Signal Processing: Image Communication 99 (2021) 116438

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Image source identification with known post-processed based on convolutional neural network

Xin Liao [a,b], Jing Chen [a], Jiaxin Chen [a,*]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
[b] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Science, Beijing 100093, China

ARTICLE INFO

Keywords:
Image source identification
PRNU
Tampered image
Operation chain
Convolutional neural network

ABSTRACT

Image source identification is important to verify the origin and authenticity of digital images. However, when images are altered by some post-processing, the performance of the existing source verification methods may degrade. In this paper, we propose a convolutional neural network (CNN) to solve the above problem. Specifically, we present a theoretical framework for different tampering operations, to confirm whether a single operation has affected photo response non-uniformity (PRNU) contained in images. Then, we divide these operations into two categories: non-influential operation and influential operation. Besides, the images altered by the combination of non-influential and influential operations are equal to images that have only undergone a single influential operation. To make our introduced CNN robust to both non-influential operation and influential operation, we define a multi-kernel noise extractor that consists of a high-pass filter and three parallel convolution filters of different sizes. The features generated by the parallel convolution layers are then fed to subsequent convolutional layers for further feature extraction. The experimental results provide the effectiveness of our method.

1. Introduction

With the development of science and technology, image acquisition devices are becoming more and more popular. Meanwhile, it is convenient for people to modify images by various graphics editors. In order to trust these images, identifying the model of the device that captured images become a paramount forensic problem with realistic significance [1]. Many forensic algorithms have been developed to solve the problem [2–9]

Due to the diversity of hardware, different imaging devices also leave different traces on the image. These subtle differences offer us a way to identify the camera model. Lukas et al. [2] used the sensor pattern noise as a fingerprint for identifying the source camera device. Color filter array (CFA) interpolation is an indispensable step in the imaging process, however, different cameras have different interpolation algorithms. The expectation maximization (EM) algorithm can be used to detect the local peak energy point in the frequency domain of digital images [3], and the authors thought the point reflected the correlation of CFA interpolation introduced to the local image. Estimating CFA interpolation coefficients exactly and then selecting suitable classifiers can also achieve higher source identification accuracy [4–6]. White Balance [7] is an important post-processing algorithm in the camera imaging system and its parameter estimation is also used in camera source identification of digital images. Kharrazi et al. [8] constructed a feature set to describe the camera models by using image quality features, color correlation, color energy ratio, and so on. In [9], the author identified digital cameras by analyzing pipeline and camera processing operations.

Recently, CNN has been widely used in the field of image source forensics. CNN learns features by training data, and these features are optimized in the process of continuous iteration. Finally, the purpose of extracting optimal features is achieved and the limitations of artificial design features are avoided. Besides, the learned features can also be fed into a support vector machine (SVM) for classification [10]. Yang et al. [11] classified the images into three subsets: saturation, smoothness, and others. Then, image patches were applied to three fusion residual networks to extract features. Finally, the images were classified by a Softmax layer. A CNN-based solution was introduced to evaluate patches of an image whether containing enough camera-related information [12]. Meanwhile, adding a preprocessing layer to improve the performance of a network is also a common method. The preprocessing layer can eliminate unnecessary noises and then obtain image residuals for the next step [13]. For example, in [14], by adding denoising high-pass filter in CNN, the interference of image edge and content information can be suppressed, and the recognition of multiple camera models can be realized. Proper preprocessing can highlight the characteristics of the tampering operation [15], which makes the network achieve better detection performance. In addition, a constrained

* Corresponding author.
E-mail address: chenjiaxin@hnu.edu.cn (J. Chen).

https://doi.org/10.1016/j.image.2021.116438
Received 26 October 2019; Received in revised form 25 June 2021; Accepted 10 August 2021
Available online 17 August 2021

convolutional layer and a nonlinear residual feature extractor are used in parallel to improve the robustness of the network [16].

Adjustments of the network structure are beneficial to improve classification accuracies too. Different network architectures directly impact the capability of learning discriminant features from the observed images [17]. By designing a CNN, David et al. [18] made camera identification not only be confined to traditional cameras, but also various mobile devices, such as mobile phones, embedded cameras, and so on. Stamm et al. [19] tested how the restriction convolution layer and high-pass filter influence results through several experiments, at the same time, the choice of activation functions also will cause some impacts on the classification accuracy. Eleni [20] improved AlexNet for source identification and united the PRNU of the camera to analyze the causes of error classification. To obtain better performance, the author enlarged differences of images from diverse models by extracting PRNU pattern noises [21]. Zhen [22] used a filter to acquire the original image and noise firstly, then fitted the CNN on the noise pattern to perform the classification.

However, in the real world, most of the images we reached may be manipulated by some post-processed or embedded secret information [23]. Thus, many researchers are detecting the traces left by image processing to reveal the processing history of an image [24]. These traces may alter the source camera information contained in the image, which reduces the accuracy of some existing image source identification approaches. Considering the actual application scenario of camera model identification techniques, it is important to devise methods to make them robust to these common post-processed operations.

