

**ROBUST HUMAN TARGET DETECTION**

A PROJECT REPORT

**Submitted By:**

**Ritesh Ghosh , Roll No: 2151160**

**Sreya Das, Roll No: 2151180**   
**Soham Ganguly, Roll No: 2151196 Aditya Baidya, Roll No: 2151194**

**Under Supervision of:**   
**Prof. Anurina Tarafdar**   
 **Department of Computer Science Engineering**

**In partial fulfillment for the award of the degree**

**Of**   
**BACHELOR OF TECHNOLOGY**

**IN**   
**COMPUTER SCIENCE AND ENGINEERING**   
**HERITAGE INSTITUTE OF TECHNOLOGY, KOLKATA**

**DECEMBER, 2024**



**HERITAGE INSTITUTE OF TECHNOLOGY, KOLKATA**

**MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY**

**BONAFIDE CERTIFICATE**   
**Certified that this project report “Robust Human Target Detection” is the bonafide work of “Mr. Ritesh Ghosh, Ms. Sreya Das, Mr. Soham Ganguly and Mr. Aditya Baidya” who carried out the project work under my supervision.**

|  |  |
| --- | --- |
|  | |
| **Prof. (Dr.) Subhashis Majumder** | **Prof. Anurina Tarafdar** |

|  |  |
| --- | --- |
| **Head of the Department Computer Science &** **Engineering** | **Assistant Professor Computer Science &Engineering** |
|  | |

**Examiner**

**DECEMBER, 2024**

1

**ACKNOWLEDGEMENT**

**The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success.**

**We would take this opportunity to thank our principal, Prof. Dr. Basab Chowdhury, for allowing us to form a group of four people and for providing us with all the necessary facilities to make our project work and of worth.**

**We will be thankful to our Head of the Department, Prof. Dr. Subhashis Majumder, for constantly supporting and giving us invaluable insights. His words of**   
**encouragement motivated us to achieve our goal and impetus to excel.**

**We are grateful to our project guide Prof. Anurina Tarafdar for the guidance, inspiration and constructive suggestions that helped us in the partial fulfillment for the award of the degree.**

**………………………………………**   
 **Ritesh Ghosh**

**………………………………………**   
 **Sreya Das**

**………………………………………**   
 **Soham Ganguly**

**………………………………………**   
 **Aditya Baidya**

2

**TABLE OF CONTENTS**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LIST OF FIGURES​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **4** |

**CHAPTER**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **I.​ Abstract​​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **5** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **II.​ Introduction​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **6** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **III.​ Literature Survey​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **7** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **IV.​ Scope of the project​** | **​** | **​** | **​** | **​** | **​** | **​** | **9** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **V.​ Input and output​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **13** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **VI.​ Methodology​​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **16** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **-​ Advantages​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **20** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **VII.​ Result​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **21** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **VIII.​ Future Plan ​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **24** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **IX.​ Conclusion** | **​** | **​** | **​** | **​** | **​** | **25** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X.​ Supplementary Material​ ​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **26** |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **XI.​ References​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **​** | **33** |

3

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Index** | **Page** |
| 1 | Proficient need for detection | 6 |
| 2 | Expected Action detection forms | 9 |
| 3 | Overall Working | 11 |
| 4 | Average Precision by Class | 13 |
| 5 | Bounding box detection (colour specific) | 14 |
| 6 | Model Performance | 16 |
| 7 | Annotated Train batch file | 17 |
| 8 | Workflow | 17 |
| 9 | User-friendly Dashboard | 19 |
| 10 (a) | Precision - Confusion Graph | 21 |
| 10 (b) | Recall - Confusion Graph | 21 |
| 10 (c) | F1 - Confusion Graph | 22 |
| 11 | Confusion Matrix | 22 |
| 12 | Heat-map | 23 |
| 13 | Results (Loss, Precision and Recall) | 23 |

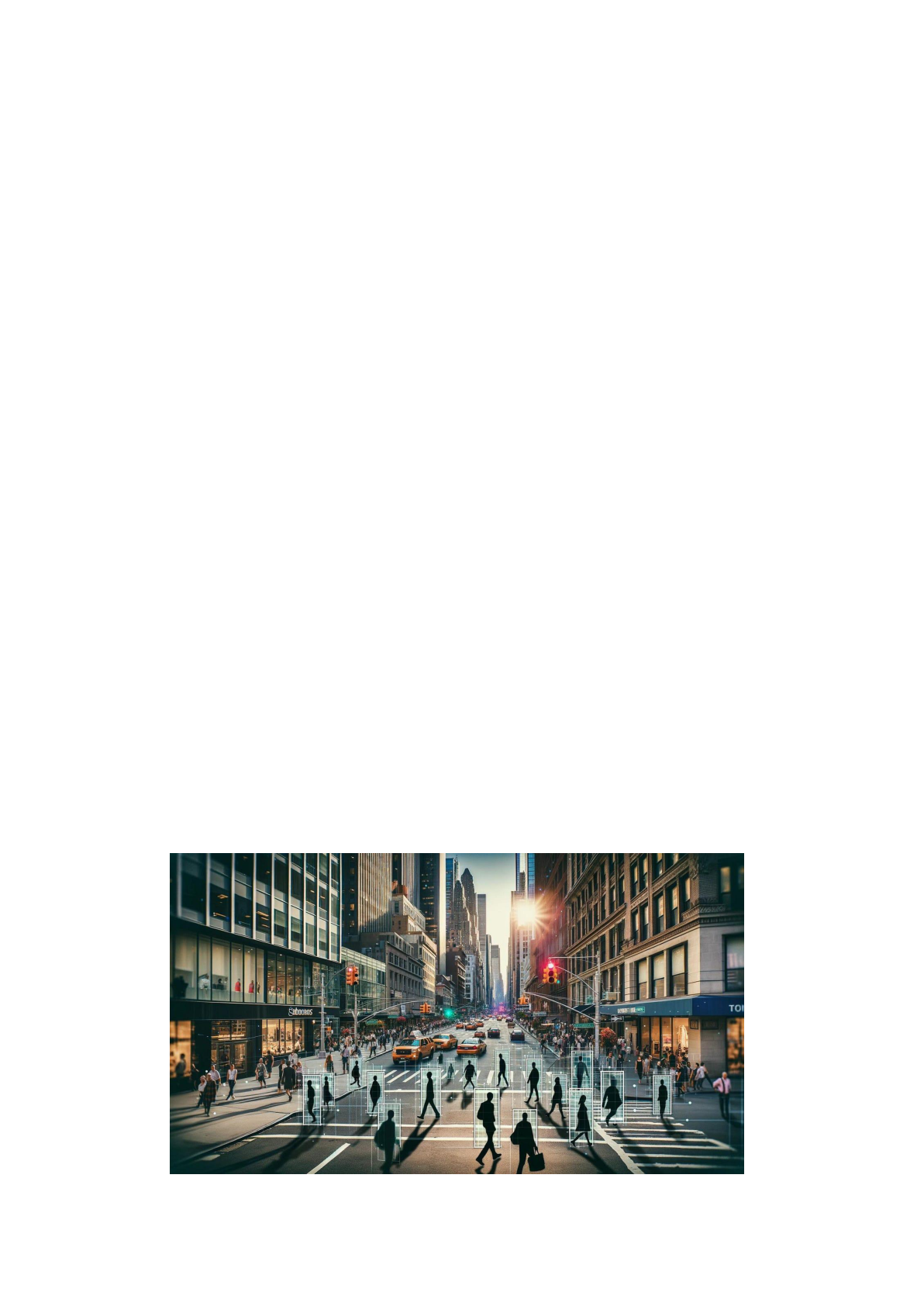
4

**Abstract of our Project**

A fundamental computer vision task with broad applications in security, healthcare, and event monitoring is abnormal activity detection in video surveillance. Surveillance systems are becoming more prevalent in crowded and dynamic environments, such as public spaces, airports, and shopping malls, which drives demand for automated systems that can identify and notify authorities about unusual or possibly dangerous activities in real time. Unlike conventional object detection assignments, abnormal activity detection seeks to find unanticipated human behavior, such as sudden running, fights, or illegal entrance, which often show infrequently and unpredictably in surveillance footage. Accurately identifying these anomalies without significant labeled data or strong computational resources is the main difficulty, which makes real-time deployment on standard hardware challenging.

Specifically using YOLOv8 (You Only Look Once, version 8) from Ultralytics and OpenCV, this project aims to solve these issues by creating a real-time abnormal activity detection system employing deep learning-based object detection and tracking methods. While preserving great accuracy and robustness in varied environments, the goal is to build a system with little computational load that can efficiently identify anomalies in video streams.

5

**​**

**Introduction**

A fast expanding area of computer vision, human action detection has uses in many different sectors including video surveillance, sports analysis, healthcare monitoring, and human-computer interaction. These systems can increase automation, boost safety, and create new user experiences by identifying and classifying human activities including walking, running, or standing.

Human action detection's complexity results from elements including variations in body posture, occlusions, lighting conditions, and different camera angles. Building a strong and precise detection system calls for sophisticated algorithms able to handle such difficulties in actual situations.

This project aims to realize an implementation of a real-time human action detection system based on the YOLOv8 object detection framework. Given its accuracy and speed, we train a custom YOLOv8 model on a dataset which is specifically tailored for three basic human actions: walking, standing, and running. The goal is to design a scalable and effective system that can reliably recognize and classify these actions within video streams.

With this project, we seek to prove the practical application of recent advances in object detection for human action recognition and lay the groundwork for further development on more complicated actions and larger datasets. The system is intended to have lightweight components for real-time processing, making it suitable for a multitude of practical implementations.

**Figure - 1**

6

**Literature Survey** 

In the scope of computer vision, human action detection is an emerging strong area of research that has seen various methods to increase accuracy and real-time response. One of the early ways of human detection was given by Dalal and Triggs with HOG [1], which stands for Histogram of Oriented Gradients. This method derives the shape and appearance of human bodies by the distribution of gradient orientation. HOG method improves detection of actions such as walking, standing and running, whether in still images or video sequences. The HOG technique is widely used in pedestrian and surveillance applications because contouring and edge description of the human body is done effectively. The technique HOG is one of the vital tools for action recognition.

The introduction of the YOLO (You Only Look Once) model by Redmon et al. [2] marked a new development in object detection technology. By implementing a single-stage detection framework, the YOLO model achieved fast processing speed, accuracy, and real-time object detection performance. This led to the model being far more efficient when timed against traditional multi-stage detectors, making it ideal for real-time applications. The evolution of YOLO is now in the stage of YOLOv8, which introduced new features such as anchor-free detection and mosaic augmentation. These refinements further enhanced YOLO’s capability to recognize objects in difficult environments—like bustling surveillance scenes—for effective human action detection in practical uses [3].

Convolutional Neural Networks (CNNs) have had tremendous impact on human action recognition. For instance, Karpathy et al. [4] showed the practicality of CNNs when they trained models using deep learning on massive datasets like Sports-1M which contains millions of sports videos. CNNs are particularly proficient at learning deep spatial and temporal dependencies within video sequences. This enabled the development of action detection systems which effortlessly classify complex actions, like sports events, using deep learning models. Similarly, Ng et al. [5] used Long Short-Term Memory (LSTM) networks to improve the modeling of time sequential dependencies in video data, like handshakes or dances. LSTMs significantly improved the performance of detecting actions where timing and sequence were crucial by capturing the temporal dynamics of evolving actions.

7

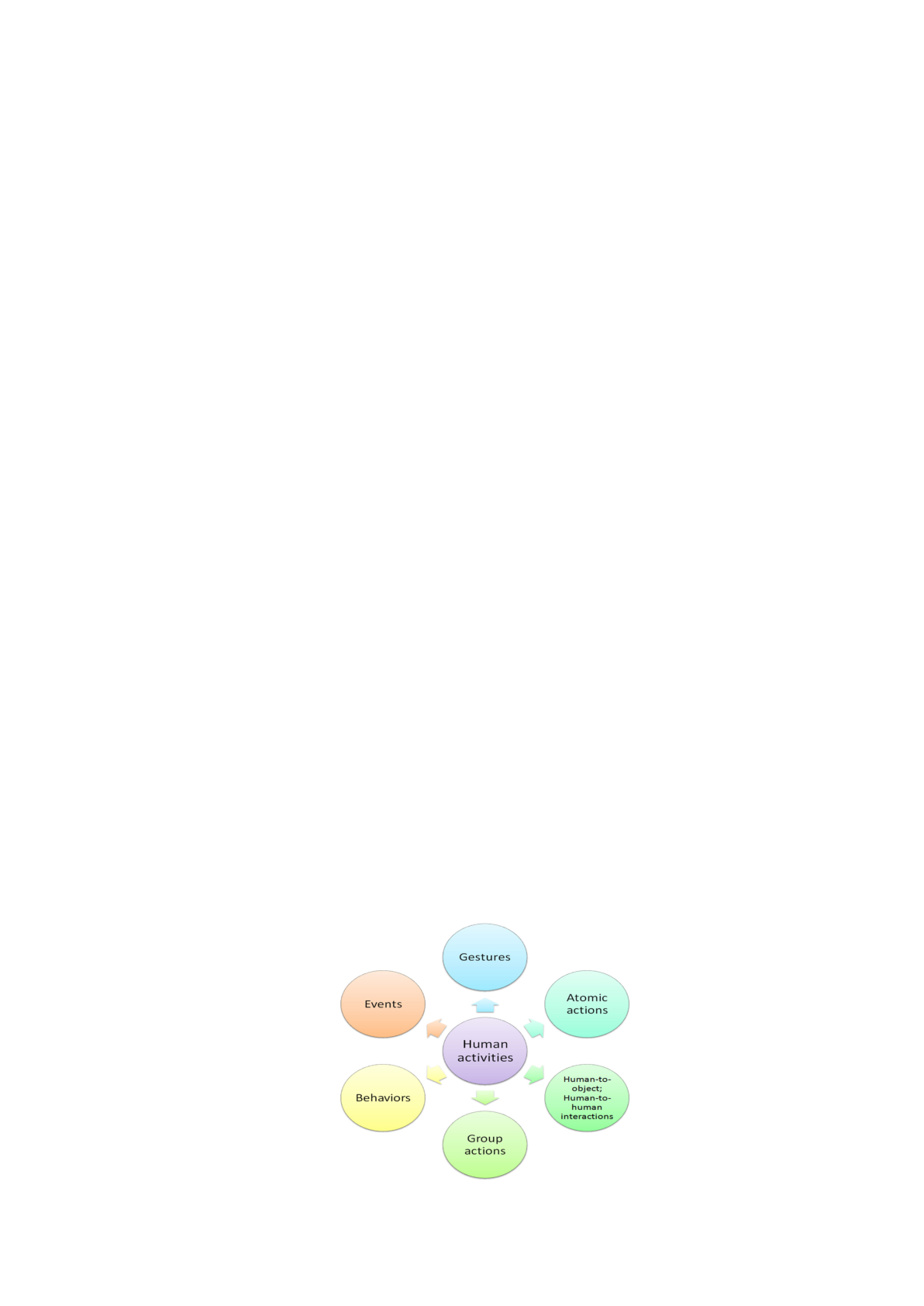
The most recent works on action detection by Liu et al. [6] integrated unsupervised clustering of motion patterns with YOLOv8, unsupervised because of the lack of considerable labeled data. This technique used YOLOv8’s object detection to cluster captured motion patterns and flagged any deviations from the normal motions as anomalies. This offered a more scalable approach for detecting unusual human actions in videos. Earlier motion detection techniques, as the optical flow introduced by Horn and Schunck [7], determined pixel shifts between video frames but were very noisy and struggled with complex activites. Lin et al. [8] built upon this by adding an optical flow and YOLO combination to enhance spatial and temporal motion capture for the detection of public abnormal behaviors.

Chen and colleagues [9] took YOLO-based action detection a step further. They added multi-scale feature extraction to YOLO models. This had an impact on detecting small-scale anomalies, like objects moving fast or slight movements even in busy or tricky scenes. YOLOv8 has also shown its worth on edge devices, as Khan and team [10] proved. By putting YOLOv8 on these devices, they spotted anomalies in real-time while cutting down on network use. This makes it useful to watch live footage in areas with poor connections or far-off places.

Sultani [11] also pushed human action detection forward. They came up with a new way to learn using multiple instance learning (MIL) without much supervision. This method could spot odd things in huge video sets with only labels for whole videos. It's much more practical in the real world where you often can't label every single frame. Because MIL can handle lots of unlabeled data, it works well to look at massive video collections.

These different studies together push forward the area of spotting human actions. They tackle issues like dealing with background noise working in real-time, and being able to handle many different settings and data sets.

8

****

**Scope of the project**

The scope of this project encompasses the design, development, and deployment of a real-time abnormal activity detection system using a custom-trained YOLOv8 deep learning model, optimized for performance, generalizability, and interpretability across a variety of real-world scenarios. The proposed system aims to go beyond traditional surveillance methods by integrating intelligent, context-aware, and visually interpretable mechanisms to monitor and respond to abnormal human behavior in dynamic environments.

This project envisions a system that is not only capable of flagging anomalies in dynamic environments—such as public spaces, healthcare facilities, or industrial areas—but is also designed to function efficiently on standard hardware or edge devices, reducing the need for high-end infrastructure or extensive retraining. The detailed goals are outlined below:

**1. Efficient Real-Time Processing**

* **Low Latency**: The system will prioritize real-time performance, ensuring that abnormal activities are detected and flagged with minimal delay. This is critical for time-sensitive applications such as live surveillance or emergency response.
* **Hardware Optimization**: By utilizing YOLOv8's lightweight architecture, the system will be designed to run efficiently on standard hardware, including edge devices like Raspberry Pi, GPUs, or CPUs, without compromising on speed or accuracy.
* **High Throughput**: The system will process multiple input streams (e.g., from CCTV cameras or drones) simultaneously, providing scalability for large-scale deployments.

**Figure - 2**

9

**2. Context-Aware Anomaly Identification**

Unlike conventional systems that only detect predefined actions, this project introduces a behavior-driven anomaly detection approach.

* **Spatial and Temporal Analysis:** Tracks motion across multiple frames to evaluate behavioral trends (e.g., someone standing in the same place too long).
* **Dynamic Crowd Awareness**: Assesses crowd density and changes in movement patterns to flag anomalies (e.g., a sudden rush or an individual sprinting in a dense crowd).
* **Proximity Detection:** Identifies close grouping of individuals in sensitive zones, potentially indicating a conflict or unauthorized gathering.
* **Environment Sensitivity:** Takes into account lighting, camera angle, and crowd density to avoid false positives in variable conditions.

**3. Data-Efficient Training and Generalization**

The system is built with efficiency in mind, particularly for real-world deployment where data collection and annotation may be limited.

* **Transfer Learning with YOLOv8:** Utilizes pre-trained weights from COCO dataset, then fine-tunes with a smaller, domain-specific Roboflow dataset to specialize in human activities.
* **Minimal Annotation Needs:** Reduces the burden of creating vast labeled datasets—abnormal behavior is inferred from context, not just labels.
* **Data Augmentation**: Enhances the training dataset with transformations such as flipping, scaling, and lighting adjustments, improving performance across diverse scenarios.
* **Robust Generalization**: Performs reliably across varied environments—indoor/outdoor, day/night, crowded or sparse—without needing re-training.

**4. Explainable Output and Visual Feedback**

Transparency is key in surveillance systems. This project includes interpretability tools that make the system's decisions clear and user-friendly.

* **Bounding Boxes and Labels**: Every detected individual is enclosed in a box with activity labels (e.g., “walking”, “fighting”) and a confidence score.
* **On-Screen Status Indicators:** Abnormal behaviors are visually marked on video feeds for real-time decision-making by operators.

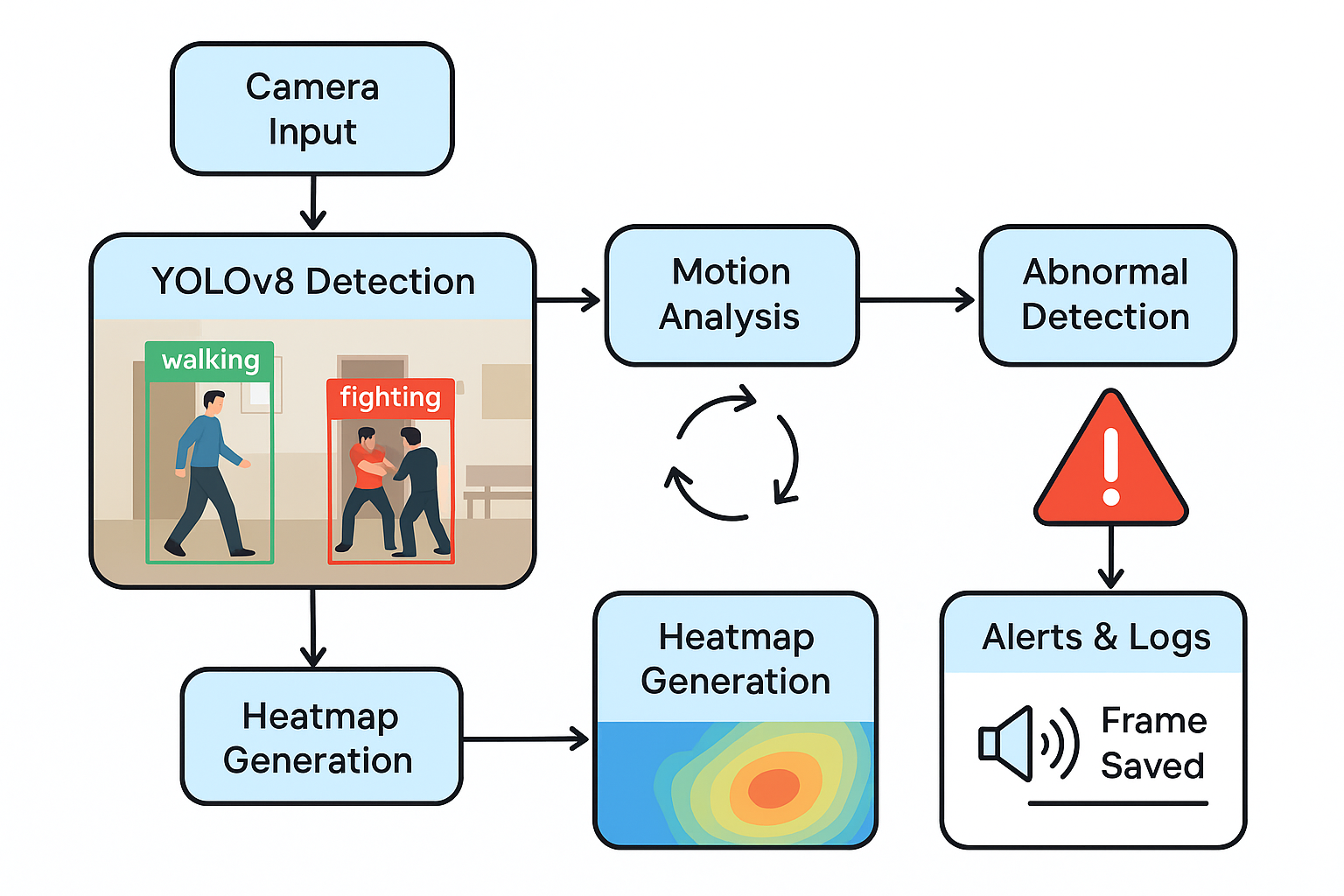
10

* **Heatmaps:** A cumulative thermal map shows the zones with highest activity or anomalies, helping in identifying hotspots or unusual movements across time.
* **Abnormal Frame Saving:** Frames with detected anomalies are saved with timestamps and activity types, creating a visual record for later reviews.

**5. Automated Alerts and Intelligent Logging**

To support rapid response and post-incident review, the system incorporates multi-modal alerting and event tracking.

* **Audio Alerts**: Beep tones are triggered when critical anomalies are detected—each event type can be assigned a distinct sound pattern.
* **Spam Control Mechanism**: Alerts include cooldown timers to avoid repeated beeping for continuous or overlapping events.
* **Logging System**: Abnormal activities are logged in a text file including:
  + Timestamp
  + Detected activity (e.g., “armed”, “loitering”)
  + Location/frame info (optional in multi-camera setups)
* **Audit-Ready Reports**: The loggedpdata, combined with saved frames, supports investigations, reporting, or ML retraining.



**Figure - 3**

11

**6. Scalable, Modular, and Cross-Domain Design**

The architecture is built to be adaptable and extensible, meeting both small-scale and enterprise-level needs across various sectors.

* **Modular Implementation**: New behaviors, rules, or detection targets can be added without restructuring the whole system.
* **Multi-Camera and Cloud Readiness**: Scales from a single camera setup (e.g., a retail shop) to large deployments (e.g., city-wide smart surveillance networks).
* **Integration with Existing Infrastructure**: Can be plugged into existing CCTV systems, IoT hubs, or cloud dashboards via standard interfaces.
* **Wide-Ranging Applications**:
  + **Public Safety**: Detecting fights, thefts, suspicious loitering in stations, streets, and airports.
  + **Healthcare**: Monitoring for patientgfalls, fainting, or unauthorized exits.
  + **Industrial Environments**: Detecting safety breaches, accidents, or restricted access in hazardous areas.
  + **Education and Retail**: Identifying crowding, aggression, or shoplifting in schools, malls, and shops.

12

**Input and Output**

**Input:**

The input to the abnormal activity detection system primarily consists of video data obtained from live surveillance feeds or pre-recorded footage. These videos may come from real-world environments such as public spaces, hospitals, offices, or educational institutions. The goalhis to detect and classify human actions within each frame in real time.

**Video Collection:**

* Videos were collected from simulated environments or sourced from surveillance systems.
* Scenarios included publiciplaces, indoor rooms, streets, and hallways under varied lighting and density conditions.

**Custom Datasets:**

To train the detection model, a custom dataset was developed using **Roboflow**, a platform that simplifies the process of annotating and managing visual data.

* Video frames were extracted and manually labeled with specific action classesglike **walking**, **running**, **standing**, **fighting**, **sitting**, **jumping**, **lying down**, **robbery**, and **armed**.
* Once labeled, theldataset was exported in YOLO format and used to train the YOLOv8s model in Google Colab.
* Roboflow also provided automated data augmentation to improve model accuracy across varied conditions.

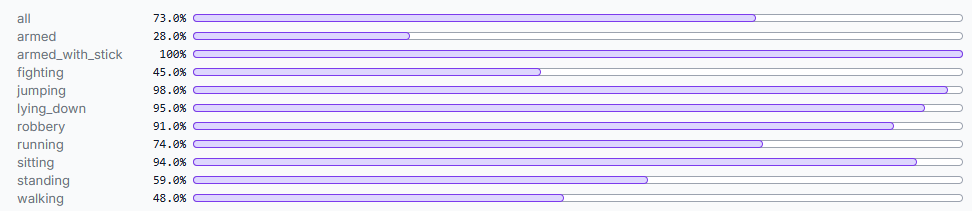


Figure : 4

13

**Output:**

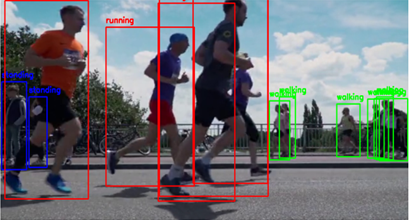
The output of the system will consist of multiple actionable results for abnormal activity detection:

* **Visual Detection and Annotation :**

The system provides real-time visual feedback by identifying individuals in the video frames and markinguthem with bounding boxes. Each detected action is labeled with a corresponding tag like "Running", "Fighting", or "Robbery", and each detection includes a confidence score indicating the system’s certainty. This helps operators interpret situations instantly and with greater accuracy.

* **Anomaly Logging and Metadata:**

In addition to visual outputs, the system maintains a structured log file that records details such as the timestamp, detected action, frame number, and the duration of the event. The log may alsojinclude contextual information like estimated crowd density or predefined thresholds for specific areas, aiding deeper analysis and better scene understanding.



**Figure - 5**

* **Saved Annotated Frames:**

Any frame in which an abnormal activity is detected is automatically saved in a designated folder. Each file is timestamped and labeled with the type of anomaly, making it easy to locate and review incidents later. The filenames follow a clear format (e.g., fighting\_12\_05\_2025\_10\_33\_21.jpg) for traceability.

14

* **Abnormality Heatmap:**

A heatmap is created based on areas with high motion or recurring abnormal activities. This visual tool helps identify hotspots or zones that require increased attention, and it can also guide future decisions about camera placement or crowd control measures.

* **Real-Time Audio Alerts:**

For high-risk actions like fights, armed threats, or robbery, the system triggers distinct audio alerts. Each type of anomaly can have a unique tone, and a built-in cooldown mechanism prevents the system from issuing repeated alerts for the same event in quick succession.

* **Performance Metrics:**

In addition to the visual outputs, the system will log performance metrics such as precision, recall, F1-score, and accuracy based on the model’s ability to correctly detect abnormal actions, which can be helpful for evaluating and fine-tuning the system.

15

**Methodology Used**

The proposed system leverages a hybrid approach combining the object detection strength of **YOLOv8** with contextual and motion-based analysis using **OpenCV**. The methodology is designed for **real-time abnormal activity detection** and operates effectively on both live surveillance feeds and recorded video streams. The system’s workflow consists of the following key components:

* **Dataset Preparation and Preprocessing:**

To train the system effectively for abnormal activity detection, a custom dataset was created using Roboflow, a platform designed for annotating and managing image datasets. Video footage was converted into frames, capturing both normal and abnormal human activities.

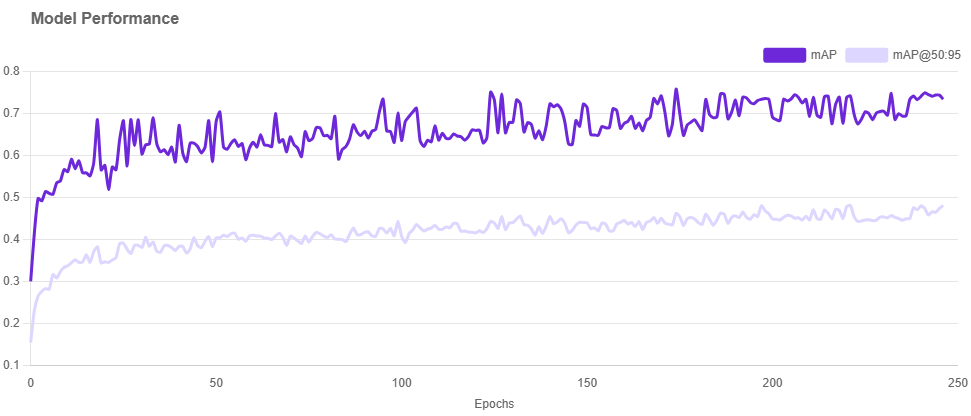
* **Classes**:

Each frame was annotated with one of nine action classes:

Walking, Standing, Sitting, Running, Fighting, Jumping, Lying Down, Robbery, and Armed.

* **Annotation Format:**
* Bounding boxes were drawn around each person.
* Labels were assigned using YOLO annotation format (text files containing object class, coordinates, and dimensions).
* **Data Augmentation**:

To improve model generalizationkand reduce overfitting, Roboflow applied several augmentation techniques like Horizontal flipping, brightness and contrast variation, rotation and scaling and random cropping.



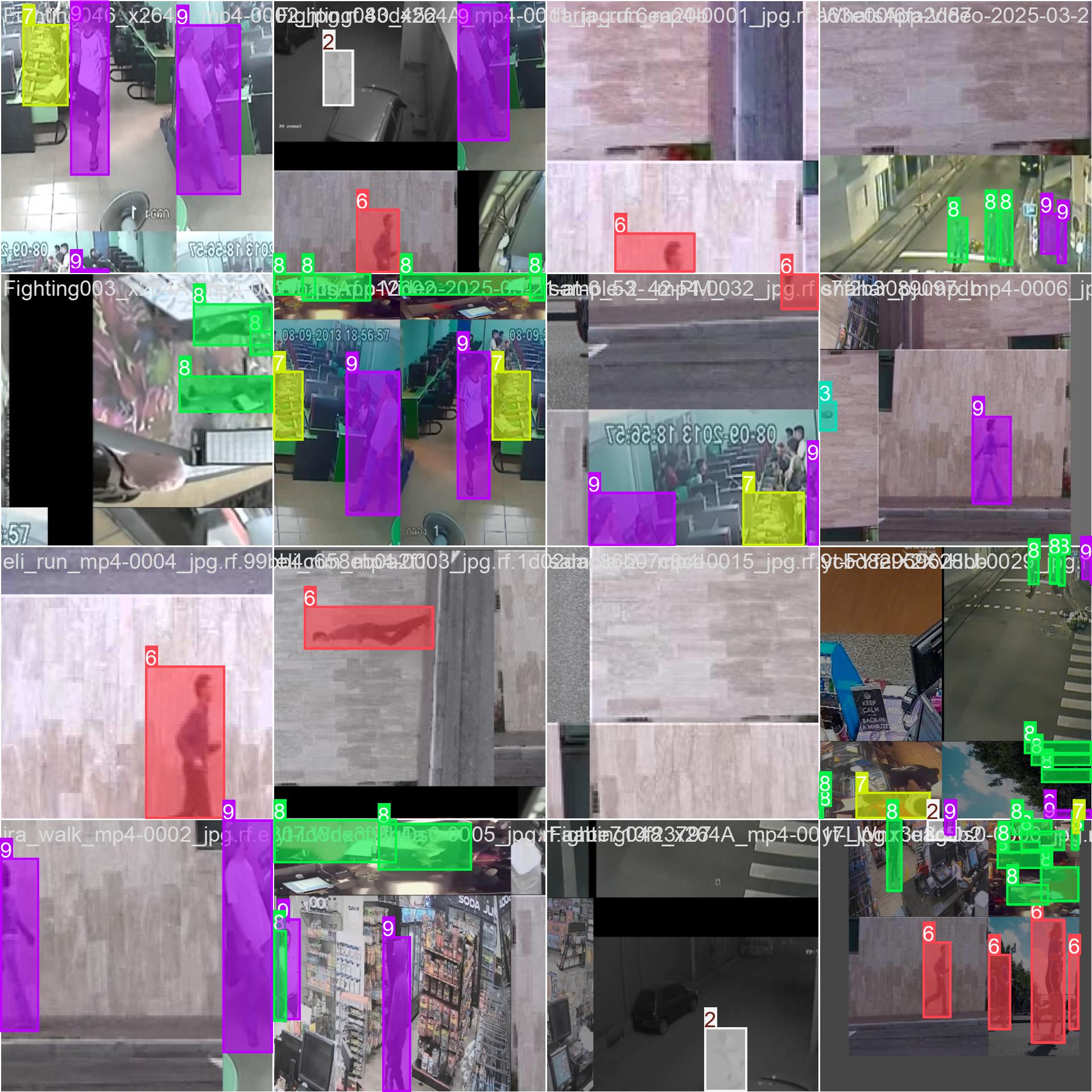
**Figure : 6**

16

* **Export Format**:

The dataset was exported in YOLOv8-compatible format, including:

* Structured folder hierarchy (train/val/test)
* Corresponding .txt annotation files
* A data.yaml file containing class names and dataset paths.

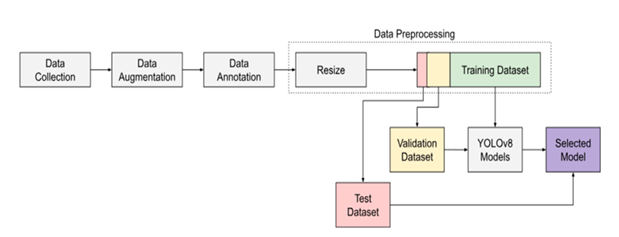


**Figure : 7**

* **Model Training Using YOLOv8:**

Model training was conducted using the YOLOv8s (small) model in Google Colab with GPU support, leveraging the Ultralytics framework.

* The training spanned 50 to 100 epochs, with a batch size of 16–32.
* Key performance metrics such as precision, recall, and mean Average Precision (mAP@0.5 and mAP@0.5:0.95) were used to monitor progress.
* The model was validated using a separate split of the dataset, and the best-performing weights were saved as **best.pt** for use in real-time detection.

****

**Figure : 8**

17

* **Contextual Preprocessing and Background Subtraction:**

To enable behavior-aware detection, the system processes environmental context using background subtraction:

* MOG2 (Mixture of Gaussians) is applied to identify moving foreground objects.
* Foreground masks are refined using morphological operations to reduce noise.
* Crowd density is calculated by comparing motion pixels to total pixels. If density exceeds pre-set thresholds (e.g., >20% during standing), the activity is flagged as potentially abnormal.
* **Abnormal Motion and Temporal Analysis:**

YOLOv8 identifies actions frame-by-frame, while OpenCV-based analysis captures behavioral anomalies across time:

* Sudden movement detection tracks the displacement of bounding box centers across consecutive frames. Large pixel shifts are interpreted as abrupt or suspicious motion.
* Temporal standing analysis monitors individuals who remain stationary beyond a defined duration (e.g., more than 10 seconds), flagging them as potentially loitering or unwell.

These methods enrich object detection with time-aware behavior context, improving abnormality detection without needing complex LSTM or 3D CNN models.

* **Group Proximity and Interaction Analysis:**

To detect potentially suspicious gatherings or social interactions:

* The system calculates pairwise distances between people using bounding box center coordinates.
* If three or more individuals cluster within a fixed radius (e.g., 100 pixels), the system marks it as a group event.
* Group proximity in sensitive areas (like near a bank counter or entrance) is treated as an anomaly.
* **Alerting, Logging, and Visual Outputs:**

To ensure usability and incident traceability, several output mechanisms are implemented:

18

* **Auditory Alerts:**

Abnormal actions like robbery, fighting, or armed threats trigger unique beep tones using the winsound.Beep() function.

* **Logging:**

All anomalies are recorded in a .txt file with timestamp and activity type.

* **Frame Capture:**

Abnormal frames are saved automatically to a designated folder (/abnormal\_frames/) using filenames that include timestamps and action labels.

* **Heatmap Generation:**

Motion masks accumulated over time are converted into thermal-style activity heatmaps using OpenCV's COLORMAP\_JET, highlighting frequent zones of movement or activity.

* **User-Friendly Dashboard:​**  
  A dashboard interface may be developed to display live video feeds alongside real-time analytics, providing operators with a centralized view of activities.

