Pattern Recognition Letters 000 (2017) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Source camera identification based on content-adaptive fusion residual networks

Pengpeng Yang[a,b], Rongrong Ni[a,b,*], Yao Zhao[a,b], Wei Zhao[a,b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Source camera identification
Content-adaptive fusion residual networks
Transform learning

ABSTRACT

Source camera identification is still a hard task in forensics community, especially for the case of the query images with small size. In this paper, we propose a solution to identify the source camera of the small-size images: content-adaptive fusion residual networks. According to the differences of the image contents, firstly, the images are divided into three subsets: saturation, smoothness and others. Then, we train three fusion residual networks for saturated images, smooth images, and others, separately, through transform learning. The fusion residual networks is formed with three paralleled residual networks and the difference of three residual networks lies in the convolutional kernel size of preprocessing layer. The features learned from the last residual blocks of three residual networks are fused and fed into softmax classifier. In particular, the residual networks is designed to learn better feature representation from the input data. The convolutional operation is added in preprocessing stage and three residual blocks are used. The experiment results show that the proposed method has satisfactory performances at three levels of source camera identification: brand level, model level, and device level.

1. Introduction

With the development of science and technology, image acquisition devices are becoming more and more pervasive. At the same time, image editing tools are becoming common and anyone can easily modify the images. The origin and the integrity of the images need to be verified reliably. Image forensics are essential techniques to prevent malicious tampering to the images for illegal benefits. One of the hot topics in multimedia forensics is source camera identification, the purpose of which is to trace where an image is from. Identifying the source camera is an important step in pointing out the owner of illicit images (e.g. crime scenes, terroristic act scenes, etc) and ensuring the security and trustworthiness of such digital data [1].

A series of operations inside the camera would be carried out when a digital image is captured. These processes could leave some inherent traces into the image, such as lens aberration [2,3], defective pixels [4,5], CFA interpolation artifacts [6], JPEG compression [7,8], image quality evaluation index and high order statistics in wavelet domain [9,10] or Sensor Pattern Nosie(SPN) [11–16]. Sensor Pattern Noise (SPN) generated by digital cameras has drawn more attention due to its uniqueness. The SPN can't be affected by the environment and arises primarily from the manufacturing imperfections and the inhomogeneity of silicon wafers. In general, two things need to be considered for these methods based on SPN. Firstly, the quality of SPN extracted from image depends on the image contents. Secondly, the detection performance could be decreased with the reduction of image size. The analysis of the small-size images provide an effective reference for the splicing forgeries operated between images coming from different cameras [17]. So it is necessary to identify the source camera of the images with small-size.

The convolutional neural network(CNN) has recently achieved better performance than traditional schemes in digital image forensics [18–22]. There are two common characters for these algorithms. Firstly, considering the different tasks between computer vision and image forensics, the researches focusing on image forensics usually add preprocessing operations into convolutional neural network architecture, which can amplify the inter-class difference and reduce the impact of the image contents. For example, median filter, Laplacian filter, and high-pass filter are applied in median filtering forensics, recapture forensics, and source camera identification, respectively. Secondly, according to the reports in these works, the convolutional neural networks is suit for dealing with small-size images and has better detection performance than the traditional schemes.

[*] Corresponding author at: Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China.
E-mail address: rrni@bjtu.edu.cn (R. Ni).

https://doi.org/10.1016/j.patrec.2017.10.016
0167-8655/© 2017 Published by Elsevier B.V.

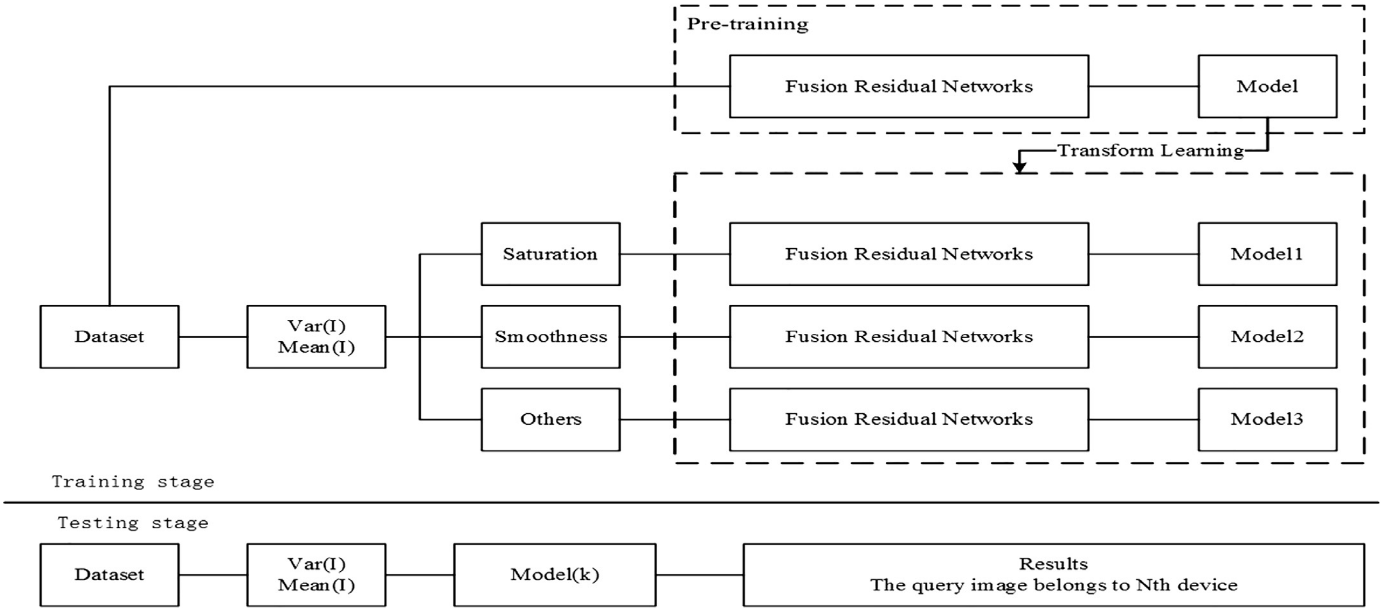


Fig. 1. The framework of proposed method

In this paper, we propose a content-adaptive fusion residual networks to achieve the source camera identification for small-size images. Firstly, we divide the images into three subsets: saturation, smoothness, and others, according to the image contents. Then the residual networks is constructed to extract the effective features. In particular, in order to improve performance of residual networks for source camera identification, the preprocessing layer is added into its architecture. The convolutional operation is implemented in preprocessing stage and the parameters of convolutional kernel need to be self-learned from the input data, which makes the convolutional neural networks self-learn the suitable convolutional kernels from the input data. Next, in order to capture more comprehensive information of the images, three residual networks are paralleled together to construct the fusion residual networks (FRN). The difference among three FRNs is the convolutional kernel size in the pre-procession. In addition, transform learning is used to deal with the limited training data. The effectiveness of proposed method is validated in the experiments and the detection performance of proposed method on four kinds of cases are discussed: camera brand identification, camera model identification, camera device identification, and source camera identification fusing different brands, models, and devices. The experimental results show that the proposed method is practicable and satisfactory.

