Thejashwini M, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Deep fake Detection in Videos Using Deep Learning

Thejashwini M ,Thejaswini K P,Vismitha K,Vivek A M,Shylaja B
Dayananda Sagar Academy of Technology and Management

Abstract- Deep fake videos—synthetically manipulated visual content created by deep learning techniques—pose an alarming threat issues in sectors involving from media and politics to healthcare. This study aims to develop a robust deep fake detec- tion framework using deep learning methods and algorithms, focusing on generalization across datasets and video types. Our approach is informed by an extensive literature review of recent advancements, including methods using xception, Mobile Net, Mask R-CNN, and affective cue-based models. The review highlights challenges such as dataset diversity, generalization issues, and real-time detection limitations. Our proposed deep learning system will address these challenges by integrating temporal and spatial video analysis, leveraging convolutional and recurrent neural networks to detect subtle manipulative traces.

Keywords: Deep fake Detection, Video Manipulation, Deep Learning, Convolutional Neural Networks, Generalization.

I. INTRODUCTION

The project titled ”Deepfake Detection in Videos Using Deep Learning” focuses on identifying forged(fake) or synthet- ically altered videos created using generative models such as GANs. These deepfakes can manipulate a person’s appearance or speech or action in video form, causing potential harm in particular domains such as media, politics, law, and personal identity. Deep fake videos pose a threat due to their increasing realism and accessibility. Malicious (cyber threat) actors can easily generate fake videos that appear genuine, potentially spreading misinformation which may influence public opinion, or damaging reputations. We can see that the effect of artificial intelligence (AI) progress also does not lay only in the scientific research field, but goes out of this scope and makes it available for everyone to apply state of the art (SOTA) AI techniques for everyone over easy-to-use applications and social networks. This paper proposes a deep learning-based framework that can effectively detect manipulations in videos through spatial and temporal inconsistencies.

The approach leverages CNNs for frame-based feature extraction and RNNs

for temporal sequence analysis, aiming to achieve high accu- racy across varied datasets.

II. LITERATURE SURVEY

From past few years, research in deep fake video detection using deep learning has advanced significantly, addressing the growing sophistication of synthetic media generation. Liu et al. (2024) provided a comprehensive survey tracing the evolution from single-modal to multi-modal detection approaches, highlight the integration of facial, audio, and behavioral cues to improve robustness. Yu et al. (2021) catego- rized the existing video-based detection methods, emphasizing the use of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and mechanisms for capturing frame- wise and temporal features. Passos et al. (2023) focused on deep learning techniques including CNN, ResNet, and VGGNet architectures, while also discussing the challenges of dataset limitations, generalizability, and real-time detection. Coccomini et al. (2021)

combined EfficientNet and Vision Transformers to effectively process facial dynamics across video frames, reducing model complexity without sacrificing accuracy. De Lima et al. (2020) introduced spatiotemporal convolutional networks capable of detecting temporal inconsis- tencies that occur due to manipulation artifacts, outperforming many frame-level detectors. Tariq et al. (2020) designed a Convolutional LSTM-based residual network (CLRNet) that learns long-term dependencies in videos, significantly im- proving generalization across different deep fake generation techniques. Vyas et al. (2024) systematically analyzed the strengths and weaknesses of existing methods, drawing atten- tion to the lack of standard benchmarks and the vulnerability of models to adversarial attacks. Kamakshi Thai et al. (2024) emphasized the role of explainable AI and forensic signal analysis in enhancing the transparency and trustworthiness of detection systems. Collectively, these studies underscore the transition toward hybrid, multi- modal, and temporally-aware

deep learning models, while also identifying critical gaps such as dataset bias, cross-dataset generalization, real-time processing, and adversarial robustness that must be addressed in future research.

III. PROPOSED METHODOLOGY

The proposed system uses a hybrid deep learning architec- ture and deep learning algorithm to detect deep fake manipu- lations in videos. It is divided into multiple modules:

Frame Extraction and Preprocessing: Input video is de- composed into individual frames, followed by normalization, resizing, and artifact enhancement. - Spatial Feature Extrac- tion: A CNN model (e.g., XceptionNet or EfficientNet) is used to detect visual oddity in individual frames. - Temporal Analy- sis: RNNs (LSTM/GRU) or Transformers track inconsistencies across sequences of frames, identifying manipulation patterns in the forged videos. - Classification: The integrated model classifies the video as Real or Fake, optionally providing frame-wise accuracy. - User Interface: An interactive UI allows uploading of videos and displays detection results along with confidence metrics.

This multi-stream system ensures detection reliability by analyzing both spatial irregularities and temporal distortions. It also provides optional training model for continuously up- dating the model with the new datasets to enhance the accuracy over the time. By preventing deep fake manipulations, the proposed system will help maintaining system integrity by improving outcomes and minimize the risks associated with fraudulent videos.

IV. SYSTEM ARCHITECTURE

Video Input and Frame Sampling
Accepts user-uploaded or forged videos and converts them into evenly spaced frames. Key frames are sampled for effi- cient processing and for accuracy.

Spatial Feature Extraction Using CNN
Every frame is processed through a CNN to extract high- dimensional visual features that highlight texture inconsisten- cies, blending artifacts, or illumination differences.

Temporal Consistency Detection Using RNN or Trans- former
Sequences of feature vectors from consecutive frames are analyzed using LSTM or Transformer layers to detect rapid changes in facial movements, mouth sync, actions or unnatural transitions.

Decision and Classification Layer
A dense classification layer processes the combined output and classifies the video into ”Real” or ”Deepfake” with a efficient score.
Web Interface

A web-based UI built using Flask that allows users to upload video files, see analysis results, and download detailed reports about their input.

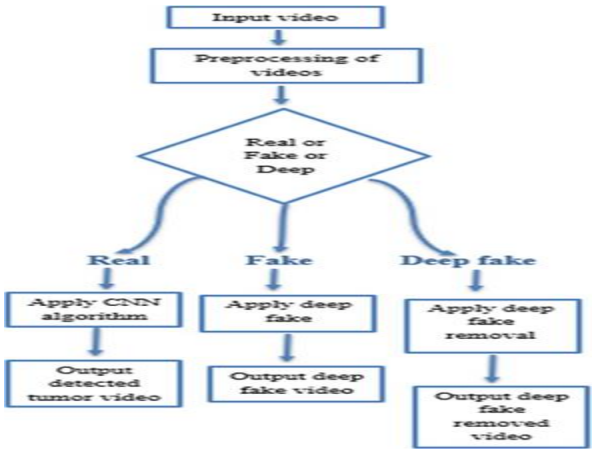

Fig. 1: System Pipeline for Tampered Video Detection

V. SYSTEM WORKING

- Video Input: A user uploads a video (e.g., MP4 format).
- Preprocessing: The system converts forged videos into evenly spaced frames and extracts frames, normalizes them, and passes them through the CNN.
- Feature Extraction: CNN detects spatial oddity; features are fed to the RNN.
- Temporal Analysis: LSTM/Transformer evaluates se- quential features which were fed to RNN and identifies frame inconsistencies.
- Classification: Based on aggregated features, the system provides the outputs "Real" or "Fake" or "Deep Fake" with accuracy.

- Result Display: UI presents results to the user, including highlighted tampered frames, with the result analysis and detailed reports.
- This pipeline of working system ensures efficient processing while maintaining accuracy and user accessibility.

- Adversarial Robustness

Training the system on adversarial and GAN-adaptive datasets will improve resistance to evasion techniques that attempt to bypass current detection methods.

- Cloud-Based Detection API

Developing a scalable cloud-based API can democratize ac- cess to deepfake detection tools, supporting wider adoption in law enforcement, journalism, and public verification platforms.

F. Domain Adaptation and Transfer Learning

Leveraging domain adaptation strategies can ensure better generalization across unseen datasets, reducing performance gaps when shifting to new video sources or manipulation techniques

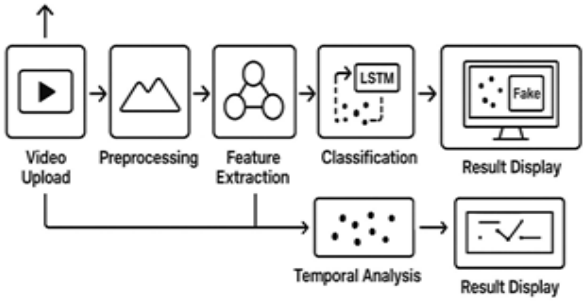

Fig. 2: Modular System Design Architecture

VI. SYSTEM DESIGN

The system is implemented with a minimalistic architecture with three layers: the user interface, the backend server and the deep learning model. The videos are uploaded by the users via the web interface, which are then forwarded to the backend to extract and preprocess the frames. The frames are processed by the trained model to determine whether the video is real or deepfake. The results are sent back to the frontend and shown to the user in a simple and easy-to-understand manner.

VIII. CONCLUSION

This research presents a comprehensive framework for deepfake video detection using deep learning techniques that combine spatial and temporal analysis. By integrating con- volutional neural networks for frame-level feature extraction with recurrent or transformer-based architectures for temporal coherence evaluation, the system demonstrates high accuracy and adaptability.

The study contributes to addressing the critical issue of synthetic media proliferation by offering a modular, extensible, and user-accessible detection system. While current results are promising, ongoing refinement, such as the inclusion of multimodal inputs and deployment optimization, is essential for widespread and reliable application. With the rising so- phistication of deepfake generation techniques, the proposed system provides a foundational approach that can be contin- uously improved to ensure digital content authenticity in an AI-driven world.

Future Work

The current framework effectively detects deepfake content using a hybrid spatial-temporal analysis approach. However, several aspects can be explored to improve the system's robustness, scalability, and generalization:

- Audio-Visual Fusion for Detection
- Future implementations could incorporate audio streams for lip-sync analysis, where mismatches between speech and facial movement serves as a tampering indicator.
- Cross-Modal and Multimodal Detection
- Integrating additional modalities such as speech signals, eye gaze tracking, and physiological cues can enhance detection reliability under varied conditions.
- Edge and Mobile Deployment
- Optimizing the model for lightweight environments such as smartphones and embedded systems will facilitate real-time and on-device detection capabilities.

REFERENCES

1. Liu, P., Tao, Q., Zhou, J. T. (2024). Evolving from Single-modal to Multi-modal Facial Deepfake Detection: A Survey. arXiv preprint arXiv:2406.06965.
2. Yu, P., et al. (2021). A Survey on Deepfake Video Detection. IET Biometrics, 10(6), 607–624. https://doi.org/10.1049/bme2.12031
3. Passos, L. A., Jodas, D., Costa, K. A. P., et al. (2023). A Review of Deep Learning-based Approaches for Deepfake Content Detection. Authorea.
4. Coccomini, D., Messina, N., Gennaro, C., Falchi, F. (2021). Combining EfficientNet and Vision Transformers for Video Deepfake Detection. arXiv preprint arXiv:2107.02612.
5. de Lima, O., Franklin, S., Basu, S., Karwoski, B., George, A. (2020). Deepfake Detection using Spatiotemporal Convolutional Networks. arXiv preprint arXiv:2006.14749.
6. Tariq, S., Lee, S., Woo, S. S. (2020). A Convolutional LSTM based Residual Network for Deepfake Video Detection. arXiv preprint arXiv:2009.07480.
7. Vyas, K., Pareek, P., Jayaswal, R., Patil, S. (2024). Analysing the Landscape of Deep Fake Detection: A Survey. International Journal of Intelligent Systems and Applications in Engineering, 12(11s), 40–55.
8. Kamakshi Thai, P., Kalige, S., Ediga, S. N., Chougoni, L. (2024). A Survey on Deepfake Detection through Deep Learning. World Journal of Advanced Research and Reviews, 21(3),2214https://doi.org/10.30574/wjarr.2024.21.3.0946