

AI-Based Abnormal Behavior Detection in Public Gatherings

Peddeni Mokshith¹, Pulla Chaitanya², Moriboyina Deekshith Kumar³, Dara Manish Kumar⁴, Ms. A. Jency⁵

^{1,2,3,4}*R.M.K College of Engineering and Technology, B. Tech Artificial Intelligence and Data Science*

⁵*Assistant Professor, R.M.K College of Engineering and Technology, B. Tech Artificial Intelligence and Data Science*

Abstract- Public environments such as railway stations, sports arenas, protest sites, and religious gatherings frequently attract large crowds where unusual behavior can rapidly develop into dangerous situations such as panic movement, physical conflicts, or stampedes. Monitoring these environments using traditional CCTV systems is challenging because security personnel must observe numerous video feeds simultaneously, which often leads to delayed reactions, operator fatigue, and reduced situational awareness. To address these limitations, this study proposes a real-time artificial intelligence framework for detecting anomalous activities in crowded public spaces through pose-based computer vision analysis. The system employs YOLOv8m-Pose to estimate human body keypoints and BoTSORT to maintain consistent multi-object tracking and identity association across consecutive frames. Each detected individual is modelled using the 17-keypoint representation defined in the COCO dataset, allowing the system to analyze body posture, movement patterns, and interactions between individuals in the scene. The proposed approach introduces a three-layer anomaly detection strategy. The first layer focuses on temporal behavioral patterns through tracked detectors, the second layer identifies immediate anomalies from individual frames using instant detectors, and the third layer analyzes collective crowd behavior through aggregate detectors. Using these mechanisms, the system can recognize ten types of abnormal events, including falling, running, loitering, stumbling, sudden stopping, crouching, physical fighting, stampede-like motion, ignored fallen persons, and abnormal crowd density. To enhance reliability, the framework incorporates a multi-signal quorum voting mechanism along with hysteresis-based state stabilization to reduce noisy predictions and maintain consistent anomaly labelling across frames. Experimental evaluation demonstrates that the proposed system can effectively perform real-time anomaly detection in crowded environments, indicating that pose-

driven behavioral analysis is a promising approach for intelligent video surveillance applications.

Keywords- *Computer vision, crowd anomaly detection, intelligent video surveillance, multi-object tracking, pose-based behavior analysis.*

I. INTRODUCTION

Large public gatherings present significant safety management challenges because the behavior of dense crowds can change rapidly and unpredictably. Places such as railway terminals, sports venues, festivals, protest locations, and religious events often attract thousands of people within confined spaces. Under such conditions, abnormal situations may emerge suddenly, including panic-driven running, physical confrontations, stampede-like movement, or individuals collapsing within the crowd. When these incidents are not identified quickly, they can escalate into dangerous situations that threaten public safety. Most existing surveillance infrastructures depend on closed-circuit television (CCTV) systems that are continuously observed by human operators. In practice, security personnel are required to monitor numerous video feeds at the same time and identify suspicious activities manually. Maintaining constant attention across multiple screens for extended periods is extremely demanding and often leads to fatigue, slower reaction times, and missed warning signs. Consequently, potentially hazardous situations may remain unnoticed until they develop into serious incidents. These challenges emphasize the need for automated surveillance solutions that can analyze crowd behavior and detect unusual activities without relying solely on manual monitoring.

Recent progress in artificial intelligence and computer vision has made it possible to analyze human activities in video streams automatically. Modern vision systems can extract visual cues related to posture, movement, and interaction patterns, allowing machines to interpret behavioral dynamics within crowded scenes. Leveraging these advancements, this research proposes a real-time framework for detecting anomalies in public gathering environments. The proposed approach performs pose-based behavior analysis by extracting seventeen human body keypoints using the YOLOv8m-Pose model and maintaining consistent identity tracking across frames through the BoTSORT multi-object tracking algorithm. By integrating pose estimation with temporal motion analysis and multi-signal reasoning, the system is able to identify abnormal patterns at both the individual and collective crowd levels. The main goal of this study is to design an efficient anomaly detection pipeline capable of producing reliable and interpretable alerts for surveillance personnel. Such a system can support faster situational awareness and improve safety monitoring in environments where large numbers of people gather.