In this paper, we analyze the reasons for the source detection accuracy decrease caused by post-processing firstly. As we know, in the imaging process, because of the imperfections of camera hardware, defects will be introduced into images which we always called noise. In the noise component, PRNU is the best characteristic that can help us to identify the camera model. Thus, depending on whether the PRNU is affected or not, these operations can be divided into two categories: non-influential operation and influential operation. As for images processed by non-influential tampering operations, this kind of operation does not affect the camera fingerprints. When images altered by influential tampering operations, due to the change of PRNU mode noise, the difficulty of camera identification is increased, thereby greatly reducing the verification accuracy. Therefore, we present a robust CNN structure based on the nature of different tampering operations and the possible impact on camera fingerprints. A multi-kernel noise extractor is inserted into our proposed CNN architecture as the first layer. Hence, the input image will be converted into noise residuals by a high-pass filter for suppressing the influence of image content. After that, three convolutions with different filter sizes are employed to obtain more comprehensive features among the various unit. The feature maps produced by the multi-kernel noise extractor are then concatenated and passed to subsequent convolutional layers for high-level abstract features extraction. Finally, image source classification results are obtained from the proposed CNN. Furthermore, we extend this framework to the operation chains. When identifying the source of images that have been modified by both influential and non-influential operations, unaffected tampering operations can be ignored. Hence, our proposed network can also obtain substantial performance improvement. Experimental results show that our proposed CNN architecture is robust to tampered image source identification.

The rest of this paper is organized as follows. Section 2 analyzes tampered image noise patterns in detail and introduces how to classify tampering operations. Section 3 shows our source identification network in detail. The multi-kernel noise extractor and the subsequent network architectures are introduced sequentially. In Section 4, we set some experiments to validate the robustness of our proposed method when images are modified by a single tampering operation or two tampering operations chain. Section 5 presents the corresponding discussions. The concluding remarks are given in Section 6.

2. Tampered image noise pattern analysis

Diversities of digital images captured by different types of imaging equipment are comprehensive reflections of the overall image acquisition process [25]. Due to the characteristics of hardware or software, defects will be introduced into the image, that is what we call noise. Noise can be divided into random read-in noise and system noise [2]. The former is unstable and highly influenced by the imaging environment. The latter mainly refers to sensor pattern noise (SPN) which is the inherent feature of the camera.

SPN consists of fixed pattern noise (FPN) and PRNU [2]. FPN is a current signal in the absence of light and can only be obtained in a completely dark shooting environment, but the images will not be taken in the dark usually. So we do not take FPN into consideration when identifying the source of images. While PRNU is caused by inconsistency of sensor response to illumination, the noise can be distinguished among different sensors. In addition, the PRNU generated due to manufacturing defects contains source information that is not affected by the photographic environment. These characteristics make PRNU become the optimum forensic features for camera model identification. PRNU can be estimated by the following equation [26].

$$I = I' + K \times I' + \theta \tag{1}$$

where I' represents an ideal noiseless image, I is the output image, and K is a zero-mean multiplicative factor responsible for PRNU. θ is a random error, for example, two cameras made with the same parts from the same manufacturer are slightly different due to defects in the manufacturing process. Then we can assume $I' = F(I)$, where $F(\cdot)$ is the filtering process. Let W denote the pattern noise which contains information of image source, we have

$$W = I - F(I) = I - I' = K \times I' + \theta \tag{2}$$

When images are subjected to post-processing, various tampering operations may have different impacts on the image pattern noise. While the tampering operation is a linear operation, the PRNU of images will not be affected. The prove is shown below,

$$I_T = T \times I = T \times I' + T \times K \times I' + T \times \theta \tag{3}$$

$$T \times I' = F(I_T) \tag{4}$$

$$W_T = I_T - F(I_T) = T \times K \times I' + T \times \theta = T \times W \tag{5}$$

$$W = W_T \times T^{-1} \tag{6}$$

where T represents a linear operation matrix, I_T is an altered image, W_T means the pattern noise of a tampered image, and T^{-1} denotes inverse matrix.

If $|T| \neq 0$, there must exist T^{-1} satisfies Eq. (6), then we can infer that W and W_T are linearly related. So we divide all tampering operations into two types:

(1) Non-influential operation: The tampering operation can be represented by a linear matrix T and $|T| \neq 0$.

(2) Influential operation: The tampering operation cannot be represented by a linear matrix.

For the images operated by non-influential operations, compared with original images, PRNU as the feature of camera model identification has no significant difference. Therefore, the tampering operations would not vary the source discrimination accuracy enormously. In the second case, the PRNU of an image is affected. Influences caused by these tampering operations make camera identification difficult because better identification performance needs to obtain high-quality PRNU. We propose a multi-kernel noise extractor to extract low-level image noise patterns and finally achieve the purpose of detecting the camera model efficiently.

