International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

A SURVEY ON DEEPFAKE VIDEO DETECTION USING DEEPLEARNING

Mrs.M.Lakshmi Prabha[1], N.Aakash[2], S.Balaganesh[3], B.Roopan[4]

[1]Assosiate Professor, Dept. of IT, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry
[2,3,4] Student, Dept. of IT, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry

---

Abstract - *In response to the growing danger that deepfake films represent to the authenticity of visual material, this study explores the effectiveness of using physiological signals heart rate, eye movement, and facial emotions in particular for detection. The identification of falsified content is significantly hampered by deepfake technology, which has led to research into new detection techniques. Using knowledge from other studies, we examine whether physiological signals can serve as trustworthy markers of dishonesty. Promising approaches for identifying irregularities in visual content include heart rate monitoring, eye movement tracking, and facial expression detection. By doing an extensive analysis of current research, techniques, and datasets, we evaluate the benefits and drawbacks of each strategy, opening the door for hybrid models that combine several physiological signals for improved accuracy. (Size 10 & Italic , cambria font)*

Key Words: **CNN, RNN, LSTM, PPG, GNN**

1.INTRODUCTION

The rise of deepfake technology has sparked widespread concern due to its ability to create highly convincing fake videos, blurring the line between reality and fiction. These videos, generated using sophisticated AI algorithms, can manipulate images and audio to make individuals appear to say or do things they never did. This poses a serious threat to the credibility of visual media and raises questions about the reliability of information shared online. Traditional methods of detecting deepfakes, such as manual inspection, are often inadequate in the face of rapidly advancing technology. In response, researchers are exploring new avenues, including the analysis of physiological signals, to detect these deceptive videos. Physiological signals like heart rate, eye movement, and facial expressions can provide valuable insights into how our bodies react to what we see. By studying these signals, researchers hope to uncover subtle differences between real and fake videos, ultimately developing automated tools for deepfake detection. This survey paper aims to explore the use of physiological signals, along with advanced deep learning techniques like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, in the fight against deepfakes. By examining existing research, methodologies, and challenges, we seek to provide a comprehensive understanding of the current state of deepfake detection. Additionally, we aim to identify areas for improvement and propose future research directions to enhance the effectiveness of deepfake detection methods. Ultimately, our goal is to address the growing threat of deepfake videos and restore trust and authenticity in digital media.

1.1 HEART RATE ANALYSIS

Heart rate serves as a fundamental physiological signal, representing the frequency of heartbeats per minute. In video analysis, heart rate can be measured through various methods, including photoplethysmography (PPG) sensors or computer vision techniques that track changes in facial blood flow. Heart rate variability (HRV), the variation in time intervals between consecutive heartbeats, holds particular significance in detecting deception. Research suggests that heightened emotional arousal, often associated with deception, can influence HRV patterns, making it a potential marker for identifying deceitful behavior.

1.2 EYE MOVEMENT ANALYSIS

Eye movement patterns provide valuable insights into cognitive processes and attentional mechanisms. Eye-tracking technology enables precise measurement and analysis of eye movements, including fixations, saccades, and smooth pursuit. In video analysis, eye-tracking data can reveal where and how individuals direct their visual attention. Correlations between eye movements and cognitive processes, such as memory retrieval and decision-making, underscore the utility of eye movement analysis in understanding human behavior and detecting deception.

1.3 FACE EXPRESSION ANALYSIS

Facial expressions play a crucial role in human communication, conveying a wide range of emotions and intentions. In video analysis, facial expression recognition techniques utilize computer vision algorithms to detect and classify facial expressions based on key facial landmarks and muscle movements. Emotional cues, such as changes in facial expressions, can provide valuable cues for detecting deception. Additionally, micro-expressions brief, involuntary facial expressions that occur within milliseconds hold particular promise in discerning concealed emotions and deceptive behavior in video content.

1.4 CNN BASED APPROACHES

CNN-based approaches are making significant strides in the realm of deepfake detection. Researchers are harnessing powerful architectures like VGGNet, ResNet, and InceptionNet to sift through visual data and pinpoint telltale

signs of manipulation. But here's where it gets really interesting. CNNs aren't just limited to visuals anymore. They're also being trained to interpret physiological signals like heart rate and eye movement. By combining these signals with visual analysis, researchers are hoping to create more accurate detection systems. To test these models, researchers train them on datasets containing both real and fake videos and compare their performance with traditional methods. This holistic approach is driving forward the field of deepfake detection, leveraging CNNs to their fullest potential in both visual and physiological signal analysis.

1.5 LSTM BASED APPROACHES

LSTM-based methods are really making a splash in the world of deepfake detection. You see, these Long Short-Term Memory (LSTM) networks are like the detectives of sequential data analysis. They're fantastic at spotting patterns over time in video sequences, which is crucial for catching those sneaky deepfakes. What's fascinating is how researchers are using LSTMs to process sequences of physiological signals, such as heart rate and eye movement data. By studying the changes in these signals over time, LSTMs can tell the difference between real and fake videos. To see how well they perform, researchers put them through rigorous testing, comparing their accuracy with traditional methods. It's all about improving the reliability of deepfake detection systems and keeping digital media trustworthy.

1.6 HYBRID CNN-LSTM BASED APPROACHES

Hybrid CNN-LSTM methods are really gaining traction in the fight against deepfake videos. They're like a dynamic duo, combining the spatial awareness of Convolutional Neural Networks (CNNs) with the time-savvy skills of Long Short-Term Memory (LSTM) networks. Together, they can spot patterns in both space and time, making them super effective at catching those tricky deepfakes. Researchers have been experimenting with different ways to blend CNNs and LSTMs, like using CNNs to pick up spatial features from each frame and LSTMs to analyze how these features change over time. What's exciting is that these hybrid models consistently outperform standalone CNN or LSTM approaches, showing that teamwork really does make the dream work when it comes to deepfake detection. It's all about leveraging the strengths of both architectures to keep our digital world honest and trustworthy.

2. LITERATURE SURVEY

A novel approach for detecting deep fake videos using graph neural network:

This recent paper proposes a novel approach using Graph Neural Networks (GNNs) for deepfake detection. GNNs excel at modeling relationships between data points[1]. In the context of deepfakes, this translates to capturing the complex relationships between facial features in videos. By analyzing these relationships, GNNs can identify inconsistencies indicative of manipulation, such as unnatural connections between facial landmarks (e.g., eyes and nose). The paper also explores how GNNs can be combined with other models (like CNNs) through fusion strategies, potentially leading to even better deepfake detection performance.

Generalization of Forgery Detection With Meta Deepfake Detection Model:

This paper introduces a groundbreaking deepfake detection methodology termed Meta Deepfake Detection (MDD), leveraging the principles of meta-learning[2]. MDD is conceived to tackle the prevalent issue of poor generalization observed in current deepfake detection models. Existing deepfake detection models often falter when confronted with unseen manipulation techniques, attributed to their limited generalizability. The paper proposes MDD as a solution to this challenge, aiming to enhance performance particularly in unexplored domains. MDD employs meta-learning, a technique enabling the acquisition of transferable knowledge from diverse source domains. Through meta-learning, the model becomes adept at adapting to novel, unseen domains without necessitating frequent model updates. Key components of MDD include meta-splitting, where source domains are split into meta-train and meta-test sets to mirror real-world domain shifts. This enables the model to simulate diverse scenarios encountered in practical applications. Data preprocessing involves using existing datasets such as DFDC, Celeb-DF-v2, and FaceForensics++, along with face extraction, resizing, and applying block shuffling transformation to augment data diversity. This comprehensive methodology, encapsulated by MDD, represents a significant advancement in the realm of deepfake detection, promising enhanced performance and robustness across varying domains.

DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms:

This paper proposes a comprehensive approach to address the growing threat of deepfake technology, which has raised significant concerns regarding misinformation and digital fraud[3]. It outlines the current landscape of deepfake research, highlighting the need for effective detection algorithms to combat the spread of fake content. Researchers have been actively exploring various methods for both generating and detecting deepfake videos. Generative models like GANs have been utilized to create realistic deepfakes, while detection techniques leverage CNNs and other methodologies to analyse visual cues and distinguish between authentic and fake videos. The integration of recurrent networks and vision transformers has further improved detection accuracy, capturing both spatial and temporal features. Specialized datasets, such as

FakeAVCeleb and CelebDF, have been developed to facilitate the training and evaluation of detection models. Overall, the paper emphasizes the multidimensional nature of deepfake research and underscores the importance of ongoing efforts to develop robust detection mechanisms to combat the harmful effects of deepfake technology on society.

An Exploratory Analysis on Visual Counterfeits Using Conv-LSTM Hybrid Architecture:

Exploratory Analysis on Visual Counterfeits Using Conv-LSTM Hybrid Architecture" delves into the intricate realm of detecting visual counterfeits, particularly in the form of synthetic face swaps known as Deepfakes[4]. By combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures, the study proposes a novel approach for identifying minute visual traces indicative of counterfeits in video frames. The paper introduces the concept of facial landmarks, where 512 landmarks are extracted and compared between real and fake frames. Parameters such as eye blinking, lip synchronization, and eyebrow movement are analysed to discern between authentic and counterfeit visual data. The recurrent nature of LSTM allows for learning based on these features, enhancing the model's ability to evaluate visual data effectively. The architecture presented in the study integrates facial landmarks movement analysis with CNN-based feature extraction and LSTM for temporal data processing. By leveraging CNN for feature extraction and LSTM for temporal memory, the model achieves competitive performance without excessive memory overhead. The facial landmark detection process involves aligning landmarks in 2D space and evaluating dense face feature descriptors, providing comprehensive insight into facial movements. These landmarks serve as crucial indicators for detecting anomalies in counterfeit videos compared to authentic ones. Additionally, CNN feature extractors are utilized to reduce spatial complexity, extracting raw abstract face features from frame images without saving the frames themselves. Transfer learning is employed to leverage pre-trained models, enhancing the efficiency of feature extraction. In the LSTM pipeline, the model memorizes patterns in video frames, utilizing mean pooling layers to extract frame-level visual features. These features are then fed into LSTM cells for further analysis, allowing the model to discern patterns indicative of counterfeit visual data. The paper outlines a comprehensive methodology for detecting visual counterfeits, emphasizing the importance of leveraging both spatial and temporal information encoded in video frames. By combining CNN and LSTM architectures, the proposed hybrid approach sets a new benchmark for detecting visual counterfeits effectively, contributing significantly to the field of counterfeit detection and cybersecurity.

Detection of DeepFake Videos Using Recurrent Neural Network (RNN):

This study proposes a new model that integrates Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and image preprocessing techniques to accurately classify fake videos from real ones[5]. The proposed model is implemented and evaluated using a MATLAB simulator with the DeepFake Images dataset, comprising 135 real videos and 677 fake videos generated through various tools. The primary aim of this research is to devise an accurate model for detecting DeepFake videos. Three sub-goals stem from this objective. Firstly, to ensure consistent findings across different sources, a dataset combining data from various distributions is prepared. Secondly, efforts are made to minimize the discrepancy between attained accuracy and validation accuracy, especially when the model is applied to videos from different distributions. Although previous CNN and RNN-based DeepFake detection systems achieved high accuracy levels in tests with videos from the same distribution, they struggled to maintain similar precision across various distributions. Thirdly, the goal is to develop a real-time platform where users can upload or select videos for DeepFake detection.

A Novel Deep Learning Approach for Deepfake Image Detection:

This paper proposes a novel deep learning approach termed the Deepfake Predictor (DFP), designed to address the challenges posed by increasingly sophisticated deepfake content[6]. The DFP approach leverages a hybrid architecture combining elements of the VGG16 and convolutional neural network (CNN) models. This pioneering strategy marks a departure from traditional transfer learning techniques and introduces a tailored framework specifically optimized for deepfake detection. The proposed mechanism involves a multi-stage process, wherein deepfake images based on fake and real human faces are utilized to construct a structured dataset. This dataset is then split into training, validation, and test sets, with the majority used for training the employed neural network techniques. Furthermore, the study extensively evaluates several transfer learning techniques such as Xception, NAS-Net, and Mobile Net alongside the proposed DFP approach. Comparative analysis reveals the efficacy of the DFP model, which outperforms existing state-of-the-art methods in terms of precision and accuracy. The research underscores the significance of advanced neural network architectures and hyperparameter optimization in enhancing the detection capabilities of deepfake media. Overall, the literature review highlights the evolution of deepfake detection methodologies, from conventional transfer learning approaches to innovative hybrid architectures tailored for the unique challenges posed by synthetic media manipulation.

