CRIME DETECTION PROJECT USING YOLOV8 AND 3D CNN

Prof. (Dr.) Deepak Uplaonkar, Associate Professor, Department of Computer Engineering, International Institute of Information Technology (I2IT), Pune.
Prof. (Dr.) Sandeep Patil, Associate Professor, Dept. Of Computer Science, International Institute of Information Technology (I2IT), Pune
Ms. Tanvi Ingale, Mr. Vedant Chaudhari, Mr. Hrutvij Kakade, Ms. Riya Ahire, Bachelor of Engineering, Dept. Of Computer Science, International Institute of Information Technology (I2IT), Pune

ABSTRACT

Intelligent Crime Detection System Using YOLOv8 and 3D CNN
This project presents an intelligent crime detection system designed using advanced machine learning techniques for object detection and 3D Convolutional Neural Networks (3D CNNs) for action recognition. The system classifies video-based data into three categories: normal, violent, or weaponized activities.The dataset preparation begins with extracting frames from video footage and annotating them based on their content. YOLOv8 is employed to detect objects and identify suspicious activities within individual frames, while 3D CNN captures the temporal progression of actions, allowing the system to recognize dynamic movements. By integrating these two models, the system ensures robust detection of both static and dynamic crime elements.The implementation is deployed in a real-time environment using a Streamlit application, enabling users to upload video feeds for live analysis. In cases of violent or weaponized activities, the system triggers email alerts to designated recipients for timely intervention.This project demonstrates that machine learning can significantly enhance public safety. The combination of object detection and temporal action recognition proves to be a highly effective approach for real-time crime detection.

Keywords: Crime detection, YOLOv8, 3D CNN, Object Detection, Action Recognition, Real-Time Surveillance, -Video Analysis, Machine Learning, Public Safety.

I. Introduction

With the increased demand for security in public spaces, it has become a necessity to develop efficient and automated crime detection systems. Traditional surveillance methods are slow and error-prone, as they require continuous human monitoring, especially in busy environments. As a result, machine learning and deep learning techniques have gained popularity due to their ability to automate real-time detection and analysis of suspicious activities.This project introduces an advanced crime detection system that leverages the YOLOv8 Object Detection Model and 3D Convolutional Neural Networks (3D CNNs) for action recognition with video data. YOLOv8 is highly effective in detecting objects and activities within individual frames, while 3D CNN captures the temporal dynamics of actions, offering a deeper understanding of events as they unfold over time. By combining these two models, the system efficiently differentiates between normal, violent, and weaponized activities, enabling faster and more accurate crime detection in real-world surveillance scenarios.The system is integrated into a user-friendly real-time application that allows seamless video feed analysis. Additionally, the application features automatic alerting, notifying authorities about potentially dangerous activities involving suspected individuals. This integration of object detection and action recognition models creates a robust crime detection solution, ultimately enhancing public safety and security.

II. Motivation

Traditional surveillance systems are the most important security tools. However, it is pretty evident that the system lacks a strong advantage in real-time crime detection. Urban crimes like theft and vandalism are increasing day by day. There is an urgent need for automation of systems. Although there are large numbers of CCTV cameras deployed throughout the cities, such systems are operated by human operators to monitor video feeds—a process vulnerable to human error and fatigue. In studies, it has shown that as much as 45% of major events may go unnoticed in just 20 minutes of continued monitoring. Advanced deep models can improve efficiency in this respect.

We wish to do this research primarily to overcome the challenges found within a manual surveillance system. Some technological advancements, like machine learning on object detection and action recognition, have opened the scope for implementing a more automated solution in identifying objects or action detection that performs at a pace faster and quicker than a human; one of them is through the real-time object detection model by YOLOv8. Finally, 3D-CNN introduces the possibility of analyzing video sequences over time; therefore, the system can determine if there is an activity that is ongoing, such as a break-in or violent confrontation.

There is the motivation of making huge reductions by creating a completely automated detection system. The purpose of these highly advanced models integrated into a real-time system is to minimize human interference, enhancing crime detection accuracy and reporting. The result will be reducing mistakes when detecting crimes and contributing to speeding up the timespan towards preventing further incident escalation. The surveillance system would be more efficient and reliable in enhancing public safety in high-risk areas, such as retail stores, ATMs, and public areas.

III. Problem Statement

Traditional surveillance systems, which are primarily manual-based, make real-time detection of violent or weaponized activities inefficient and prone to errors. This project aims to develop an automated crime detection system using YOLOv8 for object detection and 3D CNN for action recognition to perform accurate real-time analysis on video feeds. The system will also generate timely alerts for swift intervention during potential criminal incidents, enhancing overall security and response efficiency.

IV. Literature

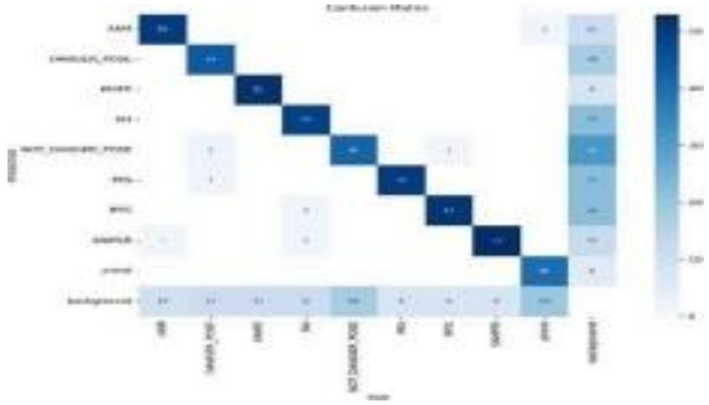

Fig. 1 Confusion Matrix

Sr.no	*Dataset Size*	*Algorithms used*	*Accuracy*
1.	Contains gun, grenade, and knife images	Proposed PELSF- DCNN, DCNN, DLNN, RNN,ANN	94.2%
2.	UCF-Crime: 1,900 videos,17.68 hours,multiple violence types	3DCNN + ConvLSTM	92.2%
3.	UCF Crime dataset	3DConvNets	45%

It acts as an important source while developing an automated system intended to identify crime. The dataset encompasses nearly 3,500 video clips, summarizing over 15 hours of video material, categorized into three main classes: Normal, Violent, and Weaponized. Diversity plays a crucial role in the machine learning aspect, ensuring effective generalization across different environments, lighting conditions, and behavioral contexts. This enhances the model's robustness against false positives and false negatives in real-world applications.

The first step in utilizing this dataset is frame extraction, a process where individual frames are extracted from video files. This transformation converts temporal video data into static images, allowing for detailed analysis. Literature suggests that high-quality input data is essential for improving model performance (Bertasius et al., 2021). Following extraction, an annotation process is conducted, where each frame is labeled as Normal, Violent, or Weaponized. These annotations serve as the ground truth for training detection models, enabling them to learn how to accurately distinguish different activities. Proper labeling is critical, as it directly impacts the learning capability of machine learning models, affecting their predictability and reliability in real-time scenarios.

"You Only Look Once (YOLO)" is one of the most renowned frameworks for real-time object detection. Its ability to perform single evaluation predictions for class probabilities and bounding boxes makes it highly efficient. YOLOv8 improves upon its predecessors by achieving greater accuracy and speed. In weapon detection, Chunchwar et al. (2024) highlight the model's efficiency in detecting and locating weapons across various scenarios. The authors note that YOLOv8's high-speed processing is crucial for analyzing surveillance footage in real time. The model's architecture ensures high detection accuracy with low computational overhead, allowing it to run efficiently even on standard hardware. This efficiency makes YOLOv8 highly suitable for real-world applications, particularly in resource-constrained environments where computational power is limited.

	Statistics of spatially Augmented UCF Crime dataset	
Number of videos	1040	80
Total Length(sec)	86.34	7.627
Min/Max length	0.6/1165.0	7.62/3600
Average length(sec)	56.60	184.90

Table 2: Statistics of Spatially Augmented UCF Crime Dataset for our Training

3D Convolutional Neural Networks (3D CNNs) offer the advantage of carrying temporal information, enabling the analysis of sequential frames to capture complex actions occurring over time. Their architecture is specifically designed to address both spatial and temporal dimensions, making them highly applicable in video classification tasks. A study by Hwang and Kang (2023) effectively demonstrates the use of 3D CNNs for anomaly detection in surveillance videos. By incorporating attention mechanisms, this approach enables the model to focus on the most relevant frames while ignoring irrelevant ones. Such mechanisms enhance efficiency during the action classification process, particularly for recognizing activities like fighting or weapon usage, which might otherwise be overlooked when processing single frames. Since 3D CNNs can process multiple frames simultaneously, they provide a deeper understanding of actions, thereby improving detection accuracy.