Furthermore, the solution can be extended to a chain consists of two tampering operations. For an operation chain mixed by a non-influential operation A and an influential operation B:

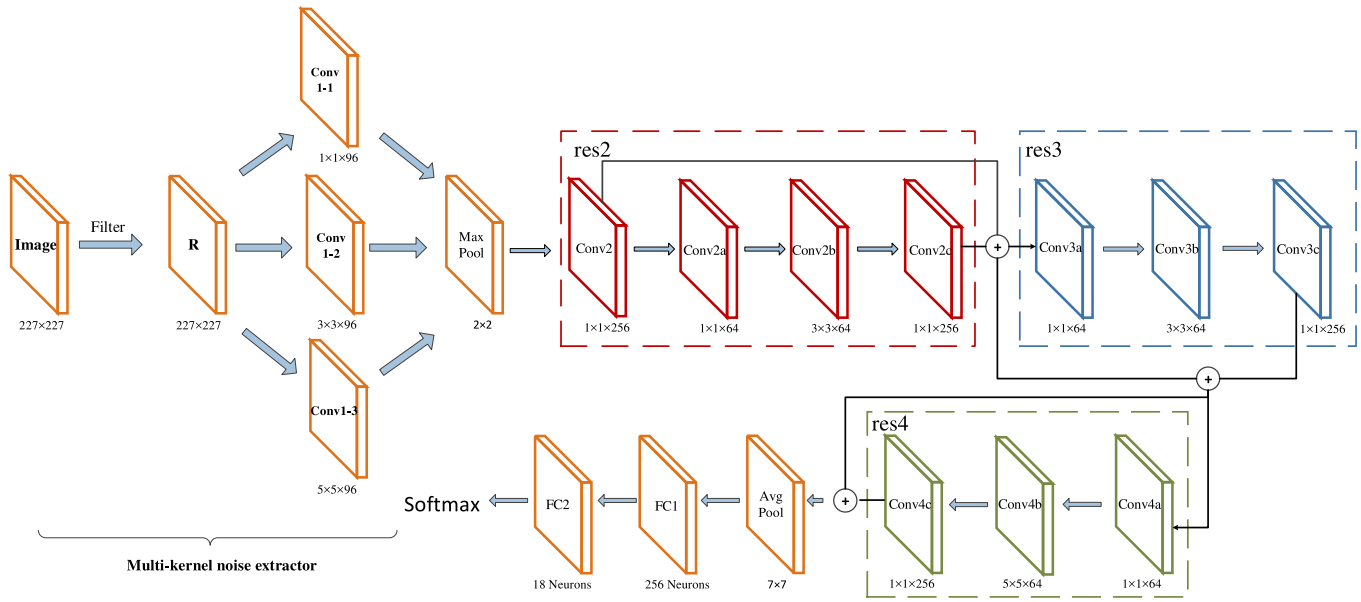

Fig. 1. Our proposed CNN architectures.

A-*B*: The original image is altered by *A* followed by *B*. Since the operation *A* has little effect on the source identification of the original image, we only need to consider the interference that tampering operation *B* may bring to the source identification. So the camera model identification of these images can share the same network with images that only undergone operation *B*.

B-*A*: The original image is first operated by *B* and then by *A*. The image modified by operation *B* can be regarded as an original image, then the subsequent process is equivalent to using single operation *A* to process a cover image. Since the influence of *A* is neglected, only interferences caused by the first tampering operation *B* are taken into consideration. We can also use the proposed network to determine the source of an image altered by *B* then *A*.

To sum up, for the camera model recognition of the image which operated by *A* and *B*, regardless of the order of manipulations, we can transform it into the source identification of images that have only been manipulated by *B*.

3. Design network for tampered image source identification

Recently, CNN has been widely applied to source camera model recognition and achieves high performance. In this paper, we combine CNN with the source feature extractor to make our method robust to post-processed images. Fig. 1 depicts the overall design of our proposed CNN architecture. Non-influential operations always occur at a single independent pixel, so we need 1×1 convolutions to retain these independent features. While the effects of influential operations on PRNU are usually generated between adjacent pixels, larger convolutions can be used to cover correlated units. Therefore, we propose a multi-kernel noise extractor to capture camera trails among various pixel stacks. Based on the following main components, this extractor can make our network robust to the post-processed image. Firstly, a high-pass filter is inserted to obtain the PRNU noise pattern. Secondly, in order to preserve the correlation of neighboring units, three paralleling convolution filters of different sizes are followed. Then the outputs of the multi-kernel noise extractor are fed into a regular convolutional layer. Deeper convolutional layers in the CNN will learn higher-level features. We present a detailed description of our architecture as below.

3.1. Multi-kernel noise extractor

Due to PRNU is the best feature for identifying the source of the camera, post-processed image source identification is typically interfered with by image content which should be suppressed. Then we can address source identification as a simple pattern noise classification problem. It has been found that PRNU is multiplicative noise and belongs to high-frequency signal [26]. So a high-pass filter [27] is set as the first part of the multi-kernel noise extractor, namely the noise extraction layer. For each image I, applying this filter is important to obtain PRNU that contain more camera fingerprints as follows:

$$R = I \times \frac{1}{12}\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix} \tag{7}$$

where R represents the image residual. Thus the PRNU noise pattern is modeled by the image residual between a pixel.

