IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

DeepFake Detection Using Machine Learning

[1]Prof.Rakhi Punwatkar , [2]Shubhangi Wadibhasme

Assistant Professor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune
PG Student, Department of Computer Engineering, Zeal College of Engineering and Research, Pune

ABSTRACT

Machine Learning and Deep learning-based software tools has facilitated the creation of credible face exchanges in videos and images that leave few traces of manipulation, in what they are known as "DeepFake"(DF) videos. Manipulations of digital videos have been demonstrated for several decades through the good use of visual effects, recent advances in deep learning have led to a drastic increase in the realism of fake content and the accessibility in which it can be created. These so-called AI-synthesized media (popularly referred to as DF).Creating the DF using the artificially intelligent tools are simple task. But, when it comes to detection of these DF, it is major challenge. Because training the algorithm to spot the DF is not simple. We have taken a step forward in detecting the DF using Convolutional Neural Network and Recurrent neural Network. System uses a convolutional Neural network (CNN) to extract features at the frame level. These features are used to train a recurrent neural network (RNN) which learns to classify if a video has been subject to manipulation or not and able to detect the temporal inconsistencies between frames introduced by the DF creation tools. Expected result against a large set of fake videos collected from standard dataset.

Keywords: Deepfake Detection , Kaggle Dataset, convolutional Neural network (CNN), recurrent neural network (RNN), Machine Learning.

I. INTRODUCTION

The increasing sophistication of smartphone cameras and the availability of good internet connection all over the world has increased the ever-growing reach of social media and media sharing portals have made the creation and transmission of digital videos more easy than ever before[2][5]. The growing computational power has made deep learning so powerful that would have been thought impossible only a handful of years ago. Like any transformative technology, this has created new challenges. So-called "DeepFake" produced by deep generative adversarial models that can manipulate video and audio clips.
Spreading of the DF over the social media platforms have become very common leading to spamming and peculating wrong information over the platform[1]. These types of the DF will be terrible, and lead to threating, misleading of common people.To overcome such a situation, DF detection is very important[4]. So, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos (DF Videos) from real videos. It's incredibly important to develop technology that can spot fakes, so that the DF can be identified and prevented from spreading over the internet[3]. The backbone of DF are deep adversarial neural networks trained on face images and target videos to automatically map the faces and facial expressions of the source to the target. A new deep learning-based method that can effectively distinguish DF videos from the real ones.. This warping leaves some distinguishable artifacts in the output deepfake video due to the resolution inconsistency between warped face area and surrounding context. It detects such artifacts by comparing the generated face areas and their surrounding regions by splitting the video into frames and extracting the features with a Convolutional Neural Network (CNN) and using the Recurrent Neural Network

(RNN) with Long Short Term Memory(LSTM) capture the temporal inconsistencies between frames during the reconstruction of the DF[6]. To train the CNN model, It simplify the process by simulating the resolution inconsistency in affine face wrappings directly.

Figure1.Analysing of Frames

II. LITERATURE SURVEY

Author	Name of the paper	Objective	Methodology	Limitation
Aarti Karandikar [2024]	Deepfake video Detection Using Convolutional Neural Network	Develop a machine learning model to detect deepfakes by identifying subtle artifacts in manipulated media.	Used CNN, real and fake datasets, preprocessing, and evaluated metrics like accuracy, precision, recall.	Challenges with generalization to unseen deepfakes, performance issues with advanced tools, and real-time detection efficiency.
Aparna Pandey [2024]	Deepfake Detection Using LSTM-Based Recurrent Neural Networks	To detect and classify deepfake videos as real or fake by using AI against AI.	Used ResNext CNN for feature extraction and LSTM-based RNN for video classification.	Limited by dataset, doesn't address real-world scenarios, and overfits on small datasets.
Aparna Bagde [2023]	Deepfake Detection using Deep Learning	Identify deepfakes and support multimedia authenticity.	Preprocessed input videos, used techniques like CNNs, RNNs, and GANs.	Focus on face-based deepfakes only, no real-time implementation, privacy concerns not addressed.
Alakananda Mitra [2021]	A ML based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction	Address challenges in detecting compressed socialmedia deepfakes.	Lightweight techniques for edge devices using partial DFDC datasets.	Limited generalizability to new datasets.

III. OBJECTIVES

- To develop an automated, accurate, and scalable system that detects deepfake media across various formats such as images, videos, and audio in real-time.
- To develop a machine learning model to detect deepfakes effectively and efficiently.
- This system aims to enhance the accuracy and robustness of deepfake detection, ensuring it can handle various types of manipulated content and adapt to new, emerging techniques.
- The system will use machine learning techniques like computer vision, CNNs, and RNNs to detect inconsistencies in images and videos, such as unnatural facial expressions, lighting issues, and irregular motion.

