iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68500

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 13 Issue IV Apr 2025- Available at www.ijraset.com

Algorithm for Deep Learning in Digital Video Forensic

Rekha Yaduvanshi[1], Saumya Raj[2], Saloni Parashar[3], Naincy Kumari[4], Dr. Sureshwati[5], Dr. Saumya Chaturvedi[6], Poonam Verma[7]

[1,2,3,4]*Department Of Computer Applications, Greater Noida Institute Of Technology (Engg. Institute), Greater Noida, India*
[5]*Assistant Professor, Department of Computer applications, Greater Noida Institute of Technology (Engg. Institute), Greater Noida, India*
[6]*Prof., Department of Computer applications, Greater Noida Institute of Technology (Engg. Institute), Greater Noida, India*
[7]*Assistant Professor, Department of Computer applications, Greater Noida Institute of Technology (Engg. Institute), Greater Noida, India*

Abstract: ***The rapid advancement of digital media has led to an increase in video-based misinformation, tampering, and forgery, posing serious challenges in legal, investigative, and journalistic domains. This project proposes a Deep Learning-Based Algorithm for Digital Video Forensics, designed to detect and analyze digital video manipulations efficiently. The algorithm leverages advanced Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to extract spatiotemporal features, detect frame-level anomalies, and identify inconsistencies indicative of video forgery. The system is implemented using Python, TensorFlow, and OpenCV, with a web-based interface built using Flask and React.js to provide an interactive and user-friendly forensic analysis platform. The results are presented with a confidence score, highlighting suspected tampered regions within the video for detailed analysis This project aims to provide a powerful forensic tool for law enforcement agencies, media organizations, and cybersecurity professionals, enabling them to verify video authenticity efficiently.***

Keywords: Video Forensic Analysis and Object based Forgery systems, Deep Neural Network, CNN and 3DCN Network, Face to face and face swap system

I. INTRODUCTION

Developing deep learning-based applications, have made portraits alive, generated unique fashion patterns, new videos and many more. This benign usage of technology has ap appreciated the researchers, while on the other hand, its malicious usage in form of forged videos has aroused concern and distrust in society. Moreover, during COVID period, when video con fencing or live videos on social media were the only means to communicate for business, education, election campaigning or any social connects, the need of video authentication increases many folds. One of the best examples Synthesizing Obama the video on social media, where former US president Barack Obama is addressing in front of camera. The video consists of context which the former president had never spoken. Similarly, surveillance videos at ATMs, banks or crime scenes are being manipulated; to conceal the truth Thus, it requires to develop a robust and efficient digital video forgery detector. The digital video forensics is a process of authenticating whether the content of a given video is manipulated or not. There are two types of forgery detection techniques, namely active forensics techniques and passive forensics technique. The active forensics techniques, utilizes the pre-embedded information is used to detect forgery. The required pre-embedded information like, digital signatures, watermarks etc. are inserted in a video at the time of video capturing. Therefore, the camera should be equipped with a hardware to embed watermark or signature in a video. The distortion in watermark or signature leads to detection of forged video. But smart phone cameras, low-cost cameras, tablets or laptop cameras are not equipped with such software/hardware to create watermarks or signature in a video. Moreover, the videos available on social media does not contain any pre-embedded information. The passive forensic techniques, exploit the intrinsic statistical characteristics of a video to detect forged video. The intrinsic characteristics of a video in forms motion residuals, gradients, moments etc. are analyzed to extract the hidden fingerprints left after performing video forgery Further these fingerprints are used to detect the forged video. The passive forensics techniques do not require any external hardware and can be applied on videos from social media as well as videos captured from a low-cost camera. Thus, the real-world forgery detection challenges can be solved using passive Forensic Technique.

