Abbreviated View of Deepfake Videos Detection Techniques

Mohammed A. Younus
Department of Computer Science
College of Science, University of Diyala
Diyala , Iraq
mohammed.akram@sciences.uodiyala.edu.iq

Taha M. Hasan
Department of Computer Science
College of Science, University of Diyala
Diyala , Iraq
dr.tahamh@sciences.uodiyala.edu.iq

***Abstract*— In the era of technological advances and a qualitative breakthrough in the artificial intelligence field and deep neural networks, a new age of hyper-realistic digital videos forgery called DeepFake has been born, with that new technology, it is difficult to distinguish between real videos and fake ones which are uploaded daily on various websites across the Internet. Many open-source DeepFake creation methods have risen, leading to a growing number of synthesized media clips over the internet. There are many efficient fast methods and techniques which have been designed to detect and spot such phenomenon. Background Comparison, Temporal Pattern Analysis, Eye blinking, Facial Artifacts, Mesoscopic Analysis, and Pose Estimation are some of those techniques. Some of these approaches designed to detect and identify the video forgery without any prior enlightenment concerning the videos under analysis. The primary scope of this study is to provide an abbreviate review of these methodologies of a method-comparison study that has been presented to assist the researcher's evaluation of such studies.**

***Index Terms*—Digital videos forgery, Deepfake Detection techniques, Mesoscopic, Facial Artifacts.**

I. Introduction

Evidence such as photographs and videos are considered to be reliable sources and often used in the courtroom and in police investigations to resolve legal cases. The advanced sophisticated technology of video and photo editing techniques made these pieces of evidence uncertain and unreliable. Due to that, this kind of evidence will require to be checked prior to be presenting it in court.

Nowadays, the most widely used tampering technique is called Deepfake. Deepfake allows any computer user to exchange the face of one person with another digitally in any video. This technique can be badly used in many cases such as produce a pornographic video of a celebrity, take revenge, or even in politic.

Advanced Deepfake videos are generated using the Generative Adversarial Network (GAN) which is a type of machine learning system that includes two neural systems, working in concert. One system creates the fake and other tries to distinguish it, with the content bouncing back and forth, and improving with each volley, it's not that easy to challenge it [1].

Deepfake or counterfeit composite videos created through deep learning using (GAN) have come into open talk because of the solid probability of focused dishonest utilizations of their use, lately, different news has been heard about scientists creating tools that can recognize Deepfake with more noteworthy than 90 percent exactness. It's soothing to believe that such researches, the damage brought about by AI-created fakes videos will be restricted. Simply run your video substance through a Deepfake detector and blast the falsehood is marked.

H. Li et. al [2], reveals that the verge of any Deepfake detection method just getting down to business for a brief timeframe. Actually, they said, sooner or later almost certainly, it won't be conceivable to distinguish AI fakes by any means. So, an alternate sort of methodology should be set up to resolve this. Yet, applications that can spot Deepfake AI-controlled videos will just ever give a halfway fix to this issue, say specialists. Likewise, with PC infections with viruses, the risk from Deepfakes is presently a perpetual element on the scene. What's more, despite the fact that it's questionable whether Deepfake is a gigantic threat from a political viewpoint, they're positively harming the lives of many incent people at this very moment through the spread of fake videos of thinks that they didn't do.

Spotting Deepfakes resembles detecting virus infections, a consistently changing challenge, it won't be long until the work is useless. Deepfake technology is developing in someway similar to a virus/antivirus dynamic.

Take the eyes flickering (blinking) method as an example, specialists found that the fact that Deepfake frameworks weren't prepared on the film of individuals with their eyes shut, the videos they delivered included unnatural flickering patterns. Artificial intelligence clones didn't blink frequently enough or, some of the time didn't flicker by any means, characteristics that could be effectively spotted with a basic calculation or a simple algorithm. What occurred next was to some degree unsurprising [4]. Not long after this forensic technique was announced to the public, the upcoming age of synthesis strategies consolidated blinking into their frameworks.

Researchers at the Artificial Inelegancy Foundation concurs that the challenge is far more prominent than straightforward identification, and says that those researches should be placed in deep perspective [5] [6].

