ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-POWERED AUTOMATED SURVEILLANCE SYSTEM FOR EARLY VIOLENCE AND THREAT DETECTION

[1] A V Murali Krishna , [2] B Meenakshi

[1,2] Assistant Professor in Department of Computer Science And Engineering Matrusri Engineering College

Abstract

Public safety in crowded and open environments has become a major concern, leading to the widespread installation of surveillance cameras for activity monitoring. Despite their presence, manually watching continuous CCTV footage is labor-intensive, inefficient, and often results in slow responses to violent situations. To address this challenge, this project introduces an automated system that detects violent behavior in surveillance videos using deep learning methods. The proposed framework utilizes the YOLO (You Only Look Once) algorithm to locate and identify people within video frames. Once individuals are detected, a Convolutional Neural Network (CNN) examines their movements and interactions to classify the activity as either violent or non-violent. To enhance the reliability and performance of the model, it is trained and evaluated using established datasets, including Real Life Violence Situations (RLVS), RWF-2000, and the Hockey Fight Dataset. When a violent event is recognized, the system automatically logs the timestamp and saves the relevant video frame as visual evidence. System performance is measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score.This solution contributes to smart surveillance by enabling rapid detection of violent incidents, allowing authorities to take prompt action and improve overall public safety.

Keywords: Violence Detection, Deep Learning, YOLO, Convolutional Neural Networks, Surveillance Systems, Computer Vision, CCTV Monitoring.

I INTRODUCTION

In recent years, ensuring safety in public places has become a major concern due to the increasing number of violent incidents occurring in crowded environments such as streets, transportation hubs, shopping malls, educational institutions, and workplaces. Surveillance systems using Closed Circuit Television (CCTV) cameras are widely deployed to monitor activities and enhance security. However, continuously observing large numbers of

surveillance cameras manually is a challenging and time-consuming task. Human operators may experience fatigue or distraction, which can lead to delayed detection of violent activities or suspicious behavior [2], [5].With the rapid advancement of Artificial Intelligence (AI) and Computer Vision, automated systems are now capable of analyzing video streams and identifying abnormal activities without constant human supervision. Deep learning techniques have shown remarkable performance in visual recognition tasks such as object detection, action recognition, and behavior analysis. These technologies enable intelligent surveillance systems that can automatically detect violent actions such as fights, physical aggression, or abnormal interactions among individuals [3], [4].

Violence detection in videos involves analyzing spatial and temporal patterns of human behavior to determine whether aggressive or harmful actions are occurring. Traditional surveillance systems mainly record video footage for later review, but they lack the ability to automatically analyze events in real time. Recent research focuses on using deep learning models such as **Convolutional Neural Networks (CNNs)** and **3D CNNs** to recognize complex human actions in videos. These models can learn visual patterns associated with violent behavior by analyzing large datasets containing examples of violent and non-violent activities [1], [11], [10].

Object detection algorithms also play an important role in surveillance analysis. The **YOLO (You Only Look Once)** model is widely used for real-time object detection because it can identify multiple objects within an image in a single pass through a neural network [27]. By detecting individuals present in video frames and analyzing their interactions, automated systems can identify suspicious actions more efficiently.

The main objective of this project is to develop a Ai-Powered Automated Surveillance System For Early Violence And Threat Detection that can automatically analyze surveillance videos and classify activities as violent or non-violent. The system extracts frames from video input, detects people using YOLO, and then analyzes their interactions using a CNN-based classification model. When violent behavior is detected, the system records the timestamp and captures the relevant frame to assist security personnel in responding quickly.

The proposed system aims to improve the efficiency of surveillance monitoring and support public safety by enabling faster identification of violent incidents. Intelligent surveillance systems based on deep learning can significantly reduce the burden on human operators and provide real-time alerts when abnormal activities occur [5], [12].

II LITERATURE SURVEY

Violence detection in surveillance videos has gained significant attention in recent years due to the growing demand for intelligent security systems. Researchers have explored various machine learning and deep learning approaches to identify violent behavior automatically from video data.

Accattoli et al. [1] proposed a deep learning-based approach for violence detection by combining 3D Convolutional Neural Networks (3D CNNs) with Support Vector Machines. The system extracts spatio-temporal features from video sequences and uses them to classify violent and non-violent activities. Their study demonstrated that combining deep learning with traditional classifiers can significantly improve detection accuracy.

Omarov et al. [2] conducted a comprehensive review of existing violence detection techniques used in surveillance systems. The study highlights various approaches including motion analysis, deep neural networks, and hybrid models for detecting abnormal events in video footage. The authors concluded that deep learning methods outperform traditional machine learning techniques due to their ability to automatically learn complex visual patterns.

Haiura and Santos [3] proposed a violence detection framework using Convolutional Neural Networks (CNNs) to classify video frames into violent and non-violent categories. Their model learns features related to human posture, motion, and interaction patterns. Experimental results showed that CNN-based approaches achieve high accuracy in recognizing violent behavior compared to conventional feature-based methods.

Moaaz and Mohamed [4] introduced a deep learning architecture for detecting violent activities in surveillance videos. The system analyzes motion patterns and interactions between individuals to identify aggressive actions such as fighting. The research demonstrates that deep learning models can effectively identify violent scenes in complex environments.

Merit et al. [5] developed an AI-based surveillance system capable of detecting violent incidents in real time. The proposed model integrates deep learning algorithms with intelligent monitoring systems to automatically identify abnormal events. The study emphasizes the importance of real-time detection to assist security personnel in preventing escalation of violent situations.

Negre et al. [6] presented a systematic review of deep learning techniques used in video-based violence detection. The research highlights the effectiveness of convolutional neural networks, recurrent neural networks, and hybrid architectures for analyzing spatial and temporal information in videos.

Cheng et al. [7] introduced the RWF-2000 dataset, which contains a large collection of real-

world violent and non-violent video clips for training and evaluating violence detection models. The dataset has become widely used in research for benchmarking violence detection algorithms.

Patel [11] proposed a CNN-LSTM model for real-time violence detection. The CNN component extracts spatial features from video frames, while the LSTM network captures temporal relationships between consecutive frames. This hybrid architecture improves the system's ability to understand motion dynamics in videos.

Andrade et al. [12] developed a deep learning-based architecture specifically designed for surveillance video analysis. Their system uses multiple neural network layers to extract hierarchical features from video sequences, enabling accurate detection of aggressive interactions.

Mahmud et al. [13] proposed a multi-model framework that combines different deep learning techniques to enhance the reliability of violence detection systems. The results showed that combining multiple models improves detection performance and reduces false alarms.

Although many studies have explored violence detection techniques, challenges still remain in achieving high accuracy in complex environments such as crowded scenes, poor lighting conditions, and real-time processing requirements. Therefore, there is a need for efficient models that can detect violent activities quickly and accurately in surveillance videos.

The proposed system in this project addresses these challenges by combining YOLO for person detection and CNN for activity classification, providing an efficient and real-time approach for identifying violent incidents in surveillance environments.

III EXISTING SYSTEM

In modern security environments, surveillance cameras such as Closed-Circuit Television (CCTV) systems are widely installed in public areas including streets, shopping malls, railway stations, educational institutions, and workplaces. These cameras continuously record video footage to monitor activities and enhance safety. In traditional surveillance setups, the recorded video streams are observed by human operators who are responsible for identifying suspicious or abnormal events.

Despite their widespread use, manual surveillance systems have several limitations. Monitoring multiple video screens for extended periods can cause fatigue and loss of concentration among security personnel. As a result, important incidents such as fights, assaults, or aggressive behavior may be missed or identified only after the incident has already occurred. This delay reduces the effectiveness of

surveillance systems in preventing or responding to dangerous situations in a timely manner.

To address these limitations, some existing systems use basic motion detection or rule-based approaches to identify abnormal activities. These systems typically rely on predefined thresholds or simple motion analysis techniques to detect unusual movements within video frames. However, such approaches are not reliable in real-world environments where normal activities such as running, sudden movements, or crowd interactions may produce false alarms. At the same time, certain violent incidents may remain undetected due to the limitations of these basic methods.

In recent years, researchers have explored the use of machine learning and deep learning techniques for activity recognition in video surveillance. Models such as Convolutional Neural Networks (CNNs) have been applied to analyze human actions and identify abnormal behaviors. Although these approaches have improved detection accuracy, many existing systems still face challenges related to high computational requirements, limited real-time processing capability, and difficulty in accurately understanding complex human interactions.

Therefore, there is still a need for more efficient and intelligent surveillance solutions that can automatically detect violent activities with higher accuracy and provide faster responses in real-time environments.