II. RELATED WORK

Progress in computer vision and deep learning has greatly expanded the ability of surveillance systems to analyze human activities automatically and identify abnormal behavior in video data. A large portion of existing research focuses on learning patterns of normal activity from surveillance footage and detecting deviations that may indicate unusual or unsafe events. One notable human-centered approach is the TSGAD (Temporal Skeleton Graph Anomaly Detection) framework introduced by Mehdi Noghre and colleagues. This method represents individuals using skeletal pose structures and models the relationships between body joints through Graph Convolutional Networks combined with Variational Autoencoders. By learning spatial dependencies among keypoints, the system attempts to reconstruct normal motion patterns and treats large reconstruction errors as indicators of anomalous behavior. While this technique is effective for capturing detailed motion dynamics, its primary output is an anomaly score rather than explicit categorization of abnormal activities. In addition, the framework does not emphasize persistent tracking of

individuals across frames, which limits its ability to examine continuous behavioral patterns of specific people within crowded scenes.

Another relevant approach is the Motion Pattern Regularity Learning (MoPRL) framework, which applies spatial-temporal transformer architectures to model motion patterns in skeletal video sequences. In this method, the system learns the distribution of typical human motion and flags anomalies when observed movements diverge from the learned pattern. Transformer-based models provide strong capabilities for capturing temporal dependencies in sequential data. However, many of these systems rely on reconstruction-based anomaly detection and are often designed for offline video analysis. As a result, they are less suited for real-time surveillance scenarios that require immediate classification of different anomaly types or reasoning about the behavior of individual persons. Object detection techniques based on convolutional neural networks have also been widely used in surveillance applications. Architectures such as YOLO (You Only Look Once) enable rapid and accurate detection of people in video frames, making them well suited for real-time environments. Despite their efficiency, these approaches primarily focus on locating objects rather than interpreting behavioral patterns. Without pose-level information and temporal motion analysis, such systems struggle to recognize complex human activities such as falling, fighting, or panic-driven movement within dense crowds.

Several studies have additionally explored weakly supervised anomaly detection methods that combine temporal feature extraction with models such as Long Short-Term Memory networks or transformer-based encoders. These methods learn representations of normal video sequences and attempt to identify irregular events through deviations in temporal embeddings. Although they can detect abnormal segments in videos, many of these approaches produce only binary anomaly decisions rather than distinguishing between specific categories of abnormal behavior. Furthermore, persistent identity tracking is often absent, making it difficult to associate behavioral patterns with individual persons across consecutive frames.

Overall, the review of existing research highlights a

key gap in current surveillance methodologies. Many systems focus either on anomaly scoring or on basic object detection, but few integrate multiple essential components required for practical deployment. In particular, real-world surveillance environments demand solutions that combine real-time processing, pose-based behavior analysis, persistent multi-object tracking, and the ability to classify different types of anomalies. Developing a unified framework that brings these capabilities together remains an important step toward more intelligent monitoring of dense public gatherings.

III. RESEARCH GAP

An examination of existing crowd anomaly detection methods highlights several challenges that limit their effectiveness when applied to real-world surveillance environments. Many pose-driven anomaly detection systems are designed to produce a single anomaly score indicating whether observed behavior deviates from normal patterns. While this strategy can identify unusual motion, it often does not determine the exact nature of the abnormal activity. In operational surveillance settings, however, distinguishing between events such as a fall, physical conflict, panic running, or stampede-like movement is essential for enabling appropriate security responses. The absence of clear anomaly categorization therefore reduces the practical value of many current solutions. Another limitation frequently observed in existing approaches is the lack of persistent multi-object tracking. Although modern object detection algorithms can locate individuals in individual video frames with high accuracy, they do not always maintain consistent identities for those individuals over time. Without stable identity tracking, analyzing behavioral patterns across multiple frames becomes difficult, and abnormal activities cannot be reliably linked to specific individuals in the crowd. This restricts the ability of the system to perform meaningful temporal behavior analysis, which is necessary for identifying events such as loitering, sudden stops, or other prolonged irregular actions.

