e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 4, April 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

International Journal Of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704183 |

Hypothesis of Forgery Identification Using Meta Deepfake Detection Model

Dr. S. Vijayaragavan, Vijayashree V, Vahini D, Nivetha T

Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: Towards the further scope of quantum computing, driven quantum algorithms such as quantum neural networks (QNNs) have a great potential for tackling the issue of the classification of bona fide or synthetic images, audio, and videos. However, the field is still evolving, and practical quantum computers may face challenges and limitations. Algorithms or models prepared for DeepFakes according to the quantum architecture depend on the specific techniques employed for classifying the pure-set and false-set from the multi-modal corpus. Consequently, classical computers may struggle to detect such DeepFakes if quantum algorithms exploit the unique properties of quantum systems. As the improvement of quantum computing is ongoing, continuous research is needed to understand its full implications in AI, especially vision intelligence. It would be another hot topic of further research for designing compatible and adaptable algorithms for DeepFakes that will work for both classical and quantum architectures or systems.

KEYWORDS: Modality Fusion In Deepfake Detection; Comprehensive Review Of Deepfake Detection

I. INTRODUCTION

DeepFakes are causing significant concern among the general public. For instance, fake videos created by fraudsters can easily deceive the general public [1]. Such fake videos can spread virally on social media, causing irreversible harm to targeted individuals or organizations (e.g., high-profile personalities or a company with significant brand value). A more sinister threat emerges when DeepFakes are used to create child pornography or sexually explicit fake content [2,3].

Generally, humans cannot distinguish a real video from a DeepFake with the naked eye (or ears) [4]. On a superficial level, DeepFakes are created by combining several techniques, such as merging, combining, replacing, and superimposing images and video clips to create fake videos [5], making them appear real. Taking advantage of more recent AI techniques such as generative adversarial networks (GANs), DeepFakes can now generate hyper-realistic content by incorporating audio into the video, thereby not only altering the visual content but also making it realistic in terms of audio [6].

Several approaches have recently been proposed to detect such manipulated content by analyzing spatial and frequency information in images, as well as temporal and frequency information from audio and video. To advance the state of the art in detecting DeepFakes, several benchmarking datasets have been made available to the public. By leveraging these databases and existing approaches, state-of-the-art methods have evolved to exploit the concept of information fusion, providing robust detection of fake media.

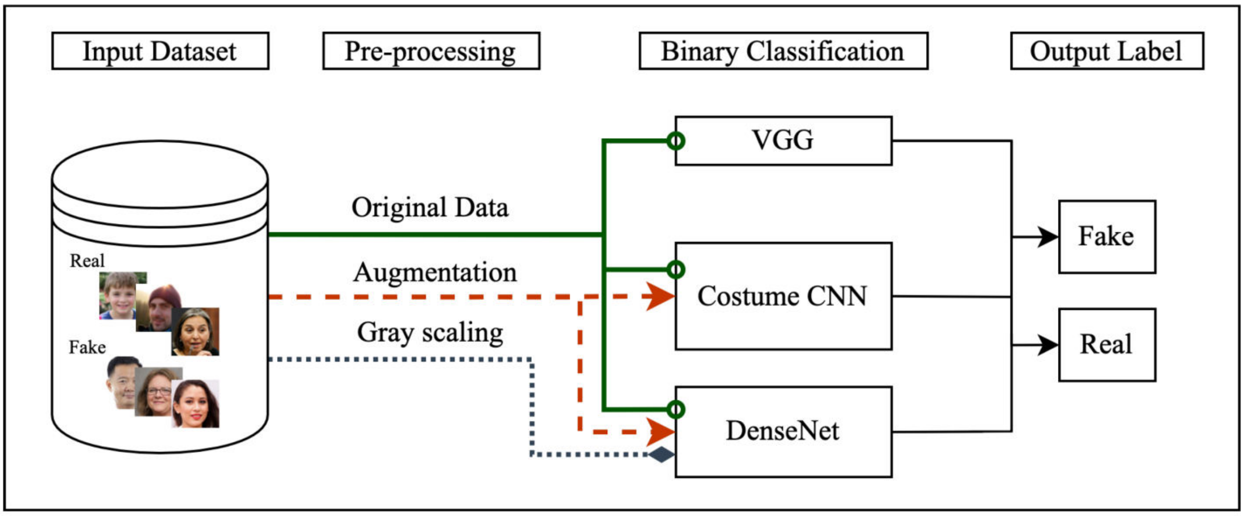

Fig 1: Comparison of Deepfake

Although there are numerous surveys on DeepFake detection (DFD) [7] detailing advances, challenges, and potential future work, this paper focuses on a sub-topic of media modality fusion in DFD, thus complementing existing review works. In this paper, we delve into existing approaches in DFD, listing relevant works and benchmarking databases and their respective results. Furthermore, we discuss challenges and potential future work aimed at pushing the boundaries of the current state of the art in DeepFake detection.

Used the DFDC dataset to train the DL models for detecting fraudulent mages or videos. In this paper, researchers put their efforts into the extraction of face features especially for false positive images, that can generate a large noisy corpus of contents for ameliorating the detection accuracy. Prior to feeding these features to three proposed deep learning architectures, MesoInception-4, XceptionNet, and EfficientNet, the authors integrated two pre-processing steps: a data augmentation layer and an image filtering layer. In the first step, they pre-processed the dataset, including transformations such as horizontal and vertical flipping, random cropping, rotation, compression, Gaussian and motion blurring, and brightness, saturation, and contrast transformation. This staging layer is used to improve the quality of the image. In the second pre-processing layer, they eliminate images whose sizes are less or equal to N/2 when it is in a connected form, where N is the number of extracted frames per video after the face extraction. At last, the DL models incorporate sigmoid activation in the final layer, Adam optimizer, and minimization in log loss error while training on DFDC.

Here, in the given figure, p2, p3, p5, and p6 measure the height, whereas p1 and p4 measure the eye width. These points are responsible for determining whether the eyes are closed or open. In this study, the average human eye blinking rate was used as a threshold to detect and count the eye blink and blink intervals because the normal blinking rate of humans is between 2 and 10 s, and each eye blink will take between 0.1 and 0.4 s. Based on this calculation, the authors classified fake and real videos. It detected 184 eye blinks per minute on real videos and 428 eye blinks per minute on fake videos, and the overall accuracy was 93.23% and 98.1% for real and fake videos, respectively.

II. RELATED WORK

Korshunov et al. [16] also evaluated baseline face-swap detection algorithms and found that the lip-sync-based approach failed to detect mismatches between lip movements and speech. They also verified that image quality measures with a Support Vector Machine (SVM) classifier can detect high-quality DeepFake videos with an 8.97% equal error rate.

Agarwal and Varshney [18] designed a statistical model based on hypothesis testing to detect face-swapped content or fraudulence in images. In this study, the authors considered a mathematical bound value corresponding to the error probability based on the detection of genuine or GAN-generated images.

Lyu [19] highlighted the key challenges in detecting DeepFakes using high-quality or definition-synthesized videos and audio clips. The author tried to raise a deep concern about one of the critical disadvantages of the current DeepFake generation methods that cannot produce a fine mapping of color shades for hair with respect to the human face.

Based on the above discussion, this paper highlighted the laconic view of the proposed DFD pipeline or mechanism and was also concerned about the future amelioration of advanced DFD. Here, the author proposed an adversarial perturbation-enabled model that will give less emphasis on DNN-based face detectors. The proposed detection model consisted of two phases: (a) face detection phase enabling the adversarial perturbation approach and (b) an AI system to detect DeepFake.

Kumar et al. implemented several DL approaches and compared their results with the context of DeepFake classification using metric learning. The authors used a Multitask Cascaded Convolutional Neural Network (MTCNN) to extract faces from images or videos. The MTCNN incorporates three networks: (a) a proposal network, (b) a refine network, and (c) output networks to suppress overlapping boxes using non-max-suppression and generate bounded faces. The Xception architecture was used for transfer learning, and sequence classification was applied using LSTM in addition to 3D convolution and a triple network. A triplet network was used with metric learning for proposing an approach that counts the number of frames in a particular video clip. The realism factor had to be accessed if the number of frames was found to be less than the actual number of frames when compared with pristine video. In this study, three types of triplet-generation methods were investigated. These were easy triplets, semi-hard triplets, and hard triplets, which are based on the distance between the anchor, positive, and negative embedding vectors. The proposed detection architecture, as shown in **Figure 7**, leverages XceptionNet for the entire process using the MTCNN. In the first phase, the FaceNet model is used to detect, extract, and generate a feature space which is 512 dimension embedding vectors for each face. Subsequently, the generated feature space is fed to semi-hard triplets that discriminate between fake frames and pristine frames through triplet loss. During validation, this approach achieved an AUC score of 99.2% on Celeb-DF and accuracy of 99.71% on a highly compressed neural texture.

