Data Analytics and Artificial Intelligence

Vol: 3(4), April 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

Fake Image Detection

Madhumitha Venkatesan

Adhiyamaan College of Engineering Hosur, Tamandu, India.

Corresponding Author Email: Vmithu33@gmail.com

Abstract. *In daily life, people come across tampered or forged images from the tabloid magazines to the business industry. Furthermore, in media outlets, scientific journals, political campaigns, courtrooms, and photo hoaxes that land in our email boxes, forged images are appearing more frequently in a unique way unable to identify the fake image with the needed sophistication. With the wide spread of digital document use in administrations, fabrication and use of forged documents have become a serious problem. This paper presents a study and classification of the most important works on image and document forgery detection. The classification is based on documents type, forgery type, detection method, validation dataset, evaluation metrics and obtained results. Most of existing forgery detection works are dealing with images and few of them analyze administrative documents and go deeper to analyze their contents.*

Keywords: *Fake Image, Neural network, Detection methods, CNN, Dense Net.*

1. INTRODUCTION

Administrative documents are forms created to establish an identity, right or authorization. And therefore, falsification of documents or identity is severely punishable by law. This concerns identity and authorization documents such as passports, identity cards and driving licenses. Fraudsters use these false documents to commit offences generally, like travelling using fake passport or visa, or to forge diplomas to get a job. Moreover, digitalization or digital transformation has created new communication and optimization tools for business and administration management. Due to new technologies, it is easy to digitize documents using a scanner and save them in pdf format or images. The new tendency of governments is to use digital documents instead of hard ones to optimize many administrative procedures. Digital images took their place in the existence of film photography with pluses Van Dijck. Moreover, unlike conventional photography, the image capture by the digital cameras is easy, besides the storage and transfer is also feasible. In the current information era, the benefits of the digital images are exploited in different fields such as military, news, media, medical diagnosis, forensics, tabloid magazines, scientific journals, fashion industries, court halls and so on Watson and Null. One of the central fields which fetched a notable gain is electronic commerce (Amazon, Snap deal etc.). Because of the advancement in the IT and internet sector, the growth rate in E-Commerce has considerably increased in recent years. Unfortunately, nowadays document images tampering has become easier with the use of sophisticated tools with recent advances in technology and multimedia. Anyone can use available and low-cost professional image forgery tools easily to modify images, such that it cannot be distinguished from authentic ones with the naked eye. It is a fact that social media has changed the way people interact and carry on with their everyday lives. Social networking sites are a prominent media phenomenon nowadays and have attracted many people. Worldwide, the number of users now exceeds three billion. In the Gulf region, growth in the number of active users has exceeded 66%. Saudi Arabia ranks seventh in the world in terms of social media use; more than 75% of its estimated 25 million people are active users of social media. Social media is based on specific foundations that bring people together and empower them to express themselves, share their interests and ideas, and forge new friendships with others who share their interests. Facebook, Twitter, and Instagram are among the most popular social networking sites of the day. It is a widespread practice to share images online through social networking services such as Instagram. At least 80 million images are currently shared via Instagram every day. Instagram enables users to take photographs, apply digital photographic filters, and upload the pictures to website for social networking together with short captions. People upload and share billions of pictures every day on social media. They depend on previous information taken from the original image. Though, these methods are not used for all kinds of documents because they need special equipment to embed watermarks or a signature like cameras. Alternatively, passive methods are used to detect forgery without previous inserted information. These methods can treat two main types of image forgery: Copy-move tampering (CMF) and image splicing tampering, where copy-move forgery is the most generally adapted by forgers. Many researchers have been interested in copy-move forgery and there exist several methods introduced in this field. Many researchers have been interested in copy-move forgery and there exist several methods introduced in this field. A huge number of people have become victims of photo forgery in this technological age. Some criminals use software to exploit and use pictures as evidence to confuse the courts of justice. This research proposes an approach that takes an image as input and classifies it, using an effective system (the CNN model). The result of this proposed research will be helpful in monitoring and tracking social media content and in discovering fraud on social networking sites, especially in the field of images.

LITERATURE REVIEW

Very little work has been finalized around detecting forge audio, images, and videos. Yet, several studies and tasks are underway to identify what can be done around the incredible proliferation of counterfeit pictures online. Adobe recognizes the way in which Photoshop is misused and has tried to offer a sort of antidote. The following provide a summary of a few of these studies:

Zheng et al. (2018), the identification of fake news and images is very difficult, as fact-finding of news on a pure basis remains an open problem and few existing models Can be used to resolve the problem. It has been proposed to study the problem of "detecting false news." Through a thorough investigation of counterfeit news, many useful properties are determined from text words and pictures used in counterfeit news. There are some hidden characteristics in words and images used in fake news, which can be identified through a collection of hidden properties derived from this model through various layers. A pattern called TI-CNN has been proposed. By displaying clear and embedded features in a unified space, TI-CNN is trained with both text and image information at the same time.

Kuruvilla et al., a neural network was successfully trained by analyzing the 4000 fake and 4000 real images error level. The trained neural network has succeeded in identifying the image as fake or real, with a high success rate of 83%. The results showed that using this application on mobile platforms significantly reduces the spread of fake images across social networks. In addition, this can be used as a false image verification method in digital authentication, court evidence assessment, etc. It develops and tests reliable fake image detection programs by combining the results of metadata analysis (40%) and neural network output (60%).

Kim's and Lee's digital forensics techniques are needed to detect manipulation and fake images used for illegal purposes. Thus, the researchers in this study have been working on an algorithm to detect fake images through deep learning technology, which has achieved remarkable results in modern research. First, a converted neural network is applied to image processing. In addition, a high pass filter is used to get at hidden features in the image instead of semantic information in the image. For experiments, modified images are created using intermediate filter, Gaussian blurring, and added white Gaussian noise.

Raturi's 2018 architecture was proposed to identify counterfeit accounts in social networks, especially on Facebook. In this research, a machine learning feature was used to better predict fake accounts, based on their posts and the placement on their social networking walls. Support Vector Machine (SVM) and Complement Naïve Bayes (CNB) were used in this process, to validate content based on text classification and data analysis. The analysis of the data focused on the collection of offensive words, and the number of times they were repeated. For Facebook, SVM shows a 97% resolution where CNB shows 95% accuracy in recognizing Bag of Words (BOW) -based counterfeit accounts. The results of the study confirmed that the main problem related to the safety of social networks is that data is not properly validated before publishing. In a 2017 study by Bunk et al, two systems were proposed to detect and localize fake images using a mix of resampling properties and deep learning.

This research develops an approach that takes an image as input and classifies it, using the CNN model. For a completely new task/problem, CNNs are very good feature extractors. It extracts useful attributes from an already trained CNN with its trained weights by feeding your data at each level and tuning the CNN a bit for the specific task. This means that CNN can be retrained for new recognition tasks, enabling them to build on pre-existing networks. This is called pre-training, where one can avoid training a CNN from the beginning and save time. CNN can carry out automatic feature extraction for the given task. It eliminates the need for manual feature extraction, since the features are learned directly by CNN. In terms of performance, CNNs outperform many methods for image recognition tasks and many other tasks where it gives a high accuracy and accurate result. Another key feature of CNNs is weight sharing, which basically means that the same weight is used for two layers in the model. Due to the above features and advantages, CNN is used in this research in comparison to other deep learning algorithms.

2. RESEARCH METHODOLOGY

This research explores a supervised machine learning classification problem [14,18], where the label or category of the input sample is known as the training phase. There are two labels or classes: the original image class and the fake image class. The researcher uses the deep learning technique via a conventional neural network (CNN). The most serious challenge in the image and video forgery detection field is the fake face image detection. Fake face images can be used to create fake identities on social media networks, thus stealing personal information illegally. For instance, the fake image generator can be used to produce images of celebrities with inappropriate content, which has hazardous consequences. In this section, the proposed deep learning framework with the pairwise learning strategy is introduced in detail. The proposed two-step learning method that combines the CFF based on pairwise learning strategy and the classifier learning is presented in Figure 1. Introducing the supervised learning strategy in the fake face image detection the problems related to both difficult collection of training samples generated by all possible GANs and the need to retrain the fake face detector to obtain an effective model for the fake face images generated by a new GAN, are addressed. Specifically, to overcome these problems, the fake and real images are paired and followed by using the pairwise information to construct the contrastive loss to learn the discriminative common fake feature (CFF) by the proposed CFFN. Once the discriminative CFF is learned,

the classification network captures the discriminative CFF to identify whether the image is real or fake. The details of the proposed method are described in the following.

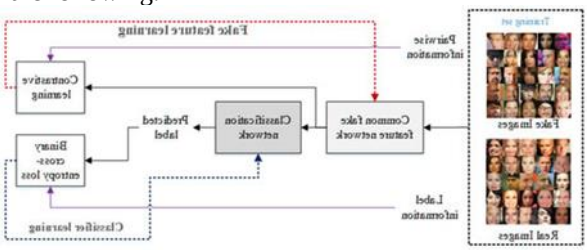

FIGURE 1. The flowchart of the proposed fake face detector is based on the proposed common fake feature network with the two-step learning approach.

Input Features for Neural Networks: Features in a neural network are the variables or attributes in the data set where extraction of features is a fundamental step in automated methods based on approaches to machine learning. The goal is to obtain useful data characteristics. In order to classify images, convolution neural networks use features. Such features are taught by the network during the training process itself. Features aim to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). Then this new simplified set of features should be able to summarize most of the details in the original set of features. Therefore, from a combination of the original set, a condensed version of the original features can be produced.

Block-based method: This method depends on dividing the image into blocks of different shapes, including squares and circles that can be non-overlapping or overlapping to be used afterwards in the preprocessing stage. Then, two main steps are performed to detect fake image:

•Features are first extracted from these blocks and compared to determine their level of similarity.

A) Feature extraction step

There exist many techniques for features extraction in block-based methods. We can cite for example:

•Discrete Cosine Transform (DCT) which is one of the most widely used techniques in CMFD. It is based on frequency transform and known for its robustness against noise addition and JPEG compression.

• Texture and intensity-based features used in images that contain certain patterns or textures, for example we mention pictures that contain landscapes characterized by containing a certain density and smoothness, such as grass, trees, ground and sky.

B) Matching step

To find similar blocks, the matching process is used to compare the features for each block and then match them to determine the manipulated area. Here is some of matching techniques for block-based methods:

• Sorting: Sorting is a way that orders the features, and it is commonly used in the matching process of block-based approaches. The sorting techniques include KD-Tree, Lexicographical, and Radix.

• Euclidean distance: Euclidean distance is measuring the distance between two points or two vectors in Euclidean space. After arranging the calculated distances, similar blocks are identified, and then the suspected regions are distinguished in the image.

C) Common Fake Feature Network

Many advanced CNN can be used to learn the fake features from the training set. Xception Network was used to capture the powerful feature from the training images in a purely supervised way. Other advanced CNNs, such as DenseNet, ResNet, Xception, can also be applied to the fake face detector training. However, most of these advanced CNNs are trained in a supervised way, so the classification performance depends on the training set. Rather than learn the fake features from all the GANs' images, we seek the CFF over different GANs. In this way, a suitable backbone network is needed for learning CFFs. However, the traditional CNNs (e.g., the DenseNet) are not designed to learn the discriminative CFF. To overcome this shortcoming, we propose integrating the Siamese network with the Dense Net, developing the CFFN to achieve discriminative CFF learning.

A dense block is a basic component in the Dense Net , which is one of the state-of-the-art CNN models for image recognition. However, it is trained by the supervised learning strategy, while the proposed pairwise learning strategy for the CFFs denotes a semi-supervised learning strategy. The proposed CFFN is a two-streamed network designed to allow the pairwise input for CFF learning. On the other hand, the traditional CNNs, which are single-streamed networks, are unable to receive the paired information; thus, the common features can be difficultly learned by the traditional CNNs. In the proposed CFFN, the backbone network can be any of the advanced CNNs, such as ResNet, Xception, or DenseNet. Once the backbone network is trained to have the best feature representation ability, the performance of the fake image recognition can be improved as well. To this end, DenseNet is selected as a backbone network of the proposed CFFN.

Moreover, it is well known that CNNs capture the hierarchical feature representation from a low level to a high level. In other words, the CNNs use only on high-level feature representation to identify whether the image is fake or not. However, the CFFs of fake face images may not exist only in the high-level representation but also in the middle-level feature representation. Inspired by this work, the cross-layer features are integrated into the classification layer to improve the fake image recognition performance, as shown in Figure 2.

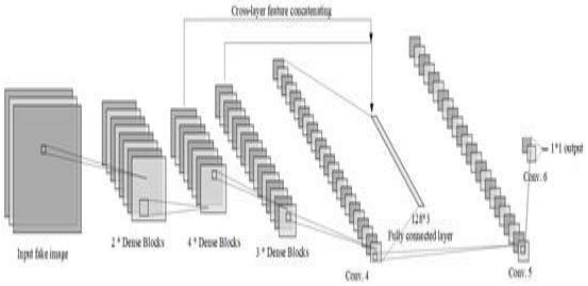

FIGURE 2. The structure of the proposed common fake feature network.

3. EXPERIMENTAL RESULT

The proposed fake-image detection with robust hashing was experimentally evaluated in terms of accuracy and robustness against image manipulations.

A) Experiment setup: In the experiment, four fake-image datasets: Image Manipulation Dataset [31], UADFV [26], CycleGAN [10], and StarGAN [11] were used. The details of datasets are shown in the figure. The datasets consist of pairs of a fake-image and the original one. As one of the state-of-the-art fake detection methods, Wang's method [23] was compared with the proposed one. Wang's method was proposed for detecting images generated by using CNNs including various GAN models, where a classifier is trained by using ProGAN.

FIGURE 3. Example of the Datasets

B) Results without additional manipulation: The proposed method had a higher performance than Wang's method in terms of both AP and Acc (fake). In addition, the performance of Wang's method heavily decreased when using the image manipulation and UADFV datasets. The reason is that Wang's method focuses on detecting fake images generated by using CNNs. The image manipulation dataset does not consist of images generated with GANs. In addition, although UADFV consists of images generated by using the method, they have the influence of video compression.

4. CONCLUSION

In this paper, we proposed a novel fake-image detection method with robust hashing for the first time. Recently, electronic attacks have spread in Saudi Arabia. There is currently no clear vision nor a unified framework to protect us against the dangers of piracy and threats, especially about the penetration of social media and the spread of false accounts. This has led Saudi Arabia to invest in information security, which is concerned with protecting the technical infrastructure from hacking and focuses on developing techniques and tools to protect social media from electronic attacks and threats. This research has contributed to the rapid detection of fraud in social media, especially in the field of images, thus solving the problem of spreading rumors and promoting false news on social networking sites and helps communities seeking to protect their technical infrastructure from piracy and cyber threats and to strengthen their information security, where the crime of image forgery poses a danger to societies. There are some problem and limitations in neural networks including it computationally expensive, requiring the use of powerful and distinct processing units. Without a good CPU, neural networks are quite slow to train for complex tasks. Although various robust hashing methods have been proposed to retrieve similar images to a query one so far, a robust hashing method proposed by Li et al was applied to various datasets including fake images generated with GANs. In the experiment, the proposed method was demonstrated not only to outperform a state-of-the-art but also to be robust against the combination of image manipulations.

REFERENCE

[1]. L. Verdoliva, "Media forensics and deepfakes: An overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.

[2]. Damian Radcliffe, Amanda Lam, "Social Media in the Middle East," [online] Available: https://www.researchgate.net/publication/32318. 5146_Social_Media_in_the_Middle_East_The_Story_of_2017 [Accessed 06 Feb 2019].

[3]. Y. Sugawara, S. Shiota, and H. Kiya, "Checkerboard artifacts free convolutional neural networks," APSIPA Transactions on Signal and Information Processing, vol. 8, p. e9, 2019.

[4]. R. Raturi, (2018). Machine Learning Implementation for Identifying Fake Accounts in Social Network. International Journal of Pure and Applied Mathematics, 118(20), 4785-4797.
[5]. J. Bunk, J. Bappy, H. Mohammed, T. M. Nataraj, L., Flenner, A., Manjunath, B., et al. (2017). Detection and Localization of Image Forgeries using Resampling Features and Deep Learning. University of California, Department of Electrical and Computer Engineering, USA.
[6]. S. Aphiwongsophon, & P. Chongstitvatana, (2017). Detecting Fake News with Machine Learning Method. Chulalongkorn University, Department of Computer Engineering, Bangkok, Thailand.
[7]. Y. Kinoshita and H. Kiya, "Fixed smooth convolutional layer for avoiding checkerboard artifacts in cnns," in Proc. in IEEE International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 3712–3716.
[8]. T. Osakabe, M. Tanaka, Y. Kinoshita, and H. Kiya, "Cyclegan without checkerboard artifacts for counter-forensics of fake-image detection," arXive preprint arXive:2012.00287, 2020. [Online]. Available: https: //arxiv.org/abs/2012.00287 .
[9]. T. Chuman, K. Iida, W. Sirichotedumrong, and H. Kiya, "Image manipulation specifications on social networking services for encryption-thencompression systems," IEICE Transactions on Information and Systems, vol. E102.D, no. 1, pp. 11–18, 2019.
[10]. Selling Stock. (2014). Selling Stock. [online] Available at: https://www. selling-stock.com/Article/18-billion-images-uploaded-to-the-web-everyd [Accessed 12 Feb 2019].
[11]. Li, W., Prasad, S., Fowler, J. E., & Bruce, L. M. (2012). Localitypreserving dimensionality reduction and classification for hyperspectral image analysis. IEEE Transactions on Geoscience and Remote Sensing, 50(4), 1185–1198.
[12]. A. Krizhevsky, I. Sutskever, & G. E. Hinton, (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 1097–1105.
[13]. K. Ravi, (2018). Detecting fake images with Machine Learning. Harkuch Journal.
[14]. L. Zheng, Y. Yang, J. Zhang, Q. Cui, X. Zhang, Z. Li, et al. (2018). TICNN: Convolutional Neural Networks for Fake News Detection. United States.
[15]. T. Chuman, K. Kurihara, and H. Kiya, "Security evaluation for block scrambling-based etc systems against extended jigsaw puzzle solver attacks," in Proc. of IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 229–234.
[16]. H. A. Alberry, A. A. Hegazy and G. I. Salama, "A fast SIFT based method for copy move forgery detection", Future Computing and Informatics Journal, Elsevier, 2018, 3, pp. 159-165.
[17]. K. R. Revi and M. Wilscy, "Scale invariant feature transform based copy-move fake image detection techniques on electronic images—A survey", 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), 2017, pp. 2315-2318.
[18]. Y. Zheng, Y. Cao, and C.-H. Chang, "A puf-based data-device hash for tampered image detection and source camera identification," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 620– 634, 2019.
[19]. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In European Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 21–37.
[20]. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 539–546.
[21]. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 685–694.
[22]. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. (IJCV) 2015, 115, 211–252.

