

Detecting GAN generated Fake Images using Co-occurrence Matrices

Lakshmanan Nataraj; Mayachitra Inc., Santa Barbara, California, USA

Tajuddin Manhar Mohammed; Mayachitra Inc., Santa Barbara, California, USA

B. S. Manjunath; Mayachitra Inc., Santa Barbara, California, USA

Shivkumar Chandrasekaran; Mayachitra Inc., Santa Barbara, California, USA

Arjuna Flenner; Naval Air Warfare Center Weapons Division, China Lake, California, USA

Jawadul H. Bappy; JD.com

Amit K. Roy-Chowdhury; University of California, Riverside, California, USA

Abstract

The advent of Generative Adversarial Networks (GANs) has brought about completely novel ways of transforming and manipulating pixels in digital images. GAN based techniques such as Image-to-Image translations, DeepFakes, and other automated methods have become increasingly popular in creating fake images. In this paper, we propose a novel approach to detect GAN generated fake images using a combination of co-occurrence matrices and deep learning. We extract co-occurrence matrices on three color channels in the pixel domain and train a model using a deep convolutional neural network (CNN) framework. Experimental results on two diverse and challenging GAN datasets comprising more than 56,000 images based on unpaired image-to-image translations (cycleGAN [1]) and facial attributes/expressions (StarGAN [2]) show that our approach is promising and achieves more than 99% classification accuracy in both datasets. Further, our approach also generalizes well and achieves good results when trained on one dataset and tested on the other.

Introduction

Recent advances in Machine Learning and Artificial Intelligence have made it tremendously easy to create and synthesize digital manipulations in images and videos. In particular, Generative Adversarial Networks (GANs) [3] have been one of the most promising advancements in image enhancement and manipulation. Due to the success of using GANs for image editing, it is now possible to use a combination of GANs and off-the-shelf image-editing tools to modify digital images to such an extent that it has become difficult to distinguish doctored images from normal ones. The field of digital Image Forensics develops tools and techniques to detect manipulations in digital images such as splicing, resampling and copy move, but the efficacy and robustness of these tools on GAN generated images is yet to be seen. To address this, we propose a novel method to automatically identify GAN generated fake images using techniques that have been inspired from classical steganalysis.

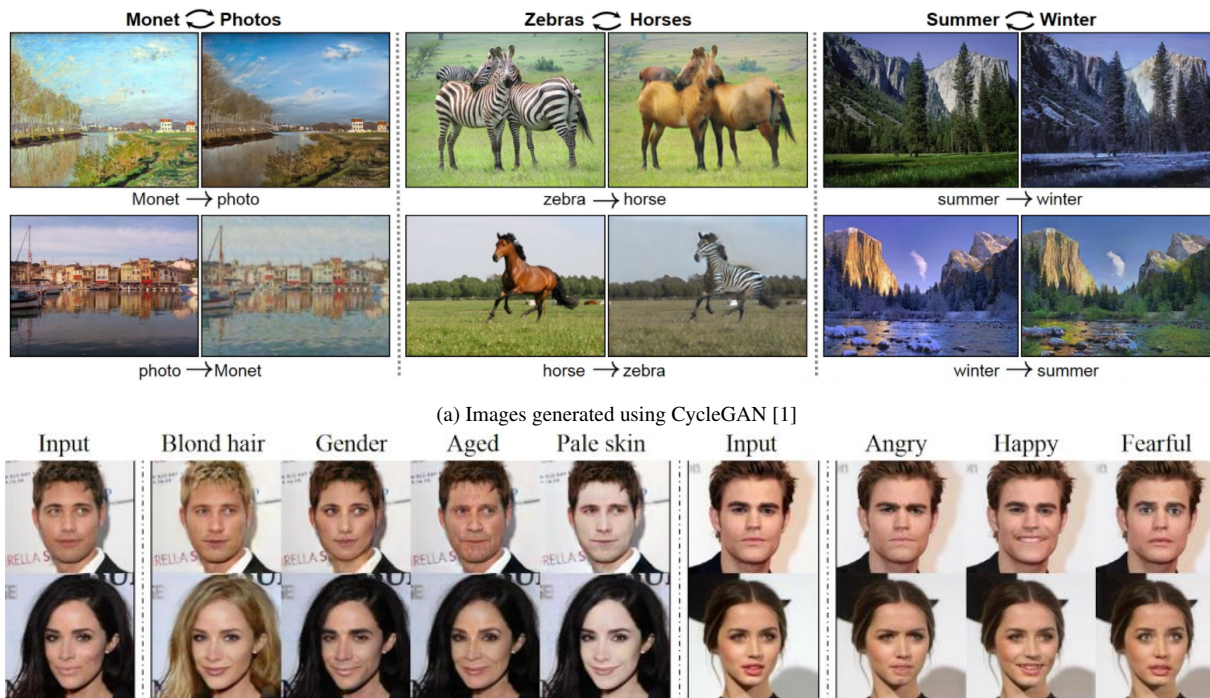
The seminal work on GANs[3] cast the machine learning field of generative modeling as a game theory optimization problem. GANs contain two networks - the first network is a generative network that can generate fake images and the second network is a discriminative network that determines if an image is real or fake. Encoded in the GAN loss function is a min-max

