Holistic Image Manipulation Detection using Pixel Co-occurrence Matrices

Lakshmanan Nataraj1, Michael Goebel2, Tajuddin Manhar Mohammed1, Shivkumar Chandrasekaran1,2, B. S. Manjunath1,2
1Mayachitra Inc; Santa Barbara, CA
2University of California, Santa Barbara

Abstract

Digital image forensics aims to detect images that have been digitally manipulated. Realistic image forgeries involve a combination of splicing, resampling, region removal, smoothing and other manipulation methods. While most detection methods in literature focus on detecting a particular type of manipulation, it is challenging to identify doctored images that involve a host of manipulations. In this paper, we propose a novel approach to holistically detect tampered images using a combination of pixel co-occurrence matrices and deep learning. We extract horizontal and vertical co-occurrence matrices on three color channels in the pixel domain and train a model using a deep convolutional neural network (CNN) framework. Our method is agnostic to the type of manipulation and classifies an image as tampered or un-tampered. We train and validate our model on a dataset of more than 86,000 images. Experimental results show that our approach is promising and achieves more than 0.99 area under the curve (AUC) evaluation metric on the training and validation subsets. Further, our approach also generalizes well and achieves around 0.81 AUC on an unseen test dataset comprising more than 19,740 images released as part of the Media Forensics Challenge (MFC) 2020. Our score was highest among all other teams that participated in the challenge, at the time of announcement of the challenge results.

Introduction

Fake images and videos are ubiquitous in today’s world. These fake media are becoming a growing threat to information reliability. With the availability of powerful apps and tools, as well as recent Artificial Intelligence (AI) based automated technologies such as Deepfakes, it has become very trivial to generate fake manipulated images. The field of Digital image forensics aims to develop methods to automatically detect whether digital images have been tampered with or not.

There are many types of image forgeries such as splicing objects from one image to another, removing objects or regions from images, creating copies of objects in the same image, and more. To detect these manipulations, researchers have proposed several methods based on a variety of techniques such as JPEG compression artifacts, resampling detection, lighting artifacts, noise inconsistencies, camera sensor noise, and many more. However, most of these techniques focus on a specific type of manipulation or groups of similar manipulation operations while in a realistic scenario, several manipulation operations are applied when creating doctored images. For example, when an object is spliced onto an image, it is often accompanied by other operations such as scaling, rotation, smoothing, contrast enhancement, to name a few (Fig. 1). Very few studies address these challenging scenarios with the aid of Image Forensics challenges and competitions such as IEEE Image Forensics challenge [1] and the recent Media Forensics challenges [2, 3, 4]. These competitions try to mimic a realistic scenario and contain a large number of doctored images which involve several types of image manipulations. In order to detect these types of realistic tampered images, it is important to look into holistic image manipulation detection techniques which are not focused on detecting a single type of manipulation but can detect a host of different manipulations. In this paper, we focus on a holistic method to determine if an image has been tampered or not.

When an image is doctored, the statistics of the pixels in the doctored image are usually different than the untampered image. Hence, methods that look for deviations from natural image statistics could be effective in holistic detection of tampered images. These methods have been well studied in the field of steganalysis which aims to detect the presence of hidden data in digital images. One such method is based on analyzing co-occurrences of pixels by computing a co-occurrence matrix. Traditionally, this method uses hand-crafted features computed on the co-occurrence matrix and a machine learning classifier such as support vector machines determines if a message is hidden in the image [5, 6]. Other techniques involve calculating image residuals or passing the image through different filters before computing the co-occurrence matrix [7, 8, 9].

Inspired by steganalysis and natural image statistics as well as recent advances in deep learning, we propose a novel method to holistically detect if images are manipulated, using a combination of pixel co-occurrence matrices and deep learning. First, we compute horizontal and vertical co-occurrence matrices on the different color channels of an image. These capture the distribution of neighboring pixels horizontally and vertically. Then we stack these matrices and pass them directly through a deep learning framework and allow the network to learn important features...
that can separate tampered images from untampered images. We also avoid computation of residuals or passing an image through various filters, but rather compute the co-occurrence matrices on the image pixels itself. Recently, we showed that this approach is effective in detection of Generative Adversarial Network (GAN) generated images [10]. In this paper, we extend this approach to the more general case of detection of manipulated images. Experimental results on the evaluation dataset of the recently concluded Media Forensics Challenge (MFC) 2020 showed that our approach is promising and a viable method for holistic image manipulation detection. Our approach was the top performing approach in the challenge [4] (as shown in slide 47).

**Related Work**

Several methods have been proposed to detect digital image manipulations (see [11, 12, 13, 14, 15] for an overview). These included detection of splicing, resampling, copy-move, object removal, JPEG compression artifacts, machine learning and deep learning techniques, to name a few. Many techniques have been proposed to detect resampling [16, 17, 18, 19, 20], copy-move [21, 22], splicing [23, 24], seam carving [25, 26] and inpainting based object removal [27, 28] or artifacts arising from artificial intelligence (AI) generated images [29, 10, 30, 31, 32]. Several approaches exploit JPEG blocking artifacts to detect tampered regions [33, 34, 35]. Recent efforts in detecting manipulations exploit deep learning based approaches [36, 37, 38, 39, 40, 41, 42]. These include detection of generic manipulations [36, 37], resampling [43], splicing [38], reflection [44, 45] and bootleg [46]. Unlike most of these works, our approach combines classical statistical approaches with modern deep learning techniques by passing co-occurrence matrices through a deep learning framework, and then holistically detecting whether an image has been manipulated or not. In prior work [10], we had shown that this approach was effective in detecting images which were generated using a GAN. In this paper, we extend this approach to the more general case of holistic detection of manipulated images. Further, we compute both horizontal and vertical co-occurrence matrices and pass a 6-D tensor through a deep learning framework, and then holistically detecting whether an image has been manipulated or not. In prior work [10], we had shown that this approach was effective in detecting images which were generated using a GAN. In this paper, we extend this approach to the more general case of holistic detection of manipulated images. Further, we compute both horizontal and vertical co-occurrence matrices and pass a 6-D tensor through a deep learning framework, and then holistically detecting whether an image has been manipulated or not. In prior work [10], we had shown that this approach was effective in detecting images which were generated using a GAN. In this paper, we extend this approach to the more general case of holistic detection of manipulated images. Further, we compute both horizontal and vertical co-occurrence matrices and pass a 6-D tensor through a deep learning framework, and then holistically detecting whether an image has been manipulated or not.

**Methodology**

**Co-Occurrence Matrix Computation**

The first step is computing pixel co-occurrence matrices on different image channels and then passing these stacked matrices through a deep learning network. The co-occurrence matrices represent a two-dimensional histogram of pixel pair values in a region of interest. The vertical axis of the histogram represents the first value of the pair, and the horizontal axis, the second value. Eqn. 1 shows an example of this computation for a vertical pair.

\[
C_{i,j} = \sum_{m,n} \begin{cases} 
1, & I[m,n] = i \text{ and } I[m+1,n] = j \\
0, & \text{otherwise}
\end{cases}
\]  

(1)

Under the assumption of 8-bit pixel depth (grayscale images), this will always produce a co-occurrence matrix of size 256x256 for a single image channel. This is a key advantage of using co-occurrence matrices since it will allow for the same network to be trained and tested on a variety of images without re-training. Fig. 2 shows an example of an image and a co-occurrence matrix computed on an image channel. In this paper, we compute horizontal and vertical co-occurrence matrices on the red, green and blue channels of an image.

**End-to-end Learning**

Now, we describe our end-to-end system to holistically detect manipulations in digital images. Fig. 3 shows the block schematic of our system. First step is to compute the horizontal and vertical co-occurrence matrices on the RGB channels to obtain two 3x256x256 tensors. These are then stacked to obtain a 6x256x256 tensor. This tensor is then passed through a multi-layer deep convolutional neural network. We use a ResNet50 network [47] to train our model, with over 50 epochs, a batch size of 64, Adam optimizer, and cross-entropy loss. After each of the epochs, the model was evaluated on the validation set and the weights corresponding to the lowest validation loss were saved, and used for the remainder of the tests.

**Experiments**

**Datasets**

We evaluate our approach on datasets provided as part of the DARPA Media Forensics (MediFor) program [48], in which we were one of the participating teams. These datasets are further divided into Development datasets and Evaluation datasets, and broadly contain two classes of images: untampered and tampered images. The tampered images are further broken down by sub-classes depending on the manipulation operations applied on the images. These operations include splicing, object removal, copy-move, contrast enhancement, blurring, cropping, antiforensics, and social media laundering, to name a few. Later releases of the datasets also included Generative Adversarial Networks (GAN) based manipulations. More details on the different types of manipulations and how the datasets were curated can be found in [49].

**Development Datasets:** These datasets comprise various releases of tampered and untampered images between 2017 and 2019, and are mainly provided to the performers for the purpose of development of algorithms (summarized in Tab. 1). We use these datasets for training and validating our models and then test them on the evaluation datasets.

**Evaluation Datasets:** These datasets were provided to the teams as part of annual challenges held between 2017 and 2020. The labels of the images were not released at the time of challenge but released a few months after the challenge is completed. Tab. 1 summarizes the distribution of images as part of the annual chal-
Figure 3: An end-to-end framework to holistically detect tampered images.

Table 1: Summary of datasets provided for development (Dev) and evaluation (Eval)

| Dataset Name       | Dev/Eval | Total No. of images | No. of untampered images | No. of tampered images |
|--------------------|----------|---------------------|--------------------------|------------------------|
| MFC18-Dev1-Ver1    | Dev      | 6,708               | 997                      | 5,711                  |
| MFC18-Dev1-Ver2    | Dev      | 5,579               | 1,163                    | 4,416                  |
| MFC18-Dev1-Ver3    | Dev      | 5,557               | 1,162                    | 4,395                  |
| MFC18-Dev2-Ver1    | Dev      | 38,339              | 1,429                    | 36,910                 |
| MFC19-Dev1-Ver1    | Dev      | 3,592               | 3462                     | 130                    |
| NC17-Eval-Part1    | Eval     | 4,073               | 2,685                    | 1,408                  |
| MFC18-Eval-Part1   | Eval     | 17,406              | 14,143                   | 3,263                  |
| MFC19-Eval-Part1   | Eval     | 16,024              | 10,276                   | 5,7408                 |
| MFC20-Eval-Part1   | Eval     | 19,747              | 6,124                    | 13,623                 |

Table 2: Results on MFC 2020 Evaluation dataset

| Training Datasets       | AUC-ROC |
|-------------------------|---------|
| Dev only                | 0.77    |
| Dev + Eval (2018-19)    | 0.81    |

Table 3: Improvements using Fusion

| Co-occurrence Matrix Direction | AUC-ROC |
|-------------------------------|---------|
| Horizontal                    | 0.80    |
| Vertical                      | 0.80    |
| Horizontal + Vertical         | 0.81    |
| Fusion                        | 0.82    |

Results on Development datasets

We combine all the datasets of the Development datasets into one dataset comprising a total of 59,774 images. Of this, we choose 90% for training and 10% for validation. Fig. 4(a,b) show the model accuracy and loss respectively for the training and validation data. We obtained a high training and validation accuracy, as well as low losses on both sets.

Results on Evaluation dataset

We evaluated our holistic detector on the MFC 2020 Evaluation dataset of 19,000+ images. Initially, we trained a model only on the Development datasets (59,000+ images) and obtained an area under the curve (AUC) of the receiver operating characteristic (ROC) curve (metric used in the challenge) of 0.77. Then, we trained a model that included both the Development datasets and Evaluation datasets (37,000+ images) of previous years (whose ground truth was made available at the time of the 2020 challenge), and obtained an AUC of 0.81 as shown in Tab. 2. This was the highest reported AUC at the time of announcement of the challenge results [4] (slide 47).

Boosting Detection Performance using Fusion

To further boost the performance, we trained two more models, one that uses only horizontal co-occurrence matrices and the other that uses only vertical co-occurrence matrices. Then, we compute the average prediction values of the three models as the final prediction. This fusion further resulted in a slight increase in the AUC to 0.82 (Tab. 3).

Results on Selective Manipulation Types

Next we evaluate the performance of our holistic image manipulation detector on selective manipulation types such as splicing, blurring, and others. This analysis helps in understanding on which manipulation types our model succeeds/fails in detecting. Tab. 4 summarizes the results on selective manipulations used in the MFC 2020 evaluation dataset. We can see that our model achieves very high AUC (greater than 0.9) on manipulations such
| Manipulation Type          | AUC-ROC |
|---------------------------|---------|
| Splice                    | 0.73    |
| Clone                     | 0.82    |
| SpliceClone               | 0.73    |
| Crop                      | 0.78    |
| Resize                    | 0.82    |
| Global Intensity Normalization | 0.94  |
| Intensity Change          | 0.81    |
| Antiforensic-PRNU          | 0.83    |
| Antiforensic-CFA           | 0.81    |
| Social Media Laundering   | 0.93    |
| Global Blurring           | 0.80    |
| Local Blurring            | 0.83    |
| GAN                       | 0.83    |
| Non-GAN CGI               | 0.85    |
| Distortion                | 0.89    |

Table 4: Results on Selective Manipulations in MFC 2020 Evaluation dataset

as Global Intensity Normalization and Social Media Laundering, and high AUC on other types such as Blurring, GAN and Computer Graphics (CGI) based manipulations, Resize, Antiforensics, to name a few. There are also manipulation types such as Splicing, Cropping for which the AUC is not high. This could be because the global statistics of the pixels do not change a lot in manipulations such as cropping or splicing. One solution to increase the performance for these manipulation types is to consider local block statistics instead of the global image statistics, which we will explore in future.

Conclusion and Future Work

In this paper, we presented a method to holistically detect tampered images which are usually created using many different manipulation operations. We combine statistical co-occurrence matrices with deep learning by extracting horizontal and vertical co-occurrence matrices on the image channels and then passing them through a deep neural network to classify if images have been tampered or not. Our approach is holistic since it does not depend on a particular image manipulation type. Experiments on a recently concluded media forensics challenge showed that our approach achieved promising results. In future, we will focus on combining our holistic image manipulation detection method with methods that focus on detecting specific manipulations.

Acknowledgements

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

[1] “IEEE IFS-TC Image Forensics Challenge Dataset,” http://ifs reco 2014.1
[2] “2017 NIST Nimble Challenge Evaluation,” https://www.nist.gov/itl/idad/mig/nimble-challenge-2017-evaluation.1
[3] “Media Forensics Challenge 2019,” https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2019-0.1
[4] “2020 MediFor Challenge Evaluation Overview,” https://mfc2020.medicyr 2019.1
[5] K. Sullivan, U. Madhow, S. Chandrasekaran, et al., “Steganalysis of spread spectrum data hiding exploiting cover memory,” in Security, Steganography, and Watermarking of Multimedia Contents VII, 5681, 38–47, International Society for Optics and Photonics (2005).
[6] K. Sullivan, U. Madhow, S. Chandrasekaran, et al., “Steganalysis for markov cover data with applications to images,” IEEE Transactions on Information Forensics and Security 1(2), 275–287 (2006).
[7] T. Pevný, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” Information Forensics and Security, IEEE Transactions on 5(2), 215–224 (2010).
[8] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” IEEE Transactions on Information Forensics and Security 7(3), 868–882 (2012).
[9] D. Cozzolino, G. Poggi, and L. Verdoliva, “Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection,” in Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, 159–164, ACM (2017).
[10] L. Nataraj, T. M. Mohammed, B. Manjunath, et al., “Detecting gan generated fake images using co-occurrence matrices,” arXiv preprint arXiv:1903.06836 (2019).
[11] H. Farid, “Image forgery detection,” IEEE Signal processing magazine 26(2), 16–25 (2009).
[12] B. Mahdian and S. Saic, “A bibliography on blind methods for identifying image forgery,” Signal Processing: Image Communication 25(6), 389–399 (2010).
[13] G. K. Birajdar and V. H. Mankar, “Digital image forgery detection using passive techniques: A survey,” Digital Investigation 10(3), 226–245 (2013).
[14] X. Lin, J.-H. Li, S.-L. Wang, et al., “Recent advances in passive digital image security forensics: A brief review,” Engineering (2018).
[15] S. Walla and K. Kumar, “Digital image forgery detection: a systematic scrutiny,” Australian Journal of Forensic Sciences, 1–39 (2018).
[16] A. C. Popescu and H. Farid, “Exposing digital forgeries by detecting traces of resampling,” IEEE Transactions on signal processing 53(2), 758–767 (2005).
[17] B. Mahdian and S. Saic, “Blind authentication using periodic properties of interpolation,” Information Forensics and Security, IEEE Transactions on 3, 529–538 (2008).
[18] M. Kirchner, “On the detectability of local resampling in digital images,” Security, Forensics, Steganography, and Watermarking of Multimedia Contents X 6819(1), 68190F, SPIE (2008).
[19] X. Feng, I. J. Cox, and G. Doerr, “Normalized energy density-based forensic detection of resampled images,” IEEE Transactions on Multimedia 14(3), 536–545 (2012).
[20] S.-J. Ryu and H.-K. Lee, “Estimation of linear transformation by an image security forensics: A brief review,” Digital Investigation 10(3), 758–767 (2013).
[21] J. Li, X. Li, B. Yang, et al., “Segmentation-based image copy-move forgery detection scheme,” IEEE Transactions on Information Forensics and Security 10(3), 507–518 (2015).
Acknowledgements

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. The paper is approved for public release, distribution unlimited.

Author Biography

Lakshmanan Nataraj received his B.E degree from Sri Venkateswara College of Engineering, Anna University in 2007, and the Ph.D. degree in the Electrical and Computer Engineering from the University of California, Santa Barbara in 2015. He is currently a Senior Research Staff Member at Mayachitra Inc., Santa Barbara, CA. His research interests include malware analysis and image forensics.

Michael Goebel received his B.S. and M.S. degrees in Elec-
Tajuddin Manhar Mohammed received his B.Tech (Hons.) degree from Indian Institute of Technology (IIT), Hyderabad, India in 2015 and his M.S. degree in Electrical and Computer Engineering from University of California Santa Barbara (UCSB), Santa Barbara, CA in 2016. After obtaining his Masters degree, he obtained a job as a Research Staff Member for Mayachitra Inc., Santa Barbara, CA. His recent research efforts include developing computer vision techniques for image forensics and cyber security.

Shivkumar Chandrasekaran received his Ph.D. degree in Computer Science from Yale University, New Haven, CT, in 1994. He is a Professor in the Electrical and Computer Engineering Department, University of California, Santa Barbara. His research interests are in Computational Mathematics.

B. S. Manjunath (F'05) received the Ph.D. degree in Electrical Engineering from the University of Southern California in 1991. He is currently a Distinguished Professor at the ECE Department at the University of California at Santa Barbara. He has co-authored about 300 peer-reviewed articles. His current research interests include image processing, computer vision and biomedical image analysis.