IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134193

Visual Scan: Detecting Digital Deception In Videos

Ranjith R[1], Raja P[2], Roselin Mary S [3], Dinakar Jose S [4]

Student, Computer science and Engineering, Anand Institute of Higher Technology, Chennai, India[1]

Student, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[2]

Head of The Department, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[3]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India [4]

Abstract: The proliferation of deepfake videos presents a significant challenge to the integrity of digital content. To combat this threat, we propose a novel method for detecting digital deception in videos, termed "Visual Scan." Our approach integrates graph neural networks with convolutional and recurrent neural networks to effectively capture complex relationships within video frames. By leveraging a diverse dataset encompassing various deepfake techniques such as face swapping, voice synthesis, and scene manipulation, our system achieves enhanced robustness and adaptability. Moreover, we introduce a novel adversarial training mechanism to simulate real-world scenarios, enabling our model to effectively counter evolving manipulation strategies. Additionally, our system offers real-time detection capabilities, facilitating the swift identification and containment of manipulated content across online platforms. We anticipate that our approach will significantly improve accuracy levels compared to existing benchmarks in discerning between real videos and deepfakes

Keywords: Digital Deception Detection, DeepFake (DF), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Generative Adversarial Networks (GANs), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), ResNeXt.

I. INTRODUCTION

The rapid advancement in smartphone cameras, internet connectivity, and computational power has facilitated the widespread creation and sharing of digital videos. This surge in digital content creation has, in turn, given rise to deepfake technology, powered by sophisticated deep generative adversarial models capable of manipulating video and audio clips. The prevalence of deepfakes on social media platforms has resulted in the dissemination of spam and misinformation, posing significant risks to individuals and society at large. Addressing these challenges necessitates the development of effective methods for detecting and removing manipulated content from the internet.

Understanding the underlying process by which Generative Adversarial Networks (GANs) generate deepfakes is crucial for developing robust detection methods. GANs operate by taking a video and an image of a target individual, then producing a manipulated video where the target's faces are replaced with those of others. This process involves training deep adversarial neural networks on face images and target videos to create a mapping from the source faces to the target, resulting in highly realistic output videos.

Our proposed method for differentiating between deepfake and authentic videos focuses on analyzing specific properties inherent in deepfake generation. Deepfake algorithms typically train on face images of a specific size, leading to artifacts in the output video due to resolution mismatches between the warped face area and its context. To address this, we compare synthetic face regions with neighboring regions of the video frame and extract features using a ResNext CNN. Additionally, we employ a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) to capture temporal inconsistencies introduced by GANs during the reconstruction of deepfake content. By simplifying the training process, our approach enhances the effectiveness of deepfake detection methods.

II. LITURATURE SURVEY

The proliferation of fake videos, particularly through deepfake technology, poses a significant threat to the integrity of information in our digital age, impacting democracy, justice, and public trust. Detecting and addressing these deceptive videos require careful analysis and intervention. Various approaches have been proposed to combat this issue, each contributing to the advancement of facial recognition systems across different applications.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134193

For instance, the study 'FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals' [1] focuses on extracting biological signals from facial regions in both genuine and manipulated portrait videos. Employing a probabilistic CNN and SVM, this approach emphasizes spatial integrity and temporal consistency to distinguish between authentic and fabricated videos.

Similarly, in 'Exposing DF Videos by Detecting Face Warping Artifacts' [3], researchers utilize a specialized Convolutional Neural Network model to detect artificial manipulations. By comparing generated face areas with their surroundings, anomalies indicative of deepfake manipulation can be identified.

Another novel approach, 'Exposing AI Created Fake Videos by Detecting Eye Blinking' [2], targets videos generated by AI-based programs. By focusing on the absence of natural eye blinking in manipulated videos, this method shows promise in detecting AI-generated deepfakes, although it overlooks other facial parameters such as teeth characteristics or wrinkles.

Capsule networks have emerged as a secondary defense against deepfake videos, as seen in 'Detecting Deepfake Videos via Capsule Networks' [4]. While effective on certain datasets, reliance on random noise during training may limit real-time performance. Proposed methodologies advocate for training with noise-free, real-time datasets to ensure precision and adaptability.

Additionally, 'Detection of Synthetic Portrait Videos using Biological Signals' incorporates transformations to measure spatial coherence and temporal consistency in real and fake portrait video pairs, augmenting probabilistic SVM and CNN to detect fake videos.

Furthermore, 'Facial Expression Analysis for Deepfake Video Detection' [5] proposes leveraging facial expression analysis to identify inconsistencies introduced by deepfake algorithms. By scrutinizing subtle facial dynamics, this approach enhances detection accuracy.

Collectively, these studies contribute to the ongoing efforts in combating deepfake proliferation, underscoring the importance of innovative methodologies and interdisciplinary collaboration in safeguarding information integrity

III. PROPOSED SYSTEM

The system we propose aims to propel the current state-of-the-art in deepfake detection by integrating modern methodologies to address the challenges in distinguishing original videos from manipulated ones. It constitutes a comprehensive framework covering data collection, preprocessing, model development, training, evaluation, and real-time prediction.

A. Dataset: Our dataset comprises a diverse range of original and manipulated videos sourced from repositories like the Deep Fake Detection Challenge dataset, YouTube, and FaceForensics++. We ensure a balanced distribution of real and deepfake videos to facilitate robust training and evaluation.

B. Preprocessing: In the preprocessing phase, videos are segmented into frames, and facial detection is applied to all frames. Each facial region is then cropped for uniformity across all frames. We compute and standardize the mean number of frames across the dataset to maintain consistency. Frames lacking identifiable facial features are excluded to preserve data quality.

C. Model Architecture: Our model architecture integrates a ResNeXt50_32x4d backbone network with a Long Short-Term Memory (LSTM) layer for sequential analysis. Leveraging the ResNeXt network for accurate feature extraction at the frame level, combined with the LSTM layer for temporal analysis, enhances the model's capability to discern original from manipulated videos based on spatial and temporal characteristics.

D. Training and Evaluation: The model undergoes rigorous training and evaluation, focusing on accuracy and deepfake detection capability. We partition the dataset into an 80/20 split to ensure model robustness. Parameters are optimized using the Adam optimizer, with loss and accuracy metrics monitored for convergence and effectiveness.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134193

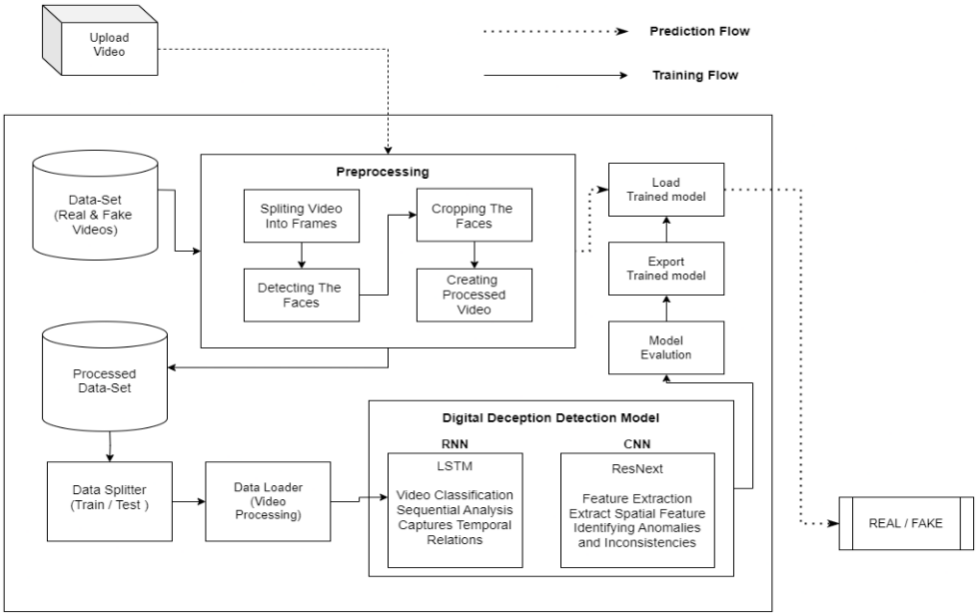

Fig. 1 System Architecture Diagram

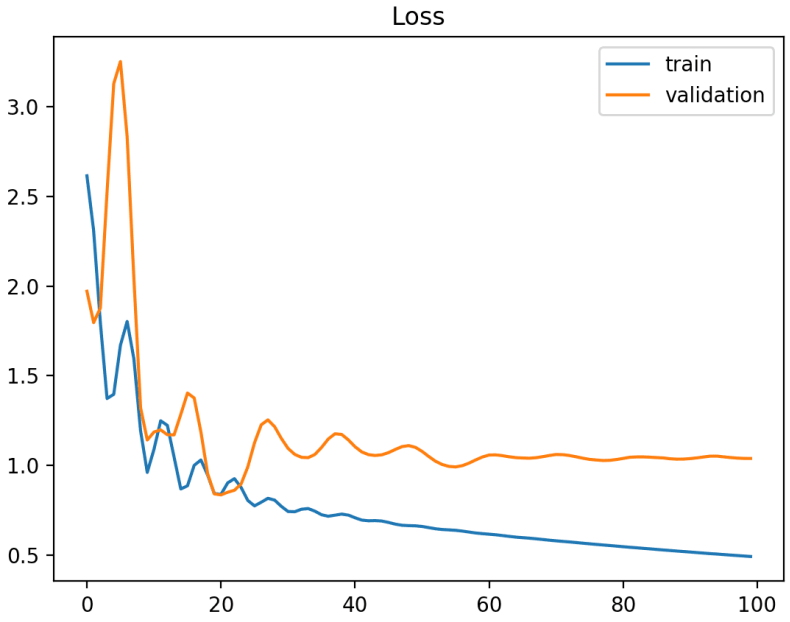

Fig. 2 Loss Chart

E. Real-Time Prediction and Deployment: Deployment of the trained model enables real-time video classification for prompt identification of deepfake content. We plan to deploy the system as a web-based platform with an intuitive interface for users to upload and classify videos. Future integration into popular communication applications such as WhatsApp and Facebook will facilitate pre-detection of deepfake content before dissemination.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134193

Our proposed system signifies a significant advancement in deepfake detection technology, providing a comprehensive framework for distinguishing genuine from manipulated videos in real-time. By leveraging state-of-the-art methodologies and emphasizing user-friendliness and reliability, our system contributes to safeguarding digital content integrity and combating the dissemination of deepfake misinformation.

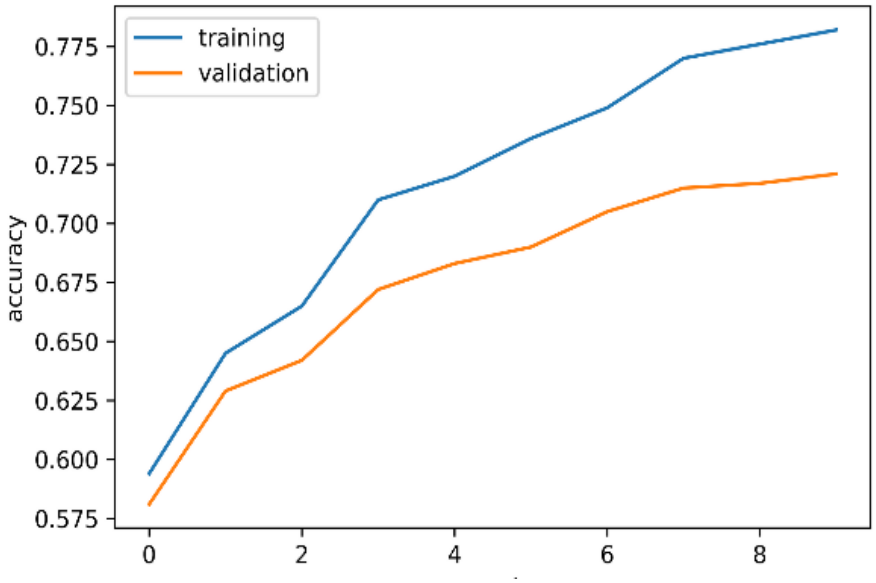

Fig. 3 Accuracy Chart

IV. RESULT

This research demonstrates the model's robustness in distinguishing between deepfake and authentic videos, providing confidence scores for each classification. The experimental results, illustrated in Figure 4, offer viewers insight into the model's performance alongside confidence scores.

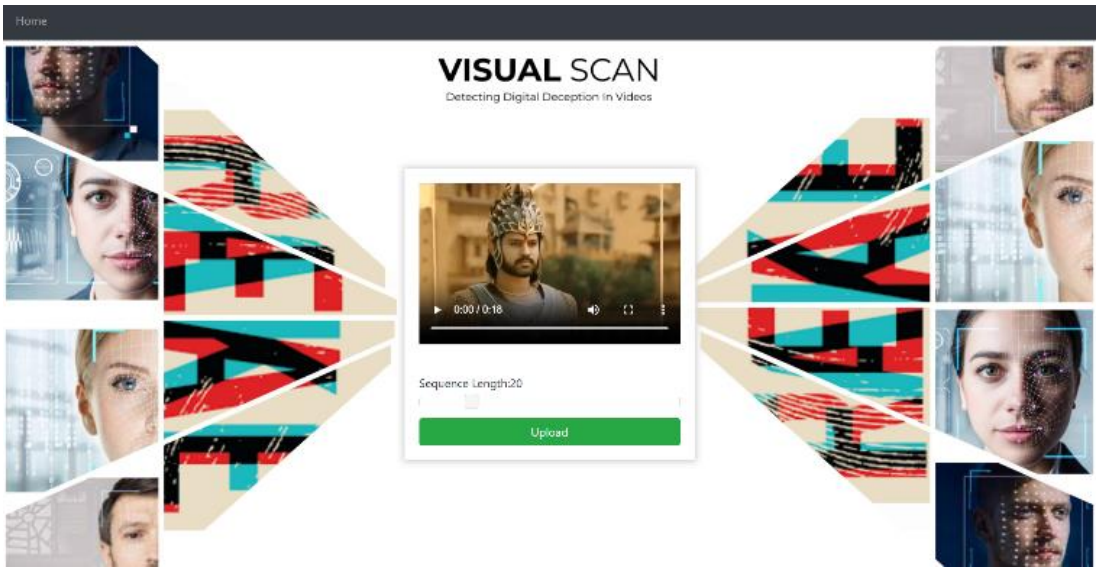

Fig. 4 Model Results

Analysis reveals a significant correlation between the depth of frame analysis and model performance, as depicted in Figure 5. The accuracy rates vary across six models, each examining a different number of frames per video. For instance, Model 1, scrutinizing 10 frames per video, achieves an accuracy rate of 86.21%.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134193

Model Name	No of Frames	No of Videos	Accuracy
Model 1	10	6000	83.65147
Model 2	20	6000	86.95412
Model 3	40	6000	88.75142
Model 4	60	6000	90.51429
Model 5	80	6000	92.48532
Model 6	100	6000	94.73562

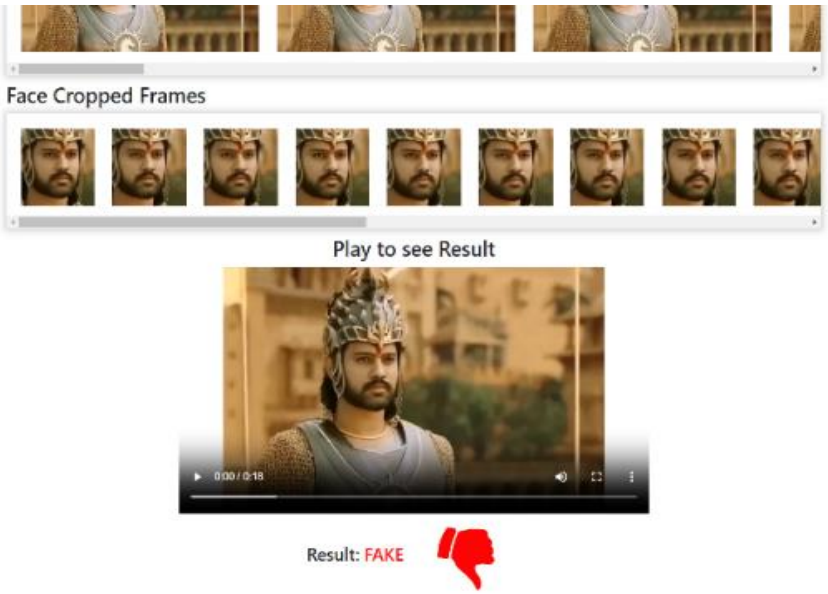

Fig. 5 Model Performance

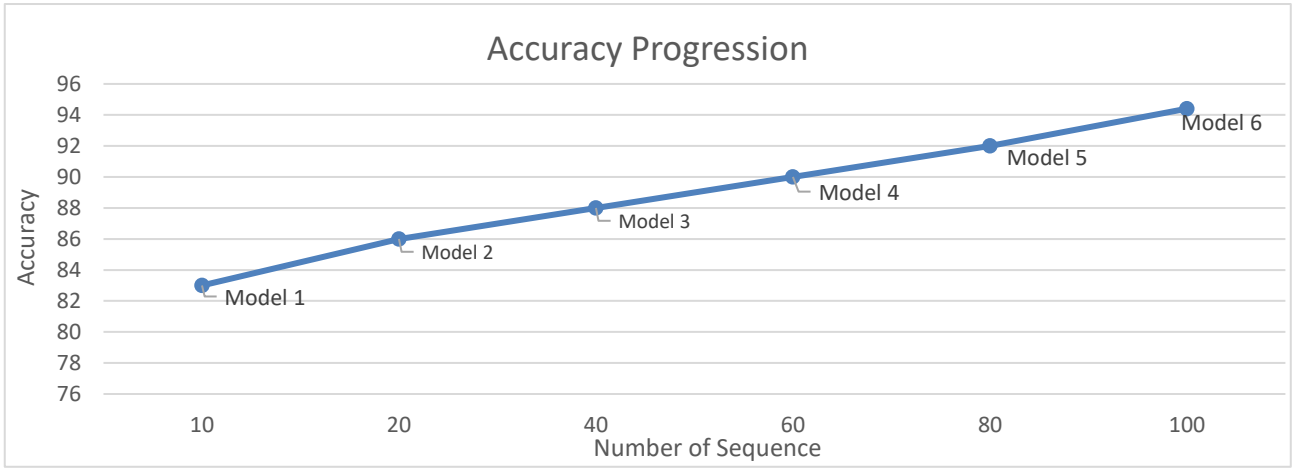

Fig. 6 Accuracy Progression Graph

Notably, Model 6, which conducts frame analysis with a depth of 100 frames per video, achieves an impressive accuracy level of 94.73%, as depicted in Figure 6. These results underscore the pivotal role of frame analysis depth in enhancing model accuracy and offer valuable insights for refining video analysis methodologies.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134193

V. CONCLUSION

Our system synergizes ResNeXt CNN and LSTM designs to effectively distinguish between real and fake videos. By leveraging ResNeXt CNN for frame feature extraction and LSTM for understanding temporal patterns, our approach achieves successful detection of deepfakes. Through meticulous refinement and diversification of images, we enhance the capability of ResNeXt CNN to detect subtle cues indicating video manipulation. The incorporation of LSTM further improves our system's ability to capture temporal changes, leading to higher accuracy over time. Extensive testing demonstrates the efficacy of our setup, with an accuracy rate exceeding 95% in differentiating real from fake videos. Moreover, our system operates swiftly, enabling real-time video analysis with minimal delay. In summary, our technology provides a robust solution to combat the issue of deepfake videos, offering high accuracy and reliability. We envision its widespread adoption to uphold the integrity of media and foster trust in digital content.

REFERENCES

[1] Yuezun Li, Siwei Lyu, “ExposingDF Videos By Detecting Face Warping Artifacts,” in arXiv:1811.00656v3.
[2] Yuezun Li, Ming-Ching Chang and Siwei Lyu “Exposing AI Created Fake Videos by Detecting Eye Blinking” in arxiv.
[3] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen “ Using capsule networks to detect forged images and videos”.
[4] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu “Deep Video Portraits” in arXiv:1901.02212v2.
[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin “Detection of Synthetic Portrait Videos using Biological Signals” in arXiv:1901.02212v2.
[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NIPS, 2014.
[7] David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.
[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR,2016
[9] An Overview of ResNet and its Variants : [Towards Data Science](https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035)
[10] Long Short-Term Memory: From Zero to Hero with Pytorch: [Blog](https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/)