IV. PROPOSED

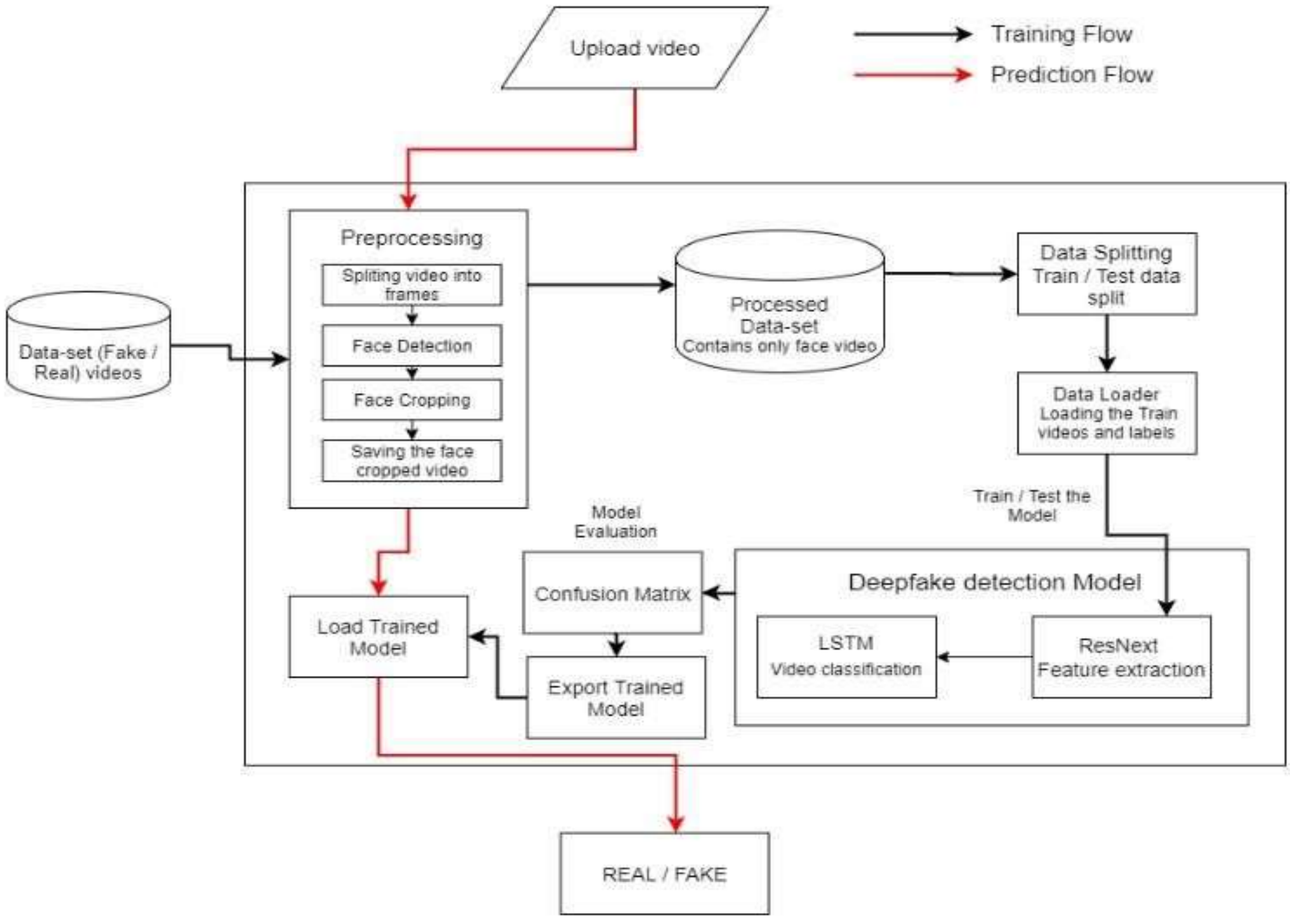

Figure1.System Architecture

V. PROPOSED METHODOLOGY

• Data collection: It involves gathering a diverse dataset of both real and fake media samples, including images, videos where the Real content is sourced from various sources.

• Data preprocessing: It is essential to standardize formats, align facial landmarks, and remove inconsistencies, ensuring the dataset is well-prepared for training.

• Model training: Training takes place on the prepared dataset using the chosen deep learning model, which is usually a Convolutional Neural Network (CNN). Data augmentation techniques are applied to enhance the dataset's diversity and the model's robustness. This phase is critical for teaching the model to recognize patterns and anomalies associated with deepfake content.

• Performance evaluation: Using known metrics like accuracy, precision, and recall, the model's performance is assessed once it has been trained. These metrics give us information about how well the algorithm works to differentiate between real and deepfake content. The goal is to maximize the accuracy of detection while minimizing false positives and false negatives.

VI. VALIDATION OF MODEL

The common performance evaluation metrics for validation of models include:

Accuracy: - It is the proportion of the total number of predictions that were correct and can be calculated from the following equation:

Accuracy =Tx+Ty Tx+Fx+Ty+Fy

Where, Tx= True Positives, Fx= False Positives, Ty= True Negatives, Fy= False Negatives

Recall: - It is defined as the percentage of total relevant results correctly classified by the algorithm.

Recall=Tx/Tx+Fx

Precision: - refers to the percentage of the results which are relevant.

Precision=Tx/Tx+Ty

F-statistics:-It is a metric that combines precision and recall and is calculated as the harmonic mean of precision and recall.

Fn=(1+n^2) *precision*recall/((n^2precision) +recall)

VII. FUTURE SCOPE

- There is always a scope for enhancements in any developed system, especially when the project build using latest trending technology and has a good scope in future.
- Web based platform can be up scaled to a browser plugin for ease of access to the user.
- Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deepfakes.
- The model can be extended to detect artificial audio and then be combined with an image processing module.

- Various other transfer learning models can be used to increase accuracy and to be able to classify the data correctly

- It is observed that low-quality images and images of larger size are giving predictions with lower accuracy; this can be rectified by training the models with more epochs for better accuracy.

VIII. CONCLUSION

In conclusion, present a neural network-based approach to classify the image as deep fake, or real, along with the confidence of the proposed model. The rapid advancement of deep learning models, neural networks, and machine learning algorithms has brought us closer to the goal of identifying and mitigating the impact of deceptive content. Deepfake detection methods have made substantial progress in recent years, offering hope in the battle against misleading and manipulated media. As continually refine these approaches and explore new avenues that are actively strengthening our defenses against the threat posed by deepfakes. The scheduled method is capable of detecting the image as a deep fake or real based on the dataset parameter. It will provide a very high accuracy of real-time data.

REFERENCES

1] Manish, Manish, Sai Kiran Reddy ‘‘DEEPFAKE DETECTION ON FACE IMAGES & VIDEOS USING DEEP LEARNING’’ International Journal of Creative Research Thoughts 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882

2] Aparna Pandey, Ruchi Soni, Nitin Kumar Sahu, Vanshaj H. Bawane, Dennis Marc Zaman ‘‘Deepfake Detection and Prevention” International Journal of Research Publication and Reviews, Vol 5, no 1, pp 1855- 1857 January 2024

3] Manumitha G P, Pragathi V, Shreya S S , Swetha B T , Rudresh N C ‘‘DEEPFAKE DETECTION USING MACHINE LEARNING’’ International Journal Trendy Research in Engineering and TechnologyVolume 8 Proceedings of NCCDS-24 ISSN NO 2582-0958

4] Soni Ragho, Onkar Divekar, Tushar Somwanshi, Suraj Salgar,Sarthak Shinde ‘‘Review On Deepfake Face DetectionUsing Machine Learning’’ © 2024 IJCRT | Volume 12, Issue 11 November 2024 | ISSN: 2320-2882

5] Kakarla Sai Priyanka,Gajawada Arihant,Akhandam Satyaditya,Yenduru Lakshmi varshitha,Dr Sanyasi Naidu Pasala “ DeepFake Detection’’ © 2023 JETIR April 2023, Volume 10, Issue 4

6] AndrewH.sung , Md Shohel Rana ‘‘Deepfake Detection Using Machine Learning Algorithms” 2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI) 978-1-6654-2420-2/21©2021 IEEE DOI 10.1109/IIAI-AAI 53430.2021 00079

7] Abhijit Jadhav, Abhishek Patange,Jay Patel, Hitendra Patil, Manjushri Mahajan ‘‘Deepfake Video Detection using Neural Networks ” International Journal for Scientific Research & Development| Vol. 8, Issue 1, 2020

8] Alakananda Mitra , Saraju P. Mohanty,Peter Corcoran ,Elias Kougianos “A Machine Learning based Approach for DeepFake Detection in Social Media through Key Video Frame Extraction’’ DOI: 10.1007/s42979-021-00495-x

9] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies,Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" in arXiv:1901.08971.

10] Deepfake detection challenge dataset : https://www.kaggle.com/c/deepfake-detection challenge/data Accessed on 26 March, 2020

11] 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]

12] Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. Mesonet: A compact facial video forgery detection network. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7. [Google Scholar]

13] Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming auto-encoders. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011; Proceedings, Part I 21. Springer: Berlin/Heidelberg, Germany, 2011; pp. 44–51. [Google Scholar]

14] Nguyen, H.H.; Yamagishi, J.; Echizen, I. Capsule-forensics: Using capsule networks to detect forged images and videos. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2307–2311. [Google Scholar]

15] Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Repubic of Korea, 27 October–2 November 2019; pp. 1–11. [Google Scholar]