The rest of this paper is organized as follows. Section 2 describes the related work; Section 3 presents the details of the algorithm proposed in this work; Section 4 includes the experimental results; conclusions are given in Section 5.

2. Source camera identification based on convolutional neural networks

The convolutional neural network(CNN) has recently achieved better performance than traditional schemes on camera device forensics. Differing from the CNN used in computer vision field, a preprocessing layer is added to the CNN architecture for image forensics. In the work [21], an algorithm based on convolutional neural network is proposed for source camera identification and two types of residuals in preprocessing stage are evaluated: high-pass filtering residual and the residual noise extracted by subtracting the de-noised version of the image from the image itself. The two residuals are shown in the following formulas, respectively.

$$R_1 = I - WF(I) \tag{1}$$

$$R_2 = I * \frac{1}{12}\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix} \tag{2}$$

Where I represents the input image; WF(I) is the de-noised image as descripted in Lukas' work [23]; * means the convolution. As reported in this work, adding high-pass filter in the pre-procession into the CNN architecture has better detection performance.

In our previous work [22], we proposed the CAF-CNN to identify the source camera. The CAF-CNN has better detection performance than the algorithms based on handcrafted features and the method in the work [21] in the case of the small-size images. The three CA-CNNs were constructed and paralleled to extract the multi-scale features. There are seven layers in the CA-CNN: a preprocessing layer, five convolutional layers, and a softmax layer.The self-learn convolutional kernel is put into preprocessing layer. The convolutional layer contains four operations: convolution, Batch-Normalization, ReLu, and average pooling. The numbers of feature maps in five convolutional layers are 8, 16, 32, 64, and 128, respectively. In order to avoid overfitting, we applied global average pooling to the last pooling layer and directly fed the output of global average pooling into softmax layer. What's more, the Batch-Normalization layer is used. It has been proved that it is an effective mode to accelerate convergence.

3. Proposed algorithm

The framework of proposed method is shown in Fig. 1, which includes two part: training and testing stage. For training stage, three steps are executed in order. Firstly, the pre-training is done as shown in the dashed frame at the top right-hand corner. For pre-training stage, all images from the training dataset are used to train a fusion residual network in end-to-end way whose architecture is presented in Fig. 2, which get a preliminary model. Secondly, the dataset for training is divided into three subset, according to the mean (m) and variance (v) of the images, as shown

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

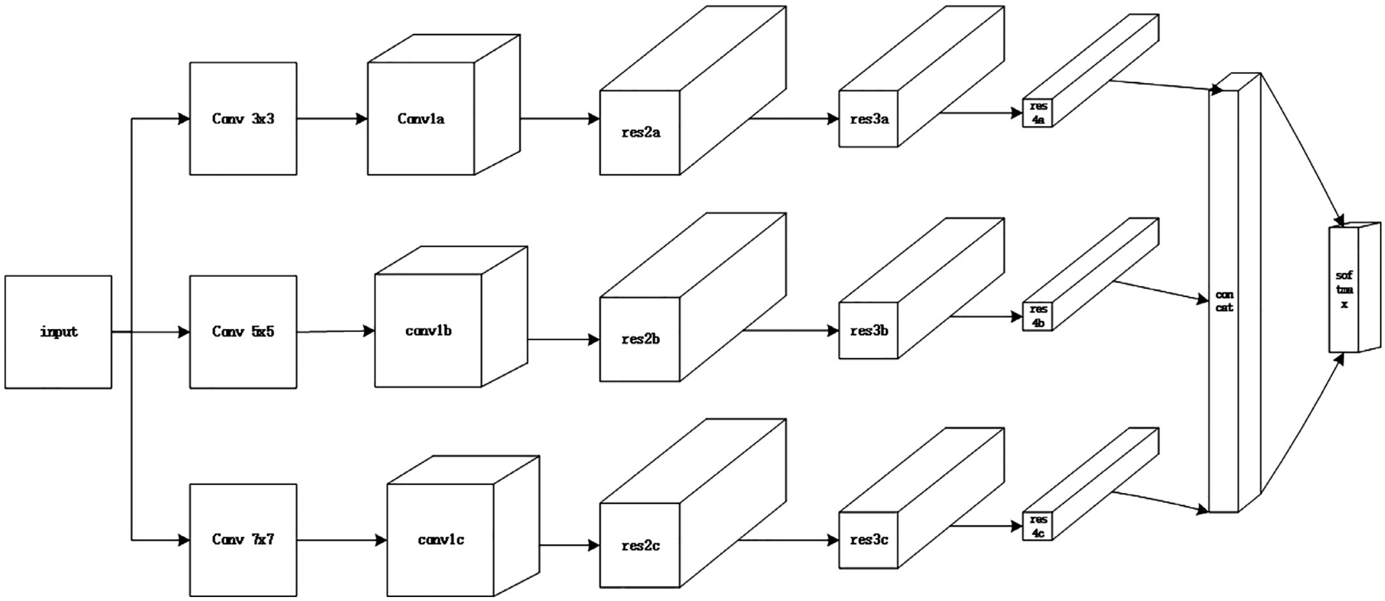
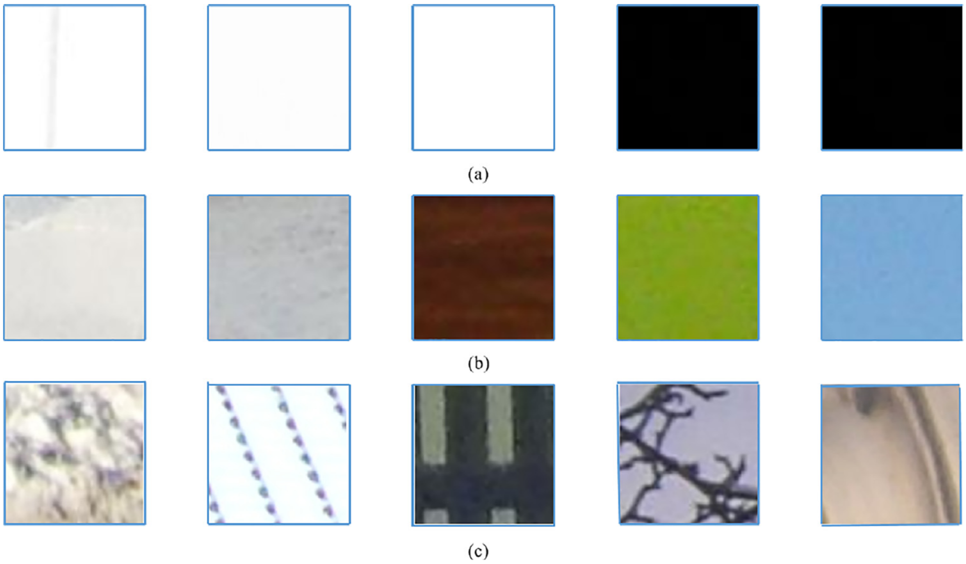


Fig. 2. The architecture of fusion residual networks.


