



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

AI into Multimedia Deepfakes Creation and Detection

Nikhil Dhanaji Chavan¹, Tejas Pramod Sonawane², Kedar Bhikaji Mane³,

Sanket Devidas Mulmule⁴, Prof. S.B.Nimbekar⁵

U.G. Students, Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, Maharashtra, India¹⁻⁴

Professor, Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, Maharashtra, India⁵

ABSTRACT: This paper presents an innovative technique for detecting digitally manipulated media through advanced deep vision methods. The proposed system harnesses the capabilities of Convolutional Neural Networks (CNNs) alongside Recurrent Neural Networks (RNNs) to examine both spatial and temporal aspects of multimedia data. In recent times, the availability of free, deep learning-based software has significantly simplified the creation of highly realistic face-to-face video interactions, commonly referred to as “Deep Fake” (DF) videos—essentially, digitally deceiving the viewer. Although visual effects have enabled video manipulation for years, recent advancements in deep learning have vastly increased the authenticity of fake content and lowered the barrier to its creation. Our research has made modest but meaningful progress through the integration of CNNs and RNNs for detecting deepfake networks. CNNs are employed to extract frame-level visual features, which are then used to train an RNN. This RNN is capable of determining whether a video has been tampered with by detecting temporal inconsistencies between frames—often introduced by deepfake generation tools.

KEYWORDS: Deepfake Video Detection, convolutional Neural network (CNN), recurrent neural network (RNN).

I. INTRODUCTION

The evolution of sophisticated image and video manipulation technologies has significantly reshaped the landscape of digital media creation and consumption. Today, advanced tools enable seamless alteration of visual content with remarkable realism. Among these, deepfake technology driven by artificial intelligence and machine learning has emerged as a leading method for generating lifelike images and videos that imitate real individuals and events. This capability is primarily powered by **Generative Adversarial Networks (GANs)**, which involve two neural networks: one that creates synthetic content and another that evaluates its authenticity, iteratively improving the output's realism. Such technological advancements have made it easier and faster to produce manipulated media that is nearly indistinguishable from authentic material. While these innovations open new avenues in fields like entertainment and advertising—where they can be used to generate virtual characters or recreate historical personalities for educational purposes—they also pose significant risks. The same methods used for creative applications are often exploited for malicious purposes, including spreading misinformation, political propaganda, social manipulation, and personal defamation. The widespread availability of these tools, many of which require little technical skill, has accelerated the dissemination of digital deception and undermined public confidence in media. As the boundary between real and fabricated content becomes increasingly blurred, it becomes essential to develop effective detection systems. This paper focuses on investigating how the integration of various deep learning techniques can contribute to the development of more intelligent, adaptive, and context-aware systems for distinguishing digitally manipulated images from authentic ones.

II. MOTIVATION

Detecting digitally manipulated media relies on a multifaceted approach that combines the analysis of visual artifacts with the identification of temporal inconsistencies. Modern detection systems are designed to uncover subtle anomalies that may indicate tampering. These include irregular eye movements, unnatural facial expressions, inconsistent lighting, and flawed rendering of facial features—imperfections that may be imperceptible to the human eye but detectable by machine learning models. One key aspect of this detection process is the analysis of audio-visual synchronization.

Algorithms are trained to recognize discrepancies such as lip movements that don't align with the spoken audio, mismatched voice tones, or unnatural pauses, all of which can signal that a video has been manipulated. Additionally, temporal analysis plays a vital role in spotting frame-to-frame inconsistencies like flickering, jittery transitions, and abrupt changes in facial geometry or expressions, which often result from frame interpolation errors during the synthesis process. Deep learning models, especially Convolutional Neural Networks (CNNs), serve as the foundation for many detection systems. These models are trained on vast datasets of both authentic and manipulated content, allowing them to learn high-level features and detect intricate patterns that may elude manual inspection. Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks are often integrated alongside CNNs to enhance the model's capability to track changes over time and improve accuracy in temporal analysis.

