Real-Time Deepfake Detection In Video Calls

Aditya Raj, Sumaira Ashfaque, Mrs Indervati

School of Computer Science Engineering, Galgotias University, Greater Noida

sumairaashfaque2@gmail.com, adityababu0004@gmail.com, indervati@galgotiasuniversity.edu.in

Abstract

The fast development of actual-time deepfake manufacturing creates protection and trust troubles throughout actual-time video conversation structures. In this paintings, we recommend a actual-time deepfake detection framework for stay video name eventualities. The laptop vision pipeline for facial regions plays face detection and alignment with deep learning techniques, and spatial and temporal functions are extracted from pre- processed snap shots. Convolutional Neural Networks (CNNs) research first-rate spatial-artifacts for face-level detection; temporal models are useful for exploiting frame-to-frame irregularities in manipulated videos. Features produced by using CNNs and temporal fashions are surpassed to supervised machine gaining knowledge of models that classify video streams to either actual or deepfake. The framework is optimized to paintings in actual-time to method inputs and make predictions with low latency, so it really works for present video conferencing programs. Experimental consequences endorse that deepfake may be categorized with excessive accuracy and low computational cost, making a contribution to the authenticity, protection and reliability of real-time virtual verbal exchange.

Keywords: Deepfake, CNN, Detection, AI, Deep Learning

Introduction

In recent years, Artificial Intelligence and Machine Learning has Grown Rapidly in both aspects innovations and challenges. The major Development among them are Deepfake Technology. Deepfake is basically a synthetic media where individual's face, voice or expression can be replaced or manipulated by another individuals using advanced neural networks. Initially, we developed this technology for the purpose of entertainment and creativity like to enhance visuals effects and to use historical Figures in movies by recreating it. But as of now Deepfake Have became a major problem/threats of concern because peoples are using it for different purpose or misleading it for political propaganda, fake news, cyber frauds and identity manipulation. The major alarming problem of deepfake is that it look too realistic that it extremely difficult to identify which is actually real and which is fake that cannot be observed by human alone. Due to which it become a more critical challenges during real-time video communication such as zoom, google meet and Microsoft teams to detect if the person in the live video is real or fake. To address this rising problem this research paper aim to developed a deepfake detection system, which is capable of identifying similarity, liveness detection and detect suspicious behavior. This system is design in a way that it can be run through a live webcam and by using OBS virtual camera, we can integrate it in zoom and google meet calls.

This Implementation enhances digital communication security by providing real-time feedback on the authenticity and suspiciousness of a participants, that will ensure trust and integrity in virtual interaction. Additionally, after COVID-19, there is a greater reliance on digital communication and virtual interactions are a way of life. Online platforms are now crucial to people for professional meetings, academic pursuits, interviews, financial operations, and personal communication. This dependency has made it very important to keep authenticity and trust during online interactions. But with the rise of AI-powered media, there are emerging vulnerabilities that allow attackers to leverage synthetic identities for deceptive live video communication. These attacks can compromise personal privacy, and pose significant problems for organizations that rely on off-site verification systems. Another problem with deep fake technology is that people in the real world are not aware of it. Attaining such synchronization of facial expressions, eye movement and lip-syncing has made most people unable to detect manipulated videos created with today's deepfake models. Sometimes, even the best cameras and video conferencing app don't detect suspicious activities during live streaming. This would make manual verification unreliable and multiply the need for an automatic intelligent detection system that can analyze the facial behavior at all times in real time. Moreover, there is an increasing need for real time detection mechanism to be an integral part to augment digital forensic systems and cybercrime prevention techniques. The AI-powered verification models can be integrated into existing communication platforms to help identify suspicious users before they can cause harm or spread false information. Such systems can also be applied in

the areas of secure data transfer such as online recruitment, exams, verification of identification for banks, and confidential corporate meetings where validation is a critical requirement. The goal of the proposed system is to increase the reliability, transparency and safety of virtual communication environment. It is a more practical and effective approach to live video analysis, as compared to the traditional media verification approaches that are only suitable for pre-recorded video. Finally, the project emphasizes the need for integrating AI technologies with cybersecurity methods to create innovative solutions that combat digital threats, as they evolve.

