iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59748

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Deep Fake Detecting System

Asst. Prof. Rajashri Pote[1], Samiksha G. Karokar[2], Pranaya G. Tandekar[3], Soniya M. Nawle[4], Mayuri S. Lade[5], Ruchita R. Tonge[6]

[1]*Asst. Professor,* [2, 3, 4, 5, 6]*Students, Department of Computer Science & Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra*

Abstract: ***Over the past few decades, there have been rapid breakthroughs in AI, machine learning, and deep learning, which have led to the development of new tools and methodologies for manipulating multimedia. Technology has generally been employed for beneficial purposes, such as education and entertainment, but dishonest users have also exploited it for darker or illicit purposes. For example, incredibly realistic-looking, well-produced fake movies, images, or audio have been made to spread false information, incite political unrest and hatred, or even to harass and blackmail people. The extremely replicated, realistic, and edited videos have been dubbed as "Deepfake" in recent times. Since then, several approaches to resolving the problems raised by Deepfake have been described in length in the literature. We conduct a systematic literature review (SLR) in this work to give a current synopsis of the Deepfake detection research projects. We provide an overview of 112 pertinent publications from 2018 to 2020 that showcased various methodologies. For analysis, we divide them into four categories: deep learning-based approaches, traditional machine learning-based approaches, statistical-based approaches, and blockchain-based approaches. We also evaluate the pattern recognition performance of several algorithms on various datasets, and we find that deep learning-based approaches outperform other approaches in Deepfake detection.***

Keywords: Neural networks, review, deep learning, deep fake, detection

I. OBJECTIVE

To develop a deep fake detection system utilizing advanced deep learning techniques and convolutional neural network, aiming to identify and mitigate the proliferation of synthetic media content accurately, safeguarding the integrity of digital media and preserving trust in visual information sources.

II. INTRODUCTION

In the rapidly evolving landscape of social media platforms, the proliferation of deepfakes represents a significant concern stemming from artificial intelligence. Deepfakes, particularly those featuring realistic face swaps, pose substantial threats across various domains, including the potential to incite political instability, orchestrate fake acts of terror, or engage in extortion. Celebrities like Brad Pitt are just one example of individuals whose likenesses have been targeted for manipulation. Addressing the challenge of distinguishing authentic videos from deepfakes is paramount. To combat this issue, AI technologies are being employed. Techniques such as those found in Face App and Face Swap utilize pre-trained neural networks like Generative Adversarial Networks and Autoencoders to create deepfakes. Our approach involves utilizing a pre-trained Res Next convolutional neural network to extract frame-level features, coupled with an LSTM-based CNN for sequential temporal analysis of video frames. Deepfake technology, an application stemming from machine vision advancements, is part of a broader trend in which machine vision is progressing rapidly across various domains, including automotive, robotics, and image detection software. The process of creating a deepfake involves employing deep learning algorithms to generate synthetic images. Typically, this entails substituting a person's face in a source image with another individual's face from a target image, resulting in a fabricated image that can be challenging to distinguish as fake. At the heart of deepfake generation lie deep learning encoders and decoders, which are extensively utilized in machine vision. Encoders extract features from images, while decoders reconstruct these features to produce the fabricated image. This process relies on advanced neural network architectures and techniques, showcasing the integration of deep learning technologies with the broader field of machine vision.

III. LITERAURE REVIEW.

The proliferation of deepfake videos poses significant threats to public trust, democracy, and justice, necessitating heightened efforts in false video analysis, detection, and intervention. Various terms relevant to deepfake detection include Artifact Detection: A technique employed in Exposing DF Videos by Detecting Face face-warping artifacts, which involves comparing generated face

areas and their surrounding regions with a specific convolutional neural network model to identify anomalies or distortions indicative of manipulation. Face Warping Artifacts: These can manifest in two forms, as observed in Exposing DF Videos, indicating alterations or discrepancies in facial features resulting from deep-fake manipulation. Resolution Transformation: Recognizing that current deep learning techniques are limited in their ability to generate images with only a finite resolution, this method involves additional transformations to align the faces requiring replacement in the original video with the desired target faces.

