International Journal of Grid and Distributed Computing
Vol. 13, No. 2, (2020), pp. 2355-2365

DfakeTect: A Review on Deepfake Detection System

[1]Ashish Kumar Mishra, [2]Sushmita Bharti, [3]Raj Singh, [4]Jyoti, Tauseef Ahmad[*]

Department of Information Technology,
Rajkiya Engineering College Azamgarh, UP, India
{[1]ashishmishra1897, [2]bhartisushmita21, [3]drajsinghvns99, [4]sjyotirec1998}@gmail.com,
[*]tauseefahmad@zhcet.ac.in

Abstract

With increase in the popularity of social media nowadays people sharing everything on social media and online media content is growing progressively. In between all of that a technology termed as "Deep Fake" has emerged which is used to manipulate media content. As this technology is accessible to everyone, now fake videos are being shared extensively which is a threat to a person and its image. It can also be used to create political instability in the country. Revenge Pornography was the first few things for which this technology was used. To tackle this situation several methods have been proposed which can detect if the video is being altered or not. The paper provides a brief summary of some of the novel works done in this area, the datasets available to support the research. We have reviewed the evolution of deep fakes and some of the methods to counter them and facilitate the further development of more improved and effective technology and methods to counter the challenges of the deep fakes.

Keywords: *Deepfake, GAN, Swapping, Face2Face, Visual Artifacts*

1. Introduction

Altering faces in images has been accessible and pulled in broad consideration recently. Substituting face on images isn't new as we can propose. Now people have begun to use this technology in their everyday life. Improvement of innovation has assumed an incredible function being developed of mankind. Among these incredible innovational progressions "DEEPFAKE" was risen. Deepfake got its name from the word Deep Learning and Fake. It is a technology which has utilized profound learning models to make counterfeit images or audios by changing the face or entire body of an individual in a picture or video by someone else. With the recent advancement in online streaming videos there is a need for a system which can be used to authenticate the forgery in the videos. And with the recent development of deep fake it has now become a serious concern for the digital content that is shared on several online video streaming platforms. As deepfakes have been used for several malpractices like defaming the politicians and celebrities by using their faces in pornography, making videos of politicians to be present in some events they never been to and many more unethical things are being used to defame the people.

To counter such unethical practices many researchers have come with several different novel approaches and algorithms. Some of these approaches have been summarized further in the upcoming sections. However, there is still a need for improvement in the work done till now for deepfake detection because as soon as detection algorithms come into play, the deepfake video generators try to counter the detection by improving their own technique. To accomplish this literature work 7 papers are focused and about 42 papers have been used to maintain the accuracy of the proposed work.

Moving further in the paper, in section 2 we have discussed the existing algorithms that are used to generate fake videos along with the available datasets for research. In section 3 approaches that many researchers have used to detect the fake videos are presented. After that there is a short discussion on what will be the future directions of work in this area is in Section 4.

ISSN: 2005-4262 IJGDC

2. Literature Review

Deep fakes have now started to impact the trust of people in media content that are being posted online or are present publicly as due to them before sharing them people fear of being it fake as they are losing their faith in them. They could even create misery and may have negative impacts, they increase disinformation and even could fire political pressure or war. This is particularly basic these days as the advances in deepfakes are progressively advancing and online media platforms can spread these contents rapidly.

It is hard for a person to detect a deep fake with naked eyes. In reaction to the increasing number of sensible altered content, researchers are trying hard to come out with a strategy for the detection of such manipulations. As a consequence, the detection of such fake videos is one of the most prominent challenges in the area of forensic analysis of online media. The quality of deepfake generated is increasing day by day so the performance of the detection algorithms needs to be improved accordingly. With several researches going on many techniques have been proposed to detect the manipulation in a video. Many other important works have been discussed in [14],[32]-[34], [37]-[42]. Some of the notable works done are as follows.

2.1. DeepFake

Deepfakes [1] are the most used and efficient face swapping technique used for video alteration. The face of a person in the video is changed by the face of a different person, preserving the background content and the expressions of the original face remain unchanged. Firstly, the encoder algorithm is fed with faces of two different people so that it can learn similarities between the two faces, and then it reduces the faces up to features they share, meanwhile compressing the face images in the process. Afterwards a second algorithm known as decoder is then fed with the compressed faces which learns to recover the faces from the compressed face images. Since there are two faces, two decoders are taken for training; one decoder recovers the face of the first person, and another decoder recovers the face of the second person. For swapping the face, encrypted images are sent to the different decoders i.e. decoders are swapped. Now the decoder reconstructs the face of the first person with the expressions and orientation of the second person. For it to be more realistic the whole process has to be performed on each frame.

2.2 Face2Face

The method Face2Face, according to [19] is able to transpose the expressions of the face from a referenced video to a desired video in actual time. The goal is to have the expressions of the face of the referenced actor on the face of the person in the desired video and present the altered output video in a realistic method. The above described technique depends on setting a Three-Dimensional mutable face model and computing the illumination. The final result is demonstrated onto the target video.

This technique also has some limitations; Some of the assumptions in the technique has sometimes led to generation of some artifacts in the video which make it traceable. The long hair and beards make it difficult for the technique to replace.

2.3. FaceSwap

A method [2] which is used to change the face of a person in the image which is taken as a recommendation, with the face and attributes of the input image faces. The face is first tailored on the basis of facial landmarks to have a consistency in between the input image face and recommended face in terms of size and its posture. And then it is presented to an image deforming method so that the face and its background should be adjusted.

2.4. Generative Adversarial Networks

Generative Adversarial Networks [18] or GANs for short, is a deep learning technique which is used to generate new contents. They are a model architecture which are used for training generative learning techniques. The model architecture of GAN involves two sub-models: one is

ISSN: 2005-4262 IJGDC

a generator model which is used for generating new contents or examples and second one a discriminator model which is used for classifying whether the generated contents of the generator model are real or they are fake. They both compete with each other and produce the best contents.

2.5. Different Datasets Available

There are number of very useful datasets available for deepfake videos. This section discusses few of very important and commonly used datasets.

2.5.1. UADFV: The UADFV dataset consists of total 98 videos out of which there 49 real videos which have been taken from publicly available YouTube videos and the other 49 are DeepFake videos. These DeepFake videos have been generated from "FakeAPP" [5] using the DNN model.

2.5.2. Deepfake TIMIT: Deepfake TIMIT is a dataset of videos where the person identity in the videos were manipulated using the Generative adversarial model [3]. This approach, in turn, was invented from the Deepfake [1] algorithm which is based on autoencoders. While creating the dataset, they have manually shortlisted sixteen pairs of people who look similar from the openly available dataset of VidTIMIT [20]. They have two models which are trained for every subject of theirs, one of the models is trained with a lower quality (64 x 64) input/output size model, and the other model is trained with a higher quality (128 x 128) size model. Per person there are a total of 10 videos in the dataset of VidTIMIT, using these videos they have generated around 320 videos which correspond to every version of the video available, which constituted a total of around a collection of 620 videos which have their faces swapped and altered. They have kept the audio intact with no alteration there.

2.5.3. Celeb –DF: The Celeb-DF [13] dataset contains a collection of around 6000 videos out of which 10 % are real videos and the rest 90% is Deep Fake videos. The collection of the videos has over two million frames from the videos. Length of the videos in the dataset is on average around thirty seconds which corresponds to a frame rate of 30 per seconds. The original videos have been collected from freely available platforms such as "Youtube", while collecting the videos they have maintained a diversification in the distribution of type of videos. The collections of interviews of 59 celebrities are divided on the basis of their genders, ages, and skin color. The videos in the dataset are evenly distributed among the above categories also keeping in mind the area of their origin taking videos of the people from all over the world. In addition to all that also this dataset has original videos that show a variety of changes which correspond to topics like face size of the person (in pixels), orientations, surrounding lighting, and the background scenes. The DeepFake videos which have been constituted has been developed by present have been generated by manipulating and altering videos by replacing the faces of every pair of the people who have been taken into consideration. The format of the final videos is in MPEG4.0.