The integration of YOLOv8 with 3D CNN represents a significant advancement in crime detection methodologies. This hybrid approach leverages the strengths of both models, enhancing overall system performance. YOLOv8 efficiently localizes and classifies objects in individual frames, while 3D CNNs capture the temporal evolution of actions across multiple frames. According to Dugyala et al. (2023), this integrated model effectively differentiates between normal and violent activities. These models are deployed within a user-friendly application, such as Streamlit, to analyze uploaded video feeds in real-time. When violent or weaponized actions are detected, the system automatically sends alerts to relevant recipients, ensuring swift intervention and enhanced public safety.

Ultimately, state-of-the-art machine learning techniques in crime detection go beyond just AI technologies; they emphasize the continuous refinement and assessment of these models. The Smart City CCTV Violence Detection Dataset, among other comprehensive datasets, provides researchers with opportunities to merge various detection algorithms, paving the way for highly sophisticated urban monitoring systems. These advancements enable real-time threat detection, ensuring efficient surveillance and enhanced security in modern cities.

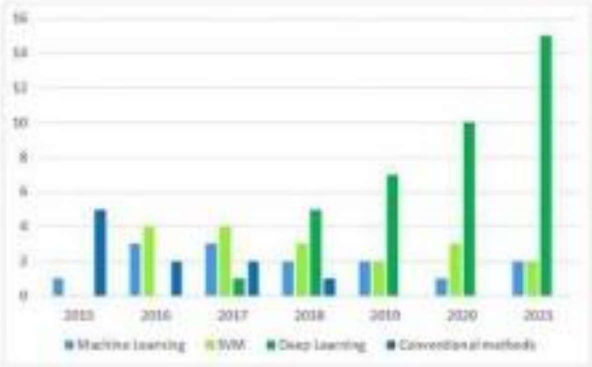

Fig. 2. Distribution of Papers on Violence Detection per Year

V. Related Work

Automated Deep Learning for Anomaly Recognition: This research focuses on multi-class anomaly recognition using a fine-tuned 3D ConvNet model. This approach addresses the gap in previous binary classification methods by capturing both spatial and temporal features within video frames. The model demonstrates superior accuracy on the UCF Crime dataset, facilitated by frame-level labeling and spatial augmentation. However, limitations include the lack of extensive dataset labeling for specific anomaly types. Future studies are recommended to investigate semi-supervised learning for improved feature extraction and generalization. [1]

3D CNN and ConvLSTM Hybrid Model (Paper 2): A hybrid model combining 3D CNN and ConvLSTM achieved high detection accuracy across five large-scale datasets, with training accuracy reaching 100% and recognition reliability rates of 98.5%, 99.2%, and 94.5%. This model demonstrates improved detection accuracy compared to standalone 3D CNN models. Future enhancements will focus on predictive anomaly detection capabilities to support broader applications in real-time surveillance. [2]

Real-Time Weapon Detection System: Utilizing YOLOv8 for real-time weapon detection, this system incorporates an email alert mechanism and a Streamlit interface, achieving a Mean Average Precision (MAP) of 0.78. The study emphasizes the importance of machine learning for public safety, with future improvements directed at enhancing adaptability to diverse environmental conditions and expanding applications in various security settings. [3]

PELSF-DCNN Model for Weapon Detection: The proposed PELSF-DCNN model achieves a high accuracy of 97.5% and precision of 96.8% for weapon detection, with YOLOv8 also providing enhanced detection performance. Currently, the model differentiates guns but is unable to distinguish between real and dummy weapons. Future work includes integrating an ensemble classifier to improve weapon-type differentiation and broaden the model's utility for various weapon detection tasks. [4]

3D CNN-Based Anomaly Detection: This study introduces a 3D CNN network with an attention module for detecting violent behaviors in real-time video surveillance, particularly focusing on extracting spatiotemporal features. The network showed improved classification performance compared to traditional 2D CNNs. A key limitation noted is memory constraints, which restrict the number of frames that can be processed per video. Future work involves exploring methods to optimize image size and video duration to enhance model performance, with potential applications in industrial safety monitoring. [5]

A study proposed a method for anomaly detection using a location-based approach, contrasting with traditional object-based methods, where objects are first identified before classification decisions are made. The proposed architecture characterizes and models object behavior at the pixel level based on motion information. To extract motion details, the background was subtracted, and a Gaussian model was used for spatiotemporal feature extraction. Anomalies were identified using a Hidden Markov Model (HMM), incorporating directional information such as speed and object size. The study achieved 91.25% accuracy using a confusion matrix. However, two key disadvantages were noted: first, optical flow methods are computationally complex for real-time applications without specialized hardware; second, background subtraction often leads to the removal of essential parts of the image. [6]

Another study proposed a framework for detecting anomalies in traffic data, analyzing vehicle movement patterns. The method used datasets from various urban areas and employed an unsupervised learning approach, eliminating the need for labeled data. The proposed system identified anomalous vehicle behavior in real-time trajectories using the Kalman Filter to mitigate occlusion problems. For spatial feature learning, a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was applied to distinguish regular from irregular movement patterns. However, the proposed anomaly detection model was inefficient in terms of computational cost at testing time and was limited to detecting only one class of anomaly (traffic patterns). [7]

Another study introduced a new feature descriptor called Histogram of Optical Flow and Magnitude Entropy (HOFME) to address various anomalous scenarios. The HOFME descriptor extracts spatiotemporal features from surveillance videos using optical flow (OF) information to analyze normal activity patterns. A nearest neighbor search algorithm was employed to classify the extracted features as normal or abnormal. This study introduced a 3D matrix representation for optical flow data, where each row and column corresponded to different orientation and magnitude ranges. [8]

VI. Requirements

A. Dataset

The dataset consists of videos categorized into normal, violent, and weaponized activities. It includes frame extraction and annotations for training machine learning models. This annotated dataset serves as the foundation for training YOLOv8 and 3D CNN, enhancing the system's ability to accurately detect and classify real-time criminal activities.

B. Libraries Used

- TensorFlow: Deep learning framework.
- Keras: High-level neural networks API.
- PyTorch: Deep learning library for research.
- OpenCV: Computer vision library.
- YOLOv8: Object detection model.
- NumPy: Numerical computations library.
- Pandas: Data manipulation tool.
- Matplotlib: Plotting and visualization library.
- Seaborn: Statistical data visualization.
- Streamlit: Web app framework.
- smtplib: Email sending library.
- scikit-learn: Machine learning library.

VII. Results

We evaluate the performance of a hybrid approach in video surveillance-based crime detection by integrating YOLOv8 and 3D CNNs. While the object detection capabilities of YOLOv8 enable real-time crime activity identification with considerable accuracy, 3D CNN enhances temporal feature extraction, improving anomaly detection efficiency.This hybrid approach significantly boosts the accuracy of crime identification, creating a robust framework suitable for smart surveillance systems in dynamic urban environments. By combining spatial and temporal analysis, this model effectively addresses the challenge of accurate and timely crime detection, paving the way for automated security solutions with real-world applications.

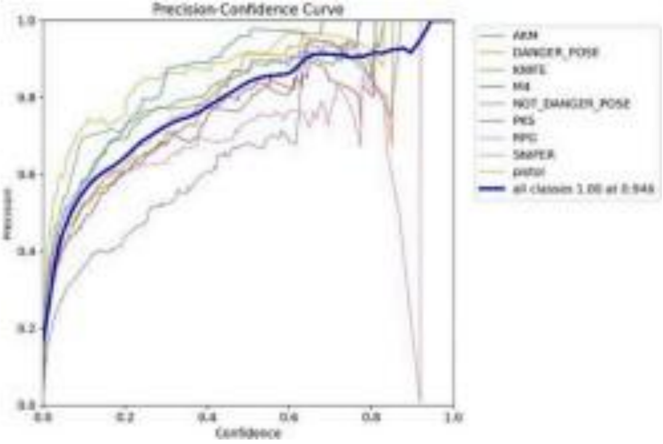

Fig. 3 Precision Confidence Curve

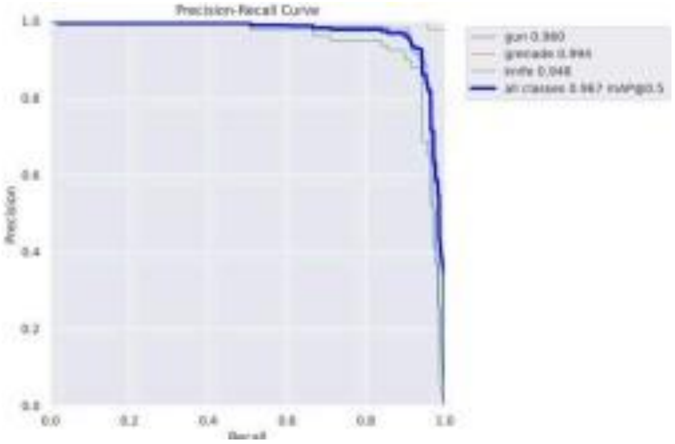

Fig. 4 Precision-Recall Curve

VIII. Conclusion

This project successfully demonstrates the development of an intelligent crime detection system by integrating YOLOv8 for object detection and 3D Convolutional Neural Networks (CNNs) for action recognition. The system efficiently classifies video-based activities into three categories: normal, violent, or weaponized, addressing the limitations of traditional surveillance systems that rely heavily on manual monitoring. By combining spatial and temporal analysis, the system ensures robust real-time crime detection, enabling timely alerts and possible intervention during dangerous incidents.The integration of both models into a real-time application not only enhances the accuracy of detection but also highlights the practical applications of machine learning in improving public safety. This project demonstrates how advanced AI techniques can contribute to the development of smart surveillance systems, providing a scalable solution for crime prevention and security enhancement across various public and private environments in the future.

References

[1] Ramna Maqsood, Usama Ijaz Bajwa (Corresponding Author), Gulsha Saleem – "Manuscript Title: Anomaly Recognition from Surveillance Videos using 3D Convolution Neural Network."

[2] Mahareek, E. A., El-Sayed, E. K., El-Desouky, N. M., & El-Dahshan, K. A. (2024). – "Detecting Anomalies in Security Cameras with 3DCNN and ConvLSTM." Al-Azhar University &

Canadian International College CIC.

[3] Chunchwar, P., Shelare, U., Nagpure, A., Patil, R., Dhole, D., & Shete, R. M. (2024). – "Real-Time Weapon Detection Using YOLOv8 and Alert Mechanism." International Journal for Research in Applied Science & Engineering Technology (IJRASET), 12(IV), April 2024. Available at real-time-weapon-detection-using-yolov8-and-alert-mechanism.

[4] Dugyala, R., Reddy, M. V. V., Reddy, C. T., & Vijendar, G. (2023). – "Weapon Detection in Surveillance Videos Using YOLOV8 and PELSF-DCNN." 3S Web of Conferences, 391, 01071. https://doi.org/10.1051/e3sconf/202339101071.

[5] Hwang, I.-C., & Kang, H.-S. (2023). – "Anomaly Detection Based on a 3D Convolutional Neural Network Combining Convolutional Block Attention Module Using Merged Frames." Sensors, 23(23), 9616. https://doi.org/10.3390/s23239616.

[6] Mahadevan, V., Li, W., Bhalodia, V., & Vasconcelos, N. (2010). – "Anomaly Detection in Crowded Scenes." In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1975–1981). IEEE.

[7] Farooq, M., Khan, N., & Ali, M. (2017). – "Unsupervised Video Surveillance for Anomaly Detection of Street Traffic." International Journal of Advanced Computer Science and Applications (IJACSA), 12(8), 270–275.

[8] Colque, R., Caetano, C., de Andrade, M., & Schwartz, W. R. (2016). – "Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos." IEEE Transactions on Circuits and Systems for Video Technology, 27(3), 673–682.