****

**Figure : 9**

19

**Key Advantages of the Methodology**

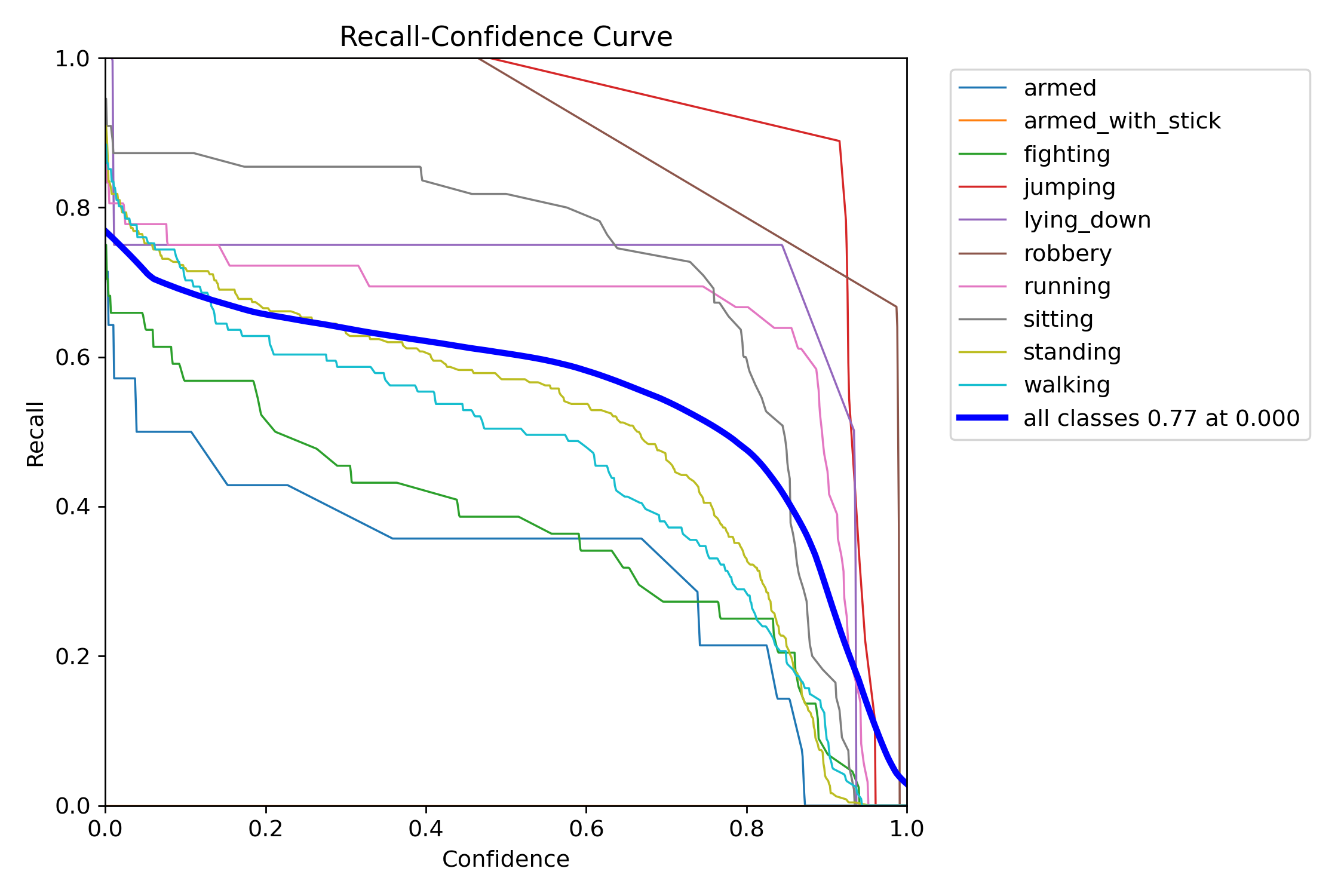
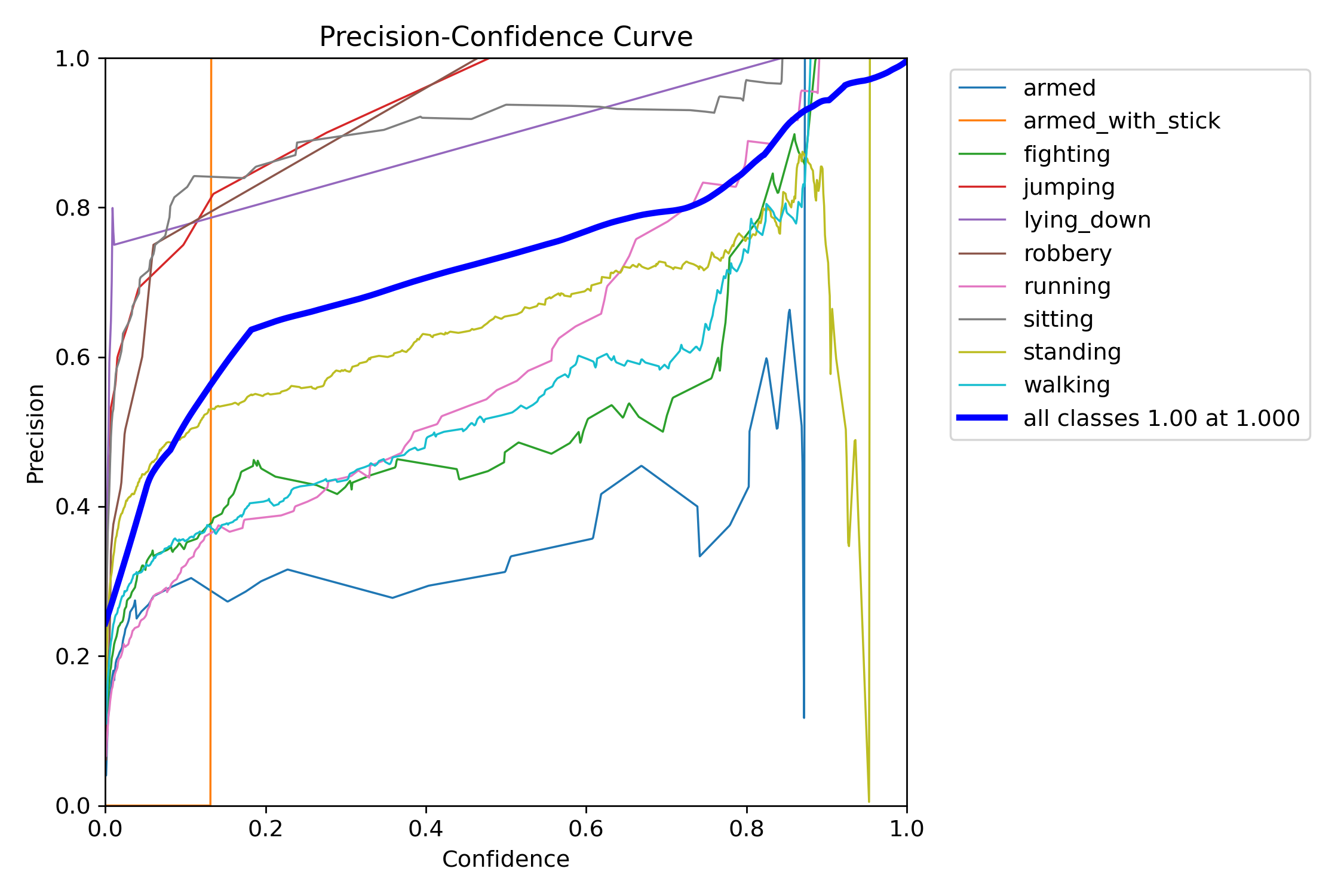
1. **Real-Time Performance:**  
   The combination of the YOLOv8s model with OpenCV-based tracking allows the system to detect and analyze human activities instantly. This makes it highly effective for live surveillance scenarios where quick response is essential, such as detecting fights, robberies, or other suspicious behavior.
2. **Interpretability and Transparency:**  
   Each detection is visually marked with bounding boxes and action labels, while all abnormal activities are logged with timestamps. This clear and interpretable output builds trust with end-users and facilitates accurate post-event analysis.
3. **Robustness Across Environments:**  
   The methodology is designed to handle real-world challenges such as variable lighting, complex backgrounds, and dynamic crowd densities. Background subtraction, noise reduction, and contextual analysis help maintain consistent performance even in difficult conditions.
4. **Scalability for Large Deployments:**  
   The system supports multiple video streams and modular integration, making it suitable for both small installations (e.g., offices or clinics) and larger infrastructures like airports, malls, or city-wide surveillance networks.
5. **Resource-Efficient Implementation:**  
   By using a lightweight version of YOLOv8 and optimizing processes with OpenCV, the system operates efficiently on standard hardware and even edge devices. This minimizes computational load and makes the solution cost-effective for a wide range of applications.

20

**Result**

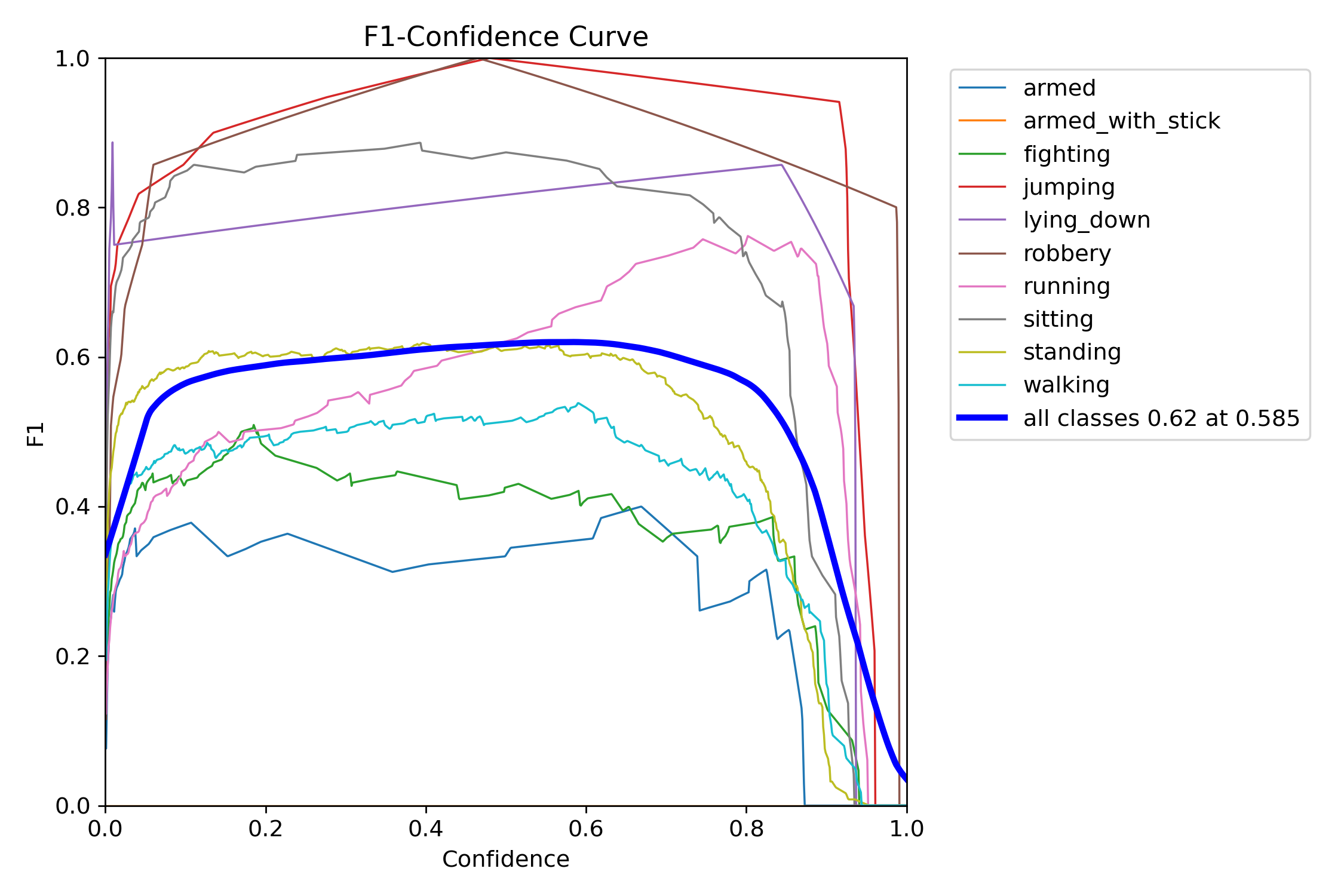
The YOLOv8s-based abnormal activity detection system was evaluated on a custom dataset comprising various human actions labeled via Roboflow. The system's performance was assessed using precision, recall, F1-score, and confusion matrix data, with visual output provided through confidence-based curves.

* **Model Accuracy**  
  The system achieved a detection accuracy of approximately 72.6%, correctly identifying both normal and abnormal human actions across multiple classes.
* **Precision-Recall-F1 Performance:**  
  The model demonstrated strong overall performance with a maximum precision of 1.00, recall of 0.77, and an F1-score of 0.62 at a confidence threshold of 0.585. Classes such as robbery, jumping, and sitting showed particularly high scores across all metrics, while slightly lower recall was observed in more visually complex classes like armed and fighting, likely due to overlap and class imbalance.



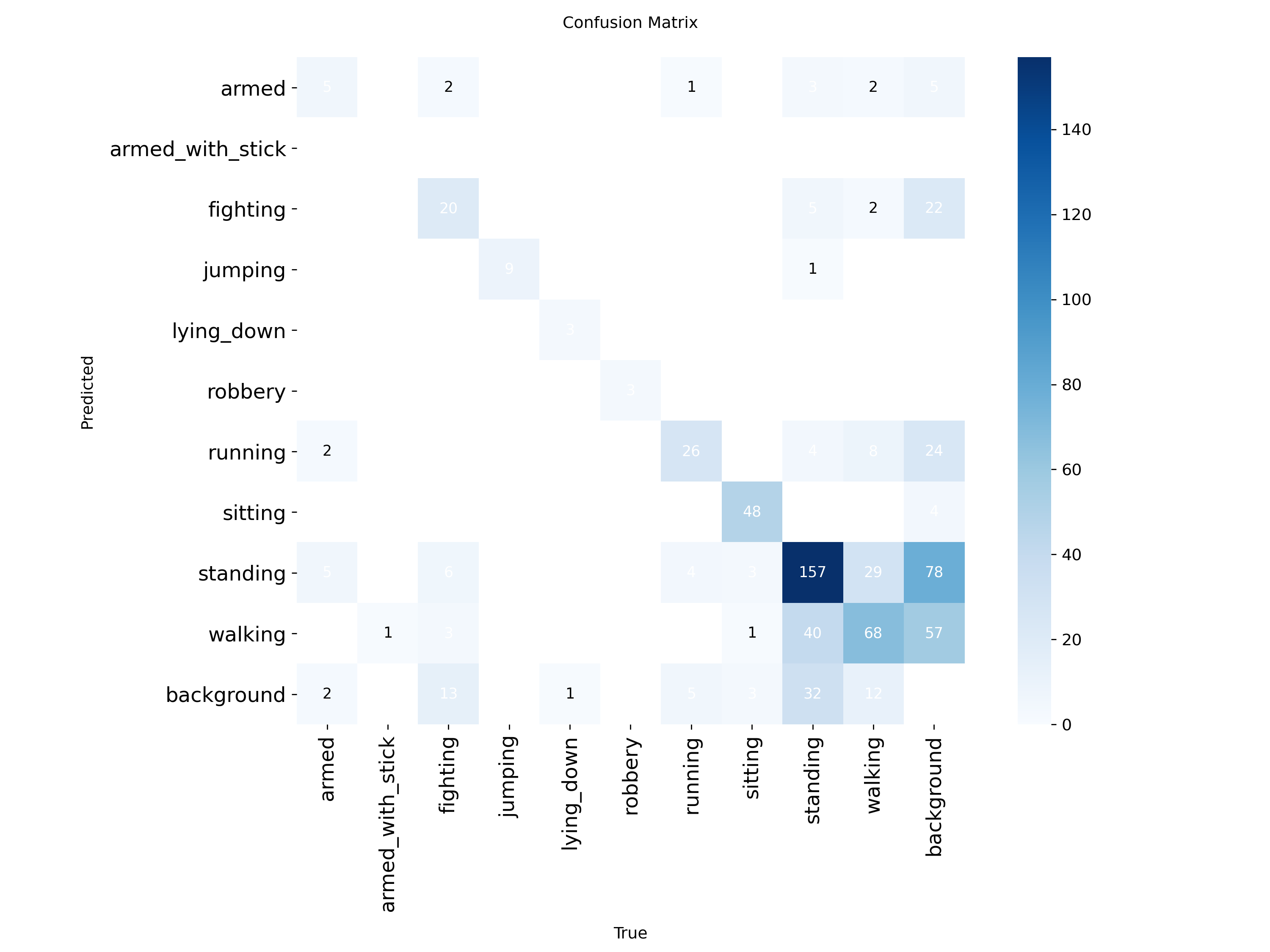
**Figure : 10(a) Figure : 10(b)**

21



**Figure : 10(c)**

* **Confusion Matrix Summary:**  
  The confusion matrix illustrates that classes like standing, walking, and sitting were detected with high accuracy (e.g., 157 true positives for standing), while some misclassifications occurred among running, armed, and fighting, which often overlap in appearance and context.

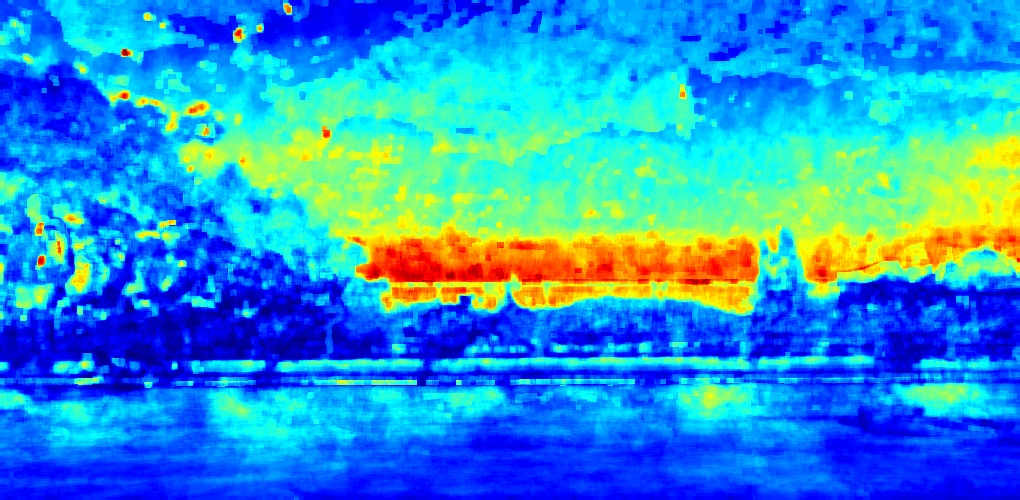


**Figure : 11**

* **Real-Time Performance**  
  The detection system consistently ran at 30–40 frames per second (FPS) on GPU, confirming real-time compatibility for surveillance applications.

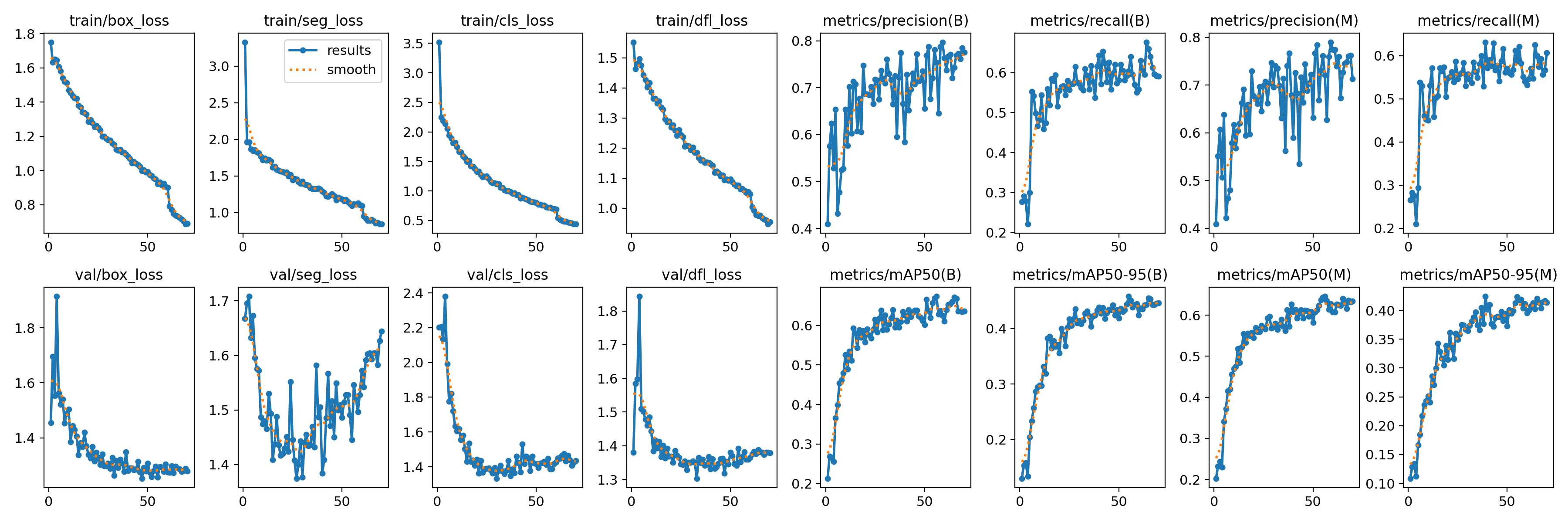
22

* **Visual Output and Alerts**  
  Bounding boxes with class labels were successfully rendered on video frames. Detected abnormal actions triggered real-time audio alerts and were logged with timestamps and labels. Annotated videos and saved abnormal frames were generated for later review.
* **Heatmap and Movement Zones**  
  The system produced motion heatmaps that visually highlighted areas of frequent or abnormal activity, assisting in zone-based surveillance and activity clustering.



**Figure : 12**

The model successfully detected actions like running and sudden stopping, with bounding boxes and labels for abnormal actions, offering reliable performance for surveillance systems.



**Figure : 13**

23

**Future Scope for Improvement**

* **Prospects for Enhancement :**  
  The anomaly detection system created with YOLOv8 provides a solid basis for real-time monitoring and can be improved with extra features and cross-domain compatibility.
* **Sophisticated Crowd Behavior Examination :**  
  Future enhancements may include clustering algorithms and density-based methods to oversee mass movements, recognize crowd formations, or pinpoint potentially aggressive group actions like riots or flash mobs—crucial in public areas during occasions or demonstrations.
* **Improved Immediate Notifications and Spatial Consciousness:**  
  Incorporating real-time heatmaps and dynamic crowd density analysis would allow the system to automatically identify areas of concern. When paired with localized alerts that include timestamps, this would facilitate quick incident identification and reaction.
* **Smart Behavior Observation:**  
  The system can be customized for specific environments such as retail shops or industrial locations to identify activities like theft, misuse of tools or machines, and other questionable or dangerous behaviors, thus enhancing safety and minimizing operational risks.
* **Implementation in Various Areas:**  
  In addition to surveillance, the approach might be applied to fields such as intelligent transportation (e.g., spotting jaywalking or traffic incidents), educational security (e.g., recognizing bullying or loitering), and wildlife observation for tracking poaching or unusual animal activity in conservation areas.
* **Automated Alerts and System Connectivity:**  
  To enhance real-time responsiveness, the system might be connected with mobile and web platforms to deliver instant alerts via SMS, email, or messaging applications. This would guarantee that essential anomalies are relayed promptly to security staff or system administrators.   
    
  These forthcoming improvements would enable the system to develop into a robust and flexible solution, aiding not just in public safety but also in workplace adherence, asset safeguarding, and environmental preservation

24

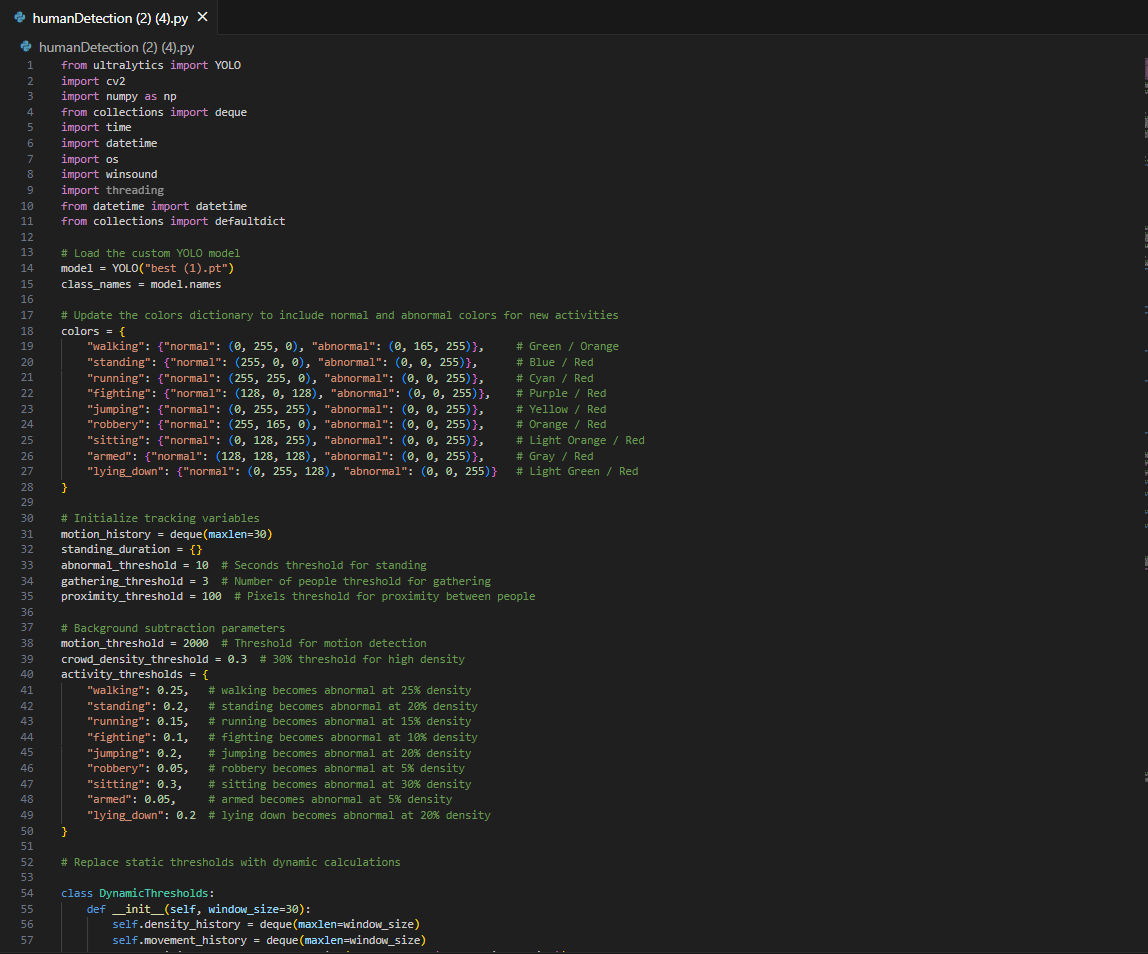
**Conclusion**

This venture presents an successful approach to anomalous movement discovery by joining the capabilities of YOLOv8 with real-time movement following and relevant investigation. The framework is planned to be both effective and interpretable, requiring negligible explained information whereas keeping up tall location exactness. By leveraging lightweight profound learning models and commonsense computer vision methods, the arrangement addresses real-world challenges in observation, particularly in high-density or energetic environments.

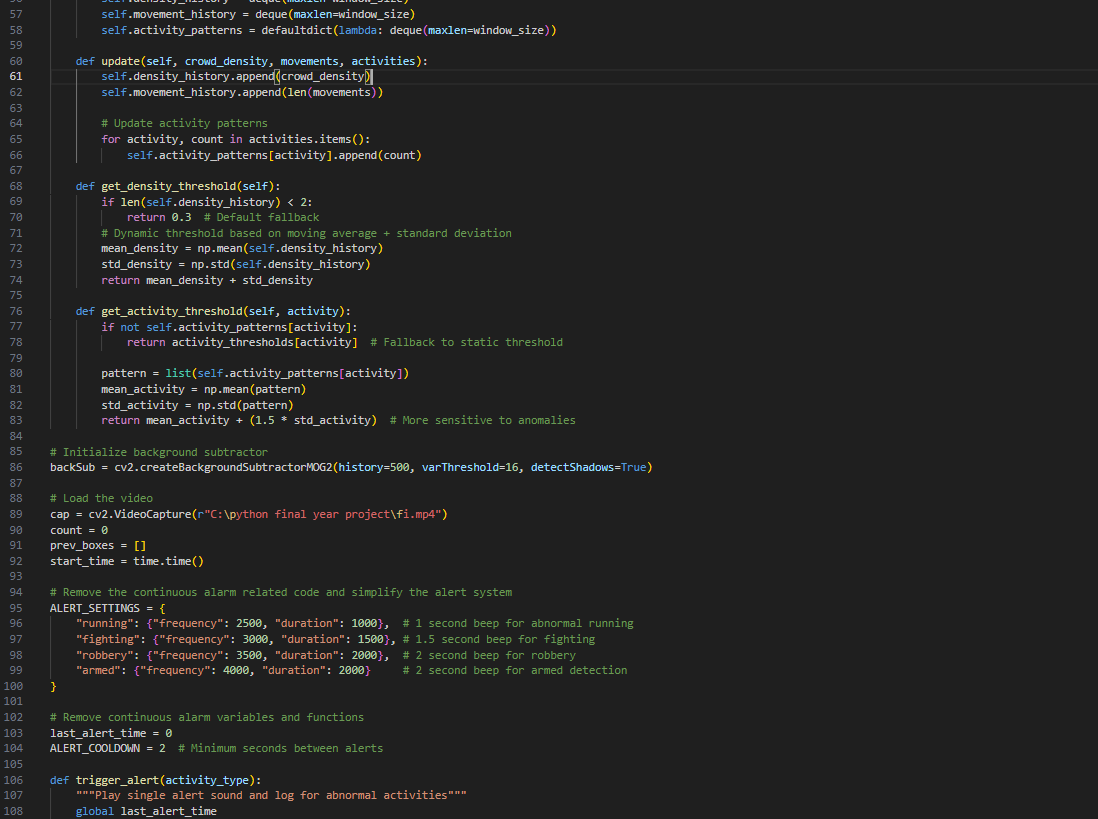
Through the combination of question discovery and behavioral examination, the framework gives solid, real-time recognizable proof of suspicious exercises, such as sudden developments, gather clustering, or potential dangers. This not as it were upgrades situational mindfulness but moreover makes a difference diminish reliance on manual observing. Once completely executed, the framework can altogether move forward reaction times and decision-making in security-sensitive applications such as open security, healthcare, and mechanical checking.

25

**Supplementary Materials (source code)**



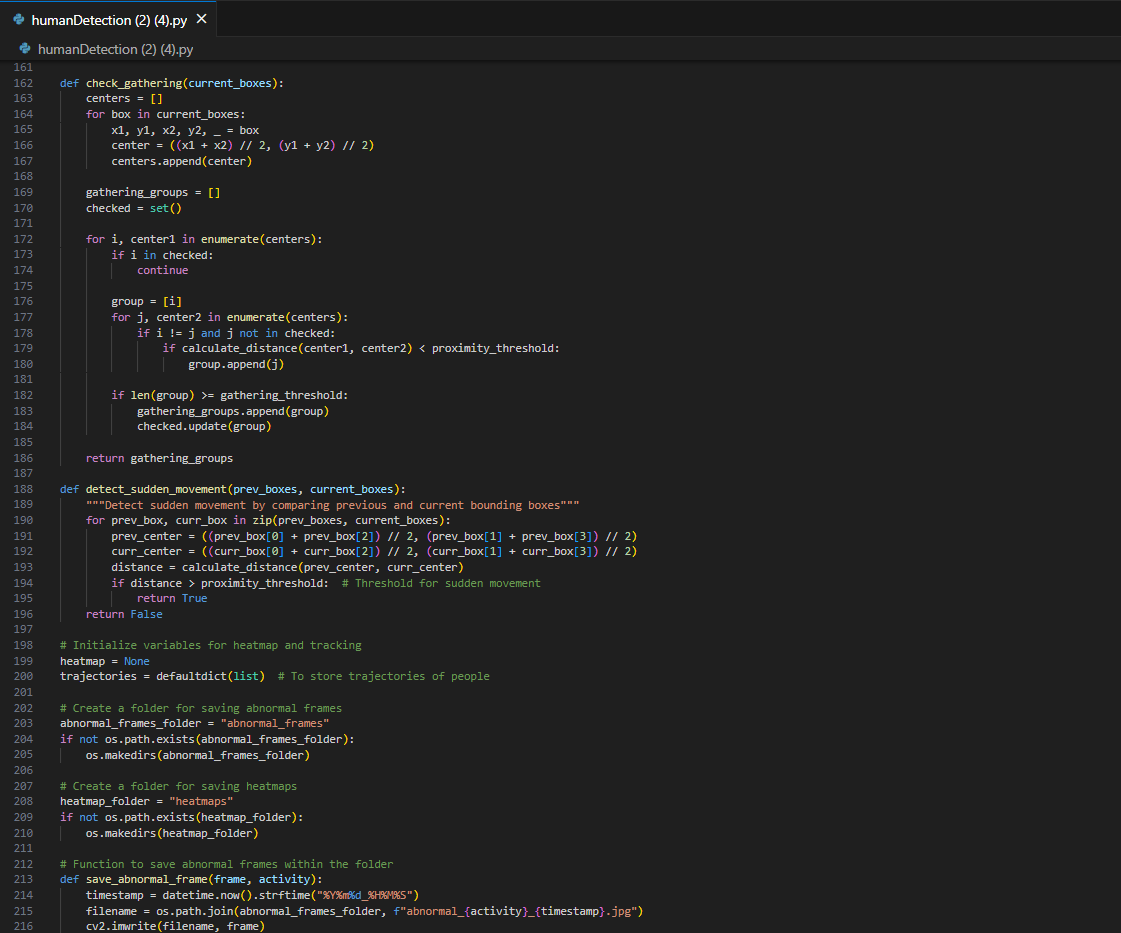
26



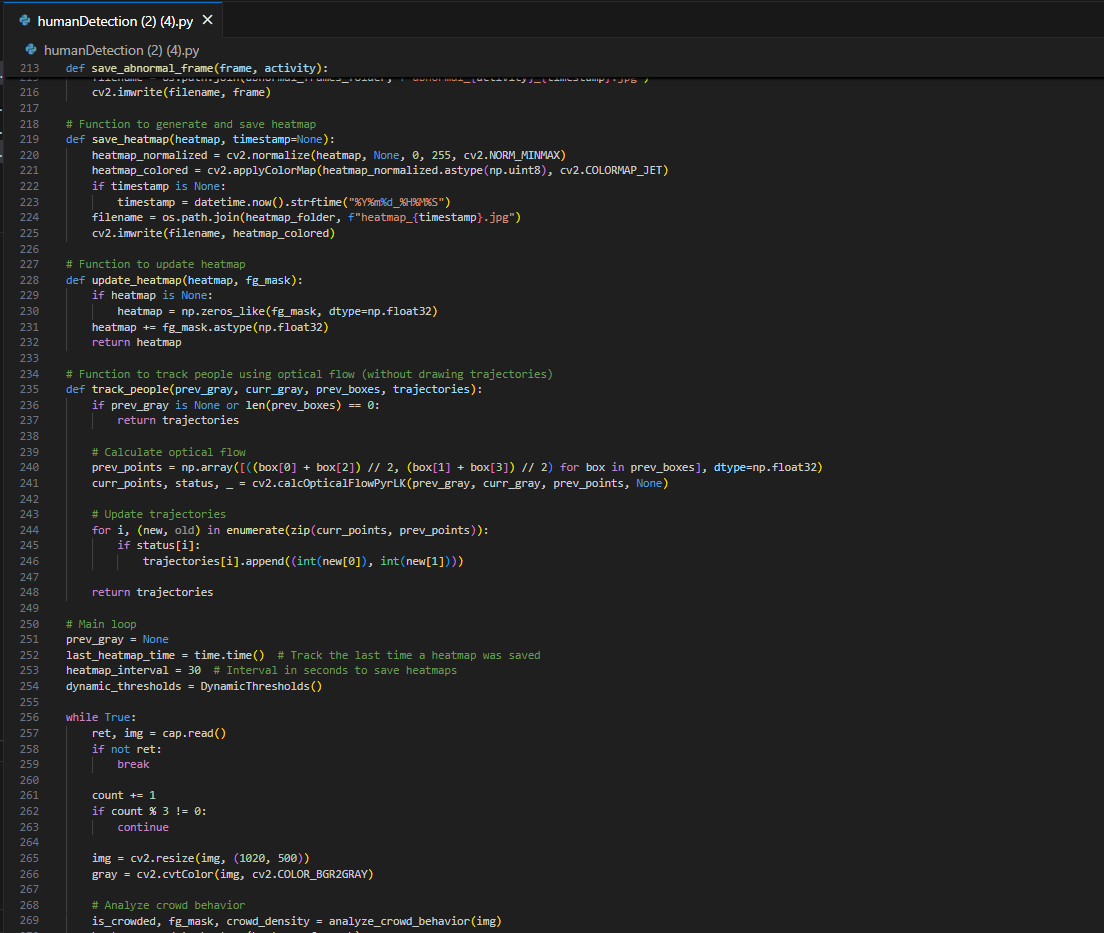
27



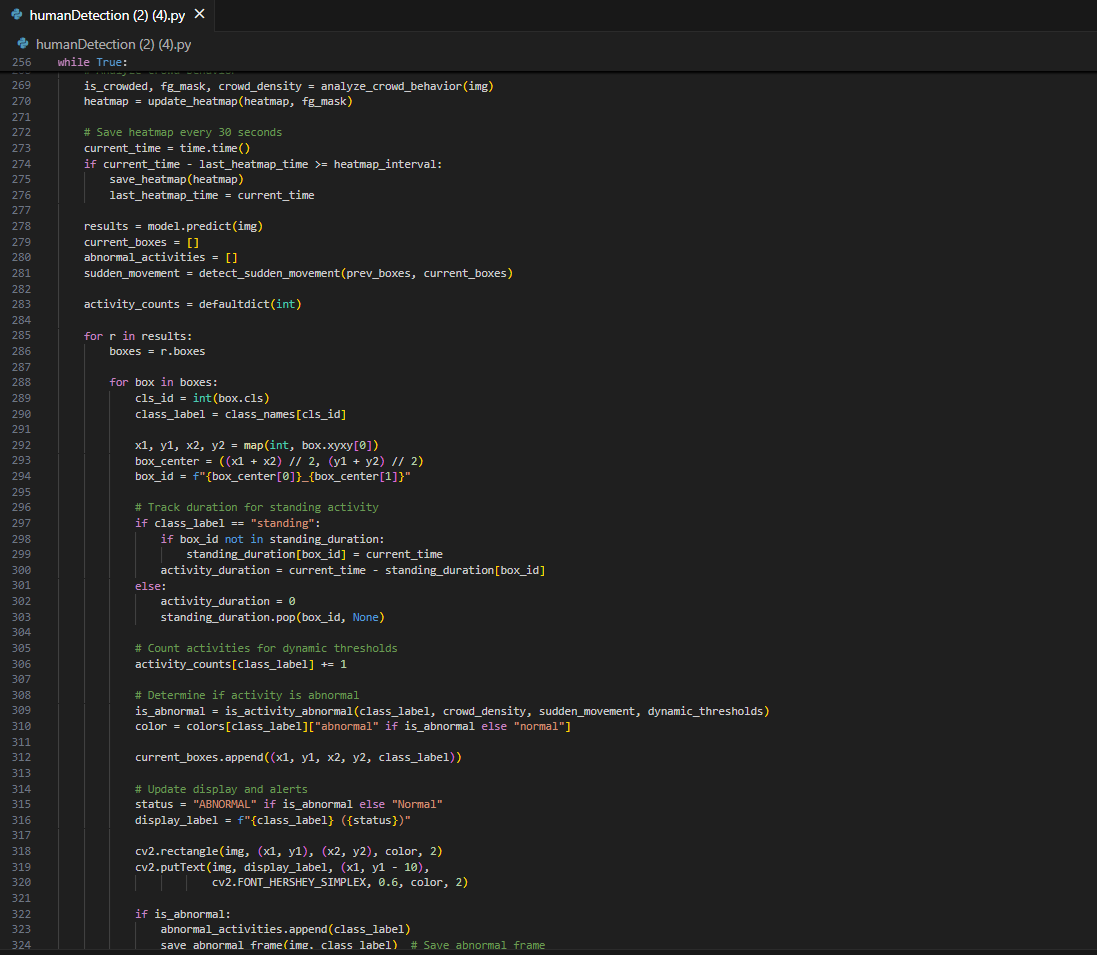
28



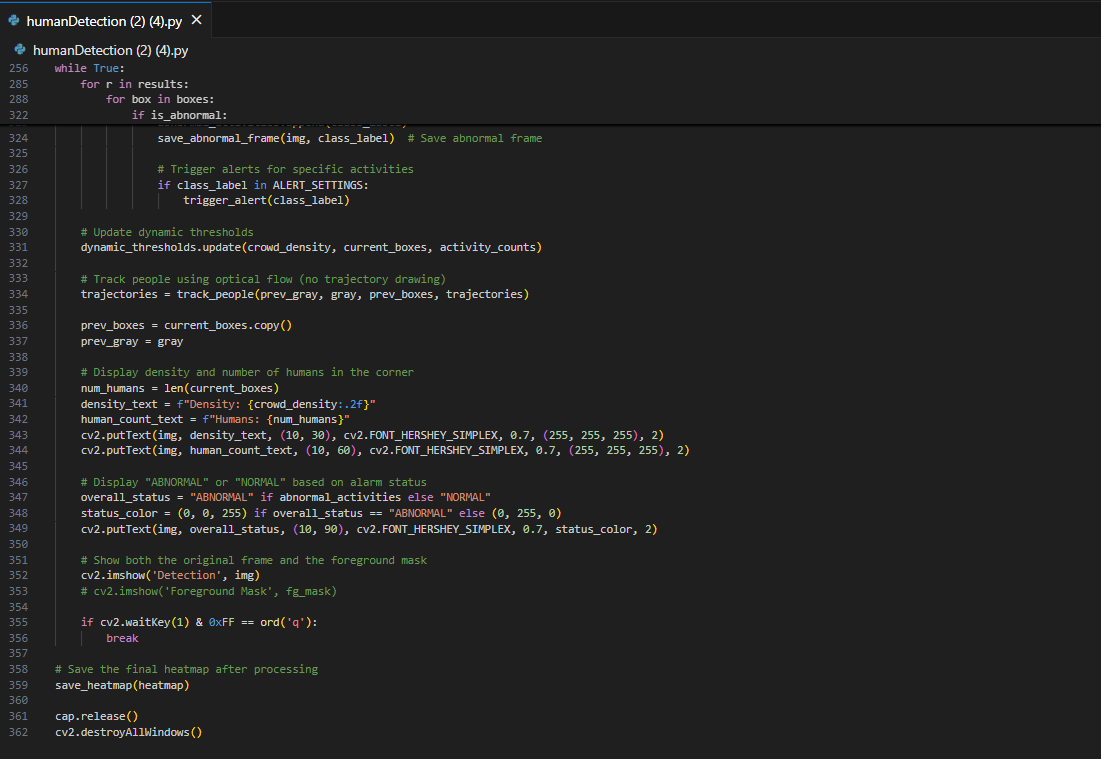
29



30



31



32

**References**

[1] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 886-893.

[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.

[3] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks.

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1725-1732.

[4] Ng, B., Zhang, Z., & Zhu, Y. (2015). Temporal sequence modeling with LSTM networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 37(4), 707-720.

[5] Liu, X., Zhang, S., & Zhang, M. (2023). Detecting abnormal behaviors using YOLOv8 and unsupervised clustering. IEEE Transactions on Circuits and Systems for Video Technology, 33(5), 1223-1231.

[6] Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17(1-3), 185-203.

[7] Lin, W., Xie, L., & Wu, Y. (2020). YOLO with optical flow for anomaly detection in dynamic environments. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8740-8749.

[8] Chen, Y., Zhou, Z., & Xu, L. (2022). Multi-scale feature extraction with YOLO for anomaly detection in crowded environments. IEEE Transactions on Image Processing, 31, 2354-2367.

[9] Khan, A., Rehman, A., & Ali, R. (2023). YOLOv8 for real-time anomaly detection on edge devices. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1761-1770.

[10] Sultani, M. (2018). Learning video-level anomaly detection with weak supervision. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7537-7546.

33