



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Enhancing Public Safety: Real-Time Crowd and Violence Detection Using YOLOv8 with Automated Alerts and Notifications

Mohammed Siraj F

Aalim Muhammed Salegh College of Engineering
Chennai, India

Sayed Abuu Backer Siddiq M

Aalim Muhammed Salegh College of Engineering Chennai, India

Mentor :

1. Kiruthika N

Aalim Muhammed Salegh College of Engineering Chennai, India

Shameer Khan I

Aalim Muhammed Salegh College of Engineering Chennai, India

2. G.Divya.ME

Aalim Muhammed Salegh College of Engineering

Chennai, India

Abstract— Public safety remains a critical concern in modern society, especially in densely populated public areas such as transportation hubs, stadiums, and marketplaces. Traditional surveillance systems often rely on manual monitoring, which is time-consuming and prone to human error. This project introduces a real-time crowd and violence detection system using the YOLOv8 (You Only Look Once, Version 8) algorithm, a state-of-the-art object detection model known for its speed and accuracy. The system monitors live video feeds from CCTV or IP cameras, identifying situations involving

crowd buildup or violent behavior such as fighting or aggressive gestures. Once detected, the system triggers an automated alert that includes real-time data like timestamp, location, and event type, which is sent to security personnel through SMS, email, or mobile notifications. This solution ensures immediate response, reduces human dependency, and enhances situational awareness. The integration of artificial intelligence and real-time communication allows for efficient crowd control and violence prevention, thereby improving overall public safety. The project demonstrates how deep learning and automation can work together to

create safer environments by providing authorities with actionable intelligence in real time.

1. INTRODUCTION

Public safety in densely populated areas is an ongoing challenge, especially in places where large crowds gather. In such environments, traditional security systems that rely on human operators to monitor live video footage are often inefficient and prone to error. Human security personnel may not be able to detect potential risks immediately, resulting in delayed responses to emergencies or violent incidents. This project addresses this challenge by developing a real-time surveillance system based on YOLOv8, an advanced deep learning model for object detection. YOLOv8 excels at detecting objects in images with remarkable speed and accuracy, making it an ideal solution for real-time monitoring of public spaces. The primary goal of this system is to automatically detect dangerous situations such as overcrowding, fights, or physical aggression and send instant alerts to authorities or emergency responders. These alerts are sent via SMS, email, or push notifications, providing real-time data including the location, type of incident, timestamp, and a snapshot of the event, enabling faster intervention. By utilizing YOLOv8, the system is capable of processing video frames rapidly and accurately, detecting crowds and violent behavior with minimal false positives. The system operates with minimal latency, making it suitable for environments that require quick decision-making, such as transportation hubs, stadiums, and public gatherings. Furthermore, the system's architecture allows for easy integration with existing surveillance infrastructure, making it a scalable solution that can be deployed in various settings with little modification. The introduction of an intelligent, AI-based monitoring system has the potential to transform how public safety is managed in crowded spaces, offering a proactive

approach that reduces response times, prevents incidents, and enhances overall security. With the growing need for advanced security systems that can respond quickly and effectively to threats, this project provides a timely and relevant solution, demonstrating how AI can improve public safety and reduce human error in high-risk environments

2. BACKGROUND

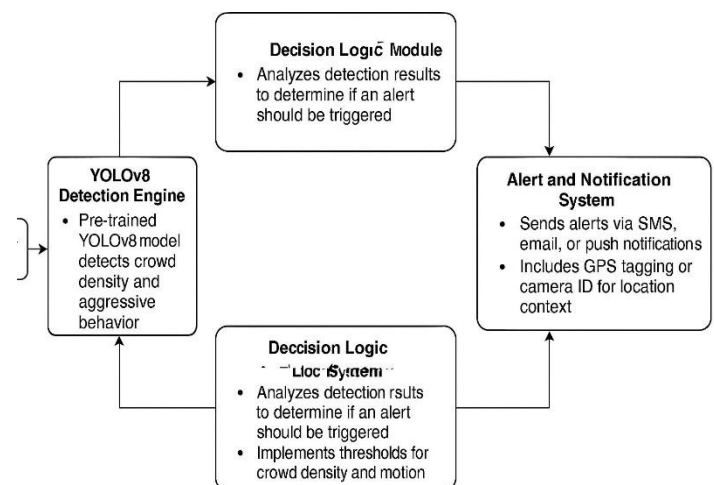
In recent years, the field of computer vision has made significant strides, particularly in the application of deep learning models for real-time object detection. YOLO (You Only Look Once) is one of the most widely used algorithms in this domain, known for its ability to perform both object classification and localization in a single pass through the network. YOLOv8, the latest version developed by Ultralytics, improves upon previous iterations by offering faster processing speeds, higher accuracy, and the ability to detect more diverse objects in complex environments. Its use in surveillance systems is particularly promising, as it allows for quick, real-time detection of various objects, making it ideal for high-stakes scenarios such as monitoring public spaces for safety threats. Previous research has mainly focused on individual aspects of surveillance, such as crowd detection and violence detection, but the integration of both capabilities into a single system remains rare. Crowd detection has traditionally involved methods such as density estimation, head counting, and optical flow analysis. These methods, while effective, are often computationally expensive and may struggle to handle real-time video feeds in dynamic environments. On the other hand, violence detection typically involves analyzing spatiotemporal features or using deep learning-based approaches to recognize aggressive behavior such as pushing, hitting, or fighting. These methods often require large amounts of labeled data and can be prone to false positives. By combining both crowd detection and violence detection into a unified framework using YOLOv8, this

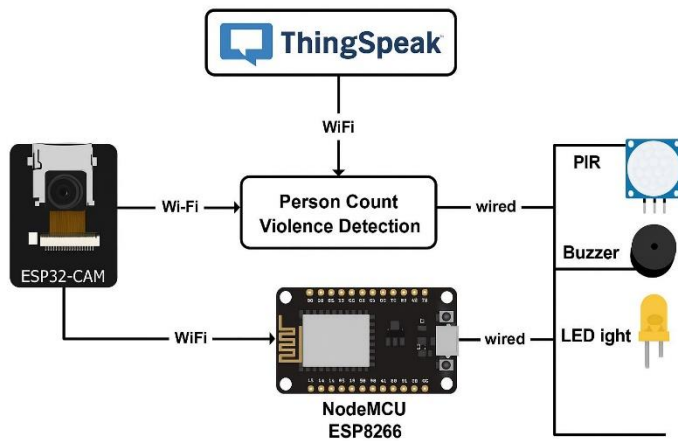
project aims to create a more efficient and accurate solution. YOLOv8's ability to process video frames rapidly and detect multiple objects in real-time is a key advantage, enabling the system to identify both crowded areas and violent behavior simultaneously. Furthermore, the addition of an automated alert system enhances the effectiveness of the solution by notifying authorities or security personnel instantly when a potential threat is detected. This combination of cutting-edge object detection, machine learning, and real-time communication represents an advanced approach to public safety surveillance, addressing the shortcomings of traditional systems and offering a scalable, automated solution for modern security needs.

3. ARCHITECTURE

The architecture of the proposed real-time crowd and violence detection system is designed to ensure seamless processing of live video feeds and rapid decision-making. The system consists of several components that work together to provide accurate, real-time threat detection and instant alerting. The first component is the **video input layer**, where live video streams from surveillance cameras are captured and sent to the system. These cameras can be installed in high-risk areas such as public transportation hubs, arenas, or crowded public events. The video feed is then passed to the **preprocessing module**, which performs essential image processing tasks such as resizing, normalization, and frame extraction to prepare the video frames for input into the YOLOv8 model. The preprocessing module ensures that the frames are in the correct format and resolution, optimizing them for detection without causing excessive latency. The core of the system is the **YOLOv8 detection engine**, which uses the YOLOv8 model to analyze each frame in real-time. YOLOv8 excels at identifying and classifying multiple objects within the frame, including crowds, fights, and other forms of violent behavior. Once an object is detected, the system generates a bounding box around it and assigns a label (e.g., "crowd" or

"fight"), along with a confidence score indicating the likelihood that the detection is accurate. The next component is the **decision logic module**, which interprets the output from YOLOv8. This module uses predefined thresholds to determine when an alert should be triggered. For example, if the number of people in a certain area exceeds a specified density, or if violent actions such as pushing or fighting are detected, an alert is generated. The **alert notification system** is responsible for sending these alerts to the relevant authorities. Alerts are sent through SMS, email, or push notifications and contain crucial information such as the type of threat, its location, a timestamp, and a snapshot of the event. Finally, an **optional monitoring dashboard** allows administrators to visualize the ongoing analysis and review historical alerts. This dashboard provides a user-friendly interface for managing multiple camera feeds, reviewing alerts, and tracking the status of detected events. The modular architecture of this system ensures scalability and flexibility, allowing it to be deployed in various environments with minimal setup. The system is also designed to work efficiently on both cloud and edge computing platforms, enabling real-time processing even in resource-constrained settings.





The architecture of the proposed public safety surveillance system is a hybrid framework that combines **IoT-based sensor devices**, **edge computing with camera modules**, **AI-powered visual analytics using YOLOv8**, and **cloud communication** via ThingSpeak. The goal of this design is to detect motion, analyze live video streams, recognize overcrowding or violent incidents, and immediately trigger physical alerts in the environment. All hardware components are integrated using a unified Wi-Fi network for efficient real-time communication.

System Overview

The system consists of the following major components:

1. **ESP32-CAM Module** – Handles live video capture and AI-based object/violence detection.
2. **NodeMCU ESP8266** – Acts as the main controller for managing input sensors and triggering alerts.
3. **PIR Motion Sensor** – Detects movement and activates the monitoring process.
4. **Buzzer** – Alerts through sound signals in case of threats.
5. **LED Light** – Acts as a visual indicator in emergency situations.
6. **ThingSpeak Cloud Platform** – Facilitates data exchange between NodeMCU and ESP32-CAM.

All these devices are connected wirelessly via

the same Wi-Fi network, enabling them to send and receive signals seamlessly. This architecture supports **event-driven automation**—meaning all operations are triggered by real-world occurrences like motion or visual anomalies.

Step-by-Step Architecture Workflow

1. Idle State and Monitoring Initialization

In its idle state, the system remains in a passive monitoring mode. The **NodeMCU ESP8266** is continuously powered and runs a loop to monitor the **PIR sensor**, which is connected to one of its digital pins (e.g., D2). The PIR sensor is responsible for detecting any motion in its field of view. It outputs a digital HIGH signal when motion is sensed, which is then processed by the NodeMCU. Upon detecting motion, the NodeMCU immediately connects to the **ThingSpeak IoT cloud platform** and updates a specific channel or field (e.g., field1 = 1) to act as a remote trigger. This update serves as a message or signal for the ESP32-CAM to activate.

2. ESP32-CAM Activation and Video Streaming

The **ESP32-CAM**, a microcontroller with a built-in camera and Wi-Fi capability, remains in a listening or polling state. It continuously checks the status of the ThingSpeak field associated with it. When it reads the value "1", it interprets this as a trigger to begin video capture and real-time processing.

Once activated, the ESP32-CAM captures live video from its environment and performs two major operations using the **YOLOv8 (You Only Look Once, version 8)** algorithm:

- **Crowd Counting:** YOLOv8 detects human figures in the video frame and counts the number of people.
- **Violence Detection:** Using a trained dataset, the model also identifies patterns or poses indicative of physical aggression, such as punching, kicking, or fighting gestures.

The model can run either on the ESP32-CAM directly (if optimized for lightweight inference) or offloaded to a server or local machine for advanced processing.

3. Threshold Logic and Decision Making

Once the ESP32-CAM completes the object detection step, it evaluates two key conditions:

- **Condition 1 – Overcrowding:**
If the number of people detected exceeds a predefined threshold (e.g., more than 5 individuals), this triggers a warning. A signal is sent to the NodeMCU via ThingSpeak or MQTT communication to activate the **buzzer** for crowd alert.
- **Condition 2 – Violence Detected:**
If any frame is classified as showing **violent activity**, this is considered a high-priority alert. In this case, the ESP32-CAM sends a signal to the NodeMCU to activate both the **buzzer and the LED light**, signaling an emergency to nearby security personnel or the public.

The NodeMCU receives these instructions and accordingly switches its GPIO pins HIGH to power the alert components.

4. Alert and Notification System

Once either or both alert conditions are met:

- The **buzzer** emits a loud sound to draw immediate attention.
- The **LED light** begins flashing or remains steadily lit (depending on the configuration) as a visual warning.

This setup ensures that even in the absence of an operator or security guard, **real-time warnings** are provided on-site to prevent escalation of a situation.

The alerts remain active for a set duration (e.g., 30 seconds) or until motion is no longer detected and the YOLOv8 model no longer detects threats. After this, the system resets and returns to its idle monitoring mode.

Component-Level Design and Connectivity

1. PIR Sensor to NodeMCU

- **VCC** → 3.3V on NodeMCU
- **GND** → GND on NodeMCU
- **OUT** → Digital pin (e.g., D2) on NodeMCU

The PIR sensor detects motion within a range of 5–7 meters and triggers the OUT pin high when activity is sensed.

2. Buzzer and LED to NodeMCU

- Buzzer connected to Digital Pin D3 (via NPN transistor if high current needed).
- LED connected to Digital Pin D4 (with a current-limiting resistor).

Both are activated by setting their pins HIGH in the NodeMCU code.

3. ESP32-CAM Setup

The ESP32-CAM is programmed using Arduino IDE with support for Wi-Fi. It reads the status from ThingSpeak to determine whether to start capturing video. For higher accuracy or complex inference, it may stream frames to a local PC or cloud server with YOLOv8 installed.

Communication Protocols Used

- **HTTP:** Used to update and read values from the ThingSpeak platform.
- **Wi-Fi (802.11):** For real-time wireless communication between ESP32-CAM, NodeMCU, and the internet.
- **GPIO Signaling:** Used to control physical components like buzzers and LEDs.
- **Optional MQTT/REST APIs:** Can be used for faster communication in advanced versions.

4. CONCLUSION

This project demonstrates the significant potential of artificial intelligence, specifically YOLOv8, in enhancing public safety through real-time crowd and violence detection. Traditional surveillance systems often rely on human operators to monitor video feeds, a process that can be inefficient and prone to errors. By utilizing YOLOv8, which offers fast and accurate object detection, the proposed system automates the detection of potential threats, such as overcrowding and violent behavior, and immediately notifies authorities. This ability to provide real-time alerts based on the analysis of video frames helps reduce response times, enabling faster intervention by security personnel or emergency responders. Furthermore, the integration of both crowd detection and violence detection into a single model is a novel approach that enhances the system's effectiveness and reduces the complexity of deploying multiple specialized systems. The architecture of the system, which includes modular components for video input, preprocessing, detection, decision-making, and alerting, ensures that it can be easily scaled and adapted to different environments, from transportation hubs to large public events. This flexibility makes the system a practical solution for a wide range of public safety applications. In the future, the system could be further improved by incorporating additional features such as facial

recognition, predictive analytics, and audio-based threat detection. These improvements would further enhance the system's ability to respond to emergencies and identify perpetrators. Overall, the project showcases the potential of AI and deep learning in transforming public safety infrastructure, offering a proactive and efficient solution to address the growing challenges of maintaining security in crowded public spaces.

REFERENCES

1. **Sridevi, B. (2019).**
A Novel Linear SVM-Based Compressive Collaborative Spectrum Sensing (CCSS) Scheme for IoT Cognitive 5G. *Soft Computing*, 23(18), 8515–8523. Springer.
<https://doi.org/10.1007/s00500-018-3534-1>
2. **Karthikeyan, S. (2019).**
A Survey on Despeckling Filters for Speckle Noise Removal in Ultrasound Images. In *Proceedings of the IEEE International Conference on Communication and Signal Processing (ICCSP)* (pp. 605–609). IEEE.
<https://ieeexplore.ieee.org/document/8697914>
3. **Sangeerani Devi, A., Prakash, S., Laavanya, K., Shali, A., & Sathish Kumar, D. (2019).**
Violence Detection and Target Finding using Computer Vision. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(5S3), 123–127.
https://www.academia.edu/125750759/Violence_Detection_and_Target_Finding_using_Computer_Vision
4. **Valarmathi, B., Kshitij, J., Dimple, R., Srinivasa Gupta, N., Harold Robinson, Y., & Arulkumar, G. (2023).**
Human Detection and Action Recognition for Search and Rescue in Disasters Using YOLOv3 Algorithm. *Journal of Electrical and Computer Engineering*.
<https://vit.ac.in/2023-2>
5. **SharathKumar, Y.H., & Naveena C. (2023).**
A Deep Learning Based System to Estimate Crowd and Detect Violence in Videos. In A. Biswas, V.B. Semwal, & D. Singh (Eds.), *Artificial Intelligence for Societal Issues* (Vol. 231). Springer, Cham.
https://doi.org/10.1007/978-3-031-12419-8_3