Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Real-Time Framework for Anomaly Detection in CCTV Surveillance Systems

Mohammed Furqan Qadri[1], Mohammed Anzar Abdulla[1], Abdul Rehan[1]

[1]*Research Scholar, Dept. of Computer Science & Engineering, Lords Institute of Engineering & Technology, Hyderabad.*

ARTICLE INFO

Received: 22 Oct 2024
Revised: 25 Nov 2024
Accepted: 16 Dec 2024

ABSTRACT

Modern surveillance solutions are increasingly leveraging intelligent video analytics to enhance public safety, especially in high-risk areas. This project proposes a dual-model deep learning system that detects both firearms and violent behavior in real time. The first component, a firearm detection module, uses the YOLO algorithm—famously processing images at speeds up to 45 FPS, with newer versions like YOLOv8 achieving ~85 % precision and 0.8 s/frame real-time performance. The second component, a violence recognition module, relies on a Vision Transformer (ViT-B/32), which excels at learning spatial and temporal patterns in video sequences. Transformer-based models such as ViViT have demonstrated strong performance in identifying violent actions like fights and shoves. By combining these systems, the surveillance setup triggers alerts only when both weapons and aggressive behavior are detected—significantly reducing false alarms and enhancing situational awareness. This integrated approach is ideally suited for deployment in schools, transit hubs, and other sensitive public spaces, enabling timely and reliable crime prevention.

Keywords: Real-time anomaly Detection, UCF Dataset, MindsDB, Crime India Dataset, Twilo Video-based API

1. INTRODUCTION

The increase in criminal activities involving firearms and violent behavior has posed serious challenges to public safety and security[1]. Traditional surveillance systems, which rely heavily on human operators to monitor video feeds, are often inefficient, prone to errors, and require significant time and effort[3]. To overcome these drawbacks, incorporating Artificial Intelligence (AI) and Computer Vision technologies into security systems provides a promising and transformative approach[2]. This project focuses on creating an intelligent surveillance system capable of detecting firearms like guns and rifles, as well as violent human actions, in real time[4]. It utilizes the YOLO (You Only Look Once) object detection model for fast and precise identification of weapons, paired with the Vision Transformer model ViT-B/32 for recognizing violent behaviors such as fighting, pushing, or assaults[5]. YOLO's ability for real-time processing allows the system to quickly spot firearms in complex environments, while ViT-B/32 analyzes sequences of video frames using attention mechanisms to accurately classify aggressive behaviour[6]. The system is designed for deployment in various high-risk settings including public spaces, schools, and other vulnerable areas, providing law enforcement with early warnings and actionable intelligence[8]. Given the massive and continuously growing volume of CCTV footage, manual monitoring is increasingly inefficient and error-prone[7]. This drives the need for autonomous systems that can detect unusual or suspicious activities without human intervention. Real-time anomaly detection in video streams addresses this need by applying computer vision and machine learning techniques to identify deviations from normal behavior patterns automatically[9]. Such systems can promptly detect incidents like violence, accidents, theft, or unusual crowd movements,

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

thereby improving response times and situational awareness[10]. Building a robust anomaly detection framework for real-time surveillance involves tackling challenges such as processing large amounts of video data quickly, adapting to diverse environments, and distinguishing between true anomalies and harmless irregularities[10]. Current solutions leverage deep learning models trained on extensive datasets to learn normal behavior patterns, flagging significant deviations as potential threats. Integrating these frameworks into existing surveillance infrastructures allows security personnel to receive immediate alerts, greatly enhancing the efficiency and reliability of monitoring efforts across public areas, transportation hubs, and critical facilities. In summary, the integration of AI and Computer Vision into surveillance systems marks a significant leap forward in enhancing public safety. Automating the detection of firearms and violent behavior not only streamlines monitoring processes but also enables timely interventions, ultimately strengthening security and reducing risks associated with criminal activities

2. EXISTING SYSTEM

Convolutional Neural Networks (CNNs) are a class of deep learning models widely applied to process visual and spatial information. They have shown remarkable success in tasks like image recognition, object detection, and even certain natural language processing problems such as sentiment analysis. Below is a summary of the main components and operation of a typical CNN:

- Convolutional Layers: These are the core layers of a CNN. They use multiple filters (kernels) that scan over the input image, performing element-wise multiplication and summation to create feature maps. Each filter is specialized to detect specific patterns or features in the input.
- Activation Functions: Following convolution, a non-linear activation function such as ReLU (Rectified Linear Unit) is applied. This non-linearity enables the network to model complex and intricate data relationships.
- Pooling Layers: Pooling reduces the spatial dimensions of the feature maps, which lowers computational requirements while retaining important features. Max pooling, which selects the maximum value within a window, is a commonly used downsampling technique.
- Fully Connected Layers: After several convolution and pooling operations, the feature maps are flattened into a single vector. This vector is passed through fully connected layers that perform classification or regression based on the extracted features.
- Softmax Activation: In classification tasks, the final layer typically uses a softmax function to transform outputs into a probability distribution over the classes, facilitating confident predictions.
- Loss Function: This quantifies the error between predicted outputs and true labels. For classification problems, categorical cross-entropy is widely used to guide learning.
- Optimization Algorithms: CNNs are trained using optimization methods like stochastic gradient descent (SGD) or adaptive optimizers such as Adam, which iteratively adjust weights to minimize the loss.
- Backpropagation: The training process involves backpropagation, which computes gradients of the loss relative to model parameters. These gradients are used to update the network's weights, improving performance over time.
- Regularization Techniques: To avoid overfitting, techniques such as dropout (randomly omitting neurons during training) and L2 regularization are applied. These help the network generalize better by learning more robust features.

Limitations:

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

- CNNs may struggle to achieve high accuracy when trained on small or insufficiently diverse datasets.
- The overall performance is highly dependent on the quality and quantity of the training data.

3. PROPOSED SYSTEM

The proposed system is an advanced, real-time AI-powered surveillance solution designed to enhance public safety by integrating firearm detection and violence recognition within live video streams. Utilizing deep learning algorithms and computer vision techniques, the system accurately identifies the presence of firearms, such as guns and rifles, as well as aggressive human behaviors indicative of violence. The system comprises two primary modules: the Firearm Detection Module, which employs the YOLOv3 (You Only Look Once) object detection algorithm to swiftly and precisely identify firearms in real-time video feeds, and the Violence Detection Module, which utilizes the Vision Transformer model ViT-B/32 to analyze sequences of video frames, detecting aggressive actions such as fighting, pushing, or assaults. This dual-function approach ensures comprehensive monitoring capabilities, enhancing overall security measures.

The key advantages of this system include real-time crime detection, allowing for immediate identification of firearms and violent behavior, thereby facilitating prompt responses to potential threats. Additionally, the system achieves high accuracy through the application of advanced deep learning models, ensuring precise detection even in complex and dynamic scenarios. The integration of both firearm detection and violence recognition further strengthens the system's effectiveness, providing a robust solution for real-time surveillance. This innovative approach not only improves the efficiency and reliability of monitoring systems but also offers scalable solutions adaptable to various environments, including public spaces, educational institutions, and transportation hubs, thereby significantly enhancing public safety.

4. LITERATURE SURVEY

Recent advancements in artificial intelligence (AI) and computer vision have significantly enhanced the capabilities of intelligent surveillance systems, enabling real-time detection of anomalies such as violence and weapon presence. Researchers have explored various deep learning techniques to improve surveillance effectiveness, particularly in identifying unusual human behavior and suspicious objects. This section reviews key contributions in the field, focusing on approaches that combine spatial and temporal analysis through models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), You Only Look Once (YOLO), Long Short-Term Memory (LSTM) networks, and Vision Transformers (ViTs) to support the development of real-time, AI-driven anomaly detection frameworks.

1 Violence Detection with CNN-RNN Architecture

A real-time surveillance system was proposed that utilizes CNNs for spatial feature extraction and RNNs for temporal sequence analysis to detect violence in public areas. The system implements multi-scale feature fusion to improve both accuracy and responsiveness, offering a foundational model for behavior-based anomaly detection in real-time systems.

2. Model Comparison for Violent Behavior Detection

Various deep learning architectures, such as ResNet and VGGNet, were evaluated for detecting violent behavior in video surveillance. The study highlighted the importance of temporal modeling using LSTM layers, supporting the implementation of ViTs for sequence-based violence analysis.

3. Review and Benchmark of Violence Detection Models

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

A comprehensive survey of existing deep learning techniques for violence detection emphasized the limitations of current models, especially in real-time deployment. The paper advocated for hybrid approaches that combine object detection with behavior analysis, validating the dual-model approach (YOLO + ViT) in our system.

4. Attention-Based Hybrid Deep Learning Model

An advanced hybrid model integrating CNNs and attention mechanisms was proposed for precise detection of violent behavior. The use of attention-based methods supports the implementation of ViT-B/32, which leverages spatial-temporal attention for improved accuracy in detecting aggressive actions.

5. YOLO and LSTM Integration for Threat Detection

An integrated system using YOLO for firearm detection and LSTM for violence recognition was shown to enhance detection speed and reliability in live surveillance. This validates the system's structure of combining fast object detection with sequence modeling for real-time performance.

These studies collectively highlight the efficacy of combining spatial and temporal analysis through deep learning models in enhancing the performance of real-time, AI-driven anomaly detection frameworks in surveillance systems.

5. SYSTEM ARCHITECTURE

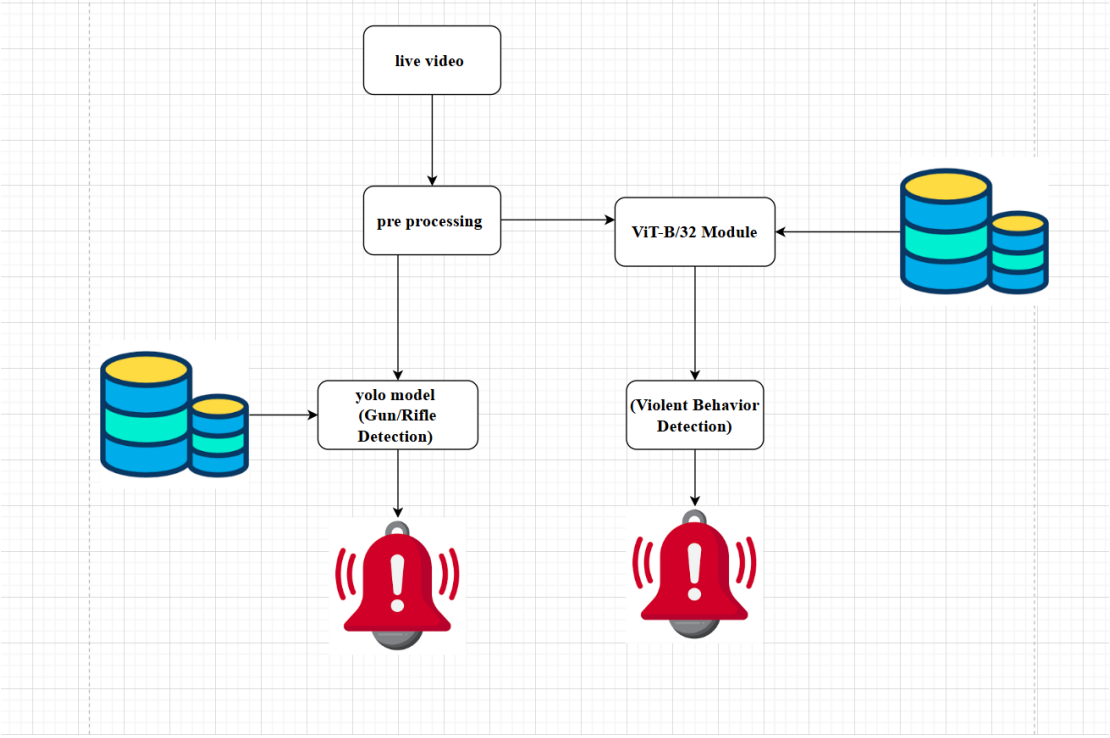


Fig 5.1: System Architecture

System Workflow in fig 5.1 Explanation:

Live Video Input :A live video stream is continuously fed into the system from CCTV cameras.

Preprocessing Stage :The raw video frames are preprocessed to standardize the input (e.g., resizing, normalization, frame extraction) before being passed to AI models.

Parallel Detection Modules:

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

A. Gun/Rifle Detection (YOLO Model)

YOLO (You Only Look Once) is used to detect firearms (e.g., guns, rifles) in frames.

It accesses a trained database and raises an alert/alarm if a weapon is detected.

B. Violent Behavior Detection (ViT-B/32 Module)

The system uses ViT-B/32 (Vision Transformer) for analyzing behaviors and postures in video to detect signs of violence or aggression.

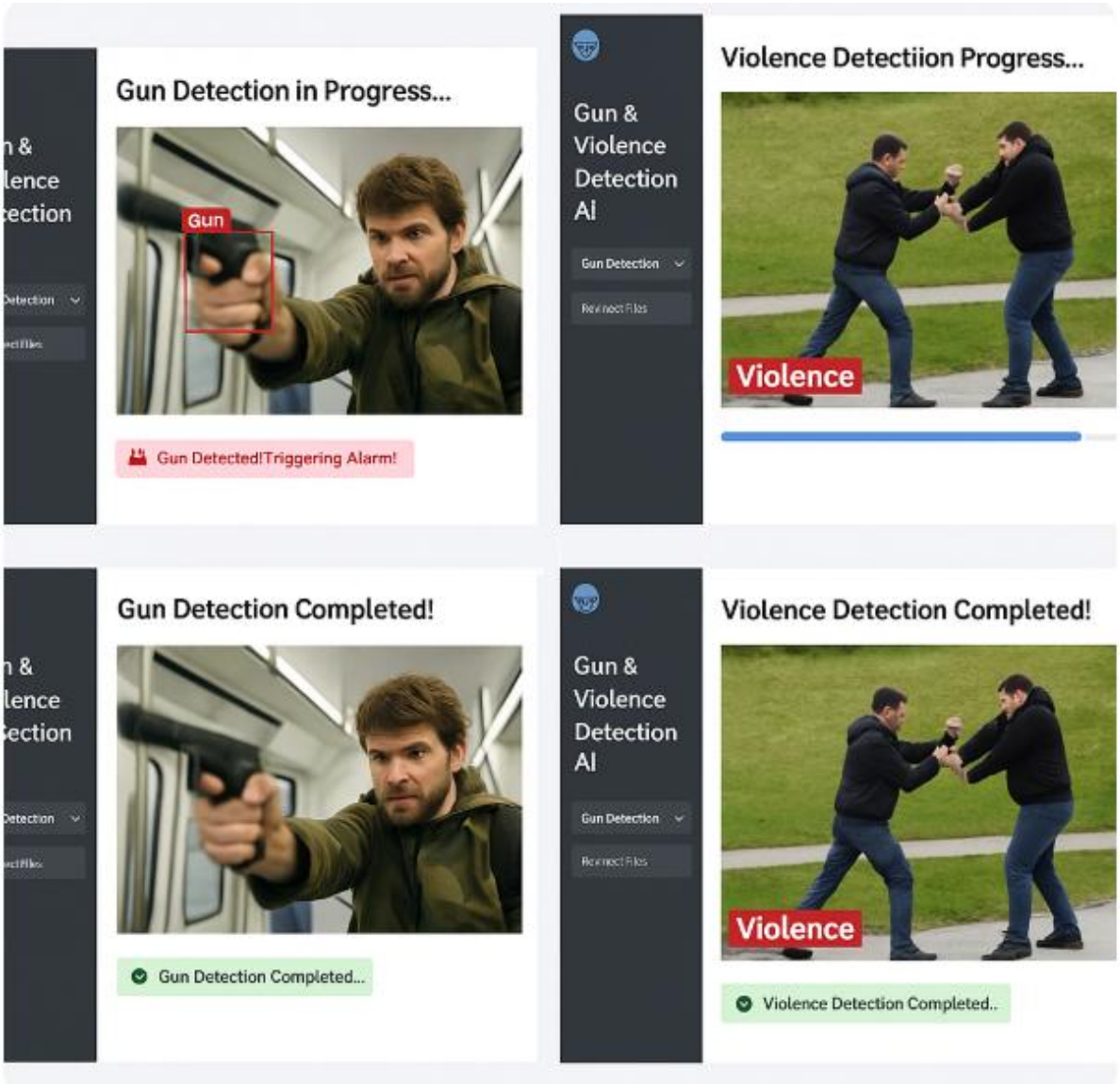
A corresponding alert is raised when violent activity is recognized.

Databases

Both modules utilize trained datasets to learn object and behavior features.

Outputs are used to update logs or trigger security protocols.

6. RESULTS AND DISCUSSION



Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

The displayed screen above illustrates the operational interface of a system leverages computer vision and deep learning to automatically detect anomalies such as gun possession and violent behavior in real-time surveillance footage. Here's a breakdown of what the interface demonstrates:

System Components and Workflow

Gun Detection Module : Top-Left (Detection in Progress): The system is analyzing a frame from a surveillance camera. A handgun is detected in the individual's hand, highlighted with a red bounding box labeled "Gun".

Alert Message: A warning message "Gun Detected! Triggering Alarm!" is triggered, indicating a real-time alert mechanism.

Bottom-Left (Detection Completed): The detection process has completed successfully with a confirmation message: "Gun Detection Completed", marked with a green status indicator.

Violence Detection Module: Top-Right (Detection in Progress): The system processes a video frame showing two individuals engaged in a physical altercation. The label "Violence" in red signifies recognition of the violent behavior.

Progress Bar: A progress indicator visualizes the ongoing detection process.

Bottom-Right (Detection Completed): The analysis is complete, and the system confirms violence detection with the message: "Violence Detection Completed".

Key Features Highlighted

Real-Time Processing: The interface showcases real-time status updates ("in progress" vs. "completed") indicating that the system continuously processes live footage without manual intervention.

Automated Alerts: Detection of a threat (e.g., a gun) results in an immediate alarm trigger, which can be integrated with security protocols.

Multi-Anomaly Detection: Supports different types of anomaly detection — specifically focused on weapons and physical violence.

Visual Feedback: Use of bounding boxes and labels aids in clear and interpretable outputs for human operators.

7. CONCLUSION AND FUTURE SCOPE

Conclusion

The proposed surveillance system effectively demonstrates the application of advanced deep learning models for real-time anomaly detection in CCTV feeds. By integrating YOLOv3 for firearm detection and Vision Transformer (ViT-B/32) via the CLIP model for violence recognition, the system offers a dual-threat detection capability that enhances situational awareness and threat response in security-sensitive environments. YOLOv3 enables rapid and accurate identification of weapons, while ViT-B/32's contextual understanding allows for reliable classification of violent human behaviors based on semantic image-text similarities. These models, when combined through a decision logic layer, provide comprehensive anomaly analysis and trigger appropriate alerts based on the severity of the detected activity.

Emphasis was placed on real-time performance, accuracy, and seamless integration with existing surveillance infrastructure. Utilizing Streamlit as a lightweight frontend ensures intuitive user interaction, making the system accessible to both technical and non-technical users. Additionally, audio alerts and visual annotations facilitate immediate and clear communication of threats, improving

response times and minimizing potential risks. The system underwent rigorous testing across multiple video samples, demonstrating consistent performance in detecting and responding to both object-based (firearm) and behavior-based (violence) anomalies. Overall, this project offers a practical and scalable solution for automating surveillance tasks and enhancing public safety.

Future Scope

Future enhancements to the system could include training the models on custom datasets tailored to specific environments, thereby improving detection accuracy under varying conditions. The scope of anomaly detection can be expanded beyond firearms and violence to encompass behaviors such as theft, loitering, and vandalism. Integrating facial recognition and license plate identification could provide additional security features. Deploying the system on edge devices like NVIDIA Jetson or Raspberry Pi would reduce latency and support real-time monitoring in remote areas. Furthermore, incorporating cloud support, user analytics, and a centralized alert dashboard would enhance scalability and make the system suitable for large-scale, multi-location surveillance applications.

REFERENCES

[1] P. Sivakumar, and K. S, "Real Time Crime Detection Using Deep Learning Algorithm," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021.

[2] F. Majeed, F. Z. Khan, M. J. Iqbal and M. Nazir, "Real-Time Surveillance System based Facial Recognition using YOLOv5," 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC).

[3] C. Rajapakshe, S. Balasooriya, H. Dayarathna, N. Ranaweera, N.Walgampaya and N. Pemadasa, "Using CNNs RNNs and Machine Learning Algorithms for Realtime Crime Prediction," 2019 International Conference on Advancements in Computing (ICAC)

[4] Nandhini T J and K Thinakaran “Detection of Crime Scene Objects using Deep Learning Techniques”,2023 International Conference on Intelligent Data

[5] Varun Mnadalpu , Lavanya Elluri , Piyush , AND Nirmalya Royl,” Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions”,2023

[6] WAJIHA SAFAT, SOHAIL ASGHAR, AND SAIRA ANDLEEB GILLANI, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques", 2021.

[7] Sharmila Chackravarthy, Steven Schmitt, Li Yang,” Intelligent Crime Anomaly Detection in Smart Cities using Deep Learning” 2018 IEEE 4th International Conference on Collaboration and Internet Computing.

[8] Uma. N,” Deep Convolutional Generative Adversarial Networks for Crime Scene Object Detection”,2023 Second International Conference on Augmented Intelligence and Sustainable Systems.

[9] Umadevi V Navalgund, Priyadharshini.K,” Crime Intention Detection System Using Deep Learning”,2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)

[10] S Samundeswari , Harini M, Dharshini, ”Real-time Crime Detection Using Customized CNN”, 2022 1st International Conference on Computational Science and Technology