Proceedings of the 2018 International Conference on Machine Learning and Cybernetics, Chengdu, China, 15-18 July, 2018

DETECTION AND CLASSIFICATION OF VARIOUS IMAGE OPERATIONS USING DEEP LEARNING TECHNOLOGY

TIAN HUANG, XIAOCHEN YUAN*

Faculty of Information Technology
Macau University of Science and Technology
Macau S.A.R., China
E-MAIL: ht1072469004@gmail.com, xcyuan@must.edu.mo

Abstract:

As one of the main medium for information transmission, the digital image can be easily tampered during transmission. It is becoming more and more important to identify whether the given image is an original image or a processed image. In this paper, we propose a compact universal feature based on spatial domain in virtue of some latest image forensic methods and design a multi-class classification scheme using the deep learning technique, to identify and furthermore classify the various normal image operations. According to the experimental results, the proposed method can well detect and classify the multiple common image post-processing operations. And the comparison with the existing feature shows the better performance of the proposed method.

Keywords:

Image post-processing Operations detection; Spatial rich model (SRM); Deep learning technique; Convolutional neural network (CNN)

1. Introduction

The post-processing technique of digital image gains fast development, which ensures an easy modification of digital images without being perceptible. At present, digital image forgeries appear extremely rampant in people's daily work and life. If forged images are used in social media, scientific discovery or law, it will cause a serious impact on political and social stability. Therefore, people pay ever-increasing attention to image forensics [1].

A large number of forensics methods have been proposed so far, like detecting JPEG compression history [2-5], exposing contrast enhancement [6, 7], resampling [8-10] and median filtering [11-14], revealing image splicing [15, 16], identifying frequency domain filtering [17] and so on. However, most of these state-of-the-art features just consider specific operation and only do binary classification. For example, in [2-5], the author worked on JPEG compression together with artifacts which came from image coding based on block. In [7], the author proposed a blind forensic algorithm to detect how to achieve digital image modification in virtue of universal contrast enhancement operations. In [8], this method was set forth on the basis of a derivative operator as well as the radon transformation capable of detecting the resampling traces and interpolation. In [18], a deep learning-based method was proposed to detect the copy-move forgeries in images. Moreover, these features exhibit low reasonability and feasibility, given that processing types of detected image are unknown in most cases. For instance, when a median filtered image is fed into a classifier to detect the Gamma Correction, it could be classified as original image or Gamma corrected image, which is totally wrong, and the feature employed in this method is specially designed for Gamma correction detecting. If we apply this classifier to identifying other image operations, the accuracy will drop significantly. Hence, it is very essential to propose an effective forensics method that is able to simultaneously identify various image operations.

In past decades, several works have considered on such an issue. Fridrich et al. [19] proposed the Spatial Rich Model (SRM) steganalysis feature. The proposed SRM employed multiple linear and non-linear filter to generate residual, then extracted the high order feature and merged it into SRM feature. Multiple sets of filter which were selected by the experiments, can be used to extract the residual information under different models from the image. This can magnify the noise caused by embedding to

* Corresponding author. This work is supported by Science and Technology Development Fund of Macau SAR under Grant 051/2016/A2.

978-1-5386-5214-5/18/$31.00 ©2018 IEEE

a certain extent and it is advantageous to the feature extraction. In [20], the authors proposed a method based on SRM feature. Firstly the dimensionality of SRM feature was reduced, then the improved feature was employed to train the ensemble classifiers to classify the various common image post-processing operations. Although the method exhibits a better performance than the most of other state-of-art methods, the performance presents an obvious downtrend with the decrease in the size of observed images and meanwhile will not be optimized iteratively. What's more, features which are manually designed may not exhibit enough effectivity to identify various image operations due to large complexity.

