN) proposed in this paper is a convolutional neural network-based image classification model. It is capable of extracting features from images at different scales and merging them to enhance classification accuracy. In the context of source camera identification tasks, this network excels in extracting sensor pattern noise from photos, providing a data foundation for the retrieval of source camera types.

MFCNN primarily consists of two modules: the multi-scale convolution module and the feature fusion module. Among them, the multi-scale convolution module is the focus of network design, and addressing the characteristics of sensor pattern noise, such as how to design an appropriate network structure and achieve effective representation, is the core issue of the research. Given that the size of the convolutional kernel directly determines the local receptive field's size, and considering sensor fingerprints as a very subtle noise, to capture more comprehensive features, features obtained by applying multi-scale convolutional kernels are fed into parallel residual blocks of different granularities. Benefiting from residual connections within the blocks, the model's expressive and generalization abilities are enhanced, and the stability of the model training process is improved. The feature fusion module combines the downsampling results from parallel residual blocks and feeds them into the final classifier, and subsequently identify the source camera.

The complete algorithmic architecture, illustrated in Figure 2, takes a 64×64 image input. It's then processed by 3×3, 5×5, and 7×7 convolutional kernels of varying sizes. Following the convolutional layers, the features undergo batch normalization, ReLU activation, and pooling, before entering three parallel residual blocks. Each residual block consists of a 1×1 convolutional layer, batch normalization layer, ReLU activation layer, 3×3 convolutional layer, batch normalization layer, ReLU activation layer, 1×1 convolutional layer, batch normalization layer, ReLU layer, and pooling layer. Each branch of the residual block ultimately outputs a 256-dimensional deep vector. In the fusion module, these three 256-dimensional vectors are stacked to form a 768-dimensional vector. This vector is then processed through a basic fully connected layer and a Softmax layer to output the final result.

3.2. **The Source Camera identification Method Based on Feature Retrieval.**

In contrast to traditional digital image-based methods that can perceive explicit camera fingerprints during the modeling process, deep learning-based methods typically do not explicitly define such noise. Instead, they directly use the model's output to predict classification results. This approach necessitates retraining and refitting to new data distributions when expanding device categories, incurring significant overhead.

As described in Section 3.1, once the neural network based on multi-scale feature fusion is fully trained, the model possesses strong capabilities for extracting sensor pattern noise. These capabilities are stored in the form of high-dimensional vectors in the forward channel of the model. This paper introduces a source camera identification method based on feature retrieval, as illustrated in Figure 3.

This method is implemented in two stages. In the first phase, the generation phase, the entire training set is processed through the trained model for inference. The resultant 768-dimensional vectors (depicted as the gray module in Figure 3) produced during this inference are extracted and stored as fingerprint indexes corresponding to their respective categories. In the second phase, the retrieval phase, an input image designated for testing undergoes model inference, and its high-dimensional feature vector is extracted. This vector is then compared with each index in the device index database by calculating the Euclidean distance to determine its category affiliation. The final search outcome is then produced based on this comparison. The specific calculation is expressed in Formula 1, where $[x_1, x_2, \ldots, x_n]$ represents the high-dimensional vector obtained from the model for the test image, $[y_1^t, y_2^t, \ldots, y_n^t]$ represents the index vector corresponding to category t

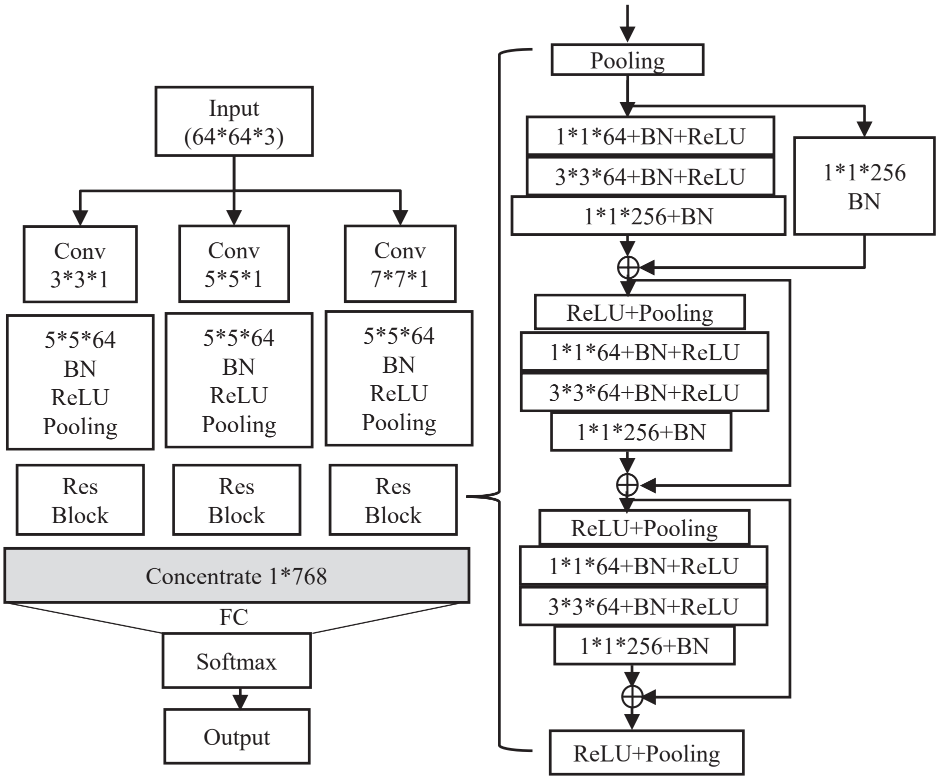

Figure 2. Architecture of the Multi-scale Feature Fusion Network

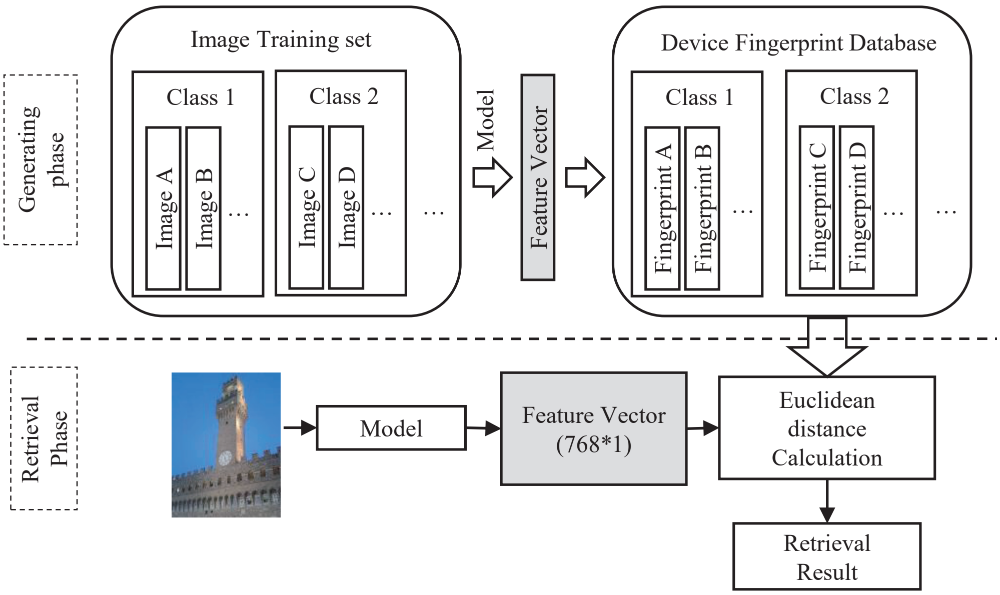

Figure 3. Flowchart of the Retrieval Method

in the fingerprint database, and D^t represents the Euclidean distance between the two vectors.

$$D^t = \sqrt{\sum_{i=1}^{n}(X_i - y_i^t)^2} \quad (1)$$

When facing an expansion of device categories, it is only necessary to repeat the process of the first stage for the new class samples. This involves generating fingerprint indexes corresponding to the newly added category. Subsequently, the source camera retrieval for the expanded device categories can be implemented as described in the second stage. As there is no need for repeated model training, the efficiency of deployment in scenarios involving device category expansion can be improved.

4. Experiment and Analysis.

4.1. Datasets. This study utilizes two datasets for performance validation testing. The first one is the VISION dataset, which consists of imaging devices from 11 brands with a total of 35 portable devices. This part of the dataset includes 34,427 JPEG images, making it a widely used dataset for image source attribution. The second one is the Daxing dataset, which pertains to digital imaging devices from 22 models comprising 90 smartphones. This dataset encompasses 43,400 JPEG images, distinguishing it as the largest current dataset for smartphone image source attribution.

To comprehensively validate the effectiveness of the method proposed in this paper, we set up experiments on three scales, containing image data from 10 device categories, 35 device categories, and 90 device categories respectively. The data for the 10 device categories is selected from the VISION dataset, which includes two iPhone 6 models with IDs D06 and D15, in addition to eight other different device models. The data for the 35 device categories comprises the complete VISION image dataset, which includes several devices of the same model. Lastly, the data for the 90 device categories is sourced from the complete Daxing image dataset. To make full use of the data in each scale of the experiment, images were uniformly segmented into non-overlapping blocks of 64×64 pixels. Subsequently, the data was split in a ratio of 8:1:1 for training, validation, and testing sets, respectively. The number of images in all three experiments is in the order of millions.

4.2. Experimental Setup. Environment and Hardware: The experimental code was constructed based on the PyTorch framework. It was run on the Ubuntu 18.04 operating system, with an Intel Xeon Gold 6230N processor, and the NVIDIA A100-80G graphics card.

Training Details: For model training, the Adam optimizer was employed in conjunction with the cross-entropy loss function. The initial learning rate was set to 0.01, with a batch size of 512. The entire training process consisted of 80 epochs.

Benchmark and Comparative Analysis: To highlight the effectiveness of the proposed method, Iris-Net [21], ResNet-18 [22] and DavNet [23] were selected as a benchmark for comparison. Based on existing research results, this method has demonstrated high accuracy in camera source tracing. Furthermore, to showcase the improvement brought about by multi-scale feature fusion compared to a single-branch network for this task, the three parallel network branches proposed in our method were each taken as individual methods and were evaluated in comparative experiments.

Category Expansion Evaluation: To demonstrate that the source camera identification method based on feature retrieval can effectively handle category expansion scenarios, the multi-scale feature fusion neural network model trained on the smallest scale dataset (comprising 10 categories) was tested across experiments of all three scales.

4.3. **Evaluation Metrics.** The primary evaluation metrics chosen for assessing the performance of the camera source identification method based on multi-scale feature fusion are accuracy and the confusion matrix.

Accuracy: Accuracy is the ratio of correctly classified samples to the total number of samples. It is generally used to evaluate the overall precision of a model.

Confusion Matrix: The columns of the confusion matrix represent predicted classes, while the rows represent the actual classes of the data. The diagonal represents the number of samples for which the model's prediction matches the actual label. This matrix can be used to analyze the classification accuracy for each individual category, thus allowing for an analysis of the model's precision at a granular level.

4.4. **Results Analysis.** (1) Analysis of Classification and Source Camera Identification Accuracy Based on MFCNN Model

This experiment initially conducts a comparative test on the source camera classification accuracy of the MFCNN model and other existing models. The experimental results are presented in Table 1. It is evident from the table that the proposed method in this paper consistently achieves the highest classification accuracy in the experiments. Thanks to the multi-scale convolution's ability to capture comprehensive image features and the feature enhancement capability of parallel residual networks, the model achieves an accuracy of 98.02% in experiments with 10 device classes. On the complete VISION dataset, it attains an accuracy of 93.2%, and even on the comprehensive Daxing dataset, it achieves an accuracy exceeding 85%. These results demonstrate the effective feature extraction capabilities of the proposed method for image pattern noise.

