International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 03 | Mar 2024 www.irjet.net p-ISSN: 2395-0072

Deepfake Detection: A Literature Review

Sayed Shifa Mohd Imran[1], Dr. Pallavi Devendra Tawde[2]

[1]Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India
[2]Assistant Professor, Nagindas Khandwala College, Mumbai, Maharashtra, India

---

Abstract- *Deepfake controls can be utilized to make persuading pantomimes of people, possibly driving to personality robbery or unapproved get to touchy information. As fake recordings and sound are utilized in different shapes of cyberattacks, counting stick phishing and social designing, having vigorous location components in put can avoid unauthorized get to to delicate data. With the restrictions famous over, deep fake discovery can be utilized to distinguish substance that has been controlled for noxious purposes. Identifying and labeling fake recordings and pictures permit people and associations to take activity to halt the spread of possibly harming deception. This can protect the notoriety and security of people and avoid the dispersal of fake news, fakes, or cyberbullying. Luckily, devices to distinguish deep fakes are moreover making strides. Deepfake location requires collaboration between specialists from different areas, such as computer science, counterfeit insights, brain research, and law. Deepfake detectors can look for obvious biometric signs inside a video, such as a person's pulse or a voice produced by human vocal organs or maybe than a synthesizer. Amusingly, the apparatuses utilized to prepare and move forward these locators nowadays might in the long run be utilized to prepare the following era of deep fakes as well. In conclusion, deep fake discovery is a quickly advancing field with noteworthy suggestions for society. Proceeded collaboration between analysts, policymakers, and the open is basic to create viable location methods, address lawful and moral concerns, and advance open mindfulness to moderate the potential hurts of deep fakes.*

Key Words: *Deepfake, Image, Audio, Video, Detection*

1. INTRODUCTION

Confront discovery, too called facial location, is a counterfeit insights (AI)-based computer innovation utilized to discover and recognize human faces in advanced pictures and video. Confront location innovation is frequently utilized for observation and following of individuals in genuine time. It is utilized in different areas counting security, biometrics, law authorization, excitement and social media. Face discovery employments machine learning (ML) and counterfeit neural arrange (ANN) innovation, and plays an imperative part in confront following, confront investigation and facial acknowledgment. In confront investigation; confront discovery employments facial expressions to distinguish which parts of a picture or video ought to be centered on to decide age, gender and feelings. In a facial acknowledgment framework, confront location information is required to create a face print and coordinate it with other put away face prints.

As a key component in facial imaging applications, such as facial acknowledgment and confront examination, confront discovery makes different focal points for clients, counting the following:

• Improved security. Confront location makes strides observation endeavors and makes a difference track down hoodlums and fear mongers. Individual security is improved when clients utilize their faces input of passwords, since there's nothing for programmers to take or change.

• Easy to coordinated. Confront discovery and facial acknowledgment innovation is simple to coordinated, and most applications are congruous with the larger part of cybersecurity software.

• Automated distinguishing proof. In the past, recognizable proof was physically performed by an individual; this was wasteful and as often as possible wrong. Confront location permits the recognizable proof prepare to be computerized, sparing time and expanding accuracy.

History of confront detection the to begin with computerized confront discovery tests were propelled in 1964 by American mathematician Woodrow W. Bledsoe. His group at All encompassing Inquire about in Palo Alto, Calif., utilized a simple scanner to filter people's faces and discover matches in an endeavor to program computers to recognize faces. The try was generally unsuccessful since of the computer's trouble with posture, lighting and facial expressions. Major enhancements to confront discovery technique came in 2001, when computer vision analysts at the Mitsubishi Electric Investigate Research facilities Paul Viola and Michael Jones proposed a system to identify faces in genuine time with tall exactness.

The Viola-Jones system is based on preparing a demonstrate to get it what is and is not a confront. Once prepared, the show extricates particular highlights, which are put away in a record so that highlights from unused pictures can be compared with the put away highlights at different stages. If the picture beneath think about passes through each organize of the highlight comparison, at that point a confront has been identified and operations can proceed. The Viola-Jones system is still utilized to recognize faces in real-time applications, but it has restrictions. For case, the system might not work if a confront is secured with a cover or scarf, or if the confront isn't legitimately situated, the calculation might not be able to discover it. Later a long time have brought progresses in confront discovery utilizing profound learning, which beats conventional computer vision methods.

Deep fakes employments two calculations -- a generator and a discriminator -- to make and refine fake substance. The generator builds a preparing information set based on the craved yield, making the introductory fake computerized substance, whereas the discriminator analyzes how practical or fake the starting adaptation of the substance is. This handle is rehashed, permitting the generator to make strides at making practical substance and the discriminator to gotten to be more talented at spotting imperfections for the generator to correct.

The taking after are a few particular approaches to making deep fakes:

• Source video deep fakes. When working from a source video, a neural network-based deep fake auto encoder analyzes the substance to get it important qualities of the target, such as facial expressions and body dialect. It at that point forces these characteristics onto the unique video. This auto encoder incorporates an encoder, which encodes the important qualities; and a decoder, which forces these traits onto the target video.

• Audio deep fakes. For sound deep fakes, a GAN clones the sound of a person's voice, makes a demonstrate based on the vocal designs and employments that demonstrate to make the voice say anything the maker needs. This method is commonly utilized by video amusement developers.

• Lip syncing. Lip syncing is another common strategy utilized in deep fakes. Here, the deep fake maps a voice recording to the video, making it show up as in spite of the fact that the individual in the video is talking the words in the recording. If the sound itself is a deep fake, at that point the video includes an additional layer of deception.

There are a few eminent illustrations of deep fakes, counting the following:

• Facebook author Check Zuckerberg was the casualty of a deep fake that appeared him gloating approximately how Facebook "claims" its clients. The video was planned to appear how individuals can utilize social media stages such as Facebook to betray the public.

• U.S. President Joe Biden was the casualty of various deep fakes in 2020 appearing him in overstated states of cognitive decrease implied to impact the presidential decision. Presidents Barack Obama and Donald Trump have too been casualties of deep fake recordings, a few to spread disinformation and a few as parody and entertainment.

• During the Russian intrusion of Ukraine in 2022, Ukrainian President Volodomyr Zelenskyy was depicted telling his troops to yield to the Russians.

2. REVIEW OF LITERATURE

Andreas et al [1] this paper examines the realism of state-of the-art image manipulations, and how difficult it is to detect them, either automatically or by humans. After the collecting data it is manipulated, then the image is detected whether it is fake or real using CNNs convutional neural networks.

Yuezun Li et al [2] The need to develop and evaluate Deep Fake detection algorithms calls for large-scale datasets. However, current Deep Fake datasets suffer from low visual quality and do not resemble Deep Fake videos circulated on the Internet. The use of DNNs has made the process to create convincing fake videos increasingly easier and faster. In this work, they present a new large-scale and challenging Deep Fake video dataset, Celeb-DF3, for the development and evaluation of Deep Fake detection algorithms.

Brian et al [3] The DFDC is the largest currently and publicly available face swap video dataset. The dataset contains over 100,000 clips from 3,426+ paid actors. The dataset is created using several Deep fakes and GAN-based and non-learning techniques.

Ricard et al [4] By analyzing a low resolution video sequence of FaceForensics++ dataset, our method detects manipulated videos with 90% accuracy. They solve the issue of detecting artiflcial image content, more specifically, fake faces. To identify the nature of these images, we introduce a novel machine learning based approach. The approach is based on a classic frequency analysis of images that detects various behaviors at high frequencies.

Ruben et al [5] This survey offers a comprehensive overview of methods to detect and manipulate face images, including Deep Fake techniques. Specifically, four categories of face manipulation are examined: i) the full face; ii) switching identities; iii) modifying characteristics; iv) switching expressions.

Nicol'o et al [6] Take up the challenge of detecting face alteration in video sequences that use contemporary facial manipulation methods. Using more than 10,000 videos, the CNN approach is used to recognize false videos.

Wanying Ge et al [7] The application of SHapley Additive exPlanations (SHAP) to obtain novel insights into spying detection is presented in this paper. A visualization tool called SHapley Additive exPlanations is used to visualize the output of a machine learning model in order to make it easier to understand. By calculating the contribution of each feature to the prediction, it can be used to explain the prediction of any model.

Chunlei Peng et al [8] By assigning distinct scores to both genuine and false face data, you can enhance the model's capacity to recognize complicated samples with greater detail. The idea of perceptual forgery fidelity should be taken into consideration given the complexity of face quality distribution of data in the real world. We replace the prior binary classification with the forgery fidelity score by mapping facial data of various attributes to discrete values.

Tianchen et al [9] Predicated on the idea that unique source features in photos can be retained and recovered following the application of cutting-edge deep fake generating techniques. Different source features at different locations can be found in the fabricated image. We can identify counterfeit photos by extracting the local source features and calculating their self-consistency.

Bojia et al [10] In this research, we offer a new dataset called WildDeepfake, which comprises of 7,314 face sequences derived from 707 deep fake videos acquired entirely from the internet, to better enhance detection against real-world deep fakes. Two Attention-based Deepfake Detection Networks (ADDNets) were presented by the researcher.

Kaede et al [11] In order to identify deep fakes, we introduce in this paper new synthetic training data dubbed self-blended images (SBIs). To replicate forging artifacts, SBIs are created by merging source and target photos that have been marginally altered from one authentic image.

Shichao et al [12] The purpose of this study is to interpret how artifact attributes of photos are learned by deep fake detection algorithms under the simple supervision of binary labels. To improve the effectiveness of forgery detection on compressed movies, use the FST-Matching Deepfake Detection Model. According to the results, this strategy performs well.

Anubhav Jain et al [13] Deepfake detection technology that avoids the requirement for any real data by utilizing synthetically created data via StyleGAN3. The final trained model demonstrates reduced bias and more interpretable characteristics.

Fatima et al [14] We trained a dataset of 9,000 images over 150 epochs and found that the ResNet50 model was the best model of network architectures utilized, with 100% training accuracy, 99.18% validation accuracy, training loss 0.0003, validation loss 0.0265, and testing accuracy of 99%.

Tiewen et al [15] Creates a range of forged faces from a masked clean one, enabling the deep fake detection model to learn general and robust representations rather than overfitting to particular artifacts. The deep fake detection model is trained using a variety of manipulated faces generated from a single clean face, allowing it to learn universal and resilient features instead of focusing on specific distortions.

Narayan et al [16] Dual shot face detector extracts faces from several photos and videos, while MesoNet, FWA, XceptionNet, and Capsule techniques are used to identify deepfakes.Detect both low and high resolution photos.Training requires a more powerful computational engine.

Khan et al [17] Introduce the Mel-frequency cepstral coefficient. We initially examine the datasets using Bispectral analysis and NTU techniques, similar to a machine learning cycle. Trained several RNN models with features, such as MFCCs, RMS,

zero crossing, chroma frequency, and spectral roll-off. Calculation should be as simple as possible. Limitations include audio segments that are just 2 seconds long.

Almutairi et al [18] With around 97% accuracy, the author uses a variety of machine learning techniques, such as SVM and KNN, to distinguish between real and false audio. CNN, another deep learning technique, has a 99% accuracy rate and just 2% misclassification rate. With high accuracy, the author employed a variety of CNN models, such as ResNet34, LSTM, and RNN. These methods can be used to identify an audio clip that has been mimicked. There aren't many non-English audio Deepfake detection techniques.

Ilyas et al [19] Make use of a revolutionary AVFakeNet framework to identify fraudulent audio and video modalities in a unified manner. Additionally, a unique Dense Swin Transformer Net for feature extraction was designed. Find multimodality using an innovative method, increased computing power is needed.

Table 1: LITERATURE REVIEW SUMMARY

TITLE	YEAR	TECHNIQUE USED	DESCRIPTION
FaceForensics++: Learning to Detect Manipulated Facial Images [1]	2019	Convolutional neural networks (CNNs). Dataset: 1000 videos.	It helps to detect fake images from videos by trained forgery detectors. Acquiring the skill of identifying altered facial photographs. Mastering the ability to recognize edited facial images.
Celeb-DF:A Large-scale Challenging Dataset for Deep Fake Forensics [2]	2020	Deep Neural Networks (DNNs) Dataset: 5639 high quality videos.	Celeb-DF is a comprehensive dataset designed to challenge deep fake forensic techniques on a large scale.
The Deep Fake Detection Challenge (DFDC) Dataset [3]	2020	GAN MTCNN Dataset: 100000 clips from 3426 paid actors.	A deep fake detection model trained solely on the DFDC dataset has the ability to generalize to authentic "in-the-wild" deep fake videos, making it a valuable tool for analyzing potentially manipulated videos.
Unmasking DeepFakes with simple Features [4]	2019	GAN Two datasets	Our novel machine learning approach accurately detects manipulated videos with a 90% success rate by analyzing a low resolution video sequence from the FaceForensics++ dataset. This method specifically focuses on identifying fake faces and artificial image content.
Video Face Manipulation Detection Through Ensemble of CNNs [6]	2020	CNN	The solution presented involves deriving various models from a foundational network (EfficientNetB4) by incorporating attention layers and siamese training. By merging these networks, we demonstrate significant advancements in detecting face manipulation across two extensive datasets comprising over 119,000 videos.
Deep Fidelity: Perceptual Forgery Fidelity Assessment for Deepfake Detection [8]	2023	SSAAFormer	The proposed DeepFidelity framework aims to dynamically identify real and fake faces by examining the fidelity of facial images, taking into account the diverse quality levels present in both categories. By leveraging advanced techniques, this system offers a comprehensive solution for detecting Deep fakes with varying image quality, enhancing the accuracy and reliability of face verification processes.
Learning Self-Consistency for Deepfake Detection [9]	2021	Pair-wise self-consistency learning (PCL)	The new Deepfake detection system is effectively differentiate between authentic and manipulated faces based on their varying image quality. This framework is

			designed to address the intricate nature of quality distribution in both real and fake faces by analyzing the perceptual forgery fidelity of facial images.
Wild Deepfake: A Challenging Real-World Dataset for Deepfake Detection [10]	2021	Attention- based Deepfake Detection Networks (ADDNets)	An intricate dataset designed to test the capabilities of deep fake detection algorithms in real-world scenarios has been developed. This challenging dataset aims to push the boundaries of current technology by presenting difficult scenarios for detection models to accurately identify manipulated media. The dataset includes a diverse range of deep fake videos that have been meticulously crafted to closely resemble authentic footage, making it a formidable test for even the most advanced detection systems.
Detecting Deep fakes with Self-Blended Images [11]	2022	Self-blended images (SBIs)	Self-Blended Images involve the use of advanced algorithms and machine learning techniques to analyze and compare various facial features within a video. By examining the subtle differences in facial expressions, skin texture, and lighting conditions, this method can identify any inconsistencies or anomalies that may indicate the presence of a deep fake.
Explaining Deepfake Detection by Analysing Image Matching [12]	2022	FST- Matching, DNNs	The process of identifying deep fake content involves a thorough analysis of image matching techniques. By examining the similarities and differences between the original and manipulated images, researchers can develop algorithms that can detect the subtle alterations made in deep fake videos. This analysis typically involves comparing pixel values, color gradients, and other visual features to determine if an image has been digitally altered.
Masked Conditional Diffusion Model for Enhancing Deepfake Detection [15]	2024	Masked Conditional Diffusion Model (MCDM) Data Augmentation	The MCDM is a cutting-edge technique that has been developed to enhance the detection of deep fake videos. Deepfake videos are manipulated videos that use artificial intelligence to replace the face of a person in an existing video with someone else's face. These videos can be incredibly convincing and pose a significant threat to the authenticity of visual media.

3. CONCLUSION

Despite the impressive performance of deep learning in detecting deep fakes, the quality of deep fake content continues to rise, necessitating enhancements in current deep learning techniques for accurate identification of fake videos and images. Moreover, there is a lack of clarity on the optimal number of layers and suitable architecture for deep fake detection within existing deep learning methods. An additional area of exploration involves integrating deep fake detection methods into social media platforms to enhance their ability to combat the widespread influence of deep fakes and minimize their consequences.

Furthermore, there is a need to explore integrating deep fake detection methods into social media platforms to enhance their ability to combat the widespread influence of deep fakes and minimize their negative consequences. This highlights the importance of continuous research and development in the field of deep fake detection to stay ahead of evolving deep fake technologies and protect against potential misuse.

REFERENCES

[1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner.

(2019). FaceForensics++: Learning to Detect Manipulated Facial Images. *IEEE Conference Publication*

[2] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu. (2020). Celeb-DF:A Large-scale Challenging Dataset for Deep Fake Forensics. *IEEE Conference Publication*

[3] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, Cristian Canton Ferrer. (2020) The Deep Fake Detection Challenge (DFDC) Dataset. *arXiv:2006.07397 Vol4*

[4] Ricard Durall, Margret Keupe, Franz-Josef Pfreundt, Janis Keuper. (2019) Unmasking DeepFakes with simple Features. *arXiv.org*

[5] Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. *ArXiv:2001.00179*.

[6] Nicol'o Bonettini, Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, Stefano Tubaro. (2020) Video Face Manipulation Detection Through Ensemble of CNNs. *25th International Conference on Pattern Recognition (ICPR)*

[7] Wanying Ge, Jose Patino, Massimiliano Todisco and Nicholas Evans. (2021) *arXiv:2110.03309v1*

[8] Chunlei Peng, Huiqing Guo, Decheng Liu, Nannan Wang, Ruimin Hu, Xinbo Gao. (2023) Deep Fidelity: Perceptual Forgery Fidelity Assessment for Deepfake Detection *arXiv:2312.04961v1*

[9] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, Wei Xia. (2021) Learning Self-Consistency for Deepfake Detection *IEEE/CVF International Conference on Computer Vision (ICCV)*

[10] Bojia Zi ,Minghao Chang ,Jingjing Chen, Xingjun Ma, Yu-Gang Jiang. (2021) Wild Deepfake: A Challenging Real- World Dataset for Deepfake Detection *arXiv:2101.01456v1*

[11] Kaede Shiohara Toshihiko Yamasaki. (2022) Detecting Deep fakes with Self-Blended Images *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR*)

[12] Shichao Dong, Jin Wang, Jiajun Liang, Haoqiang Fan, and Renhe Ji. (2022) Explaining Deepfake Detection by Analysing Image Matching. *arXiv:2207.09679v1*

[13] Anubhav Jain, Nasir Memon, Julian Togelius. (2022) A Dataless FaceSwap Detection Approach Using Synthetic Images. *IEEE International Joint Conference on Biometrics, IJCB. Institute of Electrical and Electronics Engineers Inc.*

[14] Fatima Maher Salman and Samy S. Abu-Naser. (2022) Classification of Real and Fake Human Faces Using Deep Learning. *International Journal of Academic Engineering Research (IJAER) 6 (3)*

[15] Tiewen Chen , Shanmin Yang, Shu Hu, Zhenghan Fang, Ying Fu, Xi Wu, Xin Wang. (2024) Masked Conditional Diffusion Model for Enhancing Deepfake Detection. *ArXiv, abs/2402.00541*

[16] Narayan, Kartik, et al. (2023). DF-Platter: Multi-Face Heterogeneous Deepfake Dataset. *Conference on Computer Vision and Pattern Recognition (CVPR)*

[17] Khan, Madeeha B.; Goel, Sanjay; Katar Anandan, Jaswant; Zhao, Jersey; and Naik, Ramavath Rakesh (2022). Deepfake Audio Detection. *International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*

[18] Almutairi, Z.; Elgibreen (2022). Review of Modern Audio Deepfake Detection Methods. *Algorithms, Vol 15, Issue 5. Academic Journal*

[19] Ilyas, Hafsa, Ali Javed, and Khalid Mahmood Malik (2023). AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio–visualdeepfakes detection. *arXiv:2305.01979v3*