II. RELATED WORK

The increase in deepfake technologies and the creation of synthetic content has drawn a lot of attention to video tampering and forgery detection. An outline of recent advancements in this field is provided below:

There are significant worries about the rise in manipulated digital content, particularly in cases involving law and security. To increase the precision of identifying forged content in videos, deep learning techniques have AI-based systems must be adopted for efficient monitoring and inconsistency detection in digital videos, as traditional forensic techniques frequently fall short in providing real-time analysis. The traditional approaches to video forgery detection used manually created features, which were ineffective at identifying contemporary deepfake methods. This field has been transformed by deep learning techniques like CNNs and RNNs, which enable automated feature extraction and temporal analysis. Various tampering techniques, such as splicing, copy-move forgery, and deepfake generation, affect the authenticity of videos. Strong and adaptable forensic systems that can adjust to various manipulation scenarios are therefore required. Furthermore, gathering video data alone is insufficient; combining various forensic techniques and processing them intelligently are crucial. Forensic systems can guarantee more precise detection of manipulated content and offer a clearer picture of the authenticity of digital videos by combining CNNs, LSTMs, optical flow analysis, and transformers. In order to combat misinformation and video forgery, this integrated approach is essential.

1) For video forensics tasks, such as identifying Face2Face, Deepfake, and other methods of creating fake videos, Convolutional Neural Networks (CNNs) have been utilized extensively.
2) Sequential video frame analysis for tampering detection has proven successful when done by Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.
3) By spotting irregularities in motion and spectral data, optical flow analysis and frequency domain techniques aid in the detection of video tampering
4) Because they can identify long-range dependencies in video sequences, transformer-based models, like Vision Transformers (Vites), have recently shown promise in the field of video forensics.
5) In order to identify fake videos, Generative Adversarial Networks (GANs) train deep learning models on sizable datasets to differentiate between authentic and fraudulent content.

A. Subject-Based Forgery Systems And Video Forensic Analysis

Video Forensic Analysis: - The method of looking into digital videos to find evidence of fraud, manipulation, or tampering is known as video forensic analysis.

In order to determine whether the content has been altered, this process entails taking features out of videos, examining discrepancies, and applying deep learning models. Motion estimation, frame-by-frame analysis, and metadata inspection are important methods.

Object-Based Forgery Detection: - Finding manipulated objects within a video frame is the main goal of object-based forgery detection, which entails spotting irregularities in lighting, reflections, object edges, and background alignment.

To determine whether objects have been added, taken out, or changed in an unusual way, deep learning methods like object detection networks (YOLO, Faster R-CNN) can be applied Forensic analysts can identify fraudulent video alterations with high accuracy by combining these techniques.

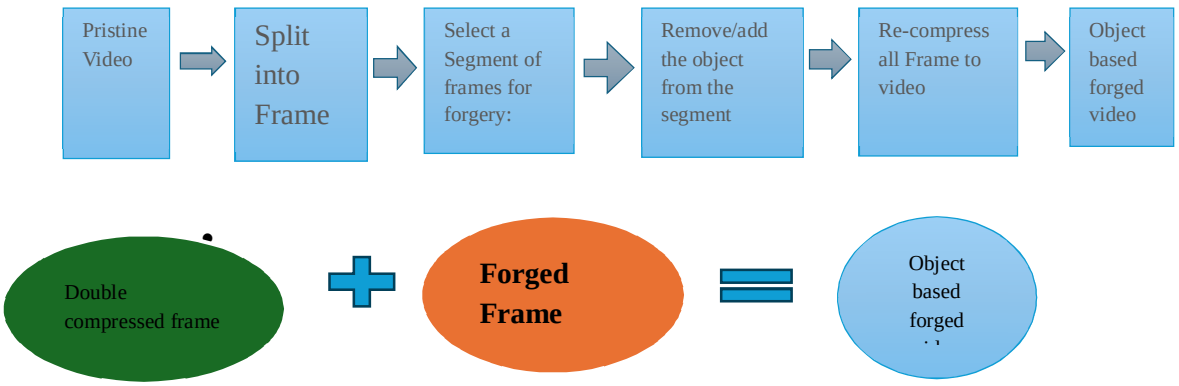


Figure 2.5: The flow diagram of object based forged video generation.

