SCIENCE CHINA
Information Sciences

• LETTER •

March 2020, Vol. 63 139109:1–139109:3
https://doi.org/10.1007/s11432-018-9492-6

# Image processing operations identification via convolutional neural network

Bolin CHEN[1], Haodong LI[2], Weiqi LUO[1*] & Jiwu HUANG[2]

[1]*Guangdong Key Laboratory of Information Security Technology, School of Data and Computer Science, Sun Yat-sen University, Guangzhou* 510006, *China;*
[2]*College of Information Engineering, Shenzhen University, Shenzhen* 518052, *China*

Received 29 March 2018/Accepted 15 June 2018/Published online 10 February 2020

**Citation** Chen B L, Li H D, Luo W Q, et al. Image processing operations identification via convolutional neural network. Sci China Inf Sci, 2020, 63(3): 139109, https://doi.org/10.1007/s11432-018-9492-6

Dear editor,

In recent years, image forensics has attracted increasing attention, and many forensic methods have been proposed for identifying image processing operations. Until now, most existing methods have been based on hand-crafted features, and only one specific operation is considered in these methods. In many forensic scenarios, however, multiple classification for various image processing operations is more practical. In addition, for some image processing operations, it is difficult to construct effective hand-crafted features. In this study, we therefore propose a new method based on convolutional neural networks (CNNs) to adaptively learn discriminative features for identifying typical image processing operations. However, directly employ CNN structures in other fields (e.g., computer vision) for image forensics cannot obtain satisfactory performance. Thus, we carefully design the network structures (including the high-pass filter bank, the channel expansion layer, the pooling layers, and the activation functions) of the proposed method. Experimental results show that the proposed method outperforms the current best method, which is based on hand-crafted features, and three related methods based on CNNs for image steganalysis and/or forensics, achieving state-of-the-art results. Furthermore, we provide supplementary results to show the rationality and robustness of the proposed model.

*The proposed model.* The proposed model is illustrated in Figure 1. We assume that the model input is a gray-scale image with the size of $M \times M$ (with $M$ being a multiple of 32). Based on our previous study [1], the artifacts introduced by various image processing operations are easier to capture in the image residual domain. Thus, the proposed model firstly transforms the input image into residuals with four high-pass filters as shown in Figure A1. We then use a "channel expansion layer" to process the resulting residuals and increase the channel number of feature maps from 4 to 32. The resulting feature maps are then input into six similar and typical layer groups to obtain high-level features. Each of these six groups contains a convolutional layer and a pooling layer. The convolutional layer serves to double the channel number, while the pooling layer (except for the last one) downsamples the width and height of the feature maps by a factor of two. It should be noted that the pooling layer in the last group is different, as it downsamples the feature maps to one, along with image width and height, via the average pooling. This type of pooling layer is known as "global average pooling". In the proposed model, all of the convolutional layers are equipped with the TanH function. Moreover, all of the convolutional layers and pooling layers use the smallest-sized kernel with a center (i.e., of size $3 \times 3$) except for the global average pooling, since a small kernel size has fewer parameters, and this helps train the network faster and prevent overfitting.

* Corresponding author (email: luoweiqi@mail.sysu.edu.cn)

 

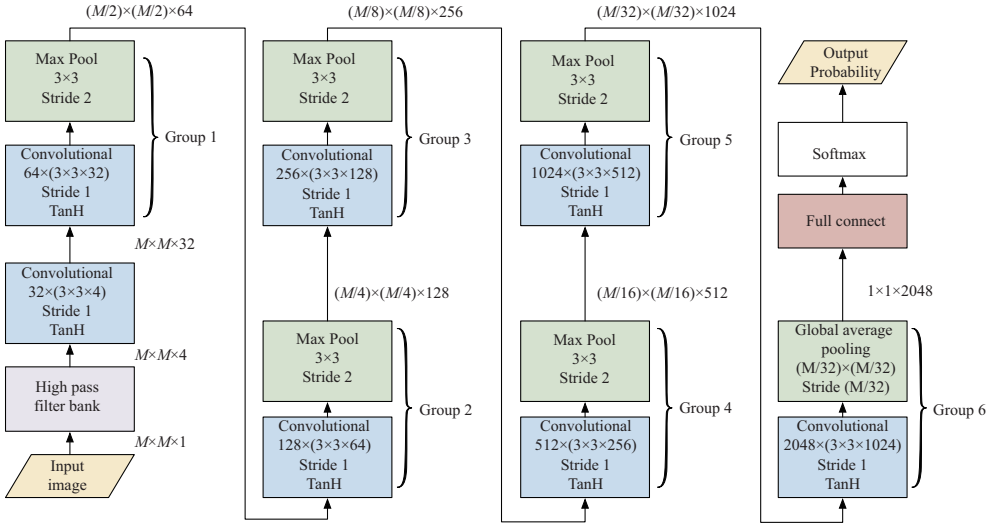


**Figure 1** (Color online) The architecture of the proposed model.

The proposed CNN-based model was implemented using the TensorFlow machine learning framework. Instead of the typical vanilla SGD, Nesterov momentum was used to train the network, since this learned faster and performed better based on our experiments. The momentum was set as 0.9, and L2 regularization was used. The corresponding weight decay was 0.0005. We initialized all the weights using a Gaussian distribution with zero mean and standard deviation of 0.01, and initialized all the biases as zero. In the training stage, the batch size was 64. Training set was shuffled between epochs. A step decay in learning rate was used in our method. The learning rate was initialized as 0.01, and it was divided by 10 when the validation accuracy stopped improving. Finally, the training was stopped after reducing the learning rate three times.

*Experimental setup.* We collected 40000 images in raw format from different cameras, and transformed them into gray-scale with a size of $512\times512$ as original images. Eleven typical image processing operations are considered, including Gamma correction (GC), histogram equalization (HE), unsharp masking sharpening (UM), mean filtering (MeanF), Gaussian filtering (GF), median filtering (MedF), Wiener filtering (WF), scaling (Sca), rotation (Rot), JPEG and JPEG 2000 (JP2). Following [1], these operations are performed on each original image with a random parameter selected from Table A1. After performing these operations, we cropped the $256\times256$ regions from the center of each image. Therefore, we obtained 12 classes of images (including the original), each of which contains 40000 images. In each experiment, we firstly divided the original images into a training data set (26000 images), a validation data set (4000 images), and a testing data set (10000 images). In the following experiments, we randomly divided the training, validation, and testing data three times and report the average results.

In order to demonstrate the effectiveness of the proposed model, we compare the proposed method with related work, including the current best method based on hand-crafted features [1], and three CNN-based methods including those of Chen et al. [2], Bayar et al. [3], and Xu et al. [4]. To achieve a fair comparison, the size of the input images needed to be the same for all methods. Thus, we modify the corresponding input layer and the fully connected layers due to the image size and/or the memory limitation, while preserving the other layers as they are for the other CNN-based methods.

*Validation of the designs.* Based on previous studies, most CNN-based models are highly dependent on the investigated problem. Thus, we conducted experiments to validate the rationality of the proposed model. Three parts of the proposed model have been considered, including the high-pass filter bank, the channel expansion layer, and the last pooling layer. In addition, we also considered the activation functions used in the proposed model; refer to Appendix B for details.

*Binary classification.* In this experiment, we try to identify whether a given image is original or modified by a certain image processing operation. The average detection results evaluated on testing data are shown in Table A2; we observe that all of the methods can obtain satisfactory results (larger than 91%) for all image processing operations except for GC. Overall, the proposed CNN-based method generally works better than the existing CNN-based methods (e.g., [2–4]), especially

for identifying GC, UM, and Sca. Compared with the current best method [1], we obtain similar results for all image processing operations. On average, the proposed CNN-based method outperforms the current best method [1] slightly for the binary classification task (around 0.5% improvement); this is shown in the final column of Table A2.

*Multiple classification.* In many forensic scenes, multiple classification is both more practical and more difficult compared with binary classification. In this experiment, we try to identify 11 typical operations as shown in Table A1. The average confusion matrices for the proposed method are shown in Tables A3; we observe that the proposed method can effectively identify most image processing operations. All values along the diagonal line of the confusion matrix are larger than 95%, and almost all values located on non-diagonal lines (i.e., false detection rates) are less than 1%. The false detection rates are larger for the GC and Sca, and this is consistent with the results for binary classification shown in Table A2. In addition, we show the average results along the diagonal values of the corresponding confusion matrices for the five methods in Table A4; we can observe that the proposed CNN-based method outperforms the other approaches, and improves on the current best method [1] by over 2%, which is a significant improvement on multiple classification for 11 image processing operations.

*Robustness analysis.* In order to validate the robustness of the proposed model, two other experiments are considered in this section: an evaluation of the performances of the proposed model for smaller images and an evaluation the pre-trained model for different data sources.

In the first experiment, we first cropped the center part of the images used previously, using three sizes: $128 \times 128$, $64 \times 64$, and $32 \times 32$. Then we trained different models for different sizes. The experimental results are shown in Table A5; we can observe that the performance of the five methods is reduced with a decrease in the image size due to insufficient statistics. However, the proposed model always outperforms the others, and the improvement is larger when the image size is small.

In the second experiment, we evaluated the proposed model using other testing image sources. To this end, we firstly collected 10000 images from BOSSbase 1.01 [5] and obtained 12 classes of images as described in experimental setup. Then we used the pre-trained model with the image set described in experimental setup to test the images from BOSSbase. The experimental results for multiple classification are shown in Table A6; we can observe that the accuracies drop slightly (less than 1% for all cases) compared with the results shown in Table A5, indicating that the generalization of the proposed model for image processing operation classification is very good.

*Conclusion and future work.* In this study, we have proposed a novel CNN-based method to identify 11 typical image processing operations. We carefully designed a CNN-based model for identifying various image processing operations and analyzed the influence of different network components. We conducted extensive comparative experiments to show that the proposed method can achieve state-of-the-art results compared with other methods. In addition, we provided experimental results to demonstrate the robustness of our model. In the future, we will extend the proposed method to detect images processed by multiple operations and identify the images processing order.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61672551, 61602318), Special Research Plan of Guangdong Province (Grant No. 2015TQ01X365), Guangzhou Science and Technology Plan Project (Grant No. 201707010167), Shenzhen R&D Program (Grant No. JCYJ20160328144421330), and Alibaba Group through Alibaba Innovative Research Program.

**Supporting information** Figure A1, Tables A1–A6, Appendix B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

**References**

1 Li H D, Luo W Q, Qiu X Q, et al. Identification of various image operations using residual-based features. IEEE Trans Circuits Syst Video Technol, 2018, 28: 31–45
2 Chen J S, Kang X G, Liu Y, et al. Median filtering forensics based on convolutional neural networks. IEEE Signal Process Lett, 2015, 22: 1849–1853
3 Bayar B, Stamm M C. A deep learning approach to universal image manipulation detection using a new convolutional layer. In: Proceedings of ACM Workshop on Information Hiding and Multimedia Security, Vigo, 2016. 5–10
4 Xu G S, Wu H Z, Shi Y Q. Structural design of convolutional neural networks for steganalysis. IEEE Signal Process Lett, 2016, 23: 708–712
5 Bas P, Filler T, Pevný T. "Break our steganographic system": the ins and outs of organizing BOSS. Inf Hiding, 2011, 6958: 59–70