Then the output image residuals are processed by three kinds of convolutional kernel size in parallel: 1×1, 3×3, 5×5. We assume that each unit from an earlier layer corresponds to some region of the input and these units are grouped into filter banks. Thus, a convolutional kernel with different sizes implies various sizes of receptive fields to protect the integrity of forensics features. Applying 1×1 convolutional kernel can reduce dimension without changing the size of the feature map. Besides, the convolution process of a 1×1 convolutional kernel is equivalent to the calculation process of a full connection layer. We can increase the nonlinearity of the network by adding an activation function directly behind the convolutional layer. This kernel can also remain correspondences between individual pixels for later convolutional layers. 3×3 and 5×5 convolutions can cover larger correlated units of input image residual so that we can capture traits between adjacent pixel stacks. For the subsequent network to accept input residuals with different ranges, the feature maps produced by the paralleling three convolutional kernels are concatenated. The output filter banks cascaded into a single output vector forming the input of the next stage.

3.2. The subsequent architectures of our network

We apply three residual blocks to receive the outputs of the multi-kernel noise extractor respectively. The residual block was proposed

by He et al. [28]. It has been widely used in many fields and achieved better performances. The residual block is implemented by 1×1 convolution, Batch-Normalization (BN) [29], ReLu, 3×3 convolution, BN, ReLu, 1×1 convolution, BN and ReLu in turn. Among them, ReLu is used to increase the nonlinearities of the network while the BN is aimed at preventing over-fitting. The numbers of feature maps of convolutional operation in the residual block are 64, 256, 256, respectively. To extract more abstract and larger features at higher levels, it is necessary to reduce their spatial aggregation. Thus, we increase the size of the convolution kernel in the middle of the third residual network from 3 to 5. The outputs of this residual block are fed into an average pooling layer, then two inner product layers are successively connected behind this pooling layer. The first one with 256 output neurons produces a 256-dimensional feature vector. The final $256 \times N$ inner product layer, where N is the number of training classes. When the features are transferred to the fully connected layer, we no longer need to increase the nonlinearity of the network through the ReLU activation function. But we hope that there can be a function to realize the classification. Softmax can map the output of multiple neurons into the (0,1), which can be understood as a probability, so as to carry out multi-classification. So we directly connect the softmax function as a classification tool behind the last inner product layers.

4. Experimental results

We conduct a set of experiments to verify the effectiveness of the multi-kernel noise extractor firstly. Then, through a series of camera model identification experiments assuming that the images have experienced single post-processing, it is proved that our proposed method can achieve excellent performance in this scenario. Besides, images are also processed by both non-influential and influential operations to show the robustness of our camera model identification method.

4.1. Experiment setting

We first download Dresden database [30] which includes more than 16,000 images captured by 27 different models. Then, 18 types of cameras with more than 500 natural images are chosen. Since the two camera models Nikon D70 and Nikon D70s are considered as a single model [31], we will get 17 types of cameras finally. The correspondences between numbers and camera models are shown in Table 1. Five 227×227 patches are randomly cut out from each image without overlapping. Eventually, we get 45,000 patches and divide the training set and testing set into a ratio of 4 to 1. Note that the images from the same source will not appear in both the training and testing set. The overall architecture was learned through Stochastic Gradient Descent on batches of 64 patches. While momentum is fixed to 0.9. The learning rate is initialized to 0.001 and scheduled to be multiplied by 20% for every 10 epochs. Furthermore, in order to evaluate the effectiveness of the proposed method, we trained two CNN [10,13] on the same database to compare with our model.

4.2. The multi-kernel noise extractor performance analysis

In order to show the role of the noise extractor more intuitively, we perform an experiment to compare the performance of the proposed network architecture with the same network without the multi-kernel noise extractor in identifying the camera model of the original images. The confusion matrices of the identification results are shown in Tables 2 and 3. It can be seen that for the cameras with lower detection efficiency, different models of cameras produced by the same manufacturer are usually misclassified. Due to the consistency of the manufacturers, different cameras under the same brand may have some similar hardware, which makes the imaging equipment leave similar traces in the image. This phenomenon leads to misjudgment of source identification. Comparing Tables 2 and 3, we observe that the accuracy of using the multi-kernel noise extractor is higher than that of not using it, indicating that the multi-kernel noise extractor is effective in detecting the image source.

Table 1
Correspondence table of number and camera models.

Number	Camera model	Number	Camera model
1	Canon_Ixus70	10	Pentax_OptioA
2	Casio_EX-Z	11	Ricoh_GX100
3	FujiFilm_FinePixJ5	12	Rollei_RCP
4	Kodak_M1063	13	Samsung_L74wide
5	Nikon_CoolPixS7	14	Samsung_NV15
6	Nikon_D70	15	Sony_DSC-H50
7	Nikon_D200	16	Sony_DSC-D77
8	Olympus_mju	17	Praktica_DCZ
9	Panasonic_DMC		

For the images that undergo post-processing operations, we compare the source identification results of contrast-enhanced images and JPEG compressed images under four different CNN models to explain the performance of the multi-kernel noise extractor. That is, the Tuama et al.'s method [13] with multi-kernel noise extractor, the Tuama et al.'s method [13], our proposed network without multi-kernel noise extractor and our proposed method. Tables 4 and 5 provide the source identification results.