B. Using Deep Neural Network (DNN)

A Deep Neural Network (DNN) is a type of artificial neural network with multiple hidden layers that enhance its ability to learn complex patterns by examining frame-level and temporal inconsistencies, DNNs are frequently used in video forensics to identify deepfake and tampered videos.

Usually, these networks include**:**

- Input Layer: Takes in video frames. Hidden Layers Several layers of neurons with REL-like activation functions that aid in feature extraction.
- Output Layer: Uses learned patterns to classify videos as authentic or tampered. DNNs are crucial for digital video forensic applications because they are highly accurate at identifying manipulated content**.**

C. Using Convolutional Neural Network (CNN)

One type of deep learning model created especially for image and video analysis is called a convolutional neural network (CNN). Their capacity to automatically extract spatial features from frames makes them extremely useful in video forensic applications.

A CNN has several layers, such as:-

1) Conventional Layer: These layers use filters to identify patterns like shapes, edges, and textures.
2) Pooling Layers: These increase computational efficiency by reducing the spatial dimensions while maintaining significant features.
3) Fully Connected Layers: These layers use extracted features to perform classification CNNs are frequently employed in deepfake analysis and frame-level video forgery detection, where they examine individual frames to identify irregularities and manipulations.

D. Using 3D Convolutional Network (3D CNNs)

3D Convolutional Networks (3D CNNs), in contrast to conventional CNNs, extend the convolution operation to three dimensions, which makes them perfect for capturing temporal and spatial features in videos. This is essential for video forensic applications where comprehending motion patterns across frames is necessary to detect tampering.

- 3D Convolutional Layers: These apply filters in three dimensions (height, width, and time) to capture motion-related features.
- 3D Pooling Layers: These lower the dimensionality while maintaining important spatiotemporal data.
- Fully Connected Layers: These complete the classification using the representations that have been learned.

3D CNNs are particularly effective in detecting video splicing, frame insertion, and deepfake videos by analyzing motion inconsistencies that standard CNNs might overlook. Digital video forensic systems can improve the accuracy of forgery detection by combining spatial and temporal analysis techniques by utilizing CNNs and 3D CNNs in tandem. This comprehensive strategy guarantees robust

E. USING Face - To - Face System

The goal of the Face-to-Face system is to create realistic-looking but phony content by altering or synthesizing a person's movements and facial expressions. Deepfake technology, which modifies a person's facial expressions to mimic a different speech or emotion, makes extensive use of this technique.

- Face Expression Manipulation: AI models alter a person's facial expressions while preserving the facial structure
- Lip-Syncing Techniques: Algorithms create the illusion that someone is saying something they never said by synchronizing lip movements with speech patterns
- Motion Transfer: Replicating realistic behavior, one person's head movements and facial expressions are applied to another person. To guarantee realistic and smooth changes, these methods frequently use deep learning architectures like Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs). Forensic examination of biometric characteristics like eye motion and blinking patterns, as well as irregular facial movements and unnatural transitions, is necessary to identify such manipulations

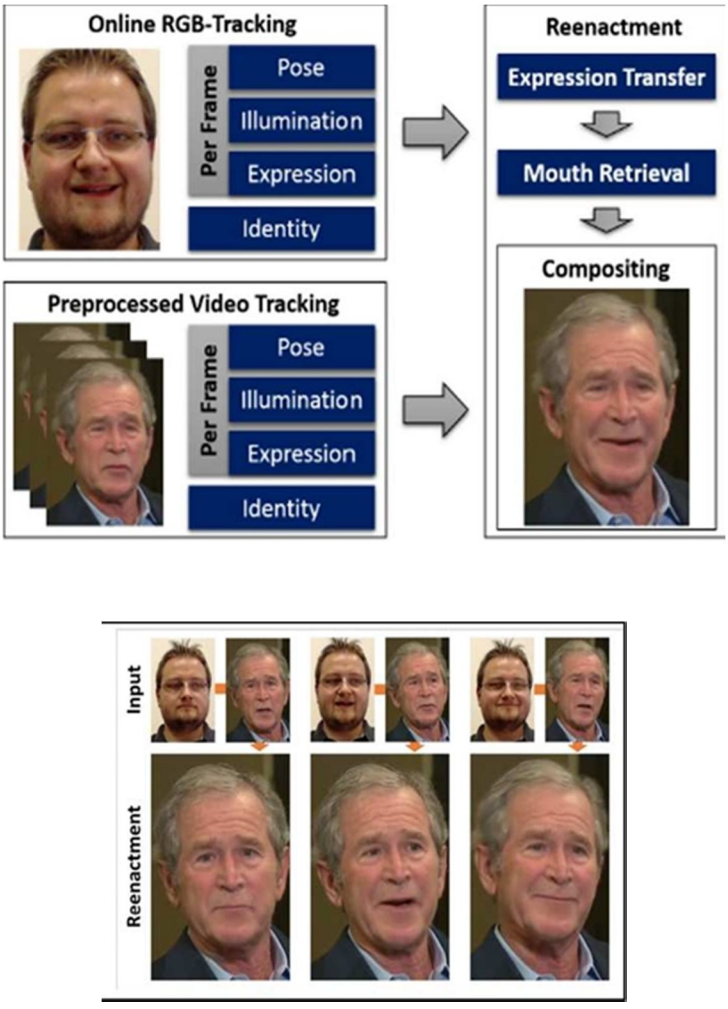


Figure: The Face2Face forgery generation method

F. Using Face Swap System

In an image or video, a face swap system is a technology that substitutes a different person's face while preserving consistent poses, lighting, and realistic facial expressions. It is frequently utilized in deepfake production, augmented reality (AR), and entertainment.

The Face Swap Procedure

- Face Recognition and Alignment: detects and aligns faces using deep learning models such as MTCNN or Retina Face. Encoding Faces and Extracting Features: uses model Face Mapping & Transformation such as Face Net or VGG-Face to extract facial features
- Face Mapping & Transformation: creates a swapped face using autoencoders or Generative Adversarial Networks (GANs). makes use of deep learning-based image synthesis to make sure the new face blends in with the surroundings
- Blending and post-processing: blends the swapped face seamlessly using methods like Poisson blending or Alpha blending

Typical Face Swap Models

Deep Fake (Autoencoders): Uses two faces to train a model, then switches them according to learned embeddings

Face swapping in real time with high realism is made possible by FSGAN (Face Swapping GAN).

Face swaps are produced by Reenact GAN using expression translation shows a example of a face created from the Face Forensics++ dataset using the Face Swap facial manipulation technique.

Object Based Forged Video Data Set

Dataset Name: SYSU-OBJFORG and SYSU-OBJFORG VFVL
Dataset Description: The SYSU-OBJFORG dataset is used to identify object-based video forgeries. It includes video clips taken from static security camera footage, both forged and unaltered. Certain segments (ranging from 25 to 150 frames long) in the forged videos have been altered.
The forged frames, which are separated into forged frames and double-compressed frames, are labeled using the pristine (original) videos as ground truth. The temporal-CNN model is trained and tested using the dataset
This original dataset is the basis for the SYSU-OBJFORG VFVL dataset, which has been altered with variable frame sizes and lengths to evaluate the model's resilience in various video scenarios.

Dataset Source:

- The 200 video clips in the SYSU-OBJFORG dataset are 100 forged and 100 pristine, and they have a 720p resolution (1280x720).
- Ten test videos with varying lengths and frame sizes make up the SYSU-OBJFORG VFVL dataset

Dataset Format:

- SYSU-OBJFORG videos have a resolution of 720p (1280x720)
- Compression: The H.264 codec is used to compress videos
- 25 frames per second is the frame rate
- 3Mbps is the bit rate.
- File Format: The.mp4 format is used to store the videos

Variable Dataset:
Because it contains videos with varying frame sizes and lengths, the SYSU-OBJFORG VFVL dataset is a variable dataset. This change makes it possible to test the model's resilience to truncated or resized videos.

