Digital image forgery detection using deep learning approach

Digital image forgery detection using deep learning approach

A Kuznetsov[1,2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: kuznetsoff.andrey@gmail.com

Abstract. This paper presents an algorithm for detecting one of the most commonly used types of digital image forgeries - splicing. The algorithm is based on the use of the VGG-16 convolutional neural network. The proposed network architecture takes image patches as input and obtains classification results for a patch: original or forgery. On the training stage we select patches from original image regions and on the borders of embedded splicing. The obtained results demonstrate high classification accuracy (97.8% accuracy for fine-tuned model and 96.4% accuracy for the zero-stage trained) for a set of images containing artificial distortions in comparison with existing solutions. Experimental research was conducted using CASIA dataset.

1. Introduction

Due to the rapid development of digital image processing technologies and the ever-growing popularity of digital recording devices, digital image processing became quite a simple operation even for an inexperienced user. Using modern software for digital images processing, any user can make changes to digital images in such a way that it will be almost impossible to visually distinguish a fake from the original data. In the past few decades, distorted digital images often appear in the media, and against this background, there are constant debates about the validity of the information provided. The fact of intentional change of information contained in the image in order to conceal it or distort it will be called artificial changes (distortions) or attacks.

The most common types of artificial distortions of digital images are copy-move, resampling and splicing. All of them are used to conceal or distort the information presented on the satellite image. The first type of distortion (embedding duplicates) means copying a fragment of a satellite image, introducing any distortion into this fragment (affine transformation, additive noise, re-quantizing brightness levels, etc.) and embedding the modified fragment into another area of the same image (that part of it, which must be hidden) [1,2]. The second popular type of artificial distortions is resampling [3] - affine transformation of digital image parts and embedding them into other images. The use of this type of distortion is relevant, for example, in the example described above for reducing the size of contaminants on the surface of water resources. The third commonly used type of distortions - splicing - is to use fragments of different satellite images to form a new satellite image or detailed distortion of an existing one (for example, forming a populated object copied from another satellite image on the territory of the forest fund) [4]. And finally, another way that attackers use is JPEG compression [5].

In this case, after embedding any information in the JPEG file and recompression, there are local differences in the properties of JPEG compression.

The described methods of embedding distortions are the most popular today, as evidenced by the huge number of publications aimed at developing solutions to detect such attacks [1-5].

2. Forgery detection algorithm

In this paper, to solve the problem of splicing 1detection, we use deep learning methods to classify images into two classes: original and containing distortion. The developed approach involves the analysis of the image in a sliding window and the classification of each image fragment corresponding to the position of the window. This approach makes it available to build discriminant functions directly on the basis of the data without any a priori knowledge about the process of extracting features. In this paper, we use convolutional neural networks (CNN), which in the past few years have achieved very good results in many computer vision applications, such as image classification, object recognition, image segmentation, face recognition, and many others. Training such networks is computationally very complicated and requires a huge amount of data. However, existing databases of images containing forgeries contain no more than a few thousand images, which is not enough to train or fine tune networks with complex architecture, such as, for example, VGG-16 [6] (Figure 1).

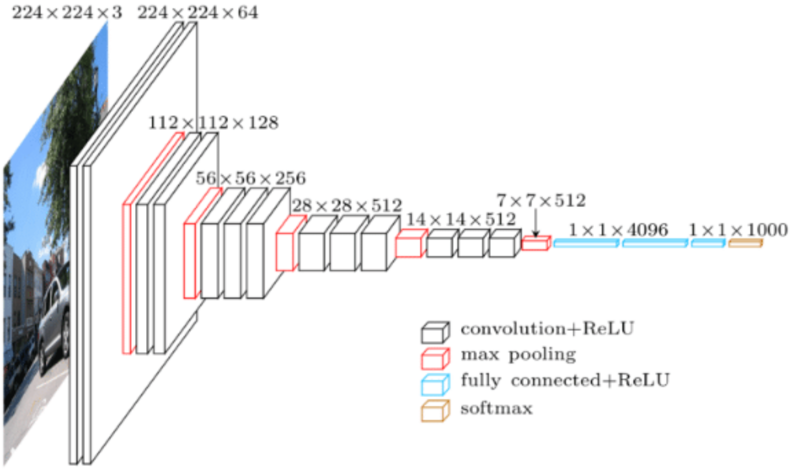

Figure 1. VGG-16 architecture.

To solve this problem, a classification strategy based on image fragments (patches) is proposed. This allows us to basically solve two problems: first, the lack of data for training. Secondly, it allows us to fix the size of the input data in accordance with the size of the patch, without using augmentation methods and without the use of additional geometric deformations. Moreover, since the CASIA dataset [7] (and similar available digital image forgery datasets) do not contain a segmented annotation, a simple and efficient method is used to automatically calculate the mask of the fake part based on the information contained in the CASIA original images.

The proposed model is a VGG-like convolutional network architecture [6]. It accepts patches of fixed size 40x40x3 as input signals and consists of two convolutional blocks and two fully connected blocks. Each convolutional block contains two convolutional layers with the activation ReLU function, followed by a pooling layer. All convolutional layers use 3x3 kernels, and the size of the pooling layer is 2x2. Between the various blocks, dropout layers [26] are used to solve the problem of overfitting. The normalization procedure is applied to the input data, which brings the input data values in the range [0, 1]. The total number of parameters that must be configured at the training stage in the proposed network is 869154.

It should be also noted that there was also implemented a fine-tuned model of the proposed network. COCO pretrained weights were used not to perform CNN training from zero-stage. Further we will compare the proposed network with other networks.

