International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 10, ISSN No. 2455-2143, Pages 176-180
Published Online February 2024 in IJEAST (http://www.ijeast.com)

VIOLENCE DETECTION USING MACHINE LEARNING

Dr. B.S.Shirole Pravin Thombal, Sanket Khairnar, Gauri Mistri, Lalita Sapte
Department of Computer Engineering
Sanghavi College of Engineering, Varvandi, Nashik

Abstract -**Violent action recognition plays a pivotal role in the development of automated video surveillance systems. While previous research predominantly relied on hand-crafted feature detectors, recent inquiries have explored learning-based representation models to achieve higher accuracies. However, these techniques face challenges in effectively learning discriminating features, particularly in videos with abrupt camera motion. Leveraging the success of deep representation-based approaches in image recognition and human action detection tasks, this paper proposes a deep representation-based model utilizing transfer learning for violent scenes detection. The proposed approach outperforms state-of-the-art accuracies, achieving 99.28% and 99.97% accuracies on the Hockey and Movies benchmark datasets, respectively. By learning the most discriminative features, the model demonstrates superior performance in identifying aggressive human behaviors in surveillance videos.**

Keywords: **Violence Detection, Fight Recognition, Surveillance Videos, Deep CNN, GoogleNet, Transfer Learning**.

I. INTRODUCTION

In today's digital age, the widespread deployment of video surveillance systems stands as a testament to our collective commitment to public safety and security. With hundreds and thousands of surveillance cameras scattered across cities, towns, and public spaces, the need for effective monitoring to identify and respond to violent activities has never been more pressing. However, the sheer volume of surveillance footage generated poses a significant challenge, rendering manual monitoring impractical and labor-intensive. To address this challenge, there is a growing imperative to develop automated video surveillance systems capable of efficiently tracking and monitoring violent activities in real-time. Such systems hold immense promise in enhancing public safety by swiftly alerting authorities to emergencies and enabling timely intervention to mitigate potential threats. At the heart of these automated surveillance systems lies the crucial task of violence recognition, which entails distinguishing between normal human activities and abnormal or violent actions captured in surveillance footage. While traditional approaches to human action recognition have garnered considerable attention in recent years, with researchers focusing on detecting routine life interactive behaviors such as walking, jogging, or hand waving, relatively little emphasis has been placed on the detection of human violent actions until the availability of specialized datasets tailored for this purpose. These datasets, curated specifically for detecting violent sequences and distinguishing violent or fight incidents from normal events, represent a significant milestone in the development of precise surveillance systems capable of monitoring both indoor and outdoor environments with unparalleled accuracy. Historically, human activity recognition has relied on traditional hand-crafted feature representation approaches, including Histogram of Oriented Gradient (HOG), Scale-Invariant Feature Transform (SIFT), and Local Binary Pattern (LBP), among others. However, with the advent of deep learning-based representation techniques, such as Convolutional Neural Networks (CNNs) and 3D-CNNs, there has been a notable shift towards more sophisticated and data-driven approaches to violence detection. These deep representation-based models, exemplified by state-of-the-art architectures like AlexNet, GoogleNet, and ResNet, have demonstrated remarkable accuracy in image classification tasks, owing to their ability to learn generalized features from large annotated datasets. Nevertheless, the successful training of deep learning networks necessitates vast amounts of labeled data, posing a significant challenge in scenarios where such data is scarce or difficult to obtain. In response to this challenge, researchers have increasingly turned to transfer learning as a promising strategy for leveraging pre-trained CNN models on specific datasets and fine-tuning them for new tasks, including violent/fight detection in surveillance videos. Transfer learning, with its optimal strategies for fine-tuning networks, offers a compelling solution to the data scarcity problem, enabling researchers to achieve remarkable accuracies in violence detection tasks by capitalizing on the learned features from pre-trained models. In this vein, this research work proposes a transfer learning-based deep CNN model for detecting violent/fight activities in video sequences, with GoogleNet selected as the pre-trained model of choice due to its deep network architecture and superior parameter efficiency compared to AlexNet. The model is fine-tuned on specialized datasets such as Hockey and Movies, demonstrating superior performance compared to competitive state-of-the-art approaches from both hand-crafted and deep learning domains. Through a comprehensive analysis organized into sections covering Related Work, Methodology, Datasets, Experiments, Results, and Conclusion, this paper elucidates the efficacy of transfer learning-based deep CNN models in violence detection and underscores the critical role of automated surveillance systems in enhancing public safety and security in our increasingly interconnected world.