Dataset Preprocessing

- Frame Size Adjustment: Superfluous areas are cropped out of videos. The redundant parts of the forged videos are found by comparing them to the original footage.
- Frame Length Reduction: To create the illusion of different video lengths, the length of each forged video is changed while keeping the forged frames intact

Dataset Type
Dataset for object-based forgery detection

- Forgeries: One or two fabricated segments can be found in every forged video
- Ground Truth: To categorize the forged and double-compressed video frames, 100 unaltered videos are used as the ground truth.

Detailed SYSU-OBJFORG VFVL Dataset (Table 3.1)

Sr. No.	Video Name	Frame Size	Total Frames
1	00003125-254.mp4	640x1024	151
2	00004000-118.mp4	640x1024	150
3	00015133-268.mp4	512x896	186
4	00018158-267.mp4	512x1024	159
5	00026000-125.mp4	512x1024	130

This dataset aids in testing the temporal-CNN's capacity to identify object-based video forgeries in a variety of scenarios, including when the video undergoes changes like frame resizing and length adjustments.

G. Benefits of Deep Learning-Based Video Forensic Analysis and Forgery Detection

Detecting manipulations in digital videos, such as deepfakes, which involve face swapping or alteration, requires the use of object-based forgery systems and video forensic analysis. By examining both temporal and spatial inconsistencies in video frames, deep neural networks (DNNs), convolutional neural networks (CNNs), and 3D CNNs are essential for this detection. CNNs are excellent at picking up on minute artifacts like misaligned features or unusual facial textures. 3D CNNs, on the other hand, monitor changes over time, like unusual lip-syncing or blinking, which aid in spotting manipulated content. Realistic fake faces are produced by face-to-face and face-swap systems, but deep learning forensic systems. can identify irregularities in movement, lighting, and facial geometry. Deepfake video forensics gains from these technologies' high accuracy, real-time detection capabilities, and resilience againstinvolving manipulation techniques, which makes it essential for thwarting video-based deception in media, security, and legal applications.

H. Difficulties with Deep Learning-Based Forgery Detection and Video Forensic Analysis

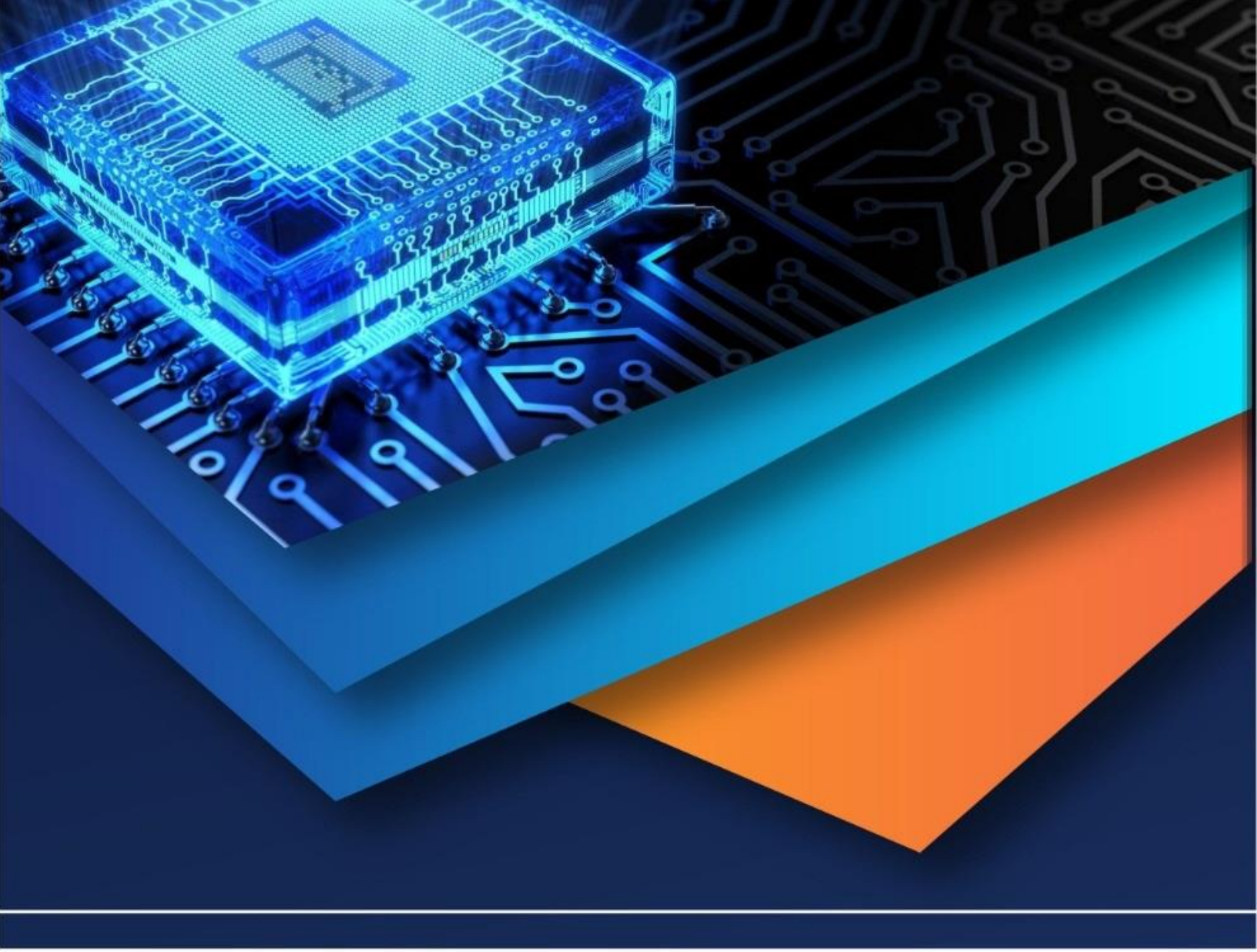
Neural networks based on deep learning, such as CNN, 3D-CNN, and Face-to-Face & Face Swap methods, confront several obstacles when identifying object-based and deepfake video forgeries. Large-scale, diverse, and high-resolution datasets containing both forged and pristine videos are necessary for training highly accurate models, making data availability and quality one of the main challenges. The generalization When models trained on particular datasets are unable to identify forgeries in unseen videos because of differences in lighting, resolution, and compression artifacts, problems occur. Another issue is computational complexity, which makes real-time deepfake detection challenging because deep networks like CNN and 3D-CNN require expensive hardware. Adversarial attacks and developing deepfake techniques make detection even more difficult because forgers are always improving their techniques, which gradually reduces the effectiveness of conventional detection models. Furthermore, it can be challenging to capture motion artifacts and temporal inconsistencies in videos using CNNs alone; for more accurate temporal analysis, 3D-CNNs or hybrid models are needed Furthermore, face-swapping and face-reenactment techniques produce nearly flawless manipulations that call for sophisticated forensic methods to identify minute discrepancies in eye blinks, facial movements, and lighting mismatches. For deepfake forensics to become more transparent and reliable, researchers must concentrate on creating strong, lightweight, and adaptable neural networks that use explainable AI techniques, adversarial training, and multi-modal detection techniques.

III. CONCLUSION

Deep learning algorithms have greatly advanced digital video forensics, achieving high accuracy in detecting tampering, deepfakes, and verifying video authenticity. These models, such as CNNs, GANs, and hybrid networks, excel in tasks like source camera identification and forgery localization, showing robust performance in many cases. However, challenges like computational demands, generalization across diverse datasets, and vulnerability adversarial attacks remain. While challenges remain, continued development of deep learning techniques promises to make digital video forensics even more robust, efficient, and reliable in the future.

REFERENCES

[1] Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation, Jianyu Xiao, Shenyang Li, And Qinglin Xu., 2019
[2] "Hashing Algorithm MD5", Shweta Mishra, Sikha Mishra, Nilesh Kumar 2013.
[3] Yakin Chang, Cheol Kon Jung, (Member, Yee), Peng Ke, Hyoseob Song, And Junge Hwang, "Automatic Contrast Limited Adaptive Histogram Equalization with Dual Gamma Correction".
[4] Muhammad, Khan, Tanveer Hussain, and Sung Wook Baik. "Efficient CNN based summarization of surveillance videos for resource-constrained devices." Pattern Recognition Letters (2018).
[5] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, ``Dual autoencoder network for retinas based low-light image enhancement," IEEE Access, vol. 6, pp. 22084-22093, 2018.
[6] W. Fan, K. Wang, C. François, and Z. Xiong, ``Median filtered image quality enhancement and anti-forensics via variational deconvolution," IEEE Trans. Inf. Forensics Security, vol. 10, no. 5, pp. 1076-1091, May 2015
[7] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, ``Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," IEEE Trans. Image Process., vol. 25, no. 12, pp. 5664-5677, Dec. 2016.
[8] S. Mandal, X. L. Dean-Ben, and D. Razan sky, ``Visual quality enhancement in optoacoustic tomography using active contour segmentation priors," IEEE Trans. Med. Image., vol. 35, no. 10, pp. 2209_2217, Oct. 2016. S. Kim, W. Kang, E. Lee, and J. Paik, ``Wavelet-domain color image enhancement using altered directional bases and frequency-adaptive shrinkage," IEEE Trans. Consume. Electron. vol. 56, no. 2, pp. 063-1070, May 2010.
[9] S. Kim, W. Kang, E. Lee, and J. Paik, ``Wavelet-domain color image enhancement using littered directional bases and frequency-adaptive shrinkage," IEEE Trans. Consume. Electron. vol. 56, no. 2, pp. 063-1070, May 2010.
[10] 0] A Survey of Deep Learning-based Object Detection, Xicheng Jiao, Fellow, IEEE, Fan Zhang, Fang Liu, Senior Member, IEEE, Shu yuan Yang, Senior Member, IEEE. 2019 Lingling Li, Member, IEEE, Zixin Feng, Member, IEEE, and Rong Qu, Senior Member, IEEE
[11] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, ``Automatic contrast limited adaptive histogram equalization with dual gamma correction," IEEE Access, vol. 6, pp. 11782-11792, 2018.
[12] M. Grega, A. Maiola's, P. Guzik, and M. Lescaut, ``Automated detection of firearms and knives in a CCTV image," Sensors, vol. 16, no. 1, p. 47, 2016.
[13] Graupe, Daniel, "Principle of artificial Neural networks", 2013, World Scientific Publishing Co Pte Ltd
[14] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, ``Automatic contrast limited adaptive histogram equalization with dual gamma correction," IEEE Access, vol. 6, pp. 11782-11792, 2018. 12.
[15] M. Grega, A. Mattioli's, P. Guzik, and M. Lescaut, ``Automated detection of firearms and knives in a CCTV image," Sensors, vol. 16, no. 1, p. 47, 2016. 13. Graupe, Daniel, "Principle of artificial Neural networks", 2013, World Scientific Publishing Co Pte Ltd
[16] "Digital Image Processing", R. C. Gonzalez & R. E. Woods, Addison-Wesley Publishing Company, Inc., 1992.
[17] Sitara, K., and B. M. Mestre. "Digital video tampering detection: An overview of passive techniques." in Digital Investigation 18 (2016): 8-22
[18] Wang, Wan, et al. "Identifying video forgery process using optical flow." International Workshop on Digital Watermarking. Springer, Berlin, Heidelberg, 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY