Sakshi Fand..2023,11:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology

An Open Access Journal

Deep fake Detection using Deep Learning

Prof. Aparna Bagde, Sakshi Fand, Kanchan Varma, Aditya Gawali
Dept. of Computer Engineering
NBNSTIC, Pune, Maharashtra, India.

Abstract- The rapid advancements in AI, machine learning, and deep learning technologies, which have given rise to new tools for manipulating multimedia. While these technologies have found legitimate applications in entertainment and education, they have also been exploited for malicious purposes. Notably, high-quality and realistic fake multimedia content, known as Deepfake, has been used to spread misinformation, incite political discord, and engage in malicious activities like harassment and blackmail.Deepfake algorithms possess the unsettling ability to craft counterfeit images and videos so convincingly that they elude human scrutiny. These algorithms adeptly fashion deceptive visual and auditory content, manipulating the appearances and behaviour of targeted individuals to such a degree that viewers instinctively place their trust in what is, in fact, a fabrication. Distinguishing these deepfakes from genuine content becomes a formidable challenge, as the human eye struggles to discern the difference. In response, this paper conducts a comprehensive exploration, delving into the array of tools and algorithms employed in the creation of deepfakes, while placing particular emphasis on the vital aspect of deepfake detection methods. Through in-depth discussions that encompass challenges, research endeavour, technological advancements, and strategic approaches linked to the realm of deepfakes, this survey scrutinizes the landscape. By tracing the evolution of deepfakes and appraising the current state of deepfake identification techniques, it offers a holistic evaluation of deepfake methodologies. This, in turn, contributes to the formulation of innovative and more resilient strategies, essential for countering the ever-growing sophistication of deepfake technology.

Keywords- Deepfake Detection, Deep Learning, Video or Image manipulation

I.INTRODUCTION

In the lead-up to the 2020 US election, the rise of deepfake videos became a major concern in the media. With the proliferation of fake news, there was a growing worry that people could no longer trust what they see online. In response to this challenge, Facebook and Instagram introduced a new policy in January 2020 to ban AI-manipulated "deepfake" videos that could mislead viewers during the election. Deepfakes are a form of synthetic media where an individual's likeness is replaced with someone else's in an existing image or video. The rapid evolution of deepfakes has prompted both the academic community and the technology industry to place significant emphasis on automatically detecting deepfake videos. As deepfakes are increasingly used for creating various forms of fake content, including celebrity pornography and fake news, there is a pressing need for effective detection methods.Deepfake technology is heavily used in the creation of adult content, with thousands of deepfake videos found on pornographic platforms. Additionally, new platforms dedicated to distributing deepfake pornography emerged.Deepfake learning models represent a critical aspect of combating the proliferation of digitally manipulated multimedia content.

Several prominent models and approaches have emerged in this domain:Variational Autoencoders (VAEs): VAEs are used to encode and decode visual content. They can be employed for deepfake detection by identifying anomalies in the encoding-decoding process, as deepfakes often struggle to maintain consistency.Convolutional Neural Networks (CNNs): Finds extensive application in the realm of image and video scrutiny. They can be employed to detect inconsistencies, artifacts, or irregularities in deepfake media by learning patterns indicative of manipulation.Recurrent Neural Networks (RNNs): RNNs are effective for sequential data analysis, making them valuable for video-based deepfake detection. They can capture temporal irregularities and anomalies in manipulated videos.
Generative Adversarial Networks (GANs): GANs, which are the foundation of many deepfake generation methods, can also be used for detection. By training a GAN to identify genuine content, discrepancies in the generated deepfake can be detected.

Capsule Networks (CapsNets): CapsNets are capable of capturing hierarchical relationships within images. They can be employed to discern anomalies or misalignments in deepfake images.Lip-sync Detection Models: Lip-sync models are trained to detect inconsistencies between audio and video in deepfake videos, as deepfakes may have difficulty synchronizing lip movements with spoken words.
Hybrid Models: Combining multiple deep learning models, such as a combination of CNNs, RNNs, and GANs, can provide a more comprehensive approach to deepfake detection, leveraging the strengths of each model type.

Siamese Networks: Siamese networks are used to compare and contrast two inputs, making them suitable for one-shot deepfake detection, where a known genuine reference is compared with the input.
Feature-Based Models: These models extract specific features from multimedia content and analyze them for irregularities or discrepancies, such as variations in eye colour, blinking patterns, or facial landmarks.
Capsule Networks (CapsNets): CapsNets have shown promise in identifying anomalies in deepfake images by capturing the hierarchical structures in images.
Meta-Learning Approaches: Meta-learning leverages a database of known deepfake and genuine content to adapt detection models to new, unseen deepfakes.