(a)

(b)

(c)

Fig. 3. Examples from three subsets. (a), (b), (c) represent the saturated (T_1), smooth (T_2), and other (T_3) images, respectively.

in Eqs. (4) and (5). Some examples from three subsets are given in Fig. 3. Thirdly, for three subsets, we train three fusion residual networks by transfer learning in end-to-end way respectively. The model trained in the first step is applied into the model training in the third step by transfer learning. Some researches has proved that transfer learning has a good effect on the model performance. All in all, during the training stage, we will get three fusion residual models for three subsets. In the testing stage, first of all, the mean and variance of the given query image is calculated. Then, according to the rules in formulas (4) and (5), we know which subset the query images belongs to and the image is feed into the model trained for the same subset to get the final classification.

3.1. Content-adaptive analysis

The transformation from the real scene to the digital image in cameras would go through four stages: generator, demoasic, processor, and JPEG compression, in order, as shown in Fig. 4. The RAW image come into being after the generator. The generator contains four basic components [24]: lenses, optical filter, CFA pattern, and sensor. An important and stable camera fingerprint, sensor pattern noise(SPN), would be emerge in this stage. Then the demosaicing operation is carried out for the RAW image, which would introduce the periodic noise in general. Next, for the better visual effect, the image would be processed by white balancing, denoising, contrast adjusting, and so on. Finally, we get the digital image(JPEG format) after JPEG compressing. According to the above statement, we build the model for the process of generating the digital image:

$$I = J(P(D(G(r)))) \quad (3)$$

where, r represents the continuous-time signal; I denotes the image generated, it is a discrete time signal; G(.) means the operation of the generator; D(.) is the demosaic; P(.) and J(.) denote the operation of the processor and JPEG compression, respectively.

As far as we know, G(.), D(.), and J(.) are related with the image contents [11,25,26]. The G(.) include a special part: sensor pattern noise. It is a unique fingerprint for each camera device. The sensor pattern noise has been proved that it is relevant to the image contents. According to the report in the work [11], the smooth and

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

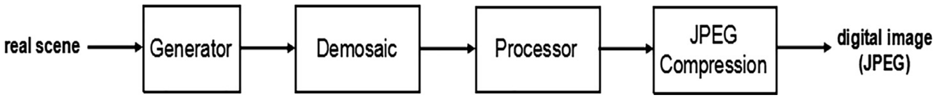


Fig. 4. The process of generating the digital image.

bright images are the better choice, comparing with the saturate images that have the limited information. Besides, considering the better perceptual image quality, the camera manufacturers would process the D(.) and J(.) according to the image contents. Therefore, the fingerprint left should be not same for the different image contents. For different image contents, the difficulty level of source camera identification should be vary. So we separated the database into three subsets according to the degree of detection difficulty: smoothness, others, and saturation. The division principle is as follows:

$$I_i \in \begin{cases} Subset1 = T_1(m, v) \\ \\ Subset2 = T_2(m, v) \\ \\ Subset3 = T_3(m, v) \end{cases} \tag{4}$$

$$\begin{cases} T_1(m, v) & m \in [0, 5] \bigcup [250, 255], v \in [0, 25] \\ T_2(m, v) & m \in [0, 5] \bigcup [250, 255], v \in [25, 50] \ || \\ & m \in (5, 250), v \in [0, 50] \\ T_3(m, v) & others \end{cases} \tag{5}$$

where, v and m denote the variance of the grayscale image. T mean the sets of the means and the variances. If the variance is bigger than 50 and the mean is bigger than 250 or the mean is smaller than 5, the image will be divided into Subset1. When the variance is less than 25, the image will be divided into Subset2. The others will be put into Subset3. The selection of the thresholds is empirical.

3.2. Content-adaptive fusion residual networks

The different image contents should be handled separately according to the above statement. In order to capture effective features for the different image contents, the fusion residual networks is designed, as shown in Fig. 2. Firstly, we construct the residual networks(RN). The self-learning filter and three residual blocks are used. Then, three residual networks are paralleled to extract the multi-scale features.

3.2.1. Residual networks

According to the image forensics algorithms based on CNN, in order to magnify the inter-class differences, the preprocessing layer is added into the CNN. And the special filter is applied in the preprocessing stage. For example, Laplacian filter and high-pass filter are used for recapture forensics, camera model identification, respectively. However, the special filter used maybe not the best choice for CNN and some important information could be lost. For instance, for source camera identification, the traditional schemes calculate the correlation between SPN and the reference SPN. It is important for detection performance to extract the high-quality SPN. And it has been proved that the SPN is related to the image contents. The researchers try to extract the precise SPN through a variety of processions. Therefore, it is maybe not the best way to preprocessing the input data using high-pass filter for source camera identification based on CNN. Considering that the convolutional neural network can self-learn better feature representations from the input data, in this work, we replace the special filter with a convolutional layer, as is shown in the following formula:

$$R = I * W_{pq} \tag{6}$$

Where, W means the parameters of the convolutional kernel and p, q denote the length and width of the convolutional kernel, respectively. It can be learned from the input data using mini-batch gradient descent. In addition, we apply the residual block to our architecture. The residual block was proposed by He et al. [27]. It has been widely used in many fields and achieved better performances. The architecture of residual networks is shown in Fig. 5.

We compare the self-learning filters of RN with high-pass filter and the results are as shown in Fig. 6. The filters learned from the RGB channel are visual. The high-pass filter is shown in Fig. 6(a) and the others are the convolutional kernels of RN for B, G, R channel, respectively. As can be seen, the self-learning convolutional kernels in the preprocessing stage are different with high-pass filter. The experiment results show that the self-learning convolutional kernels have better performance.

3.2.2. Fusion residual networks

In order to capture more comprehensive features, we parallel three residual networks, as is shown in Fig. 2. The input image is processed in preprocessing layer by three kinds of convolutional kernel size: 3 × 3, 5 × 5, 7 × 7, respectively. Then a group of layers, convolution ,Batch-Normalization, ReLU, and average pooling layer, is used. Then, three residual blocks are applied. The residual block is implemented by 1x1 convolution, Batch-Normalization, ReLu, 3 × 3 convolution, Batch-Normalization, ReLu, 1 × 1 convolution, Batch-Normalization, ReLu, and average pooling operation in order. The numbers of feature maps of convolutional operation in residual block are 64, 256, 256, respectively. In order to reduce the number of model parameters, global average pooling operation is used in res4a, res4b, and res4c layers. The output of global average pooling are put together and then fed into a softmax layer. The calculation of softmax layer is as following:

$$f(y_i) = -log\left(\frac{e^{y_i}}{\sum_{i=1}^{n} e^{y_i}}\right) \tag{7}$$

$$y_i = \sum_{k=1}^{768} W_k^i * X_k + B_i \tag{8}$$

where, i represents the class label, n is the number of the classes; X and k mean the output of global average pooling and its number, respectively; W and B are the weights and biases, which will be learned using mini-batch gradient descent. The parameters of the architecture are shown in Table 1.

4. Experiments

In order to validate the proposed algorithm, we conduct a set of experiments on the Dresden database that provide more than 16,000 images took by 74 camera devices. The images captured by thirteenth devices are chose and the list of thirteenth devices is shown in Table 2. The chose images are cut into non-overlapping 64 × 64 image patches and these image patches construct the dataset used in the experiments. The dataset is split by assigning 4/6 of the images to a training set, 1/6 to a validation set, and 1/6 to a test set, respectively. For the experiment one, the training set includes 2,757,888 image batches. For the experiment two and three, 818,748 image patches are used for the training sets. The learning rate is initialized to 0.01, and scheduled to decrease 10% for every 10,000 iterations. It should be noted that the learning

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

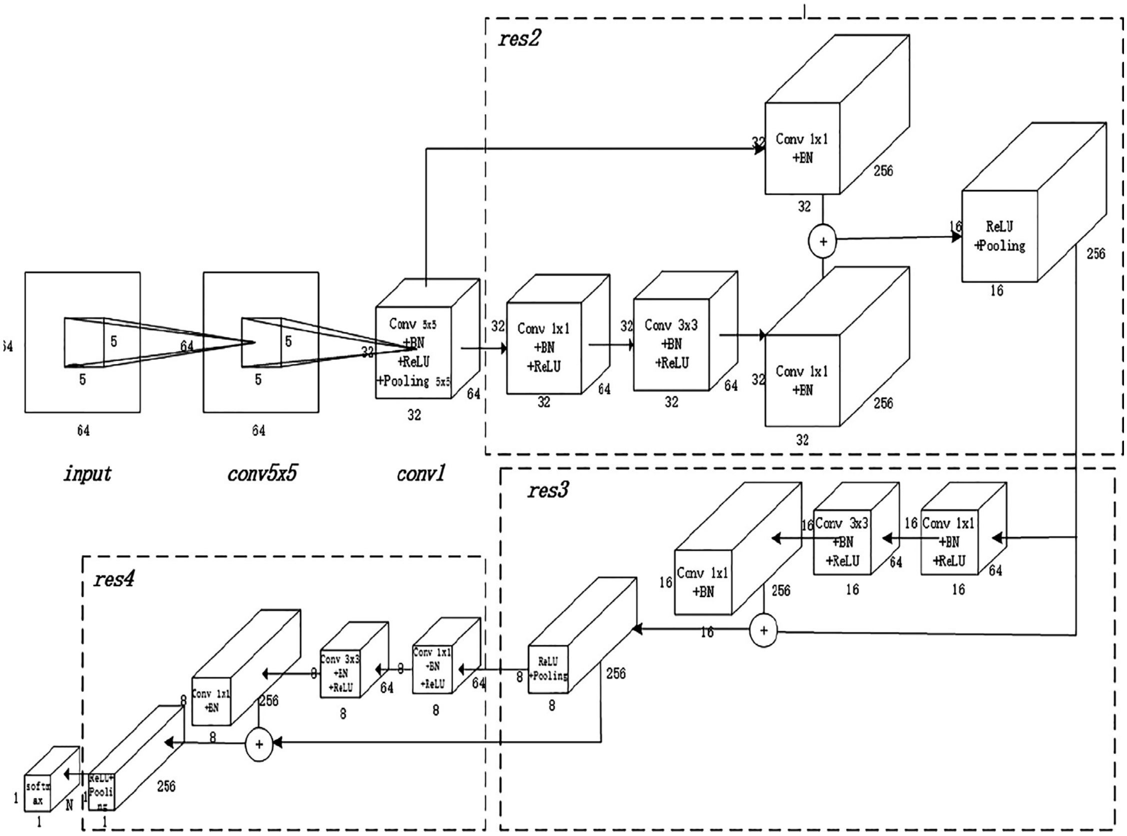


Fig. 5. The architecture of the residual networks.

Table 1
The parameters of the fusion residual networks.

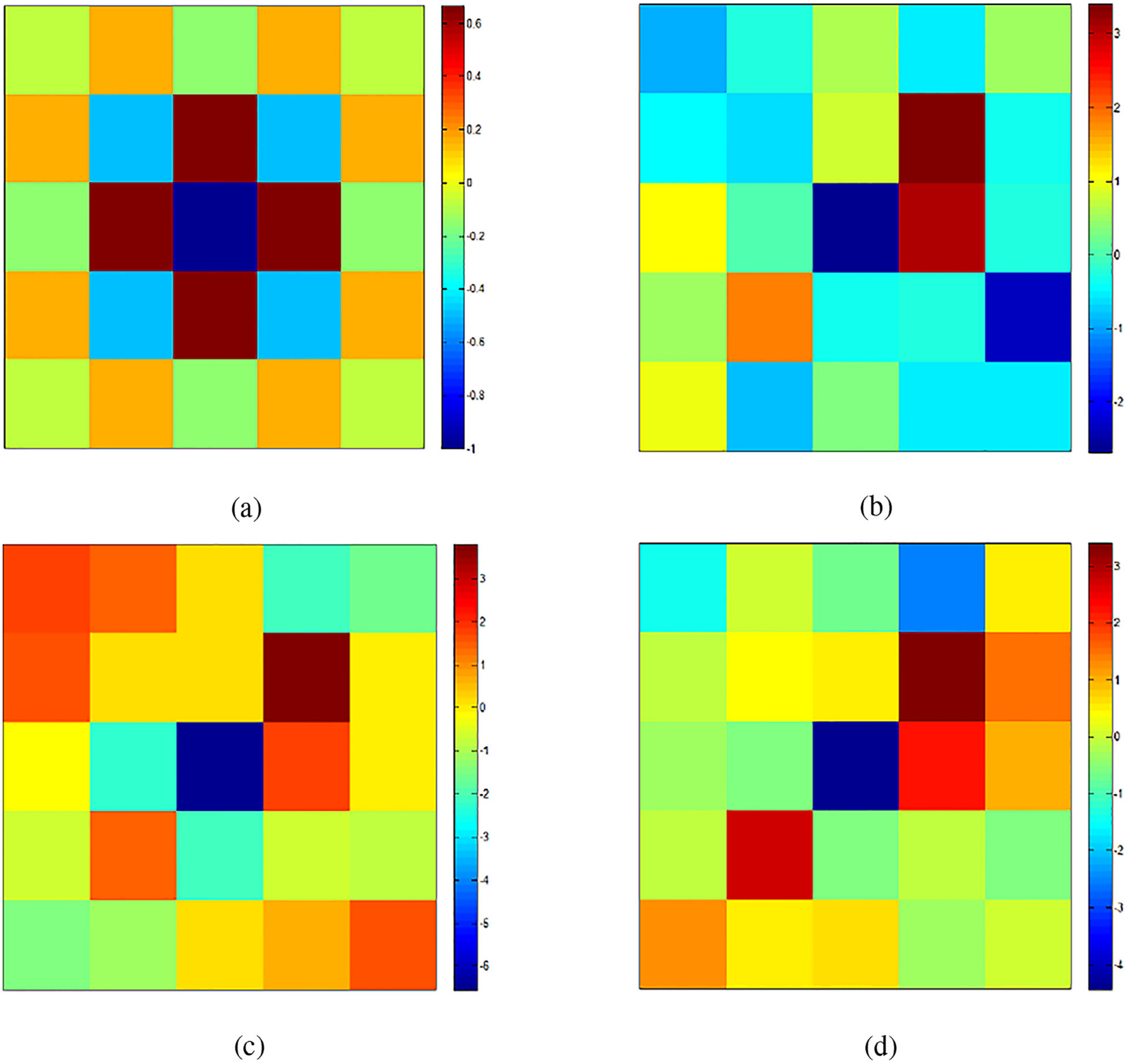
Layername	Parameters		
input	64 × 64 × 3		
conv_	3 × 3 × 1 stride:1	5 × 5 × 1 stride:1	7 × 7 × 1 stride:1
conv1_	5 × 5 × 64 stride:1	5 × 5 × 64 stride:1	5 × 5 × 64 stride:1
ave_pooling	5 × 5 stride:2	5 × 5 stride:2	5 × 5 stride:2
res2_	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1
	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1
	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1
ave_pooling	5 × 5 stride:2	5 × 5 stride:2	5 × 5 stride:2
res3_	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1
	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1
	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1
ave_pooling	5 × 5 stride:2	5 × 5 stride:2	5 × 5 stride:2
res4_	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1	1 × 1 × 64 stride:1
	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1	3 × 3 × 64 stride:1
	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1	1 × 1 × 256 stride:1
global_ave_pooling	8 × 8 stride:1	8 × 8 stride:1	8 × 8 stride:1
softmax		1 × n	