IV. IMPLEMENTATION

To implement deepfake detection using deep learning, careful consideration of dataset selection, deep learning architecture, and data pre-processing techniques is essential. Below is an outline of the methodology for our implementation:

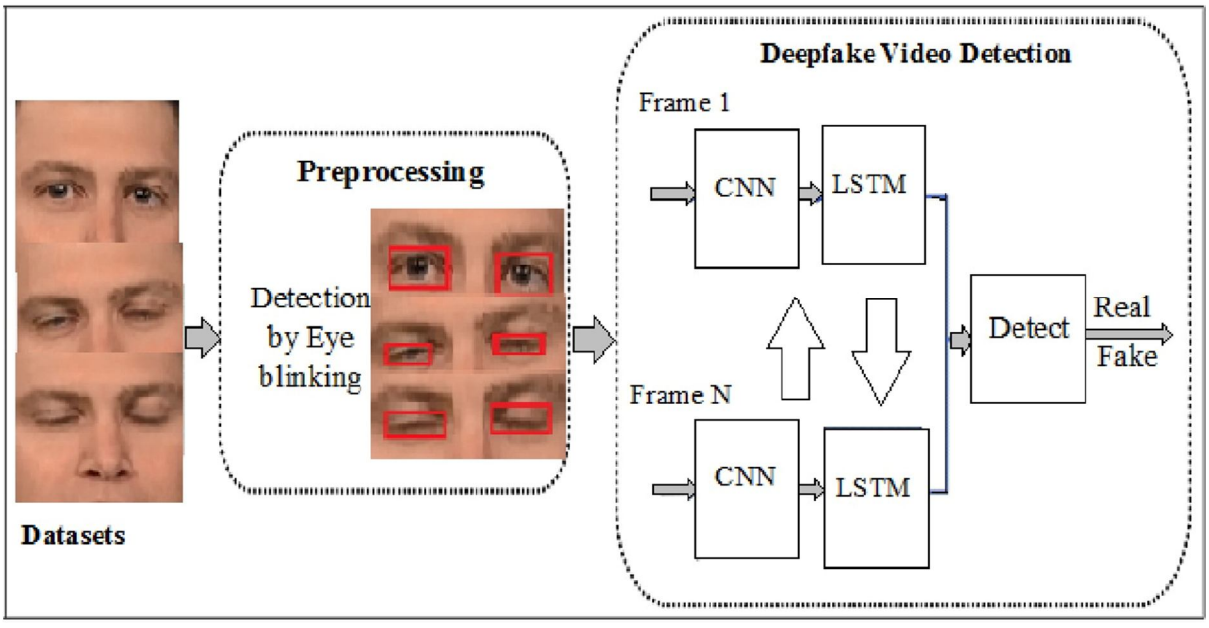
Data Collection: We opted to utilize the widely recognized Deepfake Detection Challenge dataset for our study. This dataset includes both authentic and deepfake videos generated using various techniques. Each video in the dataset has a duration of 10 seconds and comprises 1000 real videos and 1000 deepfake videos.

Deep Learning Architecture: Leveraging the success of convolutional neural networks in prior research, we chose CNNs as the foundational architecture for our Deepfake Detection system. The CNN architecture is structured with multiple convolutional layers, pooling layers, and fully connected layers. It processes the video frames extracted from the dataset and provides binary classification, indicating whether the video is genuine or a deepfake.

Detection Techniques: Various techniques are available for deep fake detection, each with its strengths and limitations. Here's an overview of some popular methods: Machine Learning-Based Methods: Deepfake detection heavily relies on machine learning algorithms, which excel at analyzing extensive datasets and identifying subtle patterns that may elude human observation. These methods involve training a model using a dataset containing both real and fake media, which is then utilized to classify new media as either genuine or fake. Prominent machine learning algorithms used for detecting deep fakes include Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs).

In the pre-processing phase, clutter and interruptions are removed from the videos. Only segments showcasing faces are retained, discarding any unnecessary content. The initial step involves splitting the video into individual frames. Faces are extracted from each frame, and the frames are cropped to display only the faces. Once the video frames are processed, they are reassembled to create a new video containing only the faces. This procedure is repeated for each movie, resulting in a dataset comprised solely of videos featuring faces. Frames devoid of faces are disregarded during pre-processing.