The effectiveness of these deepfake learning models often relies on the quality and diversity of the training data, the robustness of the model to various manipulation techniques, and its ability to generalize to different types of deepfake content. Ongoing research and collaboration among experts in machine learning and digital forensics are crucial for developing increasingly sophisticated and reliable deepfake detection models.

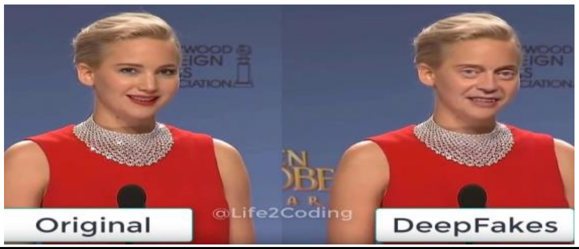

Fig.1 Example of real and deepfake.

1.1 Problem Statement

The problem at hand is the proliferation of deepfake content, which poses a substantial threat to the authenticity and trustworthiness of multimedia in various domains. Detecting these convincingly forged media using advanced machine learning techniques and ensuring their reliable identification are key challenges to mitigate the impact of deepfakes.

1.2 Objectives

The objectives of this Deepfake Detection using Deep Learning are as follows:
1. To accurately identify deepfakes while minimizing false positives and false negatives.
2. To work across different modalities, such as images, videos, and audio.
3. To operate in real-time or near real-time to prevent the spread of potentially harmful content.
4. To promote public awareness and education about deepfakes and how to identify them.
5. To ensure that the detection process respects privacy rights and doesn't violate users' privacy.

II.METHODOLOGY

Pre-processing Module Steps:
Frame Capture: The input video is split into frames. It is done by using OpenCV. Given that the project

uses single images as input, inter-frame information is not required, as it does not significantly contribute to the model's efficiency.

Face Detection: In each frame, faces are detected and labelled using the cascade classifier provided by OpenCV. The Haarcascadefrontalfacealtclassifier is chosen for its accuracy in detecting the face area. To mitigate non-face selection issues, only the largest detected face area is retained.

Saving Face Areas: The detected face areas are saved as new images. Before storing, these facial images are resized uniformly to meet the input size requirements of the deep learning models. This pre-processing module is crucial for converting video data into a format suitable for deep learning models that can then be used for deepfake detection. It addresses the challenge of working with video data when the models typically require images as input

The approaches to discerning the authenticity of digital media content in the context of deepfake detection encompass a diverse array of techniques and strategies. These methodologies are rooted in cutting-edge technologies and computational methodologies designed to distinguish genuine content from artificially generated counterparts.

These strategies often involve the utilization of advanced machine learning and deep learning models, which are trained on extensive datasets to learn patterns and anomalies associated with deepfake content. By scrutinizing visual, auditory, and even contextual cues, these models can identify inconsistencies, imperfections, or telltale signs that distinguish deepfakes from legitimate content.

Diagram:

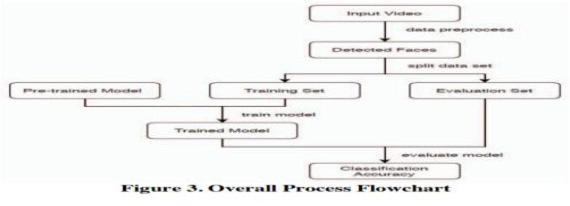

Figure 3. Overall Process Flowchart

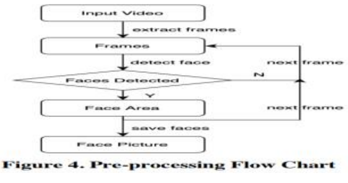

Figure 4. Pre-processing Flow Chart

III. LITERATURE REVIEW

1.Deepfakes Detection Methods(2021, 10thInternational Conference on Information and Automation for Sustainability (ICIAfS),Negambo, Sri Lanka)

Authors: M. Weerawardana and T. FernandoThis study, authored by M. Weerawardana and T. Fernando, underscores the growing urgency of addressing the Deepfake phenomenon, which has the potential to undermine trust in digital media and inflict harm in various aspects of society. The review of existing Deepfake detection methods highlights the inadequacy of current solutions in effectively combating the proliferation of these deceptive videos. Notably, the research emphasizes the prominent role of deep learning technologies, which have demonstrated superior performance in identifying Deepfakes compared to traditional methods. The paper also sheds light on the ongoing challenges in this domain, notably the scarcity of highly accurate and fully automated Deepfake detection solutions.

