REAL TIME DEEPFAKE DETECTION IN VIDEO CALLS

Aditya Chaturvedi[1]
School of Computer Science and Engineering
Galgotias University Greater Noida, India
aditya.22scse1010885@galgotiasuniversity.edu.in

Dr. Kumar Manoj[2]
School of Computer Science and Engineering
Galgotias University Greater Noida, India
kumarmanoj@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.

I. 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

Figure 1: Visual Illustration of Real and Deepfake facial representations.

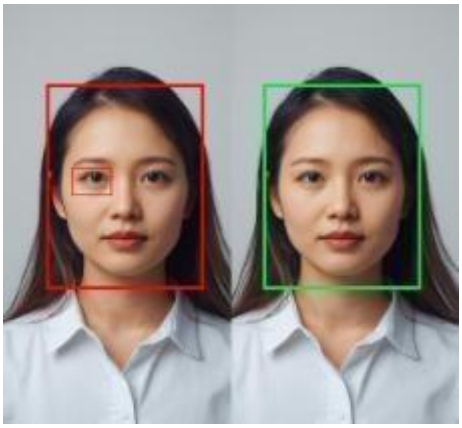

Figure 2: Illustration of AI-based face verification showing real and fake detection.

II. LITERATURE SURVEY

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.

S. No.	Author(s)	Technique / Method	Key Idea / Contribution	Strengths	Limitations
1	Afchar et al. (MesoNet)	Lightweight CNN (MesoNet)	Used mesoscopic image features to detect facial manipulation artifacts	Good detection accuracy for manipulated videos	Mainly tested offline; not suitable for real-time detection
2	Li et al.	Eye blinking-based detection using LRCN	Detected deepfakes by identifying unnatural eye-blinking patterns	Effective for early-generation deepfakes	Less effective for advanced deepfakes with realistic eye movement
3	Nguyen et al.	Survey of CNN, RNN, GAN-based methods	Provided a comprehensive review of deepfake generation and detection techniques	Identified key challenges and research gaps	High computational complexity; limited real-time applicability
4	Sabir et al.	CNN + LSTM (Temporal Deep Learning)	Captured temporal inconsistencies across video frames	Improved accuracy using spatial–temporal features	High latency due to LSTM; unsuitable for real-time use
5	Rossler et al.	CNN-based models + FaceForensics++ dataset	Introduced FaceForensics++, a large-scale benchmark dataset	Strong performance on uncompressed videos	Performance degrades with compression and unseen manipulations
6	Guera and Delp	RNN-based temporal analysis	Leveraged temporal inconsistencies as manipulation indicators	Effective for long video sequences	Poor performance on short or low-quality videos
7	Zhao et al.	Frequency-domain feature analysis	Detected GAN artifacts using abnormal frequency patterns	Robust to certain post-processing operations	Sensitive to noise and compression artifacts
8	Dolhansky et al.	DFDC Dataset & Benchmark	Introduced the DFDC dataset to address real-world deepfake challenges	Encouraged development of generalizable models	Real-time detection remains challenging

III. 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.

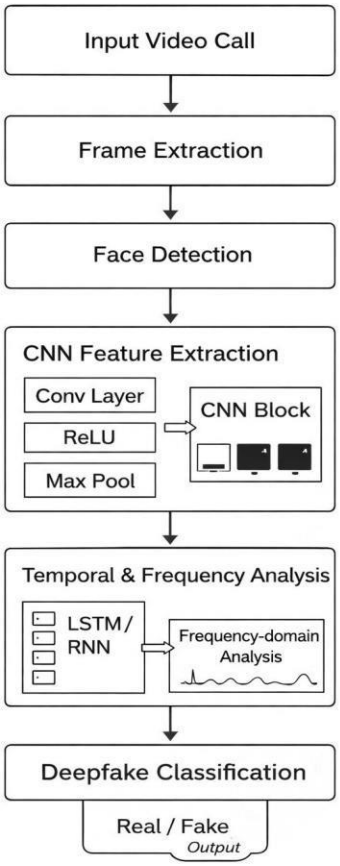

Figure 3: Computer Vision Workflow for Real-time Deepfake Detection in Video Calls

IV. IMPLEMENTATION

The Implementation of this project involved several key stages including environment setup, model- training, real-time detection through streamlit and integration using OBS for live testing on video conferencing platforms.

1 . Python environment setup-

First, we set python environment for the project.

We install python in the system.

Then install required libraries which is necessary-

tensorflow-for deep learning model

opencv-for handling video frames

NumPy, streamlit -to process data and web-interfaces

Environment was tested to ensure that all modules working perfectly .

2. Tensorflow Model training -In this phase , A convolutional neural networks was designed and trained using tensorflow to detect deepfake videos. The data was processed through resizing, normalization and frame extraction to improve training efficiency . The trained model learned to differentiate between authentic and suspicious faces, so that test set will achieve high accuracy on the test set.

3. streamlit Application for Deepfake-Detection-

To show the result in real-time we use streamlit. The application live video input from webcam and process each frame through the trained CNN model in real-time. After detecting the system will predict result

4. OBS Virtual Camera Integration-We used OBS(open broadcaster software) to convert the streamlit into a virtual camera feed. Using this we can processed live video to be used as webcam source in other application. By enabling start virtual options in OBS, Then the result from the deepfake detector can integrate into platforms like zoom and google meet.

5. Integration and Testing with Zoom and Google meet-In the testing phase, the project was integrated with zoom and google meet. We select OBS virtual camera setting in the meeting. When running the streamlit app the detection output was visible in line meeting. Then the system analyze the face and detect whether it's fake or real.

V. RESULT

The main results of the project was to differentiate between Authentic and Deepfake Faces by implementing real-time video feed in the system.

When we run streamlit in the system it execute live

Frame from the webcam and for each frame it calculate the similarity score. If the similarity is larger than threshold value that is 0.6 than it is labeled as Authentic, but if the value is lesser than the threshold value that is 0.6 than it is labelled as suspicious. All the result were updated in real-time. We also implement line chart feature in our model to shows the confidence value, which means how sure our model is that the frame is authentic or suspicious.

If the system thinks that the face is authentic, then the confidence value will be higher ranging from 80-100% and line goes in the upward direction whereas if the system think that the frame was suspicious the confidence value was low ranging from 40-60% and the line goes in the downward direction, from this we can have real-time trend which show that in which moments the system detect that the video was real and when it is doubtful. In this way, line chart works as a continuous performance indicator which shows accuracy and certainty for each frames.

As in the pie chart feature that we implemented

It give the overall summary that how much part is authentic or suspicious in the frame. If the system

Marks the system as authentic then the green part portion was bigger in the pie chart. But on the other hand if the system marks frame as suspicious

Then the red portion was bigger in the pie chart.

The system also show that if the faces was not moving it will flag as a static face, then the liveness detection triggers and model will label it as a suspicious. In the testing phase, when we integrate OBS virtual camera in the zoom and Google meet. The System will detect the same way in the live meeting. That's mean in the video conferencing environment deepfake detection system was working successfully.

In the performance terms our model was accurate and fast. On Generating and comparing the face embeddings there is as such no noticeable delay and the detection real time was working on 20-25FPS approx. This results proved that our model/system can be apply in a realistic use-case such as online meeting and interview verification.

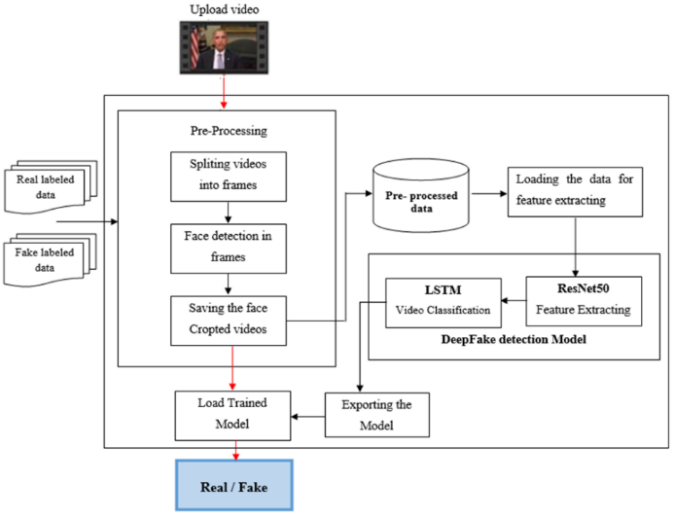

Figure 4: Real-Time Deepfake Detection Framework for Video Conferencing Platforms

VI. 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.

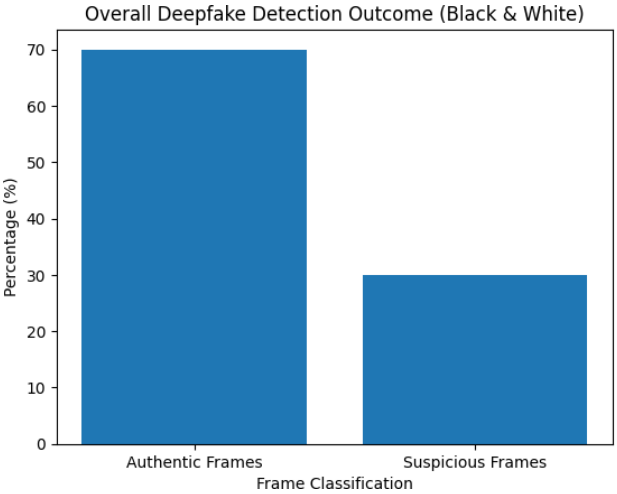

Figure 5: Deepfake Detection Outcome (Authentic Frames vs Suspicious Frames)

VII. 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] P. Zhou et al., “Two-Stream Neural Networks for Tampered Face Detection,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1831–1839.

[8] Y. Qian et al., “Thinking in Frequency: Detecting Facial Forgery through Mining Frequency-Aware Clues,” in*Proc. European Conf. Computer Vision (ECCV)*, 2020, pp. 86–103.

[9] S. Sabir et al., “Recurrent Convolutional Strategies for Face

[10] Manipulation Detection in Videos,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (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.