In addition, many previously proposed systems treat detection, motion analysis, and anomaly reasoning as separate stages rather than combining them within a unified processing pipeline. When these components operate independently, information extracted from

pose estimation and motion dynamics is not fully utilized during behavioral interpretation. Such fragmented designs reduce the system’s ability to capture complex interactions between individuals or recognize collective crowd behaviors that develop over time. A further challenge in real-time anomaly detection is the instability of frame-level predictions. Systems that classify anomalies independently for each frame often produce inconsistent outputs, where labels rapidly alternate between normal and abnormal states in consecutive frames. This phenomenon, commonly referred to as label flickering, can significantly reduce the reliability of surveillance alerts and make automated monitoring systems difficult to trust in practical deployments. Improving temporal consistency is therefore a key requirement for dependable anomaly detection.

These limitations indicate the need for an integrated framework that combines real-time pose estimation, persistent identity tracking, temporal behavioral analysis, and robust multi-category anomaly classification within a single architecture. The system proposed in this work addresses these challenges by designing a unified pipeline capable of identifying and categorizing abnormal behaviors in densely populated public environments with greater reliability.

IV. PROPOSED METHODOLOGY

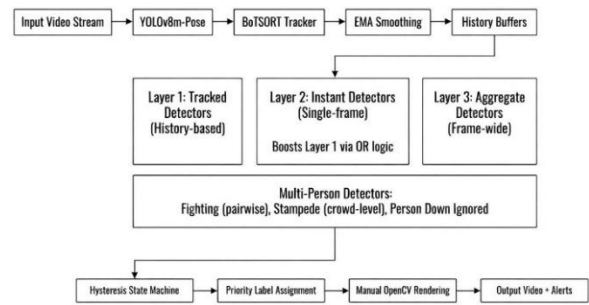


Fig. 1. Architecture of the proposed pose-based crowd anomaly detection framework

The architecture of the proposed crowd anomaly detection system is presented in Fig. 1. The framework operates as a multi-stage processing pipeline designed to analyze surveillance video streams and identify abnormal behavior in crowded environments in real time. Video data is first obtained from surveillance cameras installed in locations such as railway stations, stadiums, or other densely populated public areas.

These video frames serve as the input for the anomaly detection pipeline. Each frame is initially processed using the YOLOv8m-Pose model, which performs both human detection and pose estimation in a single stage. For every detected individual, the model generates a bounding box along with a skeletal representation consisting of 17 keypoints defined by the COCO human pose format. These keypoints represent important body joints such as shoulders, elbows, hips, and knees, allowing the system to capture detailed information about posture and body movement.

After detection, the identified individuals are passed to the BoTSORT multi-object tracking module, which assigns a unique identity to each person and maintains that identity across consecutive frames. The tracker combines motion prediction with appearance-based re-identification features, enabling stable tracking even in crowded scenes where temporary occlusions or overlaps occur between people. To improve the reliability of the pose data, a smoothing step is applied to the detected keypoints. An Exponential Moving Average (EMA) technique is used to reduce small fluctuations in the keypoint coordinates that may arise from pose estimation noise. This process stabilizes the skeletal representation and minimizes jitter that could otherwise lead to inaccurate behavioral analysis.

The refined pose information is then stored in temporal history buffers associated with each tracked individual. These buffers maintain records of keypoint coordinates, bounding box positions, centroid trajectories, and estimated motion velocities over multiple frames. By preserving temporal information, the system can analyze movement patterns and behavioral changes over time rather than relying solely on isolated frame observations. The buffered data is subsequently processed by a three-layer anomaly detection framework composed of tracked detectors, instant detectors, and aggregate detectors. Tracked detectors examine the temporal history of each individual to identify abnormal motion patterns that develop over time. Instant detectors operate on single-frame pose cues and complement the temporal analysis by detecting sudden abnormal postures or movements. Aggregate detectors analyze scene-level statistics to identify crowd-wide events, including abnormal crowd density or coordinated panic movement.

Beyond individual behavior analysis, the system also incorporates multi-person interaction detectors that evaluate spatial relationships between nearby individuals. These detectors are responsible for recognizing complex crowd events such as physical fights, stampede-like movement patterns, and scenarios where a person has fallen but remains unattended by surrounding individuals. To improve the stability of detection results, outputs from all anomaly detectors are processed using a hysteresis-based state machine. This mechanism prevents rapid switching between normal and abnormal states across consecutive frames, thereby reducing label flickering. When multiple anomaly types are detected simultaneously, a priority-based label assignment module determines the most relevant alert to display.

Finally, the processed results are visualized using an OpenCV-based rendering module. This module overlays skeletal structures, bounding boxes, anomaly labels, and alert notifications directly onto the video frames. The resulting annotated video stream provides surveillance operators with clear visual feedback and real-time alerts regarding abnormal activities occurring within the monitored environment.

V. SYSTEM ARCHITECTURE AND MODULES

Table 1 Modules of the Proposed Crowd Anomaly Detection System

Module	Description
M1	YOLOv8m-Pose model for human detection and extraction of 17 body keypoints
M2	BoTSORT algorithm for maintaining consistent identities across frames
M3	Exponential Moving Average (EMA) for smoothing keypoint coordinates
M4	Temporal buffers storing pose data, centroids, and motion-related features
M5	Tracked detectors for analyzing behavior over time
M6	Instant detectors for identifying anomalies from single-frame observations
M7	Aggregate detectors for evaluating crowd-level patterns
M8	Interaction detectors for events such as fighting, stampede, and unattended fallen individuals
M9	Hysteresis-based state module for stabilizing detection outputs
M10	OpenCV-based module for visualization and real-time rendering

The proposed crowd anomaly detection framework is organized as a set of interconnected processing modules that together handle detection, tracking, behavioral interpretation, and result visualization. Each module performs a clearly defined function within the pipeline, enabling the system to process surveillance video streams in a structured and efficient manner. This modular organization also supports flexibility, allowing individual components to be improved or replaced without affecting the entire system. The main modules and their functionalities are summarized in Table I.

The system is designed to operate as a layered pipeline in which each group of modules contributes to a specific stage of processing. The initial stage focuses on perception, where human detection and pose estimation are performed, followed by multi-object tracking to ensure identity consistency across frames. These modules convert raw video input into structured representations of individuals, including skeletal configurations and motion-related attributes. The next stage is dedicated to behavioral analysis. Here, temporal and spatial features derived from pose trajectories are examined to identify abnormal patterns. By combining both short-term observations and longer temporal sequences, the system can capture subtle behavioral changes as well as sudden anomalies.

The final stage is responsible for decision stabilization and output generation. Mechanisms such as hysteresis-based state control are used to ensure consistent anomaly labelling over time, reducing fluctuations in predictions. The processed results are then rendered visually using an optimized display module that overlays detection outputs onto the video stream. This modular and layered design enables the framework to handle complex crowd dynamics efficiently while maintaining real-time performance. At the same time, it provides a clear separation of responsibilities between components, which is important for scalability and future enhancements.

VI. ANOMALY DETECTION FRAMEWORK

The anomaly detection component of the proposed system is designed to interpret both individual behavior and collective crowd dynamics using pose-driven motion analysis. Once pose estimation, tracking, and

temporal data storage are completed, the system evaluates behavioral patterns through multiple detection mechanisms operating at different temporal and spatial scales. These mechanisms process skeletal motion, spatial relationships, and crowd-level statistics to identify abnormal activities within the scene. The overall design follows a layered approach, where each detector contributes to a specific aspect of behavioral understanding.

The first layer consists of tracked detectors, which perform temporal analysis of each individual using historical pose information. These detectors utilize data stored in temporal buffers, including keypoint trajectories, centroid displacement, bounding box variations, and velocity patterns. By examining how motion evolves over time, the system can identify behaviors such as falling, running, loitering, stumbling, sudden stopping, and crouching. Instead of relying on a single indicator, multiple features derived from pose structure and motion dynamics are combined using a voting mechanism. This multi-signal strategy ensures that anomalies are confirmed only when consistent evidence is observed, thereby reducing false positives caused by noise or short-term pose instability.

Complementing the temporal layer, instant detectors operate on individual frames without requiring historical context. These detectors focus on identifying abnormal postures or configurations that can be recognized immediately, such as sudden collapses or extreme body orientations. Since certain events occur abruptly and may not exhibit gradual temporal patterns, instant detectors provide rapid detection capability and strengthen the overall decision-making process when strong visual cues are present.

To address large-scale behavioral patterns, the framework incorporates aggregate detectors that analyze crowd-level statistics within each frame. Rather than focusing on individual actions, these detectors evaluate the overall distribution of movement across all detected persons. Situations such as sudden spikes in running activity, unusually high crowd density, or coordinated panic-like motion are identified by measuring the proportion of individuals exhibiting abnormal behavior. This enables the system to detect events that emerge from collective dynamics rather than isolated actions.

In addition to individual and group analysis, the framework includes interaction-based detectors that examine relationships between nearby individuals. These detectors analyze proximity, relative motion direction, and pose dynamics to identify complex events involving multiple participants. Examples include physical altercations, stampede-like movement patterns, and scenarios where a fallen individual remains unattended. Such behaviors require relational reasoning and cannot be reliably detected through single-person analysis alone.

By integrating temporal, instantaneous, aggregate, and interaction-based detection strategies, the proposed framework provides a comprehensive mechanism for identifying a wide range of anomalies. This multi-layered design allows the system to capture both short-duration events and long-term behavioral trends, while maintaining robustness in the presence of noisy detections and rapidly changing crowd conditions.

VII. IMPLEMENTATION DETAILS

The proposed crowd anomaly detection system was developed in Python using a combination of deep learning and computer vision libraries. Human pose estimation was implemented using the Ultralytics version of the YOLOv8 framework, while OpenCV was used for video processing and visualization tasks. All experiments were carried out using the PyTorch framework with GPU acceleration enabled via CUDA. The system was executed in the Google Colab environment on an NVIDIA T4 GPU to support real-time inference.

For pose estimation, the pretrained YOLOv8m-Pose model was used to detect individuals and extract 17 human body keypoints corresponding to the COCO format. Input video frames were resized to a resolution of 640 pixels to maintain a balance between computational cost and detection accuracy. A confidence threshold of 0.35 was applied to remove low-confidence detections. The model was deployed on the GPU through the Ultralytics interface to enable efficient frame-by-frame processing.

To ensure continuity of individual identities across frames, multi-object tracking was performed using the BoTSORT algorithm. This tracker combines motion

prediction with appearance-based re-identification to associate detections over time. Each person is assigned a unique identifier that persists throughout the video sequence, allowing the system to maintain consistent identity information and analyze behavioral patterns over extended durations.

To address noise in pose estimation, an Exponential Moving Average (EMA) smoothing technique was applied to the detected keypoints. Minor variations in keypoint positions across consecutive frames can introduce instability in motion analysis. The smoothing process reduces this effect by blending current keypoint coordinates with values from previous frames. This operation is applied independently to each joint, provided that the detection confidence exceeds a predefined threshold.

Temporal information for each tracked individual is maintained using sliding window buffers. These buffers store recent observations, including keypoint coordinates, centroid positions, bounding box dimensions, and estimated velocities. Implemented using fixed-length deque structures, the buffers automatically discard older entries as new data becomes available. This design enables efficient temporal analysis of motion trajectories and posture variations without excessive memory usage.

The anomaly detection logic is implemented through a set of modular functions that evaluate pose geometry, motion characteristics, and spatial relationships. Each anomaly type is identified using multiple behavioral indicators derived from keypoint configurations and movement patterns. For instance, fall detection considers factors such as torso inclination, changes in bounding box proportions, and relative positioning of head and hip joints. By combining multiple signals, the system reduces sensitivity to noise and improves detection reliability.

Visualization is handled using a custom OpenCV-based rendering pipeline. Instead of relying on built-in visualization utilities, the system directly draws bounding boxes, skeletal structures, and anomaly labels using optimized OpenCV functions. This approach significantly improves rendering efficiency, reducing latency from approximately 100 milliseconds per frame to nearly 1 millisecond per frame. As a result,

the system is able to maintain real-time performance while providing clear and informative visual outputs.

Overall, the implementation follows a modular design in which detection, tracking, smoothing, anomaly analysis, and visualization operate as coordinated components within a single pipeline. The combination of GPU-accelerated inference and optimized rendering allows the system to process surveillance video streams efficiently while maintaining stable and reliable anomaly detection performance.

Table II Implementation Configuration

Parameter	Value
Programming Language	Python
Framework	PyTorch
Pose Model	YOLOv8m-Pose
Tracker	BoTSORT
Input Resolution	640 px
GPU	NVIDIA T4

VIII. EXPERIMENTAL EVALUATION AND RESULTS

The proposed crowd anomaly detection system was evaluated using surveillance-style video sequences that simulate real-world public gathering conditions. The dataset includes scenes from environments such as railway stations, pedestrian pathways, and other crowded public spaces where multiple individuals are present within the camera's field of view. The evaluation focuses on assessing the system's ability to accurately detect and classify abnormal behaviors while sustaining real-time processing performance.

Video frames were processed sequentially through the complete pipeline, which includes pose estimation, multi-object tracking, temporal feature extraction, and anomaly detection. Human pose estimation was performed using the YOLOv8m-Pose model, which generates bounding boxes and 17 skeletal keypoints for each detected individual. These detections were then linked across frames using the BoTSORT tracking algorithm to preserve identity consistency. The resulting pose and motion information was analyzed by the anomaly detection framework to identify deviations from normal behavior patterns.

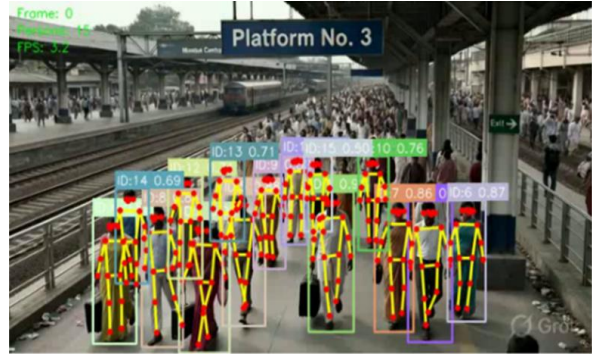


Fig. 2. Normal crowd monitoring output of the proposed system showing pose estimation, bounding boxes, and persistent tracking IDs without detected anomalies.

Fig. 2 presents the system output under normal crowd conditions. In this scenario, individuals exhibit regular movement without any abnormal activities. The system successfully detects and tracks multiple individuals simultaneously, assigning unique identifiers to each person. The rendered output includes skeletal structures and bounding boxes, demonstrating stable pose estimation and reliable tracking performance even in moderately dense environments.

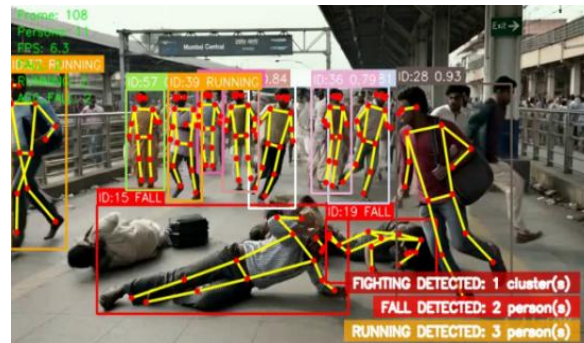


Fig. 3. Abnormal crowd behavior detection illustrating anomaly alerts such as running, falling, and irregular posture detected using the proposed framework.

Fig. 3 illustrates the system's behavior in the presence of abnormal activities. Events such as sudden running, falling, and irregular postures are detected and highlighted using labeled bounding boxes and visual alerts. These detections are generated based on the analysis of pose configurations, motion dynamics, and interactions between individuals. The results indicate

that the system can effectively identify anomalies at the individual level while continuing to track surrounding persons without disruption.

In terms of computational performance, the system achieves real-time processing by combining efficient pose estimation with optimized visualization. Experiments conducted in a Google Colab environment using an NVIDIA T4 GPU show that the system processes video at approximately 15 to 20 frames per second. Performance gains were primarily achieved by replacing default visualization utilities with a custom OpenCV-based rendering pipeline, which reduced rendering latency from around 100 milliseconds per frame to nearly 1 millisecond per frame.

Overall, the experimental results demonstrate that the proposed framework is capable of accurately detecting both normal and abnormal behaviors in crowded environments while maintaining real-time execution. The integration of pose-based analysis, persistent identity tracking, and multi-layer anomaly detection enables the system to generate clear and interpretable alerts, making it suitable for practical surveillance applications.

IX. DISCUSSION AND ANALYSIS

The experimental observations indicate that the proposed pose-based framework is effective in analyzing complex human behavior within densely populated environments. By integrating pose estimation, multi-object tracking, and layered anomaly detection, the system is able to capture both individual-level actions and collective crowd dynamics. Unlike conventional surveillance approaches that rely primarily on bounding boxes or motion intensity, the use of skeletal keypoints enables a more detailed understanding of posture and body orientation. This allows the system to identify subtle behavioral patterns such as falling, crouching, or irregular movement that may not be distinguishable through object-level features alone.

A key strength of the framework lies in its use of persistent multi-object tracking. By assigning consistent identities to individuals using the BoTSORT algorithm, the system maintains continuous behavioral

records across frames. This capability supports reliable temporal analysis and eliminates ambiguity when multiple individuals are present in the scene. It also ensures that detected anomalies can be associated with specific individuals, which is important for practical monitoring and response in real-world surveillance scenarios.

The use of multi-signal reasoning further improves detection robustness. Each anomaly is evaluated using a combination of features derived from pose structure, motion characteristics, and spatial relationships, rather than relying on a single indicator. This approach reduces sensitivity to noise and minimizes false detections caused by temporary pose variations or minor tracking inconsistencies. As a result, the system maintains stable performance even in challenging conditions such as occlusions, dense crowd formations, and rapid movement.

Another important contribution is the introduction of temporal stabilization through a hysteresis-based state mechanism. In many real-time systems, independent frame-level predictions can lead to inconsistent outputs, where anomaly labels fluctuate rapidly between normal and abnormal states. The proposed approach addresses this issue by enforcing consistency over time, allowing anomalies to be confirmed only when sufficient evidence is accumulated. This improves both the reliability and interpretability of the generated alerts.

Overall, the proposed architecture demonstrates strong potential for deployment in real-world surveillance environments. The combination of pose-based analysis, identity tracking, multi-signal detection, and temporal stabilization provides a comprehensive solution for monitoring complex crowd behavior. These capabilities make the system suitable for scenarios characterized by high density, frequent occlusions, and dynamic interactions, where accurate and timely detection of abnormal events is critical.

X. CONCLUSION AND FUTURE WORK

This work presented a real-time crowd anomaly detection framework for monitoring abnormal activities in densely populated public environments. The system combines pose-based human analysis,

persistent multi-object tracking, and a multi-layer anomaly detection strategy to interpret both individual behavior and overall crowd dynamics. By utilizing the YOLOv8m-Pose model for extracting skeletal keypoints and the BoTSORT algorithm for maintaining identity consistency, the framework is able to continuously track multiple individuals and analyze their motion patterns over time. The use of structured pose information allows the system to capture detailed behavioral cues, enabling more precise interpretation of human actions within complex scenes.

The anomaly detection process is designed as a layered mechanism that integrates tracked, instant, aggregate, and interaction-based detectors. This combination allows the system to evaluate both short-term and long-term behavioral patterns while accounting for spatial relationships between individuals. The use of multi-signal decision strategies improves detection reliability by reducing the impact of noise and transient variations. In addition, the incorporation of a hysteresis-based stabilization mechanism ensures consistent anomaly labeling across frames, addressing the common issue of prediction instability in real-time systems. Experimental results indicate that the framework is capable of identifying multiple categories of abnormal behavior at the individual level while maintaining real-time processing performance. The optimized visualization pipeline further supports practical deployment by enabling efficient and clear alert generation.

Despite these strengths, certain limitations remain. The system's performance can be affected by challenging visual conditions such as severe occlusions, poor lighting, or extreme camera angles. In addition, the current implementation relies on single-camera input, which restricts its ability to maintain identity continuity across larger surveillance areas.

Future work can address these limitations by extending the framework to multi-camera environments, enabling cross-camera tracking and improved coverage of large public spaces. Incorporating predictive models for early anomaly forecasting could further enhance the system's ability to prevent incidents before they escalate. Moreover, training and evaluating the system on larger and more diverse real-world datasets would improve its robustness and generalization capability.

These enhancements would contribute toward the development of more advanced and proactive surveillance systems for ensuring public safety in complex environments.

REFERENCES

- [1] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 6479–6488.
- [2] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," ACM Comput. Surv., vol. 54, no. 2, pp. 1–38, 2019.
- [3] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011, pp. 3449–3456.
- [4] B. Zhou, S. Tang, and M. Shah, "Learning temporal regularity in video sequences," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 733–742.
- [5] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2009, pp. 935–942.
- [6] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed cameras," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 555–560, 2008.
- [7] T. Hirschorn, Y. Averbuch-Elor, and D. Cohen-Or, "STG-NF: Normalizing flows for human pose anomaly detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2023.
- [8] J. Liu, W. Luo, and S. Gao, "Future frame prediction for anomaly detection: A new baseline," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.
- [9] L. Sun, J. Liu, and X. Tang, "Graph convolutional networks for skeleton-based action recognition," in Proc. AAAI Conf. Artif. Intell., 2018.
- [10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.
- [11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [12] G. Jocher et al., "Ultralytics YOLOv8," GitHub

- repository, 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [13] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2017.
- [14] A. Bewley et al., "Simple online and realtime tracking," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2016.
- [15] Y. Aharon et al., "BoT-SORT: Robust associations multi-pedestrian tracking," arXiv preprint arXiv:2206.14651, 2022.
- [16] Z. Zhang, M. Shi, and X. Chen, "Crowd counting via scale-adaptive convolutional neural network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.
- [17] S. Georgescu, A. Barbalau, R. T. Ionescu, and M. Shah, "Anomaly detection in video: A survey," IEEE Trans. Pattern Anal. Mach. Intell., 2023.
- [18] Y. Liu, Y. Wang, and X. Zhang, "Pose-based abnormal behavior detection in crowded scenes using deep graph networks," IEEE Access, 2023.
- [19] H. Wang, Z. Chen, and Y. Li, "Skeleton-based human behavior analysis for video anomaly detection," Pattern Recognit. Lett., 2024.
- [20] X. Zhang, T. Chen, and L. Zhao, "Real-time crowd behavior analysis using pose estimation and multi-object tracking," IEEE Access, 2024.
- [21] A. Barbalau et al., "Object-centric video anomaly detection using multi-scale feature learning," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2024.
- [22] K. Kim and J. Lee, "Pose-guided abnormal activity detection in surveillance videos," IEEE Sensors J., 2024.