International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

DEEPFAKE DETECTION USING DEEP LEARNING (CNN+LSTM)

Mohd Salim Shaikh[1], Lucky Nirankari[2], Vasant Pardeshi[3], Rupesh Sharma[4] Prof. Sunil Kale[5]

[1,2,3,4]Student [5]Professor [1,2,3,4,5]Department of Computer Engineering

[1,2,3,4,5]Sandip Institute of Technology and Research Center®, Nashik, India

---

Abstract - Artificial intelligence advancements have led to the development of deepfake technology, which seriously jeopardises the integrity of visual media material. Robust detection algorithms are becoming more and more necessary as deepfake creation techniques become more complex. This study combines Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs) to present a novel method for deepfake identification. The suggested CNN+LSTM architecture makes use of LSTMs' temporal modelling capabilities and CNNs' spatial feature extraction capabilities. While the LSTM component analyses the temporal connections between frames to identify patterns suggestive of deepfake manipulation, the CNN component concentrates on capturing local features and patterns in individual frames. The combination of these two networks improves the model's capacity to identify minute anomalies and inconsistencies that are indicative of deepfake content. To extract frame-level characteristics, we use Res-Next Convolutional Neural Networks. These attributes are then used to train a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) to determine whether a video has been manipulated, i.e., whether it is a deepfake or a genuine video. We intend to train our deepfake detection model on a varied set of public datasets in order to improve its real-time performance. We improve the model's adaptability by learning features from different photos. Face-Forensic++, Deepfake Detection Challenge, and Celeb-DF datasets are used to extract videos. Furthermore, to assure competitive performance in real-world scenarios, our model will be assessed against a large amount of real-time data, including the YouTube dataset.

Key Words: Temporal modelling, Deepfake technology
, Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs)

1. INTRODUCTION

Deepfake technology involves the use of neural network techniques such as GANs (Generative Adversarial Networks) or Auto Encoders to synthesize human images. These techniques make surprisingly convincing deepfake films by layering target pictures onto source videos, which can be difficult to distinguish from genuine ones with the naked eye. In this paper, we describe a novel deep learning strategy for distinguishing AI-generated bogus videos from legitimate ones. We leverage discernible artefacts left in frames during the deepfake generation process by taking advantage of the limitations of current deepfake creation tools. While these artefacts are invisible to humans, trained neural networks can detect them.

To extract frame-level characteristics, we use Res-Next Convolutional Neural Networks. These attributes are then used to train a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) to determine whether a video has been manipulated, i.e., whether it is a deepfake or a genuine video. We intend to train our deepfake detection model on a varied set of public datasets in order to improve its real-time performance. By learning features from various photos, we boost the model's versatility. Face-Forensic++, Deepfake Detection Challenge, and Celeb-DF datasets are used to extract videos. Furthermore, to assure competitive performance in real-world scenarios, our model will be assessed against a large amount of real-time data, including the YouTube dataset.

2. LITERATURE SURVEY

"Faceforensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces" by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, from the Technical University of Munich, University Federico II of Naples, and University of Erlangen-Nuremberg in March 2018, presents a significant contribution to the field of forgery detection in video content. [1]

The "Deepfake Detection Challenge (DFDC) Preview Dataset" paper, authored by Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer from the AI Red Team at Facebook AI, was published in October 2019. The paper introduces a dataset created for the Deepfake Detection Challenge, aiming to spur advancements in deepfake detection technologies. The dataset includes manipulated videos with facial deepfakes, providing a preview for researchers and participants in the challenge. [2]

"Celeb-DF: A Large-Scale Challenging Dataset for Deepfake Forensics" was authored by Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu in March 2020. The research was conducted at the University at Albany, State University of New York, USA, and the University of Chinese Academy of Sciences, China. [3]

"An Improved Dense CNN Architecture for Deepfake Image Detection," authored by Yogesh Patel, Sudeep Tanwar (Senior Member, IEEE), Innocent Ewean Davidson (Senior Member, IEEE), and Thokozile F. Mazibuko, and published in March 2023, introduces an advanced convolutional neural network (CNN) architecture tailored for detecting deepfake images. The authors leverage their expertise to propose enhancements, addressing contemporary challenges in deepfake detection. [4]

"Exposing Deepfake Videos by Detecting Face Warping Artifacts," authored by Yuezun Li and Siwei Lyu from the University at Albany, State University of New York (May

2019), focuses on detecting face warping artifacts as a means of unveiling deepfake videos. The authors contribute to the field of deepfake detection by proposing methods to identify distortions in facial features. This research addresses the challenges of early deepfake detection, offering insights into the limitations of existing methods. [5]

"DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," authored by Tackhyun Jung, Sangwon Kim, and Keecheon Kim in 2020, likely explores a novel approach to deepfake detection. The focus appears to be on leveraging the human eye blinking pattern as a distinctive feature for identifying deepfake videos. The authors may present a method that utilizes deep learning techniques to analyze and detect anomalies in eye blinking patterns, aiming to distinguish authentic content from artificially generated deepfake videos. This innovative approach could contribute to the advancement of reliable and effective methods for identifying manipulated content through the analysis of subtle physiological cues. [6]

"Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos," authored by Latha M S, Abdul Samad, Alekhya B, Harshitha M, and Ms. Rakshitha P in June 2022, likely presents a novel approach to detecting manipulated or forged images and videos. The focus appears to be on utilizing Capsule Networks, a type of neural network architecture, for the purpose of forensic analysis. [7]

"Deep Fake Video Detection Using Recurrent Neural Networks" by Abdul Jamsheed V. and Janet B., published in April 2021, likely focuses on employing Recurrent Neural Networks (RNNs) for the purpose of detecting deep fake videos. The primary objective appears to be the development and application of a model that utilizes the temporal information in videos, a characteristic addressed by RNNs, to identify manipulated content. The authors likely discuss the methodology involved in training the Recurrent Neural Network for deep fake detection, emphasizing the importance of considering sequential patterns and temporal dependencies within video data. The paper may also present an evaluation of the proposed model's performance, possibly using relevant datasets containing both authentic and deep fake videos. [8]

"FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals," authored by Umur Aybars Ciftci, ˙ilke Demir, and Lijun Yin, and published in July 2020, likely presents a method for detecting synthetic or fake portrait videos. The emphasis appears to be on leveraging biological signals as a means of identifying manipulated content. [9]

"Deepfake Detection: A Systematic Literature Review," authored by Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung, and published in March 2022, likely provides a comprehensive review of existing literature on the topic of deepfake detection. The authors likely systematically analyze and summarize various methods, techniques, and advancements in the field of detecting deepfake content. [10]

3. PROPOSED SYSTEM

Numerous tools facilitate the creation of Deepfakes (DF), yet the availability of tools for DF detection is limited. Our innovative approach to DF detection is poised to make a significant contribution, preventing the proliferation of DF across the World Wide Web. We are developing a web-based platform enabling users to upload videos and classify them as authentic or manipulated. This project has the potential for scalability, evolving from a web-based platform to a browser plugin for seamless automatic DF detection.

Furthermore, major applications such as WhatsApp and Facebook could seamlessly integrate our project for preemptive DF detection before sharing with other users. A key objective is to assess its performance and acceptance based on criteria such as security, user-friendliness, accuracy, and reliability. Our method is designed to detect various DF types, including replacement DF, retrenchment DF, and interpersonal DF. Figure 1 illustrates the straightforward system architecture of the proposed system.

In addition to providing a robust web-based platform and potential integration into major applications like WhatsApp and Facebook, our project envisions scalability by evolving into a browser plugin. This plugin would empower users with automatic DF detection capabilities, ensuring a more streamlined and proactive approach to preventing the spread of manipulated content. The proposed system's architecture, as depicted in Figure 1, illustrates its simplicity, emphasizing user accessibility and ease of use.

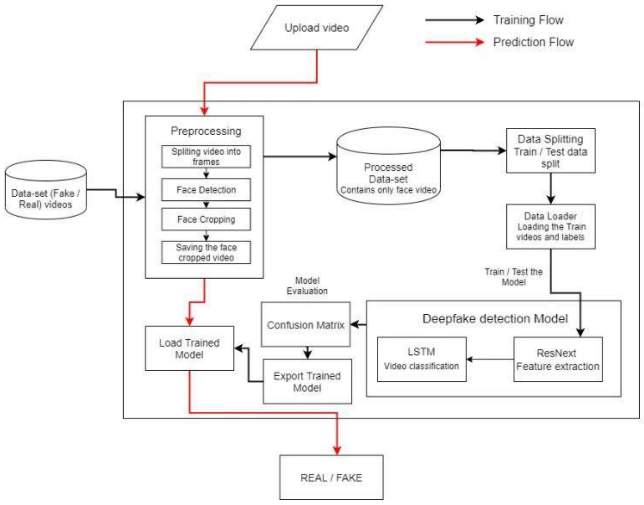

Fig 3.1: System architecture

Moreover, our focus extends beyond mere technical functionality. We are committed to evaluating the system's performance and acceptability on multiple fronts, including security measures to safeguard against evolving threats, user-friendliness to encourage widespread adoption, accuracy in distinguishing between authentic and manipulated content, and reliability in providing consistent results. By addressing these essential criteria, we aim to develop a comprehensive solution that not only advances the field of DF detection but also aligns

with the expectations and needs of end-users across different platforms and applications.

1. Dataset:

We are using a mixed dataset which consists of equal amount of videos from different dataset sources like YouTube, FaceForensics++, Deep fake detection challenge dataset. Our newly prepared dataset contains 50% of the original video and 50% of the manipulated deepfake videos. The dataset is split into 70% train and 30% test set.

2. Preprocessing:

Dataset preprocessing includes the splitting the video into frames. Followed by the face detection and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the new processed face cropped dataset is created containing the frames equal to the mean. The frames that doesn't have faces in it are ignored during preprocessing. As processing the 10 second video at 30 frames per second i.e total 300 frames will require a lot of computational power. So for experimental purpose we are proposing to used only first 100 frames for training the model.

3. Model:

The model consists of resnext50_32x4d followed by one LSTM layer. The Data Loader loads the preprocessed face cropped videos and split the videos into train and test set. Further the frames from the processed videos are passed to the model for training and testing in mini batches.

4. ResNext CNN for Feature Extraction:

Instead of writing the rewriting the classifier, we are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the gradient descent of the model. The 2048-dimensional feature vectors after the last pooling layers are then used as the sequential LSTM input.

5. LSTM for Sequence Processing:

Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. The key challenge that we need to address is the de- sign of a model to recursively process a Deepfake Video Detection using Neural Networks sequence in a meaningful manner.

6. Predict:

A new video is passed to the trained model for prediction. A new video is also preprocessed to bring in the format of the trained model. The video is split into frames followed by face cropping and instead of storing the video into local storage the cropped frames are directly passed to the trained model for detection.

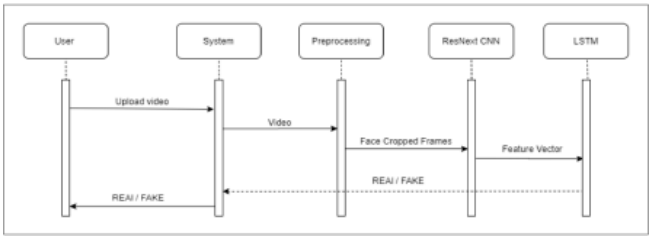

Fig 3.2: Sequence Diagram

4. FUTURE SCOPE

Fighting the spread of manipulated media requires integrating deepfake detection into popular platforms like Facebook and WhatsApp. Using a Deepfake (DF) Detection system to perform a preliminary scan of videos prior to uploading is part of this proactive approach. The content is quickly classified by the system as "Real" or "DF" (meaning it's a DeepFake). A video is quickly flagged and blocked from being uploaded if it is determined to be fraudulent.

This integration has a number of noteworthy benefits. First off, it offers protection in real time by promptly detecting and stopping the spread of potentially dangerous deepfake content. By preventing users from coming across misleading media, this protection preserves their faith in the platform.

Legally, this integration can establish a robust defense against the dissemination of fraudulent content. It enables legal action to be taken against those who create or distribute malicious deepfakes, reinforcing accountability in the digital space. Integrating a deepfake detection system into widely used applications offers real-time defense, user education, adaptive learning, privacy protection, and legal support. This implementation can serve as a crucial step towards ensuring the responsible use of technology in the digital age.

5. CONCLUSIONS

We presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. Our method is capable of predicting the output by processing 1 second of video (10 frames per second) with a good accuracy. We implement the model by using pre-trained ResNext CNN model to extract the frame
level features and LSTM for temporal sequence processing to spot the changes between the t and t-1 frame. Our model can process the video in the frame sequence of 10,20,40,60 ,80,100.

The deepfake detection method that is being presented makes use of a neural network architecture and is centred on identifying videos as deepfake or real, along with a model-derived confidence measure. With only one second of video footage—roughly the equivalent of processing ten frames per second—the method demonstrates the ability to make accurate predictions.

ACKNOWLEDGEMENT

First and foremost, we wish to record our sincere gratitude to the Management of this college and our Respected Principal **Prof. (Dr) M. M. Patil**.

Our sincere thanks to **Prof. (Dr) Ankita V. Karale**, Head, Department of Computer, Sandip Institute of Technology and Research Centre, Nashik.

We express our sincere gratitude to our guide, **Prof. Sunil Kale** for guiding us in the investigations of this project and in carrying out experimental work.

REFERENCES

1. Faceforensics: A Large-Scale Video Dataset For Forgery Detection In Human Faces Andreas Rossler ¨ 1 Davide Cozzolino2 Luisa Verdoliva2 Christian Riess3 Justus Thies1 Matthias Nießner1 1technical University Of Munich 2university Federico Ii Of Naples 3university Of Erlangen-Nuremberg -- Mar 2018

2. The Deepfake Detection Challenge (Dfdc) Preview Dataset Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, Cristian Canton Ferrer Ai Red Team, Facebook Ai -- Oct 2019

3. Celeb-Df: A Large-Scale Challenging Dataset For Deepfake Forensics Yuezun Li1 , Xin Yang1 , Pu Sun2 , Honggang Qi2 And Siwei Lyu1 1 University At Albany, State University Of New York, Usa 2 University Of Chinese Academy Of Sciences, China – Mar 2020

4. An Improved Dense Cnn Architecture For Deepfake Image Detection Yogesh Patel 1 , Sudeep Tanwar 1 , (Senior Member, Ieee), Innocent Ewean Davidson4 , (Senior Member, Ieee), And Thokozile F. Mazibuko5 – Mar 2023

5. Exposing Deepfake Videos By Detecting Face Warping Artifacts Yuezun Li, Siwei Lyu Computer Science Department University At Albany, State University Of New York, Usa – May 2019

6. Deepvision: Deepfakes Detection Using Human Eye Blinking Pattern Tackhyun Jung1 , Sangwon Kim2 , And Keecheon Kim3 – 2020

7. Capsule-Forensics: Using Capsule Networks To Detect Forged Images And Videos Latha M S1 , Abdul Samad2 , Alekhya B3 , Harshitha M4 , Ms Rakshitha P5 – June 2022

8. Deep Fake Video Detection Using Recurrent Neural Networks Abdul Jamsheed V. 1* , Janet B. 2 – April 2021

9. Fakecatcher: Detection Of Synthetic Portrait Videos Using Biological Signals Umur Aybars Ciftci, ˙ilke Demir, And Lijun Yin, Senior Member, Ieee – July 2020

10. Deepfake Detection: A Systematic Literature Review Md Shohel Rana 1,2, (Member, Ieee), Mohammad Nur Nobi3 , (Member, Ieee), Beddhu Murali2 , And Andrew H. Sung2 , (Member, Ieee) – Mar 2022

11. Fighting Deepfake By Exposing The Convolutional Traces On Images Luca Guarnera 1,2, (Student Member, Ieee), Oliver Giudice 1 , And Sebastiano Battiato 1,2, (Senior Member, Ieee) – Sept 2020

12. Generalization Of Forgery Detection With Meta Deepfake Detection Model Van-Nhan Tran 1 , Seong-Geun Kwon2 , Suk-Hwan Lee3 , Hoanh-Su Le4 , And Ki-Ryong Kwon 1 – Jan 2023

13. Deep Detection For Face Manipulation Disheng Feng 1 , Xuequan Lu1 – Sept 2020

14. Deep Learning For Deepfakes Creation And Detection: A Survey Thanh Thi Nguyena , Quoc Viet Hung Nguyenb , Dung Tien Nguyena , Duc Thanh Quoc-Viet Phamf , Cuong M. Nguyeng – Aug 2022

15. Exposing Deep Fakes Using Inconsistent Head Poses Xin Yang? , Yuezun Li? And Siwei Lyu University At Albany, State University Of New York, Usa – Nov 2018

16. Coarse-To-Fine Copy-Move Forgery Detection For Video Forensics Shan Jia 1,2, Zhengquan Xu1,2, Hao Wang 1,2, Chunhui Feng3 , And Tao Wang1,2 – May 2018

17. Recurrent Convolutional Strategies For Face Manipulation Detection In Videos Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael Abdalmageed, Iacopo Masi, Prem Natarajan Usc Information Sciences Institute, Marina Del Rey, Ca, Usa – May 2019

18. Deepfakes And Beyond: A Survey Of Face Manipulation And Fake Detection Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales And Javier Ortega-Garcia Biometrics And Data Pattern Analytics - Bida Lab, Universidad Autonoma De Madrid, Spain – June 2020

19. Deepfakes: A New Threat To Face Recognition? Assessment And Detection Pavel Korshunov And Sebastien Marcel – Dec 2018

20. Mesonet: A Compact Facial Video Forgery Detection Network Darius Afchar Ecole Des Ponts Paristech ´ Marne-La-Vallee, France ´ -- Sept 2018