III. METHODOLOGY

Detecting digitally manipulated media relies on a multifaceted approach that combines the analysis of visual artifacts with the identification of temporal inconsistencies. Modern detection systems are designed to uncover subtle anomalies that may indicate tampering. These include irregular eye movements, unnatural facial expressions, inconsistent lighting, and flawed rendering of facial features—imperfections that may be imperceptible to the human eye but detectable by machine learning models. One key aspect of this detection process is the analysis of audio-visual synchronization. Algorithms are trained to recognize discrepancies such as lip movements that don't align with the spoken audio, mismatched voice tones, or unnatural pauses, all of which can signal that a video has been manipulated. Additionally, temporal analysis plays a vital role in spotting frame-to-frame inconsistencies like flickering, jittery transitions, and abrupt changes in facial geometry or expressions, which often result from frame interpolation errors during the synthesis process. Deep learning models, especially Convolutional Neural Networks (CNNs), serve as the foundation for many detection systems. These models are trained on vast datasets of both authentic and manipulated content, allowing them to learn high-level features and detect intricate patterns that may elude manual inspection. Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks are often integrated alongside CNNs to enhance the model's capability to track changes over time and improve accuracy in temporal analysis.

Following Steps Shows the Flow of system Design :

- 1) Data Gathering.
- 2) Data Pre-Processing
- 3) Developing Model
- 4) Model Evaluation
- 5) Load Trained Model
- 6) Output

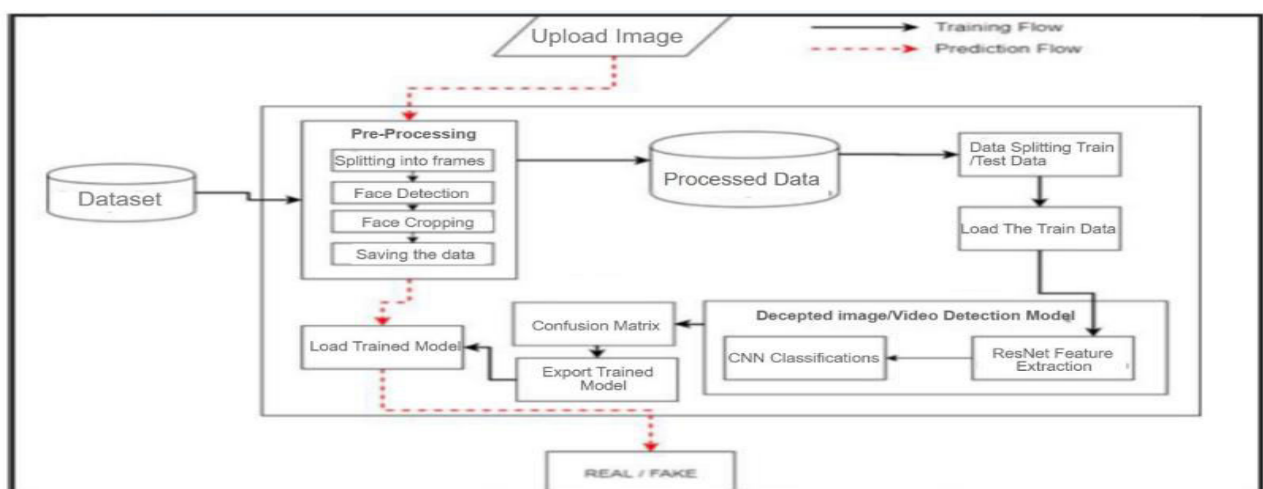


Fig. Proposed Architecture 1

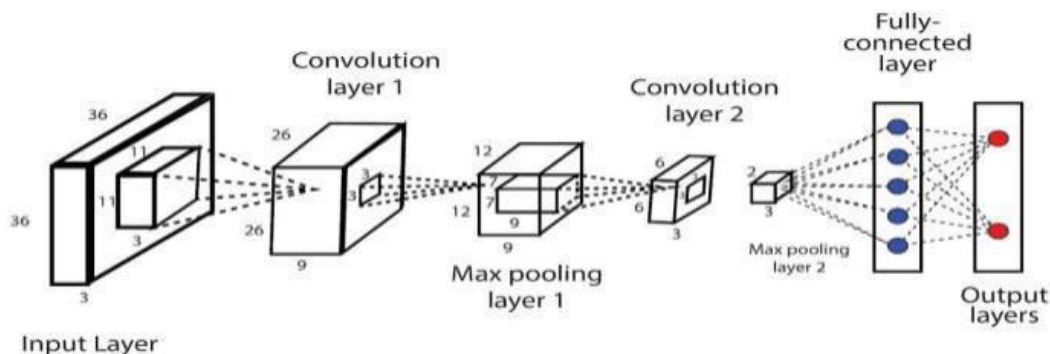
Algorithms Used are As Follows:

1) Convolutional Neural Network:

CNNs are well-suited for image-based tasks due to their ability to capture spatial hierarchies in visual data. In deepfake detection, CNNs extract high-level features from image data, capturing subtle artifacts or irregular patterns introduced during the creation of fake images or videos. CNNs are effective in identifying spatial artifacts, such as discrepancies in pixel distributions, inconsistencies in lighting, and unnatural textures or edges. Layers within CNNs (e.g., convolutional, pooling, and fully connected layers) enable the network to differentiate real from manipulated content. CNN-based models, pre-trained on large datasets like ImageNet, can be fine-tuned for deepfake detection, providing a strong foundation with significant image features already learned.

Model and Material which are used is presented in this section. Table and model should be in prescribed format.

Fig. CNN Algorithm Working 1



2) ResNet (Residual Network) :

ResNet is a CNN variant that uses "residual blocks," which allow networks to be very deep (e.g., ResNet-50, ResNet-101) without suffering from vanishing gradients. Residual blocks bypass certain layers through shortcut connections, enabling the model to learn identity mappings easily. ResNet's architecture, particularly its ability to go deeper without degradation, makes it powerful for identifying more complex features in high-resolution images. For deepfake detection, this means it can uncover intricate inconsistencies that may not be visible at shallower levels. ResNet models are highly effective in image classification and are known for their robustness and high performance, even on challenging image recognition tasks like deepfake detection. Their depth allows for greater feature extraction, making them suitable for identifying subtle signs of image manipulation.

IV. APPLICATIONS

Deepfakes have emerged as a significant means of spreading disinformation, using manipulated videos to deceive audiences. Tools designed for detection play a crucial role in identifying fake content, thereby supporting the credibility of genuine media sources. These synthetic videos are frequently weaponized against individuals, fabricating footage that can harm reputations or breach privacy. As deepfakes become increasingly widespread, they pose a threat to public trust in digital media, causing people to doubt even legitimate content. Enhancing the capacity to tell real media apart from fake helps rebuild trust in online information and protects the public from being misled.

V. CONCLUSION

We have initiated the foundational work for creating deepfake detection models aimed at recognizing artificially generated media, including images produced via deep learning methods. Although the models themselves are yet to be developed, our research has successfully entered its preliminary stages, with a well-defined approach and clearly outlined objectives to steer our future work. At present, we are engaged in a comprehensive review of existing literature, which offers critical perspectives on current techniques and highlights the challenges involved in detecting deepfakes. In summary, although our project is still in its early phase, the research strategy and framework we have put

in place are both solid and systematic. We look forward to the upcoming stages, where we will focus on building and assessing detection models—ultimately supporting the integrity and trustworthiness of digital content..

REFERENCES

- [1] DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS Siwei Lyu Computer Science Department University at Albany, State University of New York 978-1-7281-1485-9/20/\$31.00@2020 IEEE.
- [2] Deepfake Video Detection Based on Spatial ,Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning, John K. Lewis, Imad Eddine Toubal, Helen Chen .
- [3] Deepfake Video Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning 2020 IEEE Applied Imagery Pattern Recognition Workshop.
- [4] Deepfake video detection: challenges and opportunities Achhardeep Kaur¹ · Azadeh Noori Hoshyar² · Vidya Saikrishna³ · Selena Firmin¹ · Feng Xia⁴ Accepted: 16 May 2024 / Published online: 29 May 2024 © The Author(s) 2024
- [5] Deepfake Detection and the Impact of Limited Computing Capabilities Paloma Cantero Arjona, Alfonso S´anchez-Maci´a 8 Feb 2024. S. Strobel, “Streamlit: Turn Data Scripts into Shareable Web Apps”, Journal of Open Source Software, Vol. 5, No. 55, pp. 2654, 2020. Overview of Sequential Pattern Mining Algorithms, Kirti S. Patil, Feb 2014, IJARCET



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details