Copy-Move Forgery Detection Based on Deep Learning

Junlin Ouyang, Yizhi Liu, Miao Liao
School of Computer Science and Engineering,
Hunan University of Science and Technology,
Xiangtan, China

Abstract—**Copy-move forgery detection (CMFD) is probably one of the most active research areas within the blind image forensics field. Among existing algorithms, most of them are based on block and key-point methods, or combination of them. Recently, some deep convolutional neural networks methods have been applied in the image classification, image forensic, image hashing retrieval, and so on, which have shown better performance than the traditional method. In the work, a novel copy-move forgery detection method based on convolutional neural network is proposed. The proposed method uses existing trained model from large database as ImageNet, and then adjusts slightly the net structure using small training samples. Experimental results show that the method we proposed obtains satisfactory performance to the forgery image generated automatically by computer with simple image copy-move operation**

Keywords- Tamper detection, copy-move forgery, image forensics, convolutional neural network.

I. INTRODUCTION

Copy-move forgery is the most common tamper which consists in copying one part of an image and then pasting in another part of the same image. Copy-move forgery detection (CMFD) is probably one of the most active research fields in blind image forensics. A large number of CMFD methods have been reported in the literature [1-5]. They can be roughly divided into two classes: key-point based [1-2] methods and block-based methods [3-5]. Scale-invariant feature transform (SIFT) and speeded up robust feature (SURF) are commonly used techniques by the first category methods. Features of key point are first extracted from the whole image. Each key point is then matched in terms of these features so as to find feature similar point. A forgery region may be detected if a clustering region generated by matching pairs with the same affine transformation is large enough. The key-point-based methods can efficiently find duplicated regions and achieve good performance to the geometric distortion such as rotation, scaling and translation. However, the drawback of the key-point based methods is that duplicated regions with little visual structures or key-points are hard to detect. Block-based methods divide the image into overlapping blocks, and extract some features from each block and look for block feature matching. These matching pairs are considered to be part of duplicated regions if the number of matching pairs with the same "shift vector" exceeds a certain threshold.

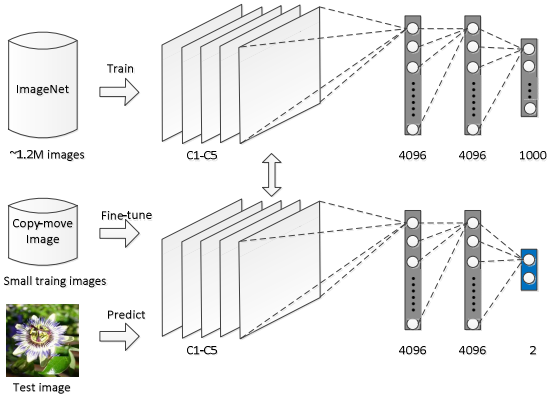

Fig. 1. The proposed image copy-move forgery detection framework based deep convolutional neural network. First the parameters of model are transferred from the existing trained model as ImageNet, and then are adjusted slightly the net structure using small copy-move training samples. Finally, the test image may be identified by the trained model.

In order to design a robust copy-scaling-move detection method, Ou et al. [6] proposed a three-dimensional detection model based on pyramid. The proposed method can detect the random tampering patch that is scaled, but is not robust to tampering patch that is rotation or the combination of rotation and scaling. Park et al. [7] proposed an up-sampled log-polar Fourier descriptor, which is robust to geometric transformations including rotation, scaling, and reflection, and applied it to the copy-move method. For the exiting problem between key-point based methods and block-based methods, Zheng et al. [8] proposed a fusion method. First, an adaptive initial size of region is segmented; then the key-points based on SIFT method are extract. The ratio of key-point is used to classify the region into smooth or non-smooth. For the smooth region, the block-based approach is used to detect the forgery region; otherwise, the key-points method is used. Bi et al. [9] proposed a feature extraction and hierarchical feature matching method using multi-level dense descriptor. The extracted dense descriptor consists of two parts: the Color Texture Descriptor and the Invariant Moment Descriptor. Experimental results

978-1-5386-1937-7/17/$31.00 ©2017 IEEE

show that the scheme can achieve much well results to common signal processing but large angle rotation and scaling.

Recently, Wang et al. [10] proposed a copy-move method that combined feature point matching technology and block matching technology. The detector based on Harris is used to extract image key-point, and the statistical features of key-point neighborhoods are looked as forensics features. A region growth technology and a mismatch checking approach are developed to improve detected accuracy. Li et al. [11] proposed a copy-move detection method that first segments the image into semantically independent patches, and then extracted key-point. In the matching process, the suspicious pairs of patches are roughly estimated an affine transform matrix, then the EM algorithm is used to refine the estimated matrix. Zandi et al. [12] proposed a new interest point detector by using the advantages of both block-based and traditional key-point based methods. A filtering algorithm is used to prune the falsely matched regions and the key-point density is adjusted by an iterative improvement strategy. Pun et al. [13] used over-segmentation and feature point matching and proposed a forgery detection method. First, an over-segmentation method is used to segment the image into some non-overlapping blocks; then, the feature points in each block are extracted, and a feature matching method based on block is used to detect the suspected forgery regions. Finally, the morphological operation is used to detect the more accurate forgery regions.

Convolutional Neural Network (CNN) is a well-known deep learning architecture, and is a great success in the task of image classification and recognition. In the digital image passive forensics field, many researchers try to deal with these problems by CNN. For instance, Kang et al. [14] first applied the CNN into the median image filter forensics. Baroffio et al. [15] applied it into camera type identification of image forensics. Belhassen et al. [16] proposed a general image tampering model. Xu et al. [17] applied the CNN to the image steganography analysis.

Inspired by these works, we will explore the application of image copy-move forgery detection by the CNN. The idea is that the proposed method uses existing trained model from large database as ImageNet, and then adjusts slightly the net structure using small training samples.

II. THE PREPOSED METHOD

An end-to-end deep learning architecture for image copy-move forgery detection is proposed in the section. In general, we obtain an image classification or recognition model by deep learning or CNN that needs thousands of images with labeled as image training sample. However, most existing copy-move forgery image databases may only contain several hundred or less ones. The technology of fine-tuning can solve this problem. It uses the existing train model from large train database as ImageNet, and then adjusts slightly the net structure using small training samples. The reason is that the training model from ImageNet has the very strong generalization ability, and it can achieve well performance when we apply ours small number of images on the model.

A. Model

Most existing fine-tuning model from ImageNet training database includes AlexNet, VGGNet, GoogLeNet, ResNet and so on, as shown in Table I.

Table I The number level of different model

Model	AlexNet	VGG	GoogLeNet	ResNet
Year	2012	2014	2014	2015
Number of level	8	19	22	152

From the Table I, it can be see that the number of level of AlexNet model is eight; other model has larger number of level. Especially, the model of ResNet is up to 152 levels. The number of level of these models is larger, the performance of them is better, but their construction is more complex, and they need more memory and training time. In this work, we use the pre-trained model proposed by Krizhevsky et al. [20] from the Caffe CNN library. The parameter configuration of the model is shown in Table II.

To better understand the role of the used layers, some operators are descripted as follows:

Table II Parameters configuration of the AlexNet model

	Convolution				RELU	Pooling		Normal /Dropout
	Kernel size	stride	pad	group		size	stride	RLN
C1	11×11	4	0	1	Yes	3×3	2	5×5
C2	5×5	1	2	2	Yes	3×3	2	5×5
C3	3×3	1	1	1	Yes	-	-	-
C4	3×3	1	1	2	Yes	-	-	-
C5	3×3	1	1	2	Yes	3×3	2	-
C6	Inner Product				Yes	-	-	Drop6
C7	Inner Product				Yes	-	-	Drop7
C8	Output:1000							

- *Convolution*: the convolution layer is a core of convolution neural Network. The kernel of convolution is a filter, and also is called local receptive field, which can reduce the number of neuron because it is impractical to connect neurons to all neurons in the previous neural Network.
- *RELU:* The rectifier function, f(x) = Max(0, x), is an activation function which can be used by neurons. The rectifier activation function is used instead of a linear activation function in order to add non linearity.
- *Pooling:* A pooling operator deals with individual feature channels, coalescing nearby feature values into one. It can reduce the number of neurons and computation complexity. At the same time, it also is a nonlinear operation. Common pooling operators include max-pooling or sum-pooling. For instance, max-pooling is defined as:

$$y_{ijk} = \max\{y_{i'j'k} : i \le i' < i+p, j \le j' < j+p\}$$

- *Normal:* The Normal operator normalizes the feature channels vector of each spatial location, and is channel-wise normalization. The form of the normalization operator is defined as:

$$y_{ijk'} = \frac{x_{ijk}}{(\kappa + \alpha \sum_{k \in G(k')} x_{ijk}^2)^{\beta}}$$

- *Dropout:* In the training process of the inner-product, half of the nodes are ignored. The reason is that it prevents inter-dependencies from emerging between nodes; this allows the network to avoid over fitting and to learn a more robust relationship. The dropout operation has much the same performance, but less in both time and storage required.

B. Method

To achieve a model of copy-move forgery detection, the proposed method performs the following step:

1. Build copy-move forgery image database. The number of forgery images should have 10000 ones in order to fine-tune the CNN. However, the number of forgery images is very small for most existing forgery image database built by handcraft in real scenario. Moreover, it is time-consuming and tedious work if we build the forgery images contained about 10000 ones by handcraft. Therefore, we simply simulated the copy-move tampering operation using the rectangle block. The forgery images are generated by moving the rectangle block from the upper left corner to the center randomly. The specific process reference experimental section.
2. Initialize the CNN network. We initialize the parameters of CNN based on Caffe architecture. The weight coefficients of CaffeNet model, which have been trained by ImageNet, are transferred into the initialized CNN. Then, the output number of network is modified into two kinds, i.e., the original image or tampering one.
3. Fine-tuning the CNN network. The images dataset including training images, validation one, and the test ones are inputted into the CNN. In the process of running, some parameters should be adjusted according to the accuracy of prediction, for example, study ratio, the size of batch, and so on.
4. Identify image. If the training process has been completed from the upper fine-tuning step, the identification results can be obtained by inputting test image into the obtained training model.

III. EXPERIMENTAL RESULT

In this section, we first introduce the datasets including UCID [18], OXFORD flower [19], and CMFD [5]. And then, we present our experimental results. To better understand these experimental, some results of visualization on CNN also are presented. Finally, we analyze the reasons and give the future work.

A. Datasets

The following experiments have been conducted in the UCID image dataset including 1338 color images. The second dataset is much larger from OXFORD flower, consisting of 8189 images and 102 different categories of flowers. The third dataset is CMFD image database, which contains 48 source images and various forged images.

For the image dataset UCID and OXFORD, we simply simulate copy-move operation. They rely on copying a portion of an image on the same image and moving it from the upper left corner to the bottom right corner as shown respectively in the first and second row of Fig. 2. The size and position of copied region are randomly selected, but they are in certain range. For each dataset of them, 10000 copy-move forgery images are built by computer. They are named as Data1 and Data2, respectively. Some samples of them are illustrated in the first and second row of Fig. 2. The CMFD image database is named as Data3 as illustrated in the third row of Fig. 2.

Fig. 2. Some examples of copy-move forgery image dataset. Top row: forged versions of images from OXFORD flower dataset, named as *Data1*. Second row: forged versions of images from UCID dataset, named as *Data2*. Third row: from CMFD dataset by handcraft in real scenario named as *Data3*.

B. Experimental results

We fine-tune the CNN by using the image datasets Data1 and Data2. For each dataset, 3000 forgery images are trained the network, 3000 forgery images are validated, and the rest of dataset are used as test set. In order to balance the ratio of original images and forgery ones, we add some original images by data argument method as translation, flip etc. The detection results are shown in Table III

Table III Performance comparison of detection results on different copy-move forgery datasets.

Datasets	Test Error (%)
Data1	2.32
Data2	2.43
Data1+ Data2	3.56
Data3	42

From the Table III, it can be seen that the Data1 and Data2 or combination of them obtain better performance for the fine-tuning model. However, when the trained model is applied into the dataset Data3, the performance is very poor. The reasons are that although the tampering operation is randomly for the

position and size within datasets Data1 and Data2, which are in a certain range. The trained model can recognize the single tampering operations, but not recognize the real scenario since their tampering operations are not trained by the model. If we want achieve better performance for the real scenario tampering, we need a large number of real scenario tampering images.

C. Visualization and Analysis

To better understand the nature of CNN work and analysis the experimental results, we give the feature map of CNN as shown in the Fig. 3-4. From the feature map, it can be seen that the feature map describe the feature of image content rather than the image tampering trace. Although the signal of tampering trace is very weak, the accuracy is high from the experimental results of Table III when the dataset is Data1 or Data2. This shows that the capacity of CNN in classifying the images with simple copy-move forgery operation is enough large. On the other hand, the performance is very low when we transfer the model of CNN from Data1 or Data2 to the tampering image of real scenario from Data3. This experimental results further approve that the CNN not always achieve good performance when it is directly applied others image classification tasks.

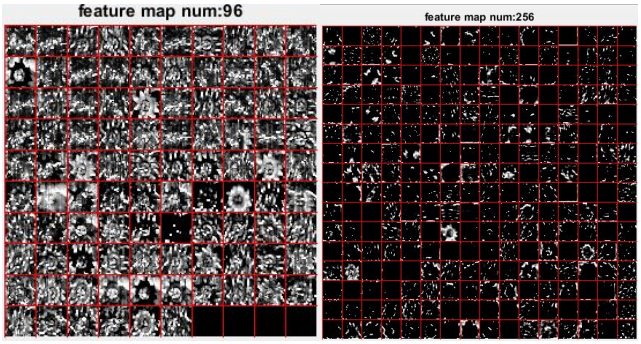

Fig. 3. Feature maps of visualization. Left: visualization result of first level convolution. Right: visualization result of second level convolution.

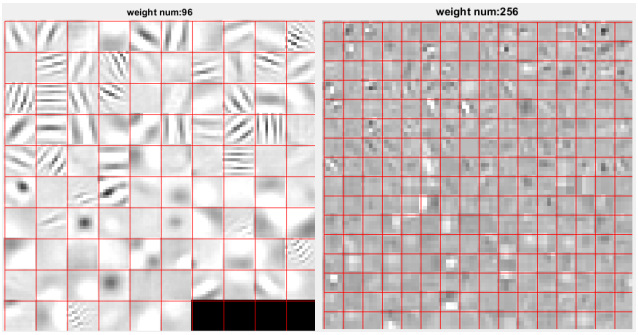

Fig. 4. Weight maps of visualization. Left: visualization result of first level convolution. Right: visualization result of second level convolution.

In order to achieve better performance for the copy-move forgery detection in future work, we can build large copy-move forgery dataset by handcraft to train the model of CNN. Unfortunately, it is almost impossible because the work is very time-consuming and tedious in order to generate various forgery images. Therefore, we will modify the structure of CNN by adding a level of feature transform in the future work. By doing so, the input of CNN is image tampering trace rather than image nature feature.

IV. CONCLUSION

A novel copy-move forgery detection method based on convolutional neural network was proposed. The proposed method uses existing trained model from large database as ImageNet, and then adjusts slightly the net structure using small training copy-move samples. Experimental results show that the proposed method obtains good performance to the forgery image generated automatically by computer with simple image copy-move operation, but is not robust to the copy-move forgery image of real scenario. We analyzed the reason of this case, and visualized the feature map of CNN. In the end, we gave some suggestions and methods to solve this issue. Although the proposed method is not perfect, it is first applied the CNN to copy-move forgery detection. Therefore, in order to achieve better performance in various real scenarios, the copy-move forgery detection method based on CNN deserves further research, and still has a long distance to go in the future work.

ACKNOWLEDGEMENTS

This work was supported by the natural science foundation of Hunan province under Grants 2017JJ2099, 2017JJ3091, by the Hunan province education department under Grants 16C0642, 17C0645, by the Doctor Fund University of Science and Technology of Hunan under Grants E51684, and by National Science and Technology Support Project of China, under grant number 2015BAF32B01.

REFERENCES

[1] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Transactions on Information Forensics and Security*, vol. 5, no.4, pp. 857-867, 2010.

[2] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo and G. Serra, "A sift-based forensic method for copy–move attack detection and transformation recovery," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099-1110, 2011.

[3] S.-J. Ryu, M.-J. Lee and H.-K. Lee, "Detection of copy-rotate-move forgery using Zernike moments," in *Proc. Information Hiding Conference*, 2010.

[4] H.-J. Lin, C.-W. Wang and Y.-T. Kao,"Fast copy-move forgery detection," *WSEAS Transactions on Signal Processing*, vol. 5, no. 5, pp. 188-1975, 2009.

[5] V. Christlein, C. Riess, J. Jordan and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches", *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1841-1854, 2012.

[6] J. L. Ouyang, J. Wu, G. Coatrieux , Z. Shao, , H.Z. Shu. "A roubst copy-scale-move forgery detection method based on pyramid model", Journal of southeast university (Natural science edition), vol. 45, no. 06, pp.1116-1120, 2014.
[7] C.-S. Park, C. Kim, J. Lee, and G.-R. Kwon, "Rotation and scale invariant upsampled log-polar fourier descriptor for copy-move forgery detection," Multimedia Tools and Applications, pp. 1-19, 2016.
[8] J. Zheng, Y. Liu, J. Ren, T. Zhu, Y. Yan, and H. Yang, "Fusion of block and keypoints based approaches for effective copy-move image forgery detection," Multidimensional Systems and Signal Processing, pp. 1-17, 2016.
[9] X. Bi, C.-M. Pun, and X.-C. Yuan, "Multi-Level Dense Descriptor and Hierarchical Feature Matching for Copy–Move Forgery Detection," Information Sciences, vol. 345, pp. 226-242, 2016.
[10] X. Wang, G. He, C. Tang, Y. Han, and S. Wang, "Keypoints-Based Image Passive Forensics Method for Copy-Move Attacks," International Journal of Pattern Recognition and Artificial Intelligence, vol. 30, no. 03, pp. 1655008, 2016.
[11] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," IEEE Transactions on Information Forensics and Security, vol. 10, no. 3, pp. 507-518, 2015.
[12] M. Zandi, A. Mahmoudi-Aznaveh, and A. Talebpour, "Iterative copy-move forgery detection based on a new interest point detector," IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2499-2512, 2016.
[13] C.-M. Pun, X.-C. Yuan, and X.-L. Bi, "Image forgery detection using adaptive oversegmentation and feature point matching," IEEE Transactions on Information Forensics and Security, vol. 10, no. 8, pp. 1705-1716, 2015.
[14] J. S. Chen, X. G. Kang, Y. Liu, and Z. J. Wang, "Median Filtering Forensics Based on Convolutional Neural Networks," IEEE Signal Processing Letters, vol. 22, no. 11, pp. 1849-1853, Nov, 2015.
[15] L. Baroffio, L. Bondi, P. Bestagini, and S. Tubaro, "Camera identification with deep convolutional networks," arXiv preprint arXiv:1603.01068, 2016.
[16] B. Bayar, "A Deep Learning Approach To Universal Image Manipulation Detection Using A New Convolutional Layer," Acm Workshop on Information Hiding & Multimedia Security, pp.5-10, 2016.
[17] G. S. Xu, H. Z. Wu, Y. Q. Shi. Structural Design of Convolutional Neural Networks for Steganalysis[J]. IEEE Signal Processing Letters, vol. 23, no.5, pp. 708-712, 2016.
[18] Schaefer G, Stich M UCID: an uncompressed color image database. Electronic Imaging, pp: 472-480, 2004.
[19] Automated flower classification over a large number of classes. Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing, 2008.
[20] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.