rate is initialized to 0.001 for transfer learning. The max iteration is set to 500,000 and the momentum is fixed to 0.9.

About the weights initialization, the Gaussian filler with expected value $u = 0$ and standard deviation $\delta = 0.01$ is used in the convolutional layers, including the preprocessing layer. And the Xavier filler is applied into softmax layer, which is a common setting for the image forensics algorithms based on Convolutional Neural Networks. We have special consideration for the weights initialization in the preprocessing layer. It is well known that it is important to have good weights initialization for convolutional neural networks and some researches focusing on image forensics have proposed that it is a better choice to initialize weights with high-pass filter for the preprocessing layer. However, as can be seen from our experimental results (Table 3), high-pass filler in the preprocessing layer has worse performance than Gaussian filler. Therefore, in this work, we use the Gaussian filler to initialize weights of the preprocessing layer. The more suitable weights initialization in the preprocessing layer for source camera identification need to further research and exploration.

For each group of experiments, firstly, we pre-train the fusion residual networks in the whole training dataset. Then the training dataset is divided into three subsets and the fusion residual networks for three subsets are trained by transform learning. In the testing stage, the query images are fed into the trained models according to the formulas (4) and (5). The detection accuracies are averaged over 3 random experiments.

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016



(a)

(b)

(c)

(d)

Fig. 6. filter visualization. high-pass filter (a), self-learning convolutional kernels of residual networks for B (b), G (c), R (d), respectively.

Table 2
The list of camera devices used.

ID	Camera devices	Original resolution
1	Kodak_M1063_0	3664 × 2748
2	Pentax_OptioA40_0	4000 × 3000
3	Nikon_CoolPixS710_1	4352 × 3264
4	Sony_DSC-H50_0	3456 × 2592
5	Olympus_mju_1050SW_2	3648 × 2736
6	Panasonic_DMC-FZ50_1	3648 × 2736
7	Agfa_Sensor530s_0	2560 × 1920
8	Ricoh_GX100_0	3648 × 2736
9	Samsung_NV15_0	3648 × 2736
10	Sony_DSC-W170_0	3648 × 2736
11	Sony_DSC-T77_0	3648 × 2736
12	Sony_DSC-T77_1	3648 × 2736
13	Sony_DSC-T77_2	3648 × 2736

Table 3
The detection accuracy for camera brand identification.

Type	Preprocessing	Ave_acc
CA-CNN	HP	81.62%
	Conv 3 × 3	87.72%
	Conv 5 × 5	90.11%
	Conv 7 × 7	90.68%
GoogleNet [21]	HP	91.60%
ResNet [27]	None	96.20%
RN	Conv 3 × 3	95.58%
	Conv 5 × 5	96.21%
	Conv 7 × 7	96.03%
CAF-CNN [22]	Conv3 5 7	94.17%
FRN	Conv3 5 7	96.26%
CA-FRN	Conv3 5 7(Res)	97.03%

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

Table 4
The detection accuracy of content-adaptive fusion residual networks and fusion residual networks. The best results are highlighted in bold.

	Smoothness		Saturation		Others	
	FRN	CA-FRN	FRN	CA-FRN	FRN	CA-FRN
1	99.27%	99.09%	84.16%	80.83%	99.13%	99.57%
2	99.03%	99.76%	27.33%	42.55%	97.90%	99.46%
3	93.99%	97.51%	97.66%	93.92%	96.16%	97.49%
4	96.07%	98.02%	51.44%	66.93%	96.82%	95.11%
5	91.14%	89.95%	61.05%	61.18%	96.38%	97.65%
6	94.25%	96.27%	78.45%	85.63%	96.02%	98.19%
7	97.33%	94.9%	20.42%	97.16%	96.38%	98.15%
8	96.96%	98.27%	11.50%	97.08%	97.11%	97.83%
9	96.66%	96.59%	31.40%	46.61%	98.3%	96.95%
AVE	96.14%	**96.73%**	68.5%	**76.89%**	97.12%	**97.8%**

4.1. Camera brand identification

In the first experiment, in order to assess the performance of the proposed algorithm in the case of camera brand identification, nine camera devices from ID 1 to ID 9 are selected. The detection accuracies are shown in Table 3. The CA-CNN represents the method proposed in the work [22]. The Laplacian filter in the preprocessing layer is replaced with high-pass filter or three kinds of self-learn convolutional kernels: conv3 × 3, conv5 × 5, and conv7 × 7. The RN denotes the residual networks and three kinds of self-learn convolutional kernels in pre-procession are tested. The FRN and the CAF-CNN mean three paralleled residual networks and three paralleled CA-CNNs, respectively. The difference between three paralleled networks is the kernel size of the convolution in preprocessing stage. The CA-FRN is the content-adaptive fusion residual networks proposed in this work. The Googlenet [28] with the high-pass filter in the pre-procession are compared with proposed algorithm. It has the best detection performance according to the report in the work [21]. Besides, since the proposed method only has 12 layers, for fair comparison, the architecture with similar network depth should be used, ResNet18 [27] has been tested. It can be confirmed that content-adaptive fusion residual networks has the best detection performance. The self-learn convolutional kernel in pre-procession is an effective way. The detection accuracy of the CNN-HP is 81.62%, which is much lower than CNN-Conv3 × 3, CNN-Conv5 × 5, CNN-Conv7 × 7. The different kernel sizes in preprocessing stage have the positive effect on the detection performances. The fusion networks can achieve a far better detection performance and its detection accuracy is higher than CNN-Conv and RN-Conv. By comparison with GoogleNet-HP, the self-learn convolution in pre-procession of residual networks, fusion networks, and content-adaptive fusion residual networks proposed in this work have the better performances.

For validate the effectiveness of the content-adaptive fusion residual networks, we compare the detection performance of CA-FRN with FRN in three subsets: saturation, smoothness, and others. The detection accuracies for each camera brand are shown in Table 4. The CA-FRN has the better performance than FRN. What needs to be explained is that there are some significant differences between the performance of the smoothness and the others. The reason is that the training data for three subsets are unbalanced and the training data of the saturation is much less than the others. The detection performance for the saturation can be improved by increasing the training data.

4.2. Camera model, device identification

In the second experiment, the performances of proposed method for camera model and device identification are evaluated. For the case of camera model identification, three camera models are selected: Sony_DSC-H50, Sony_DSCW170, and Sony_DSC-T77. For the case of camera device identification, three camera devices from the same model are chose: Sony_DSC-T77_0, Sony_DSC-T77_1, and Sony_DSC-T77_2. Instead of re-training a new CAF-CNN model, we finetune the model trained in the first experiment.

For camera model identification, the detection accuracy is above 87.55%. For camera device identification, three devices of Sony_DSC-T77 is used. There is no doubt that camera device identification is a hard task. The average detection accuracy of the proposed method in this case is 73.27%. The confusion matrixes of proposed method in both cases are shown in Fig. 7.

In order to further estimate the feasibility of the algorithm, we test it in the mixing dataset, including different camera brands, same camera brand but different camera models, and same camera model but different camera devices. The cameras used in this stage are Sony_DSC-T77_0, Sony_DSC-T77_1, Sony_DSC-H50_0, Olympus_mju_1050SW_2, Panasonic_DMC-FZ50_1, Agfa_Sensor530s_0, Ricoh_GX100_0, Samsung_NV15_0, Kodak_M1063_0. The average detection performance is near 92%, which demonstrate that it is practicable and satisfactory to identify source camera for small-size images. The confusion matrix of proposed method in this cases is shown in Fig. 8.

4.3. Comparison with PRNU based algorithm

PRNU based methods have been study for many years in the community of image forensics and have good performance on source camera identification and clustering tasks. We compare with PRNU based method [29] to confirm the effectiveness of the proposed algorithm. The ROC and overall ROC are used to compare the performances for camera brand, model, device, and mixing identifications of two methods. To obtain the overall ROC curve, for a given detection threshold, the numbers of true positives and false positives are counted for each camera, respectively. Then these numbers are summed up and used to calculate the true positive rate (TPR) and false positive rate (FPR), which was used in the work [30]. The results are shown in the Fig. 9 below. The ROC and overall ROC of PRNU based and proposed method are indicated by a dotted and solid line, respectively. In particular, the black lines mean the overall ROC curve and color lines are the ROC curve for each camera. As we can see from the results, proposed method has better performance than PRNU based method. It should be noted that the image patches used for PRNU based method are center crop from the original images by 64 × 64. The reference PRNU is extracted by 100 training images.

4.4. Image tamper detection

Tampered image detection is a hot topic in image forensics community. There are a lot of algorithms have been proposed. The source camera identification can provide an effective reference for the splicing forgeries. Supposing that the tampered image includes multiple contents from different camera devices. It is likely that the attacker want to make a composite image and the components of composite image originate from multi-images token by different cameras. In this situation, the source camera identification for small-size images can detect the tampered image and locate the tampered areas. We test proposed algorithm in this case. The results are shown in Fig. 10 and there is a good detection performance.

In order to get a better experimental validation, we follow the protocol of image localization in Korus work. The images captured by Kodak_M1063_0 and Pentax_OptioA40_0 are chose to synthesize the tampered images. Firstly, in order to simplify, the images were central cropped by 1024 × 1024. Then synthetic forgeries were generated by replacing a randomly located square re-

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

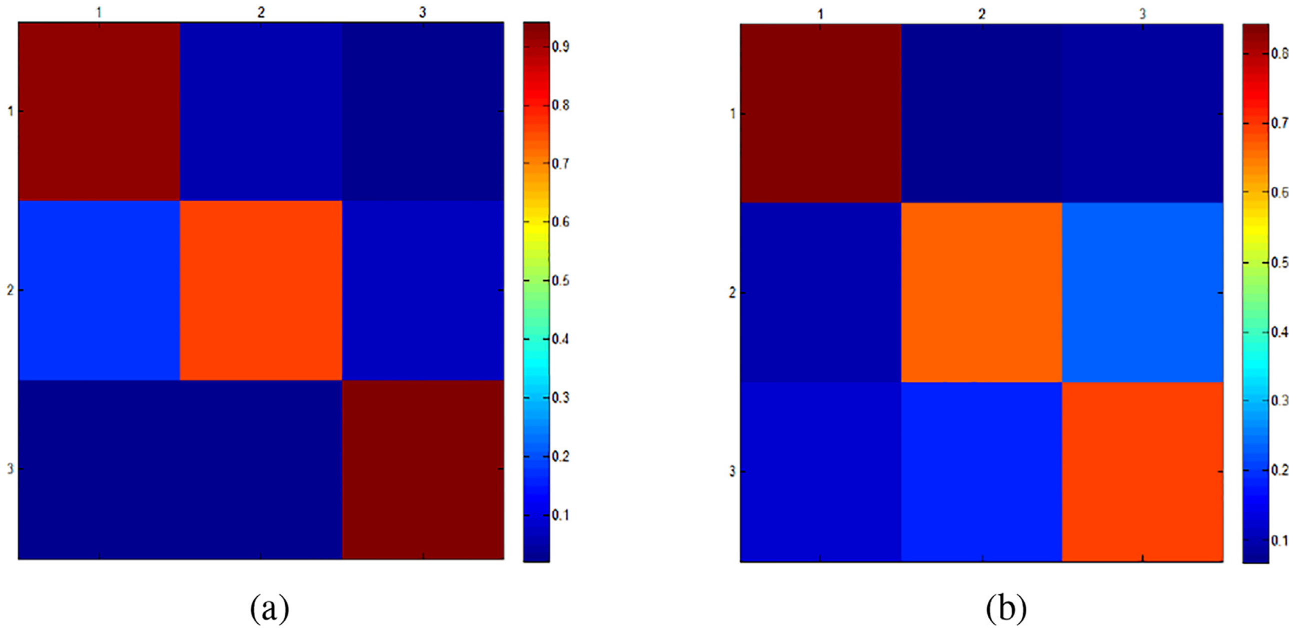


Fig. 7. Confusion matrixes of proposed method. (a) (b) show the confusion matrix in the case of the camera model identification and camera device identification, respectively.

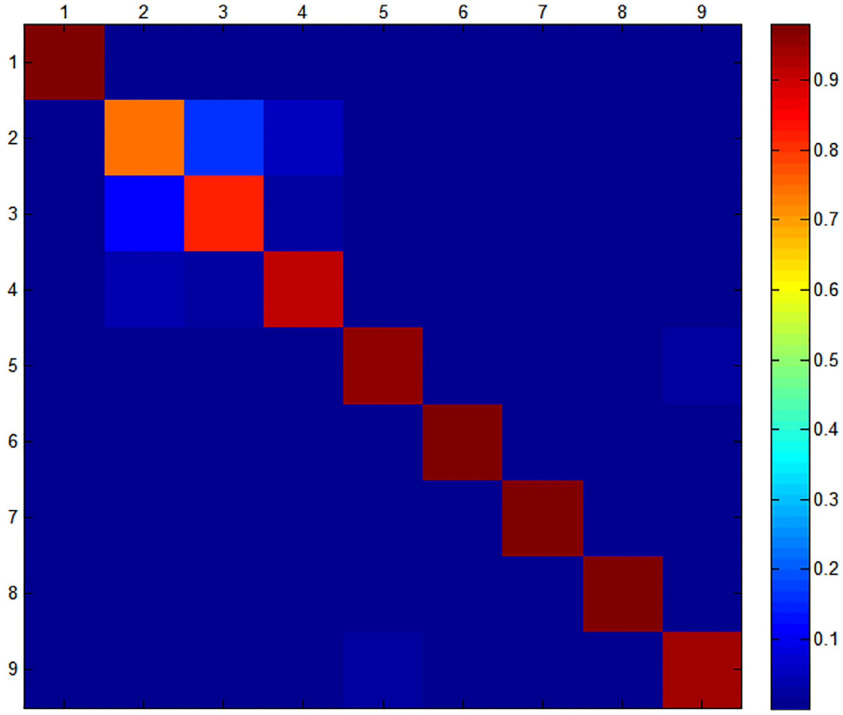


Fig. 8. Confusion matrix of proposed method in the case of mixing dataset.

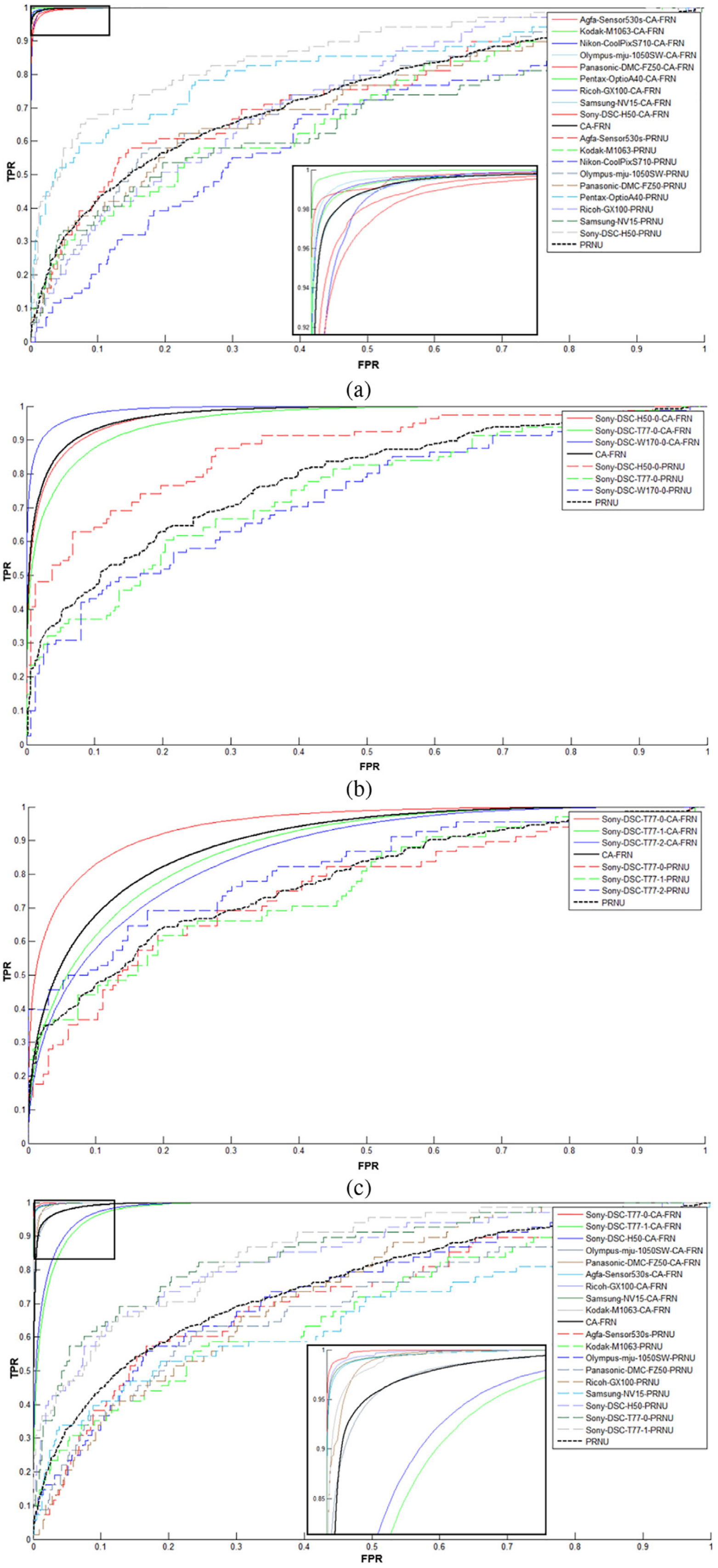
gion of the input image with a randomly chosen patch from another camera. For example, the image from the input camera, Kodak_M1063_0, is replaced a randomly located square region with a patch from the tampering camera, Pentax_OptioA40_0. Then the synthetic image is generated. The tampered blocks is set to three scales: 64, 128, 256. For each tampered block scale, two groups of tampered images were generated and each group includes 150 images. Finally, 900 tampered images were generated. For each group, we conduct the tampering location experiments and the evaluation criterions are the ACC and TPR:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$TPR = \frac{TP}{TP + FP} \tag{10}$$

Where, TP and TN mean the block numbers that the blocks are correctly classified into tampered and original blocks. FP and FN represent the block numbers that were misclassified into tampered and original blocks. The results for six groups are shown in Table 5. For the second column, A_B is the input camera _ tampering camera. The ACC and TPR are above 98%. It can be concluded that proposed method can provide a good way for tampering localization.

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016



(a)

(b)

(c)

(d)

Fig. 9. ROC and Overall ROC for PRNU based and proposed methods. (a), (b), (c), (d) mean the result for camera brand, model, device, and mixing identification, respectively.

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016

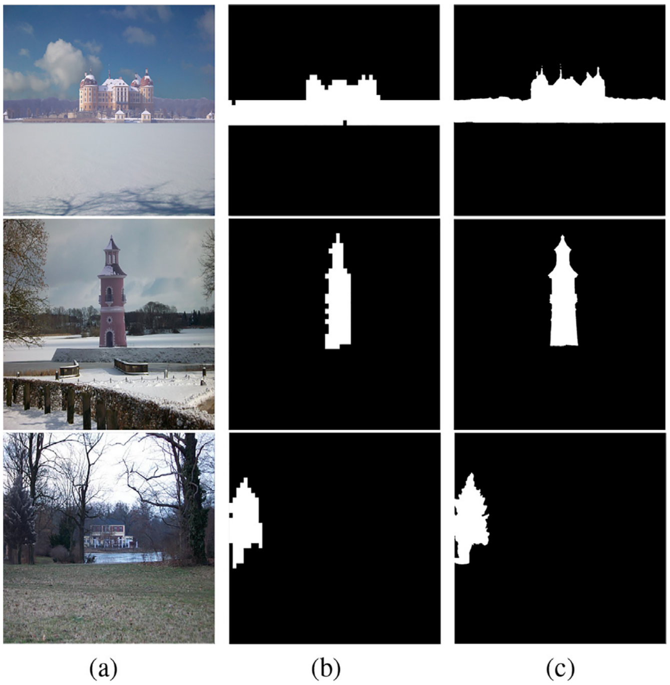


Fig. 10. Image tamper detection. The first column(a) represents the tampered images, the second column (b) show the detection results of proposed method, the last column (c) is the ground truth.

Table 5
The detection performance for tampering localization.

Tampering block size	Cameras	ACC	TPR
64	Kodak_Pentax	99.18%	98%
	Pentax_Kodak	98.34%	100%
128	Kodak_Pentax	99.19%	98.83%
	Pentax_Kodak	98.54%	100%
256	Kodak_Pentax	99.18%	98.96%
	Pentax_Kodak	98.55%	99.96%

5. Conclusions

In this paper, we propose a content-adaptive fusion residual networks for small-size images to achieve the source camera identification. According the differences of the image contents, the images are divided into three subsets: saturation, smoothness and the others. In order to learn better feature representation from the input data, we construct a residual networks. The convolutional layer is added in preprocessing stage; three residual blocks are used in architecture. Then three special residual networks are paralleled. The features learned from the last residual blocks are fused and fed into softmax classifier. The difference of three residual networks lies in the convolutional kernel size of preprocessing layer. Moreover, considering the influence of different image contents for the algorithm, we train three fusion residual networks for smooth images, saturated images, and the others, separately. The experimental results show that the proposed method has satisfactory performances at three levels of source camera identification: brand level, model level, and device level in the case of the small-size images. We believe that the architecture of CA-FRN could be applied to other image forensics scenarios.

Acknowledgments

This work was supported in part by National NSF of China (61672090, 61332012), the National Key Research and Development of China (2016YFB0800404), Fundamental Research Funds for the Central Universities (2015JBZ002, 2017YJS054). We greatly acknowledge the support of NVIDIA Corporation with the donation of the GPU used for this research.

References

[1] M. Stamm, M. Wu, K. Liu, Information forensics: an overview of the first decade, Access IEEE (2013).
[2] S. Choi, E.Y. Lam, K.K.Y. Wong, Source camera identification using footprints from lens aberration, in: Proc. SPIE, 2006.
[3] M.K. Johnson, H. Farid, Exposing digital forgeries through chromatic aberration, ACM Multimedia and Security Workshop, 2006.
[4] K. Kurosawa, K. Kuroki, N. Saitoh, Ccd fingerprint method-identification of a video camera from videotaped images, in: Proc. IEEE Int. Conf. on Image Processing, 1999.
[5] Z.J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, N. Saitoh, Methods for identification of images acquired with digital cameras, Enabling Technologies for Law Enforcement, Int. Society for Optics and Photonics, 2001.
[6] A.C. Popescu, H. Farid, Exposing digital forgeries in color filter array interpolated images, IEEE Trans. Signal Process (2005).
[7] M.J. Sorrell, Digital camera source identification through jpeg quantisation, Multimedia Forensics and Security, Information Science Reference, Hershey, NY, USA, 2008.
[8] E.J. Alles, Z.J.M.H. Geradts, C.J. Veenman, Source camera identification for heavily jpeg compressed low resolution still images, J. Forensic Sci. (2009).
[9] B. Sankur, O. Celiktutan, I. Avcibas, Blind identification of cell phone cameras, in: Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, 2007.
[10] P. Sutthiwan, J. Ye, Y.Q. Shi, An enhanced statistical approach to identifying photorealistic images, in: Proc. Int. Workshop on Digital Watermarking, 2009.
[11] M. Chen, J. Fridrich, M. Goljan, J. Lukas, Determining image origin and integrity using sensor noise, IEEE Trans. Inf. Forensics Secur. (2008).
[12] Y. Hu, B. Yu, C. Jian, Source camera identification using large components of sensor pattern noise, in: Proc. Int. Conf. on Computer Science and its Applications, 2009.
[13] C.-T. Li, Source camera identification using enhanced sensor pattern noise, IEEE Trans. Inf. Forensics Secur. (2010).
[14] G. Wu, X. Kang, K. Liu, A context adaptive predictor of sensor pattern noise for camera source identification, in: Proc. of the 19th IEEE Int. Conf. on Image Processing, 2012.
[15] M. Goljan, Digital camera identification from images estimating false acceptance probability, Digital Watermarking, 2009.
[16] X. Kang, Y. Li, Z. Qu, J. Huang, Enhancing source camera identification performance with a camera reference phase sensor pattern noise, IEEE Trans. Inf. Forensics Secur. (2012).
[17] D. Cozzolino, D. Gragnaniello, L. Verdoliva, Image forgery localization through the fusion of camera-based, feature-based and pixel-based techniques, in: IEEE International Conference on Image Processing, 2015.
[18] J. Chen, X. Kang, Y. Liu, Z. Wang, Median filtering forensics based on convolutional neural networks, IEEE Signal Process Lett. (2015).
[19] P. Yang, R. Ni, Y. Zhao, Recapture image forensics based on Laplacian convolutional neural networks, International Workshop on Digital-forensics and Watermarking, 2016.
[20] H. Li, S. Wang, A. Kot, Image recapture detection with convolutional and recurrent neural networks, IST International Symposium on Electronic Imaging, 2017.
[21] A. Tuama, F. Comby, M. Chaumont, Camera model identification with the use of deep convolutional neural networks, IEEE International Workshop on Information Forensics and Security, 2016.
[22] P. Yang, W. Zhao, R. Ni, Y. Zhao, Source camera identification based on content-adaptive fusion network, arXiv:1703.04856 (2017).
[23] J. Lukas, M. Goljan, Digital camera identification from sensor noise, IEEE Trans. Inf. Secur. Forensics (2006).
[24] A. Piva, An overview on image forensics, ISRN Sig. Process. (2013).
[25] S. Bayram, H. Sencar, N. Memon, Improvements on source camera-model identification based on cfa interpolation, Proc of Wg, 2006.
[26] T. Gloe, R. BÈµhme, The ‘dresden image database’ for benchmarking digital image forensics, in: Proceedings of the 25th Symposium On Applied Computing, 2010.
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[29] J.F.M. Goljan, T. Filler, Large scale test of sensor fingerprint camera identification, in: Proceedings of SPIE-The International Society for Optical Engineering, 2009.
[30] X. Lin, C. Li, Enhancing sensor pattern noise via filtering distortion removal, IEEE Signal Process Lett. (2016).

Please cite this article as: P. Yang et al., Source camera identification based on content-adaptive fusion residual networks, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.016