III. METHODS

Drew attention to the growing danger posed by DeepFake videos, which are realistic but artificially produced videos that might trick viewers by depicting things or people that do not actually happen or exist. Owing to their increasingly complex generating processes, modified films are sometimes difficult for traditional DeepFake detection systems to recognize correctly. The authors suggested a multi-attentional strategy that combines self-attention, spatial attention, and temporal attention mechanisms to overcome this difficulty. These attention processes enabled the model to focus on essential regions and patterns while filtering out extraneous data, allowing it to effectively capture both global and local contextual information within videos.

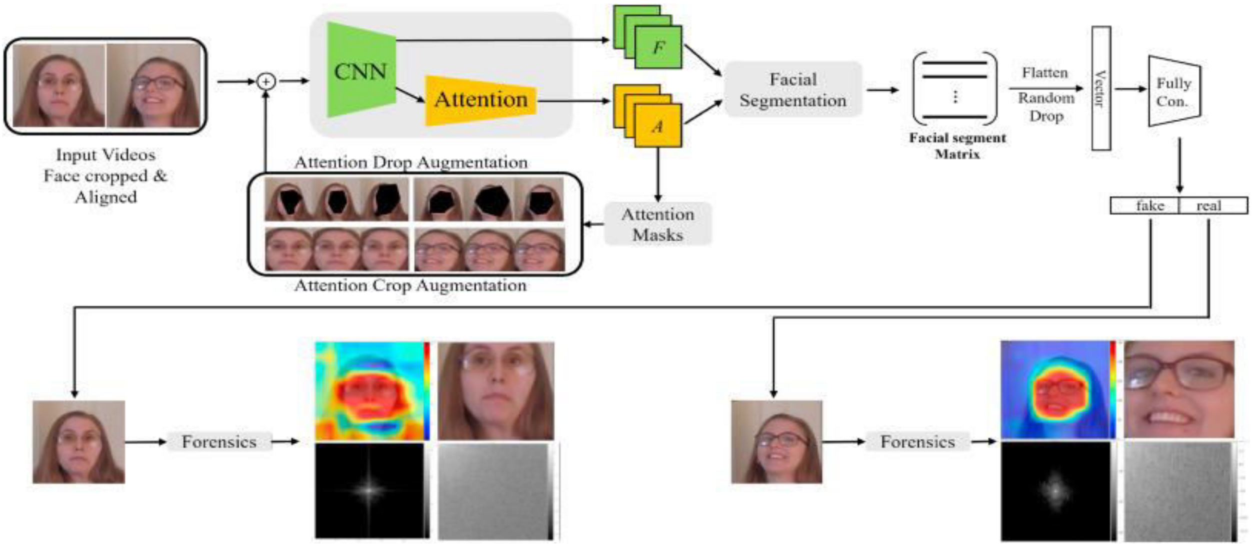

Fig 2: Deepfake forensics analysis

The proposed DeepFake detection model can identify artifacts, inconsistencies, or anomalous patterns that point to DeepFake manipulation by incorporating these attention mechanisms. For decision-making, the model examines visual and temporal clues, such as facial expressions, eye movements, and motion patterns. The authors stressed the value of using sizable datasets that include a variety of DeepFake variations while training the multi-attentional DeepFake detection model. The generalization and resilience of the model against unknown manipulation approaches were enhanced using this method. To excel on various DeepFake video formats, the model can also benefit from transfer learning and domain-adaptation techniques. However, the authors noted that there is still competition between DeepFake production techniques and detection approaches. To remain ahead of harmful actors and ensure the development of efficient DeepFake detection techniques, they emphasized on the necessity for ongoing research, innovation, and collaboration among the scientific community, industry, and governments.

Highlighted that, while visual cues have been extensively utilized in DeepFake detection, audio information can provide valuable complementary signals. Manipulated videos often exhibit discrepancies between audio and visual components because of the challenges of synchronizing fake audio with manipulated visual content. To address this, the authors proposed a joint audio-visual DeepFake detection approach that simultaneously analyzes both audio and visual aspects of videos. The model leverages deep learning techniques to extract relevant features from both modalities and integrates them to make a joint decision regarding the authenticity of the video. The visual component of the model utilizes Convolutional Neural Networks (CNNs) to extract visual features from frames or facial regions of the input video. These features capture visual cues such as facial expressions, inconsistencies in facial movements, or artifacts introduced during DeepFake manipulation.

Simultaneously, the audio component of the model employs audio processing techniques, such as spectrogram analysis, to extract relevant audio features. These features capture acoustic cues such as speech patterns, speaker characteristics, and anomalies in audio quality. The extracted audio and visual features are then fused using fusion mechanisms, such as concatenation or attention mechanisms, to create a joint representation that captures combined information from both modalities. This joint representation is fed into a classification model that determines whether the video is genuine or manipulated.

IV. RESULT ANALYSIS

MTFF-Net used a variety of visual elements retrieved from DeepFake videos to improve detection. Color histogram, optical flow, Convolutional Neural Networks (CNNs), and long short-term memory (LSTM) features were the four main visual features of the network. Each element served as a representation of a different aspect of the video content and offered helpful hints for differentiating between real and fake videos.

The color histogram feature recorded statistical data on color distributions in frames, allowing for the detection of anomalies or inconsistencies caused by DeepFake manipulation. Using the optical flow function, it was possible to identify anomalies that may be present in DeepFake videos by capturing the motion patterns between frames. The CNN features were extracted using pre-trained CNN models, which selected high-level representations from the input frames. These features could distinguish between real content and staff that have been altered, and record intricate visual patterns.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704183 |

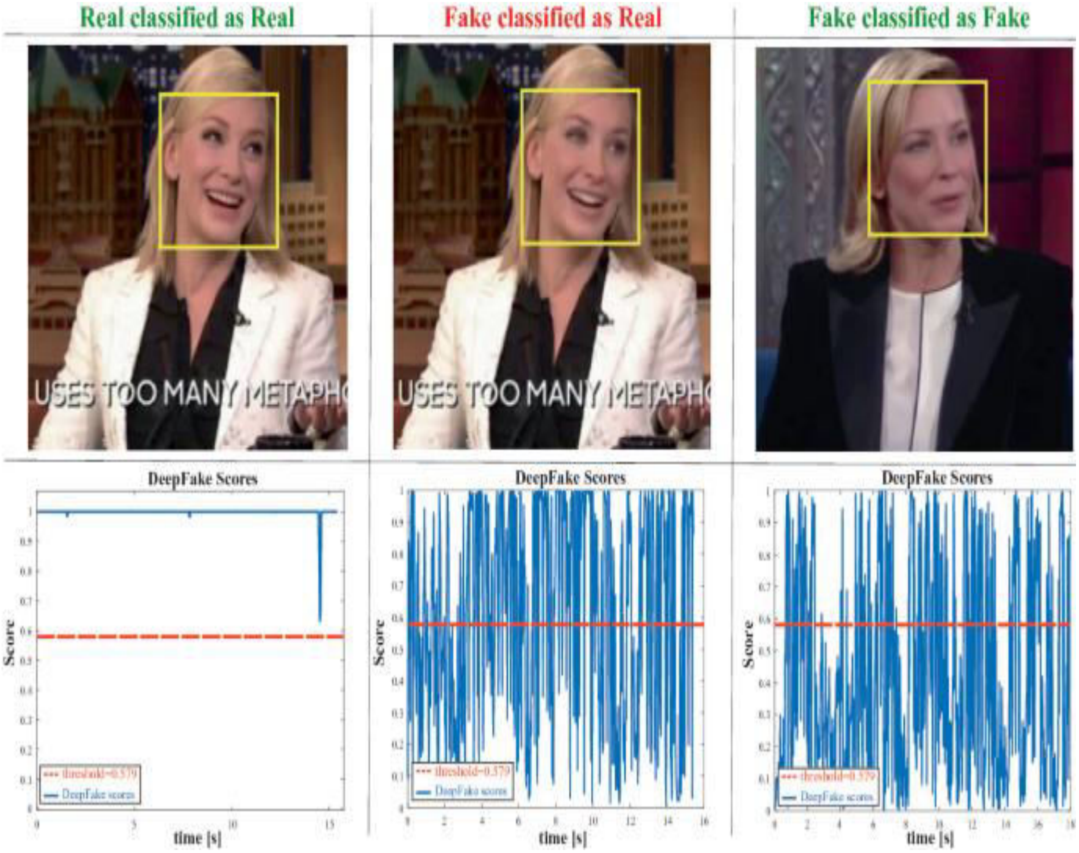

Fig 3: result analysis of DeepFakes Detection

The LSTM features were obtained from LSTM networks, which considered the sequential data between frames and captured the temporal dynamics of the video. These characteristics were particularly helpful for spotting temporal irregularities that DeepFake videos frequently contained. The authors' multi-branch architecture extracted discriminative representations from each modality by processing each piece of visual information separately. The features were then combined at several levels, enabling the network to take advantage of the complementary data offered by each feature. The authors employed a sizable dataset that contained a wide variety of DeepFake movies to train MTFF-Net. To optimize the network parameters and facilitate precise detection, they used proper loss functions and optimization approaches. The experimental findings in this study showed that when compared to single-feature-based approaches and other cutting-edge DeepFake detection models, MTFF-Net performed better. The network can capture a thorough grasp of the video content owing to the multi-feature fusion strategy, improving the detection accuracy and robustness against various DeepFake manipulation approaches.

V. CONCLUSION

