A HYBRID APPROACH FOR DETECTION OF DEEP FAKE VIDEOS

Mr.K.Jeevan Ratnakar[1], Anil Arudra[2], Jaswanth Bolamala[3], Srinivas Reddy Burrumukku[4], Kotesh Naik Dheerevath[5]

[1]Assistant Professor, IT Department, Vasireddy Venkatadri Institute of Technology, Namburu, Guntur, Andhra Pradesh -5225208

[2,3,4,5]UG Students, IT Department, Vasireddy Venkatadri Institute of Technology, Namburu, Guntur, Andhra Pradesh -5225208

Mail Id: jeevanratnakar@vvit.net

ABSTRACT

The rise of face-swap deepfake videos poses significant challenges to personal privacy, public trust and media integrity. This research focuses on developing an advanced AI/ML based solution leveraging Graphical Neural Network (GNN) and Transformers for the detection of face-swap based deepfake videos. GNNs are employed to model relational inconsistencies in facial landmarks, capturing spatial dependencies and subtle anomalies introduced by face-swapping techniques. Transformers are integrated to analyse temporal inconsistences across video frames, effectively detecting motion anomalies such as un natural blinking or misaligned expressions. The hybrid framework combines strengths of GNNs for spatial anomaly detection and transformers for temporal feature extraction ensuring high accuracy ad robustness against diverse deepfake generation methods. Experimental results demonstrate that the proposed system achieves state of the art performance, with better accuracy levels on benchmark dataset.

Keywords: Deepfake detection, Face-swap videos, Graph Neural Networks (GNN), Transformers, Temporal inconsistencies, Spatial anomalies.

I INTRODUCTION

The face is the most distinctive feature of human beings. With the tremendous growth of face synthesis technology, the security risk posed by face manipulation is becoming increasingly significant. Individual's faces may often be swapped with someone else's faces that appear authentic because of the myriads of algorithms based on deep learning technology. Deepfake is an emerging subdomain of artificial intelligence technology in which one person's face is overlaid over another person's face. More specially, multiple methods based on generative adversarial networks produce high-resolution deepfake videos. Unfortunately, due to the widespread usage of cell phones and the development of numerous social networking sites, deepfake content is spreading faster than ever before in the twenty first century, which has turned into a global danger. Initially, deepfake videos were discernible with the human eye due to the pixel collapse phenomena that tend to create artificial visual inconsistencies in the eyes blinking, skin tone or facial shape. Not only videos, but also audios can be turned into deepfakes. Deepfakes have grown to be barely distinguishable from natural videos as technology has progressed over the years. Consequentially, people all across the world are experiencing inescapable complications.

Because of deepfake technology, people may choose their fashion more quickly, which benefits the fashion and e-commerce industries. Furthermore, this technology aids the entertainment business by providing artificial voices for artists who cannot dub on time. Additionally, filmmakers can now recreate many classic sequences or utilize special effects in their films because of deepfake technology. Deepfake technology can potentially let Alzheimer's patients communicate with a younger version of themselves, which might help them retain their memories. The faces of many celebrities and other well-known individuals have been grafted onto the bodies of pornographic models, and these images are widely available on the Internet. Deepfake technology may create satirical, pornographic, or political content about familiar people by utilizing their pictures and voices without their consent.

A deep fake video of the former American president Barack Obama is being circulated on the Internet these days where he is uttering things that he has never expressed. Furthermore, deepfakes have already been used to alter Joe Biden's footage showing his tongue out during the US 2020 election. Besides, Taylor Swift, Gal Gadot, Emma Watson, Meghan Markle, and many other celebrities have been victims of deepfake technology. In the United States and Asian societies, many women are also victimized by deep fake technologies, scams, deception, and insecurities from society, researchers have been relentlessly trying to detect deepfakes. The identification of deepfakes would reduce the number of crimes that are currently occurring around the world. Therefore, researchers have paid attention to the mechanism for validating the integrity of deepfakes. In reaction to this trend, some multinational companies have started to take initiatives. For instance, Google has made a fake video database accessible for academicians to build new algorithms to detect deepfake, while Facebook and Microsoft have organized the Deepfake Detection Challenge

The dataset for this work was obtained from Face Forensic++ from Kaggle.

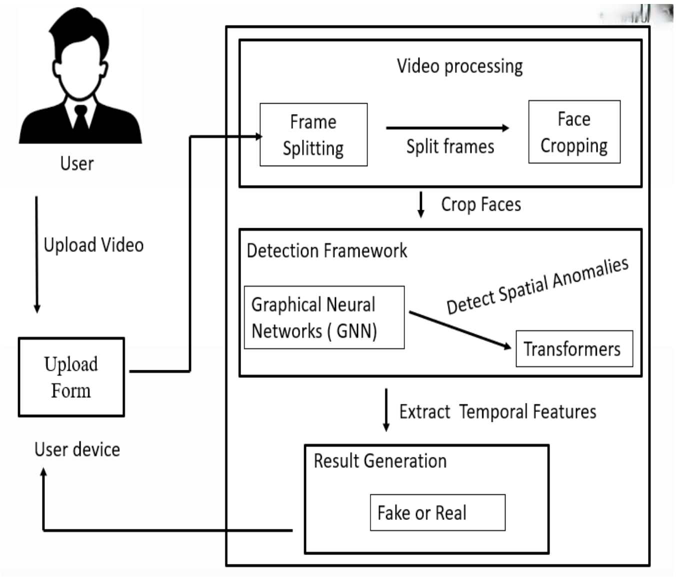


Fig 1: System architecture

II LITERATURE SURVEY

A. Existing Research:

In existing system, Autoencoders have been widely used for deepfake video detection due to their ability to learn compressed representations of data and reconstruct input frames. In this approach Autoencoders capture the key features of input video frames and compare reconstructed frames with the originals to identify anomalies introduced by deepfake manipulations. Reconstruction errors, such as distortions in facial landmarks or pixel-level inconsistencies, serve as indicators of manipulated content.

- Learning a compressed latent space representation of facial features.
- Identifying anomalies in reconstructed images to detect manipulated content.
- Leveraging reconstructions errors as in indicator of deepfake manipulation.

B. Limitations:

While autoencoder models have some success, they suffer from several limitations:

1.Autoencoders primarily focus on frame-wise analysis, often failing to capture temporal inconsistencies across consecutive frames.

2. Facial manipulations in deepfake videos often disrupt relationships between facial landmarks where autoencoders struggle to model these dependencies.

3. Autoencoder models are highly dependent on the training dataset. They may fail to generalize well when exposed to unseen deepfake variations.

C. Proposed System:

To overcome the drawbacks of autoencoder-models, **Graph Neural networks (GNNs)** and **Transformers** have emerged as promising approaches due to their ability to model complex dependencies and high-dimensional representations.

1.Graph Neural Networks (GNN):

- GNNs are well-suited for capturing structural relationships between facial components.
- Instead of analyzing individual pixels, GNNs represent facial landmarks as nodes and their spatial relationships as edges in a graph.

- A message-passing mechanism allows the model to learn the connectivity patterns between different facial features.

2.Transformers

Unlike the traditional approaches, **Vision Transformers (ViTs)** architecture capture long-range dependencies in video sequences.

By applying self-attention mechanisms, Transformers can detect inconsistencies in head movements, lip-sync, and eye blinking patterns.

III IMPLEMENTATION

Creating a deep fake video detection system, first we have to train the model on the dataset. We have taken the FaceForensic++ dataset for model training.

Gathering and Preparing Data

The first task in building the system is collecting data. For this, we rely on publicly available dataset like FaceForensics++. This dataset contains the two sub video folders, one of the contains fake videos and another one contains real videos for the model training.

Frame extraction and Face Cropping

The next step is extracting the frames from the video and cropping the faces from the frames using OpenCV library. The system determines the cropped faces as landmarks and train on the features of the face like eyes blinking, face symmetry etc.

Spatial Anomaly Detection

After cropping the faces, the next step is detection of spatial anomalies from the faces. Graph Neural Networks (GNNs) in deep fake video detection focuses on identifying inconsistencies in spatial relationships within video frames. Deep fake videos often exhibit irregularities in facial structures, lighting, and texture coherence. GNNs analyse spatial dependencies between pixels or facial landmarks, capturing topological discrepancies that standard convolutional networks may overlook. By modelling facial features as a graph, where nodes represent key points and edges encode spatial relationships, GNNs effectively highlight subtle anomalies.

Temporal Features Extraction:

The next step in the process is extraction of Temporal features using Vision Transformers **(ViT)**. ViT processes the frames and extracts the featues like unusual eye blinking, face textures etc from the video.

IV.METHODOLOGY

The proposed deep fake video detection system uses deep learning models like GNN and transformers to detect whether the given video is deep fake or not. It uses the python libraries like OpenCV for video frame extraction, torch vision is used for the processing of the frames. The main model is a comabination Graph Neural networks and transformers. This system handles the both spatial and temporal features for the detection.

V ALGORITHMS

Data Collection: Collecting the fake and real videos from the dataset like FaceForensics++ from the Kaggle platform for the model.

Preprocessing: extracting the frames from the video and cropping the faces from the frames and take as facial land marks.

Spatial Anomalies: Graph Neural Network detects the spatial anomalies like inconsistency in lighting, asymmetric in face structure etc.

Temporal features: Vision Transformers are used to extract the temporal features like lip moments, eye blinking etc.

Display Results:

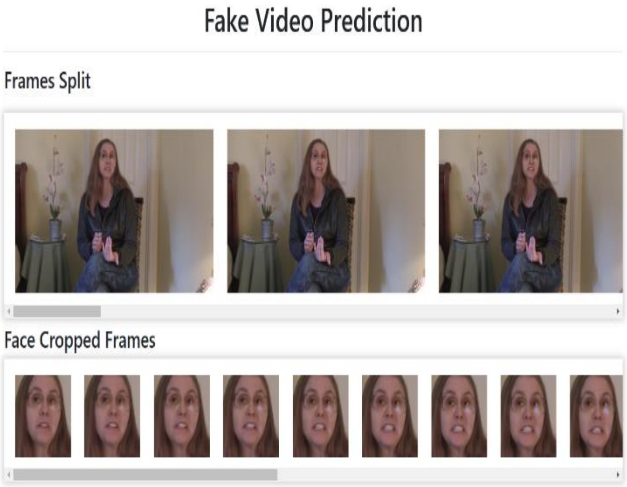
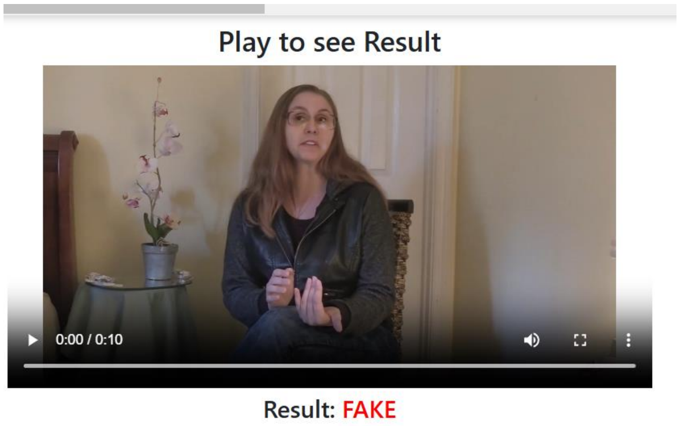


Fig 1: Frame extraction and face cropping



Play to see Result

Result: FAKE

Fig 2: displaying the video is Fake or Real

CONCLUSION

The hybrid approach effectively captures both spatial and temporal artifacts, leveraging GNNs for relationship modelling and Transformers for powerful feature extraction. The combination enhances robustness against adversarial attacks and improves detection.

References

A. Trabelsi, M. M. Pic, and J.-L. Dugelay. Improving deepfake detection by mixing top solutions of the dfdc. in, 2022:30, 2022.

Naveed Ur Rehman Ahmed, Afzal Badshah, Hanan Adeel, Ayesha Tajammul, Ali Daud, and Tariq Alsahfi. Visual deepfake detection: Review of techniques, tools, limitations, and future prospects. IEEE Access, 13:1923–1961, 2025.

P. N. Vasist and S. Krishnan. Engaging with deepfakes: a meta-synthesis from the perspective of social shaping of technology theory. Internet Research, 2022.

V. Wesselkamp, K. Rieck, D. Arp, and E. Quiring. Misleading deep-fake detection with gan fingerprints, arxiv. 2022.

R. Wightman and GitHub PyTorch Image Models. 2019.

Z. Yan, P. Sun, Y. Lang, S. Du, S. Zhang, W. Wang, and L. Liu. Multimodal graph learning for deepfake detection, arxiv. 2023.

X. Yang, Y. Li, and S. Lyu. Exposing deep fakes using inconsistent head poses. In Icassp Ieee, editor, International Conference on Acoustics, Speech and Signal, pages 2019–2019. Processing (ICASSP, 2019.

N. Yu, V. Skripniuk, S. Abdelnabi, and M. Fritz. Artificial fingerprinting for generative models: Rooting deepfake attribution in training data. 2022.

G. P. Zachary. Digital manipulation and the future of electoral democ racy in the u.s. IEEE Transactions on Technology and Society, 1:104 112, June 2020.

D. Zhang, F. Lin, Y. Hua, P. Wang, D. Zeng, and S. Ge. Deepfake video detection with spatiotemporal dropout transformer, arxiv. 2022.