From Table 4, it can be seen that our proposed method containing the multi-kernel noise extractor shows obvious advantages compared to the network without the multi-kernel noise extractor. The average accuracy of source detection for contrast-enhanced images has been increased by 2.0%. It is worth noting that even if the proposed network does not use the multi-kernel noise extractor, the detection performance is still better than that of the network [13]. When applying our multi-kernel noise extractor to the compared network [13], we can find that the accuracy dropped. This is because the multi-kernel noise extractor is designed for our specific problem, rather than a general network layer. It needs to be combined with the network structure we proposed to effectively detect the source of the images.

Similarly, for the problem of detecting the source of JPEG compressed images, we can observe from Table 5 that using the multi-kernel noise extractor can significantly improve the detection effect. What is more, compared with the comparison method [13], our method with and without multi-kernel noise extractor has improved accuracy by 9.0% and 1.9%, respectively. The experimental results show that the multi-kernel noise extractor, through the combination of high-pass filters and convolution kernels of different sizes, further realizes the efficient extraction of the camera fingerprints contained in the image and improves the overall detection efficiency.

When comparing Tables 4 and 5, it can be also observed that the multi-kernel noise extraction has different effects on the source detection of images undergoing different post-processing. The reasons for the above results are as follows.

(1) Because PRNU is utilized in our work to identify the source of post-processed images, which is multiplicative noise and belongs to high-frequency signal, applying a high-pass filter in the multi-kernel noise extractor can retain high-frequency information and suppress low-frequency noise information caused by post-processing operations. For the images altered by non-influential operations, such as contrast enhancement, the PRNU features are not significantly different from those of the original images. Hence, the effect of the high-pass filter on the post-processed image and the original image is similar. For the images altered by influential operations, such as JPEG compress, the PRNU features are weakened by traces of post-processing operations. The use of the high-pass filter can effectively reduce the interference of post-processing traces. Therefore, with the help of the multi-kernel noise extractor, the source identification performance can be improved.

(2) In the multi-kernel noise extractor, we then use three parallel convolution filters of different sizes to obtain multi-scale forensics features. Convolution kernels of different sizes can extract different receptive fields, which can capture features between adjacent

Table 2
The confusion matrix (in percentage points %) of the original image source identification results in the network without noise extractor. The total accuracy is 97.2%. "-" means zero.

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	98.4	–	–	–	–	0.2	–	–	–	–	–	–	0.4	–	–	1	–
2	–	100	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3	–	–	97.6	–	1.4	–	–	–	–	0.2	–	–	–	0.2	–	–	0.6
4	–	–	–	98	–	–	–	–	–	–	–	–	–	2	–	–	–
5	–	–	–	–	95.2	2.2	0.8	–	–	–	–	–	–	–	1.8	–	–
6	–	–	–	–	2	97.6	0.4	–	–	–	–	–	–	–	–	–	–
7	–	0.2	0.2	–	0.8	4.2	92.8	–	–	0.6	–	–	1.2	–	–	–	–
8	–	–	1.6	–	–	–	–	97.4	–	0.6	–	0.4	–	–	–	–	–
9	1	–	–	–	0.4	–	–	–	98.2	–	–	0.4	–	–	–	–	–
10	–	–	–	–	–	–	–	–	–	98.8	–	–	0.6	0.4	–	–	0.2
11	–	–	–	–	0.4	–	–	–	–	–	99.6	–	–	–	–	–	–
12	–	–	–	–	–	–	0.6	–	–	–	–	99.4	–	–	–	–	–
13	0.6	–	–	–	–	–	–	–	0.2	–	–	–	97.8	1.4	–	–	–
14	–	–	–	–	–	–	–	0.4	–	–	–	–	3.4	96.2	–	–	–
15	2.2	1	–	–	–	–	–	–	–	–	–	1.4	–	–	88.8	5.8	0.8
16	–	–	–	–	–	–	–	–	–	–	0.2	–	–	–	0.2	99.6	–
17	–	–	–	1.4	–	–	–	–	–	–	–	1	–	–	–	–	97.6

Table 3
The confusion matrix (in percentage points %) of the original images source identification results in the proposed network. The total accuracy is 98.1%. "-" means zero.

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	99.4	–	–	–	–	–	–	–	0.2	–	–	–	–	–	–	0.4	–
2	–	99.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.8
3	0.8	–	98.4	–	–	–	–	–	–	–	–	0.8	–	–	–	–	–
4	–	–	–	98.8	–	–	–	–	0.2	–	1	–	–	–	–	–	–
5	–	–	–	–	97.2	0.8	1.4	–	–	–	–	–	–	–	–	0.6	–
6	–	–	–	–	1.6	94.8	3.6	–	–	–	–	–	–	–	–	–	–
7	–	–	–	–	–	1.2	98.4	–	–	0.4	–	–	–	–	–	–	–
8	–	–	–	–	–	–	–	99.6	–	0.4	–	–	–	–	–	–	–
9	–	–	–	–	0.2	–	–	–	99.8	–	–	–	–	–	–	–	–
10	–	–	–	–	–	–	0.6	–	–	97.4	–	–	–	–	–	2	–
11	–	–	0.4	–	–	–	–	–	–	–	99.6	–	–	–	–	–	–
12	–	–	–	0.2	–	–	1	–	–	–	–	98.8	–	–	–	–	–
13	–	–	–	–	–	–	0.2	–	–	–	–	–	95.6	4.2	–	–	–
14	–	–	–	–	1.2	–	–	0.8	–	–	–	–	4.6	93.4	–	–	–
15	–	–	–	–	–	–	–	–	–	–	–	–	–	–	99.6	0.4	–
16	–	–	–	–	0.2	–	–	–	–	–	–	–	–	–	0.6	99.2	–
17	0.4	–	–	–	–	–	–	–	–	–	–	1.4	–	–	–	–	98.2

Table 4
The effect (in percentage points %) of multi-kernel noise extractor on source identification in contrast-enhanced images when using different identification methods.

Parameter	0.6	0.8	1.2	1.4	Average
Tuama et al.'s method [13] with multi-kernel noise extractor	52.5	60.0	55.0	47.5	53.8
Tuama et al.'s method [13]	94.3	91.2	93.6	92.4	92.9
The proposed network without multi-kernel noise extractor	96.5	94.9	95.0	94.2	95.2
The proposed method	96.5	98.1	96.9	97.3	97.2

Table 5
The effect (in percentage points %) of multi-kernel noise extractor on source identification in JPEG compressed images when using different identification methods.

Parameter	70	80	90	Average
Tuama et al.'s method [13] with multi-kernel noise extractor	40.0	55.0	45.0	46.7
Tuama et al.'s method [13]	74.3	77.3	83.2	78.3
The proposed network without multi-kernel noise extractor	75.5	78.1	86.9	80.2
The proposed method	85.4	84.8	91.6	87.3

pixel stacks of different units. Based on the integration of multi-scale forensics features, the information utilization rate is improved, and the post-processed image source identification performance can be improved.

(3) By using the multi-kernel noise extraction, the low-level image noise patterns can be learned, which is useful to detect camera models. Meanwhile, these noise patterns are received by three residual blocks to extract more abstract and larger features at higher levels. By combining three residual blocks, spatial aggregation can be reduced, thereby reducing the loss of forensics information, which helps improve source identification performance.

4.3. The source identification of images tampered by single tampering operation

Four editing operations, contrast enhancement, resizing, median filtering and JPEG compressed, are performed on our unaltered database. Among them, contrast enhancement and resizing belong to linear tampering operation, median filtering and JPEG compressed belong to non-linear tampering operation.

Tables 6 and 7 show the source identification results of contrast-enhanced images and resized images under different CNN models. In Table 6, we consider the contrast enhancement as a processing operation with four different factors, 0.6, 0.8, 1.2 and 1.4. The parameter of the last column is 1.0, which represents the original image. It can be observed that the performances of networks under different tampering parameters are similar. In Table 7, the range of resizing strengths are

Table 6
Average accuracies of image source identification under contrast enhancement in different identification methods.

Parameter	0.6	0.8	1.2	1.4	1 (original image)
Proposed method	96.5%	98.1%	96.9%	97.3%	98.7%
Bondi et al.'s method [10]	94.1%	92.2%	94.9%	95.5%	95.7%
Tuama et al.'s method [13]	94.3%	91.2%	93.6%	92.4%	95.3%

Table 7
Average accuracies of image source identification under resizing in different identification methods.

Parameter	0.6	0.8	1.2	1.4
Proposed method	93.7%	94.1%	95.2%	94.3%
Bondi et al.'s method [10]	89.3%	90.2%	91.6%	89.4%
Tuama et al.'s method [13]	88.3%	86.2%	89.6%	87.4%

Table 8
Average accuracies of image source identification under median filtering in different identification methods.

Parameter	3×3	5×5	7×7
Proposed method	84.4%	72.1%	69.2%
Bondi et al.'s method [10]	65.3%	58.3%	57.2%
Tuama et al.'s method [13]	61.5%	52.9%	48.5%

Table 9
Average accuracies of image source identification under JPEG compressed in different identification methods.

Parameter	70	80	90
Proposed method	85.4%	84.8%	91.6%
Bondi et al.'s method [10]	76.9%	75.1%	87.9%
Tuama et al.'s method [13]	74.3%	77.3%	83.2%

from 0.6 to 1.4, and the interval is 0.2. It is shown that the fluctuations of accuracy in the same row do not exceed 4%. These experimental results validate that the linear tampering operation affects the result of classification slightly. We can achieve high accuracy of tampered image source identification as the original image. Compared with the other two existing methods, the average accuracies of image source identification are increased by 4.5%.

Besides, our method shows obvious superiority when detecting images processed by an influential operation. Tables 8 and 9 show the results of median filtering and JPEG compressed. It can be seen that, compared to original images, the detection accuracy of images altered by an influential operation is substantially reduced under Bondi et al.'s method [10] and Tuama et al.'s method [13]. That is, nonlinear tampering operation will result in low accuracy of image source recognition, while our proposed network could obtain high accuracy of camera model identification. Specifically, we can achieve a 75.9% average identification rate with median filtered images and 87.2% with JPEG compressed images. Our method improves performance by nearly 7.23% and 8.93% compared to the other two methods [10,13] respectively.

Experimental results show that our proposed method is evidently robust to tampered image source identification, particularly when images are modified by influential tampering operations. Feature maps produced by the multi-kernel noise extractor can typically improve the overall identification rate with all possible tampering operations.

4.4. The source identification of images tampered by the combination of non-influential and influential operations

In these experiments, we assess the performance of our source identification method in a more complex scenario where images are modified by an operation chain consists of two tampering operations. Specifically, the operation chain contains a non-influential operation and influential operation. The combinations of two tampering operations we choose are JPEG compressed and contrast enhancement, median filtering and contrast enhancement. In this case, the non-influential operation refers to contrast enhancement which does not affect the accuracy of source recognition. Then, our network is applied to identify the image manipulated by JPEG compressed and contrast enhancement, median filtering and contrast enhancement respectively.

Table 10
Average accuracies of image source identification under median filtering and contrast enhancement.

Parameter	0.6	0.8	1.2	1.4
3×3	83.9%	83.2%	82.8%	83.3%
5×5	71.3%	70.2%	72.4%	71.8%
7×7	66.3%	68.1%	69.1%	67.8%

Table 11
Average accuracies of image source identification under JPEG compressed and contrast enhancement.

Parameter	0.6	0.8	1.2	1.4
70	85.9%	84.2%	86.8%	83.3%
80	83.1%	87.0%	85.4%	84.6%
90	91.3%	89.7%	92.2%	90.0%

In Tables 10 and 11, the first row lists the parameters of contrast enhancement, and the first column lists the parameters of median filtering and JPEG compressed. When the parameters of median filtering or JPEG compressed are fixed, the identification rates fluctuate slightly between different contrast enhancement factors and the average accuracies are similar to images only processed by median filtering or JPEG compressed. The results show that our source identification approach is robust to the chain consists of two tampering operations.

5. Discussions

When detecting an image in the real world, we usually do not know the post-processing operations it has undergone. Because different post-processing operations will leave different processing traces in the image, it would become more difficult for a forensic investigator to determine the image source when diverse post-processing operations are applied.

For example, if the model trained with contrast-enhanced images is utilized to detect images that have not undergone post-processing, the source of the images can still be effectively identified. The reason is that the contrast enhancement operation changes the source noise contained in the image slightly, so there are some similar camera fingerprints between the contrast-enhanced images and the original images. Utilizing the model generated by the contrast-enhanced images can also obtain the source features from the original image, so the identification performance is better. However, for tampered images with other post-processing operations, such as JPEG compression, resizing and median filtering, the detection efficiency will be significantly reduced.

When the network model trained with contrast-enhanced images is used to identify the source of the images processed after JPEG compression, JPEG compression has a different effect on image noise than contrast enhancement. Therefore, the network model cannot extract high-quality features, and the detection rate will be greatly reduced. When applying this model to identify the source of the resized images, because reducing the image size will lose part of the image pixel information, and expanding the image size will add part of the pixel information to the image. Both of these situations will have a different effect on the camera fingerprint in the image than the contrast enhancement, resulting in lower detection accuracy. And when detecting the median filtered images, due to highly non-linear, median filtering operation will destroy linear correlations among neighboring pixels thus leading to different processing traces from contrast enhancement.

Thus, when the contrast enhancement model is used to detect the source of the median filtered images, the accuracy will be poor.

We have to admit that our paper only focuses on source identification when the kinds of manipulations are known to a forensic detector. In the future, we will improve our network and enhance its robustness in the real scenario where the tampering operations experienced by the image are unknown.

6. Conclusions

Source camera identification is still a real issue in the field of image forensics, especially in the case of the investigated images are modified by some operations. In this paper, we propose a robust network to achieve post-processed image source identification based on image noises. The main contributions of the paper are as follows:

(1) By analyzing post-processed image noises, tampering operations can be divided into two classes: non-influential operation and influential operation. Moreover, we found that in the issue of source forensics, the images altered by both non-influential and influential operations can be regarded as the images only modified by influential operations.

(2) A multi-kernel noise extractor is presented to capture camera fingerprints. This layer consists of a high-pass filter to extract noise residual and three paralleling convolution filters to extract forensics features from noise maps in different size kernels.

(3) Experimental results demonstrate that our proposed source identification method is robust to a single tampering operation and two tampering operations chain. Compared with the existing works, our network shows obvious superiority.

In the future, we will try to present a strengthened network to detect the source of images tampered with two different influential operations. Meanwhile, some open-set forensics scenarios, such as the tampering operations experienced by the images are unknown, are also under consideration to make our research more realistic.

CRediT authorship contribution statement

Xin Liao: Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision, Funding acquisition. **Jing Chen:** Conceptualization, Experimental design and Data collection, Data interpretation, Writing – original draft. **Jiaxin Chen:** Conceptualization, Experimental design, Data interpretation, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant Nos. 61972142, 61772191), Hunan Provincial Natural Science Foundation of China (Grant No. 2020JJ4212), Key Lab of Information Network Security, Ministry of Public Security(Grant No. C20611), Key Lab of Forensic Science, Academy of Forensic Science, Ministry of Justice, China (Grant No. KF202118), Science and Technology Program of Changsha, China (Grant No. kq2004021).

References

[1] M. Kirchner, T. Gloe, Forensic camera model identification, in: Handbook of Digital Forensics of Multimedia Data and Devices, 2015.
[2] J. Lukas, J. Fridrich, M. Goljan, Digital camera identification from sensor pattern noise, IEEE Trans. Inf. Forensics Secur. 1 (2) (2006) 205–214.
[3] A.C. Popescu, H. Farid, Exposing digital forgeries in color filter array interpolated images, IEEE Trans. Signal Process. 53 (10) (2005) 3948–3959.
[4] Y. Long, Y. Huang, Image based source camera identification using demosaicking, in: IEEE Workshop on Multimedia Signal Processing, 2006, pp. 419–424.
[5] A. Swaminathan, M. Wu, K.J.R. Liu, Nonintrusive component forensics of visual sensors using output images, IEEE Trans. Inf. Forensics Secur. 2 (1) (2007) 91–106.
[6] C. Chen, M.C. Stamm, Camera model identification framework using an ensemble of demosaicing features, in: IEEE International Workshop on Information Forensics and Security, 2015, pp. 1–6.
[7] Z. Deng, A. Gijsenij, J. Zhang, Source camera identification using auto-white balance approximation, in: International Conference on ComputerVision, 2011, pp. 57–64.
[8] M. Kharrazi, H.T. Sencar, N. Memon, Blind source camera identification, in: IEEE International Conference on Image Processing, Vol. 1, 2004, pp. 709–712.
[9] F. Meng, X. Kong, X. You, A new feature based method for source camera identification, in: Advances in Digital Forensics IV, IFIP International Federation for Information Processing, 2008, pp. 207–218.
[10] L. Bondi, L. Baroffio, D. Guera, P. Bestagini, E.J. Delp, S. Tubaro, First steps toward camera model identification with convolutional neural networks, IEEE Signal Process. Lett. (2017) 259–263.
[11] P. Yang, R. Ni, Y. Zhao, W. Zhao, Source camera identification based on content-adaptive fusion residual networks, Pattern Recognit. Lett. (2017) 1–10.
[12] D. Guera, K. Yarlagadda, Reliability map estimation for CNN-Based camera model attribution, in: Winter Conference on Applications of Computer Vision, 2018, pp. 964–973.
[13] A. Tuama, F. Comby, M. Chaumont, Camera Model Identification With The Use of Deep Convolutional Neural Networks, in: IEEE International Workshop on Information Forensics and Security, 2016.
[14] B. Bayar, M.C. Stamm, A deep learning approach to universal image manipulation detection using a new convolutional layer, in: ACM Workshop on Information Hiding and Multimedia Security, 2016, pp. 5–10.
[15] J.Y. Sun, S.W. Kim, S.W. Lee, S.-J. KO, A novel contrast enhancement forensics based on convolutional neural networks, Signal Process., Image Commun. 63 (2018) 149–160.
[16] B. Bayar, M.C. Stamm, Augmented convolutional feature maps for robust CNN-based camera model identification, in: International Conference on Image Processing, 2017, pp. 4098–4102.
[17] L. Bondi, D. Guera, L. Baroffio, P. Bestagini, E.J. Delp, S. Tubaro, A preliminary study on convolutional neural networks for camera model identification, in: Media Watermarking, Security, and Forensics, 2017, pp. 67–76.
[18] D. Freire, F. Narducci, Deep learning for source camera identification on mobile devices, Pattern Recognit. Lett. (2018) 1–6.
[19] B. Bayar, M.C. Stamm, Design principles of convolutional neural networks for multimedia forensics, Electron. Imaging (2017) 77–86.
[20] A. Eleni, G. Zeno, V.E. Erwin, Camera recognition with deep learning, Forensic Sciences Research (2018) 210–218.
[21] D. Cozzolino, L. Verdoliva, Camera-based image forgery localization using convolutional neural networks, in: European Signal Processing Conference, 2018, pp. 1372–1376.
[22] Z. Zhen, Camera model identification with convolutional neural networks and image noise pattern, http://hdl.handle.net/2142/100123.
[23] X. Liao, Zh. Qin, L.P. Ding, Data embedding in digital images using critical functions, Signal Process., Image Commun. 58 (2017) 146–156.
[24] Z. Chen, Y. Zhao, R. Ni, Detection of operation chain: JPEG-Resampling-JPEG, Signal Process., Image Commun. 57 (2017) 8–20.
[25] W.J. Zeng, H. Yu, C.Y. Lin, Multimedia Security Technologies for Digital Rights Management, Academic Press, 2006, pp. 383–412.
[26] M. Chen, J. Fridrich, M. Goljan, Digital imaging sensor identification(further study), in: Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, 2007, pp. 16–28.
[27] Y. Qian, J. Dong, W. Wang, T. Tan, Deep learning for steganalysis via convolutional neural networks, in: Proc. SPIE, Vol. 9409, 2015, pp. 94090J-94090J-10.
[28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[29] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, pp. 448–456.
[30] T. Gloe, R. Bohme, The dresden image database for benchmarking digital image forensics, J. Digit. Forensic Pract. (2010).
[31] M. Kirchner, T. Gloe, Forensic camera model identification, in: Handbook of Digital Forensics of Multimedia Data and Devices Chichester, 2015, pp. 329–374.