

Anti Spoofing and Deepfake Detection

Ambresh Bhadrashetty¹, Manjulabai Bhadrashetty², Laxmiputra³

¹Assistant Professor, Department of Computer Science and Engineering (MCA)
Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India

²Associate Professor, Department Of Computer Science, Government First Grade College for Women's
,Jewargi Colony, Kalaburagi, India

³Student, Department of Computer Science and Engineering (MCA)
Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India

Abstract - The rapid advancement of artificial intelligence has enabled the creation of highly realistic synthetic videos, commonly known as deepfakes. These manipulated videos pose significant challenges to digital security, media authenticity, and public trust, as they can be used to spread misinformation, commit fraud, and influence public opinion. Detecting such forged content has become an important research area in computer vision and multimedia forensics.

This paper presents an intelligent deepfake video detection system based on a hybrid deep learning framework. The proposed approach combines a ResNeXt-based Convolutional Neural Network (CNN) for extracting spatial features from video frames with a Long Short-Term Memory (LSTM) network for analyzing temporal relationships across consecutive frames. This integration enables the model to identify both visual inconsistencies and sequential anomalies that are commonly present in manipulated videos.

To improve detection performance and generalization, the model is trained on a diverse dataset consisting of authentic and manipulated videos collected from multiple publicly available sources. The preprocessing stage includes face detection, face cropping, frame extraction, and normalization to enhance the quality of input data. Experimental evaluation demonstrates that the proposed framework effectively distinguishes genuine videos from deepfakes while maintaining computational efficiency.

The developed system provides a practical solution for automated deepfake detection and can be applied in social media monitoring, digital content verification, cybersecurity, and forensic investigations. The results indicate that combining spatial and temporal feature learning significantly improves the reliability of deepfake identification, contributing to the development of trustworthy digital media environments.

Key Words: Deepfake Detection, Deep Learning, Computer Vision, ResNeXt, LSTM, CNN, Video Analysis, Artificial Intelligence, Digital Forensics, Cybersecurity.

1. INTRODUCTION

The widespread use of digital media and social networking platforms has transformed the way information is created, shared, and consumed. Along with these advancements, Artificial Intelligence (AI) has introduced powerful technologies capable of generating highly realistic synthetic content. One of the most notable developments is deepfake technology, which uses deep learning algorithms to manipulate videos and images by altering facial expressions, replacing identities, or generating entirely artificial content. While such technology has beneficial applications in entertainment, education, and content creation, it also presents serious challenges related to misinformation, identity theft, cybercrime, and digital fraud.

Deepfakes are commonly created using advanced neural network architectures such as Generative Adversarial Networks (GANs), Autoencoders, and Diffusion Models. These techniques can produce manipulated videos that appear authentic to human observers, making manual verification increasingly difficult. The rapid growth of deepfake content on online platforms has raised concerns regarding media authenticity, public trust, and information security.

To address this challenge, an effective automated detection mechanism is required. This research proposes a deep learning-based framework for identifying manipulated videos by analyzing both spatial and temporal characteristics. The system combines a ResNeXt-based Convolutional Neural Network (CNN) for extracting visual features from video frames and a Long Short-Term Memory (LSTM) network for capturing sequential patterns across frames. By integrating these techniques, the proposed model can effectively distinguish between genuine and manipulated videos.

The objective of this work is to develop a reliable, efficient, and scalable deepfake detection system that supports digital content verification and contributes to the protection of online information ecosystems. The proposed approach aims to strengthen multimedia security and assist in combating the growing threat of AI-generated video forgeries.

2. PROBLEM STATEMENT

The rapid advancement of artificial intelligence has made it possible to create highly realistic manipulated videos, commonly known as deepfakes. These videos can alter a person's appearance, speech, or actions in a way that appears genuine, making it difficult for viewers to distinguish between authentic and fabricated content. The increasing availability of deepfake generation tools has accelerated the spread of misleading and fraudulent media across digital platforms.

Deepfake videos pose significant risks in areas such as misinformation, identity theft, financial fraud, cybercrime, political manipulation, and reputational damage. Since modern deepfake techniques generate highly convincing visual content, traditional methods of manual verification are often ineffective. As a result, there is a growing need for automated systems capable of accurately detecting manipulated videos.

The main challenge is to identify subtle visual artifacts and temporal inconsistencies introduced during the deepfake generation process while maintaining high detection accuracy across different video qualities and manipulation techniques. Therefore, this research focuses on developing an intelligent deep learning-based framework that can effectively distinguish genuine videos from manipulated ones, helping to improve digital media authenticity, security, and trustworthiness.

3. OBJECTIVES

The primary objective of this research is to develop an intelligent and reliable system for detecting deepfake videos using advanced deep learning techniques. The proposed framework aims to identify manipulated content by analyzing visual and temporal patterns present in video sequences. By combining spatial feature extraction with sequential data analysis, the system seeks to accurately differentiate between authentic and forged videos.

The specific objectives of the study are to improve the accuracy of deepfake detection, reduce the impact of misinformation spread through manipulated media, and provide an automated solution for digital content verification. The system is designed to process uploaded videos, extract meaningful features, and generate classification results with confidence scores. Additionally, the research aims to enhance the security and trustworthiness of digital media platforms by supporting applications in cybersecurity, multimedia forensics, and online content authentication. The overall goal is to create a scalable and efficient detection framework capable of adapting to evolving deepfake generation techniques.

4. METHODOLOGY

The proposed deepfake detection system follows a structured methodology that combines video preprocessing, feature extraction, temporal analysis, and classification. Initially, video datasets containing both authentic and manipulated samples are collected from publicly available sources. The videos are then

processed to extract frames, and facial regions are identified and cropped to focus the analysis on the most relevant areas. This preprocessing step reduces noise and improves the quality of the input data.

After preprocessing, the extracted facial frames are passed through a ResNeXt-based Convolutional Neural Network (CNN), which learns important visual features such as texture irregularities, blending artifacts, and inconsistencies introduced during the deepfake generation process. The resulting feature representations are then provided to a Long Short-Term Memory (LSTM) network, which analyzes temporal relationships among consecutive frames and captures motion-based anomalies that may not be visible in individual images.

The combined CNN-LSTM architecture is trained using a balanced dataset containing both real and fake videos. During training, optimization techniques and loss functions are applied to improve model performance and reduce classification errors. Finally, the trained model evaluates unseen videos and classifies them as either authentic or manipulated, along with a confidence score indicating the reliability of the prediction. This methodology enables accurate and efficient detection of deepfake content while maintaining robustness across different video sources and manipulation techniques.

5. LITERATURE SURVEY

Deepfake detection has emerged as an important research area due to the rapid growth of artificial intelligence-based media manipulation techniques. Researchers have proposed various approaches to identify forged videos by analyzing visual artifacts, facial inconsistencies, and temporal irregularities. Early studies primarily focused on detecting image-level anomalies introduced during the face-swapping process. These methods utilized convolutional neural networks to identify distortions around facial boundaries, unnatural textures, and blending artifacts. While effective in controlled environments, their performance often decreased when applied to high-quality deepfakes.

Subsequent research explored physiological and behavioral characteristics, such as eye blinking patterns, head movements, and facial expressions, to distinguish manipulated content from authentic videos. These approaches demonstrated that synthetic videos frequently fail to reproduce natural human behaviors consistently. However, advancements in deepfake generation techniques gradually reduced the reliability of individual physiological indicators.

Recent studies have adopted deep learning architectures that combine spatial and temporal analysis for improved detection accuracy. Convolutional Neural Networks (CNNs) have been widely used to extract discriminative visual features, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been employed to model frame-to-frame relationships and temporal dependencies. These hybrid approaches have shown better performance in identifying subtle inconsistencies present across video sequences.

Although significant progress has been achieved, challenges such as varying video quality, compression effects, and continuously evolving generation methods still affect detection performance. Therefore, there is a need for more robust and scalable frameworks capable of learning complex patterns from diverse datasets. The proposed research addresses these challenges by integrating advanced feature extraction and temporal modeling techniques to enhance the reliability and effectiveness of deepfake video detection systems.

6.SYSTEM DESIN

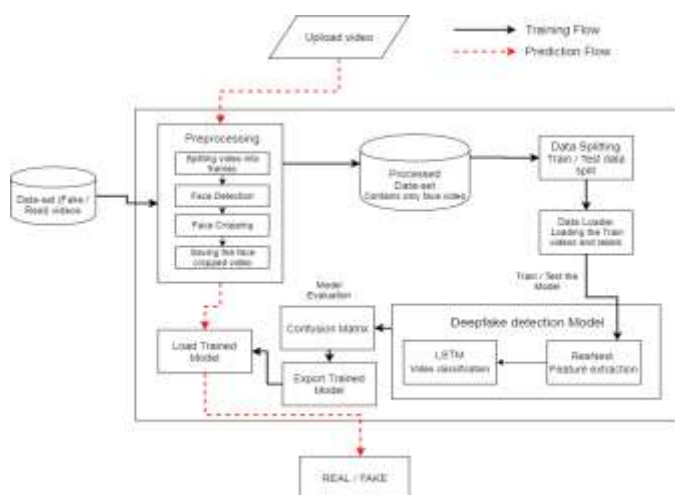


Figure 1: System Architecture

illustrates the overall architecture of the proposed deepfake video detection system. The process begins with a dataset containing both real and manipulated videos. During the preprocessing stage, videos are divided into frames, facial regions are detected and cropped, and the processed face-centered videos are stored in a new dataset. The dataset is then split into training and testing subsets.

The training data is loaded through a data loader and passed to the deepfake detection model, which consists of a ResNeXt Convolutional Neural Network (CNN) for feature extraction and a Long Short-Term Memory (LSTM) network for temporal sequence analysis and video classification. The model is evaluated using a confusion matrix, and the trained model is subsequently saved for future predictions.

During the prediction phase, a user uploads a video, which undergoes the same preprocessing operations. The processed video is then analyzed using the trained model, and the system generates a final classification result indicating whether the video is Real or Fake. The architecture supports both model training and real-time inference, providing an efficient framework for automated deepfake detection.

7. SCREENSHOTS

1. Home Page

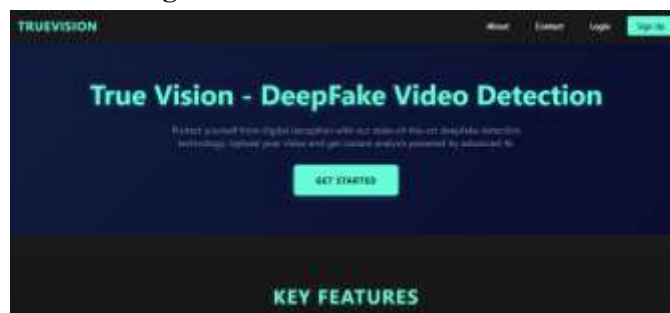


Figure 2: Home Page

2. Uploading Real Video



Figure 3: Uploading Real Video

3. Real Video Output



Figure 4: Real Video Output

4. Uploading Fake Video



Figure 5: Uploading Fake Video

5. Fake video Output



Figure 6: Fake video Output

8. CONCLUSION AND FUTURE SCOPE

The rapid growth of artificial intelligence has enabled the creation of highly realistic deepfake videos, creating significant challenges for digital security, information authenticity, and public trust. To address this issue, this research proposed a deep learning-based deepfake detection framework that combines a ResNeXt Convolutional Neural Network (CNN) for spatial feature extraction with a Long Short-Term Memory (LSTM) network for temporal sequence analysis. The system processes video data through preprocessing, face extraction, feature learning, and classification stages to accurately distinguish between authentic and manipulated videos. Experimental results demonstrate that the proposed approach effectively detects visual artifacts and temporal inconsistencies commonly found in deepfake content while maintaining reliable performance across diverse video samples.

The developed framework provides an automated, efficient, and scalable solution for multimedia authentication, digital forensics, and cybersecurity applications. It can assist in reducing the spread of misinformation, identity fraud, and other malicious activities involving manipulated media. In the future, the system can be enhanced by incorporating advanced architectures such as Vision Transformers (ViT), EfficientNet, and multimodal learning techniques that analyze visual, audio, and behavioral cues simultaneously. Further improvements may include real-time detection of live video streams, deployment on mobile and web platforms, integration with social media monitoring systems, and training on larger, more diverse datasets to improve robustness against emerging deepfake generation techniques. These advancements will contribute to building more secure and trustworthy digital media environments.

9. REFERENCES

- [1] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.
- [2] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The Deepfake Detection Challenge (DFDC) Dataset," *arXiv preprint arXiv:2006.07397*, 2020.
- [3] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [6] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [8] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [9] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [11] H. Farid, "Image Forgery Detection: A Survey," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, 2009.
- [12] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *International Conference on Learning Representations (ICLR)*, 2014.