Deep Fake Video Detection Using Transfer Learning Approach:

This paper proposes a novel framework for the detection of fake videos, commonly known as deepfakes, which have become a significant concern due to their potential for spreading misinformation and causing social and political unrest[7]. As the internet serves as a rapid medium for the dissemination of fake news, the development of computational tools to combat this phenomenon becomes increasingly crucial. Deepfake generation algorithms, readily available on cloud platforms at low computational costs, enable the creation of highly realistic fake content, further exacerbating the challenge of detecting such manipulations. In response to this pressing issue, the proposed framework integrates transfer learning in autoencoders with a hybrid model of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). This approach aims to enhance the accuracy and robustness of deepfake detection by leveraging transfer learning techniques and combining the strengths of CNNs and RNNs. By incorporating transfer learning, the framework can adapt pre-trained models to the task of deepfake detection, thereby improving performance even in the face of unseen attacks or changes in the domain.

Deepfakes Creation and Detection Using Deep Learning:

This paper surveys deepfakes, exploring how they are created, detected, and potentially even enhanced, all through the lens of deep learning[8]. Deep learning has revolutionized many fields, but it has also led to the creation of deepfakes – synthetic media used to fabricate events or actions. Deepfakes, especially those that swap faces, can be very realistic and pose a threat to privacy and security. To address this, researchers have focused on two areas: deepfake creation and detection. Deepfakes are created using techniques like Generative Adversarial Networks (GANs), which train models on real images to produce realistic forgeries. For detection, Convolutional Neural Networks (CNNs) have proven powerful. One such CNN architecture is MesoNet, designed specifically to detect deepfakes. It achieves high accuracy even on low-quality videos found on social media. This is because it avoids focusing on high-level semantic analysis or microscopic noise, instead opting for an intermediate strategy using a simpler CNN architecture. The key to MesoNet's success is its ability to extract features. It uses alternating convolutional and pooling layers, followed by a dense network with a single hidden layer. Convolutional layers extract features, while pooling layers create smaller versions of the feature maps, allowing for efficient analysis. Additionally, the network uses techniques to improve its performance, such as ReLU activation functions and Batch Normalization. Dropout regularization is also used to improve the robustness of the fully-connected layers.

A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods:

This paper provides an in-depth exploration of deep learning architectures utilized in the detection of deepfakes[9]. It delves into Convolutional Neural Networks (CNNs), which serve as foundational tools in the analysis of images and videos. Architectures such as VGG and ResNet are frequently employed for extracting features from deepfake content, enabling effective detection. In addition to CNNs, the paper investigates Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, which play a crucial role in examining temporal information embedded within videos. By capturing subtle inconsistencies in facial movements, RNNs aid in identifying potential manipulation within video sequences. Moreover, the paper explores Fusion Methods, which involve combining the strengths of CNNs and RNNs. Techniques such as early fusion, which involves merging features before classification, and late fusion, which integrates classification outcomes from individual models, are examined for their efficacy in deepfake detection. By examining and analysing these deep learning architectures, this paper contributes to a comprehensive understanding of the tools and methodologies employed in the detection of deepfakes, paving the way for enhanced detection and mitigation strategies.

A Review of Deep Learning-based Approaches for Deepfake Content Detection:

This paper sheds light on the specific challenges posed by deepfakes and elucidates how deep learning tackles them[10]. It begins by scrutinizing Face Swapping Techniques commonly utilized in deepfake generation, such as Face2Face and Deepfakes. Through deep learning models, the paper examines how inconsistencies in facial features, skin texture, and lighting, which are indicative of manipulation, can be detected and analysed. Furthermore, the paper delves into the critical issue of Dataset Bias, emphasizing that the effectiveness of deep learning models is heavily contingent on the quality and diversity of training data. It underscores the significance of datasets encompassing a broad spectrum of deepfake creation techniques alongside authentic videos. Such diverse datasets are crucial for training robust models capable of effectively discerning between genuine and manipulated content.