Table 1. Summary of source camera classification accuracy

Models	10-device Dataset	35-device Dataset (VISION)	90-device Dataset (Daxing)
Iris-Net	75.60%	63.36%	39.56%
ResNet-18	94.37%	90.99%	80.68%
DavNet	87.91%	70.47%	57.68%
Ours(3×3)	73.51%	91.13%	78.94%
Ours(5×5)	83.37%	90.62%	79.43%
Ours(7×7)	94.25%	91.00%	58.11%
Ours	**98.02%**	**93.20%**	**85.44%**

Subsequently, the classification module of the aforementioned model is transformed into a retrieval form, i.e., the original model's fully connected layer is removed, and high-dimensional vectors are used to represent camera fingerprints for retrieval matching. All methods employ pre-trained models obtained on data from 10 device classes. The experimental results are presented in Table 2. From Table 2, it can be observed that the proposed source camera identification method based on MFCNN achieves favorable traceability accuracy.

Table 2. Summary of source camera identification accuracy

Models	10-device Dataset	35-device Dataset (VISION)	90-device Dataset (Daxing)
Iris-Net	66.93%	47.92%	50.85%
ResNet-18	96.99%	79.11%	58.53%
DavNet	68.69%	45.00%	35.79%
Ours	**97.34%**	**86.90%**	**76.38%**

(2) Analysis of Category Expansion

From Tables 1 and 2, it is evident that as the number of devices increases, there is a noticeable decline in both the classification and source camera identification prediction accuracy for almost all comparative models. In some cases, the prediction accuracy falls below 50%. The reason for this outcome is the diminishing gap between different categories as the number of device types increases, especially when considering the growing diversity within the same device type. The limited expressive capabilities of the compared models result in a significant decrease in accuracy. In contrast, the use of the multi-scale feature fusion network in this paper shows a minimal decrease in accuracy as the device scale increases.

In Table 2, although all methods use pre-trained models obtained on a 10-device-class dataset, comparing Tables 1 and 2 reveals that the proposed method experiences the smallest decline in identification accuracy. It even achieves a traceability accuracy of 76% on a dataset with 90 device classes. This represents a decrease of only 9 percentage points compared to Table 1. Additionally, since it does not involve secondary training, the retrieval-based method exhibits a more cost-effective response to category expansion scenarios.

(3) Analysis of Confusion Matrix

10-device Test Performance: The exact accuracy rate for each category can be observed from the confusion matrix. Figure 4 showcases the prediction results of the multi-scale feature fusion network for the first experiment, while Figure 5 presents the results using the retrieval method. Both methods achieved remarkably high tracing accuracy rates in the 10-device test. However, there were some prediction errors for the D06 and D15 devices, which was anticipated. These two devices are the only ones of the same model in the first experiment. Devices of the same model have certain similarities in their sensor pattern noise, leading to reduced accuracy.

Performance on VISION and Daxing Datasets: The top row of Figure 6 displays the prediction results of the multi-scale feature fusion network for the second and third experiments, while the bottom row shows the results of the retrieval method. It can be observed that most of the dark pixel blocks are concentrated along the diagonal, proving the effectiveness of the camera tracing methods proposed in this study. However, it's noticeable that there's some feature coupling among certain devices. Devices labeled as 31 and 37 as well as 53 and 57 had certain misidentifications. These devices are all iPhones but of different models. This suggests that while the proposed method performs well, there's room for further optimization, especially for large-scale camera tracing tasks.

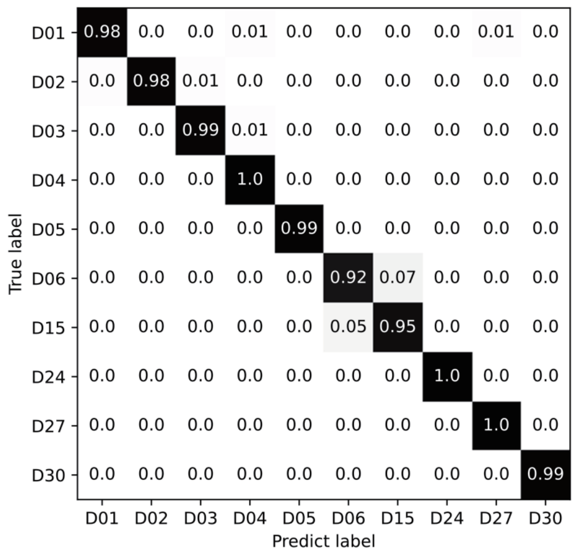

FIGURE 4. Confusion Matrix (Performance of feature fusion network on 10 devices)

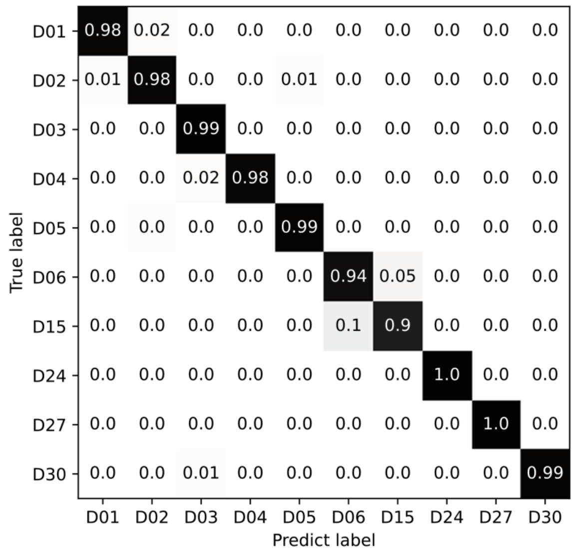

FIGURE 5. Confusion Matrix (Performance of retrieval method on 10 devices)

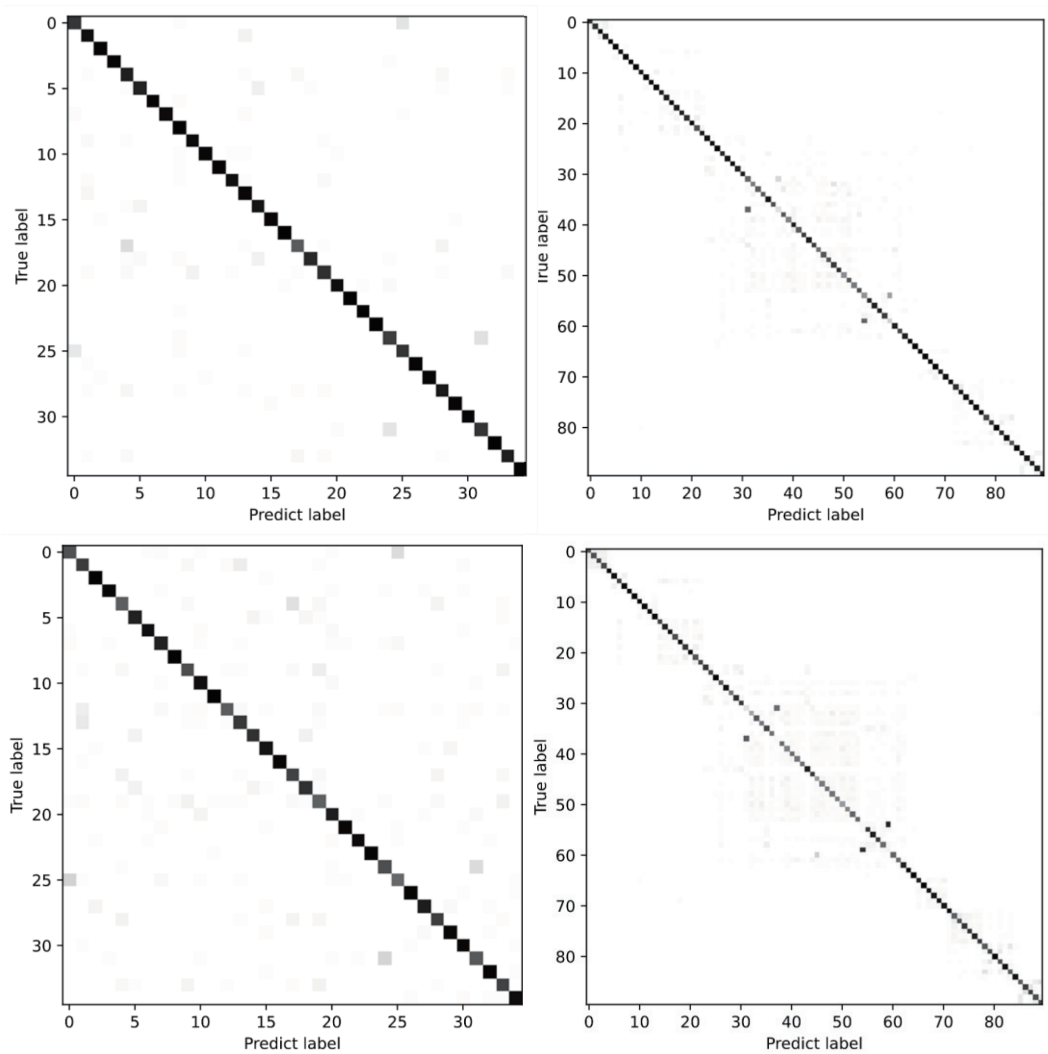

FIGURE 6. Confusion Matrix (Top row: Performance of feature fusion network on VISION and Daxing datasets; Bottom row: Performance of retrieval method on VISION and Daxing datasets)

5. **Conclusions.** This paper presents an effective method for extracting residual camera fingerprints in images based on multi-scale feature fusion. Compared to previous methods, this approach is adaptable to a larger scale of devices. By convoluting the image through various sizes, features of different granularities are extracted. These are then passed through parallel residual networks to amplify the features. This forms a feature vector which is introduced into the feature fusion layer, obtaining a 768-dimensional vector that represents sensor pattern noise. Thanks to feature fusion, the representational power of this vector is significantly superior to a single branch network, thereby achieving a higher accuracy rate for camera traceability. At the same time, by leveraging the superior representational ability of the high-dimensional vector, a device index database is constructed. When faced with category expansion, this allows for efficient fingerprint retrieval of devices, ensuring traceability accuracy close to a retrained model.

While the current work has shown promising results, there are still some challenges to address. For instance, there is a substantial decline in classification accuracy when the number of categories sharply increases. Additionally, questions remain about the accuracy of source camera identification when images undergo compression, cropping, or other processing. In future work, we plan to explore the introduction of alternative model structures and methods to enhance the accuracy and robustness of source camera identification.

Acknowledgment. This work is partially supported by the National Key Research and Development Program of China under Grant No.2020AAA0140000.

REFERENCES

[1] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14, no. 5, pp. 910–932 ,2020.
[2] H. Xiong, M. Yang, T. Yao, J. Chen, and S. Kumari, "Efficient Unbounded Fully Attribute Hiding Inner Product Encryption in Cloud-Aided WBANs," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5424-5432, 2022.
[3] H. Xiong, Z. Zhou, L. Wang, Z. Zhao, X. Huang, and H. Zhang, "An Anonymous Authentication Protocol With Delegation and Revocation for Content Delivery Networks," *IEEE Systems Journal*, vol. 16, no. 3, pp. 4118-4129, 2022.
[4] H. Xiong, Q. Mei, Y. Zhao, L. Peng and H. Zhang, "Scalable and Forward Secure Network Attestation with Privacy-Preserving in Cloud-Assisted Internet of Things," *IEEE Sensors Journal*, vol. 19, no. 18, pp. 8317-8331, 2019.
[5] K. Wang, C.-M. Chen, Z. Liang, M. M. Hassan, G. ML Sarne, L. Fotia, and G. Fortino, "A trusted consensus fusion scheme for decentralized collaborated learning in massive IoT domain," *Information Fusion*, vol. 72, pp. 100-109, 2021.
[6] J. Gao, H. Zou, F. Zhang, and T.-Y. Wu, "An intelligent stage light-based actor identification and positioning system," *International Journal of Information and Computer Security*, vol. 18, no. 1/2, 204-218, 2022.
[7] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human Motion Recognition Based on SVM in VR Art Media Interaction Environment," *Human-centric Computing and Information Sciences*, vol. 9, 40, 2019.
[8] F. Zhang, T.-Y. Wu, and G. Zheng, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP Journal on Image and Video Processing*, vol. 2019, 58, 2019.
[9] E. K. Wang, X. Zhang, F. Wang, T.-Y. Wu, and C.-M. Chen, "Multilayer dense attention model for image caption," *IEEE Access*, vol. 7, pp. 66358-66368, 2019.
[10] J. Bernacki, "A survey on digital camera identification methods," *Forensic Science International: Digital Investigation*, vol. 34, 300983, 2020.
[11] X. Jiang, S. Wei, R. Zhao, Y. Zhao, X. Du, and G. Du, "An Overview of Device Traceability Technology for Digital Images," *Journal of Beijing Jiaotong University*, vol. 43, no. 2, pp. 48–57, 2019.

[12] K. C. San, E. Y.Lam, K. Kenneth and Y. Wong, "Source camera identification using footprints from lens aberration," in *Digital photography II*, SPIE, 2006, pp. 172–179.

[13] A. E. Dirik, H. T. Sencar, and N. Memon. "Source camera identification based on sensor dust characteristics," in *2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, IEEE, 2007, pp. 1–6.

[14] J. Lukas, J.-J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[15] M. Goljan, "Digital camera identification from images–estimating false acceptance probability," in *Digital Watermarking: 7th International Workshop, International Workshop on Digital Watermarking (IWDW 2008)*, Springer Berlin Heidelberg, 2009, pp. 454-468.

[16] Y. Quan, and C.-T Li, "On addressing the impact of ISO speed upon PRNU and forgery detection," *IEEE Transactions on Information Forensics and Security*, vol 16, pp. 190–202, 2020.

[17] A. C. Popescu, and H. Farid, "Exposing digital forgeries in color filter array interpolated images,‘’ *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

[18] C. Gao, Y. Xiao, and H. Tian. "Image camera fingerprint extraction algorithm based on multi-stage progressive neural network," *Information Network Security*, vol. 22, no. 10, pp. 15–23, 2022.

[19] A. M. Rafi, U. Kamal, R. Hoque, A. Abrar, S. Das, R. Laganière, and Md. K. Hasan, "Application of DenseNet in Camera Model Identification and Post-processing Detection," in *CVPR Workshops 2019*, IEEE, 2019, pp. 19–28.

[20] P. Yang, R. Ni, Y. Zhao, and W. Zhao, "Source camera identification based on content-adaptive fusion residual networks,", *Pattern Recognition Letters*, vol. 119, pp. 195–204, 2019.

[21] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "A deep learning approach for iris sensor model identification," *Pattern Recognition Letters*, vol. 113, pp. 46–53, 2018.

[22] Y. Chen, Y. Huang Y, and X. Ding, "Camera model identification with residual neural network," in *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017, pp. 4337–4341.

[23] D. Freire-Obregón, F. Narducci, S. Barra, and M. C. Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognition Letters*, vol. 126, pp. 86–91, 2019.

[24] D. Shullani, M. Fontani, M. Iuliani, O.-A. Shaya, and A. Piva, "Vision: a video and image dataset for source identification," *EURASIP Journal on Information Security*, vol. 2017, no. 1, pp. 1–16, 2017.

[25] H. Tian, Y. Xiao, G. Cao, Y. Zhang, Z. Xu, and Y. Zhao, "Daxing smartphone identification dataset," *IEEE Access 7*, vol. 7, pp. 101046–101053, 2019.