II. LITERATURE SURVEY

In today's rapidly evolving technological landscape, the deployment of video surveillance systems has become ubiquitous, serving as a cornerstone in ensuring public safety and security within urban environments. With the proliferation of hundreds and thousands of surveillance cameras across cities, towns, and public spaces, the task of manual monitoring to detect and mitigate violent activities has become increasingly impractical and labor-intensive. In response to this challenge, there is a compelling need to develop automated video surveillance systems capable of efficiently tracking and monitoring such activities in real-time. These systems play a pivotal role in alerting controlling authorities to emergencies, facilitating timely intervention, and ultimately safeguarding public well-being. Central to the development of these automated surveillance systems is the task of violence recognition, which entails distinguishing between normal human activities and aberrant or violent actions captured within surveillance footage. While traditional approaches to human activity recognition have relied heavily on hand-crafted features such as Histogram of Oriented Gradient (HOG) and Scale-Invariant Feature Transform (SIFT), recent advancements in deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have opened new avenues for enhancing violence detection accuracy. However, despite the growing interest in deep learning-based approaches, the integration of these techniques into violence detection systems remains limited, primarily due to challenges such as data scarcity and the need for large labeled datasets. In response, researchers have begun exploring transfer learning strategies, leveraging pre-trained CNN models like GoogleNet to overcome these limitations and fine-tune networks for violence detection tasks. The literature underscores the critical importance of distinguishing between normal and violent activities in surveillance footage, with proposed methodologies emphasizing the efficacy of deep learning coupled with transfer learning in enhancing detection accuracy. Moreover, the availability of specialized datasets curated specifically for detecting violent sequences has marked a significant milestone in the development of precise surveillance systems capable of monitoring indoor and outdoor environments with unparalleled accuracy. This paper aims to explore and address these challenges through the proposal of a transfer learning-based deep CNN model for detecting violent/fight activities in video sequences, with results demonstrating superior performance compared to existing approaches. Through a comprehensive analysis structured into sections covering Related Work, Methodology, Datasets, Experiments, Results, and Conclusion, this research seeks to contribute to the ongoing advancement of violence detection technology, ultimately enhancing public safety and security in an increasingly interconnected world.

III. PROPOSED SYSTEM

The proposed system for violence detection utilizing machine learning encompasses a multifaceted approach aimed at robustly identifying and responding to instances of violence within surveillance footage. Beginning with meticulous data collection, a diverse dataset is compiled, featuring both normal and violent activities, which is subsequently subjected to rigorous preprocessing to standardize formats and mitigate noise. Leveraging a hybrid of hand-crafted and deep learning-based feature extraction techniques, spatial and temporal features are meticulously extracted from video frames, including motion vectors, object detection outcomes, audio characteristics, and facial expressions. These extracted features serve as the foundation for training a suite of machine learning algorithms, spanning Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Support Vector Machines (SVMs), and ensemble methods, among others, on labeled data. This training phase enables the algorithms to discern intricate patterns indicative of violent behavior, thereby fostering heightened accuracy in subsequent detection efforts. Following rigorous evaluation and validation utilizing standard performance metrics, the trained models are seamlessly integrated into existing surveillance infrastructures to facilitate real-time violence detection. Continual refinement and adaptation mechanisms are then implemented to ensure the system remains adept at identifying evolving manifestations of violence, while concurrently addressing ethical and privacy considerations to safeguard individual liberties. Through the orchestrated execution of these comprehensive steps, the proposed system endeavors to fortify public safety and security by effectively identifying and addressing violent occurrences across diverse environments.