game which creates a competition between the generative and discriminative networks. As the discriminative network becomes better at distinguishing between real and fake images, the generative model becomes better at generating fake images.

GANs have been applied to many image processing tasks such as image synthesis, super-resolution and image completion. Inspired by the results of these image processing tasks, GANs have brought in novel attack avenues such as computer generated (CG) faces [4], augmenting faces with CG facial attributes [2], and seamless transfer of texture between images [1], to name a few. Two of the most common applications of GANs include texture or style transfer between images and face manipulations. An example of GAN generated texture translation between images, such as horses-to-zebras and summer-to-winter, is shown in Fig. 1(a). These techniques manipulate the entire image to change the visual appearance of the scene [1]. There has been also tremendous progress in facial manipulations - in particular, automatic generation of facial attributes and expressions. Fig. 1(b) shows one such recent example where various facial attributes such as hair, gender, age and skin color, and expressions such as anger, happiness and fear are generated on faces of celebrities [2].

While the visual results are promising, the GAN based techniques alter the statistics of pixels in the images that they generate. Hence, methods that look for deviations from natural image statistics could be effective in detecting GAN generated fake 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, we propose a novel method to identify GAN generated images using a combination of pixel co-occurrence matrices and deep learning. Here we pass the co-occurrence matrices directly through a deep learning framework and allow the network to learn important features of the co-occurrence matrices. We also avoid computation of residuals or passing an image through various filters, but rather compute the co-occurrence matrices on the image pix-



(a) Images generated using CycleGAN [1]
(b) Images generated using StarGAN [2]
Figure 1: Examples of images that have been generated using GANs.

els itself. Experimental results on two diverse and challenging datasets generated using GAN based methods show that our approach is promising and generalizes well when the GAN model is not known during training.

Related Work

Since the seminal work on GANs [3], there have been several hundreds of papers on using GANs to generate images. These works focus on generating images of high perceptual quality [10, 11, 12, 13, 14, 15, 4], image-to-image translations [13, 16, 1], domain transfer [17, 18], super-resolution [19], image synthesis and completion [20, 21, 22], and generation of facial attributes and expressions [23, 24, 18, 2]. Several methods have been proposed in the area of image forensics over the past years [25, 26, 27, 28, 29]. Recent approaches have focused on applying deep learning based methods to detect tampered images [30, 31, 32, 33, 34, 9, 35]

The detection of GAN images is a new area in image forensics and there are very few papers in this area [36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. Related fields also include detection of computer generated (CG) images [46, 47, 48, 49]. The most relevant work is a recent paper [36] on detecting GAN based image-to-image translation generated using cycleGAN [1]. Here the authors compare various existing methods to identify cycleGAN images from normal ones. The top results they obtained using a combination of residual features [50, 9] and deep learning [51]. Similar to [36], the authors in [38] compute the residuals of high pass filtered images and then extract co-occurrence matrices on these residuals, which are then concatenated to form a feature vector that can distinguish real from fake GAN images. In contrast to these approaches, our approach does not need any image resid-

uals to be computed. Rather, our method directly computes co-occurrence matrices on the three color channels which are then passed through a deep convolutional neural network (CNN) to learn a model that can detect fake GAN generated images.

Methodology

To detect GAN images, we compute co-occurrence matrices on the RGB channels of an image. Co-occurrence matrices have been previously used in steganalysis to identify images that have data hidden in them [5, 6, 7, 8] and in image forensics to detect or localize tampered images [50, 9]. In prior works, co-occurrence matrices are usually computed on image residuals by passing the image through many filters and then obtaining the difference. Sometimes features or statistics are also computed on the matrices and then a classifier is trained on these features to classify data hidden or tampered images [5].

However, in this paper, we compute co-occurrence matrices directly on the image pixels on each of the red, green and blue channels and pass them through a convolutional neural network, thereby allowing the network to learn important features from the co-occurrence matrices. An overview of our approach is shown in Fig. 2. Specifically, the first step is to compute the co-occurrence matrices on the RGB channels to obtain a $3 \times 256 \times 256$ tensor. This tensor is then passed through a multi-layer deep convolutional neural network: conv layer with 32 3×3 convs + ReLu layer + conv layer with 32 5×5 convs + max pooling layer + conv layer with 64 3×3 convs + ReLu layer + conv layer with 64 5×5 convs + max pooling layer + conv layer with 128 3×3 convs + ReLu layer + conv layer with 128 5×5 convs + max pooling layer + 256 dense layer + 256 dense layer + sigmoid layer. A variant of adaptive

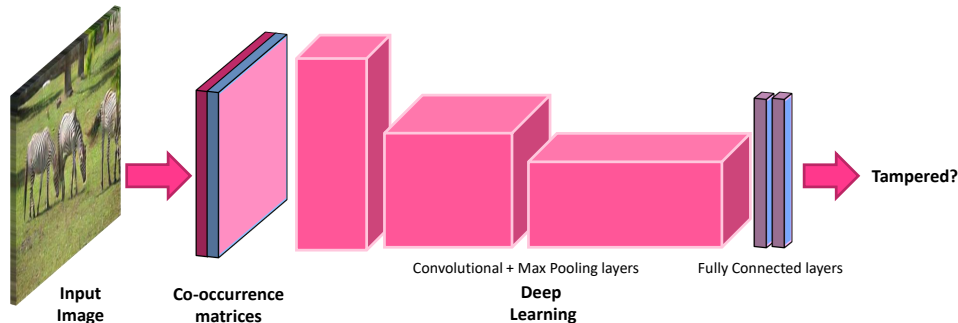


Figure 2: An end-to-end framework to detect GAN generated images

stochastic gradient descent is used as the optimizer.

Experiments

We evaluate our method on two diverse and challenging datasets which contain GAN generated images such as image-to-image translation, style transfer, facial attributes and expressions.

Datasets

CycleGAN dataset: This dataset contains unpaired image-to-image translations of various objects and scenes such as horses-to-zebras, summer-to-winter, images to paintings (Monet, Van Gough), style transfer such as labels to facades, and others that were generated using a cycle-consistent GAN framework [1]. We followed the instructions provided by the authors as detailed in <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>, and obtained 36,302 images (18,151 GAN and 18,151 non-GAN images). The distribution of GAN images are as follows: apple2orange (2014), horse2zebra (2401), summer2winter (2193), cityscapes (2975), facades (400), map2sat (1096), Ukiyoe (1500), Van Gogh (1500), Cezanne (1500), Monet (2752). We obtained this distribution from the authors of [36]¹ and we compare our approach with their results in the Experiments section.

StarGAN dataset: This dataset consists of 19,990 images of which 1,999 were from faces taken from the celebA dataset [52] of celebrity faces and the remaining 17,991 images were GAN generated images with 5 varying facial attributes such as black hair, blond hair, brown hair, gender change, aged and 4 combinations of the same. We followed the instructions in <https://github.com/yunjey/stargan> to generate the images.

Evaluation

We first evaluate our approach on two datasets separately and then perform cross evaluation on the two datasets (one dataset as training and other as testing) to see the generalizability of our approach. For both datasets, 50% of the data is used for training, 25% for validation and 25% for testing. We train the network for 50 epochs with a batch size of 40 and use a variant of adaptive stochastic gradient as optimizer. Fig. 3(a,b) and Fig. 3(c,d) show the model accuracy and loss on cycleGAN dataset and StarGAN dataset, respectively. We obtained a high training and validation accuracy of 99.90% and 99.40% on cycleGAN dataset, and 99.43% and 99.39% on StarGAN dataset respectively. We

Table 1: Experiment on Generalizability

Training dataset	Testing dataset	Accuracy
cycleGAN	StarGAN	99.49
StarGAN	cycleGAN	93.42

then evaluated the model on the held-out test sets and obtained a testing accuracy of 99.71% on the cycleGAN dataset and 99.37% on the StarGAN dataset.