2.5.4. FaceForensics++: It is a rhetorical dataset which consists of one thousand original video sequences. These video frames have been altered with the techniques mentioned in the paper above like: Faceswap, Face2Face, etc. This collection of data has been constituted from around 1000 videos which have been taken from the open platform like youtube and all of these videos have aligned frontal face and does not contain any barriers which helps in a way that enables mechanized manipulation methods to develop realistic forged videos. They have also provided a binary mask which enables video and image classification and their segmentation. Apart from this they have given around thousands of other deepfake contents for further generation to increase data collection.

2.5.5. DFD: The DFD DeepFake detection dataset consists of 3, 068 DeepFake videos which have been generated from 363 original videos of twenty-eight individuals who have consented to the use of their videos. The individuals are from various categories of different genders, different ages and belong from different ethnic groups. They have not disclosed the details of the manipulation technique they have used, but most likely it has to be an improved version of the

ISSN: 2005-4262 IJGDC

starting DeepFake generation technique. It was generated by Google in collaboration with Jigsaw.

2.5.6. DFDC: The DFDC dataset was generated by running a campaign where certain actors were used. The dataset is diversified by ensuring that the actors' crowd sourced has variability in their gender, age and color. A rough estimation of the approximate distribution of the videos on the basis of gender and race across this dataset is around 74% of them is female and 26% male; the ethnic group distribution varies as 68% of them are Caucasian, 20% African-American, 9% east-Asian, and 3% of them are from south-Asia. Another dataset is presented in [7].

3. Common Approaches for Deep Fake Videos Detection

This section discusses many famous approaches used by the researchers for detecting the deep fake videos.

3.1. Deepfake Video Detection Using Recurrent Neural Networks [12]

Using the Fake Apps to generate fake videos leads to some anomalies in the videos i.e. inconsistencies in between the frames and temporal inconsistencies. These defects in the techniques can be utilized to detect if the provided video is real or not. These anomalies are introduced during the generation pipeline of such videos. Different lightning conditions, camera angles make it hard for AI encoders to create real look alike faces. Boundary effects near the face regions are present in the newly generated face and the other regions. The defect of temporal unawareness is due to the fact that the auto encoder is not aware of any of its previous works which leads to multiple anomalies. Due to inconsistent illumination flickering phenomenon near the face region can be seen in many videos.

The method used in [12] is: Firstly, they proposed a detection system which involves a convolutional LSTM for processing frames. In convolutional LSTM there are two essential components to capture the inconsistencies between different frames due to the swapping process, one is CNN which has been used for extracting features of frames and the other one is LSTM used for analyzing the temporal sequences.

A collection of 600 videos was used for evaluation of the method. From which features for each frame that are developed by the CNNs were obtained. Later, the features of successive frames are linked together and then for further analysis it was passed to the LSTM. Finally, an estimate is produced whether a sequence is either a deepfake or not. They got a 94% accuracy through there method

3.2. MesoNet: A Compact Facial Video Forgery Detection Network [9]

In [9] the author proposes a method which has the functionality which without human effort trace the face manipulation in the videos, and mainly focuses on some prominent techniques which used to create such realistic and fused videos such as Face2Face and Deepfake. They both have a similar nature of the falsifications which have made a single technique useful for both of them. The nature of falsification present in fake videos of such types are:

- Input data is compressed on a less encoding space which makes the output blurred.
- No facial re-enactment is present in some of the frames.

For the detection proposed two architectures which exploits features at mesoscopic level, i.e. in between macro and micro. Meso4, a network that begins with a series of four layers of consecutive convolutions and pooling, and further it has a big network with a hidden layer. MesoInception4, an alternative network structure with a change in convolutional layers of Meso4 by a makeover of the inception module [30]. The results show that there is a decrease in accuracy with rise of compression level in the videos. It also gives us a clue that an aggregate frame images from a video will improve the accuracy of detection of videos Also that the mouth and eyes plays an important role in the detection of fake videos that are developed by Deep fake technique. They claim that the technique has a rate of detection of around 98% for Deepfake videos and for Face2Face generated videos it is around 95 % which is a very good result.

ISSN: 2005-4262 IJGDC

3.3. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking [11]

This work in [11] has focused on the physiological signals of a human behavior as some of these physiological signals are not well represented in the synthesized videos. They have utilized the fact that there is a recognizable absence of such physiological signs identified in individuals which are not fully caught in such recordings. Signs may incorporate unconstrained and natural exercise like flickering of eye (flickering alludes to fast shutting and opening development of eye lead), moment of eyes, facial expression etc.

Deep fake generated images lack eye blinking function, as the datasets provided for training do not contain images with eyes closed. This non-presence of eye blinking is thus a significant sign to tell that the videos are coming from a different source than a cam recorder. The algorithm they have presented encompasses two parts, one is pre- processing which includes face extraction and alignment and then it is sent to the other step, Long-term Recurrent Convolutional Networks (LRCN) model. The model is composed of three parts; feature extraction (which convert the eye region into discriminative features) implemented using CNN, sequence learning implemented with a recursive neural network (RNN) along with a Long Short-Term Memory (LSTM) and prediction of the state which is a fully connected layer which gives a probability of opening and closing state. To test and train their method they used their own generated dataset (Eye Blinking Video dataset) and generated good results.

However advanced forgery developers can still create realistic blinking of the eye with improved models and improved training dataset. So, we should be focused on exploring other types of signals to counter the fake videos and improve their detection.

3.4. Image Feature Detectors for Deepfake Video Detection [8]

Authors in [8] propose a unique method to trace deepfake videos using the machine learning algorithm i.e. Support Vector Machine (SVM) [23]. It is one of the most useful and powerful methods which is used as a classification as well as regression machine learning algorithm. It has four different types of kernels, these are linear, sigmoidal, radial basis function, and polynomial function. The kernels are used to solve a lot of different types of problems in different areas of domain. SVM is applied on linear and nonlinear data sets.

They utilize the fact that there are some inconsistencies in the generated images like differences in lightning conditions, boundary effects due to swapping between the newly generated face and the rest of the region etc. they utilized this defect by grabbing the features points. The features are edge points in an image which can be traced using different class feature points detector or edge detectors such as FAST [29], KAZE [27], SURF [28], ORB [25], BRISK [26], and HOG [24] algorithms. SVM classifiers can be trained with these classifiers. Table 1 Of all the different methods confusion matrix is used for comparing different detectors with the SVM classifier. The performance of HOG comes out to be the best among the other all with an accuracy of around 94.5%.

Table 1: The Values of The Confusion Matrix.

	TR	**FR**	**TF**	**FF**	
BRISK	75	25	99	1	87%
HOG	95	5	94	6	94.5%
FAST	74	26	99	1	86.5%
KAZE	64	36	89	11	76.5%
ORB	83	17	99	1	91%

ISSN: 2005-4262 IJGDC

SURF	90	10	91	9	90.5%

3.5. Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations [10]

Authors in [10] have presented a method which exploits several limitations of computer vision and graphics which lead to some characteristic's artifacts in the produced content. Research progress is making it hard to pint out the differences between the original and fake image. However, some of the artifacts can still be pointed out. Authors have categorized these artifacts into different types as computer vision problems that are yet to be fully solved. These are Global Consistency, Illumination Estimation and Geometry Estimation.

They have proposed several different sets of features for detecting manipulated videos of different methods i.e. generated images, deepfakes, face2face. For generating faces, they used the method of [21] i.e. ProGAN and Kingma and Dhariwal's Glow method [22]. These methods generate high resolution images. There is still a lack of global consistency. Lack of similarity between the eyes i.e. high difference in color between the left and right eyes. It is the characteristics they used for detecting generated images. Manipulation of facial attributes is done by different methods, they have mainly focused on the method Face2Face in [19] which is able to put expressions of the face in the target face from a source face. For detection of fake videos by this method they have focused on features for the border of the face and the tip of the nose such as shading around the nose and artifacts along the boundary of the face region.

Face swapping is one of the most prominent methods for fake video creation since the method is publicly available. For deepfakes videos detection they have exploited missing of the reflections, and some details in the teeth and eye areas. The proposed method follows a straightforward way. Generated images have certain visual artifacts which can be used against them for their detection.

3.6. Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches [6]

Detecting fake videos on the basis of spatial and temporal manipulation needs a large set of data and clarity. Authors in [6] have used the shape of the mouth to detect the fake videos. Phonemes (the sound produced when we pronounce any word), visemes (shape of the mouth) have been considered as the detection technique. The phonemes M (mumma), P (parrot), B (brother) have grabbed their attention. In pronouncing these letters once in a while our mouths completely get close. This mismatch in phonemes and visemes they have utilized to distinguish even spatially little and transiently restricted manipulated videos from the real one.

3.6.1 Extracting phonemes: In order to extract the phonemes M, P, B associated with words they have used Google –speech –to- text API [4], which automatically transcribes the audio from video. They have gone through the recording physically to eliminate any mistakes and afterward adjusted to the sound, utilizing the P2FA [17]. This arrangement produces a succession of phonemes alongside their beginning and end time in the info sound/video.

- **Measuring Visemes:** For measuring the visemes they have used three approaches; manual, profile and third CNN. In the first approach a person expert in image analysis is given the six frames of the video and one closed mouth frame as a reference and asked to label each frame. In the second approach is an automatic detector which uses the intensity frame to tell a mouth is open or it is closed. CNN gave input of colored images, cropped the area around the mouth region and then it is resized to 128x128 pixels. The output was of the form, corresponding to (0) is open mouth and (1) is closed mouth. The trained convolutional network is used to detect open or close mouths. Table II The method was used on four deep fake data sets that are T2V-S [16], A2V [15], T2V-L, and in-the-wild.

ISSN: 2005-4262 IJGDC

Table 2. The Accuracy of The Techniques (Profile and CNN).

Dataset	Profile	CNN
A2V	96.6%	96.9%
In the wild	93.9%	97.0%
T2V-L	83.7%	71.1%
T2V-S	89.5%	80.7%
Original	99.4%	99.6%

3.7. Exposing deep Fake Videos by Detecting Face Warping Artifacts [31]

Authors [31] uses the fact that DeepFake algorithm right now is able to only generate compressed images of only few resolutions, which needs furthermore attention to match the faces to be replaced in the source video. They have shown that such videos can be easily traced by a convolutional neural network. The method they produced has saved a significant time by generating negative examples using some simple image processing technique rather than training the classifier with a large amount of real and fake videos dataset. Also helps in avoiding overfitting cases.

They trained using their own training data 4 different CNN models VGG16 [35], ResNet50 [36], ResNet101 [36] and ResNet152 [36]. For better results they have cropped the region of interest for each training example by 10 times. Table 3 shows the final fake probability. The author has taken the average of all the predictions.

Table 3. The AUC performance of their method and other methods on UADFV dataset and Deepfake TIMIT datasets.

Methods Used	UADFV Dataset	DeepFake TIMIT	DeepFake TIMIT
		LQ	**HQ**
Meso 4 **MesoInception 4**	84.3 82.1	87.8 80.4	68.4 62.7
Two Stream NN	85.1	83.5	73.5
Head Pose	89.0	--	--
Theirs VGG16 **Theirs ResNet50** **Theirs ResNet101** **Theirs ResNet152**	84.5 97.4 95.4 93.8	84.6 99.9 97.6 99.4	57.4 93.2 86.9 91.2

They claim that their method has surpassed state of the art methods present for the detection.

The Table 4 presents few findings which are very important to take care while proposing any scheme or methods for Deepfake detections.

ISSN: 2005-4262 IJGDC

Table 4. Comment on the methods.

Method	Limitations/Shortcomings
RNN + LSTM [12]	LSTM focuses on temporal inconsistencies but what about the intraframe inconsistencies?
Mesonet [9]	Did having some attention layer enhance the detector?
Eye Blinking [11]	Blinking effect of the eye can be improved with further training of data.
SVM [8]	Accuracy can be further improved by using more features.
Visual Artifacts [10]	The technique mentioned lacks generality.
Phoneme-Viseme Mismatches [6]	Including more visemes for analysis can improve the result.
Face Warping Artifacts [31]	It focuses on particular artifacts, with improving deepfake technology it may fail.

4. Conclusion and Future Work

The new developments in deep fake have improved the quality of fake videos. And it is creating a distrust among the people for the online media content. Since the technology to generate fake content is publicly available and is in reach of everyone, we need some technique which can counter them. Knowing the situation due to the deepfake, researchers have come forward on developing techniques for countering deepfake and have been posting several results related to this.

The quality of the fake contents is getting advanced data by day and so with such advancement going on, the performance of their countering methods has to be improved accordingly. As we have seen that a lot of work has been done for the detection videos in which face is altered but only few works are there which focuses on the fake videos in which the audio is also altered. So, in the near future we hope to see different methods which will mainly focus on audio alteration. Algorithms with the ability to counter both types of deep fake content i.e. having the ability to identify and detect both types of manipulation in a single model. It will be like merging two effective techniques of their different domains, the audio altered videos and the faces altered videos.

The suggested approach may consist of three stages: First we will test the videos by using the method which mainly focuses on faces altered in the videos. After going through this classifier then it will be sent to the other classifier which will test its audio alteration. And on the basis of the result from both classifiers we will say that the video is fake or not? This approach may be effective as nowadays videos are both audio and video altered and we need a technique which can test and detect both.

References

[1] Deepfakes faceswap. https://github.com/deepfakes/faceswap

[2] Faceswap. https://github.com/MarekKowalski/FaceSwap/.

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

ISSN: 2005-4262 IJGDC

[4] Googlespeech-to-text.https://cloud.google.com/speech-to-text/docs.

[5] FakeApp. https://www.malavida.com/en/soft/ fakeapp/, Accessed Nov 4, (2019).

[6] S. Agarwal, H. Farid, O. Fried and M. Agrawala, "Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, (2020), pp. 2814-2822, doi: 10.1109/CVPRW50498.2020.00338.

[7] Nicholas Dufour, Andrew Gully, Per Karlsson, Alexey Victor Vorbyov, Thomas Leung, Jeremiah Childs, and Christoph Bregler, “Deepfakes detection dataset” by google & jigsaw.

[8] F. F. Kharbat, T. Elamsy, A. Mahmoud and R. Abdullah, "Image Feature Detectors for Deepfake Video Detection," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, (2019), pp. 1-4, doi: 10.1109/AICCSA47632.2019.9035360.

[9] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, Hong Kong, (2018), pp. 1-7, doi: 10.1109/WIFS.2018.8630761.

[10] F. Matern, C. Riess and M. Stamminger, "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations," 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, (2019), pp. 83-92, doi: 10.1109/WACVW.2019.00020.

[11] Y. Li, M. Chang and S. Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking," 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, Hong Kong, (2018), pp. 1-7, doi: 10.1109/WIFS.2018.8630787.

[12] D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, (2018), pp. 1-6, doi: 10.1109/AVSS.2018.8639163.

[13] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwey Lyu, “Celeb-DF: A New Dataset for Deepfake Forensics”, arXiv preprint arXiv:1909.12962, (2019).

[14] Ohad Fried, Ayush Tewari, Michael Zollh¨ofer, Adam Finkelstein,Eli Shechtman, Dan B Goldman, Kyle Genova, ZeyuJin, Christian Theobalt, and Maneesh Agrawala. “Text-based editing of talking-head video”, ACM Transactions on Graphics, Article No.: 68 (2019).

[15] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman, “Synthesizing Obama: Learning lip sync from audio” ACM Transactions on Graphics, (2017).

[16] Tadas Baltruˇsaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In IEEE Winter Conference on Applications of Computer Vision, pages 1–10, (2016).

[17] Steve Rubin, Floraine Berthouzoz, Gautham J Mysore, Wilmot Li, and Maneesh Agrawala. Content-based tools for editing audio stories. In Proceedings of the 26th annual ACM symposium on User interface software and technology, (2013).

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Proceedings of Advances in Neural Information Processing Systems, pp. 2672–2680, December (2014), Montreal, Canada.

[19] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt and M. Nießner, "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, (2016) pp. 2387-2395, doi: 10.1109/CVPR. (2016).262.

ISSN: 2005-4262 IJGDC

[20] VidTIMIT database (http://conradsanderson.id.au/vidtimit/).

[21] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation.In International Conference on Learning Representations, (2018).

[22] D. P. Kingma and P. Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. arXiv preprint arXiv:1807.03039, (2018).

[23] Guenther N, Schonlau M. Support vector machines. The Stata Journal. (2016) December;16(4):917-37.

[24] Yao S, Pan S, Wang T, Zheng C, Shen W, Chong Y. A new pedestrian detection method based on combined HOG and LSS features. Neurocomputing. (2015) March 3; 151:1006-14.

[25] ORB E. Rublee, "ORB: An efficient alternative to SIFT or SURF," in IEEE International Conference on Computer Vision, Barcelona, ICCV, (2011), pp. 2564-2571.

[26] BRISK S. Leutenegger, "BRISK: Binary robust invariant scalable key points," in IEEE International Conference on Computer Vision,Barcelona, ICCV, (2011), pp. 2548-2555.

[27] KAZE P. F. Alcantarilla, "KAZE features," in European Conference on Computer Vision, Berlin, ECCV, (2012), pp. 214-227.

[28] SURF H. Bay, "Speeded-up robust features (SURF)," Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, (2008).

[29] Fast E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in Computer Vision–ECCV 2006, (2006), pp. 430–443.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al. Going deeper with convolutions. Cvpr, (2015).

[31] Yuezun Li and Siwei Lyu. Exposing deepfake videos by de-tecting face warping artifacts. arXiv: 1811.00656, (2018).

[32] Peng Zhou, Xintong Han, Vlad I. Morariu, and Larry S.Davis. Two-stream neural networks for tampered face detection. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2017).

[33] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes using inconsistent head poses. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 8261–8265, (2019).

[34] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. Deepfakes and beyond: A Survey of face manipulation and fake detection, (2020).

[35] S. Liu and W. Deng, "Very deep convolutional neural network-based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, (2015).

[36] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, (2016).

[37] Ning Yu, Larry Davis, and Mario Fritz "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints" In IEEE International Conference on Computer Vision, (2018).

[38] T. Ali, T. Ahmad and M. Imran, "UOCR: A ligature based approach for an Urdu OCR system," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 388-394.

ISSN: 2005-4262 IJGDC

[39] Tameem Ahmad, Sayyed Usman Ahmed, Syed Omar Ali & Rifa Khan, "Beginning with exploring the way for rumor free social networks", Journal of Statistics and Management Systems, 23:2, 231-238, DOI: 10.1080/09720510.2020.1724623, (2020)

[40] Babak Mahdian and Stanislav Saic "Blind authentication using periodic properties of interpolation", IEEE Transactions on Information Forensics and Security, (2008).

[41] Matthias Kirchner and Thomas Gloe, "On resampling detection in re-compressed images" In Information Forensics and Security, 2009 WIFS 2009 First IEEE International Workshop on. IEEE, (2009).

[42] Matthias Kirchner and Rainer Bohme, "Hiding traces of resampling in digital images", IEEE Transactions on Information Forensics and Security, (2018).

ISSN: 2005-4262 IJGDC