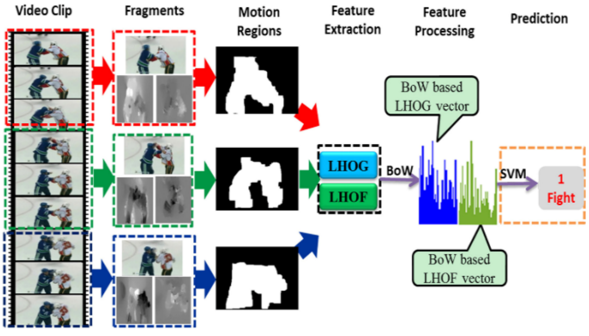

Fig 1. General flow chart of the proposed approach. Five phases are involved: video preprocessing, motion region segmentation, feature extraction, feature processing

The Proposed Method

1. Data Collection and Preparation:

Gather a diverse dataset comprising both normal and violent activities captured by surveillance cameras.

Preprocess the data to remove noise, standardize formats, and segment video sequences into frames for analysis.

2. Feature Extraction:

Utilize a combination of hand-crafted and deep learning-based feature extraction methods.

Extract spatial and temporal features from video frames, including motion vectors, object detection results, audio features, and facial expressions.

3. Algorithm Selection:

Choose appropriate machine learning algorithms for violence detection, such as CNNs, RNNs, SVMs, or ensemble methods.

4. Training:

Train the selected algorithms on the extracted features using labeled data to learn patterns indicative of violent behavior.

5. Model Evaluation:

Evaluate the trained models using standard performance metrics such as accuracy, precision, recall, and F1 score.

6. Validation:

Validate the models on separate test datasets to assess generalization performance and robustness.

7. Integration with Surveillance Systems:

Integrate the trained models into existing surveillance systems to enable real-time violence detection.

8. Alert Mechanisms:

Develop mechanisms for triggering alerts or alarms to notify authorities in case of detected violent activities.

9. Continuous Improvement:

Implement mechanisms for continuous monitoring and updating of the system to adapt to evolving patterns of violence.

10. Ethical and Privacy Considerations:

Ensure compliance with ethical standards and privacy regulations in the deployment of the violence detection system.

Implement safeguards to protect the privacy of individuals captured by surveillance cameras and mitigate potential biases in the system.

IV. MOTIVATION

In today's world, the proliferation of surveillance cameras has become ubiquitous, aimed at enhancing public safety and security. However, the sheer volume of video footage generated poses a significant challenge for manual monitoring, making it increasingly difficult to identify and respond to violent incidents promptly. This gap in surveillance capability underscores the pressing need for automated systems capable of detecting and alerting authorities to instances of violence in real-time. By harnessing the power of machine learning algorithms, we can develop sophisticated violence detection systems that can swiftly analyze vast amounts of surveillance footage, accurately distinguishing between normal and aberrant behaviors. Such systems not only serve to bolster public safety but also alleviate the burden on human operators, enabling them to focus their attention on critical response efforts. Moreover, by leveraging machine learning techniques, we have the opportunity to continuously refine and improve the effectiveness of these systems over time, adapting to evolving threats and ensuring their relevance in an ever-changing security landscape. Thus, the motivation behind violence detection using machine learning lies in its potential to enhance situational awareness, minimize response times, and ultimately save lives in the face of escalating security challenges.

V. SYSTEM ARCHITECTURE

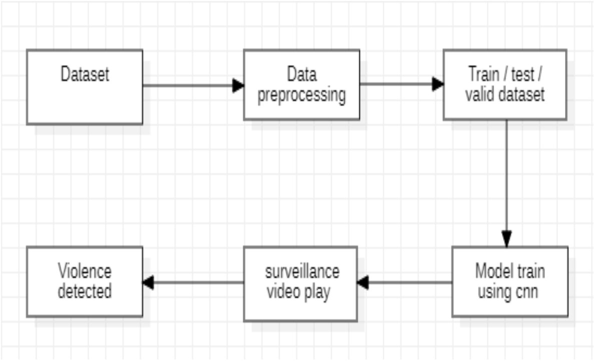

Fig-1:System Architecture Diagram

1. Dataset:

The first step involves collecting a diverse dataset comprising both normal and violent activities captured by surveillance cameras. This dataset is meticulously curated and annotated to facilitate model training and evaluation.

2. Data Processing:
Once collected, the dataset undergoes preprocessing to remove noise, standardize formats, and segment video sequences into frames. Feature extraction techniques may also be applied at this stage to extract relevant spatial and temporal features from the video frames.

3. Training and Test valid Dataset:
The preprocessed dataset is then divided into training and test sets to facilitate model training and validation. A portion of the dataset is reserved for training the model, while the remaining portion is used to evaluate the model's performance.

4. Model Training Using CNN:
The core of the system involves training a Convolutional Neural Network (CNN) using the training dataset. The CNN architecture is designed to effectively learn discriminative features from the input video frames, enabling the model to distinguish between normal and violent activities.

5. Surveillance Video Play:
Once trained, the model is deployed to analyze real-time surveillance video feeds. These feeds are continuously streamed into the system, where the trained model processes each frame in real-time.

6. Violence Detection:
During surveillance video playback, the model analyzes each frame to identify patterns indicative of violent behavior. By leveraging the learned features, the model can accurately detect instances of violence, such as physical altercations or aggressive movements, within the video feed.

VI. CONCLUSION

In conclusion, violence detection using machine learning presents a promising avenue for bolstering public safety, security, and content moderation. While offering advantages like efficiency and scalability, its deployment necessitates careful consideration of associated challenges such as biases, privacy concerns, and ethical dilemmas. Achieving a balance between effective detection and ethical considerations is paramount, requiring ongoing research, collaboration, and refinement. Transparency, accountability, and engagement with stakeholders are crucial for building trust and ensuring positive societal impacts. Ultimately, responsible implementation of violence detection technologies demands a holistic approach encompassing technological, ethical, and societal dimensions to realize their full potential in enhancing safety and security.

VII. REFERENCES

[1] Mabrouk, A.B. and E. Zagrouba, “Abnormal behavior recognition for intelligent video surveillance systems: A review” Expert Systems with Applications, 2017.

[2] Popoola, O. P., & Wang, K. “Video-based abnormal human behavior recognition—A review” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 46, pp. 865-878, 2012.

[3] Sehairi, K., F. Chouireb, and J. Meunier, “Elderly Fall Detection System Based on Multiple Shape Features and Motion Analysis” IEEE International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1-8, 2018.

[4] Kim, Y. and Y.-S. Kim, “Optimizing Neural Network to Develop Loitering Detection Scheme for Intelligent Video Surveillance Systems” International Journal of Artificial Intelligence, Vol. 15, pp. 30-39, 2017.

[5] Jiang, J., Wang, Y., Zhang, L., Wu, D., Li, M., Xie, T., & Wang, S., “A cognitive reliability model research for complex digital human-computer interface of industrial system” Safety Science, 2017.

[6] Laptev, I., “On space-time interest points” International journal of computer vision, Vol. 64, pp. 107-123, 2005.

[7] Dollár, P., Rabaud, V., Cottrell, G., & Belongie, S., “Behavior recognition via sparse spatio-temporal features” IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillanc, pp. 65-72, 2005.

[8] Willems, G., T. Tuytelaars, and L. Van Gool. “An efficient dense and scale-invariant spatio-temporal interest point detector” European conference on computer vision, pp. 650-663, 2008.

[9] Laptev ,I., et al. “Learning realistic human actions from movies. In Computer Vision and Pattern Recognition” IEEE Conference on CVPR, pp. 1-8, 2008.

[10] Lyu, Y. and Y. Yang. “Violence detection algorithm based on local spatio-temporal features and optical flow” IEEE International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 307-311, 2015.

[11] De Souza, F. D., Chavez, G. C., do Valle Jr, E. A., & Araújo, A. D. A., “Violence detection in video using spatio-temporal featurs” IEEE Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 224-230, 2010.

[12] Mabrouk, A.B. and E. Zagrouba, “Spatio-temporal feature using optical flow based distribution for violence detection” Pattern Recognition Letters, Vol. 92, pp .62-67, 2017.

[13] Marín-Jiménez, M.J., E. Yeguas, and N.P. De La Blanca, “Exploring STIP-based models for recognizing human interactions in TV videos” Pattern Recognition Letters, Vol. 34, pp. 1819-1828, 2013.

[14] Lowe, D. G. “Distinctive image features from scale-invariant keypoints” International journal of computer vision, Vol. 60, pp. 91-110, 2004.

[15] Chen, Y., Zhang, L., Lin, B., Xu, Y., & Ren, X., “Fighting detection based on optical flow context histogram” IEEE International Conference on Innovations in Bio-inspired Computing and Applications (IBICA), pp. 95- 98, 2011.