Generalizability

Next, we evaluate the generalizability of our approach by training on one dataset and testing on the other. First we train on all the images in the cycleGAN dataset (35,302 images) and test the model on all images of the StarGAN dataset (19,990 images), and then we reverse the experiment where we train on StarGAN and test on cycleGAN. We train the network till 50 epochs and report on the model that gave the highest accuracy. As shown in Tab. 1, our method still maintains a high accuracy even across diverse datasets. The model trained on cycleGAN dataset has a higher accuracy of 99.45% in comparison with the model trained on StarGAN dataset which got 93.42%. The lower accuracy for the model trained on StarGAN dataset could be because of the non-uniform distribution of class samples in the StarGAN dataset, and due to the diverse image sources/categories in the cycleGAN dataset.

Comparison with State-of-the-art

We compare our approach with the results presented in [36]. Here the authors conduct a study on the detection of images manipulated by GAN-based image-to-image translation on the cycleGAN dataset. For evaluation, they adopt a leave-one-manipulation-out strategy on the categories in the cycleGAN dataset, where at each iteration images belonging to one category are set aside for validation and the images from other categories are used for training. The methods evaluated are based on steganalysis, generic image manipulations, detection of computer graphics, a GAN discriminator used in the cycleGAN paper, and generic deep learning architecture pretrained on ImageNet [53], but fine tuned to the cycleGAN dataset. Among these the top performing ones were from steganalysis [8, 50] based on extracting features from high-pass residual images, a deep neural network designed to extract residual features [9] (denoted by Cozolino2017) and XceptionNet [51] deep neural network trained

¹We thank the authors for providing the dataset distribution

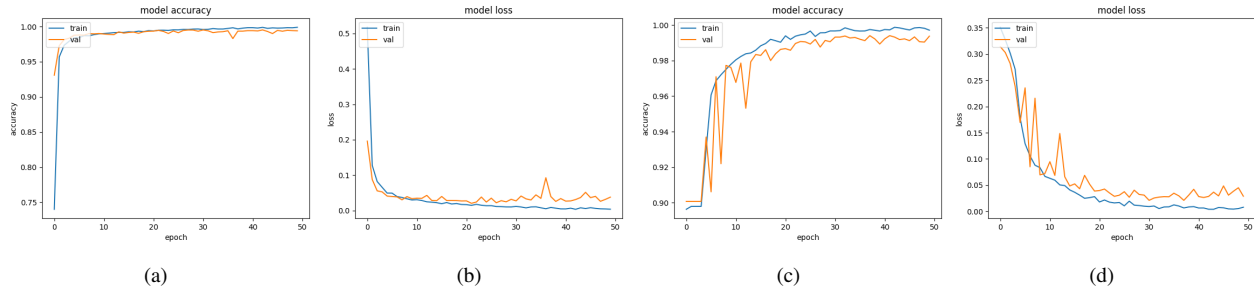


Figure 3: Model accuracy and loss on CycleGAN dataset [1] (a,b) and StarGAN dataset [2] (c,d)

Table 2: Comparison with State-of-the-art

Method	ap2or	ho2zeb	wint2sum	citysc.	facades	map2sat	Ukiyoe	Van Gogh	Cezanne	Monet	Average
Steganalysis feat.	98.93	98.44	66.23	100.00	97.38	88.09	97.93	99.73	99.83	98.52	94.40
Cozzalino2017	99.90	99.98	61.22	99.92	97.25	99.59	100.00	99.93	100	99.16	95.07
XceptionNet	95.91	99.16	76.74	100.00	98.56	76.79	100.00	99.93	100.00	95.10	94.49
Proposed	99.78	99.75	99.72	92.00	80.63	97.51	99.63	100.00	99.63	99.16	97.84

Table 3: Effect of JPEG compression

JPEG Quality Factor	Trained on original images	Trained on JPEG compressed images
95	74.5	93.78
85	69.46	91.61
75	64.46	87.31

on ImageNet but fine-tuned to this dataset. We report the results as mentioned in the paper only from these three methods and compare with our approach.