IV PROBLEM STATEMENT

With the rapid growth of surveillance technologies, a large number of cameras are deployed in public and private locations to monitor activities and improve security. These cameras generate a huge amount of video data every day. Monitoring such a vast volume of video footage manually is extremely difficult and inefficient. Security personnel are often required to observe multiple camera feeds simultaneously, which can lead to reduced attention, fatigue, and delayed recognition of critical incidents.

Violent activities such as fights, assaults, and aggressive interactions usually occur suddenly and may escalate quickly if not addressed in time. Traditional surveillance systems mainly focus on recording video footage for later investigation and do not have the capability to automatically detect violent behavior while it is happening. As a result, authorities may only become aware of such incidents after significant damage or harm has already occurred.

Another major challenge in surveillance monitoring is the complexity of real-world environments. Conventional motion-based detection methods are unable to differentiate between normal activities and violent behavior effectively. Actions such as running, playing, or crowd movement may trigger false alarms, while

actual violent activities may go unnoticed. Factors such as poor lighting conditions, crowded scenes, and occlusions further complicate the detection process.

To overcome these challenges, there is a need for an intelligent surveillance system that can automatically analyze video streams, identify individuals present in the scene, and accurately detect violent activities. Such a system should be capable of processing video data efficiently, detecting violent events in real time, and generating alerts that can assist security personnel in taking immediate action.

This project aims to address this problem by developing an automated violence detection system using deep learning techniques. The proposed system utilizes the YOLO algorithm for detecting people in video frames and Convolutional Neural Networks (CNN) for classifying activities as violent or non-violent. By integrating these technologies, the system can improve the efficiency of surveillance monitoring and contribute to enhanced public safety.

V PROPOSED SYSTEM

To overcome the limitations of traditional surveillance systems, this project proposes an intelligent violence detection system that automatically identifies violent activities in surveillance videos using deep learning techniques. The proposed system integrates real-time object detection and activity classification to accurately detect aggressive behavior in video streams.

The system uses the YOLO (You Only Look Once**)** object detection algorithm to identify and locate individuals present in each video frame. YOLO is known for its high speed and accuracy in detecting objects in real-time applications. By detecting human presence in the scene, the system focuses on analyzing interactions between individuals to determine whether violent behavior is occurring.

After detecting people in the video frames, the extracted frames are processed using a Convolutional Neural Network (CNN) model for activity classification. The CNN analyzes spatial features in the frames to differentiate between violent and non-violent activities. The model is trained using publicly available datasets such as Real Life Violence Situations (RLVS), RWF-2000, and Hockey Fight Dataset, which contain examples of both violent and normal activities. Training the model on these datasets helps the system learn visual patterns associated with aggressive actions such as fighting, pushing, or physical attacks.

The proposed system processes surveillance videos by first converting them into individual frames. These frames are then passed through the YOLO model to detect persons present in the scene. Once the human objects are identified, the CNN model analyzes their behavior and classifies the activity as either violent or non-

violent. If violent behavior is detected, the system automatically records the timestamp of the event and captures the corresponding frame image for further investigation.

Another important feature of the proposed system is its ability to operate in near real-time environments. The combination of YOLO for fast object detection and CNN for accurate activity classification allows the system to analyze video streams efficiently. This helps in generating timely alerts that can assist security personnel in responding quickly to potentially dangerous situations.

Overall, the proposed violence detection system aims to enhance surveillance monitoring by providing automated analysis of video footage. By reducing the dependency on manual observation and enabling faster detection of violent incidents, the system can significantly improve safety and security in public spaces such as transportation hubs, educational institutions, shopping centers, and workplaces.

VI METHODOLOGY

The proposed violence detection system follows a systematic approach to analyze surveillance videos and automatically identify violent activities using deep learning techniques. The process begins with collecting video input from surveillance cameras or recorded video files that contain various human activities occurring in different environments such as public places, streets, educational institutions, and workplaces. These videos serve as the primary data source for detecting abnormal or aggressive behavior.

The input video is first converted into a sequence of frames so that each moment of the video can be analyzed individually. Frame extraction helps in breaking the video into manageable image units for further processing. After extraction, preprocessing operations such as resizing and normalization are applied to the frames in order to reduce computational complexity and improve the efficiency of the system. These preprocessing steps help the model process the data faster while maintaining essential visual information.

Once the frames are prepared, the YOLO (You Only Look Once) object detection algorithm is applied to detect people present in the video frames. YOLO is a real-time object detection model that can identify multiple objects in a single pass through the neural network. The algorithm generates bounding boxes around detected individuals, allowing the system to focus specifically on human activities and interactions within the scene.

After detecting the individuals in the frame, the system analyzes the detected regions using a Convolutional Neural Network (CNN) to classify the activity. The CNN extracts important visual features from the frames, such as body posture, movement patterns, and interactions between individuals. Based on these features,

the model determines whether the activity represents violent or non-violent behavior.

The CNN model is trained using datasets that contain examples of both violent and normal activities, including Real Life Violence Situations (RLVS), RWF-2000, and Hockey Fight Dataset. Training the model with these datasets allows it to learn the patterns and characteristics associated with aggressive actions such as fighting, pushing, or physical attacks.

When the model detects violent behavior in the video, the system records the timestamp of the event and captures the corresponding frame image. This information can help security personnel quickly identify and respond to the situation. Through this methodology, the system enables automated monitoring of surveillance footage and improves the ability to detect violent incidents in real-time environments.

VII IMPLEMENTATION

The implementation of the proposed violence detection system involves developing a deep learning framework capable of processing video data and identifying violent activities automatically. The system is implemented using the Python programming language due to its strong support for machine learning and computer vision applications. Various libraries and frameworks are used to handle different components of the system.

OpenCV is used for handling video input, frame extraction, and image processing tasks. It enables the system to read video files, capture frames, and perform image preprocessing operations efficiently. Deep learning frameworks such as TensorFlow or PyTorch are used to build and train the Convolutional Neural Network model that performs the activity classification.

The training process begins with preparing the dataset. Videos from datasets such as RLVS, RWF-2000, and Hockey Fight Dataset are converted into frames, and the images are labeled as violent or non-violent. These labeled frames are then divided into training and testing sets. During training, the CNN model learns to recognize patterns associated with violent behavior by analyzing the visual features present in the frames.

The YOLO object detection model is integrated into the system to detect human objects in each video frame. During execution, the system processes the video frame by frame. YOLO first detects the presence of individuals in the frame and draws bounding boxes around them. The detected regions are then passed to the CNN model for activity classification.

The CNN model analyzes the detected regions and determines whether the activity in the frame is violent or non-violent. If violent behavior is detected, the system records the timestamp of the event and saves the corresponding frame image. This feature allows security personnel to

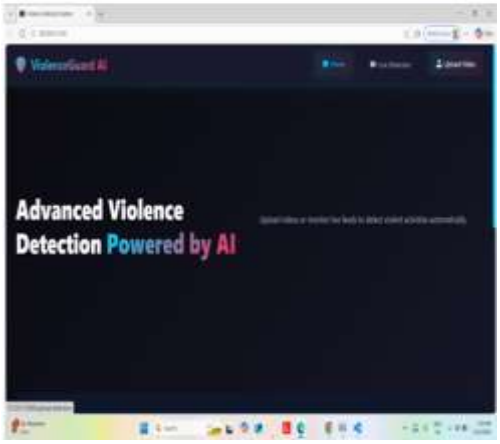
review the detected incident and take necessary action.

The overall system is tested using different video samples to evaluate its performance. The effectiveness of the model is measured using evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics help determine how well the system can correctly detect violent activities while minimizing false alarms.

Through this implementation, the system provides an automated solution for analyzing surveillance videos and detecting violent activities. The combination of YOLO for object detection and CNN for activity classification enables the system to monitor video streams efficiently and support improved safety in surveillance environments.

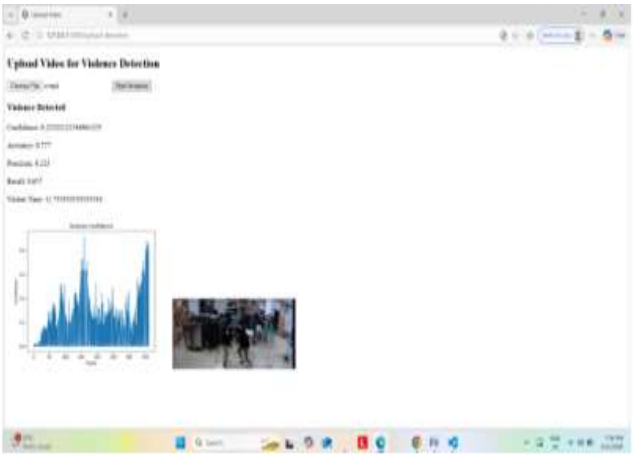
Results and Discussion

The performance of the proposed violence detection system was evaluated using several surveillance video samples containing both violent and non-violent activities. The system processes the input videos by extracting frames, detecting human objects using the YOLO object detection algorithm, and classifying the actions using a Convolutional Neural Network (CNN). The evaluation focuses on determining how accurately and efficiently the system can identify violent behavior in real-world video scenarios.



Home page of the output

Uploading the Dataset



Detection of Voilence

During testing, the YOLO model successfully detected individuals present in the video frames by generating bounding boxes around the detected persons. This allowed the system to

concentrate on human interactions within the scene. The detected regions were then passed to the CNN model, which analyzed the visual patterns and classified the activity as either violent or non-violent. The CNN model was trained using datasets such as RLVS, RWF-2000, and Hockey Fight Dataset, which helped the model learn patterns associated with aggressive behaviors like fighting and pushing.

To measure the effectiveness of the system, standard evaluation metrics such as accuracy, precision, recall, and F1-score were used. These metrics help assess how well the model performs in identifying violent activities while minimizing false detections.

The results indicate that the system achieved a high accuracy rate in detecting violent activities in surveillance videos. The precision value shows that most of the detected violent events were correctly classified by the model, while the recall value indicates that the system was able to identify the majority of actual violent incidents present in the dataset. The F1-score provides a balanced evaluation of the system's performance.

In addition to performance metrics, the system was also evaluated on different types of video activities to understand its behavior in various scenarios.

The experimental results show that the system performs well in detecting both violent and non-violent activities. Violent activities such as fighting were successfully recognized by the CNN model, while normal activities like walking or standing were classified as non-violent. The integration of YOLO and CNN enabled efficient detection and classification of activities within the video frames.

However, certain limitations were observed during testing. In complex environments such as crowded scenes, poor lighting conditions, or partial occlusion of individuals, the system occasionally produced incorrect classifications. These factors can affect the clarity of visual features and make it more difficult for the model to accurately interpret human interactions.

Despite these challenges, the proposed system significantly reduces the need for continuous manual monitoring of surveillance footage. By automatically detecting violent activities and capturing the corresponding frame along with the timestamp, the system can assist security personnel in quickly identifying and responding to critical incidents.

Overall, the results demonstrate that the combination of YOLO for human detection and CNN for activity classification provides an effective and reliable approach for automated violence detection in surveillance systems.

VIII CONCLUSION

a violence detection system based on deep learning techniques was developed to automatically identify violent activities in

surveillance videos. The increasing number of surveillance cameras installed in public and private environments has made manual monitoring difficult and inefficient. Therefore, an intelligent system capable of automatically analyzing video streams and detecting violent behavior is essential for improving public safety.

The proposed system integrates the YOLO object detection algorithm with a Convolutional Neural Network (CNN) to detect and classify human activities in video frames. YOLO is used to identify individuals present in the video, while the CNN model analyzes their interactions to determine whether the activity is violent or non-violent. The system was trained and tested using publicly available datasets such as Real Life Violence Situations (RLVS), RWF-2000, and the Hockey Fight Dataset.

The experimental results show that the proposed system can effectively detect violent activities in surveillance videos with good accuracy. The system is capable of automatically recording the timestamp and capturing the frame when a violent event is detected. This feature can assist security personnel in identifying incidents quickly and responding to dangerous situations in a timely manner.

Although the system performs well in many cases, certain challenges remain when dealing with complex environments such as crowded scenes, low lighting conditions, and partial occlusions. Future improvements can include the use of more advanced deep learning models and larger datasets to enhance the accuracy and robustness of the system.

Overall, the proposed violence detection system demonstrates that combining object detection and deep learning-based activity recognition can significantly improve the efficiency of surveillance monitoring. The system can help reduce the workload of human operators and support faster detection of violent incidents in real-world environments.

REFERENCES

[1] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[2] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in *Advances in Neural Information Processing Systems*, 2014.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[7] M. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2012.

[8] M. Cheng, J. Cai, and M. Yang, "RWF-2000: An open large-scale video database for violence detection," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2020.

[9] S. Sudhakaran and O. Lanz, "Learning to detect violent videos using convolutional long short-term memory," in *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017.

[10] M. Z. Zaheer, A. Mahmood, M. Astrid, and S. I. Lee, "Claws: Clustering assisted weakly supervised learning with normalcy suppression for anomaly detection in surveillance videos," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[13] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, and D. Cremers, "FlowNet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[14] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2018.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.