3. CONCLUSIONS

In conclusion, we've discussed how combining CNNs and LSTMs improves deepfake detection. Utilizing physiological signals like heart rate and eye movement enhances accuracy. Challenges include dataset biases, but addressing them will strengthen detection systems. Future research should focus on integration and model interpretability to stay ahead in combating deepfake proliferation.

REFERENCES

[1] El-Gayar MM, Abouhawwash M, Askar SS, Sweidan S. A novel approach for detecting deep fake videos using graph neural network. Journal of Big Data. 2024 Feb 1

[2] Tran VN, Kwon SG, Lee SH, Le HS, Kwon KR. Generalization of forgery detection with meta deepfake detection model. IEEE Access. 2022 Dec 26

[3] Alnaim NM, Almutairi ZM, Alsuwat MS, Alalawi HH, Alshobaili A, Alenezi FS. "DFFMD: a deepfake face mask dataset for infectious disease era with deepfake detection algorithms" (IEEE Access. 2023 Feb 20)

[4] Hashmi MF, Ashish BK, Keskar AG, Bokde ND, Yoon JH, Geem ZW. "An exploratory analysis on visual counterfeits using conv-lstm hybrid architecture" (IEEE Access. 2020 May 28).

[5] Albazony, Ali Abdulzahra Mohsin, et al. "DeepFake Videos Detection by Using Recurrent Neural Network (RNN)." 2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT). IEEE, 2023

[6] Raza A, Munir K, Almutairi M, "A novel deep learning approach for deepfake image detection" Applied Sciences 2022 Sep 29

[7] Suratkar S, Kazi F, "Deep fake video detection using transfer learning approach" Arabian Journal for Science and Engineering. 2023 Aug 2021

[8] Khalil, Hady A., and Shady A. Maged. "Deepfakes creation and detection using deep learning." 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) IEEE, 2021

[9] Gupta G, Raja K, Gupta M, Jan T, Whiteside ST, Prasad M. "A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods" Electronics. 2023 Dec 25 2020

[10] Passos LA, Jodas D, Costa KA, Souza Júnior LA, Rodrigues D, Del Ser J, Camacho D, Papa JP. "A review of deep learning-based approaches for deepfake content detection" Expert Systems. 2022

[11] Chen B, Li T, Ding W. "Detecting deepfake videos based on spatiotemporal attention and convolutional LSTM". Information Sciences. 2022 Jul 1

[12] Al-Dhabi, Yunes, and Shuang Zhang. "Deepfake video detection by combining convolutional neural network (cnn) and recurrent neural network (rnn)." 2021 IEEE international conference on computer science, artificial intelligence and electronic engineering (CSAIEE). IEEE, 2021

[13] Masud U, Sadiq M, Masood S, Ahmad M, Abd El-Latif AA. "LW-DeepFakeNet: a lightweight time distributed CNN-LSTM network for real-time DeepFake video detection" Signal, Image and Video Processing. 2023 Nov;17

[14] Saikia, Pallabi, et al. "A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features." 2022 international joint conference on neural networks (IJCNN). IEEE, 2022

[15] Al-Dhabi, Yunes, and Shuang Zhang. "Deepfake video detection by combining convolutional neural network (cnn) and recurrent neural network (rnn)." 2021 IEEE international conference on computer science, artificial intelligence and electronic engineering (CSAIEE). IEEE, 2021

[16] Zhang T. "Deepfake generation and detection, a survey" Multimedia Tools and Applications. 2022 Feb

[17] Rebello, Lian, et al. "Detection of Deepfake Video using Deep Learning and MesoNet." 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2023

[18] Aduwala, Sai Ashrith, et al. "Deepfake Detection using GAN discriminators." 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2021

[19] Vidya, K., et al. "Compressed Deepfake Detection using Spatio-Temporal Approach with Model Pruning." Procedia Computer Science 230 (2023)