Results on original images: Here we consider the images generated from the cycleGAN dataset as it is and do not perform any postprocessing on the images. Tab. 2 summarizes the results of our proposed approach along with the three top performing approaches in [36]. On average our method outperformed other methods and was able to achieve an accuracy of **97.84**. On most categories, our approach was better than or on-par with the other top methods. The only categories where our method performed poorly were ‘cityscapes’ and ‘facades’. These could be because the original images in these categories were JPEG compressed which could have affected the classification accuracy. In the next section, we study the effect of compression on our method.

Effect of JPEG compression: In [36], the authors investigated the sensitivity of their methods on compression. Using a compression method similar to Twitter, they trained their detection methods on original uncompressed images and tested on JPEG compressed images. The objective of this study was to test the robustness of the detection techniques when images are posted in social networks such as Twitter. In the second scenario, they train and test on the JPEG compressed images. We performed both of these experiments on the cycleGAN dataset. Since we are not aware of the exact JPEG quantization tables used in Twitter, our approach was similar but we tested on three different JPEG quality factors (QF): 95, 85 and 75. We used 50% of the data for training, 25% for validation and 25% for testing. The results are reported on the 25% testing data of the cycleGAN dataset (9,076 images). As shown in Tab. 3, the accuracy progressively drops as the QF decreases from 95-75, when trained on the original im-

ages. But when the JPEG compressed images are used for training, the accuracy shows a substantial increase. Even at a QF of 75, the accuracy is still 87.31%. This is also consistent with the results reported in [36], where they report close to a 10% drop in accuracy on Twitter-like compressed images.

Conclusions

In this paper, we proposed a novel method to detect GAN generated fake images using a combination of pixel co-occurrence matrices and deep learning. Co-occurrence matrices are computed on the color channels of an image and then trained using a deep convolutional neural network to distinguish GAN generated fake images from real ones. Experiments on two diverse GAN datasets show that our approach is both effective and generalizable. In future, we will consider localizing the manipulated pixels in GAN generated fake images.

Acknowledgments

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.

References

- [1] J.-Y. Zhu, T. Park, P. Isola, *et al.*, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision*, (2017).
- [2] Y. Choi, M. Choi, M. Kim, *et al.*, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” in

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8789–8797 (2018).

- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, “Generative adversarial nets,” in *Advances in neural information processing systems*, 2672–2680 (2014).
- [4] T. Karras, T. Aila, S. Laine, *et al.*, “Progressive growing of gans for improved quality, stability, and variation,” *arXiv preprint arXiv:1710.10196* (2017).
- [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. Kodovsky, “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] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784* (2014).
- [11] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434* (2015).
- [12] T. Salimans, I. Goodfellow, W. Zaremba, *et al.*, “Improved techniques for training gans,” in *Advances in Neural Information Processing Systems*, 2234–2242 (2016).
- [13] P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134 (2017).
- [14] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein gan,” *arXiv preprint arXiv:1701.07875* (2017).
- [15] I. Gulrajani, F. Ahmed, M. Arjovsky, *et al.*, “Improved training of wasserstein gans,” in *Advances in Neural Information Processing Systems*, 5767–5777 (2017).
- [16] Z. Yi, H. Zhang, P. Tan, *et al.*, “Dualgan: Unsupervised dual learning for image-to-image translation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2868–2876, IEEE (2017).
- [17] Y. Taigman, A. Polyak, and L. Wolf, “Unsupervised cross-domain image generation,” *arXiv preprint arXiv:1611.02200* (2016).
- [18] T. Kim, M. Cha, H. Kim, *et al.*, “Learning to discover cross-domain relations with generative adversarial networks,” in *International Conference on Machine Learning*, 1857–1865 (2017).
- [19] C. Ledig, L. Theis, F. Huszar, *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681–4690 (2017).
- [20] Y. Li, S. Liu, J. Yang, *et al.*, “Generative face completion,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, **1**(2), 3 (2017).
- [21] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and locally consistent image completion,” *ACM Transactions on Graphics (TOG)* **36**(4), 107 (2017).
- [22] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, *et al.*, “High-resolution image synthesis and semantic manipulation with conditional gans,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8798–8807 (2018).
- [23] M.-Y. Liu and O. Tuzel, “Coupled generative adversarial networks,” in *Advances in neural information processing systems*, 469–477 (2016).
- [24] G. Perarnau, J. van de Weijer, B. Raducanu, *et al.*, “Invertible conditional gans for image editing,” *arXiv preprint arXiv:1611.06355* (2016).
- [25] H. Farid, “Image forgery detection,” *IEEE Signal processing magazine* **26**(2), 16–25 (2009).
- [26] B. Mahdian and S. Saic, “A bibliography on blind methods for identifying image forgery,” *Signal Processing: Image Communication* **25**(6), 389–399 (2010).
- [27] G. K. Birajdar and V. H. Mankar, “Digital image forgery detection using passive techniques: A survey,” *Digital Investigation* **10**(3), 226–245 (2013).
- [28] X. Lin, J.-H. Li, S.-L. Wang, *et al.*, “Recent advances in passive digital image security forensics: A brief review,” *Engineering* (2018).
- [29] S. Walia and K. Kumar, “Digital image forgery detection: a systematic scrutiny,” *Australian Journal of Forensic Sciences*, 1–39 (2018).
- [30] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, 5–10 (2016).
- [31] B. Bayar and M. C. Stamm, “Design principles of convolutional neural networks for multimedia forensics,” in *The 2017 IS&T International Symposium on Electronic Imaging: Media Watermarking, Security, and Forensics*, IS&T Electronic Imaging (2017).
- [32] Y. Rao and J. Ni, “A deep learning approach to detection of splicing and copy-move forgeries in images,” in *Information Forensics and Security (WIFS), 2016 IEEE International Workshop on*, 1–6, IEEE (2016).
- [33] J. Bunk, J. H. Bappy, T. M. Mohammed, *et al.*, “Detection and localization of image forgeries using resampling features and deep learning,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, 1881–1889, IEEE (2017).
- [34] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, *et al.*, “Exploiting spatial structure for localizing manipulated image regions,” in *Proceedings of the IEEE International Conference on Computer Vision*, (2017).
- [35] P. Zhou, X. Han, V. I. Morariu, *et al.*, “Learning rich features for image manipulation detection,” *arXiv preprint arXiv:1805.04953* (2018).
- [36] F. Marra, D. Gragnaniello, D. Cozzolino, *et al.*, “Detection of gan-generated fake images over social networks,” in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 384–389, IEEE (2018).
- [37] R. Valle, U. CNMAT, W. Cai, *et al.*, “Tequilagan: How to easily identify gan samples,” *arXiv preprint arXiv:1807.04919* (2018).
- [38] H. Li, B. Li, S. Tan, *et al.*, “Detection of deep network generated images using disparities in color components,” *arXiv preprint arXiv:1808.07276* (2018).
- [39] S. McCloskey and M. Albright, “Detecting gan-generated imagery using color cues,” *arXiv preprint arXiv:1812.08247* (2018).
- [40] H. Li, H. Chen, B. Li, *et al.*, “Can forensic detectors identify gan

generated images?,” in *APSIPA Annual Summit and Conference 2018*, (2018).

- [41] A. Jain, R. Singh, and M. Vatsa, “On detecting gans and retouching based synthetic alterations,” in *9th International Conference on Biometrics: Theory, Applications and Systems*, (2018).
- [42] S. Tariq, S. Lee, H. Kim, *et al.*, “Detecting both machine and human created fake face images in the wild,” in *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, 81–87, ACM (2018).
- [43] F. Marra, D. Gragnaniello, L. Verdoliva, *et al.*, “Do gans leave artificial fingerprints?,” *arXiv preprint arXiv:1812.11842* (2018).
- [44] H. Mo, B. Chen, and W. Luo, “Fake faces identification via convolutional neural network,” in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, 43–47, ACM (2018).
- [45] N.-T. Do, I.-S. Na, and S.-H. Kim, “Forensics face detection from gans using convolutional neural network,” in *ISITC’2018*, (2018).
- [46] A. E. Dirik, S. Bayram, H. T. Sencar, *et al.*, “New features to identify computer generated images,” in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 4, IV–433, IEEE (2007).
- [47] R. Wu, X. Li, and B. Yang, “Identifying computer generated graphics via histogram features,” in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 1933–1936, IEEE (2011).
- [48] B. Mader, M. S. Banks, and H. Farid, “Identifying computer-generated portraits: The importance of training and incentives,” *Perception* **46**(9), 1062–1076 (2017).
- [49] N. Rahmouni, V. Nozick, J. Yamagishi, *et al.*, “Distinguishing computer graphics from natural images using convolution neural networks,” in *Information Forensics and Security (WIFS), 2017 IEEE Workshop on*, 1–6, IEEE (2017).
- [50] D. Cozzolino, D. Gragnaniello, and L. Verdoliva, “Image forgery detection through residual-based local descriptors and block-matching,” in *Image Processing (ICIP), 2014 IEEE International Conference on*, 5297–5301, IEEE (2014).
- [51] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *arXiv preprint*, 1610–02357 (2017).
- [52] Z. Liu, P. Luo, X. Wang, *et al.*, “Deep learning face attributes in the wild,” in *Proceedings of the IEEE International Conference on Computer Vision*, 3730–3738 (2015).
- [53] J. Deng, W. Dong, R. Socher, *et al.*, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, (2009).