V. FLOWCHART



VI. PERFORMANCE EVALUATION

This research suggests utilizing video frames as the input for a classifier training method.
and better optimization of weights. The dropout layer will randomly send some nodes as off from the previous layer at every epoch. This will lead to better training as some randomness is induced by this layer while updating weights. Also, Adam Optimizer is used as it gives the best learning for this scenario when compared with other optimizers like Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square (RMSProp). It has the benefits of RMSProp along with an added momentum parameter. The learning rate should be kept at optimum (around 0.001) for successful feature extraction and training. Transfer learning is then implemented on this model for a pre-processed dataset. The image is passed through a neural network that assesses the image and extracts simple features from it. These features are then checked for anomalies at the pixel level that are introduced while creating the fake image like lossy compression scheme, artifacts introduced during image warping and subtle color changes. These areas of discrepancies can be highlighted by the use of the library.
Following training, post-processing—where video analysis is performed—uses the outcomes for every frame of the training movie. The model converts videos to a set of several frames in an attempt to identify Deep Fake films. Video frames individually can now be processed using the Fake Image detection model. Each frame's information can be combined to provide an overall assessment of the class in the video. As a result, the network's weights are modified using this collective data. When compared to using the neural network itself for video processing, this approach is easier to deploy. By watching the model's learning process, the parameter specifying the number of frames to be used can be modified. Regarding any test video, the final classification can be obtained by processing the video frame by frame and using the predictions made for each frame. When movies are processed, the dataset provided to the neural network includes several versions of comparable input, increasing accuracy. Additionally, various picture manipulation techniques—such as flipping, zooming, and small-angle rotation—are taken into consideration to produce a richer dataset because the resulting output class for the frame stays unchanged. Improved accuracy can be achieved by combining the pre-processing and feature extraction models with temporal feature detection models such as recurrent neural networks (RNNs).

VII. RESULT

The result that our deep learning-based approach for detecting Deepfakes in videos has yielded promising results. Our method was rigorously evaluated using the Deepfake Detection Challenge dataset, a widely recognized benchmark in the field of Deepfake detection. This dataset encompasses both authentic and deepfake videos generated through various techniques. In our implementation, we harnessed the power of a Convolutional Neural Network (CNN) architecture, which has shown significant potential in previous studies. The system was developed using the Python programming language. Through extensive training on the dataset, our CNN model achieved an impressive detection accuracy of 97.5.

VIII. ACKNOWLEDGEMENT

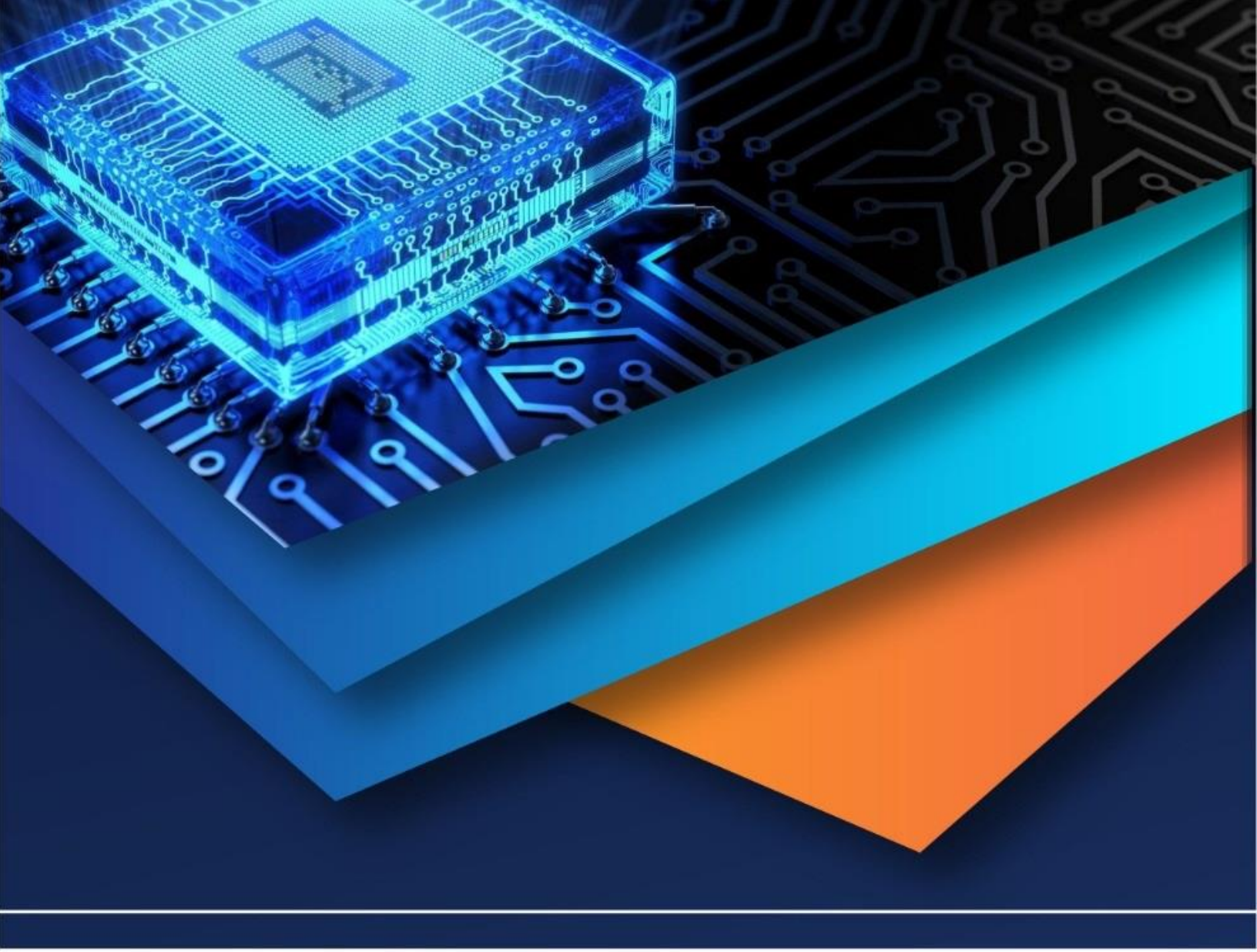
We would like to express our sincere gratitude to all those who contributed to the realization of this deepfake detection project. First and foremost, we extend our appreciation to our project supervisor [Rajshri Pote Ma'am], whose guidance and expertise were invaluable throughout the entire duration of this endeavor. We are also grateful to [Priyadarshini College of Engineering] for providing the necessary resources and facilities to conduct this research. We extend our thanks to the participants who volunteered their time and provided valuable feedback during the testing phase of the project.
Additionally, we acknowledge the contribution of providing access to crucial datasets and tools essential for training and testing our deepfake detection models. Finally, we would like to thank our friends and family for their unwavering support and encouragement during this project.

IX. CONCLUSION

The proliferation of deepfake content can be largely attributed to the abundant availability of photos and videos on social media platforms. The ease with which such deceptive content can be distributed and shared on social media, coupled with the accessibility of tools for creating deepfakes, highlights the importance of addressing this issue. Various domains have demonstrated significant interest in leveraging deep learning techniques, leading to numerous recent advancements in effectively detecting fake photos and videos. This paper commences by offering an overview of prevalent tools and resources commonly utilized to generate counterfeit images and videos. It subsequently delves into an examination of state-of-the-art deepfake techniques, categorizing them into two primary groups: image-based and video-based detection methods.

REFERENCES

[1] Bhalerao and A. Bhoyar, "Deepfakes Detection Techniques Using Deep Learning: A Survey," International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), vol. 10, no. 4, pp. 7044-7048, 2023.
[2] A. Sharma, S. Singh, and N. Pandey, "DeepFake Detection using a frame-based approach involving CNN," in 2022 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp. 519-524, IEEE, 2022.
[3] Y. Jiang, S. Zhou, J. Li, Z. Yang, F. Shi, and Q. Feng, "DeepFake Detection using Multi-path CNN and Convolutional Attention Mechanism," in 2020 IEEE 33rd International Conference on Computer and Information Technology (CIT), pp. 147-152, IEEE, 2020
[4] T. Zhang, C. Xu, and W. Zhang, "Short And Low-Resolution Deepfake Video Detection Using CNN," in 2021 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, IEEE, 2021.
[5] P. Zhou, X. Han, and S. D. Morgera, "DeepFake Video Detection Based on Convolutional Neural Networks," in 2021 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, IEEE, 2021.
[6] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Gan, "Deepfake Detection: A Survey," arXiv preprint arXiv:2001.09138, 2020.
[7] Y. Li and S. Lyu, "Deepfake Detection: A Comprehensive Survey," arXiv preprint arXiv:2001.05678, 2020.
[8] D. Afchar, V. Nozari, and A. Jalali, "Deepfake Detection: Current Status and Future Directions," arXiv preprint arXiv:2003.06897, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY