IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 8, 2018, accepted January 12, 2019, date of publication January 24, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894981

Robust Median Filtering Forensics Using Image Deblocking and Filtered Residual Fusion

WUYANG SHAN[1], YAOHUA YI[1], JUNYING QIU[2], AND AIGUO YIN[3]
[1]School of Printing and Packaging, Wuhan University, Wuhan 430079, China
[2]Institute of Fashion and Art Design, Sichuan Normal University, Chengdu 610101, China
[3]Zhuhai Pantum Electronics Co., Ltd., Zhuhai 610101, China

Corresponding author: Yaohua Yi (yyh@whu.edu.cn)

This work was supported in part by the Key Laboratory of Satellite Mapping Technology and Application and National Administration of Surveying, Mapping and Geoinformation under Grant KLSMTA-201702 and Grant KLSMTA-201604, in part by the National Science and Technology Major Project under Grant 2017ZX01030102, and in part by the National Natural Science Foundation of China under Grant 61601335 and Grant 61170101.

ABSTRACT Median filtering (MF) is frequently applied to conceal the traces of forgery and therefore can provide indirect forensic evidence of tampering when investigating composite images. The existing MF forensic methods, however, ignore how JPEG compression affects median filtered images, resulting in heavy performance degradation when detecting filtered images stored in the JPEG format. In this paper, we propose a new robust MF forensic method based on a modified convolutional neural network (CNN). First, relying on the analysis of the influence on median filtered images caused by JPEG compression, we effectively suppress the interference using image deblocking. Second, the fingerprints left by MF are highlighted via filtered residual fusion. These two functions are fulfilled with a deblocking layer and a fused filtered residual (FFR) layer. Finally, the output of the FFR layer becomes input when extracting multiple features for further classification using a tailor-made CNN. The extensive experimental results show that the proposed method outperforms the state-of-the-art methods in both JPEG compressed and small-sized MF image detection.

INDEX TERMS Median filtering, convolutional neural networks, robust forensics, JPEG compression, image deblocking.

I. INTRODUCTION

As the software for image processing improves, malicious users can now achieve their purposes by easily generating forged images. As a result, the originality, authenticity, and integrity of images are increasingly a public concern. Image forensics aims to restore trust in images and increase the difficulties when forging images. Recently, many kinds of forensic techniques for digital images were invented [1], such as techniques for detecting compositing [2]–[9], and retouching [10]–[29].

Compositing forensic techniques reveal manipulations that add or remove objects in composite images. These approaches use fingerprints, such as inconsistent direction of illumination [2]–[4], shading and shadows [5], [6], color [7], [8], and photo response non-uniformity [9] to detect image compositing. These fingerprints however, can be hidden by retouching that improves the visual quality of the forged images. Thus, compositing forensic techniques face challenges when detecting retouched composite images.

Retouching forensic techniques are designed to detect the adjustment of the visual quality of images, such as median filtering [10]–[23], contrast enhancement [24]–[27], compression [28], and resampling [29]. Each of these types of adjustments is probably used to conceal traces of image compositing. When retouching manipulations appear in images, they alert users that an image might be altered. Therefore, retouching forensic techniques identify a type of retouching manipulation, but also provide indirect evidence when investigating a composite image subjected to retouching manipulations.

This paper focuses on median filtering (MF) forensics. Previous researchers offered a lot of contribution for MF forensics. There exist three categories of MF forensic methods, i.e., threshold-based methods [10], [11], SVM-based methods [12]–[19], and CNN-based methods [20]–[23]. In the threshold-based methods, Kirchner and Fridrich [10] invented an image feature derived from histogram bins of the first-order difference image. Cao *et al.* [11] designed an

image feature based on the probability of zero values on the first-order difference map in texture regions. Then they both employed the features to detect MF images by thresholding classification. While these methods are of splendid simplicity—just relying on a scalar feature—it turned out that they cannot detect MF in images which are filtered and then JPEG compressed.

In order to improve the robustness of MF forensic methods, support vector machine (SVM) [30], [31], an advanced machine learning technique was combined with higher dimensional feature sets to detect MF images. Kirchner *et al.* [10] combined the subtractive pixel adjacency matrix [12] features with SVM to detect MF images. Yuan [13] found MF significantly affects either the order or quantity of the gray levels in every filtered image window. Such a finding motivated his construction of a 44-D median filtering forensics feature set, which contained five groups of features over all image blocks. Chen *et al.* [14] presented the global probability feature set and local correlation feature set, relying on the empirical cumulative distribution function and correlation between adjacent difference pairs in the difference domain, respectively. Kang *et al.* [15] utilized the first ten coefficients of the autoregressive model (AR) for median filtered residual as a feature set. Zhang *et al.* [16] applied the second-order local ternary pattern to investigate MF from the angle of microtexture structures. Niu *et al.* [17] fused local binary patterns and pixel difference matrix to generate a feature set. Motivated by the work in [15], Yang *et al.* [18] combined the 2-D AR coefficients of median filtered residual (MFR), average filtered residual (AFR), and gaussian filtered residual (GFR). Wang *et al.* [19] used the coefficients in image frequency domain as a feature set. The SVM-based methods are more robust than the threshold-based methods when dealing with small size and JPEG compressed MF images. But their performance still suffers heavy degradation when the quality factor of JPEG compression was lower than 90.

In recent years, motivated by the successful instances applying convolutional neural networks (CNNs) [32], [33] in image forensics [34]–[40], some researchers proposed CNN-based methods to detect MF images and achieved better performance. Chen *et al.* [20] first attempted to introduce CNNs into MF forensics. They added a MFR layer in front of a classical CNN to make it suit to MF forensic tasks. Bayar and Stamm [21], [22] designed a constrained convolutional layer for general forensic purpose. Jin *et al.* [23] explored the dark channel residual and a generative adversarial network [41] to detect RGB color MF images.

The above mentioned methods focus to improve the robustness when dealing with small size and JPEG compressed MF images. In daily life, we often save images as JPEG format. Thus, median filtered images are probably performed JPEG compression. In addition, the tampering localization relies on the detection of small size MF images. Existing MF forensic methods however, just depend on high dimensional feature sets or trendy machine learning techniques to improve their JPEG-robustness, without analyzing how the manipulation of JPEG compression affects median filtered images. As shown in Fig. 2 (c), the manipulation of JPEG compression generates the visible blocking artifacts extending all over the compressed image. MF classifiers confuse the blocking artifacts with the image texture, while the latter contains the key image features in MF forensic tasks. Therefore, the MF image features cannot be accurately extracted, resulting in poor performance on MF detection under JPEG compression. In order to improve the robustness of MF forensic methods, we propose to detect MF not only in small size images, but also in JPEG compressed ones. The main strategy relies on image deblocking methods [42]–[46], which significantly reduce the blocking artifacts caused by JPEG compression.

Besides powerful preprocessing methods, selecting appropriate machine learning techniques is also important. Among existing MF forensic methods, CNN-based methods are the most advanced, because they extract multiple features automatically and outperform other methods. However, when we choose classical CNNs to learn MF image features, it is necessary to modify the networks, because they were tailored for target recognition [20]–[23]. Therefore, in our work, we propose a fused filtered residual (FFR) layer to combine MFR, AFR, and GFR into a filtered residual image, which helps our CNN extract more accurate features. Additionally, we elaborately design a CNN framework, making it available to perform our MF tasks.

In this paper, we propose a novel robust MF forensic method using image deblocking and filtered residual fusion. In our proposed MF forensic CNN, we add a deblocking layer and FFR layer in front of a tailor-made CNN framework. The main contributions of our work are summarized as follows:

1) Unlike existing MF forensic methods, our method starts from the analysis of the influence on median filtered images caused by JPEG compression and then effectively reduces the influence via image deblocking. This strategy can be used for other JPEG-robust forensic tasks.
2) Thanks to our proposed FFR layer, our CNN extracts more accurate features and achieves performance improvement.
3) Our method outperforms the existing MF forensic methods in detecting both small size and JPEG compressed MF images.

II. MEDIAN FILTERING FORENSIC CNN

A. OVERALL ARCHITECTURE

In our preliminary experiments, we accepted potentially-filtered images directly as the input of classical CNNs, such as AlexNet and VGG, but did not get the expected performance. There are three reasons causing the poor results: Classical CNNs were tailored for recognizing objects in images, but MF just changes the texture of an image, rather than

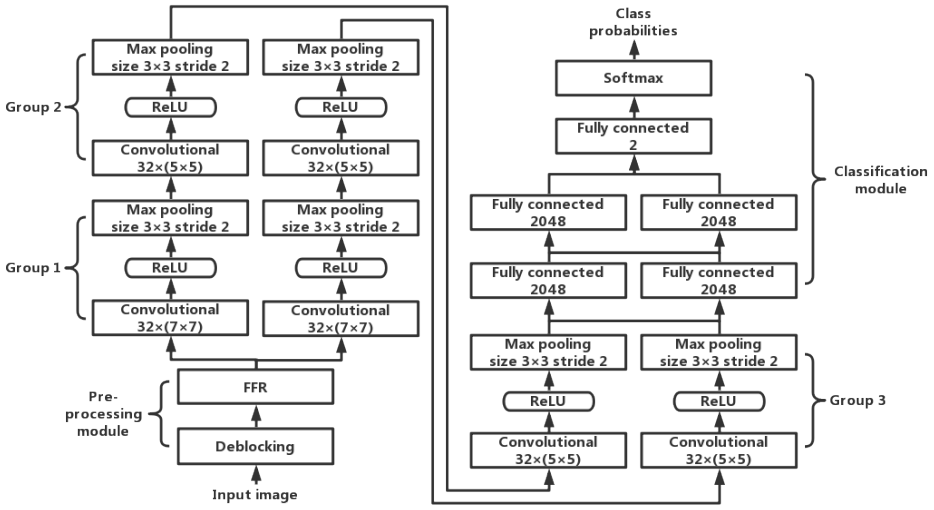


FIGURE 1. Overall architecture of our proposed MF forensic CNN. Layer types and parameter sizes (choices) are displayed inside boxes. Sizes of convolution kernels in the boxes follow (number of kernels) × (height × width). To increase the convergence rate and decrease the information loss when information travels through multiple layers, every convolutional layer is followed by rectified linear unit (ReLU) activations [46]. The first and second fully connected layers both have 4096 neurons and are equipped with a "dropout" function [32] that inhibits overfitting. The last fully connected layer has two neurons. Its output feeds into a two-way softmax layer, to produce class probabilities. Using the group technique [32], the training module was divided into two groups. The weights in the network are initialized with Gaussian initialization, Stochastic Gradient Descent (SGD) optimizer is used as the optimization solver.

the image content, as shown in Fig. 2 (b). In other words, in a potentially-filtered image, image content becomes redundant information passively with strong influence on CNNs when training on MF image features. Secondly, as shown in Fig. 2 (c), JPEG compression produces the block-ing artifacts and therefore strongly confuses classifiers. Thirdly, the frameworks of classical CNNs need to be optimized for MF forensic tasks. Hence, we designed a deblocking layer, a FFR layer and a tailor-made framework as described in detail in Section B and C, based on the three following ideas: The FFR layer suppresses the interference caused by image content, so the fingerprints left by MF are successfully exposed. Secondly, the deblocking layer import-antly reduce the blocking artifacts in potentially-filtered images after JPEG compression. Last but not least, the opti-mized CNN framework increases the efficiency and perfor-mance of our method. Figure 1 shows the overall architecture and parameters of the proposed MF forensic CNN.

As shown in Fig. 1, the workflow preprocesses input images in the deblocking and FFR layers. A tailor-made CNN learns image features from the processed images; three groups of convolutional modules and a classification module comprise the tailor-made CNN. In the classification module, the softmax function, which returns the class probability, is defined as

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad for \quad j = 1, ..., K \tag{1}$$

where K is the number of classes, z is the vector consisting of K output values of the last layer, and the softmax probability of the j-th class, $\sigma(z_j)$, is in the range of [0, 1].

Our network is optimized by minimize the following objective loss function

$$L = -\sum_{j=1}^{K} y_j \log \sigma(z_j) \tag{2}$$

where $y_j \in \{0, 1\}$ is the ground truth label of the j-th class.

Unlike classical CNNs, the first module of our CNN is a pre-processing module containing the deblocking and FFR layers.

B. DEBLOCKING LAYER

Median filtering, an order statistic operation is implemented by replacing each pixel value in an image with the median of its neighboring pixels. When performing MF, the values of neighboring pixels in filtered images are non-linear mapping to closer values, and therefore the image is strongly smoothed [48]. This results a wide gap between the image texture of MF images and original images, as shown in Fig. 2 (a)-(b). Thus, image texture contains the key MF image features.

JPEG compression is a block-based transform coding [49], which usually introduces annoying blocking artifacts in compressed images [42]–[46]. As seen in Fig. 2 (c), there are visible blocking artifacts extending all over the compressed images. MF classifiers confuse the blocking artifacts with the image texture, and therefore extract inaccurate MF image features, resulting in poor performance on MF detection under JPEG compression.

In order to reduce the blocking artifacts in JPEG compressed images while preserving image quality, we designed a deblocking layer that accepts a potentially-filtered image as input and outputs its deblocked version.

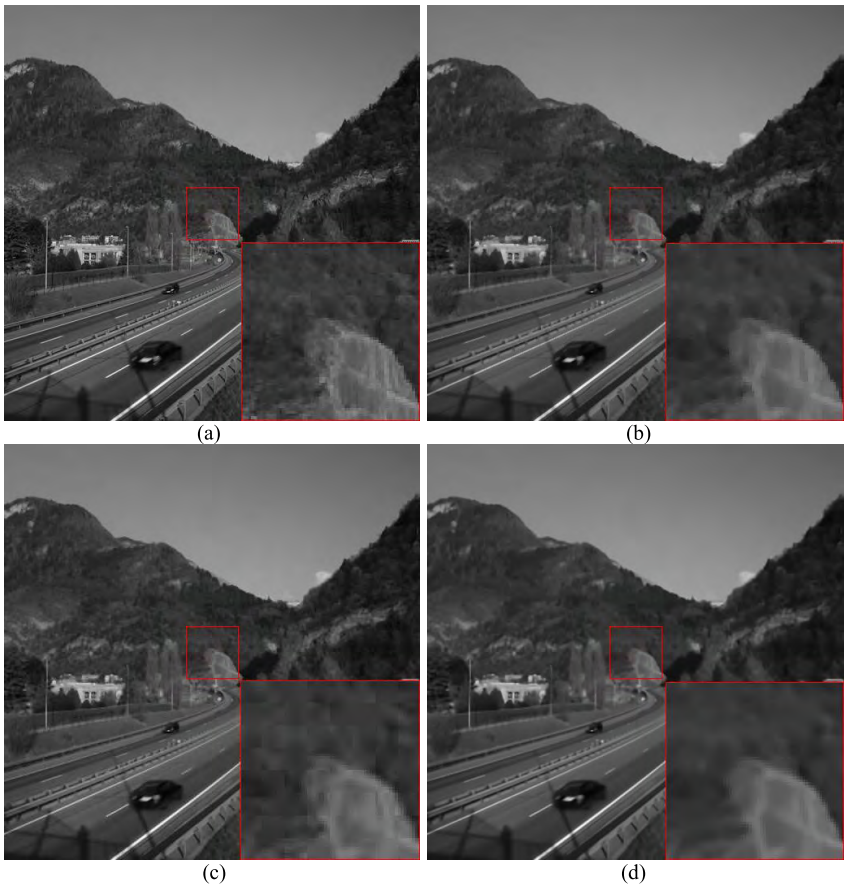


FIGURE 2. The influence on an original image caused by median filtering, JPEG compression, and image deblocking. (a) an original image; (b) its 3 × 3 median filtered image; (c) a compressed image generated by performing JPEG compression on the filtered image with a quality factor (*QF*) of 70; (d) this compressed image were performed image deblocking.

Image deblocking aims to restore a JPEG compressed image by approximating its original image, and therefore can reduce the blocking artifacts in JPEG compressed images while preserving image quality. In this paper, we solved the problem of image deblocking through maximum a posteriori (MAP) framework. Given an original image O and its JPEG compressed image J, O can be obtained by

$$\tilde{O} = \arg\max_{O} p(O\,|J\,) \tag{3}$$

Based on the Bayesian rule, Eq. (1) can be expressed as

$$\tilde{O} = \arg\max_{O} In(p(J\,|O\,)) + In(p(O)) \tag{4}$$

where $J = In(p(J\,|O\,))$ is the likelihood function, describing the relationship between J and O, $O = In(p(O))$ is the prior probability. Adopting the Gaussian quantization noise model [42] to depict the first term in Eq. (4), we can formulate it as

$$In(p(J\,|O\,)) = -\frac{1}{2\sigma_e^2}\,\|O - J\|_2^2 \tag{5}$$

where σ_e is the variance of Gaussian model, estimated by the following empirical formulation [45]:

$$\sigma_e^2 = 1.195 * (\dot{s})^{0.6394} + 0.9693,\ \dot{s} = \frac{1}{9}\sum_{i,j=1}^{3} M^q_{[i,j]} \tag{6}$$

where M^q is the 8×8 quantization matrix.

Incorporating the above quantization noise model, the non-convex low-rank prior model [43], [44] and the quantization constraint prior model [50] into Eq. (4), it is formulated as an optimization problem

$$\begin{aligned}(\tilde{O}, \tilde{Z}_{G_k}) = \arg\min_{O, Z_{G_k}} & \frac{1}{2\sigma_e^2}\,\|O - J\|_F^2 \\ & + \frac{\alpha}{2\sigma_s^2}\sum_{k=1}^{K}\left\|R_{G_k}O - Z_{G_k}\right\|_F^2 \\ & + \beta\sum_{k=1}^{K} F(Z_{G_k}) \quad s.t.\ O \in \Omega\end{aligned} \tag{7}$$

where Z_{G_k} is a low-rank matrix, R_{G_k} is a matrix operator that extracts patch from O, $F(Z_{G_k})$ is a non-convex surrogate function, Ω is the restricted solution space. Using the alternatively

minimizing strategy [45], we get the closed-form solution, i.e.

$$\dot{O} = (I + \frac{\sigma_e^2 \alpha}{\sigma_s^2} \sum_{k=1}^{K} R_{G_k}^T R_{G_k})^{-1} (J + \frac{\sigma_e^2 \alpha}{\sigma_s^2} \sum_{k=1}^{K} R_{G_k}^T Z_{G_k}) \quad (8)$$

where I is the identity matrix. By applying the projection operation, we find the optimal feasible solution, that is

$$\tilde{O} = A^{-1} P(A(\dot{O}), \hat{l}, \hat{u}) \quad (9)$$

where A is a block discrete cosine transform (DCT) matrix operator, A^{-1} represents the inverse process, $\hat{l}$ is a lower bound vector, $\hat{u}$ is a upper bound vector [45].

Algorithm 1 Algorithm in the Deblocking Layer

1: **Input:** potentially-filtered images stored in JPEG format
2:**Initialization:**
3: Get J, M^q from the input images;
4: Set parameters α, β;
5: Set $O^{(0)} = J$;
6: Calculate σ_e by Eq. (6);
7: Determine the feasible solution space Ω;
8: **For** $t = 1, ...T$
9: Calculate $\dot{O}^{(t)}$ by Eq. (8);
10: Update $O^{(t)}$ by Eq. (9);
11: **End For**
12: **Output:** Final deblocking result $\tilde{O} = O^{(T)}$

In light of all derivations above, a detailed description of algorithm in the deblocking layer is provided in Algorithm 1. After five iterations, the optimized result is the final output of the deblocking layer. As seen in Fig. 2 (d), the blocking artifacts disappear after image deblocking.

C. FFR LAYER

After a compressed potentially-filtered image is processed through the deblocking layer, the blocking artifacts are significantly reduced. The redundant image content (as described in Section A) in the processed image however, needs to be removed, as it heavily affects CNNs when training on MF image features. Motivated by Yang *et al.* [18], we designed a FFR layer for the output images of deblocking layer to suppress the interference caused by image content, so the fingerprints left by MF are successfully exposed.

At first, the FFR layer extracts the MFR, AFR, and GFR of an input image, which can be described as

$$R_f(x, y) = I(x, y) - \hat{I}_f(x, y) \quad (10)$$

where $\hat{I}_f \in \{median\ filtered\ image, average\ filtered\ image, gaussian\ filtered\ image\}$, the subscript f is the index of the filtered image, I is the original image.

Then, the three filtered residual images are combined into a fused filtered residual image. The fusion procedure can be formulated as the following energy function [51]:

$$F_i = \left(\prod_{f=1}^{S} E_\tau(R_i^{(f)}) \right)^{1/S} \quad (11)$$

where $R^{(f)}$ is the f-th image in S input images with identical size of all images $N = N_x \times N_y$, F_i is the i-th pixel value in fused image, $i = 1, 2, ..., N$, E_τ is defined as

$$E_\tau(R) = -\log \max(\psi_{\min}, 1 + \frac{R}{2(1-\tau)} - \frac{1}{2(1-\tau)}) \quad (12)$$

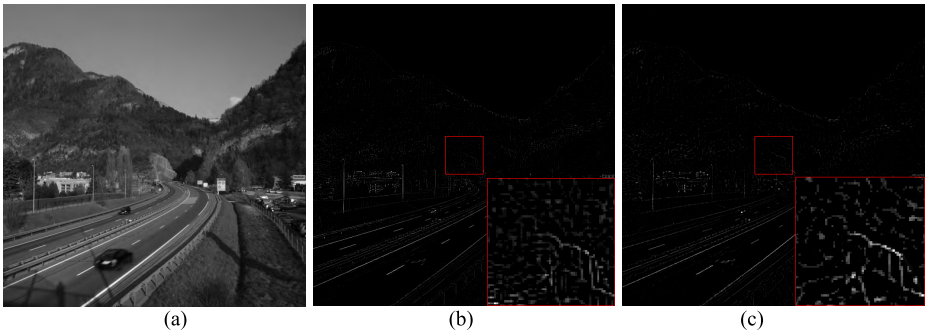
where $\psi_{\min}$ is set by users. Setting a minimal value $\psi_{\min}$ prevents the nodes from becoming fixed to certain decisions.

We record an individual threshold $\tau_i^{(f)}$ for each pixel in input images, and update it as follows:

$$\tau_i^{(f)} = \begin{cases} \tau^{(1)} \\ \tau_i^{(f-1)} + \delta & if\ f > 1\ and\ R_i^{(f-1)} \le \tau_i^{(f-1)}, \\ \tau_i^{(f-1)} - \delta & if\ f > 1\ and\ R_i^{(f-1)} > \tau_i^{(f-1)}, \end{cases} \quad (13)$$

where $\tau^{(1)}$ is an initial value, δ is the strength of the threshold drift. The above algorithm in the FFR layer is summarized in Algorithm 2.

Figure 3 shows the output image of the FFR layer. We first compress an original image with a quality factor of 70 and then perform image deblocking on the compressed image. In contrast, the original image is first median filtered and then JPEG compressed, finally performed image deblocking. As seen in Fig. 3, the interference caused by image content



(a) (b) (c)

FIGURE 3. An original image and the output images of FFR layer following the deblocking layer. (a) an original image; (b) the FFR of the image which is JPEG compressed with a *QF* of 70 and then performed image deblocking; (c) the FFR of the image which is 3 × 3 median filtered in JPEG compression with a *QF* of 70, and then performed image deblocking.

Algorithm 2 Algorithm in the FFR Layer

1: **Input:** the output images of deblocking layer
2: **Initialization:**
3: Calculate the MFR, AFR, and GFR of input images by Eq. (10);
4: Set parameters δ, $\psi_{\min}$, and $\tau^{(1)}$;
5: **For** $f = 1, ...S$
6: Update $\tau^{(f)}$ by Eq. (13);
7: Calculate $E^{(f)}$ by Eq. (12);
8: **End For**
9: Calculate F by Eq. (11);
10: **Output:** Final FFR result F

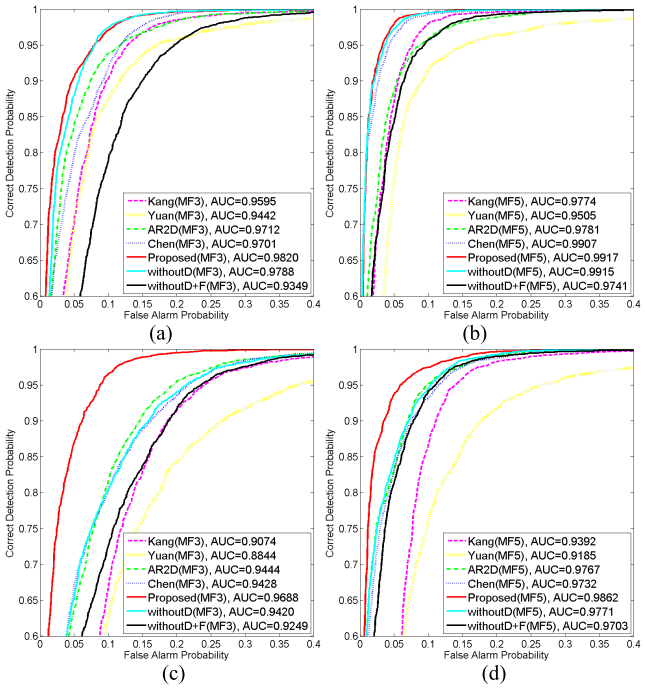


FIGURE 4. ROC curves of the proposed and four existing methods for distinguishing the median filtered images from the non-filtered images in 64 × 64 blocks. (a) JPEG90 vs MF3-JPEG90; (b) JPEG90 vs MF5-JPEG90; (c) JPEG70 vs MF3-JPEG70; (d) JPEG70 vs MF5-JPEG70.

is eliminated in the output images of FFR layer, and the difference between the median filtered image and unfiltered image can be investigated.

III. EXPERIMENTAL RESULTS

In this section, extensive experiments are performed to evaluate comprehensively the performance of the proposed method, with comparing to the state-of-the-art MF forensic methods.

A. DATABASE AND EXPERIMENTAL SETUPS

To perform the experiments, we used the grayscale version [10]–[22] of 10000 images randomly selected from five frequently-used databases as the Original dataset. The UCID database [52], the BOSS RAW database [52], the Dresden Image Database (DID) [53] and the NRCS Photo Gallery (NRCS) database [54] contribute 1000 images,

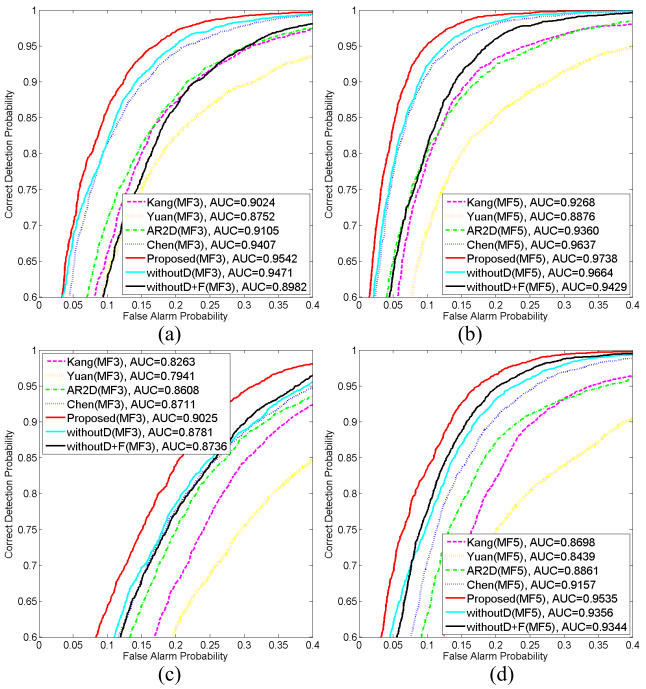


FIGURE 5. ROC curves of the proposed and four existing methods for distinguishing the median filtered images from the non-filtered images in 32 × 32 blocks. (a) JPEG90 vs MF3-JPEG90; (b) JPEG90 vs MF5-JPEG90; (c) JPEG70 vs MF3-JPEG70; (d) JPEG70 vs MF5-JPEG70.

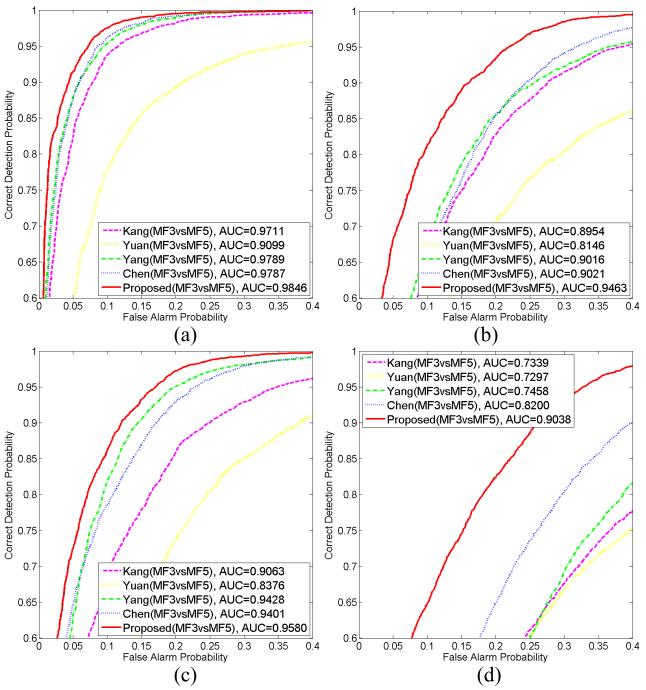


FIGURE 6. ROC curves of the proposed and four existing methods for distinguishing between 3 × 3 and 5 × 5 median filtering. (a) MF3-JPEG90 vs MF5-JPEG90 in 64 × 64 blocks; (b) MF3-JPEG90 vs MF5-JPEG90 in 32 × 32 blocks; (c) MF3-JPEG70 vs MF5-JPEG70 in 64 × 64 blocks; (d) MF3-JPEG70 vs MF5-JPEG70 in 32 × 32 blocks.

respectively, and the BOSSbase 1.01 [52] contributes 6000 images. Then, these images were processed as follows:

1) The Original dataset was compressed into JPEG format with quality factors (QFs) $QF_1 \in \{90, 70\}$.

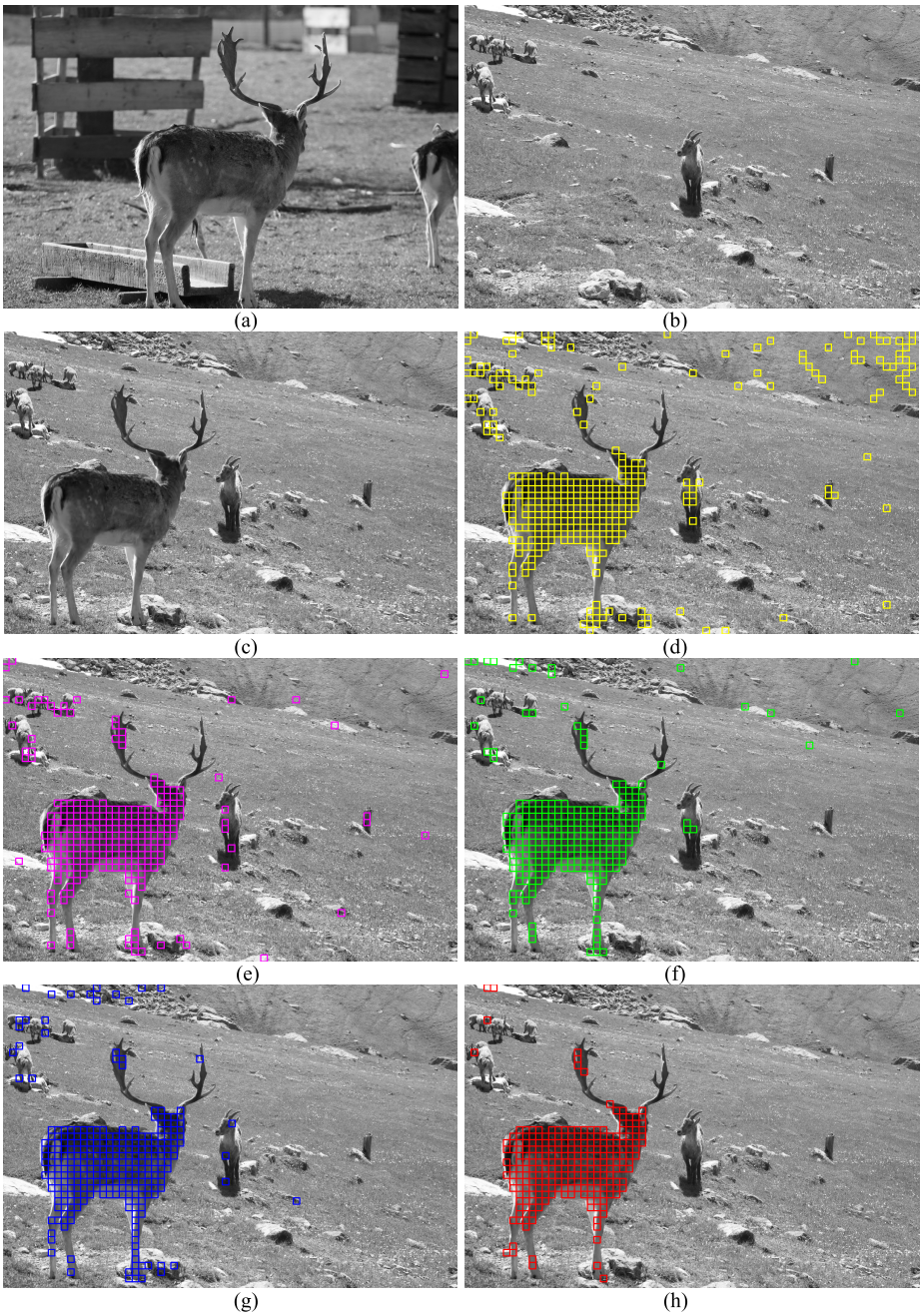


FIGURE 7. An example of a composite image detection. (a) A RAW image from which an object (the buck) is cropped; (b) A RAW image into which the cropped object is pasted; (c) An composite image compressed with a *QF* of 70. The detected blocks are marked (d) using the method in [13]; (e) using the method in [15]; (f) using the method in [18]; (g) using the method in [20]; and (h) using the proposed method.

These com-pressed images comprised the JPEG90, and JPEG70 datasets. Each of them contained 10000 images.

2) 3×3 MF and 5×5 MF were performed on the Original dataset. These filtered images comprised the MF3 and MF5 datasets, respectively. Each of them contained 10000 images.
3) The MF3-JPEG90, MF3-JPEG70, MF5-JPEG90, and MF5-JPEG70 datasets were obtained by JPEG compression performed on the MF3 and MF5 datasets with quality factors $QF_2 \in \{90, 70\}$, respectively.

The methods in [13], [15], [18], and [20] were used for comparison with our proposed method. For the method in [20] and the proposed MF forensic CNN, all experiments were performed on a modified Caffe toolbox [56]. For the other methods, we used C-SVM as classifiers.

B. MEDIAN FILTERING DETECTION IN JPEG COMPRESSED AND SMALL SIZE IMAGES

In order to detect the MF in JPEG compressed and small size images, we prepared four positive-negative pairs: {JPEG90, MF3-JPEG90}, {JPEG70, MF3-JPEG70},

{JPEG90, MF5-JPEG90}, and {JPEG70, MF5-JPEG70}. The image blocks consisting of 64 × 64, and 32 × 32 pixels were cropped from the center of every image in the positive-negative pairs. For each pair, the training set contained half number of images, while the other half of images composed the testing set. The receiver operating characteristic (ROC) curves were generat-ed for evaluation. The overall performance of the classifier is summarized by the area under the ROC curve (AUC).

MF detection ROC curves are shown in Figure 4 and Figure 5. It indicates that the performance of proposed method was still excellent and outperformed the methods in [13], [15], [18], and [20], even if the positive samples contained the small size images which were median filtered and then JPEG compressed. This experiment suggests that the proposed method is strongly robust when dealing with JPEG compression in small size images. In addition, the results for the network without the new layers have been reported in Fig. 4 and Fig. 5. The ''withoutD'' denotes that the network with the FFR layer but without the deblocking layer, ''withoutD+F'' denotes that the network without the FFR and deblocking layer. The results show that our proposed deblocking layer and FFR layer play a key role. Our method got a remarkable performance progress using the new layers. Notably, a little improvement is also important under high performance.

C. DIFFERENTIATION BETWEEN 3 × 3 AND 5 × 5 MEDIAN FILTERING

Once forensic investigators have identified that an image has been median filtered, they may wish to determine the window size used during median filtering. So, differentiating between 3 × 3 MF and 5 × 5 MF is also a valuable work. In order to test the performance of the proposed method in this scenario, we prepared two positive-negative pairs: {MF3-JPEG90, MF5-JPEG90}, and {MF3-JPEG70, MF5-JPEG70}. The image blocks consisting of 64 × 64, and 32 × 32 pixels were cropped from the center of every image in the positive-negative pairs. For each pair, the training set contained half number of images, while the other half of images composed the testing set.

MF detection ROC curves are shown in Figure 6. It is shown that our proposed method works well and outperforms the methods in [13], [15], [18], and [20]. We can conclude that our proposed method has an excellent capability to distinguish between 3 × 3 MF and 5 × 5 MF.

D. COMPOSITE IMAGES DETECTION

Figure 7 shows the stages when generating composite images and the corresponding forensic detection results based on MF. Fig. 7 (a) is the image from which the object (the buck) was cropped. Fig. 7 (b) is the image into which the cropped object was pasted. Fig. 7 (c) is the composite image. Median filtering with a filtering window size of 5 × 5 was performed on Fig. 7 (a) to adjust the inconsistent smoothness between Fig. 7 (a) and Fig. 7 (b). The object cropped from Fig. 7 (a) was then pasted into Fig. 7 (b). Finally, the composite image was stored in JPEG format with a QF of 70. In order to locate the composite positions, the test image was segmented into several blocks consisting of 32 × 32 pixels, and each block was detected one by one. The detection results of the methods in [13], [15], [18], and [20] and proposed method are shown in Fig. 7 (d) to Fig. 7 (h) respectively. The figures show that comparing with the methods in [13], [15], [18], and [20], our method achieves much lower false alarm probability and almost the same correct detection probability.

E. COMPUTATIONAL COMPLEXITY

The MF detection workflow is implemented on our computer, with an Intel i7 core CPU, 64GB memory, and two NVIDIA Titan X GPUs. The first step in this workflow is the image deblocking which reduces the blocking artifacts in potentially-filtered images after JPEG compression. The deblocking layer takes 8 seconds to generate the deblocked version of an image block consisting of 64 × 64 pixels, and takes 2 seconds to generate the deblocked version of an image block consisting of 32 × 32 pixels. Secondly, the FFR layer costs 30 milliseconds to produce a fused filtered residual. Thirdly, our network takes less than 20 minutes to train on two NVIDIA Titan X GPUs. Then, our trained model takes around 10 milliseconds to detect a potentially-filtered image.

IV. CONCLUSION

We proposed a robust MF forensic method using image deblocking and filtered residual fusion. Unlike existing MF forensic methods, our method begins with the analysis of the influence on median filtered images caused by JPEG compression and then effectively reduces the influence via image deblocking. On top of that, we suppress the interference caused by image content, so the fingerprints left by MF are successfully exposed. Experimental results demonstrate that our proposed method achieves remarkable improvements in both JPEG compressed and small size MF image detection. Furthermore, we believe that the preprocessing strategy of our method provides reference for other forensic tasks.

ACKNOWLEDGMENT

The authors would like to thank G. Zhu for sharing the source code with them. They would also like to thank the editors and anonymous reviewers who made constructive comments and improvements on this paper.

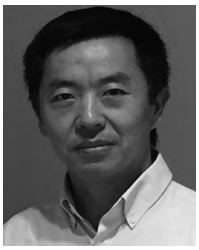
REFERENCES

[1] H. Farid, ''Image forgery detection,'' *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 16–25, Mar. 2009.
[2] M. K. Johnson and H. Farid, ''Exposing digital forgeries by detecting inconsistencies in lighting,'' in *Proc. 7th Workshop Multimedia Secur.*, 2005, pp. 1–10.
[3] M. K. Johnson and H. Farid, ''Exposing digital forgeries in complex lighting environments,'' *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 450–461, Sep. 2007.
[4] E. Kee and H. Farid, ''Exposing digital forgeries from 3-D lighting environments,'' in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2010, pp. 1–6.

[5] E. Kee, J. F. O'Brien, and H. Farid, "Exposing photo manipulation with inconsistent shadows," *ACM Trans. Graph.*, vol. 32, no. 3, pp. 28:1–28:12, Jul. 2013.
[6] E. Kee, J. F. O'Brien, and H. Farid, "Exposing photo manipulation from shading and shadows," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 165:1–165:21, Aug. 2014.
[7] T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1182–1194, Jul. 2013.
[8] T. Carvalho, F. A. Faria, H. Pedrini, R. da S. Torres, and A. Rocha, "Illuminant-based transformed spaces for image forensics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 4, pp. 720–733, Apr. 2016.
[9] P. Korus and J. Huang, "Multi-scale analysis strategies in PRNU-based tampering localization," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 4, pp. 809–824, Apr. 2017.
[10] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," *Proc. SPIE*, vol. 7541, Jan. 2010, Art. no. 754110.
[11] G. Cao, Y. Zhao, R. Ni, L. Yu, and H. Tian, "Forensic detection of median filtering in digital images," in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2010, pp. 89–94.
[12] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 215–224, Jun. 2010.
[13] H.-D. Yuan, "Blind forensics of median filtering in digital images," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1335–1345, Dec. 2011.
[14] C. Chen, J. Ni, and J. Huang, "Blind detection of median filtering in digital images: A difference domain based approach," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4699–4710, Dec. 2013.
[15] X. Kang, M. C. Stamm, A. Peng, and K. J. R. Liu, "Robust median filtering forensics using an autoregressive model," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 9, pp. 1456–1468, Sep. 2013.
[16] Y. Zhang, S. Li, S. Wang, and Y. Q. Shi, "Revealing the traces of median filtering using high-order local ternary patterns," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 275–279, Mar. 2014.
[17] Y. K. Niu, Y. Zhao, and R. Ni, "Robust median filtering detection based on local difference descriptor," *Signal Process., Image Commun.*, vol. 53, pp. 65–72, Apr. 2017.
[18] J. Yang, H. Ren, G. Zhu, J. Huang, and Y. Shi, "Detecting median filtering via two-dimensional AR models of multiple filtered residuals," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 7931–7953, Apr. 2018.
[19] D. Wang, T. Gao, and F. Yang, "A forensic algorithm against median filtering based on coefficients of image blocks in frequency domain," *Multimedia Tools Appl.*, vol. 77, no. 18, pp. 23411–23427, Sep. 2018.
[20] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, "Median filtering forensics based on convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1849–1853, Nov. 2015.
[21] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur.*, 2016, pp. 5–10.
[22] B. Bayar and M. C. Stamm, "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2691–2706, Nov. 2018.
[23] X. Jin, P. Jing, and Y. Su, "AMFNet: An adversarial network for median filtering detection," *IEEE Access*, vol. 6, pp. 50459–50467, 2018.
[24] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 492–506, Sep. 2010.
[25] G. Cao, Y. Zhao, R. Ni, and X. Li, "Contrast enhancement-based forensics in digital images," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 3, pp. 515–525, Mar. 2014.
[26] A. D. Rosa, M. Fontani, M. Massai, A. Piva, and M. Barni, "Second-Order Statistics Analysis to Cope With Contrast Enhancement Counter-Forensics," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1132–1136, Aug. 2015.
[27] W. Shan, Y. Yi, R. Huang, and Y. Xie, "Robust contrast enhancement forensics based on convolutional neural networks," *Signal Process., Image Commun.*, vol. 71, pp. 138–146, Feb. 2019, doi: 10.1016/-j.image.2018.11.011.
[28] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 480–491, Sep. 2010.
[29] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, Feb. 2005.
[30] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, Jul. 1992, pp. 144–152.
[31] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[32] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. San Mateo, CA, USA: Morgan Kaufmann, 1990, pp. 396–404.
[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1090–1098.
[34] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2014, pp. 1–4.
[35] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," *Proc. SPIE*, vol. 9409, pp. 94090J-1–94090J-10, Mar. 2015.
[36] L. Pibre, J. Pasquet, D. Ienco, and M. Chaumont, "Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch," *Proc. SPIE*, pp. 14–18, Feb. 2016.
[37] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 708–712, May 2016.
[38] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "First steps toward camera model identification with convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 259–263, Mar. 2017.
[39] H. Yao, T. Qiao, M. Xu, and N. Zheng, "Robust multi-classifier for camera model identification based on convolution neural network," *IEEE Access*, vol. 6, pp. 24973–24982, 2018.
[40] R. Xiao, R. Cui, and M. Lin, "SOMDNCD: Image change detection based on self-organizing maps and deep neural networks," *IEEE Access*, vol. 6, pp. 35915–35925, 2018.
[41] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[42] D. Sun and W.-K. Cham, "Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2743–2751, Nov. 2007.
[43] J. Ren, J. Liu, M. Li, W. Bai, and Z. Guo, "Image blocking artifacts reduction via patch clustering and low-rank minimization," in *Proc. IEEE Data Compress. Conf. (DCC)*, Mar. 2013, p. 516.
[44] M. Li, J. Liu, J. Ren, and Z. Guo, "Patch-based image deblocking using geodesic distance weighted low-rank approximation," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.*, Dec. 2014, pp. 101–104.
[45] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "CONCOLOR: Constrained non-convex low-rank model for image deblocking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1246–1259, Mar. 2016.
[46] C. Zhao, J. Zhang, S. Ma, X. Fan, Y. Zhang, and W. Gao, "Reducing image compression artifacts by structural sparse representation and quantization constraint prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2057–2071, Oct. 2017.
[47] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[48] I. Pitas and A. N. Venetsanopoulos, "Order statistics in digital image processing," *Proc. IEEE*, vol. 80, no. 12, pp. 1893–1921, Dec. 1992.
[49] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, 1991.
[50] S. H. Park and D. S. Kim, "Theory of projection onto the narrow quantization constraint set and its application," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1361–1373, Oct. 1999.
[51] P. Korus and J. Huang, "Multi-scale fusion for improved localization of malicious tampering in digital images," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1312–1326, Mar. 2016.
[52] G. Schaefer and M. Stich, "UCID: An uncompressed color image database," *Proc. SPIE*, vol. 5307, pp. 472–480, Dec. 2004.

[53] T. Gole and R. Böhme, "Dresden image database for benchmarking digital image forensics," in *Proc. ACM Symp. Appl. Comput.*, Mar. 2010, pp. 22–26.
[54] United States Department of Agriculture. *Natural Resources Conservation Service Photo Gallery*. Accessed: 2002. [Online]. Available: http://photogallery.nrcs.usda.gov
[55] P. Bas, T. Filler, and T. Pevny, "Break our steganographic system: The ins and outs of organizing boss," in *Proc. Int. Workshop Inf. Hiding*. Springer, 2011, pp. 59–70.
[56] Y. Jia *et al.* (Jun. 2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/-1408.5093

WUYANG SHAN received the B.S. and M.S. degrees in packaging engineering from the School of Packaging and Material Engineering, Hunan University of Technology, Zhuzhou, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the School of Printing and Packaging, Wuhan University. His current research interests include multimedia forensics and security and data hiding.



YAOHUA YI received the Ph.D. degree from Wuhan University, in 2004, where he is currently a Professor with the School of Printing and Packaging. He has published more than 50 papers in refereed conferences and journals. His research interests include image intelligent processing, information extraction technology, imaging quality detection, and analysis technology.

JUNYING QIU received the M.S. degree in art from the Institute of Fashion, Wuhan Textile University, Wuhan, China, in 2016. She is currently a Professional Teacher with the Institute of Fashion and Art Design, Sichuan Normal University. She also manages the scientific research work of the Institute of Fashion and Art Design and is an excellent young Teacher. Her research interests include digital art media intelligent processing, analysis, and security.

AIGUO YIN received the bachelor's degree from the Huazhong University of Science and Technology, in 1991. He is currently a Vice General Manager of Zhuhai Pantum Electronics Co., Ltd. His research interests include image processing and electronic photograph technology.