Author Biography

Lakshmanan Nataraj received his B.E degree in Electronics and Communications Engineering from Sri Venkateswara College of Engineering (affiliated to 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, where he leads research projects on malware detection and media forensics. His research interests include multimedia security, malware detection and image forensics.

Tajuddin Manhar Mohammed received his B.Tech (Hons.) degree in Electrical Engineering 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 deep learning and computer

vision techniques for image forensics and cyber security.

B. S. Manjunath (S’88M’91SM’01F’05) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1991. He is currently a Distinguished Professor of Electrical and Computer Engineering at the University of California at Santa Barbara, where he directs the Center for Multimodal Big Data Science and Healthcare. He has authored or coauthored about 300 peer-reviewed articles and served as an Associate Editor for the *IEEE Transactions on Image Processing*, the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, the *IEEE Transactions on Multimedia*, the *IEEE Transactions on Information Forensics and Security*, and the *IEEE Signal Processing Letters*. His research interests include image processing, machine learning, computer vision, and media forensics. He is a fellow of IEEE and ACM.

Shivkumar Chandrasekaran received his M.Sc. (Hons.) degree in physics from the Birla Institute of Technology and Science (BITS), Pilani, India, in 1987, and his Ph.D. degree in Computer Science from Yale University, New Haven, CT, in 1994. He was a Visiting Instructor at North Carolina State University, Raleigh, in the Mathematics Department, before joining the Electrical and Computer Engineering Department, University of California, Santa Barbara, where he is currently a Professor. His research interests are in Computational Mathematics

Arjuna Flenner received his Ph.D. in Physics at the University of Missouri-Columbia located in Columbia MO in the year 2004. His major emphasis was mathematical Physics. After obtaining his Ph.D., Arjuna Flenner obtained a job as a research physicist for NAVAIR at China Lake CA. He won the 2013 Dr. Delores M. Etter Navy Scientist and Engineer award for his work on Machine Learning.

Jawadul H. Bappy received the B.S. degree in Electrical and Electronic Engineering from the Bangladesh University of Engineering and Technology, Dhaka in 2012. He received his Ph.D. in Electrical and Computer Engineering from the University of California, Riverside in 2018. He is currently working as a scientist at JD.Com in Mountain View, CA. His main research interest includes media forensics, deep generative models, and advanced machine learning techniques for real-life applications.

Amit K. Roy-Chowdhury received the Bachelors degree in Electrical Engineering from Jadavpur University, Calcutta, India, the Masters degree in Systems Science and Automation from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park. He is a Professor of Electrical and Computer Engineering and a Cooperating Faculty in the Department of Computer Science and Engineering, University of California, Riverside. His broad research interests include computer vision, image processing, and vision-based statistical learning, with applications in cyber-physical, autonomous and intelligent systems. He is a coauthor of two books: *Camera Networks: The Acquisition and Analysis of Videos over Wide Areas*, and *Recognition of Humans and Their Activities Using Video*. He is the editor of the book *Distributed Video Sensor Networks*. He has been on the organizing and program committees of multiple computer vision and image processing conferences and is serving on the editorial boards of multiple journals. He is a Fellow of the IEEE and IAPR.