IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, May 2024

Face Manipulation Detection

M. Nagaraj[1], K Sri Pavan[2], Charan Narasimha M[3], Goudru Pandunaik[4], Ashu Mishra[5]

Assistant Professor, Department of Computer Science and Engineering[1]

Students, Department of Computer Science and Engineering[2,3,4,5]

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract*: The rise of digital manipulation techniques has led to the proliferation of deepfakes and other manipulated facial images, posing significant challenges to online trust and security. This paper proposes a novel deep learning model aimed at efficiently and effectively detecting face manipulations. The architecture combines the strengths of Efficient Net, a convolutional neural network (CNN) renowned for its accuracy and efficiency, with Long Short-Term Memory (LSTM) networks. Efficient Net is utilized for extracting high-level features from facial images, enabling the model to capture subtle inconsistencies that may indicate manipulation. These features serve as crucial inputs to the subsequent analysis performed by LSTM networks. LSTMs excel at capturing temporal dependencies within sequences of data, making them particularly well-suited for detecting manipulations in video sequences. By leveraging the power of CNNs for feature extraction and the sequential learning capabilities of LSTMs, the proposed hybrid approach aims to achieve superior performance in face manipulation detection. This combination allows the model to effectively analyse both spatial and temporal aspects of facial images, enhancing its ability to detect various forms of manipulation accurately*

Keywords: Face Manipulation, deep fakes, Efficient Net, LSTM

I. INTRODUCTION

In the modern digital landscape, the widespread availability of advanced image editing tools has raised concerns about the authenticity of visual content. Of particular interest is the manipulation of facial images, which can significantly impact domains such as media forensics, biometric authentication, and online identity verification. Detecting such manipulations is crucial for upholding trust and authenticity in visual information.

Recent advancements in deep learning have shown promise in automating the detection of manipulated images. In this study, we propose a novel method for detecting face manipulation by combining Efficient Net and Long Short-Term Memory (LSTM) networks. Efficient Net, known for its excellence in image classification tasks, serves as the backbone architecture for feature extraction. Its ability to balance model complexity and computational efficiency makes it well-suited for processing large-scale image datasets, aiding in capturing discriminative features from manipulated facial images effectively.

Complementing Efficient Net, we integrate LSTM networks to exploit temporal dependencies within sequences of facial features. Unlike traditional convolutional neural networks (CNNs) that treat input images as independent entities, LSTM networks excel in modelling sequential data by retaining memory of past observations. This enables our model to capture subtle temporal patterns indicative of image manipulation techniques.

The hybrid architecture proposed offers several advantages over existing methods. By leveraging both Efficient Net and LSTM networks, our model can effectively detect both spatial and temporal anomalies associated with manipulated facial images. Furthermore, the hybrid approach facilitates end-to-end learning, enabling the model to automatically adapt to diverse manipulation strategies without the need for handcrafted features. This paper provides a comprehensive overview of our methodology, including the design and implementation of the hybrid model. We evaluate the performance of our approach on benchmark datasets for face manipulation detection, demonstrating its efficacy in accurately identifying manipulated images across various scenarios.

II. LITRATURE SURVEY

"Fighting Deepfake by Exposing the Convolutional Traces on Images" By Luca Guarnera, Oliver Giudice, Sebastiano Battiato Year – 2020 :

The paper presents a pioneering strategy to counter the escalating threat of deepfake imagery by spotlighting convolutional traces embedded within such falsified content. Deepfake technology, powered by advanced deep learning

DOI: 10.48175/IJARSCT-18222

algorithms, poses a formidable challenge across numerous sectors including politics, media, and personal privacy, given its ability to fabricate convincingly realistic visuals. By directing attention to the convolutional imprints left behind during the generation process, the proposed method offers a fresh perspective on combating this pervasive issue. The research likely upholds a robust foundation of technical rigor and practical expertise in mathematics, computer science, and image processing. Successful identification and exposure of convolutional traces hold the potential to significantly enhance the development of more resilient deepfake detection systems, thereby fortifying defences against misinformation and manipulation in various contexts. However, challenges such as limited information accessibility, potential generalization constraints, ethical considerations, and practical adoption hurdles underscore the need for continued research, experimentation, and ethical scrutiny to effectively harness the promise of this innovative approach while navigating its associated complexities..

DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern By Tackhyun Jung, Sangwon Kim , And Keecheon Kim - 2020

The paper introduces "Deep Vision," a novel approach to detecting deepfake images by leveraging the distinctive blinking patterns of human eyes. Authored by Tackhyun Jung, Sangwon Kim, and Keecheon Kim from Konkuk University in South Korea, the research proposes an innovative method that capitalizes on a fundamental biological trait to discern manipulated visual content. Advantages of this approach include its reliance on a biometric marker that is difficult to fabricate artificially, potentially making it more resilient against sophisticated deepfake techniques. Additionally, the use of human eye behaviour adds a layer of authenticity to the detection process, aligning with natural human perception cues. However, there are also notable limitations to consider. For instance, the effectiveness of the proposed method may be contingent upon the quality and resolution of input images, potentially limiting its applicability in scenarios with low-quality or obscured visual data. Moreover, the reliance on a single biometric cue may introduce vulnerabilities to adversarial attacks or circumvention strategies developed by deepfake creators. Despite these challenges, "Deep Vision" represents a noteworthy advancement in the ongoing effort to combat the proliferation of deepfake content, offering a unique perspective that warrants further exploration and refinement in future research endeavours.

"Deepfake Detection using a Two-Stream Capsule Network" By : Z. Joseph and C. Nyirenda -2021

This paper, presented at the 2021 IST-Africa Conference, introduces a method for detecting deepfake images utilizing a Two-Stream Capsule Network. Authored by Z. Joseph and C. Nyirenda, the research proposes an innovative approach that employs capsule networks, a type of neural network architecture, to identify manipulated visual content. By utilizing two streams of information, the system aims to capture both spatial and temporal features, enhancing its ability to discern subtle inconsistencies characteristic of deepfake alterations. The method incorporates deep learning techniques and error level analysis to bolster its detection capabilities. Keywords associated with the paper include training, analytical models, computational modelling, neural networks, computer architecture, particle swarm optimization, and optimization. The focus on deepfake detection, face tampering, and the integration of capsule networks underscores the authors' commitment to exploring cutting-edge technologies in the ongoing fight against the proliferation of synthetic media. While the paper proposes an intriguing method for deepfake detection using a Two-Stream Capsule Network, several potential disadvantages warrant consideration. Firstly, the implementation of such a network introduces complexity to the detection process, demanding substantial computational resources and expertise in neural network architecture. This complexity could impede the method's scalability and accessibility, especially in contexts lacking sufficient computational infrastructure or specialized knowledge.

"Deepfake Video Detection Using Recurrent Neural Networks" By D. Guera and E. J. Delp - 2018

Presented at the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) in Auckland, New Zealand, this paper authored by D. Guera and E. J. Delp proposes a method for detecting deepfake videos using Recurrent Neural Networks (RNNs). The research aims to address the growing threat of manipulated video content by leveraging RNNs, a type of neural network architecture capable of processing sequential data. By training RNNs on features extracted from videos, the proposed approach seeks to identify patterns indicative of deepfake manipulation. Keywords associated with the paper include face, training, feature extraction, decoding, and streaming media, highlighting the key elements of the detection methodology. Through the integration of RNNs and specialized feature extraction techniques, the authors contribute to the advancement of deepfake detection technology, offering a potential solution to the challenges.

While the paper introduces a promising approach to deepfake detection using Recurrent Neural Networks (RNNs), several limitations should be considered. RNNs may struggle with capturing long-range dependencies in videos, potentially limiting their effectiveness in detecting subtle manipulations. Additionally, the method's performance may heavily rely on the quality and diversity of the training data, posing challenges in scenarios with limited access to representative datasets.

2.1 Proposed System

The proposed system aims to develop an advanced deepfake detection model using a combined (efficient Net)CNN+LSTM architecture. We will be providing a web-based platform for the user to upload the video and classify it as fake or real.

Dataset:

We are using a dataset Deep fake detection challenge dataset which was released by Facebook. This 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.

Model :

EfficientNet Backbone: The model architecture begins with an Efficient Net backbone, renowned for its efficacy in image classification tasks. We leverage a pre-trained Efficient Net model, such as EfficientNet-B0, EfficientNet-B1, etc., as the feature extractor. Transfer learning is employed to fine-tune the pre-trained Efficient Net on face manipulation detection datasets.

LSTM Integration: Following the Efficient Net backbone, an LSTM layer is introduced to capture temporal dependencies within sequences of facial features extracted from consecutive frames. This integration enables the model to learn dynamic patterns indicative of manipulation across time.

Prediction :

A fresh video undergoes preprocessing to conform to the format required by the trained model. Initially, the video is divided into individual frames, and facial detection algorithms are applied to crop and isolate facial regions from each frame. Unlike conventional methods that involve storing the entire video locally, only the cropped frames are directly fed into the trained model for analysis. Subsequently, the model assesses the input frames to determine whether the video depicts genuine content or is a deepfake. Alongside this determination, the model provides a confidence score indicating the reliability of its prediction. By adopting this streamlined approach, the system efficiently evaluates videos in real-time without the need for extensive local storage. The immediate assessment of individual frames facilitates rapid decision-making regarding the authenticity of the video, bolstered by the model's confidence assessment."

III. SYSTEM DESIGN

Deepfake detection systems analyse videos to determine if they've been manipulated. This system design tackles this challenge through several stages. First, users upload videos which are then split into individual frames. Faces are extracted from these frames in the training data. Next, a deep learning model called Net CNN analyses the frames to extract features. Finally, another deep learning model, a Long Short-Term Memory (LSTM) network, uses these extracted features to classify the video as real or fake. By training these deep learning models on a large dataset of real and manipulated videos, the system can identify patterns indicative of deepfakes in new videos.

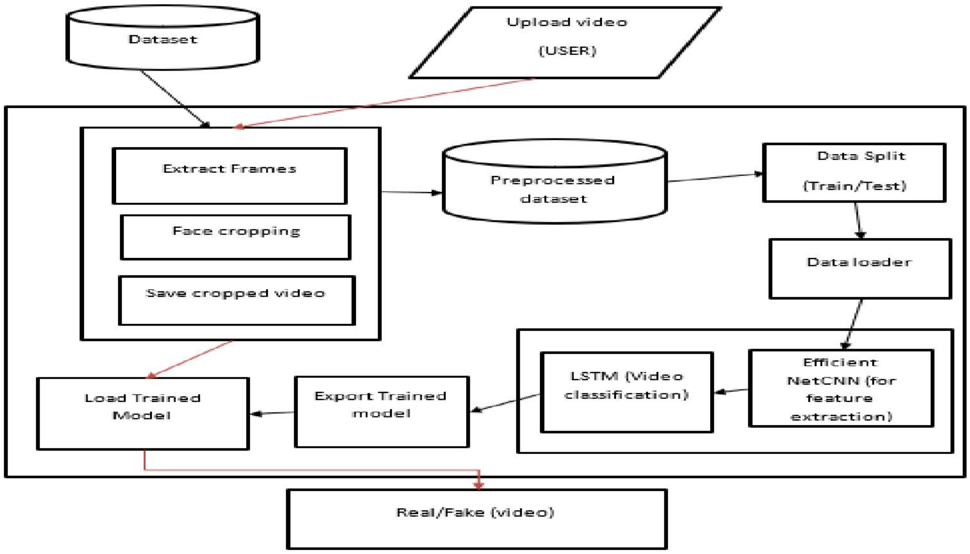

Fig.1.Overview of the System Design

IV. CONCLUSION

In conclusion, the development of a deepfake video detection system using the Efficient Net-LSTM hybrid model not only represents a technological advancement but also holds significant implications for the public. By effectively identifying manipulated videos, this system contributes to the preservation of trust and authenticity in digital content, thereby safeguarding individuals, organizations, and society as a whole from the potential harms of misinformation, manipulation, and exploitation. By deterring the spread of malicious deepfake content, the system helps to mitigate the social, political, and economic consequences associated with its dissemination, thereby contributing to a safer and more secure online environment for all. The development and deployment of deepfake detection systems not only serve the interests of individual users but also uphold the collective well-being and integrity of society at large.

REFERENCES

[1]. 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.

[2]. Al-Dulaimi, Omar Alfarouk Hadi Hasan, and Sefer Kurnaz. "A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning." *Electronics* 13.9 (2024): 1662.

[3]. Yavuzkiliç, Semih, et al. "DeepFake face video detection using hybrid deep residual networks and LSTM architecture." *AI and Deep Learning in Biometric Security*. CRC Press, 2021. 81-104.

[4]. Kharbat, Faten F., et al. "Image feature detectors for deepfake video detection." *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2019.

[5]. Malik, Asad, et al. "DeepFake detection for human face images and videos: A survey." *Ieee Access* 10 (2022): 18757-18775.

[6]. Guera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2018.

[7]. Suratkar, Shraddha, et al. "Deep-fake video detection approaches using convolutional–recurrent neural networks." *Journal of Control and Decision* 10.2 (2023): 198-214.

[8]. 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.

DOI: 10.48175/IJARSCT-18222

[9] 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.

[10 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.

DOI: 10.48175/IJARSCT-18222