3. Experimental research

As part of the first experiment, the proposed approach was analyzed as a classifier algorithm for two classes: original and forgery. For this, the dataset of images was divided into training and test samples in the ratio of 80:20. For every image a set of patches were formed that correspond to original images and forgeries. The original patches were selected from original parts of images, whereas distorted patches were selected from the borders of the embedded areas. The size of the patch was selected as 40x40 pixels. Patches were selected from the image with half-patch overlap in 20 pixels. Patches pixels were normalized before training. The proposed network was trained for 300 epochs (Figure 2 shows the training process). At the testing stage, patches were extracted using the same methodology used for training, while the final decision on image classification is made by voting on the majority of patches of the first or second class.

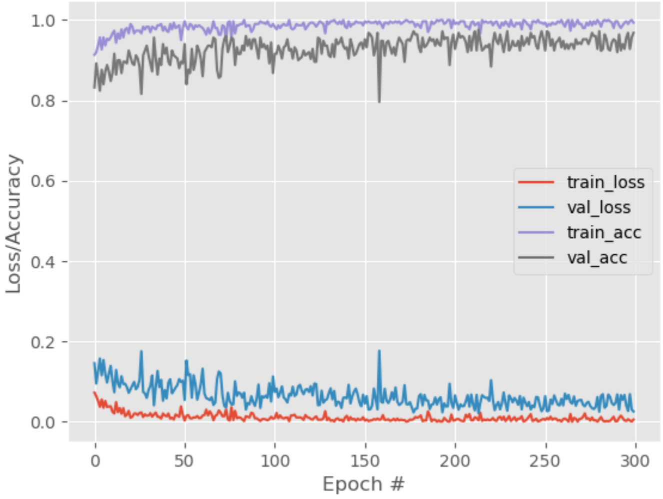

Figure 2. Training process visualization.

As a result of experiments, we obtained that the proposed CNN based solution has significantly higher accuracy for the zero-stage trained model and for the pretrained also. Table 1 presents the results of the first experiment and a comparison with some existing solutions, which are positioned as one of the best for solving the problem of splicing detection.

Table 1. Splicing detection algorithms comparison using CASIA v2 [7] dataset.

Method	Accuracy, %	Precision, %	Recall, %
Markovian rake transform [8]	79.74	-	-
Image chroma [9]	96.8	-	-
Markov chain [10]	95.6	-	-
DCT coefficients analysis [11]	90.1	-	-
The proposed approach	96.4	95.0	98.1
The proposed approach (fine-tuned)	97.8	97.1	96.8

During the second experiment, all images were additionally compressed by the JPEG algorithm to assess the effect of this type of post-processing of forgery images on the classification result. The results of the experiments are shown in Table 2.

Table 2. Dependency of the proposed CNN based solution accuracy on compression quality of forgeries using CASIA v2 [7] dataset.

Data	Accuracy, %	Precision, %	Recall, %
Initial	96.4	95.0	98.1
Compressed (Q=90)	67.1	78.1	46.3
Compressed (Q=80)	66.3	76.4	53.1

4. Conclusion

The article describes a new method for artificial distortions detection of digital images using the VGG-16 CNN. The results obtained showed a high quality of image classification (97.8% accuracy for fine-tuned model and 96.4% accuracy for the zero-stage trained) and the possibility of applying the method under conditions of repeated compression of distorted images by the JPEG algorithm in a narrow range. In the future, it is planned to conduct a detailed comparison with other methods of detection of splicing and to implement detection of distorted areas. We also plan to compare the proposed solution with Mobilenet and Resnet-50 CNN models.

5. References

[1] Cao Y, Gao T, Fan L and Yang Q 2012 A robust detection algorithm for copy-move forgery in digital images *Forensic Sci. Int.* **214** 33-43

[2] Kuznetsov A, Myasnikov V 2016 A Copy-Move Detection Algorithm Using Binary Gradient Contours *International Conference on Image Analysis and Recognition, ICIAR* **9730** 349-357

[3] Bayar B, Stamm M C 2017 On the robustness of constrained convolutional neural networks to jpeg post-compression for image resampling detection *Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing*

[4] Rao Y, Ni J 2016 A deep learning approach to detection of splicing and copy-move forgeries in images *IEEE International Workshop on Information Forensics and Security (WIFS)* 1-6

[5] Amerini I, Uricchio T, Ballan L, Caldelli R 2017 Localization of JPEG double compression through multi-domain convolutional neural networks *IEEE Conference on Computer Vision and Pattern Recognition Workshops* 1865-1871

[6] Simonyan K, Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556*

[7] CASIA Tampered Image Detection Evaluation Database 2010 URL: http://forensics.idealtest.org/casiav2/

[8] Sutthiwan P, Shi Y Q, Zhao H, Ng T-T and Su W 2011 Markovian rake transform for digital image tampering detection *Transactions on data hiding and multimedia security* **VI** 1-17

[9] Wang W, Dong J and Tan T 2009 Effective image splicing detection based on image chroma *ICIP. IEEE* 1257-1260

[10] Wang W, Dong J and Tan T 2010 Image tampering detection based on stationary distribution of markov chain *ICIP. IEEE* 2101-2104

[11] Lin Z, He J, Tang X and Tang C-K 2009 Fast, automatic and finegrained tampered jpeg image detection via DCT coefficient analysis *Pattern Recognition* 42(11) 2492-2501

Acknowledgements

The work was partly funded by the Russian Federation Ministry of Science and Higher Education within a state contract with the "Crystallography and Photonics" Research Center of the RAS under agreement 007-ГЗ/Ч3363/26 in part of "Introduction" and "Forgery detection algorithm", and Russian Foundation for Basic Research (Projects No. 19-07-00138 and 19-07-00474) in part of "Experimental research".