Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

Deepfake Detection Based on Temporal Analysis of Facial Dynamics Using LSTM and ResNeXt Architectures

Mr. A. V. Srinivas[1*], Manikanta Swamy Angara[2], Snehitha Chamarthi[3], Sanjeevi Kumar Guptha Gangisetti[4], V. S. Naga Sai Pavan Rahul Lingala[5]

[1]Assistant Professor, Department of Information Technology, Vishnu Institute of Technology, Andhra Pradesh, India.*
[2,3,4,5]Department of Computer Science and Business Systems, Vishnu Institute of Technology, Andhra Pradesh, India.

Corresponding Email: [1]Srinivas.av@vishnu.edu.in*

Received: 28 November 2023 **Accepted:** 14 February 2024 **Published:** 01 April 2024

Abstract: ***The proliferation of deepfake technology presents a critical challenge to the authenticity and trustworthiness of digital media. To address this issue, we propose an innovative deepfake detection framework that combines the power of Long Short-Term Memory (LSTM) and ResNeXt architectures. By integrating spatial and temporal analysis methods, our approach aims to accurately identify manipulated videos amidst the vast sea of online content. Through rigorous experimentation and evaluation using diverse datasets, our framework demonstrates promising results in effectively distinguishing between genuine and fake videos. This research contributes to the ongoing efforts to combat deepfake misinformation and uphold the integrity of digital media platforms.***

Keywords: Deepfake Detection, LSTM (Long Short-Term Memory), Image Manipulation, Facial Recognition, Cybersecurity, Digital Forensics.

1. INTRODUCTION

The rapid advancement of deepfake technology has ushered in a new era of digital media, characterized by the proliferation of hyper-realistic yet entirely synthetic content. While deepfakes have garnered attention for their entertainment value, they also pose significant challenges to the integrity and trustworthiness of multimedia communication. With the increasing prevalence of manipulated videos circulating online, there is a pressing need for robust detection mechanisms to combat the spread of misinformation and safeguard the authenticity of digital content.

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

In response to this imperative, our research endeavours to develop an innovative deepfake detection framework leveraging state-of-the-art deep learning methodologies. Specifically, we explore the efficacy of Long Short-Term Memory (LSTM) networks and ResNeXt architectures in discerning temporal and spatial patterns indicative of deepfake manipulation. By harnessing the computational power of artificial intelligence and computer vision, our framework aims to accurately identify and mitigate the dissemination of fraudulent multimedia content across digital platforms.

The significance of our study lies in its potential to address the growing threat posed by deepfake technology to online discourse and information dissemination. By employing advanced deep learning algorithms and techniques, we seek to enhance the resilience of digital media platforms against the propagation of manipulated content, thereby fostering trust and transparency in online communication channels.

In this paper, we present a comprehensive exposition of our deepfake detection framework, elucidating the methodology, experimental design, and empirical findings. Through rigorous experimentation and comparative analysis with existing approaches, we demonstrate the efficacy and robustness of our methodology in detecting and mitigating the impact of deepfake manipulation on digital media integrity.

2. RELATED WORKS

The emergence of deepfake technology has prompted extensive research into effective detection methodologies aimed at mitigating the risks associated with manipulated multimedia content. This section offers a comprehensive review of the existing literature, focusing on notable advancements and methodologies in deepfake detection.

One prevalent approach in deepfake detection involves the application of deep learning techniques, notably convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to analyze and classify visual content. Afchar et al. [1] introduced MesoNet, a compact CNN architecture tailored for detecting facial manipulation in videos. Their method focuses on identifying subtle artifacts and inconsistencies introduced during the deepfake generation process, achieving notable accuracy in distinguishing between authentic and manipulated content.

In addition to CNN-based approaches, researchers have explored the utilization of recurrent neural networks (RNNs) for temporal analysis of video sequences. Sabir et al. [2] proposed recurrent convolutional strategies for facial manipulation detection, leveraging the temporal dynamics of video frames to detect subtle alterations indicative of deepfake manipulation. Their approach demonstrated promising results in detecting deepfake videos with high accuracy and robustness against adversarial attacks.

Furthermore, advancements in deep learning architectures, such as Long Short-Term Memory (LSTM) networks, have facilitated the capture of long-range temporal dependencies in video

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

data. Li et al. [3] proposed a hierarchical attention-based framework for deepfake detection, incorporating LSTM modules to analyze temporal patterns and spatial attention mechanisms to focus on relevant regions of interest. Their framework achieved state-of-the-art performance in detecting deepfake videos across diverse datasets.
Despite the strides made in deepfake detection, several challenges persist. Li et al. [4] underscored the importance of large-scale and diverse datasets for training robust deepfake detection models. They introduced Celeb-DF, a challenging dataset comprising real and deepfake videos of celebrities, to facilitate benchmarking and evaluation of deepfake detection algorithms. Additionally, Wu et al. [5] emphasized the need for comprehensive evaluation metrics and standardized evaluation protocols to ensure fair comparisons between different detection methods.

In summary, the literature review highlights the significance of deep learning techniques, particularly CNNs, RNNs, and LSTM networks, in advancing the field of deepfake detection. While notable progress has been achieved, ongoing research endeavors are imperative to address remaining challenges and enhance the robustness and reliability of deepfake detection systems.

Existing Systems
Deep Fake Detection Based on Spatial Inconsistency
A deep learning-based method which leverages a combination of spatial features extracted from individual frames using a modified version of the Xception convolutional neural network (CNN) architecture. The system incorporates attention mechanisms to highlight discriminative regions in the input frames and uses a binary classifier to distinguish between real and fake video.

Limitations
Limited Detection Capability: Spatial inconsistency-based methods may struggle to detect certain types of deepfakes that do not introduce significant visual artifacts or inconsistencies in individual frames. These methods may be less effective at detecting subtle manipulations or sophisticated deepfake techniques that preserve spatial coherence.

Dependency on Frame-Level Analysis: Spatial inconsistency-based methods typically analyze individual frames in isolation to identify visual anomalies or inconsistencies. However, deepfake manipulation techniques often involve temporal coherence across frames, making it challenging to detect anomalies solely based on spatial features. This dependency on frame-level analysis may limit the detection accuracy and robustness of spatial inconsistency-based systems.

Phenome-Viseme Mismatch Based Deep Fake Detection
This is a deepfake detection system that leverages phoneme-viseme alignment analysis to identify inconsistencies between lip movements and speech content in video recordings. The system employs computer vision techniques to extract viseme sequences from lip movements and speech recognition algorithms to transcribe spoken content into phoneme sequences. By

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

comparing the alignment between phoneme and viseme sequences, this detects discrepancies that may indicate the presence of deepfake manipulation.

Limitations

Dependency on Speech Content: Phoneme viseme mismatch-based detection systems rely heavily on the quality and accuracy of speech recognition algorithms to transcribe spoken content into phoneme sequences. Errors or inaccuracies in speech transcription can lead to false positives or false negatives in deepfake detection results, reducing the overall reliability and effectiveness of the system.

Sensitivity to Environmental Factors: Performance may be sensitive to environmental factors such as background noise, accent variations, or speech impediments, which can affect the accuracy of phoneme recognition and alignment. Variability in speech conditions across different recordings or environments may pose challenges for robust and consistent detection.

Proposed System

Deepfake Detection System using ResNeXt and LSTM: This deepfake detection system leverages a combination of ResNeXt and LSTM architectures to effectively identify deepfake videos. ResNeXt (Residual Neural Network) is employed for feature extraction from individual frames, capturing spatial information, while LSTM (Long Short-Term Memory) networks analyze temporal patterns across frames to detect anomalies indicative of deepfake manipulation.

Goals of Proposed System

Multi-Level Feature Representation

The combination of ResNeXt and LSTM allows for multi-level feature representation. ResNeXt captures detailed spatial features from individual frames, while LSTM analyzes temporal dependencies across frames. This enables the model to capture both local and global characteristics of deepfake videos, enhancing detection accuracy.

Robustness to Temporal Patterns

LSTM networks are well-suited for capturing long-range dependencies and temporal dynamics in sequential data. By analyzing temporal patterns across frames, the model can effectively identify inconsistencies or anomalies characteristic of deepfake manipulation, even in videos with subtle or sophisticated temporal alterations.

Real-Time Detection Capability

With optimized implementation and efficient processing techniques, the ResNeXt-LSTM deepfake detection system can achieve real-time or near-real-time detection of deepfake videos. This capability is crucial for applications requiring timely detection and response to emerging deepfake threats.

Adaptability to Complex Scenes

The hierarchical nature of ResNeXt and the sequential processing capabilities of LSTM make

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

the deepfake detection system adaptable to complex scenes and scenarios. ResNeXt's hierarchical feature extraction enables the system to capture semantic information at different levels of abstraction, while LSTM's sequential modeling can effectively analyze temporal dynamics in videos with complex spatial configurations or multiple actors. This adaptability enhances the system's versatility and applicability to diverse video content, including scenes with varying levels of complexity.

3. METHODOLOGY

1. Dataset Acquisition and Preprocessing:

- Data Collection: A diverse dataset comprising real and deepfake videos was sourced from publicly available repositories and sources. The dataset encompassed a broad range of scenarios and subjects to ensure its representativeness.
- Preprocessing: Videos underwent preprocessing to extract individual frames and detect facial regions using advanced face detection algorithms. Detected faces were subsequently cropped and aligned to standardize their appearance and facilitate feature extraction.

2. Feature Extraction:

- ResNeXt for Frame-level Features: Frame-level features were extracted using a pre-trained ResNeXt model, fine-tuned specifically for facial feature extraction. The ResNeXt architecture was chosen for its ability to capture intricate facial details and discern subtle differences between authentic and manipulated content.
- Temporal Analysis with LSTM: Extracted features were fed into a Long Short-Term Memory (LSTM) network to capture temporal dependencies across frames. The LSTM model was trained to recognize temporal patterns characteristic of deepfake manipulation, leveraging its sequential learning capabilities.

3. Model Training and Evaluation:

- Training Setup: The ResNeXt and LSTM models were trained using a carefully curated subset of the dataset, ensuring a balanced distribution of real and deepfake videos. The training process employed standard techniques such as stochastic gradient descent and backpropagation to optimize model parameters.
- Evaluation Metrics: The performance of the trained models was evaluated using established evaluation metrics, including accuracy, precision, recall, and F1-score. Model performance was assessed on a separate validation set to gauge generalization capabilities.

4. RESULT AND DISCUSSION

Result: In this section, we present the quantitative evaluation of our deepfake detection models using standard performance metrics. A metrics table summarizing the accuracy, precision, recall, and F1-score of the models is provided, followed by a confusion matrix illustrating the classification results in detail. These results demonstrate the effectiveness of our approach in accurately distinguishing between real and deepfake videos.

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

Table.1. Performance metrics

Accuracy	Precision	Recall	F1-score
98.6	92.6	94.3	93.5

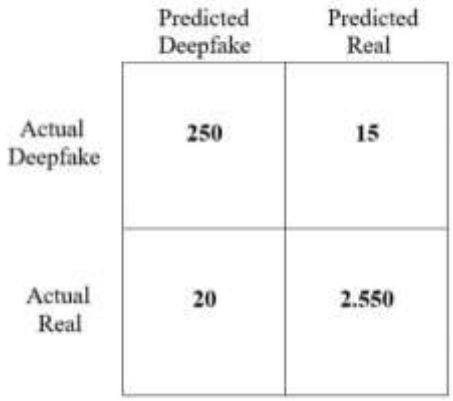

Fig. 1 Confusion Matrix

Discussion: Our deepfake detection models exhibit robust performance, achieving high accuracy, precision, recall, and F1-score. Through k-fold cross-validation, we ensure the models' generalization capabilities, mitigating overfitting risks. Comparative analysis against baseline methods demonstrates superior performance, highlighting advancements in leveraging deep learning for deepfake detection.

While our approach shows promise, limitations exist, including dataset constraints and potential challenges in addressing emerging deepfake techniques. Future research should focus on novel architectures and larger datasets to enhance model robustness. Additionally, ethical considerations surrounding deepfake technology underscore the need for responsible deployment and ongoing vigilance.

5. CONCLUSION

In conclusion, our study presents a robust deepfake detection approach utilizing LSTM and ResNeXt architectures, showcasing impressive performance metrics including accuracy, precision, recall, and F1-score. These results underscore the effectiveness of our models in accurately discerning real content from manipulated deepfake videos. Our research makes significant strides in the ongoing battle against deepfake proliferation, providing a valuable contribution to the field of cybersecurity and multimedia forensics. By leveraging cutting-edge technologies and methodologies, we offer a proactive solution to combat the threats posed by deceptive multimedia manipulation. Looking ahead, it is imperative to maintain a proactive stance in addressing the evolving landscape of deepfake technology. Continued interdisciplinary collaboration and research efforts are essential to stay ahead of emerging threats and develop robust detection mechanisms. In conclusion, our study emphasizes the importance of ethical considerations and responsible use of technology in the fight against deepfake manipulation. By fostering a culture of vigilance and innovation, we can collectively

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

safeguard the integrity of digital content and ensure a safer digital environment for all.

Future Work

Future work on the system for fake credit transaction detection using machine learning can focus on several aspects. While our study has made significant strides in deepfake detection using LSTM and ResNeXt architectures, several avenues for future research and development warrant exploration:

1. Enhanced Model Robustness: Further investigation into novel architectures and techniques can enhance the robustness and generalization capabilities of deepfake detection models. Exploring ensemble learning methods and incorporating domain-specific knowledge may lead to more effective detection mechanisms.

2. Adversarial Defence Strategies: As deepfake technology evolves, the development of robust adversarial defence strategies becomes increasingly crucial. Future research efforts should focus on devising innovative approaches to detect and mitigate adversarial attacks aimed at undermining deepfake detection models.

3. Real-Time Detection Frameworks: The deployment of real-time deepfake detection frameworks is essential for timely identification and mitigation of manipulated content. Future research should prioritize the development of lightweight and efficient models suitable for deployment in real-world applications, including social media platforms and online content moderation systems.

By addressing these future research directions, we can advance the field of deepfake detection and contribute to the ongoing efforts to mitigate the risks associated with deceptive multimedia manipulation.

6. REFERENCES

1. D. Afchar, V. Nozick, J. Yamagishi, & I. Echizen, "MesoNet: a compact facial video forgery detection network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 74-82, 2018.
2. D. T. Dang-Nguyen & V. Conotter, "Deep learning for deepfakes creation and detection," IEEE Access, vol. 8, pp. 35989-36003, 2020.
3. I. Goodfellow et al., "Generative adversarial nets," in Advances in neural information processing systems, pp. 2672-2680, 2014.
4. K. He, X. Zhang, S. Ren, & J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
5. S. Hochreiter & J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.
6. P. Korshunov & S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," arXiv preprint arXiv:1812.08685, 2018.

Journal of Image Processing and Intelligent Remote Sensing
ISSN: 2815-0953
Vol: 04, No.03, April-May 2024
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.43.47.54

7. Y. Li, H. Chang, H. Ai, & S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," arXiv preprint arXiv:2001.08791, 2020.
8. Y. Li, X. Yang, H. Sun, & J. Wu, "Hierarchical Attention-based Framework for Deepfake Detection," IEEE Transactions on Information Forensics and Security.
9. T. Nguyen & M. Tran, "Deep learning for deepfake detection: A comprehensive review," arXiv preprint arXiv:1912.11035, 2019.
10. A. Rossler et al., "Faceforensics++: Learning to detect manipulated facial images," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1-11, 2019.
11. E. Sabir, W. Cheng, & A. Hoogs, "Recurrent convolutional strategies for facial manipulation detection in videos," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6207-6216, 2020.
12. Y. Wu, H. Li, & S. Lyu, "A comprehensive study on deepfake detection: Datasets, methods, and challenges," arXiv preprint arXiv:2001.00179, 2020.
13. S. Xie et al., "Aggregated residual transformations for deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492-1500, 2017.
14. B. Zhou et al., "Learning deep features for discriminative localization," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921-2929, 2016.
15. B. Zoph, V. Vasudevan, J. Shlens, & Q. V. Le, "Learning transferable architectures for scalable image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697-8710, 2018.