A Detect method for deepfake video based on full face recognition

Kai Feng[1] ,Jan Wu[1] ,Min Tian[2]
1. School of Cyberspace Security,Shandong University of Political Science and Law, Jinan, China
2. Zhi Zhe Si Hai (Beijing) Technology Company, Ltd., Beijing, China
fkdhy@163.com, jinanwujian@163.com, huimiao_zz@163.com
Corresponding Author: Kai Feng Email: fkdhy@163.coms

***Abstract*—In recent years, with the continuous upgrading of computer hardware and the continuous development of deep learning technology, new multimedia tampering tools can make it easier for people to tamper with faces in videos. Tampered videos produced by these new tools may hardly be detected by human, so we need effective method to detect these face-tampered videos. Current popular video face tampering technologies mainly include Deepfake technology based on self-encoder and Face2face technology based on computer graphics. In this paper, we propose a new method for tamper video detection based on the full faces. Facenet algorithm is used here to compare the similarity between real and fake video faces. Finally, in the experimental part, the results showed a significant effect.**

Keywords—deepfake, facenet, convolution network, machine learning

I. INTRODUCTION

With the rapid development of Internet communication, at the same time, the image processing technology and artificial intelligence technology is rapid development, people can easily make the video, and can quickly posted on the Internet. These multimedia materials have become a major source of we-media information and play a great role in promoting the real-time transmission of information. At the same time, convenient shooting technology also brings convenience to the record of social activities, such as traffic accident handling, insurance claims and judicial evidence collection. However, on the other hand, the convenience of the production and dissemination of images and videos has also brought some negative effects. For example, some people have made some improper tampering of the images and videos they have taken and then transmitted them to the Internet. If we do not identify and limit these tampered images and videos, they will quickly spread across platforms online and could have a very serious impact on society. Therefore, how to detect these tampered images and videos is becoming more and more important and attracts more and more attention.

Photoshop and drawing tools, such as image editing software is usually used in image tampering, while video modify software tools are also countless, these are convenient for people to change the video, so as to achieve a certain personal purpose, if only engage in recreational activities, that is no problem, but in many cases, can use these technologies to achieve the purpose of some illegal, it is for the video's authenticity has brought some challenges. And previous image and video tampering technology is mainly through the way of manual operation, to achieve the perfect result, and this kind of professionals are not much more special, but the situation has changed over the years, with the development of science and technology, artificial intelligence algorithm can easily forge out all kinds of effect, make people become very easy to tamper with the video.

With the continuous upgrade of multimedia processing hardware facilities and the continuous progress of new technologies in recent years, especially the explosive growth of big data processing technology and artificial intelligence technology, new multimedia forgery and tampering technologies have emerged. Among them, the most popular technologies are Generative adversarial Networks (GAN)[1],Deepfake[2-3] and Face2face[4] tamper technology. GAN is a new type of network structure, in which the discriminant network and generating network promote each other. A trained GAN can generate a large number of counterfeit samples that are difficult to be distinguished by naked eyes. Deepfake uses technologies such as auto-encoders to replace faces in video; Face2face uses computer graphics to transfer the facial expressions of the characters in the video. These techniques make it easier and harder to fake videos, and pose a huge challenge to video identification.

In this paper, a face tampering video detection method is proposed, takes the difference of the full face image of the video as the feature to carry out the tampering detection, which can fully improve the accuracy of face detection and is not easy to be disturbed by the video content information. First of all, we use face recognition technology to identify faces, and then judge the Angle of faces. We do the same processing for faces in real videos and fake videos, and then conduct comparison experiments with the processed data. According to the experimental results, the selected faces are more different and differentiated.

978-1-7281-5224-0/20/$31.00 ©2020 IEEE

II. PREVIOUS WORK

A. Forgery image synthesis technology based on GAN

GAN[1] is a neural network framework proposed by Goodfellow et al in 2014. GAN is mainly composed of two parts: identifying and generating network, discriminant network used to distinguish which is the image and counterfeit images, generation network is used to generate the image, discriminant network and generated network used clever loss function to confrontation and promote each other, eventually trained GAN can generate realistic fake samples.

The training process of GAN is mainly divided into the following two stages: (1) The training stage of generating network: In this stage, it is necessary to fix the parameters of discriminating network and only train generating network. After the random noise passes through the generated network, the generated counterfeit samples are put into the discriminant network, and the loss function is used to optimize the generated network. (2) Discriminant network training stage: In this stage, network parameters need to be fixed and only discriminant network is trained. The real sample and the forged sample generated by the generated network are respectively used as the input to optimize the parameters of the discriminant network.

B. Face tampering detection technology in the video

One traditional methods to detect tamper image are based on Color filter array (Color filter array (CFA) detection method [5-8], based on the camera response nonuniformity (Photo response non - uniformity, PRNU) testing method of noise [9-11], detection method based on fuzzy inconsistencies [12, 13], based on the imaging detection method of shading [14-16] and trace based on geometry transform and interpolation [34-35] detection method, and so on. However, these technologies basically rely on some features of the equipment for detection, which has certain limitations. After all, there are many types of equipment, and not all images have certain equipment information. Moreover, such detection method is more complex, which is not as simple and effective as deep learning algorithm.

In paper [17], a detection method specifically targeted at Deepfake tampering videos was proposed. The paper pointed out that, since the human face image samples in the closed state of human eyes were rarely selected during the early training of Deepfake GAN network, the characters in the fake videos generated in this way would always be in an open state. Therefore, this paper proposes that tamper detection can be carried out by detecting the blinking condition of the person in the video. However, if the experimenter adds the sample of the face image in the state of closed eyes in the training process of GAN network, the detection method proposed in paper [17] will be invalid. Paper [18] pointed out that because of output Deepfake facial transformation network will face image feature points location with the input of the location of the face image feature points there is a deviation, and because Deepfake areas to tamper with the inside of a face, so deepfake tamper with the characters in video face angle difference between inside and outside will be significantly greater than the real video angle, inner and outer faces of the characters in the paper [18] proposes to use video characters in the size of the inner and outer face angle to tamper detection. However, as the face transformation network used by deepfake continues to evolve, input-output face feature point inconsistencies will gradually diminish, and the detection effect of the method used in the paper [18] will don't work[24].

Paper [19] uses the inconsistency of color component distribution in HSV channel and YCbCr channel to distinguish real face images and false face images generated by artificial intelligence tools. Paper [20] pointed out that the distribution of feature points in real face images and fake face images generated by GAN is inconsistent, so this inconsistency can be used to distinguish real and fake faces. Similarly, paper [21] uses various facial details extracted from face images to detect tampered faces. The facial image extracted from the video frame is detected by Convolutional neural network (CNN) of the paper [22-23]. However, none of the methods mentioned in the above papers takes into account the temporal relationships between video frames. Paper [24] makes use of Recurrent neural network (RNN) to detect face images extracted from video sequences. Compared with the detection method using CNN directly, the timing relationship between video frames can be utilized by using RNN for detection. However, since RNN directly extracts content information from video frame sequence, the detection method using RNN will be interfered by video content to some extent.

In addition, due to the connectivity of face recognition and face tampering detection, some typical face recognition methods can be used to deal with face tampering detection, such as FaceNet. FaceNet can directly obtain the mapping relationship from the input face image to Euclide space, and can take advantage of the Triplet loss to improve the network performance.

III. FULL FACE DETECTION ALGORITHM

A. Full face selection

Current face recognition technology has been more mature, in this article we uses the MTCNN facial recognition technology, which is the judgment of the face is mainly rely on eyes to do this, first identify the position of the eyes, and then based on the relative distance between the eyes to determine the angle of the face, though this operation will have a certain roughness, but also has certain reference value, can be found in subsequent trials have some significant differences, has distinguish effect.

B. *Introduction of algorithm model*

In this paper, the network structure of face recognition mainly includes feature extraction module and classification module. The network structure design is mainly shown in Figure 1.

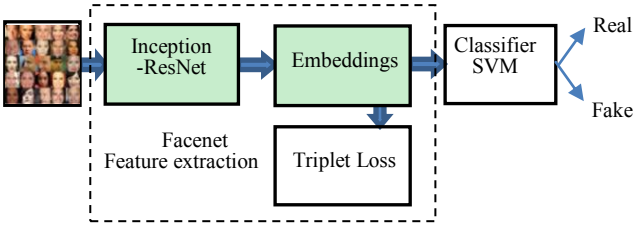

Fig. 1. The forms of large-scale network survivability associations.

Among them, feature extraction uses the facenet network proposed by Google, and its main function is to map the facial image to the image represented by multi-dimensional space vector in space, and then calculate the spatial Euclidean distance between feature vectors to represent the degree of acquaintance between two images [25]. The spatial Euclide distance between the same face is smaller, and the spatial Euclide distance between different faces is longer, so face recognition can be realized by spatial mapping of face images. Facenet uses image mapping method based on deep neural network and loss function based on triple losses to train the neural network. The loss function is shown in Equation (1).

$$Loss = \sum_{i=1}^{N}\left[\| f(A^{(i)}) - f(P^{(i)}) \|_2^2 - \| f(A^{(i)}) - f(N^{(i)}) \|_2^2 + \alpha\right]_+$$

Equation（1）

In network training, a group of face images are first input into a convolutional neural network structure (such as VGGnet Googlenet ResNet) for feature extraction, and then the feature L2 is normalized so that all image features are mapped to a hypersphere, the difference caused by the imaging environment of the sample is avoided, and the image embedded space L2 is finally obtained after normalization. Eigenvectors (embedded) take the loss of triplet as the loss function of model optimization. The guiding principle of the optimization is to make the distance between the same face samples less than the distance between different face samples. By using a large number of face image updates and network parameters optimization, the final Facenet network can directly output multidimensional feature vectors based on face images, and can compare the differences of different images. Facenet was originally used in traditional face recognition and achieved good recognition effect. In our work, the trained network structure 20170512-110547 was directly used for feature extraction. The main model adopted deep network sensor-RESnet, and the output face feature dimension was 128.

In traditional face-based face recognition applications, when the eigenvector Euclide distance of two face images is greater than 1.1, the two faces are considered to belong to different people. But in the deep forged image, the forged image often has a high similarity with the original image, so the replication threshold cannot be used as the classification standard. Therefore, in order to determine a reasonable Euclidean distance as the classification standards, classification module, the characteristics of the facial feature extraction is the embedded space characteristic vector (inset) as input, using machine learning algorithm (SVM) learning Euclidean distance is the characteristic of the difference between true and false face images, and through machine learning to choose reasonable classification standard, realize the true and false face image of two kinds of classification.

IV. EXPERIMENT

The experimental platform was windows10/Opencv/ Python3.7/Tensorflow1.14. The dataset used in this paper is Celeb-df[26], which is a high quality deepfake dataset.To prove true video and fake video gap between successive frames, verify the feasibility of the solution, we first choose real video and based on comparing the same fake video, in which we can choose two groups of data, the data we choose is only full face, the different about full face and random face is show shown in the figure 2:

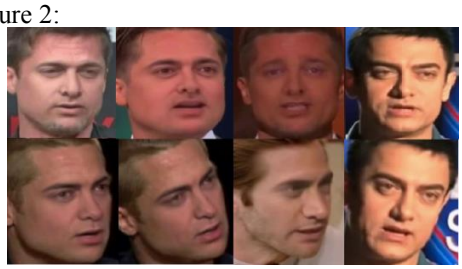

Fig. 2. The face of the dataset

In early tests, mainly focus on the differences between real face image and fake face image, select the part of the real face and fake face to do comparison, in the process of selection, this paper chose to compare multiple sets of data, a set of data is random intercept face, another set of data is intercepted face after face Angle, only choose is face to compare. After data acquisition, the images used in the test were first mapped to a certain dimensional space through the deep learning network, and then the distance between real-real and real-fake was calculated. It was found that using this method, the positive face Angle was more accurate. The statistical results of the facenet algorithm were shown in the table 1:

TABLE I. STATISTICAL OF FACE DISTANCE

Id	A	B	C
mean	1.1877	1.1088	0.891
var	0.00575	0.065	0.0839
std	0.0758	1.1998713	0.9767425

From above table, column A is the statistic between the real full face and fake full face, B column is the statistic about random real face and random fake face, C column is the statistic about full real face the distance between the other random face, intuitively look at the data, there is a visible line, by giving A column data and C column data tested for p, the p value is 0.000002, far less than 0.05, the two groups of data with a very large degree of differentiation, draw the uniform line of the above data distribution, as shown in the Fig. 3:

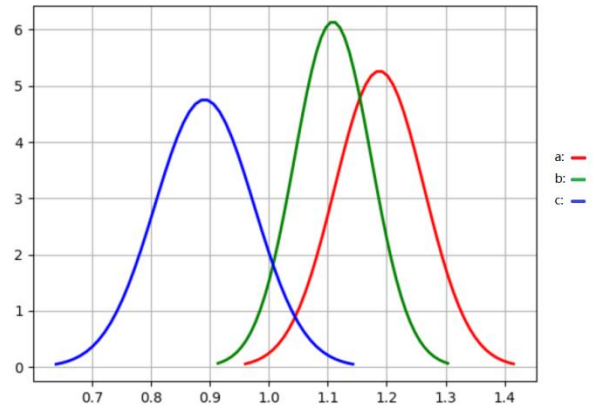

Fig. 3. the distribution of the data uniform

From above Fig, the normal distribution of three group data is completely different, it is also can be used in the identification of fake video features. Used normal distribution to test the classification effect of real and fake images, in which 100 real face images and 50 fake face images were used as training data to obtain the normal distribution model, and then 100 real and fake images were used for testing, with an accuracy rate of 86%. The specific confusion matrix is shown in the table Ⅱ :

TABLE II. CONFUSION MATRIX

	T	F
P	96	24
N	76	4

Finally, the vector generated by Facenet was used as training and testing data, the dimension is 512. SVM was used for classification, and get the accuracy rate was 99.1%, this is a state-of-the art performance.

V. CONCLUSION

In this paper, a method of forgery video detection based on positive face is proposed. Combining facenet and other face recognition technologies, a high quality face data set CeleB-DF was tested, and the experimental results show that the method has statistical differentiation. In addition, the relatively small amount of data that is used in the experiment, and the difference between fake video because of the different data, this method can be used as a reference or currently dimension to handle, used to distinguish words directly to adjust the parameters according to different data, if a feature as a combined with deep learning that can obtain the corresponding results better, the other on the selection is face, need further improve the technology, the main is to use eye recognition technology to detect face, we would consider follow-up training network face Angle is face to more accurate judgment, believes that the direction will have more applications.

ACKNOWLEDGMENT

This work is supported by Program for Young Innovative Research Team and Big Data and Artificial Intelligence Legal Research Collaborative Innovation Center in Shandong University of Political Science and Law; Projects of Shandong Province Higher Educational Science and Technology Program under Grant No.J16LN19, J18KA357, J18KA383; Shandong Province Soft Science Research Project under Grant 2019RKB01369; Open Fund of the Key Lab of Forensic Science KF202015, Ministry of Justice, China (Academy of Forensic Science).

REFERENCES

[1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al. Generative adversarial nets[C]. International Conference on Neural Information,Processing Systems, 2014: 34-42.

[2] Deepfake-Faceswap, https://github.com/deepfakes/faceswap

[3] Faceswap-GAN, https://github.com/shaoanlu/faceswap-GAN.

[4] Thies J, Zollhofer M, Stamminger M, et al. Face2Face: Real-TimeFace Capture and Reenactment of RGB Videos[C]. 2016 IEEEConference on Computer Vision and Pattern Recognition (CVPR),June 27-30, 2016. Las Vegas, NV, USA. Piscataway, NJ: IEEE,2016: 2387-2395.

[5] Y. Li, S. Lyu. Exposing DeepFake Videos By Detecting FaceWarping Artifacts. 2018: arXiv preprint arXiv:1811.00656.

[6] Dirik A E, Memon N. Image Tamper Detection Based on Demosaicing Artifacts[C]. 2009 16th IEEE International Conference onImage Processing (ICIP), November 7-10, 2009. Cairo, Egypt.Piscataway, NJ: IEEE, 2009: 1497-1500.

[7] Ferrara P, Bianchi T, de Rosa A, et al. Image Forgery LocalizationVia Fine-Grained Analysis of CFA Artifacts[J]. IEEE Transactionson Information Forensics and Security, 2012, 7(5): 1566-1577.

[8] X.L. Zhang, Z. Fang, X.P. Zhang, Forgery Detection via Inter-channel Correlation of CFA Images[J]. JOURNAL OFAPPLIED SCIENCES—Electronics and Information Engineering,2015, 33(1): 87-94.

[9] S. Peng, Y.Y. Peng, C.Y. Xiao, Image tampering detection algorithm based on CFA interpolation[J]. Transducer and MicrosystemTechnologies, 2015, 34(6): 141-144. [J].

[10] Chierchia G, Parrilli S, Poggi G, et al. PRNU-based Detection ofSmall-size Image Forgeries[C]. 2011 17th International Conference on Digital Signal Processing (DSP), July 6-8, 2011. Corfu,Greece. Piscataway, NJ: IEEE, 2011: 1-6.

[11] Chierchia G, Poggi G, Sansone C, et al. A Bayesian-MRF Approachfor PRNU-Based Image Forgery Detection[J]. IEEE Transactionson Information Forensics and Security, 2014, 9(4): 554-567.

[12] Lin X F, Li C T. Refining PRNU-based Detection of Image Forgeries[C]. 2016 Digital Media Industry & Academic Forum (DMIAF), July 4-6, 2016. Santorini, Greece. Piscataway, NJ: IEEE,2016: 222-226.

[13] Bahrami K, Kot A C. Image Tampering Detection by ExposingBlur Type Inconsistency[C]. 2014 IEEE International Conferenceon Acoustics, Speech and Signal Processing (ICASSP), May 4-9,2014. Florence, Italy. Piscataway, NJ: IEEE, 2014: 2654-2658.

[14] Bahrami K, Kot A C, Li L D, et al. Blurred Image Splicing Localization by Exposing Blur Type Inconsistency[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(5):999-1009.

[15] Johnson M K, Farid H. Exposing Digital Forgeries through Chromatic Aberration[C]. Proceeding of the 8th workshop on Multimedia and security - MM&Sec '06, September 26-27, 2006. Geneva,Switzerland. New York, USA: ACM Press, 2006: 48-55.

[16] Mayer O, Stamm M C. Accurate and Efficient Image Forgery Detection Using Lateral Chromatic Aberration[J]. IEEE Transactionson Information Forensics and Security, 2018, 13(7): 1762-1777.

[17] Vazquez-Padin D, Perez-Gonzalez F, Comesana-Alfaro P. A Random Matrix Approach to the Forensic Analysis of Upscaled Images[J]. IEEE Transactions on Information Forensics and Security,2017, 12(9): 2115-2130.

[18] Y. Li, M.C. Chang, S. Lyu, In ictu oculi: Exposing ai generatedfake face videos by detecting eye blinking. 2018: arXiv preprintarXiv:1806.02877.

[19] R. Wang, L. Ma, F. Juefei-Xu, et al. FakeSpotter: A Simple Baseline for Spotting AI-Synthesized Fake Faces.2019: arXiv preprintarXiv:1909.06122.

[20] H. Li, B. Li, S. Tan et al. Detection of deep network generated images using disparities in color components. 2018: arXiv preprintarXiv:1808.07276.

[21] X. Yang, Y. Li, H. Qi et al. Exposing GAN-synthesized Faces Using Landmark Locations. 2019: arXiv preprint arXiv:1904. 00167.

[22] Matern F, Riess C, Stamminger M. Exploiting Visual Artifacts toExpose Deepfakes and Face Manipulations[C]. 2019 IEEE WinterApplications of Computer Vision Workshops (WACVW), January7-11, 2019. Waikoloa Village, HI, USA. Piscataway, NJ: IEEE,2019: 83-92.

[23] Afchar D, Nozick V, Yamagishi J, et al. MesoNet: A Compact Facial Video Forgery Detection Network[C]. 2018 IEEE International Workshop on Information Forensics and Security (WIFS),December 11-13, 2018. Hong Kong, China. Piscataway, NJ: IEEE,2018: 1-7.

[24] Nguyen H H, Yamagishi J, Echizen I. Capsule-forensics: UsingCapsule Networks to Detect Forged Images and Videos[C].ICASSP 2019 - 2019 IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP), May 12-17, 2019. Brighton, United Kingdom. Piscataway, NJ: IEEE, 2019: 2307-2311.s

[25] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. CVPR, 2015

[26] Li, Yuezun, et al. "Celeb-DF: A New Dataset for DeepFake Forensics." arXiv: Cryptography and Security (2019).