Figure 1 : Visual Illustration of Real and deepfake facial features

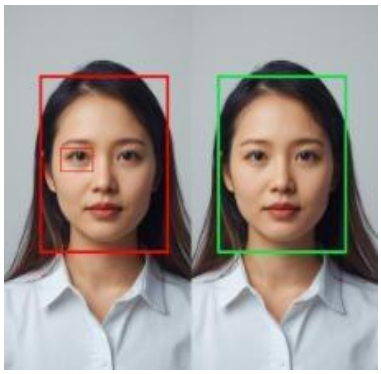

Figure 2 : Illustration of AI- based face verification showing real and fake detection.

Literature review

Deepfake detection has recently become an active area of research considering these AI or deep learning based video manipulation techniques. The realistic generation of deepfake videos based on generative adversarial networks (GANs) has raised various concerns involving identity impersonation, disinformation, social engineering attacks, and the degradation of trust in digital media. As a result, many researchers proposed deep learning-based methods to detect manipulated facial videos.

MesoNet was proposed by Afchar et al. [1] as a lightweight convolutional neural network that takes as input mesoscopic image features to detect facial manipulation artifacts that are present when creating a deepfake. MesoNet has generally good results for detecting the manipulation in videos but was mainly tested under the offline scenario and was not real-time capable.

Li et al. [2] notes that early deepfake videos lacked natural eye blinking behavior because their training datasets did not contain the binary eye states. In the same work, the authors proposed a deepfake detection algorithm based on eye blinking patterns. The authors (LRCN) to model eye states. It was successful for early deepfakes, but was less effective against advanced techniques which synthesized realistic eye movements.

Nguyen et al. [3] provide a survey on deep learning approaches to generation and detection of deepfakes, such as CNNs, RNNs, and GANs. The authors list several weaknesses in existing approaches, including computational complexity, lack of robustness across datasets, and lack of efficacy in real-time scenarios. To LSTM networks to capture the temporal inconsistency from video frames. It outperforms previous architectures when integrating spatial and temporal features. However, the dependency of LSTM models on prior observations incurs severe latency in real-time implementations.

Rossler et al. [5] introduced the dataset FaceForensics++, which became one of the most influential benchmarks to train and evaluate deepfake detection models. The authors benchmarked several state-of- the-art models on the dataset. Their results indicate that good performance can be achieved on uncompressed videos in controlled conditions, but if the videos are compressed or manipulated in unseen ways, the model performance reduces considerably Guera and Delp [6] proposed another deepfake detection method based on temporal features using recurrent neural networks and also showed that temporal inconsistencies between frames might be an indicator of video manipulation. However, the approach requires long video sequences, and it works poorly for short or low-quality video streams.

Zhao et al. [7] explored frequency-domain features that can detect deepfakes based on analyzing abnormal patterns in videos generated by GANs and can be resilient to certain modifications. However, frequency-based methods are susceptible to noise and compression artifacts commonly found in real-time video calls.

Proposed Methodology

Deepfake detection has recently become an active area of research considering these AI or deep learning based video manipulation techniques. The realistic generation of deepfake videos based on generative adversarial networks (GANs) has raised various concerns involving identity impersonation, disinformation, social engineering attacks, and the degradation of trust in digital media. As a result, many researchers proposed deep learning-based methods to detect manipulated facial videos. MesoNet was proposed by Afchar et al. [1] as a lightweight convolutional neural network that takes as input mesoscopic image features to detect facial manipulation artifacts that are present when creating a deepfake. MesoNet has generally good results for detecting the manipulation in videos but was mainly tested under the offline scenario and was not real-time capable. Li et al. [2] notes that early deepfake videos lacked natural eye blinking behavior because their training datasets did not contain the binary eye states. In the same work, the authors proposed a deepfake detection algorithm based on eye blinking patterns. The authors used a Long-term Recurrent Convolutional Network (LRCN) to model eye states. It was successful for early deepfakes, but was less effective against advanced techniques which synthesized realistic eye movements. Nguyen et al. [3] provide a survey on deep learning approaches to generation and detection of deepfakes, such as CNNs, RNNs, and GANs. The authors list several weaknesses in existing approaches, including computational complexity, lack of robustness across datasets, and lack of efficacy in real-time scenarios. To address temporal information, Sabir et al. [4] propose a temporal deep learning architecture based on CNNs and LSTM networks to capture the temporal inconsistency from video frames. It outperforms previous architectures when integrating spatial and temporal features. However, the dependency of LSTM models on prior observations incurs severe latency in real-time implementations.

Rossler et al. [5] introduced the dataset FaceForensics++, which became one of the most influential benchmarks to train and evaluate deepfake detection models. The authors benchmarked several state-of-the-art models on the dataset. Their results indicate that good performance can be achieved on uncompressed videos in controlled conditions, but if the videos are compressed or manipulated in unseen ways, the model performance reduces considerablyGuera and Delp [6] proposed another deepfake detection method based on temporal features using recurrent neural networks and also showed that temporal inconsistencies between frames might be an indicator of video manipulation. However, the approach requires long video sequences, and it works poorly for short or low-quality video streams.

Zhao et al. [7] explored frequency-domain features that can detect deepfakes based on analyzing abnormal patterns in videos generated by GANs and can be resilient to certain modifications. However, frequency-based methods are susceptible to noise and compression artifacts commonly found in real-time video calls.

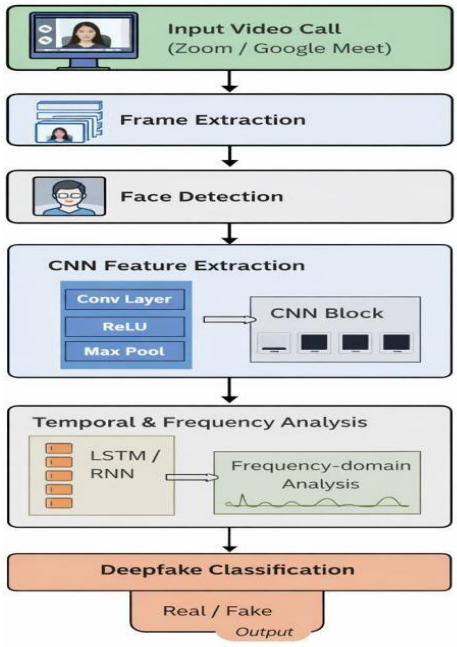

Figure 3 : Computer Vision Workflow for real-time Deepfake Detection in video calls.

System Workflow:

1. Input image or video upload
2. Frame extraction (for videos)
3. Image preprocessing
4. CNN-based feature extraction
5. Binary classification (Real/Fake)
6. Result display through web interface

Model Architecture

The proposed deepfake detection model is a binary classification system which uses CNN for detection of deepfakes by categorizing facial frames into Authentic and Suspicious. CNN architecture built with TensorFlow/Keras for learning discriminative visual features of manipulated facial media automatically. The model is designed to detect minor inconsistencies across facial texture, sync and visual artefacts that are often created during the deepfake generation process. The input for the model are the facial frames taken from the live webcam or virtual camera streams. All the frames are resized and normalized to a fixed resolution to enhance the stability of training and computational efficiency before being fed into the network. All the pre-processing operations are done using opencv.

The architecture comprises of several convolutional layers with Rectified Linear Unit (ReLU) activation functions. These convolutional layers are used for extracting low-level and high-level facial features including edges, texture abnormalities, blending artifacts and unusual visual patterns. Max-pooling layers are added after the convolution layers to decrease the dimension and computation of the network whilst still capturing critical feature information. The feature maps obtained after feature extraction are flattened and fed to fully connected dense layers. These layers fuse the removed facial images and help the model to learn the intricate relationship between the real and manipulated facial features. To prevent overfitting and enhance the model's generalizing ability, dropout regularization is employed.

The last output layer has the Sigmoid activation function, making the output probabilities fall in the range of 0 to 1. The system predicts the confidence score of the input frame and classifies the input frame as an authentic frame or suspicious frame based on the predicted confidence score. A Binary Cross-Entropy loss function was used to train the model, while an Adam optimizer was used to optimize the model for stable convergence and efficient learning performance. CNN lightweight model was chosen for real time inference during live video communication sessions with an acceptable detection accuracy and low computational overhead.

SYSTEM ARCHITECTURE AND DEPLOYMENT

The Real-Time Deepfake Detection System is a proposed browser-based tool that can analyze the authenticity of human faces of users on video communication applications on-the-fly. The system architecture features integration of Streamlit, OpenCV, TensorFlow/Keras, OBS Virtual Camera and real-time visualization modules with a unified framework for detection.

The User interface Layer is built out of the Streamlit streamlit library. The interface features webcam preview areas, detections status indicator, confidence graphs, pie-chart analysis and buttons to start/stop live detection. A similarity threshold slider is also included to fine-tune the detection sensitivity when running.

The Video Acquisition Layer receives live “video streams” from Webcams and OBS Virtual Camera sources. OBS integration enables the system to join external platforms like Zoom and Google Meet to monitor the live virtual session in real time as a deepfake. The captured frames are passed to the Processing Layer which is developed in OpenCV. This layer will extract the frames from the video and resize, normalize and pre process the facial image before it reaches the CNN model for prediction. The Deep Learning Detection Layer is a feature that utilizes the trained CNN model provided by TensorFlow/Keras to detect facial texture, visual inconsistencies, synchronization and manipulation artefacts related to deepfake media. The model keeps computing prediction levels of confidence as it is executed in real time. The outputs of the analytical process are presented dynamically by the Visualization Layer to enhance interpretability and interaction in real-time in the form of, among others: frame-wise confidence graphs, pie-chart distributions, detection labels, and confidence percentages. Finally, when the live detection process has been stopped, the Report Generation Layer generates automatically PDF reports with the summary of the predictions, with graphics analysis and with the general result of the detection of the session. Allows documentation, result storage and future forensic verification purposes. The overall architecture was assessed in terms of the practical real-time detection performance, through the experiment by using webcam stream, OBS Virtual Camera integration, Zoom meeting, and Google Meet session in various environmental conditions.

EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section describes the experimental validation of the proposed Real-Time Deepfake Detection System for real-time video communication situations. Response time, predictive capability in real time, graphical visualization performance, ease of use during live time webcam-based detection sessions are the focus points of this analysis.

Experimental Setup

The system was developed using Python and Streamlit for creating a website-based interface in the browser that can detect deepfakes in real-time. It comes with a live webcam interface, backend frame processing module, confidence visualization graphs, pie-chart based prediction analytics and automated PDF report generation functionality.

The Experimental workflow starts when the user accesses the Streamlit application via its user interface. At that time, the camera's image will be shown and continuous frame acquisition will begin (if you click on a “Start Live Detection” button). The deep learning and computer vision techniques applied in this work take each frame in the video one-by-one and analyse possible inconsistent facial features and manipulated visual artefacts in deepfake media.

An interface also features a slider for adjusting the similarity threshold which can be adjusted for sensitivity

throughout the run. The coupled real-time prediction confidence values are continuously displayed using graphical analytics, and the detection label/claim Authentic/Suspicious are included with the confidence percent to give prompt user feedback. The system was validated with Webcams stream and OBS Virtual Camera feed shots within various lighting, motion, facial orientation and background disturbance conditions. The build was able to perform repetitive real time detection without changing the performance of its browser based execution.

Effect of Real-Time Frame Analysis

Real-time frame analysis mechanism constituted an important piece of the puzzle to enhance the responsiveness of the proposed system. The application will not only process static images but also process live webcam images while running. Effects of Real-timr frame analysis-This seamless frame-by-frame surveillance facilitates detecting suspicious facial incongruities, manipulation artefacts and abnormal optical patterns with higher precision during live communications. The frame-wise analysis also enhances the visualisation of the prediction, through dynamically updating of the confidence graphs and the detection percentages in real time. Varying facial movements, lighting and facial orientation were eminently captured in the analytical results and thus allowed for more detailed observation of the authenticity fluctuations over the entire live session. The use of live frame processing, thus, enhanced the interactivity and usability of the system in the case of browser-based use.

Figure 4 : frame-wise confidence analysis during live detection.

Real-Time System Performance

The system is developed and was tested during real HWE. The system was tested in real HWE using webcam and is shown to be stable. The browser-based version could process incoming frames without any noticeable delay, as well as timestamp and update confidence graphs, pie-chart analytics and confidence labels, all the time. Streamlit's seamless integration made it easier to integrate the tool and provide real-time interaction, using a lightweight user interface. A successful automated PDF reporting system was also created to provide a detection summary and analytical results following the live session. The results of these experiments suggest: The proposed implementation is fit for practical real-time monitoring applications like virtual meetings, online interviews, and remote identity verification systems.

Figure 5 : Pie-chart representation of detection result

Error Analysis

Although the proposed system performed effectively during most testing scenarios, certain environmental conditions affected prediction consistency. Detection fluctuations were primarily observed under low lighting conditions, excessive facial motion, partial facial visibility, and low webcam quality. In some cases, background disturbances and sudden illumination changes also influenced prediction confidence values. Despite these limitations, the proposed implementation maintained practical usability and demonstrated reliable live detection capability for real-time virtual communication environments.

Discussion

The experimental results demonstrate that the proposed Streamlit-based deepfake detection system can effectively perform live facial authenticity analysis within a browser environment. Unlike conventional offline detection approaches, the proposed implementation supports continuous webcam-based monitoring, real-time graphical analytics, and automated report generation within a single integrated framework. The lightweight browser-based deployment improves practical usability and makes the system suitable for virtual meetings, online interviews, remote verification systems, and secure digital communication applications.

Conclusion

This research supplied a actual-time deepfake detection system designed for live video communication platforms together with Zoom and Google Meet. The proposed gadget integrates deep getting to know–based facial feature extraction with similarity evaluation and liveness detection to identify manipulated or synthetic video content. By combining face embeddings generated the usage of a pre-educated convolutional neural network with frame-by means of-body analysis, the system changed into in a position to differentiate among true and suspicious video frames correctly. The implementation

validated that deepfake detection may be performed in real time without requiring complicated hardware or platform-unique integration. The use of a Streamlit-primarily based interface enabled stay monitoring, confidence visualization via line charts, and statistical analysis the use of pie charts. Additionally, the mixing of a virtual digital camera the usage of OBS allowed the system to be tested in real video conferencing environments, validating its sensible applicability. The computerized technology of detection reports in PDF layout in addition greater the usability of the device for documentation and evaluation purposes.

Experimental consequences showed that the device turned into able to identifying suspicious styles including static facial conduct and coffee similarity ratings, that are commonplace indicators of deepfake content material. The visual analytics furnished clean insights into frame-level self assurance versions and standard detection distribution, making the system appropriate for both technical evaluation and non-technical demonstration. These results verify that the proposed method can function a dependable initial solution for real-time deepfake detection in online conferences. However, the system also has positive limitations. Its performance may be tormented by bad lighting situations, low video best, or extreme facial angles. In addition, the reliance on pre-trained models limits adaptability to newly rising deepfake technology techniques. Despite those demanding situations, the challenge effectively demonstrates the feasibility of deploying deepfake detection tools in actual-global video conferencing situations. In future work, the device can be similarly stepped forward via incorporating greater superior temporal models, multimodal evaluation such as audio–video synchronization, and education on larger and greater diverse datasets. Integration as a browser plugin or native application feature could also beautify accessibility. Overall, this task contributes to ongoing studies in virtual media forensics and highlights the importance of real-time deepfake detection for keeping trust and protection in virtual communication structures.

References

1. A. Rössler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc. IEEE Int. Conf Computation. View. (ICCV), 2019, pages 1–11.
2. B. Dolhansky et al., "The Deepfake Detection Challenge(DFDC) Dataset", arXiv preprint arXiv:2006.07397, 2020.
3. D. Guerra and E.J. Delp, “Deepfake video detection using recurrent neural networks,” in Proc. used. IEEE Int. Conf. Council. Video Signal-Based Surveillance (AVSS), 2018, pages 1-6.
4. Y. Lee, M.-C. Chang and S. Lu, "In ictu oculi:by knowing AI created fake videos by detecting eye blinks", in Proc. IEEE Int. Workshop Inf. Forensic Security (WIFS), 2018, pp. From 1.-7.
5. Y. Lee and S. Lew, "Uncovering deep fake Videos by detecting artifacts from Facial Bias," appears in Proceedings. IEEE Conf. calculation. View. Pattern recognition. workshops (CVPRW) (2019), pages 1-7.
6. D. Afcher, V. Nozick, J. Yamagishi, and I. Echizen, “Mesonet: A compact facial video forgery detection network,” appearing in Proc. IEEE Int. Workshop Inf. Forensic Security (WIFS), 2018, pp. From 1.-7.
7. H. Nguyen, J. Yamagishi and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos", appearing in Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing. (ICASSP), 2019, at, pages 2307–2311.
8. P. Zhou et al., “Two-stream neural networks for tempered face detection,” appears in Proc. IEEE Conf. calculation. View. Pattern recognition. Workshops CVPRW 2017 on pages 1831 to 1839.
9. Y. Qian et al., "Thinking in frequency: detecting facial forgery through mining frequency-aware clues", proceed within. EUR. range of convention ECCV, 2020, pages 86–103.
10. S. Sabir et al., "Recurrent convolutional strategies for detecting facial manipulation in videos", Proc. IEEE Conf. Calculation. see pattern reputation. workshop (CVPRW), 2019, pp. 1-10.
11. S. Singh, R. Sharma, and A.F. Smeaton, "Using GANs for synthesizing minimal training data for deepfake generation", arXiv preprint arXiv:2011.05421, 2020.
12. P. Saikia et al., “A Hybrid CNN-LSTM Model for Video deepfake detection,” arXiv preprint arXiv:2208.00788, 2022.

13. A. Dosovitskiy et al., “An image for transformers of image detection, at scale,” in Proc. Int. Conf. Learn. Represent.(ICLR), 2021.
14. K. Simonyan and A. Zisserman, “very deep complex networks for big Image recognition,” arXiv preprint arXiv:1409.1556, 2014.
15. T. Karras et al., “A style-based generator architecture for generative adversarial networks,” in Proc. IEEE Conf. Conf. Vis. pattern Recognit. (CVPR), 2019, pp. 4401–4410.
16. N. Bansal et al., “Real-Time Advanced computational intelligence for deep fake video detection,” appl. sci., vol. thirteen, no. Five, p. 3095, 2023.
17. A. Hashmi et al., “AVTENet: Audio-Visual transformer-Based ensemble network for video deepfake detection,” arXiv preprint arXiv:2310.13103, 2023.
18. I. Goodfellow et al., “Generative opposed nets,” in new style in neural information processing Systems (NeurIPS), vol. 27, 2.