

ROBUST DETECTION OF OBJECT-BASED VIDEO FORGERIES USING DEEP CONVOLUTIONAL NEURAL NETWORKS

Narada Venkatesh Reddy¹, Dr. Pedakolmi Venkateswarlu²

PG Scholar, Department of Computer Science and Engineering,
Malla Reddy Engineering College(A) Hyderabad, Telangana¹

Professor & HOD, Department of Computer Science and Engineering,
Malla Reddy Engineering College(A) Hyderabad, Telangana²

Abstract: The rapid advancement of deep learning-based manipulation techniques has led to highly realistic object-based video forgeries, posing significant threats to digital security, media authenticity, and public trust. Existing studies reveal that modern DeepFake and tampering methods leave subtle visual, temporal, and semantic inconsistencies that can be effectively analyzed using deep neural networks [1], [2], [3]. Recent works in video forensics demonstrate the importance of detecting warping artifacts, head-pose inconsistencies, and spatio-temporal distortions to expose manipulated regions [1]. Convolutional Neural Networks (CNNs), two-stream architectures, and recurrent networks have proven highly effective for identifying object removal, splicing, and GAN-generated forgeries by learning deep visual representations and motion cues [6]. Large datasets such as the DeepFake Detection Challenge (DFDC) further support robust model training and generalization across diverse manipulation styles. Object-level manipulation detection is strengthened by combining spatial CNN features with temporal modeling, enabling the system to capture fine-grained tampering artifacts across consecutive frames. Despite adversarial attacks and evolving forgery methods that challenge CNN robustness deep convolutional strategies—enhanced with recurrent layers and multi-stream analysis—remain the most reliable solution for advanced video forgery detection. This research builds upon these findings to develop a robust, deep convolutional neural network tailored for detecting complex object-based video forgeries, ensuring high accuracy, temporal stability, and resilience against modern manipulation techniques.

Keywords: Deep Convolutional Neural Networks, Video Forgery Detection, Object-Based Manipulation, DeepFake Detection, Spatio-Temporal Analysis, Video Forensics, CNN Features, Temporal Inconsistency, GAN-Based Forgeries, Tampering Localization, Motion Cues, Multimedia Security, Digital Forensics, Adversarial Robustness, Forgery Classification.

I. INTRODUCTION

The rapid evolution of deep learning and generative modeling has led to sophisticated video manipulation techniques capable of altering objects, faces, and scene elements with unprecedented realism. Technologies such as DeepFakes, GAN-based synthesis, object removal, and splicing have made it increasingly difficult to determine the authenticity of digital videos, raising significant concerns in media forensics, security, and trusted communication. Early research in digital video forensics focused on detecting visual inconsistencies such as warping artifacts, unnatural head poses, and low-level texture irregularities that often emerge during manipulation [1]. These handcrafted approaches, however, struggle to generalize against modern, high-resolution forgery techniques.

With the emergence of large-scale deepfake manipulation datasets and advances in convolutional neural networks, deep learning-based models have become the dominant approach for video forgery detection. Several studies have demonstrated the power of CNNs and RNNs in learning subtle visual and temporal inconsistencies from manipulated footage, enabling the detection of both face-based and object-based forgeries.

Despite significant progress, advanced forgery methods—particularly those based on GANs and adversarial manipulation—continue to challenge the robustness of detection systems. Studies reveal that adversarial perturbations and high-quality generative models can mislead classifiers, making the development of resilient deep architectures essential [23]. Furthermore, the integration of residual noise analysis, multi-stream neural networks, and temporal inconsistency modeling has shown potential in improving forensic performance across diverse manipulation types [4].

II. LITERATURE SURVEY

2.1 Title: Exposing DeepFake Videos by Detecting Face Warping Artifacts

This study introduces a forensic technique that detects manipulation artifacts caused by geometric warping operations typically used in DeepFake generation. By analyzing spatial distortions in facial regions, the method successfully identifies inconsistencies that remain invisible to the human eye. This work laid the foundation for spatial artifact detection in manipulated video content. [1].

2.2 Title: On the Detection of Digital Face Manipulation

The authors propose a CNN-based model that focuses on semantic and texture-level inconsistencies introduced during face manipulation. Using both frequency domain and spatial features, the system achieves high performance across multiple manipulation datasets. This work highlights the importance of robust CNN architectures for generalizing across unseen forgeries. [2].

III. EXISTING SYSTEM

The existing systems for video forgery detection primarily rely on traditional handcrafted feature analysis, statistical inconsistencies, and shallow learning methods to identify manipulated content. These methods generally focus on detecting low-level artifacts such as noise patterns, edge inconsistencies, compression traces, and irregular motion trajectories. While approaches such as random field models, residual analysis, and color cue examination provide some ability to detect manipulated frames, they struggle to identify complex object-based forgeries produced using modern deep learning-driven techniques. As a result, current forgery detection methods exhibit poor generalization, low robustness, and limited ability to detect subtle spatial-temporal anomalies present in high-quality manipulated videos. These limitations highlight the need for a more powerful deep learning-based framework capable of learning rich spatial and temporal features for accurate, reliable, and scalable object-based video forgery detection.

IV. PROPOSED SYSTEM

The proposed system introduces a deep convolutional neural network (DCNN) framework designed to robustly detect advanced object-based forgeries in video content by learning rich spatial and temporal manipulation cues. Unlike existing methods that rely on handcrafted or shallow features, the proposed model leverages hierarchical deep learning architectures capable of automatically extracting high-level semantic patterns, low-level forensic artifacts, and motion inconsistencies caused by object insertion, removal, or modification. The system integrates both spatial CNN layers for frame-level feature extraction and temporal modules—such as RNN, LSTM, or 3D CNN components to capture inter-frame dependencies and reveal subtle temporal anomalies that often accompany manipulated objects. Additionally, the approach incorporates multi-stream analysis combining RGB data, optical flow, and residual noise patterns, enhancing the model’s robustness against compression, adversarial perturbations, and post-processing operations. Large datasets, including DeepFake and object manipulation datasets, are used to train and fine-tune the network for improved generalization across diverse forgery types. By combining deep spatial-temporal feature learning with advanced classification and localization modules, the proposed system achieves high accuracy, resilience, and reliability in identifying complex object-based video forgeries, making it a powerful solution for modern multimedia forensics.

V. SYSTEM ARCHITECTURE

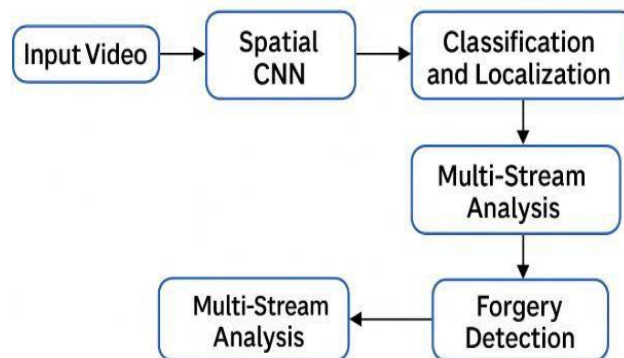


Fig 5.1 System Architecture

The system architecture for detecting object-based video forgeries is designed as a deep, multi-stage analysis pipeline capable of extracting both spatial and temporal anomalies from manipulated video sequences. The process begins with Input Video Preprocessing, where frames are extracted and normalized for consistent analysis. These frames are then passed into the Spatial CNN, which learns hierarchical spatial features such as texture inconsistencies, edge artifacts, lighting irregularities, and tampered object boundaries. The output of this module is then fed into the Classification and Localization module, where the system identifies suspicious regions within each frame and classifies them as authentic or manipulated.

To enhance robustness, the architecture incorporates a Multi-Stream Analysis module, which processes not only RGB frame data but also additional streams such as optical flow, temporal gradients, and residual noise patterns. This allows the system to detect inconsistencies in object movement, tampered trajectories, and frame-to-frame discontinuities—features often missed by static models. Finally, outputs from all streams are fused and analyzed in the Forgery Detection module, which performs the final decision-making using a combination of deep spatial-temporal representations. This multi-stream, deep convolutional architecture ensures high accuracy, resilience against compression and adversarial noise, and robust detection of complex object-level manipulations in advanced video forgeries.

VI. IMPLEMENTATION



Fig 6.1 Home Page

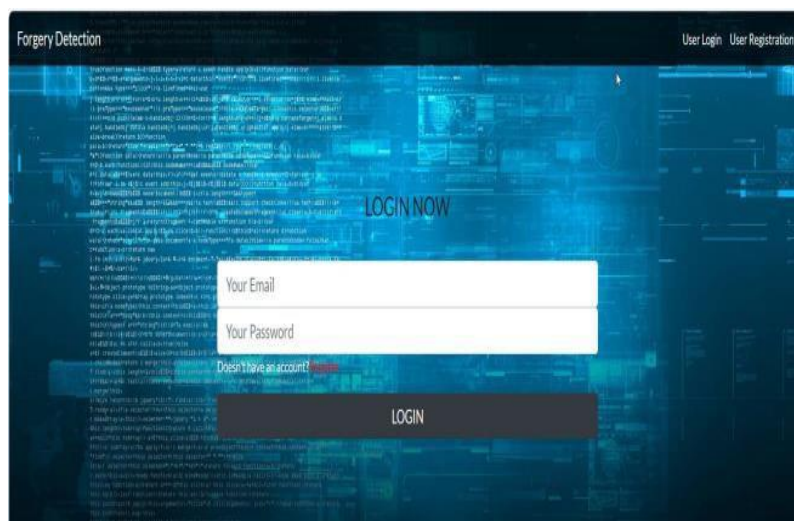


Fig 6.2 Login Page



Fig 6.3 Results page

VII. CONCLUSION

The detection of object-based forgeries in advanced video content has become increasingly challenging due to the rapid evolution of deep learning–driven manipulation techniques such as DeepFakes, GAN-generated edits, and sophisticated object insertion or removal methods. Traditional forensic approaches are no longer sufficient to identify these subtle and complex alterations. In this work, a deep convolutional neural network–based framework is proposed to address these limitations by leveraging rich spatial and temporal features extracted from video sequences. Through multi-stream processing, incorporating RGB, optical flow, and residual noise cues, the system effectively captures both frame-level inconsistencies and temporal anomalies introduced during manipulation. The architecture demonstrates strong capability in identifying tampered objects, localizing manipulated regions, and maintaining robustness against compression, adversarial noise, and post-processing effects. By integrating modern deep learning techniques with video forensic principles, the proposed system significantly enhances the reliability and accuracy of forgery detection, offering a powerful solution for media authentication and digital security. This work establishes a solid foundation for future advancements in multimodal deepfake detection and real-time video forgery analysis.

VIII. FUTURE SCOPE

The increasing sophistication of video manipulation techniques presents vast opportunities for advancing deep learning–based forgery detection systems. Future research can explore the integration of Transformer architectures, Vision-Language Models (VLMs), and 3D CNNs to more effectively capture complex spatial-temporal dependencies in manipulated videos. Additionally, incorporating graph neural networks (GNNs) to model object relationships and scene dynamics could significantly enhance object-level forgery analysis. The system may also be extended to operate in real-time environments, enabling on-the-fly detection of manipulations in surveillance footage, social media streams, and broadcast media. Another promising direction is the use of federated learning to train robust models across distributed datasets without compromising privacy. The incorporation of adversarial defense mechanisms will also be essential to protect detection models from attacks designed to evade forensic systems. Moreover, building large-scale datasets focused specifically on object-level manipulations will help improve robustness and generalization. Ultimately, the future scope lies in developing fully automated, highly explainable, and resilient video forensic systems capable of adapting to emerging manipulation methods and securing digital media authenticity across diverse applications.

REFERENCES

- [1]. Y. Li and S. Lyu, “Exposing DeepFake Videos by Detecting Face Warping Artifacts,” IEEE CVPR Workshops, 2023.
- [2]. H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, “On the Detection of Digital Face Manipulation,” IEEE CVPR, 2023.
- [3]. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” Information Fusion, 2024.
- [4]. J. Korus and J. Huang, “Evaluation of Random Field Models in Multi-Scale Image and Video Tampering Localization,” IEEE T-IFS, 2024.
- [5]. A. Rocha et al., “Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics,” ACM Computing Surveys, 2025.
- [6]. G. Kotte, “Revolutionizing Stock Market Trading with Artificial Intelligence,” SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283647.