In order to avoid the above mentioned complexity and to identify the various image operations effectively, we propose a novel approach in this paper to detect and identify the various image operations. The proposed approach is able to get image feature representations in an automatic manner on the ground of the deep learning framework. Major contributions we have made are as follows: (1) The *Sub_SRM* feature from the original images and the processed images are firstly extracted, which contributes to the improvement of acceleration of network convergence. (2) The 5-layer Conventional Neural Networks (CNN) is employed, by scanning extracted *Sub_SRM* feature sets in virtue of a patch-sized sliding window. Feature fusion is adopted to aggregate CNN features in order to get discriminated feature of an image. (3) A Softmax classifier is trained to conduct detection and classification. The structure of the rest part of the paper is shown as follows. Section II present the proposed method for detection and classification of various image operations. Section III gives the experimental results and analysis. And Section IV concludes this paper.

2. Proposed Approach for Detection of Various Image Operations

Fig. 1 shows the framework of the approach proposed which is constituted of two major steps, feature extraction and feature learning. To train the original images, various operations including spatial enhancement, spatial filtering, frequency filtering, and lossy compression, are applied to the original images, thus producing the corresponding processed images. Then the *Sub_SRM* is employed to extract the feature sets. During the feature learning, we employ a 5-layer CNN to extract the CNN features from the extracted *Sub_SRM* feature sets. The *Sub_SRM* feature sets are scanned using a patch-sized sliding window. Feature fusion is adopted to aggregate CNN features thus obtaining the discriminated feature off an image. Afterwards, the Softmax classifier of CNN is trained for detection and classification of various operations. By employing *Sub_SRM* feature as input of CNN instead of pixel values in common strategies, the proposed approach enhances the generalization ability as well as fastens the network convergence.

2.1. Feature Extraction with Sub_SRM

The general steps of the *Sub_SRM* feature extraction is: 1) Residuals Computation; 2) Quantization and Truncation; and 3) Co-occurrences Calculation. The details of each of the steps are as follows.

To calculate the residuals, given the input image, for each pixel, firstly the product of its neighboring pixels and the corresponding elements of filter matrix are calculated, followed by the calculation of the accumulative sum as residual value of the pixel position. Then, by traverse the host image with the filter, the residual value of each pixel in the image can be calculated which accordingly forms the residual map. For two-dimensional image filtering and convolution operation, if the filter exhibits rectangular symmetry, the same image with the same filter can get the same results. Therefore, we obtain the calculation method of the residual map by means of convolution here.

Given the host image $I(a,b)$ and a filter f, and the residual value calculation can be expressed using Eq. (1).

$$R_{a,b}=I\otimes f=\sum_{i=-M}^{M}\sum_{j=-N}^{N}f(i,j)\cdot I(a-i,b-j) \quad (1)$$

where R is the residual map calculated from the host image I and filter f; $M=\lfloor A/2 \rfloor$, $N=\lfloor B/2 \rfloor$, and A and B indicates size of the host image; and $\otimes$ denotes convolution process.

By multiplying and summing the corresponding pixel and its adjacent pixels by the filter coefficients, the corresponding residual of the pixel can be obtained. It should be noted that for the image edge pixels, the pixel value may not be adequately changed accordingly with the filter coefficients. Therefore, before applying filtering operations, mirror symmetric filling operations can be conducted to images, that means the images are filled adjacently by border pixel values. After this operation, the size of the residual map will not be changed.

With the residual map, the quantization and truncation process can be formed using Eq. (2).

$$R_{q,hT}=\text{trunc}_{hT}\left(\left\lfloor \frac{R}{q\cdot c}+\frac{1}{2}\right\rfloor\right) \quad (2)$$

where $R_{q,hT}$ is the residual map after truncation and quantization; c is the absolute value of the central position coefficient of the filter; q denotes the self-defined quantization factor; hT denotes the self-defined truncation coefficients; and $\text{trunc}_{hT}(\cdot)$ indicates the truncation function, which is defined in Eq. (3).

$$\text{trunc}_{hT}(a)=\begin{cases} a, & \text{if } (|a|<hT) \\ hT, & \text{if } (|a|\geq hT) \end{cases} \quad (3)$$

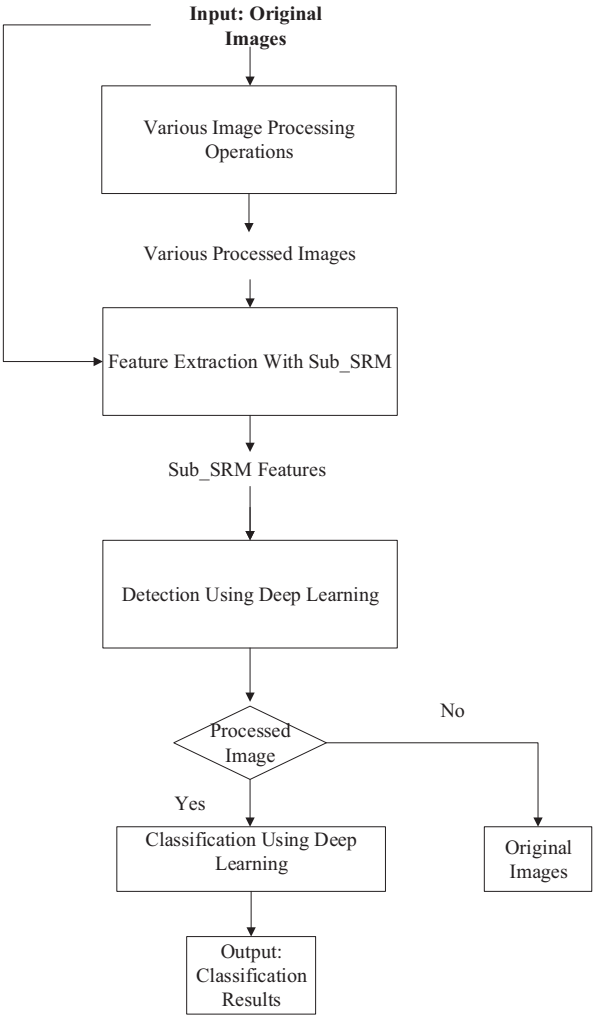

FIGURE 1. Framework of the proposed approach.

Subsequently, the forth-order Co-occurrence Matrix features can be drawn from the residual maps $R_{a,b}$ after the quantization and truncation operations. One co-occurrence distribution or matrix is that defined over an image, to be the co-occurring pixel values' distribution at a given offset, here is the way to extract *Sub_SRM* feature sets:

$$C^{SRM}_{d_0d_1d_2d_3}=\sum_{a,b=1}^{n_1,n_2-3}\left[R_{a,b+k}=d_k,\forall k=0,\ldots,3\right] \quad (4)$$

where $[\cdot]$ is the sign function of value 1 when the content is true; and otherwise, it is 0; and $d_k \in \{-hT,-hT+1,...,hT\}$.

Same as in [19], hT is set to be 2. Therefore, each co-occurrence matrix C is a four-dimensional array index with $d=[d_0,d_1,d_2,d_3] \in \{-hT,...,hT\}^4$, which gives the array with $(2hT+1)^4=625$ elements.

Finally, symmetrical combining operations will be carried out to Co-occurrence Matrix features obtained so as to acquire the SRM features. In this way, the *Sub_SRM* feature sets are extracted from the host image, and the dimensionality of the extracted *Sub_SRM* feature is of 1×325.

2.2. The Methodology of Classification

With the extracted *Sub_SRM* feature, the 5-layer CNN, includes two convolutional layers, two pooling layers and one fully-connected layer followed by one Softmax classifier, is employed for further detection and classification of the various image operations. Fig. 2 shows the architecture of our employed CNN.

In Fig. 2, the *Input layer* of the employed CNN is used to input the initial data. In common method, pixel values of the images will be the input data; rather than the conventional method, in our method, we apply the *Sub_SRM* feature of 20×20 extracted from the host images as input. Using features instead of the pixels contributes to the improvement of generalization ability and the acceleration of network convergence. In the previous stage, the dimensionality of the generated *Sub_SRM* feature is 1×325, thus we employ padding operation to meet the requirement.

The subsequent *convolutional layers* aim to extract the feature maps. There are two convolutional layers in our employed CNN model, as displayed in Fig. 2. The *convolutional layer 1* have six kernels with receptive filed 5×5, and each feature map is of size 16×16; while the *convolutional layer 2* have twelve kernels of size 5×5, and each feature map is of size 4×4.

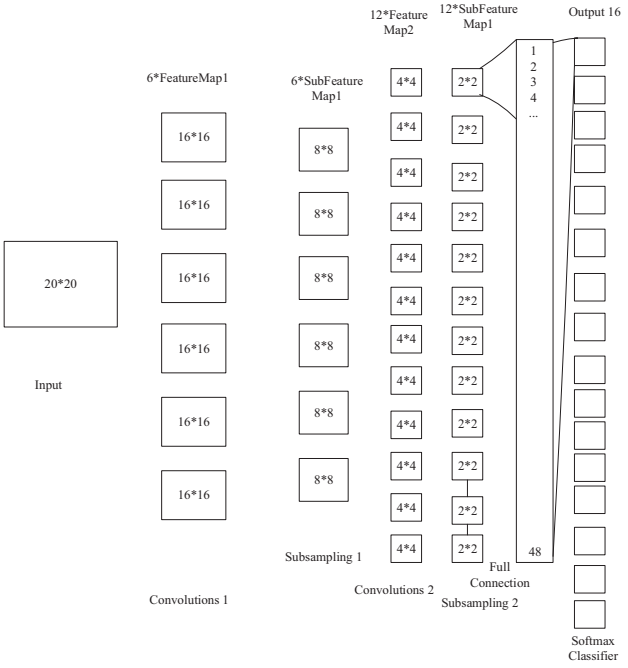

FIGURE 2. Architecture of our employed CNN.

As an important component of the CNN, *Sub-sampling layer* is usually between the two convolutional layers, as shown in Fig. 2. *Sub-sampling layer* reduces the calculation complexity through the reduction of the connections between the two *convolutional layers*. Our employed CNN model contains two Max-sampling layers, the two layers both have one kernel of size 2×2 which resamples the input spatially and reduces 75% of the activations.

After the two *convolutional layers* and the two *sub-sampling layers*, there is one *full connection layer* followed by a Softmax classifier in our employed CNN. The *full connection layer* which is connected with each neuron in the former layer, transforms the feature map extracted from the former layer into a vector of size 1×48, and feeds this vector into Softmax classifier to perform classification.

3. Experiments and results

In our experiments, thousands of images are randomly selected from the dataset Boss Base v1.0 [21]. All of the images are randomly divided into two types: half of them are utilized in training and the other in testing. In the following experiments, for each original image, seven different image operations are applied and the corresponding processed images are created. Table 1 gives the tested seven image operations and the range of the parameters of each of the corresponding operations.

TABLE 1. Types of tested image processing operations.

Type of operation		Parameters
Spatial enhancement	Gamma correction (GC)	γ:1.0,1.2,1.4,1.6,1.8,2
	Histogram equalization (HE)	n/a
Spatial filtering	Mean filtering (MeanF)	Windows size: 3×3,5×5,7×7
	Gaussian filtering (GF)	Windows size: 3×3,5×5,7×7; σ:0.8-1.6
Lossy compression	JPEG	Quality:75-90
Frequency filtering	Low-pass filtering (LPF)	Cut-off frequency: 80 (Hz)
	High-pass filtering (HPF)	Cut-off frequency: 30 (Hz)

To achieve the balance between the computation expenses and the detection accuracy, we set the number of the iterations to be 200 by experiment, that is, $num\_epochs = 200$; and by comparison, we select the pool type as *Max-pooling*. In the following, Table 2 shows the accuracies of the various operations detection using the Sub_SRM feature. It can be easily seen that the detection accuracies are above 98% for all the tested operations, which indicates that our method is very successful for detecting various image post-processing operations.

TABLE 2. Accuracy of operations detection using *Sub_SRM* feature.

Operations	GC	HE	MeanF	GF	JPEG	LPF	HPF
Accuracy	98.3%	99.7%	98.7%	99.8%	99.7%	99.9%	99.1%

In addition to detecting whether the images have been processed, this method further classify the various operations. Table 3 shows the confusion matrices of the multi-class classification using the *Sub_SRM* feature. In

Table 3, the '*' indicates that the predicted percentage is under 1.0%, that means the classification can be ignored. The results in the diagonal show the multi-class classification results. According to the results in Table 3, the average identification accuracy with *Sub_SRM* feature is calculated as 97.25%. The good results indicate the effectiveness of our method.

Furthermore, we have compared our employed *Sub_SRM* feature with the *SPAM* feature proposed in [22]. Table 4 shows the comparison results of employing the *Sub_SRM* and *SPAM* to detect and classify the various type of image post-processing operations. In addition, the results of using *Mean-polling* and *Max-polling* are respectively given, indicating the better performance of *Max-polling*. It can be easily seen that the *Sub_SRM* feature outperforms the *SPAM* feature in all the tested operations.

TABLE 3. Confusion matrix for image operation identification using *Sub_SRM* feature with 200 epoch.

Real/ Predicted	Ori	GC	HE	MeanF	GF	JPEG	LPF	HPF
Orig	**96.3 %**	0.4%	*	0.1%	*	3.1%	*	0.1%
GC	2.5 %	**90.8 %**	0.1%	0.1%	*	0.5%	*	6%
HE	*	0.4%	**99.5 %**	*	*	0.1%	*	*
MeanF	0.6 %	*	*	**99.4 %**	*	*	*	*
GF	0.1 %	0.1%	*	*	**95.0 %**	0.1%	4.3%	0.4%
JPEG	*	0.4%	*	*	0.4%	**98.7 %**	0.3%	0.2%
LPF	*	0.1%	*	*	0.1%	*	**99.8 %**	*
HPF	*	0.3%	*	*	*	0.3%	*	**99.4 %**

TABLE 4. Comparison between the proposed method using Sub_SRM feature and using the SPAM feature.

Feature	*SPAM* [22]		*Sub_SRM*	
Pooling Type	*Mean-polling*	*Max-polling*	*Mean-polling*	*Max-polling*
Ori	84.6%	91.2%	92.7%	**96.3%**
GC	84.5%	90.4%	88.6%	**90.8%**
HE	96.1%	98.4%	98.8%	**99.5%**
MeanF	60.3%	61.2%	99.1%	**99.4%**
GF	88.2%	90.4%	95.8%	**95.0%**
JPEG	72.5%	72.8%	97.0%	**98.7%**
LPF	85.3%	90.5%	99.2%	**99.8%**
HPF	90.9%	91.2%	98.7%	**99.4%**

4. Conclusions

In this paper, we propose to employ the spatial rich model *Sub_SRM* feature and the deep learning technique CNN to detect and classify the various image post-processing operations. The major contributions of this study are listed as follows: (1) the binary classifications and multi-class classifications of various image post-processing operations are put into consideration in our study. The extensive experiments which are evaluated on the various image processing operations indicate the effectiveness of the proposed method. (2) Instead of pixel values, we choose *Sub_SRM* feature as our CNN input which enhances the generalization ability and promote the network convergence. (3) Comparing with the conventional method using classifier such as SVM, we propose to employ 5-layer CNN as the classifier. The experimental results indicate the good performance of the propose method. In our future works, more image operations will be tested, different image features will be considered and the current employed CNN structure will be improved for a better identification results in this application.

Acknowledgment

This work is supported by Science and Technology Development Fund of Macau SAR under Grant 051/2016/A2.

References

[1] H. Farid, "Digital image forensics," *Scientific American,* vol. 298, p. 66, 2008.

[2] Z. Fan and R. L. D. Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,* vol. 12, pp. 230-235, 2003.

[3] H. Farid, "Exposing digital forgeries from JPEG ghosts," *IEEE Transactions on Information Forensics & Security,* vol. 4, pp. 154-160, 2009.

[4] W. Luo, J. Huang, and G. Qiu, "JPEG Error Analysis and Its Applications to Digital Image Forensics," *IEEE Transactions on Information Forensics & Security,* vol. 5, pp. 480-491, 2010.

[5] T. Bianchi and A. Piva, "Detection of Nonaligned Double JPEG Compression Based on Integer Periodicity Maps," *IEEE Transactions on Information Forensics & Security,* vol. 7, pp. 842-848, 2012.

[6] G. Cao, Y. Zhao, R. Ni, and X. Li, "Contrast Enhancement-Based Forensics in Digital Images," *IEEE Transactions on Information Forensics & Security,* vol. 9, pp. 515-525, 2014.

[7] M. Stamm and K. J. R. Liu, "Blind forensics of contrast enhancement in digital images," in *IEEE International*

Conference on Image Processing, 2008, pp. 3112-3115.

[8] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on Signal Processing,* vol. 53, pp. 758-767, 2005.

[9] B. Mahdian and S. Saic, "Blind Authentication Using Periodic Properties of Interpolation," *IEEE Transactions on Information Forensics & Security,* vol. 3, pp. 529-538, 2008.

[10] L. Li, J. Xue, Z. Tian, and N. Zheng, "Moment feature based forensic detection of resampled digital images," pp. 569-572, 2013.

[11] C. Chen, J. Ni, and J. Huang, "Blind Detection of Median Filtering in Digital Images: A Difference Domain Based Approach," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,* vol. 22, pp. 4699-710, 2013.

[12] X. Kang, M. C. Stamm, A. Peng, and K. J. R. Liu, "Robust median filtering forensics based on the autoregressive model of median filtered residual," in *Signal & Information Processing Association Summit and Conference*, 2012, pp. 1-9.

[13] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Media Forensics and Security II*, 2010, p. 754110.

[14] H. D. Yuan, "Blind Forensics of Median Filtering in Digital Images," *IEEE Transactions on Information Forensics & Security,* vol. 6, pp. 1335-1345, 2011.

[15] Y. Q. Shi, C. Chen, G. Xuan, and W. Su, "Steganalysis Versus Splicing Detection," *Digital Watermarking,* vol. 5041, pp. 158-172, 2007.

[16] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Pattern Recognition,* vol. 45, pp. 4292-4299, 2012.

[17] X. Zhao, S. Wang, S. Li, and J. Li, "Passive Image-Splicing Detection by a 2-D Noncausal Markov Model," *IEEE Transactions on Circuits & Systems for Video Technology,* vol. 25, pp. 185-199, 2015.

[18] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *IEEE International Workshop on Information Forensics and Security*, 2017, pp. 1-6.

[19] J. Fridrich and J. Kodovsky, "Rich Models for Steganalysis of Digital Images," *IEEE Transactions on Information Forensics & Security,* vol. 7, pp. 868-882, 2012.

[20] H. Li, W. Luo, X. Qiu, and J. Huang, "Identification of Various Image Operations Using Residual-based Features," *IEEE Transactions on Circuits & Systems for Video Technology,* vol. PP, pp. 1-1, 2016.

[21] P. Bas, ""Break our steganographic system": the ins and outs of organizing BOSS," in *International Conference on Information Hiding*, 2011, pp. 59-70.

[22] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by Subtractive Pixel Adjacency Matrix," *IEEE Transactions on Information Forensics & Security,* vol. 5, pp. 215-224, 2010.