Some of the Deepfake detection algorithm guarantees that they flaunted 97 percent accuracy, however assuming this is the case, 3 percent could even now be harming when thinking at the size of web platforms. That implies with each false

978-1-7281-5910-2/20/$31.00 ©2020 IEEE

positive from the model, you are trading off the trust of the clients. At that base, it won't be long, however, before the fakes are undetectable. Given the potential ramifications of this issue, what alternatives are available to us?. Here in this paper, a brief explanation of each method used to cope with this phenomenon, and a comparison is shown in Table I among most of these algorithms.

II. Deepfake Creation Pipeline

The general procedure of the Deepfakes creation is shown in Fig. 1, a Deepfake can be created by feeding the algorithm with frames from media clips that contain the source face to be supplanted. The face is identified and bounded in a box by face identifier function, trailed by the discovery of facial points by facial landmarks. The area of the face is distorted into a standard configuration by means of a relative change M, by limiting the arrangement errors of focal facial points (red dots shown in Fig. 1 (c)) to a lot of standard milestone areas, this procedure is called face arrangement. Then by cropping the image into the size of 64 × 64 pixels, and fed into the profound generative neural system to make matching. The face which has been synthesized is turned back with M−1 to coordinate the source face. At long last, with post-handling, limited smoothing, a Deepfake videos casing are made.

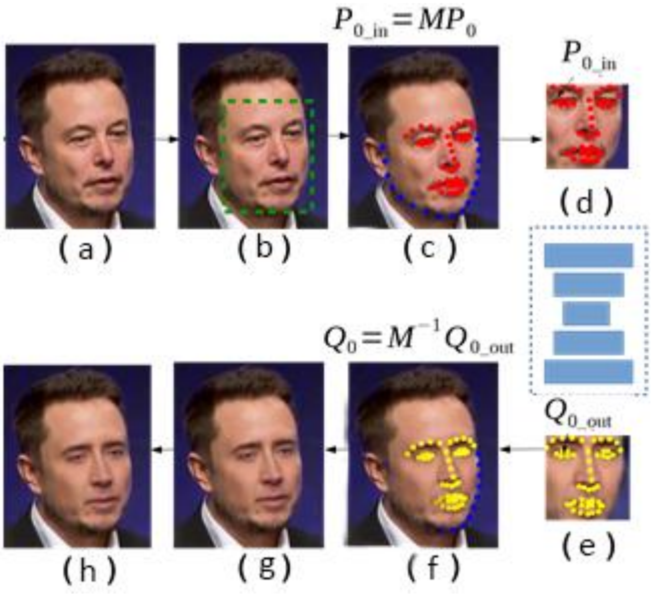

Fig. 1. Overview of Deepfake algorithm. (a) The image of the source face. (b) Face boundary in the image have been detected. (c) Landmarks have been spotted. (d) Using an affine transformation M, the face is cropped then warped to a standardized face. (e) By deep neural network, the Deepfake face is synthesized. (f) Using M−1 Deepfake face is metamorphose back. (g) The face which has been synthesized is fine-tuned based on landmarks in (c). (g) The new face is integrated into the source image. (h) The end result.

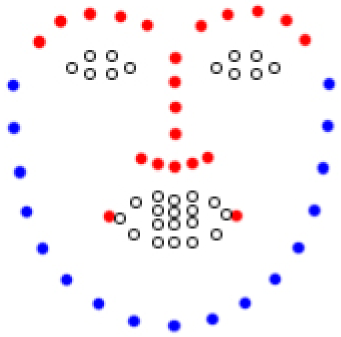

Fig. 2. The landmarks of facial. The central face section uses red dots. The whole face shown by blue and red landmarks. The head pose evaluation gets the use of the empty landmarks.

The transformation from (d) to (e) is done with the GAN, the most important knowledge that characterizes GAN is to arrange this demonstrating issue as a sort of challenge. That is the name “Adversarial" where originates from. The key point is to build not one, but two contending networks: the first on called the generator and the second one called discriminator. The first one which is the generator segment aims to make random synthetic outputs (images of faces), at the same time the second one which is the discriminator part aims to tell the difference from real. The expectation is that as the two systems go head to head, they'll both show signs of improvement with the final product being a generator organize that produces reasonable realistic outputs [1]. To summarize: GANs is a neural network system that learns to choose samples from a significant distribution (the "generative" part), and that is done by setting up a challenge (the "adversarial" part).

