IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PixelTruth: AI-Powered Deepfake Forensic Analyzer

Suresh V. Reddy[1], Prof. Harshali Bodkhe[2], Swaraj Kedari [3], Durvesh Shinde[4], Rahul Sutar[5],Sumedh Hajare [6]

[1,2,3,4,5,6] Department of Computer Engineering, SRTTC Kamshet, Pune, India

ABSTRACT

Deepfakes have emerged as a significant threat in today's digital age, enabling the creation of highly realistic manipulated videos and images that are difficult to identify without special tools.

These fake media can lead to issues like spreading false information, fraud, political manipulation, and reduced confidence in digital evidence. PixelTruth introduces an AI-driven forensic tool that uses advanced machine learning and deep learning techniques to detect and expose deepfakes. It uses CNN models, frequency analysis, and mixed feature extraction methods to spot inconsistencies in faces, lip movements, and texture patterns. This system can accurately and quickly detect deepfakes and is useful in journalism, law enforcement, and content moderation.

Keywords— Deepfake, Forensic Analysis, Artificial Intelligence, Machine Learning, CNN, Digital Trust

I. INTRODUCTION

The advancement of deep learning has made it possible to create very realistic fake media known as deepfakes.

Though this technology has positive uses in entertainment and education, it also brings up serious ethical and security issues. Deepfakes have been used for misinformation, political propaganda, identity theft, and online harassment. Traditional methods of detecting these fakes are not sufficient due to the complexity of modern generating models. There is a pressing need for AI-based forensic systems that can detect deepfakes quickly and accurately.

II. LITERATURE REVIEW

Many studies have looked into detecting deepfakes using machine learning and deep learning methods.

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been effective in identifying inconsistencies in time and space. Researchers have also looked into frequency-based detection methods that use anomalies in compression artifacts. Transformers such as Vision Transformers (ViT) and multimodal approaches that combine audio and visual cues have recently attracted attention. Despite these advancements, challenges remain in applying these methods to new types of deepfake generation, making them robust against attacks, and ensuring they work in real-time.

III.METHODOLOGY

3.1 Data Collection

We used publicly available datasets such as FaceForensics++, DFDC, and Celeb-DF for training and evaluation.

3.2 Preprocessing

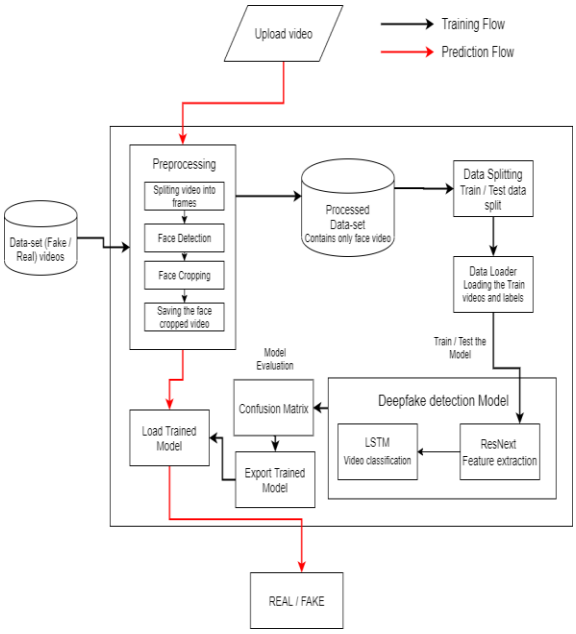
We extracted video frames and normalized them. We used MTCNN to detect facial regions and aligned them to ensure consistency.

3.3 Model Development

We built CNN architectures such as EfficientNet and XceptionNet for detecting spatial anomalies. Frequency domain analysis was used to find inconsistencies related to compression. A hybrid model combining CNN and LSTM was used to capture both spatial and temporal features.

3.4 3.4 System Architecture

The system allows for real-time processing and uses a streaming pipeline to take in video data. The system uses TensorFlow for training and can be deployed using Docker and Kubernetes.



IV. EXPERIMENTAL SETUP

We tested PixelTruth using standard datasets and real-time video feeds. The setup included:

- Data Sources: FaceForensics++, DFDC dataset, and Celeb-DF.
- Preprocessing Tools: OpenCV, Dlib, and TensorFlow preprocessing pipelines.
- Model Training: CNN, LSTM, and hybrid models trained with the Adam optimizer and early stopping.
- Validation: Five-fold cross-validation for accurate performance evaluation.
- Metrics: Accuracy, Precision, Recall, F1-Score, AUC, and Latency.

V. RESULT

The PixelTruth system showed the following results:

- CNN (XceptionNet): 92.4% Accuracy, 91.8% F1-Score
- CNN+LSTM Ensemble: 95.7% Accuracy, 95.3% F1-Score

- Frequency-based Detection: 90.1% Accuracy
- Average real-time latency: ~320 ms per frame

These results show strong performance, especially from the ensemble model, which was effective against various types of deepfake creation techniques.

VI. DISCUSSION

The evaluation shows that combining AI methods is effective for detecting deepfakes. The ensemble model successfully detected both spatial and temporal changes, helping to spot subtle manipulations. Using features from the frequency domain helped to detect artifacts caused by compression. However, ongoing challenges involve dealing with new types of deepfake creation and attacks on existing models. Future work should focus on integrating data from multiple sources, improving security against attacks, and introducing explainable AI for better understanding in forensic contexts.

VII. CONCLUSION AND FUTURE WORK

7.1 Conclusion

PixelTruth is an AI-based forensic tool that can accurately detect deepfakes with low latency. By combining CNNs, LSTMs, and frequency-based techniques, the system performed well on standard datasets and real-time situations. Its use in journalism, law enforcement, and digital content verification makes it a useful tool in fighting false information.

7.2 Future Work

- Create datasets in multiple languages and across cultures for wider use.
- Extend the analysis to include different types of multimedia inputs, such as audio and video consistency.
- Implement adversarial training to strengthen the system's robustness.
- Explore explainable AI techniques to improve transparency in forensic applications.

VIII. REFERENCES

[1] A. Rossler, et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," Proc. IEEE Int. Conf. Computer Vision (ICCV), 2019.

[2] B. Dolhansky, et al., "The DeepFake Detection Challenge Dataset," arXiv preprint arXiv:2006.07397, 2020.

[3] Y. Li, et al., "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020.

[4] D. Guera and E. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," Proc. IEEE Int. Conf. Advanced Video and Signal-Based Surveillance (AVSS), 2018.

[5] H. Tran, et al., "Learning Temporal Features for Deepfake Detection," Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), 2021.

[6] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.

[7] R. Tolosana, et al., "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[8] P. Zhou, et al., "Two-Stream Neural Networks for Tampered Face Detection," Proc.IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2017.

[9] S. Agarwal, et al., "Protecting World Leaders Against Deep Fakes," Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.

[10] F. Matern, et al., "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations," Proc. IEEE Winter Conf. Applications of Computer Vision Workshops (WACVW), 2019.

[11] P. Korshunov and S. Marcel, "Deepfakes: A New Threat to Face Recognition? Assessment and Detection," arXiv preprint arXiv:1812.08685, 2019.

[12] E. Sabir, et al., "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," Proc. IEEE Interfaces, 2019.

[13] A. Haliassos, et al., "Lips Don't Lie: A Generalisable Audio-Visual Method for Deepfake Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2021.

[14] S.-Y. Wang, et al., "CNN-generated Images Are Surprisingly Easy to Spot… for Now," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020.

[15] T. Jung, et al., "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020.