2.Deepfake Detection: A Systematic Literature Review (2022 in IEEE Access)

Authors: M. S. Rana, M. N. Nobi, B. Murali and A. H. SungTo gain a comprehensive understanding of the landscape of research in Deepfake detection, this paper conducts a systematic literature review, encompassing 112 pertinent articles published between 2018 and 2020. These articles present a diverse array of methodologies designed to address the challenges posed by Deepfakes. Our analysis categorizes these methods into four distinct groups: deep learning-based techniques, classical machine learning-based approaches, statistical methodologies, and blockchain-based solutions. Furthermore, we assess the effectiveness of these various detection methods across different datasets, and our findings underscore the superior performance of deep learning-based approaches in the realm of Deepfake detection.

In essence, this paper offers a holistic perspective on the evolving field of Deepfake detection, serving as a valuable resource for researchers and practitioners. It highlights the critical need to stay ahead of the ever-evolving Deepfake threat and reinforces the prominence of deep learning as a robust defense against the proliferation of convincing counterfeit multimedia content.

Analysis of Deepfake Detection Techniques(2023 International Conference on Circuit Power and Computing Technologies(ICCPCT), Kollam, India)

Authors: B. Puri, J. Kumar, S. Mukherjee and B. S. V

This study endeavours to delve into the multifarious techniques employed for the detection of deepfakes and assess their effectiveness in discerning manipulated content. It serves as a clarion call for sustained innovation within this domain, emphasizing the imperative of staying one step ahead in the ongoing battle against the proliferation of deceptive deepfake content. Ultimately, the aim of our research endeavour is to contribute to the collective efforts aimed at mitigating the dissemination of deepfakes and fostering the creation of trustworthy media content that faithfully mirrors reality.

Deepfake Detection: Current Challenges and Next Steps (2020)
Authors: Lyu, Siwei
In this research paper, the authors Lyu and Siweithe continual and iterativetraining of AI models on vast datasets enables the generation of material that closely emulates human craftsmanship. This convergence of AI-generated content with human-originated creations poses a formidable challenge, as it blurs the lines between synthetic and genuine content, leaving humans vulnerable to a host of predicaments, including framing, fraudulent activities, and political manipulation, among others.

Deepfake Detection through Deep Learning (2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK)
Authors: D. Pan, L. Sun, R. Wang, X. Zhang and R. O. SinnottThis paper specifically explores two deepfake detection technologies, namely Xception and MobileNet, within the context of classification tasks designed to automatically identify deepfake videos. To rigorously evaluate these methods, the research leverages training and assessment datasets derived from FaceForensics++, encompassing datasets generated through four distinct and prevalent deepfake technologies. The findings demonstrate a remarkable level of accuracy across all datasets, with detection rates ranging from 91% to 98%, contingent upon the particular deepfake technologies under scrutiny. Moreover, the paper introduces an innovative voting mechanism that extends beyond a single detection method, capitalizing on the aggregation of all four techniques. This research significantly advances our capabilities in countering the proliferation of deepfake technology by harnessing the power of deep learning.

IV. CONCLUSION

In conclusion, deepfake detection methodology contribute to the ongoing efforts to combat the proliferation of manipulated multimedia content, safeguarding the integrity of digital media in a rapidly evolving technological landscape.
The effectiveness of these deepfake detection models depends on various factors, including the quality and diversity of the training data, the robustness of the model to different manipulation techniques, and its ability to generalize to various types of deepfake content. Ongoing research and collaboration in the fields of machine learning and digital forensics are crucial for the development of increasingly sophisticated and reliable deepfake detection systems.

REFERENCES

1. Analysis of Deepfake Detection Techniques | IEEE Conference Publication | IEEE Xplore
2. Deepfake Detection: A Systematic Literature Review | IEEE Journals & Magazine | IEEE Xplore
3. Deepfake Detection: Current Challenges and Next Steps (researchgate.net)
4. Deep Fake Generation and Detection: Issues, Challenges, and Solutions | IEEE Journals & Magazine | IEEE Xplore
5. Deepfakes Detection Methods: A Literature Survey | IEEE Conference Publication | IEEE Xplore
6. Deepfake Detection through Deep Learning | IEEE Conference Publication | IEEE Xplore