The image and video feature sections on DFD elicited readers to become familiar with all the novel efforts that have been made by researchers from late 2017 to date. Although the work conducted on DeepFakes by the researchers or research groups has indeed progressed a lot towards the refinement and betterment in the existing models, there is still a large scope of further research for improving the detection pipeline in terms of precision, time efficiency, cost efficiency, and ease of interaction with real-world applications, which can curtail and act as fuel for this DeepFake detection challenge. DeepFake detection models often struggle to generalize across diverse datasets, leading to reduced effectiveness in real-world scenarios with variations in lighting conditions, facial expressions, and video quality. Another main challenge is raised due to the "unseen class of some facial datasets" in the testing dataset with respect to the training dataset. Augmenting training datasets with diverse samples, employing transfer learning from pre-trained models, and integrating attention mechanisms can enhance generalization capabilities

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704183 |

REFERENCES

1. Ajao, O.; Bhowmik, D.; Zargari, S. Sentiment aware fake news detection on online social networks. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2507–2511. [**Google Scholar**]
2. Vaccari, C.; Chadwick, A. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Soc. Media+ Soc.* **2020**, *6*, 2056305120903408. [**Google Scholar**] [**CrossRef**]
3. Eelmaa, S. Sexualization of Children in Deepfakes and Hentai: Examining Reddit User Views. *SocArxiv* **2021**, *10*. [**Google Scholar**]
4. Nguyen, T.T.; Nguyen, Q.V.H.; Nguyen, D.T.; Nguyen, D.T.; Huynh-The, T.; Nahavandi, S.; Nguyen, T.T.; Pham, Q.V.; Nguyen, C.M. Deep learning for deepfakes creation and detection: A survey. *Comput. Vis. Image Underst.* **2022**, *223*, 103525. [**Google Scholar**] [**CrossRef**]
5. Yu, P.; Xia, Z.; Fei, J.; Lu, Y. A survey on deepfake video detection. *Iet Biom.* **2021**, *10*, 607–624. [**Google Scholar**] [**CrossRef**]
6. Brock, A.; Lim, T.; Ritchie, J.M.; Weston, N. Neural photo editing with introspective adversarial networks. *arXiv* **2016**, arXiv:1609.07093. [**Google Scholar**]
7. Afzal, S.; Ghani, S.; Hittawe, M.M.; Rashid, S.F.; Knio, O.M.; Hadwiger, M.; Hoteit, I. Visualization and Visual Analytics Approaches for Image and Video Datasets: A Survey. *ACM Trans. Interact. Intell. Syst.* **2023**, *13*, 5. [**Google Scholar**] [**CrossRef**]
8. Akhtar, Z. Deepfakes Generation and Detection: A Short Survey. *J. Imaging* **2023**, *9*, 18. [**Google Scholar**] [**CrossRef**]
9. Narayan, K.; Agarwal, H.; Thakral, K.; Mittal, S.; Vatsa, M.; Singh, R. DF-Platter: Multi-Face Heterogeneous Deepfake Dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 9739–9748. [**Google Scholar**]
10. Li, Y.; Chang, M.C.; Lyu, S. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 11–13. [**Google Scholar**]

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com