III. Deepfake Detection Techniques:

Given the potential information and implications of this problem, an abbreviation is given for the options and latest techniques available up-to-date. See Table I for comparison.

A. Background Comparison

Background Comparison method is one of the very primitive ways to detect Deepfake videos [7], established scalable AI-based counter-Deepfake efforts have been shown by the online GIF repository website “Gyfcat”. The website utilizes facial acknowledgment models to spot irregularities in the rendering of the facial zone of an uploaded video. Suspected phony videos are additionally investigated by masking the facial territory looking at the database for a comparable video with a similar background and body. Similar references are then reviewed for facial likenesses to close on the credibility of the video. In any case, this methodology has several weaknesses: backgrounds could be completely faked, and composites produced using totally new footage would not be distinguished, not to forget the requirement for enormous databases of video footage besides time and recourse consuming.

B. Temporal pattern analysis

As a person's behavior can be best characterized as a period arrangement of movements, sequence-based information has been investigated as a validation approach. Specialists have consolidated a convolutional neural network (CNN) together with a Long Short-Term Memory (LSTM) to process temporal successions. Bypassing each frame of a video into a CNN and creating a succession of highlight maps for the LSTM, the network has the option to learn explicit development based practices of its subjects.

Having knowledge about human activities in videos motions is a very complex task since videos are regularly framed by a succession of persistent activities which might be additionally broken down into sub-activities. The struggles in this matter are capturing scenes dependencies between sub-activities and learning the structure of activities.

To begin with, the successiveness of activities and localize temporal bounds have to be distinguished, at that point activities can be classified in spite of everything for their irregular length. Activities that happened at the same time have to be captured to manufacture an exhaustive and reasonable comprehension of the video. Associated activities in a single scene might be hindered by other independent actions due to cross-cutting, which is a regularly utilized altering strategy that the camera removes starting with one activity then onto the next activity. This will give rise to the neglect of learned data so the single scene reliance is lost. The different deficient appearances of a similar activity likewise make it difficult to distinguish and get the structure learned.

Individual different assortments of (CNNs) have been suggested for video activity recognition. However, existing models are for the most part conceived for cropped videos which can just examine short fragments and capture per-segment motion details; consecutive learning and logics are not appropriate. This is not normal for human intelligence when a human being sees an action, he recollects it and become aware of the following activity dependent on the specific situation. This focuses on the way that videos have not to be handled independently or fragmented. For composite videos having numerous activities, it should consider the complicated connection among these activities. Models of recurrent convolutional have been demonstrated to be powerful for visual sequences, however, the present exactness of best in class models tried on complicated video datasets is still less than that on single scene datasets. This may be brought about by the absence of a comprehensive awareness of the temporal reasoning and structures of the activity [8]. Recurrent and convolutional functioning tasks handled at a close neighborhood will most likely be unqualified to figure out the long-extend temporal structure. It needs another design to learn the interaction of temporal in complex videos that have correlated actions.

Moreover, serious difficulties have experienced by recurrent networks in learning long sequences of video scenes due to the fading and gradients of noise. A type of recurrent neural networks (RNNs) known as (LSTM) is an outstanding unique network to address this issue [8]. Additionally, truncated backpropagation through time can help to take care of this issue, where at a fixed number of time steps, gradient update is carried out occasionally. Truncated backpropagation tear a sequence x into disjoint sub-sequences of length k, which in this case may not be useful for learning the global details of the video’s content. Neuroscientists and psychologists have found that "interleaving" strategy, which means contemplating related abilities or ideas in parallel, is a successful method to prepare human minds [9]. Motivated by this interleaving impact, interleaved back diffusion through time is offered, where interleaving connections are made to reduce the propagation length so as to escalate the learning of recurrent neural networks.

C. Facial Artifacts

Recently, specialists from UC Berkeley collaborated with Adobe to produce a tool fit for distinguishing manually manipulated images [10] by recognizing low-level facial warping. Via preparing a CNN on instances of images controlled utilizing the well-known Face-Aware liquefy feature, an approval exactness of 99% could be accomplished versus that of a human eyewitness (53%). However, while such outcomes are encouraging, the absence of GAN-produced models normally implies that such a network is only capable of recognizing manually-manipulated images, and are thus ineffectively appropriate for combating Deepfakes.

Methodical facial warping in Deepfakes was recognized as a deliberate example in Deepfakes by specialists at the University of Albany [12]. Their methodology depended on the propensity of Deepfake algorithms to make the lower resolution of fixed sizes for computational proficiency due to the constraint of computation resource and process time, so the algorithm of Deepfake has the ability to synthesize face images of a fixed size only. These images would then experience upscaling and affine transformation, for example, scaling and pivot to the poses of the objective faces coordinate that they will be displaced in the synthetic composite. Depending on that fact, it should experience an affine warping to coordinate the source's face configuration [11]. This warping makes significant artifacts because of the resolution inconsistency between warped face context and surrounding territory. So, recognizing Deepfake videos can make use of these artifacts. The strategy distinguishes such relics by looking at the created face area (ROI) and adjacent territory locales with a committed (CNN) model. Preparing and training the CNN model, by streamlining the procedure by reproducing the inconsistency of the resolution in affine face warping squarely. First, distinguish the faces and landmarks and it should be extracted afterward to the transform matrices to adjust the standard configuration to the faces. Adjusting the face by applying Gaussian obscuring, then the original image is back from the affine warped face using the reverse of the assessed change grid (estimated transformation matrix which has been inverted). So as to gain reproduce increasingly extraordinary resolution instances of the affine warped face, position faces into various scales to build the data diversity Fig. 3 and Fig. 4.

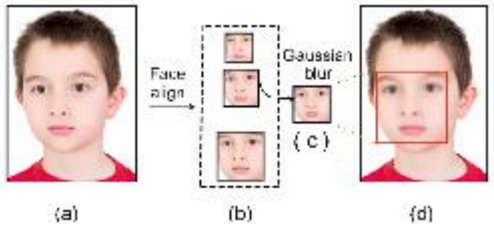

Fig. 3. Generation of negative data. (a) Genuine image. (b) Different scales of aligned face, (c) Arbitrarily picked scale of face and applied Gaussian blur to it. (d) Affine warped to source image.

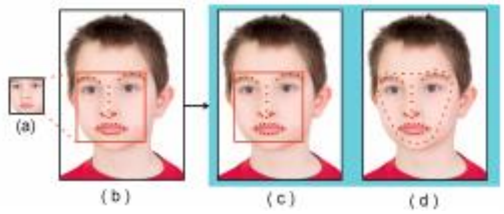

Fig. 4. Negative examples of face shape. (a) The blurred face aligned. (b) Affine warped back. (c, d) The shape of the face area refined.

D. Mesoscopic Analysis

Specialists from the National Institute of Informatics in Tokyo have shown how profound neural networks built with an artificial low number of layers are fit for recognizing fine discrepancies that are seen in Deepfakes at high computational effectiveness, accomplishing an exactness of generally 90%. The presentation of straightforward networks depends on the way that a low number of layers and a surprisingly low number of parameters encourage the identification of smaller, increasingly rudimentary examples with the diminished number of convolutions layers of the image during a forward pass.

This technique utilizes the Mesoscopic level [26] of investigation to recognize manipulated faces in videos. Surely, microscopic analysis examinations based on the noise of the image can't be applied in a compacted video status where the noise of the image is firmly degraded. Likewise, at a higher level of the semantic, the human eye combat to recognize rigged images [12], particularly when the human face is delineated in an image [13][14]. The reason is this strategy proposes to receive the middle of the road approach utilizing a profound neural network with fewer number of layers. The best classification scores among all tests have accomplished are the following two models, with a low level of representation and a significantly low number of parameters. They depend on best-performing networks for image classification [15][16] that substitute convolution's layers and pooling for highlight extraction and a network for classification which is much denser.

The prime idea of this module is the output of a few convolutional layers is stacked with various kernel shapes, therefore the capacity space increases in which the model is enhanced. It proposes to utilize [3×3] enlarged convolutions rather than the [5×5] of original module convolutions [17] so as to avert high semantic. The concept of utilizing dilated convolutions with the initiation module can be established in [18] as an intend to manage multi-scale data, so far by including convolutions of [1×1] before enlarged convolutions for measurement decrease and an extra convolution of [1×1] in parallel that acts as skip-association among progressive modules Fig. 5 shows further details. Supplanting multiple layers with commencement modules didn't offer better outcomes for classification.

