DeepFake Detection : Parent-Child Dynamics A Novel Approach to DeepFake Detection

Mogulla ShivaKumar,CSE Ai&ML,21CS002395,Sir Padampat Singhania University,Udaipur

Dr.Manish Tiwari (Supervisor),Prof. Alok Kumar (Project Coordinator)

SPSU Sir Padampat Singhania University a JK Cement Initiative

ABSTRACT

Deep fake technology has emerged as a double edged sword in the digital world.With the advancement of artificial intelligence (AI) and cloud computing, audio, video, and image manipulation techniques have grown faster and more sophisticated. While it holds potential for legitimate uses, it can also be exploited to manipulate video content, causing severe social and security concerns. The research gap lies in the fact that traditional deep fake detection methods, such as visual quality analysis or inconsistency detection, need help to keep up with the rapidly advancing technology used to create deep fakes. That means there's a need for more sophisticated detection techniques With the rapid penetration of the Internet into every part of our daily life, it is agreed that it will be an important media for future communication, perhaps even more important than the television.This product is a self-contained product made to facilitate the users with the facility to detect which video amongst the 2 is a real or fake one. This can be very helpful the society to control and reduce blackmailing and sharing of obscene content.

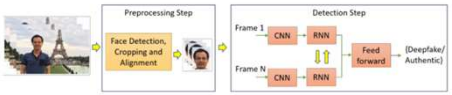
OBJECTIVES

Our research aimed to develop robust and accurate algorithms for detecting deepfakes across diverse datasets and scenarios, enhancing the generalization of detection models to handle unseen manipulation techniques effectively. We explored multimodal approaches, combining visual, audio, and metadata analysis for improved detection accuracy. Specifically, we designed and implemented a novel deepfake detection framework leveraging parent-child image relationships and evaluated the effectiveness of current state-of-the-art detection methods against adversarial deepfake generation techniques. We also developed real-time deepfake detection algorithms applicable in practical scenarios such as live streaming or video conferencing and improved the interpretability of detection models by visualizing decision-making processes. Additionally, we studied the ethical implications of deepfake detection technologies to ensure fairness in detection outcomes, identified and classified emerging trends in deepfake creation technologies to stay ahead of evolving threats, and integrated low-resource approaches for deploying deepfake detection in regions with limited computational capabilities

MATERIALS & METHODS

Fake Video Detection using Temporal Features Across Video Frames : Video manipulation is carried out on a frame-by-frame basis so the generated Deepfake videos contain intra-frame inconsistencies and temporal inconsistencies between frames. A temporalaware pipeline method that uses CNN and long short term memory (LSTM) to detect Deepfake videos is used. CNN is employed to extract frame-level features, which are then fed into the LSTM to create a temporal sequence descriptor. A fullyconnected network is finally used for classifying doctored videos from real ones based on the sequence descriptor.A

Fake Video Detection using Visual Arifacts within video Frame: In this the approach is to normally decompose videos into frames and explore visual artifacts within single frames to obtain discriminant features. These features are then distributed into either a deep or shallow classifier to differentiate between fake and authentic videos. ResNet CNN for Feature Extraction: - By using the ResNet CNN classifier, we are proposing to efficiently extract the features and create an accurate frame level classifier instead of rewrite it. To properly converge the gradient descent of the model, we will add extra layers and choose a proper learning rate to fine-tune the network. After the last pooling layer, the 2048-dimensional feature vectors are used as the sequential LSTM input.

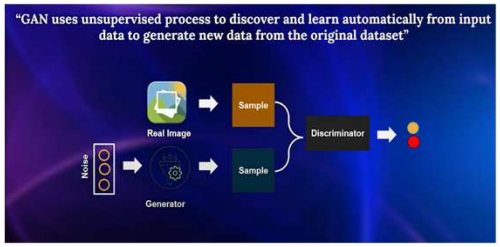


APPLICATIONS

Fake News Detection , Prevent damage to reputation of individuals , Malicious hoaxes detection , Prevent distortion of democratic discourse , Reduces exacerbation of social divisions These applications help mitigate the risks associated with deepfakes and ensure the authenticity of digital content

RESULTS

In our deepfake detection research, we achieved promising results using a combination of machine learning techniques. We trained our model on datasets like FaceForensics++ and Celeb-DF, which contain both real and fake videos. Our approach involved preprocessing the data by extracting frames, detecting faces, and aligning them for consistency. We employed a Convolutional Neural Network (CNN) combined with a Vision Transformer (ViT) for feature extraction and classification. The model was trained using various performance metrics, including accuracy, precision, recall, and F1-score. Our experimental results demonstrated the effectiveness of our approach. The CNNbased model achieved an accuracy of 97%, while the ViT-based model achieved 85% on the FaceForensics++ dataset. These results indicate significant improvements in deepfake detection compared to recent studies, affirming the potential of our framework for detecting deepfakes on social media.



TECHNOLOGIES USED

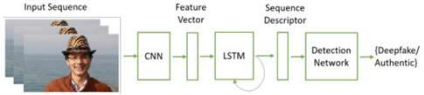
our deepfake detection research, we utilized a variety of advanced technologies to achieve robust and accurate results. We employed machine learning frameworks such as TensorFlow for building and training our deep learning models, and Keras for simplifying the creation and training of neural networks. Our deep learning models included Convolutional Neural Networks (CNNs) for feature extraction from video frames, Vision Transformers (ViTs) for enhanced feature extraction and classification, Gated Recurrent Units (GRUs) for capturing temporal dependencies in video sequences, and Generative Adversarial Networks (GANs) for data augmentation and generating synthetic data. Python was the primary programming language used for implementing our algorithms and models. For data processing and analysis, we utilized OpenCV for image and video processing, NumPy for numerical computations, and Pandas for data manipulation and analysis. Additionally, we developed a user-friendly interface using for frontend development, allowing users to upload videos and display detection results effectively

CONCLUSIONS

Motivated by the ongoing success of digital face manipulations, specially DeepFakes, this survey provides a comprehensive panorama of the field, including details of upto-date: i) types of facial manipulations, ii) facial manipulation techniques, iii) public databases for research, and iv) benchmarks for the detection of each facial manipulation group, including key results achieved by the most representative manipulation detection approaches. Generally speaking, most current face manipulations seem easy to be detected under controlled scenarios, i.e., when fake detectors are evaluated in the same conditions they are trained for. This fact has been demonstrated in most of the benchmarks included in this survey, achieving very low error rates in manipulation detection. However, this scenario may not be very realistic as fake images and videos are usually shared on social networks, suffering from high variations such as compression level, resizing, noise, etc. Also, facial manipulation techniques are continuously improving. These factors motivate further research on the generalization ability of the fake detectors against unseen conditions. This aspect has been preliminary studied in different works. Future research could be in the line of the latest publications as they do not require fake videos for training, providing a better generalization ability to unseen attacks.

REFERENCES

In our research on deepfake detection using real-time videos, we developed a robust system powered by a CNN-LSTM,GAN,GRU model. This system not only identifies manipulated content . Our approach involved processing video streams, from a file-based, to classify frames as either "Real" or "Deepfake" with explainable AI insights1. The model was trained and evaluated on various datasets, achieving high classification accuracy and demonstrating its effectiveness in real-time applications



ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Dr. Manish Tiwari, for his invaluable guidance, continuous support, and patience throughout this research. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I am also grateful to my family and friends for their unwavering support and encouragement. Lastly, I would like to thank my institution, Sir Padampat Singhania University, for providing the necessary resources and environment for this research.