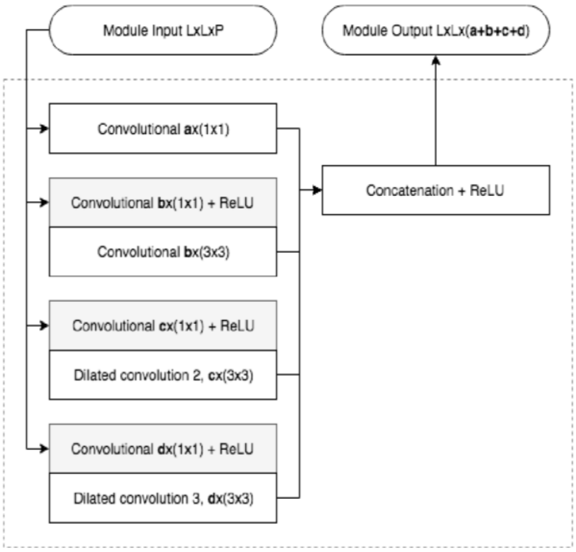
Module Input LxLxP
Module Output LxLx(a+b+c+d)
Convolutional ax(1x1)
Convolutional bx(1x1) + ReLU
Convolutional bx(3x3)
Concatenation + ReLU
Convolutional cx(1x1) + ReLU
Dilated convolution 2, cx(3x3)
Convolutional dx(1x1) + ReLU
Dilated convolution 3, dx(3x3)

Fig. 5. Inception modules used in MesoInception-4 architecture. The parameters used in the module is *a,b,c,d* ∈ N.

E. Head Pose Estimation

Specialists at the University of Albany [19] have shown that systematic differences between simple Deepfake outputs and the objective face pose could be distinguished utilizing landmark focuses and classified utilizing basic help the vector machine models. However, the presentation of the model was seen as poor with blurrier images, as landmark assignment would also be degraded.

The strategy used relies upon the Deepfakes are made by splicing synthesized human face region into the source image observations, performing as such, presenting blunders that can be uncovered when 3-dimension head poses are evaluated from the human face images. They implement tests to exhibit this marvel and further build up a classification technique dependent on this sign. Utilizing highlights dependent on this prompt, a support vector machine (SVM) classifier is assessed utilizing a lot of Deepfakes and genuine images.

Further trained SVM classifiers dependent on the variations between head poses evaluated utilizing the entire set authority landmarks and those in the focal face districts to separate Deepfakes from genuine videos or images. The features are extricated in the following methods:

1) A face detector is run and concentrates 68 facial landmarks for every video frame or image, programming bundle DLib [20] is used for that purpose.

2) After that, the equivalent 68 points from OpenFace2 with the standard 3D facial landmark model [21], the entire face is evaluated separately and the head poses from focal face locale. They inexact the camera central length here as the width of the image, camera focus as image focus, and the effect of lens distortion is disregarded.

3) The acquired rotation matrices and interpretation vectors differences are compacted into a vector, which is institutionalized by subtracting its mean and detached by its standard deviation for classification.

F. Eye blinking

This technique reveals Deepfake videos by detecting the eye blinking paucity in fake faces. Eye blinking [4] refers to the momentary of the opening and closing of the eyelid. The blink is controlled by the stem of the brain and happens without conscious. The human being blinks between intervals of 2 to10 seconds, the normal blink lasts for 0.1 to 0.4 seconds for each blink. If the video scene length average of 1/30 second, it means there is a probability of capturing a photo with someone blinking is about 7.5%. Most published photos on the internet do not show a person with their eyes closed, therefore, the sign of a Deepfake video is the lack of eyelid blinking.

This method uses long-term recurrent CNN (LRCN) [22] which is a combination of RNN and CNN, to recognize close and open eye states with the previous temporal knowledge consideration. Sukno et al. [23] invested the active shape models with invariant optimal features to determine the eyes outline and computed the eye vertical distance to figure the state of the eyes. Torricelli et al. [24] uses the consecutive frames difference to analyze eyes state. Many other algorithms are employed for blinking analysis. So far, there are no known Deep Neural Network algorithms for eye blinking detection.

Briefly, this method works on CNN-based classifier extended to LRCN [22], where the temporal relationship is incorporated between consecutive frames, as the period from opening to closed of eyelid blinking is a temporal process, where LRCN can memorize the long term dynamics to mend the artifact's effects introduced from a single image. Fig. 6 shows a general overview of the algorithm.

The blinking of eye method is a comparatively easy cue in detecting synthesized faces in Deepfake videos, sophisticated counterfeiters can have the ability to create realistic effects of blinking by using advanced post-processing models trained with extra more data.

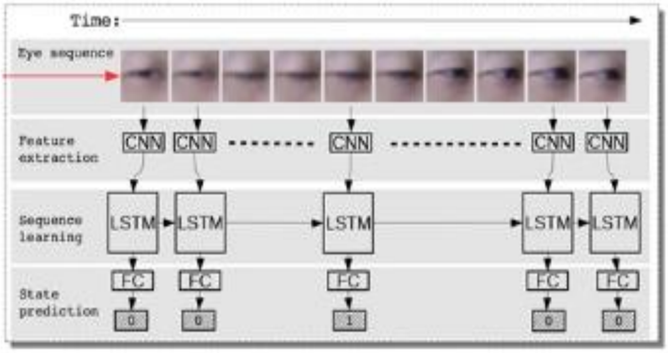

Fig. 6. LRCN method overview.

TABLE I: COMPARISON OF PROMINENT DEEPFAKE DETECTION METHODS

Methods	Reference	Classifiers/ Techniques	Features	Data Sets Used
Eye blinking	[4]	LRCN	-Use LRCN to learn the temporal patterns of eye blinking. -Based on the observation that the blinking frequency of Deepfakes is much smaller than normal.	Consist of 49 interview and presentation videos and their corresponding generated Deepfakes.
Background Comparison	[6]	RCN	Temporal discrepancies across frames are explored using RCN that integrates convolutional network DenseNet and the gated recurrent unit cells	FaceForensics++ data set, including 1,000 videos.
Temporal pattern analysis	[25]	CNN and LSTM	CNN is employed to extract frame-level features, which are distributed to LSTM to construct sequence descriptors useful for classification.	A collection of 600 videos obtained from multiple websites.
Facial Artifacts	[11]	VGG16 ResNet50, 101 or 152	Artifacts are discovered using CNN models based on resolution inconsistency between the warped face area and the surrounding context.	-UADFV, containing 49 real videos and 49 fake videos with 32752 frames in total. -DeepfakeTIMIT
Mesoscopic Analysis	[26]	CNN	-Two deep networks, i.e. Meso-4 and MesoInception-4 are introduced to examine Deepfake videos at the mesoscopic analysis level.	Two data sets: Deepfake one constituted from online videos and the FaceForensics one created by the Face2Face approach.
Head Pose Estimation	[19]	SVM	-Features are extracted using 68 landmarks of the face region. -Use SVM to classify using the extracted features.	-UADFV consists of 49 deep fake videos and their respective real videos. 241 real images and 252 deep fake images from DARPA MediFor GAN Image/Video Challenge.

.

IV. CONCLUSIONS

While it would appear as though we're in another data cold war with numerous entertainers and no treaties to direct stockpiles, the ascent of Deepfakes isn't an entirely morbid phenomenon. A blend of Deepfakes and advanced NLP (Neuro-Linguistic Programming) models, for example, OpenAI's GPT-2 (Generative Pretrained Transformer 2) could open the best approach to making increasingly human-like digital assistants, with applications in the administration and medicinal services ventures industry.

In the long term, we could possibly build up the ability to transfer a mimicry of our awareness into the cloud, making intelligent symbols long after we've passed on, without the need of genuine on-screen characters. It is both alarming and energizing that the foundations of such technologies are inside our grasp. Early executions of such thoughts have just been shown by Samsung's AI group, through "living" representations of well-known people.

As a general public, we remain at another intersection for truth. At last, the answer for Deepfakes may lay on societal adaptation, regardless of whether we can build up an improved ability to basically examine, process and assess data during a time of information overload, decentralized sources, and diminished capacities attention spans.

Detection techniques are still in their beginning stage and different strategies have been proposed and assessed yet utilizing fragmented data sets. A way to deal with improve the presentation of detection techniques is to make a developing refreshed benchmark data set of Deepfakes to approve the continuous advancement of detection strategies. This will encourage the preparation procedure of detection models, particularly those dependent on deep learning, which requires an enormous training set. Then again, current detection techniques mostly focus on drawbacks of the Deepfake generation pipelines, for example, finding weaknesses of the contenders to assault them. This sort of data and learning isn't constantly accessible in adversarial environments where aggressors regularly endeavor not to uncover such Deepfake creation advances. This is a real challenge for detection technique improvement and future research needs to concentrate on introducing more robust, scalable and generalizable strategies.

REFERENCES

[1] A. Radford, L. Metz, and S.J.a.p.a. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks. 2015.

[2] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 38-45

[3] B. Guo, Y. Ding, Y. Sun, S. Ma, and K. J. a. p. a. Li, "The Mass, Fake News, and Cognition Security," 2019.

[4] Li, Y., M.-C. Chang, and S.J.a.p.a. Lyu, In Ictu Oculi: Exposing ai generated fake face videos by detecting eye blinking. 2018.

[5] M. H. Khudhur, J. Waleed, H. Hatem, A. M. Abduldaim and D. A. Abdullah, "An Efficient and Fast Digital Image Copy-Move Forensic Technique," 2018 2nd International Conference for Engineering, Technology and Sciences of Al-Kitab (ICETS), Kirkuk, Iraq, 2018, pp. 78-82.

[6] J. Waleed, D. A. Abdullah and M. H. Khudhur, "Comprehensive Display of Digital Image Copy-Move Forensics Techniques," 2018 International Conference on Engineering Technology and their Applications (IICETA), Al-Najaf, 2018, pp. 155-160.

[7] M. Koopman, , A.M. Rodriguez, and Z. Geradts. "Detection of Deepfake Video Manipulation." in Conference: IMVIP. 2018.

[8] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in Fifteenth annual conference of the international speech communication association, 2014.

[9] D.J.E.P.R Rohrer,., Interleaving helps students distinguish among similar concepts. 2012. 24(3): p. 355-367.

[10] A. Sutardja, and Y. Zhao, Digital Image Manipulation Forensics. 2015.

[11] Li, Y. and S.J.a.p.a. Lyu, exposing deepfake videos by detecting face warping artifacts. 2018. 2.

[12] V. Schetinger, M. M. Oliveira, R. da Silva, T. J. J. C. Carvalho, and Graphics, "Humans are easily fooled by digital images," vol. 68, pp. 142-151, 2017.

[13] Balas, B. and C.J.P. Tonsager, Face animacy is not all in the eyes: Evidence from contrast chimeras. 2014. 43(5): p. 355-367.

[14] S. Fan, R. Wang, T.-T. Ng, C. Y.-C. Tan, J. S. Herberg, and B. L. J. A. T. o. A. P. Koenig, "Human perception of visual realism for photo and computer-generated face images," vol. 11, no. 2, pp. 1-21, 2014.

[15] A.Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in Advances in neural information processing systems. 2012.

[16] K.Simonyan, and A.J.a.p.a. Zisserman, Very deep convolutional networks for large-scale image recognition. 2014.

[17] F. Yu, and V.J.a.p.a. Koltun, Multi-scale context aggregation by dilated convolutions. 2015.

[18] W. Shi, F. Jiang, and D. Zhao. Single image super-resolution with dilated convolution based multi-scale information learning inception module. 2017 IEEE International Conference on Image Processing (ICIP). 2017. IEEE.

[19] X. Yang, Y. Li, and S. Lyu. "Exposing deep fakes using inconsistent head poses." in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019. IEEE.

[20] D.E.J.J King,., M.L.R., Dlib-ml: A machine learning toolkit. 2009. 10(Jul): p. 1755-1758.

[21] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 59-66: IEEE.

[22] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In CVPR, pages 2625–2634, 2015.

[23] F. M. Sukno, S.-K. Pavani, C. Butakoff, and A. F. Frangi. Automatic assessment of eye blinking patterns through statistical shape models. In International Conference on Computer Vision Systems, pages 33–42, 2009.

[24] D. Torricelli, M. Goffredo, S. Conforto, and M. Schmid. An adaptive blink detector to initialize and update a view based remote eye gaze tracking system in a natural scenario. Pattern Recognition Letters, 30(12):1144–1150, 2009.

[25] D. Guera, and E. J. Delp, (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

[26] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, (2018, December). MesoNet: a compact facial video forgery detection network. In 2018 IEEE Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE.