International Journal of Engineering Trends and Technology Volume 74 Issue 7, 493-521, July 2026
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I7P130

Original Article

Intelligent Crowd Analysis: A Learning-Based Approach for Detection and Behavior Analysis in Surveillance Videos

Ganta Raju[1*], GS Naveen Kumar[2]

[1]*Department of Information Technology, Malla Reddy University, Hyderabad, Telangana, India.*
[2]*Department of CSE (Data Science), Malla Reddy University, Hyderabad, India.*

**Corresponding Author : rishi.shinny@gmail.com*

Received: 21 August 2025 Revised: 09 April 2026 Accepted: 20 April 2026 Published: 28 July 2026

Abstract - *Crowd monitoring and analysis is a recent initiative that brings public safety as a result (especially if your business is in a crowded area) This work present a novel Intelligent Crowd Analysis (ICA) system in this paper that takes advantage of a Modified YOLOv4-tiny object detection model with Non-Maximum Suppression (NMS), Deep SORT object tracking and a Crowd Monitoring and Behavior Analysis (CMBA)module that is designed by us. The system enables the detection of real-time violations in social distancing, entry in restricted areas, and abnormal crowd behavior. In order to tackle these challenges, such as overlapping objects, dense crowds, and dynamic conditions, some domain-specific changes, such as setting anchor boxes manually and adding attention to the split detector networks, have been made. Experiments show that the system is effective and robust in different surveillance situations. When comparing the results of the Modified YOLOv4 with the state-of-the-art models like YOLOv3, Faster R-CNN, SSD, and FairMOT, the Modified YOLOv4 achieved the best precision (96.87%), recall (95.31%), F1-score (96.08%), and accuracy (97.46%) within no time while the real-time video/frame or image processing was performed through deep learning on individual vehicle counting providing better performance of results over these methods [153]. The system runs with an average of 25 FPS; thus, real-time capabilities are guaranteed. In addition to these totals, the performance achieved a 92% detection accuracy, identified situations of social distance violations, Restricted Area Violations, and abnormal motion with highly Reliant Identification. This proves that this intelligent surveillance system operates in very complex environments and in real-time, which makes it suitable for public safety, smart cities, and event management-related applications. The ICA system demonstrates improved performance in crowd analysis and monitoring.*

Keywords - *Crowd Detection, Behavior Analysis, Public Surveillance, Learning-Based Methodology, Intelligent Video Analysis.*

1. Introduction

Crowd behavior analysis from surveillance videos is a significant research challenge with a variety of applications in public safety, managing over-crowded big events, and minimizing risk in population-dense environments. The current rising smart cities with surveillance systems require an intelligent framework that can analyse the behaviour of the crowd, detect the anomalous behaviour, and act in a timely manner. Now, automated systems that can perform robustly in the face of the uncertainties and noise presented by the real world are being sought for applications such as monitoring compliance with social distancing legislation, expediting the protection of secure areas, and monitoring for cybersecurity anomalies. However, huge occlusions, people density, the phenomenon of varying motion models, and the versatility of the environment pose difficulties for conventional methodologies of crowd monitoring that render them impractical.

Over the past couple of years, a number of advances in machine learning and deep learning have opened up new avenues for crowd behavior analysis. CNNs, RNNs, and hybrid architectures have often been used in many tasks such as anomaly detection, motion prediction, and density estimation [1, 2]. As an example, Conv-LSTM-based models (known to be used for capturing spatiotemporal dependencies for anomaly detection [3, 4]) and CNN-LSTM hybrid framework have been shown to be effective in [5]. In a similar fashion, dual-stream architectures [6] and attention mechanisms [5] have been proposed to mitigate against challenges like occlusion and scale variation. Efficient real-time human detection has been made possible by object detection frameworks, especially based on YOLO [6], as well as domain adaptation methods that seek to generalize across environments better [7]. However, existing methods show some essential limitations. However, various deep learning models are able to obtain high accuracy rates, but they have high computational complexity, which makes them not applicable for real-time applications. Several other models focus on detection or tracking, but not on an integrated system for behavior analysis.

Furthermore, robustness in extremely dense and occluded environments and generalisation ability towards various surveillance scenarios and multi-camera situations are limited in existing systems [8]. Also, current literature lacks research on proactive capabilities, including the detection of an ongoing violation, how a violation appears, and anomaly explanation.

It identifies the key research gap: no single framework is capable of achieving accurate detection, robust tracking, and intelligent behavior analysis simultaneously in a real-time manner while being computationally efficient. The present-day alternatives prefer accuracy while compromising on speed or isolating individual tasks while minimizing the broader view of the entire crowd behavior. Thus, there is an increasing necessity for a seamless system that efficiently detects and tracks people, as well as understands their actions to provide real-time insights.

In order to fill this gap, this paper presents an Intelligent Crowd Analysis (ICA) system, which comprises a modified YOLOv4-tiny object detection model, Non-Maximum Suppression (NMS), Deep SORT-based tracking, and a newly introduced Crowd Monitoring and Behavior Analysis (CMBA) module.

Compared to the current methods, the proposed approach improves the performance on the detection, especially in dense and occluded environments via the design of the specific channel and spatial attention mechanisms, and the optimisation of the anchor boxes.

The incorporation of Deep SORT provides a robust multi-object tracking capability that couples motion prediction with appearance-level feature embeddings, allowing for better identity persistence in the presence of occlusions and Choppy scenes.

Unlike previous models, which target detection or tracking, the introduced CMBA module makes additional enhancements by inspecting spatial relations and motion behaviour of each individual for detecting social distancing violations, forbidden area violations, and abnormal crowd behaviours. On the other hand, this integration of detection tracking and behavioral analysis in such a unified way makes this framework line by line continuous with the previous approaches and makes crowd behavior monitoring in real-time and an end-to-end manner.

The major contributions of this work can be briefly summarized as follows:

1. This new architecture is a novel customized YOLOv4-tiny Model with newly developed custom anchor boxes and also the attention block, which focuses on bettering the detection accuracy for dense and obstructed crowded scenarios.
2. A fast implementation of Non-Maximum Suppression (NMS) to refine the detection output and eliminate any needless bounding box from the output.
3. A Deep SORT-based framework for motion prediction and appearance-based embeddings for reliable identity conservation
4. We propose a spatio-temporal and energy-efficient Crowd Monitoring and Behaviour Analysis (CMBA) module which can effectively detect social distancing and restricted areas violations, as well as abnormal crowd behaviour.
5. A model that has been jointly trained in a single time for full cycle detection, tracking, and behaviour analysis, which has yielded output in real-time better than the existing state of the art.

The remainder of this paper is organized as follows. Section 2: Related Work and Research Gaps Section 2 elaborates on the related work and the current research gaps. Section 3 introduces the proposed approach and system architecture. Sec. explains the performance analysis and experimental results. Results and limitations are discussed in Section 5, while Section 6 is the conclusion and points to future lines of research.

2. Related Work

Since crowd analysis is one of the major tasks in public safety and surveillance, it involves multiple challenges such as anomaly detection, behavior recognition, density estimation, and emotional analysis. The following subsections of this paper detail the literature written on crowd behaviour analysis using AI, focusing on different facets of the problem.

Crowd analysis has become an essential component of modern intelligent surveillance systems due to its applications in public safety, traffic management, disaster prevention, and smart city monitoring. Sreenu and Durai [1] presented a comprehensive review of deep learning-based intelligent video surveillance, highlighting the importance of automated crowd analysis for large-scale monitoring. Sánchez et al. [2] further categorized crowd behavior analysis into anomaly detection, emotion recognition, and behavior understanding while identifying several open research challenges.

Real-time crowd behavior recognition has also received considerable attention, where Rezaei and Yazdi [3] developed a deep learning framework for recognizing crowd activities in surveillance videos, whereas Rezaee et al. [4] summarized recent advances in deep learning-based crowd anomaly detection and discussed the limitations of existing surveillance systems. Similarly, Pawar and Attar [5] demonstrated the effectiveness of deep learning for abnormal crowd event detection, while Bendali-Braham et al. [6] reviewed recent developments in crowd detection, tracking, and anomaly recognition, emphasizing the need for integrated surveillance frameworks. To improve public safety, Amrutha et al. [7] proposed an intelligent suspicious activity detection model, whereas Ammar and Cherif [8] introduced the DeepROD framework for panic behavior detection using LSTM networks.

Furthermore, Gupta et al. [9] developed CrowdVAS-Net for abnormal crowd motion prediction by utilizing motion-based features, while Bour et al. [10] discussed crowd behavior analysis from both fixed and moving camera systems for surveillance applications. The importance of semantic understanding of crowd behavior was highlighted by Zitouni et al. [11], who surveyed socio-cognitive crowd analysis methods. In addition, Matkovic et al. [12] proposed a CNN-based abnormal behavior recognition framework incorporating tracking and fuzzy logic, whereas Nayan et al. [13] employed optical flow correlation analysis for efficient anomaly detection in crowded scenes. Although these studies significantly advanced intelligent surveillance, most of them focused on individual tasks such as detection, tracking, or anomaly recognition, with limited attention given to unified real-time crowd understanding.

Machine learning and deep learning techniques have significantly improved the accuracy and automation of crowd behavior analysis by enabling robust feature extraction, anomaly recognition, and spatio-temporal modeling. Direkoglu [14] proposed a CNN-based framework using Motion Information Images (MII) to detect abnormal crowd events, demonstrating superior performance on benchmark surveillance datasets.

Cheong et al. [15] developed an automated video analytics framework for crowd monitoring and counting, showing its practical applicability in real-world surveillance environments. To improve city-wide surveillance, Singh et al. [16] introduced a deep learning-based human crowd detection system capable of identifying individuals in complex urban environments. Furthermore, Zitouni et al. [17] proposed a probabilistic framework for socio-cognitive crowd behavior analysis, providing a structured approach for understanding collective human activities from surveillance videos.

A comprehensive review by Nayak et al. [18] analyzed various deep learning-based video anomaly detection techniques and identified challenges such as limited datasets, computational complexity, and poor cross-domain generalization. Object detection and tracking were further enhanced by Juyal and Sharma [19], who integrated SSD with the Deep SORT algorithm to improve people localization and trajectory estimation in surveillance videos. Similarly, Singh et al. [20] introduced an aggregation of fine-tuned convolutional neural networks for crowd anomaly detection, achieving higher detection accuracy than conventional deep learning models. Mohan et al. [21] presented a machine learning-based anomaly and activity recognition framework capable of automatically detecting suspicious events from surveillance footage.

Beyond anomaly detection, Tripathi et al. [22] investigated crowd emotion analysis using two-dimensional convolutional neural networks, demonstrating that emotional understanding can support crowd management during public events. Ramachandran and Sangaiah [23] proposed an unsupervised deep learning framework for anomaly event detection in crowded scenes, reducing dependence on manually labeled training data. In addition, Hu [24] designed a spatial-temporal deep learning architecture for abnormal behavior detection in massive surveillance systems, while Aljaloud and Ullah [26] introduced an irregularity-aware semi-supervised learning model capable of identifying unusual crowd events using only limited labeled data.

Complementing these approaches, Al-Shaery et al. [25] provided an extensive survey of crowd detection, monitoring, and management systems, emphasizing the importance of integrating artificial intelligence, computer vision, and real-time analytics to improve surveillance efficiency. Although these studies demonstrate remarkable progress in machine learning and deep learning for crowd analysis, many existing methods remain specialized for individual tasks such as detection, tracking, or anomaly recognition, thereby highlighting the need for unified frameworks capable of simultaneously performing multiple crowd analysis functions under real-world surveillance conditions.

Anomaly detection has become one of the most important research areas in intelligent surveillance because it enables the automatic identification of unusual crowd activities, thereby supporting timely intervention and public safety. Varghese and Thampi [27] emphasized the application of cognitive computing for smart crowd management, highlighting the importance of intelligent decision-making in dynamic environments. Salim et al. [28] proposed a video-based framework for crowd detection, tracking, and counting, demonstrating its effectiveness in surveillance applications while suggesting improvements in density estimation.

To recognize abnormal crowd activities, Fu et al. [29] introduced a hybrid deep learning model combining CNN and LSTM with an attention mechanism, achieving promising results in complex surveillance scenarios. Similarly, Ilyas et al. [30] developed a hybrid deep neural network integrating handcrafted and deep features for crowd anomaly detection, showing improved performance on benchmark datasets such as UMN and PETS2009. Guo et al. [31] proposed an intelligent crowd abnormal behavior analysis method for video service robots, enabling rapid identification of unusual crowd movements. Meanwhile, Kajabad and Ivanov [32] demonstrated the effectiveness of YOLOv3 for people detection and Gaussian Mixture Models for identifying high-density crowd regions, thereby improving surveillance efficiency.

A comprehensive review by Elharrouss et al. [33] summarized recent developments in intelligent video surveillance systems and discussed future directions involving artificial intelligence and deep learning. Ullah et al. [34] introduced a multi-feature-based crowd video modeling approach that combined appearance and motion information to improve visual event detection under challenging surveillance conditions. In addition, Zhang et al. [35] proposed a fuzzy inference model for evaluating crowd emotions using arousal and valence features, providing valuable insights into behavioral understanding. To classify different crowd types, Wei et al. [36] developed a very deep two-stream neural network capable of learning complementary spatial and temporal representations. Li et al. [37] comprehensively reviewed data fusion techniques for intelligent crowd monitoring and management systems, emphasizing the importance of integrating heterogeneous sensor information for improved surveillance performance.

Furthermore, Gong et al. [38] investigated the use of social media images for crowd counting during city events, demonstrating the potential of alternative visual data sources for crowd management. Yao et al. [39] proposed a residual neural network for realistic crowd behavior simulation and movement prediction, while Gao et al. [40] introduced a feature-aware domain adaptation framework for crowd counting that effectively reduced the performance gap between synthetic and real-world surveillance datasets. Although these approaches significantly improved anomaly detection and crowd behavior analysis, most studies focused on specific surveillance tasks independently, limiting their ability to simultaneously perform accurate detection, reliable tracking, behavioral reasoning, and real-time decision-making within a unified intelligent surveillance framework.

Accurate crowd density estimation and counting are fundamental components of intelligent surveillance systems, providing essential information for crowd management, public safety, and urban planning. Pustokhina et al. [41] proposed an automated deep learning framework combining Mask R-CNN and DenseNet for anomaly detection in pedestrian walkways, demonstrating improved safety monitoring capabilities. Wu et al. [42] introduced a temporal-aware neural network for fast video crowd counting that efficiently exploited temporal dependencies while reducing computational complexity. Tay et al. [43] developed a CNN-LSTM model with an attention mechanism for abnormal behavior recognition, effectively capturing both spatial and temporal information in surveillance videos. During the COVID-19 pandemic, Rezaee et al. [44] proposed a smart visual sensing framework based on modified deep transfer learning for detecting overcrowding in public places, achieving high detection accuracy in real-world environments. Similarly, Alotibi et al. [45] designed a CNN-based crowd counting model integrated with IoT infrastructure to improve monitoring in public locations, demonstrating its applicability in smart city environments. Zhang et al. [46] comprehensively reviewed edge video analytics for public safety and highlighted the advantages of deploying artificial intelligence models directly on edge devices for low-latency surveillance applications.

To enhance anomaly detection, Lin et al. [47] introduced the Social Multiple Instance Learning (Social-MIL) framework, which effectively modeled social interactions among pedestrians for improved crowd anomaly recognition. A detailed review by Li et al. [48] summarized recent advances in crowd counting and density estimation techniques, discussing conventional approaches, deep learning architectures, benchmark datasets, and future research opportunities.

Hu et al. [49] proposed a weakly supervised framework for abnormal behavior detection and localization, reducing the dependency on extensive manual annotations while maintaining competitive performance. Furthermore, Lin et al. [50] utilized synthetic datasets together with domain adaptation techniques to improve anomaly detection in crowded scenes and enhance model generalization across different surveillance environments.

Wong et al. [51] investigated pedestrian trajectory recognition using deep learning to support intelligent movement analysis and urban planning applications. Gruosso et al. [52] proposed a deep learning-based human segmentation framework for surveillance videos that improved person localization under challenging background conditions. Overall, these studies significantly advanced crowd counting, localization, and density estimation; however, most approaches primarily concentrated on quantitative crowd analysis and offered limited capability for comprehensive behavioral understanding, long-term tracking, and integrated real-time surveillance decision support.

The rapid advancement of artificial intelligence, Internet of Things (IoT), edge computing, and intelligent sensing technologies has transformed conventional surveillance systems into proactive crowd monitoring platforms capable of supporting smart city applications. Negri et al. [57] demonstrated the potential of image-based social sensing by combining artificial intelligence with crowd-sourced social media data to evaluate public policy adherence and crowd activities. Similarly, Qasim and Bhatti [58] proposed a hybrid swarm intelligence-based framework for abnormal event detection in crowded environments, achieving improved detection performance through optimized feature selection. Aberkane and Boudihir [59] explored deep reinforcement learning for anomaly detection in surveillance videos, illustrating the capability of reinforcement learning to adaptively recognize abnormal crowd behaviors. The importance of large-scale data analytics was highlighted by Subudhi et al. [60], who discussed the role of big data technologies in modern video surveillance systems for processing massive multimedia streams efficiently.

Focusing on transportation environments, Hsien and Atmosukarto [61] introduced a deep learning-based video analytics framework for train cabin crowd monitoring that enhanced passenger density estimation. Castellano et al. [62] proposed a fully Convolutional Neural Network for crowd detection in aerial images, demonstrating the applicability of deep learning in drone-assisted surveillance systems.

To evaluate crowd stability, Zhao et al. [63] developed an improved multi-column convolutional neural network capable of analyzing crowd dynamics under dense conditions. Likewise, Alotaibi et al. [64] conducted a comparative analysis of convolutional neural network architectures for large-scale crowd counting, highlighting their strengths and limitations across different benchmark datasets. Khan et al. [65] comprehensively reviewed supervised learning approaches for human detection and activity recognition, emphasizing the effectiveness of deep neural networks in surveillance applications.

Nogueira et al. [66] proposed RetailNet, a deep learning framework for people counting and hotspot detection in retail environments, enabling real-time customer flow analysis. Furthermore, Mohammed et al. [67] introduced a motion pattern-based scene classification framework using adaptive synthetic oversampling and deep neural networks for improved crowd movement analysis, while Basalamah et al. [68] developed a synthetic data-driven deep learning framework for congestion detection in public places, demonstrating enhanced generalization through artificial data generation. Ruchika and Purwar [69] proposed a crowd density estimation approach based on Hough Circle Transform and background segmentation techniques, offering an efficient alternative for video surveillance applications.

Finally, Singh et al. [70] presented a neural network-based framework for real-time anomaly recognition using CCTV surveillance, illustrating the growing adoption of intelligent vision systems for automated public safety monitoring.

Although these studies have significantly strengthened intelligent surveillance through advanced artificial intelligence techniques, IoT integration, and large-scale analytics, many existing frameworks still address individual surveillance tasks independently and lack comprehensive solutions capable of integrating detection, tracking, behavioral analysis, and decision support within a unified real-time surveillance architecture.

Recent advancements in artificial intelligence have significantly improved intelligent crowd analysis by integrating deep learning, edge computing, attention mechanisms, and real-time surveillance technologies. Zia-ul-Rehman et al. [71] proposed an AI and IoT-based framework for real-time crowd monitoring that combined intelligent sensing, communication, and automated decision-making to enhance public security in smart environments. Similarly, Alharbi [72] introduced the DeepCAMS framework, which employed spatial-temporal deep learning to detect suspicious crowd behaviors from surveillance videos while improving monitoring accuracy under dynamic environmental conditions.

To further enhance anomaly detection performance, Altowairqi et al. [73] developed a hybrid C3D-LSTM architecture integrated with attention mechanisms for learning complex spatio-temporal crowd representations, demonstrating improved abnormal behavior recognition in dense surveillance scenarios. The adoption of modern object detection architectures has also gained considerable attention, where Siva et al. [74] proposed a scalable YOLOv8-based surveillance framework capable of simultaneously detecting crowds and security threats with high computational efficiency.

Likewise, Nasir et al. [75] presented an enhanced YOLOv8 framework specifically designed for abnormal behavior detection in dense crowds, showing improved localization accuracy and faster inference compared to earlier YOLO-based models. Beyond behavior detection, Xu et al. [76] utilized a Graph Convolutional Network combined with Recurrent Neural Networks (GCN-RNN) to predict passenger flow by modeling both spatial relationships and temporal crowd dynamics, demonstrating the importance of predictive crowd analytics.

Furthermore, Shathik et al. [77] developed a smart vision system integrating edge computing and deep learning for real-time crowd monitoring and anomaly detection in urban environments, thereby reducing communication latency and improving response time.

Finally, Akhtar [78] provided a comprehensive review of recent developments in AI-based computer vision systems, discussing emerging challenges related to explainability, ethics, privacy preservation, and trustworthy surveillance while emphasizing the need for intelligent, transparent, and socially responsible monitoring systems. Although these recent studies demonstrate remarkable progress in deep learning, attention mechanisms, edge intelligence, and modern object detection frameworks, most existing approaches continue to focus on individual surveillance components such as detection, prediction, or anomaly recognition.

Consequently, there remains a significant need for a unified intelligent crowd analysis framework capable of simultaneously performing accurate person detection, robust multi-object tracking, comprehensive behavioral analysis, and real-time decision support within a single end-to-end surveillance system. Despite substantial progress in intelligent crowd surveillance, several critical limitations remain unresolved across existing studies.

Most traditional approaches primarily focus on a single task, such as crowd detection, anomaly recognition, crowd counting, density estimation, or behavior classification, without providing an integrated framework capable of performing these operations simultaneously. Deep learning-based detection models have significantly improved object localization accuracy; however, they often experience performance degradation under dense crowd conditions, severe occlusions, scale variations, and complex background environments.

Similarly, although modern tracking algorithms have enhanced identity preservation, maintaining consistent trajectories during prolonged occlusions and crowded interactions remains challenging. Recent advances in anomaly detection and behavior analysis have demonstrated promising performance using attention mechanisms, recurrent neural networks, and transformer-based architectures, yet many methods rely heavily on computationally intensive models that limit real-time deployment on resource-constrained surveillance systems. Furthermore, several intelligent surveillance solutions lack comprehensive behavioral reasoning, explainability, and unified decision-making capabilities, making them less suitable for practical smart city applications.

Emerging AI, IoT, and edge-computing frameworks have improved real-time monitoring; however, they generally address isolated surveillance problems instead of providing complete end-to-end crowd intelligence. Therefore, there remains a strong need for an integrated surveillance framework that combines efficient person detection, reliable multi-object tracking, intelligent crowd behavior analysis, violation detection, and real-time alert generation within a computationally efficient architecture.

Addressing these limitations provides the primary motivation for developing the proposed Intelligent Crowd Analysis (ICA) framework, which integrates a Modified YOLOv4-tiny detector, Non-Maximum Suppression, Deep SORT tracking, and a dedicated Crowd Monitoring and Behavior Analysis (CMBA) module to achieve comprehensive, real-time, and intelligent crowd surveillance.

3. Proposed Framework

Here, a framework is proposed that combines a fused work, best suited for identifying crowd behavior state in order to accurately detect crowd behavior in real time and at a certain distance. This has low-early-onset video sections, the CCT, HLT/VET, and spatial categorization modules; various methods act independently. It uses as input a video that it processes, of sorts, to find frames to analyze. This preprocessing further prepares the data input for the subsequent stage of object detection. The YOLOv4-Tiny detection model is adapted for detecting people in the crowd.

This is modified to YOLOv4-tiny, which, in real-time surveillance scenarios, confers the high-density, overlapping objects in an order distributed over edge devices for low computation but high result objects. Post object detection, Non-Maximum Suppression (NMS) is applied to improve detections by removing overlapped bounding boxes and keeping only the best score predictions. These detections are then filtered and sent to the Deep SORT tracking module that keeps track of people even in complicated Scenes over time, frame by frame. Deep SORT: The system used in this solution to monitor the behavior and trajectory of each individual across frames provides robust tracking in high-density scenarios.

Crowd monitoring and analysis, the next level. Application of hand-action recognition on oriented/centered hand images (bottom) on explicit gesture detection images (middle) instead of whole-body input, which also allows additional forms of behavior recognition (top) from detected/tracked people (module). Used to track abnormal activity through crowd energy levels, violations of social distancing are calculated according to the distance among individuals using edge-based or center-based measuring modes, based on the scene. Restricted area detection allows you to detect any individual who enters a forbidden zone in an instant. It also generates optical flow, heatmaps, and energy graphs from crowd movement for advanced analytics in Figure 1.

For instance, learnt that Optical Flow visualizes the direction of movement of a crowd, heatmaps show non-moving or stationary points on the image, and Dynamic Energy graphs quantify dynamic energy levels, respectively (You can check in the paper for these). Crowd Behavior Analysis Reports Can Help in Making Effective Decisions with Respect to Public Safety. Finally, data logging and alerting; the system either logs the important data for later or creates an alert to the people that it has a safety problem in real-time. The all-encompassing combination of modified YOLOv4-tiny, Deep SORT, and behavior analysis results in the system being worthwhile for usage in CCTV surveillance systems, which takes effects from human errors out of crowd monitoring applications, with a fast and dependable action on sustaining public order.

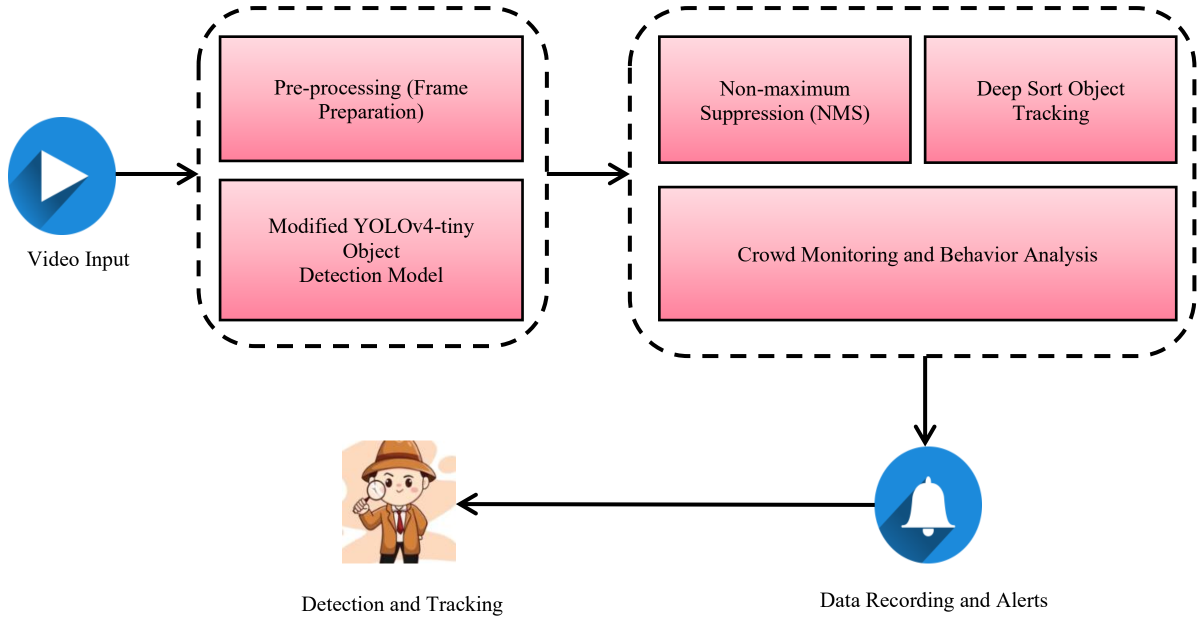

Fig. 1 Proposed framework for a learning-based approach for detection and behavior analysis in surveillance videos

3.1. Modified YOLOv4

Modified YOLOv4-tiny model. The proposed system applies a modified aspect of the compact YOLOv4-tiny model, and it is demonstrated in Figure 2, which is suited to real-time crowd monitoring and behavior analysis, focusing on the unique behavior challenges of surveillance scenarios, including high-density crowds, dynamic migration effects occurring, and limited hardware resources. As a miniature version of the YOLOv4 object detection model, YOLOv4-tiny is optimized for quick inference, taking some hit on accuracy to achieve a higher speed target; it is a good candidate designed for edge devices and real-time applications.

However, it is very similar to AlexNet, with some changes applied to improve crowd performance in surveillance applications. The first change on YOLOv4-tiny is to optimize it to detect humans in crowded environments. This means fine-tuning the model with a specific dataset designed for surveillance settings, where this dataset consists of labeled images showing crowded scenes, people of different scales, and complex conditions (occlusions or low light). It ensures that the model can detect individuals even in complex and dynamic backgrounds, thus improving the overall robustness compared to a generic pre-trained model when re-trained on such a dataset.

The anchor box sizes were optimized. YOLOv4-tiny also uses a concept of anchor boxes, which are fixed sizes of bounding boxes that are used to predict the location and scale of objects. It can localize objects more precisely than a vanilla YOLOv4 tiny. It pushes the dataset and tunes the size of the container box to be more appropriate to the scale and shape of the regular people found in surveillance images. This is important in processes where humans have different sizes (for example, perspective warp occurs with wide-angle or overhead shots, and they need to be corrected). Moreover, feature extractors in YOLOv4-tiny can be enhanced with attention mechanisms to increase efficiency and performance of the model. Attention modules, such as the Squeeze-and-Excitation (SE) block, can be incorporated within the backbone of the model to get better at focusing on relevant features and suppressing background clutter [11].

But it is this fine-tuning that changes the model to identify people in challenging situations, with multiple humans in a field of view. A third important one is that the model is computationally efficient. Although the YOLOv4-tiny can be pruned or quantized very easily from this point, it will always continue running several times faster than any full-size model on low-resource edge devices (i.e., surveillance cameras or embedded systems).

These techniques enable real-time computation by either pruning similar parameters or operating upon low-precision arithmetic, functionality that is necessary for surveillance computation. The post-processing mechanism on the modified YOLOv4-tiny is also tailored to how it fits the needs of the proposed system.

For instance, its outputs are processed to estimate the distance between individuals and find out whenever they have breached the social distance level. Bounding box predictions are then filtered by a method called Non-Maximum Suppression (NMS), which discards duplicate detections of the same object across neighbouring pixels, resulting in cleaner input for downstream stages such as tracking and crowd analysis. To sum up, YOLOv4-tiny brings all these improvements, and therefore can be seen as a light-weight yet great backbone of the designed surveillance framework. These lagging factors can be overlooked, as a high-precision real-time detection of people atleast compensates for that, ensuring the system does what it is intended to do, being an effective pro-active crowd monitoring and behaviour tracking system of high accuracy. This model, with these updates, represents a core part of any solution for a safety-critical surveillance application needing high performance.

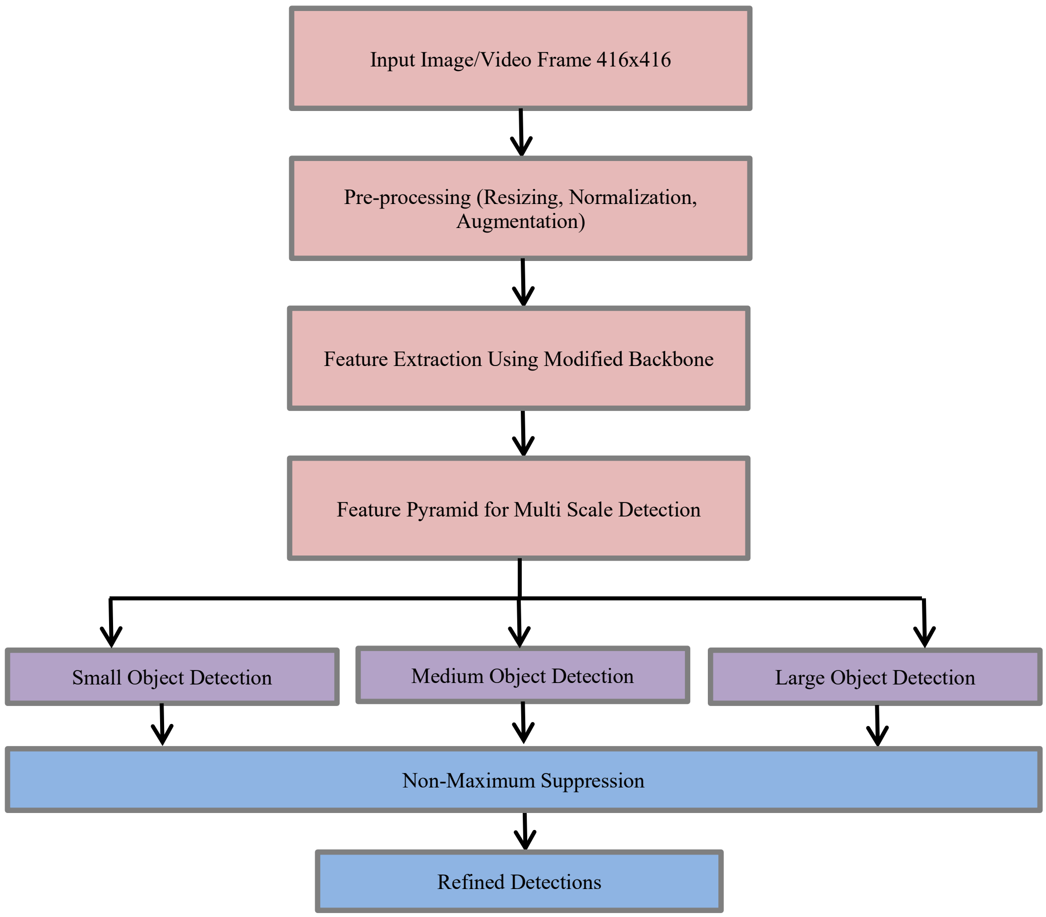

Fig. 2 Modified YOLOv4 used in the proposed framework for a learning-based approach for detection and behavior analysis in surveillance videos

3.2. Non-Maximum Suppression

Non-Maximum Suppression (NMS) is a type of post-processing that optimizes the output of the YOLOv4-tiny model, like the proposed system, by removing overlapped or similar bounding boxes. This NMS keeps only the most precise detection per object, minimizing false positives and preventing overlapping of separate detections within crowded scenes. Upon detecting a frame from the video, YOLOv4-tiny will draw multiple bounding boxes around objects along with a confidence score that indicates how confident the network is the detection.

This specific box contains an object of interest. The model tends to return many overlapping boxes for the same object in situations where objects overlap or when the crowd is dense, and this redundancy can lead to problems. Non-Maximum Suppression (NMS) solves this by only keeping the detection with the highest score and removing other significantly overlapping detections. The New Non-Maximum Suppression (NMS) process starts by sorting all detected bounding boxes based on their confidence scores in descending order, as shown in Figure 3.

NMS starts from the box with the highest confidence score and calculates IoU between this box and all boxes of the same object class remaining. Intersection over Union (IoU) The IoU is a metric that measures the overlap between two bounding boxes as the area of intersection divided by the area of union. If the primary box has an IOU greater than a specific threshold (0.5 in this study) compared to another box, then other boxes will be suppressed (eliminated) from consideration. The suppression is repeated for all the boxes (and all the detections) one at a time in the same way, and only the box with the highest confidence score (which is still not suppressed) is moved on to the next by the same loop of this solitary visual word algorithm based on the duplicate IoU. In non-overlapping bounding boxes that represent different objects, this guarantees a unique maximum score. The resulting output then passes on to the next stage, such as Deep SORT for tracking, so that the algorithm can operate on accurate and non-redundant input.

In early NMS (Non-Maximum Suppression), where multiple detections occur for people in overlapping positions in high-density crowd use cases, NMS is [6] less critical. NMS helps avoid processing redundant boxes with low scores and thus boosts computational efficiency at the tracking stage, while it also makes downstream analytics more reliable in two important tasks: social distance violation detection and abnormal activity monitoring. An IoU threshold is adjustable depending on the application, balancing sensitivity (detecting all possible objects) with precision (not detecting positive objects incorrectly). In other words, NMS is an important step in the entire object detection pipeline that reduces the outputs from the YOLOv4-tiny model to suitable locations to be fed into the proposed advanced tracking and behavior analysis tasks afterwards in the surveillance framework. This aligns with the goals behind the design of the system, which focuses on real-time monitoring and crowd management.

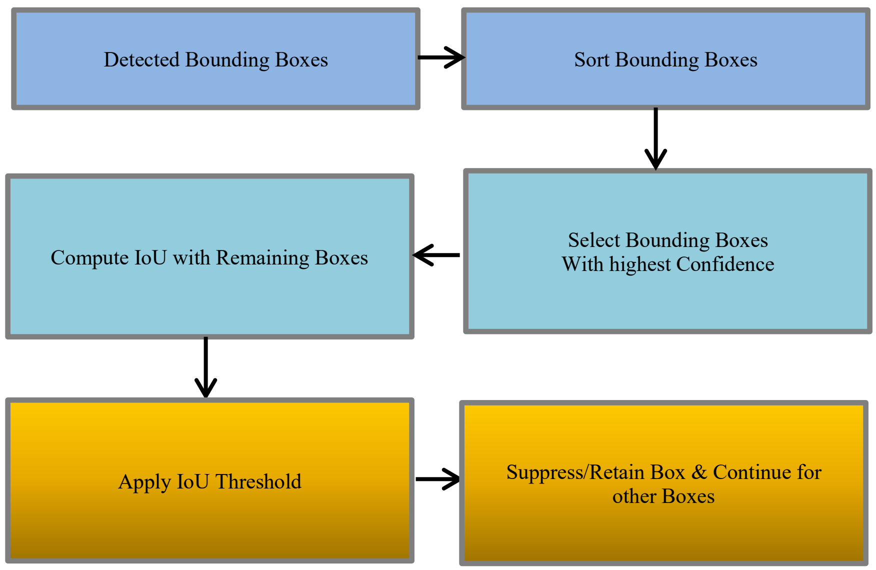

Fig. 3 Functional flow of non-maximum suppression

3.3. Deep Sort-Object Tracking

As depicted in Figure 4, Deep SORT (Simple Online and real-time tracking with a Deep Association Metric) is an advanced object tracking approach that extends the earlier SORT framework with the spirit of deep learning-based association metrics. Deep SORT ensures accurate and robust person tracking in highly crowded and dynamic environments as part of the detection and tracking in this proposed framework. Tracking without loss of follow-up is crucial for ongoing crowd behavior monitoring and analysis over time. Deep SORT mainly links detected objects across multiple frames in a video stream.

Once the YOLOv4-tiny model detects objects (e.g., people in a scene), use Deep SORT to assign the following unique identities and track these objects across the frames. Kalman filtering is combined with a Hungarian algorithm, ensuring the objects are correctly associated with their previous detections. The significant contribution of Deep SORT is the integration of a deep learning-based feature extractor to enhance the association metric. In contrast to classical SORT, which uses only motion and position information in correspondence, Deep SORT extracts appearance features from detection using a pre-trained Convolutional Neural Network (CNN). This can be useful to further discriminate visually similar objects, especially in cluttered scenes, where occlusions and overlaps are common.

Deep SORT combines motion and appearance features for greater accuracy and robustness. This includes behavior analysis, such as breaches of social distance, entering a restricted area, and abnormal crowd behaviors, which are critical in the proposed framework, Deep SORT. That is to say, if it does have the trajectory of every single individual, the system could be able to measure distances between individuals over time and know when an individual has entered a prohibited area, like crossing from one red zone or another within a white zone (during night-time) not being that far away from each other at least in terms of distance or intermittent movements based on periods that prove some sudden change which indicates deviations–potential anomalies. Moreover, as Deep SORT maintains unique identities, it ensures proper data collection for further analyses like optical flow generation, heatmaps, and energy graphs.

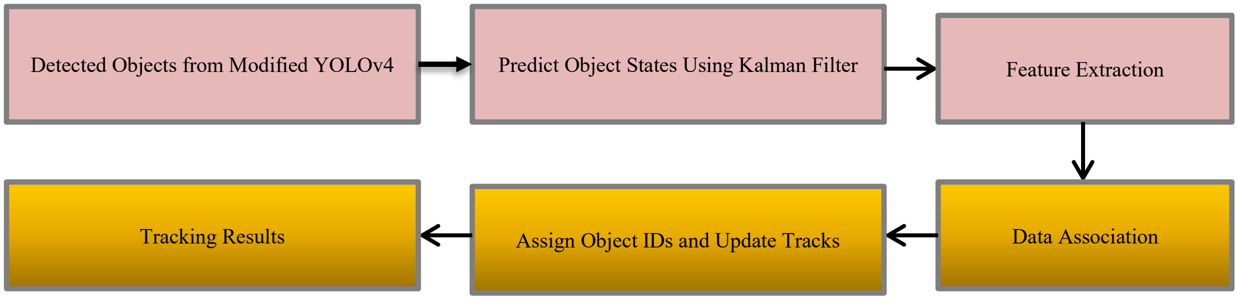

Fig. 4 Functional flow of deep sort-object tracking

Deep SORT also has the benefit of running in real time. Nevertheless, while this technique applies deep feature extraction to improve the precision of detecting an object in general and humans in particular, it is still computationally inexpensive to afford its use in real-time surveillance applications. The efficiency is also improved by applying YOLOv4-tiny as the object detection algorithm, which can achieve a high frame rate for tracking while maintaining good accuracy.

Thus, deep SORT ensures reliable multi-object tracking when integrated with the framework, providing one of the critical components for effective crowd monitoring and performance analysis of behavior. Due to the challenges faced by physically intelligent surveillance, such as occlusions, re-identification, and motion prediction in crowded scenarios, tracking is considered a vital element of the proposed system.

3.4. Proposed Algorithms

They discuss a number of algorithms for implementing the proposed system. It requires the ICA algorithm, which is proposed here to build a strong and timely running framework for crowd anomaly detection, violation detection, and crowd behavior analysis in various public surveillance applications. The proposed algorithm integrates the latest in cutting-edge technology for object detection and tracking and behavior analysis to ensure safety while performing these tasks in high-density human environments.

In this, it detects basic interventions such as the breaking of social distance, entering prohibited areas, and abnormal crowd movement for proactive interventions and decision-making. In addition, the algorithm provides comprehensive performance analytics enabling deployment in smart cities, mass gatherings, and security-sensitive deployments.

Algorithm 1: Intelligent Crowd Analysis (ICA)
Input: Video stream V
Output: Detected crowd abnormalities or violations R, performance statistics P

```
Begin
Initialized flagged events vector E
V'←Preprocess(V)
For each frame f in V.'
  f'←Normalization(f)
  D←Human Detection(f')
  D'←Non-Maximum Supression(D)
  T←Deep Sort Object Tracking(D')
  E←Empty
  For each tracked object t in T
    Violation Flag←Crowd Monitoring and Behaviour Analysis)
    Add violation Flag to E
  End For
  For each violation, Flag in E
    Generate Alert for violation Flag
    Update R
  End For
End For
P←Performance Evaluation (R, ground truth)
Print R
Print P
End
```

Processing the Video Streams of Surveillance for Crowd Abnormality Analysis and Violation Detection Using an Intelligent Crowd Analysis (ICA) Algorithm for Behavior Analysis. The algorithm works as it first prepares a vector structure called E in which events such as violations or abnormalities will be recorded [8].

It could help the system detect real-time occurrences and interactively respond. Next, the input video stream V passes through a series of pre-processing steps (e.g, frame resize, pixel normalization), producing an intermediate video V' that represents a sufficiently pre-processed version of the original video ready to be processed by the algorithm. Then handle f, each frame in V' one at a time.)

Whereas the algorithm normalizes the input frame at all the pixel intensities in such a way that it well distributes and fetches the accuracy for the next detection purposes. As its input, the human detection module accepts the aforementioned normalized frame, f', and applies a slightly modified YOLOv4-tiny model that outputs bounding boxes around the people in the crowd.

Hence, this stage will result in a matrix [D] with detected bounding boxes, object class labels, and confidence scores. NMS is then applied on D, producing D' NMS, which reduces the same bounding boxes (with high IOU (intersection-over-union)) to a bounding box and, as an additional, unique aliasing operation, leaving the least noisy tracking data.

The refined detections are then sent to the Deep SORT object tracking module, which assigns a unique ID to each detected person and tracks their associated states frame by frame. This outputs T (tracked objects with trajectories and IDs). Variables in the proposed algorithms are shown in Table 1.

Table 1. Variables used in the proposed algorithms

Variable	Description
V	Input video stream
f	Current video frame
B	Detected bounding boxes
D	List of detections with bounding boxes, labels, and scores
D'	Final filtered detections after NMS
P	Predicted object states from the Kalman filter.
T	Object tracker maintaining trajectories and IDs
Z	Restricted zone coordinates
$D_{i,j}$	Distance between two objects
E_i	Movement energy for object i

The system monitors the crowd and analyzes behavior for each tracked object t in T. These include the computation of distances between individuals to determine if social distancing has been violated, detection of breaches in restricted areas using overlaps with predefined regions, and movement analysis for abnormal behavior detection, such as rapid exit from a crowded area.

A flag called violation flag, which is then appended to the list of events that happened (E), provides a clear indication for each violation or even an anomaly detected all tracked objects in a frame performed some action that flagged their violation, at this point algorithm goes through the event vector E, extracts every individual violation flag separately and generates adequate alert about a specific vision violation from each operator.

At the same time, the algorithm records all three violations that occurred in the background in the results vector R while ensuring that the information related to every event is maintained. After processing every frame in the video, the algorithm handles the performance evaluation of R against ground truth. This step extracts the performance metrics P to confirm the system's efficiency based on detection accuracy and processing [9]. Finally, the algorithm prints R (the detected violations) and P (the performance statistics) on how well the system functions. The ICA algorithm is scalable and can be run in real time, making it suitable for many surveillance applications. The algorithm employs novel detection, tracking, and behavior analysis techniques to deliver a reliable crowd-monitoring system that enables preventive measures in multiple public environments. The ICA algorithm calls several other algorithms for pre-processing, human detection, non-maximum suppression, object tracking, and crowd behavior analysis.

Preprocessing is a basic but essential algorithm of the Intelligent Crowd Analysis (ICA) system, which prepares video input for further processing. This model accepts low-level video stream information (V) and outputs the sequence of normalized video frames through V′ so that every input data is favorably uniform for object detection and analysis. Preprocessing is crucial to eliminating heterogeneity in downstream algorithms because, without it, the size and resolution of the frames and pixel intensity would be distributed non-linearly.

Algorithm 2: Preprocessing
Input: Video Stream V
Output: Normalized Video Frames V'

1. Begin
2. For each frame f in V
3. Resized Frame←Resize Frame (416 x 416)
4. Normalized Frame←Normalization (resized Frame)
5. Add normalized Frame to V'
6. End For
7. Return V'
8. End

The algorithm first loops through each frame (f) possibly contained in the input video stream (V). It does resize here, so every frame is resized to 416×416 pixels. Resize to 416×416 pixels - This resizing was performed for the input images so that they are suitable for the mean, while I trained the variant of the YOLOv4-tiny object detection model, since it works best with this size of square inputs. Another advantage of resizing is that it will reduce the computational overhead for larger frames, thus making it possible to use it in real-time applications. After that, the frames are normalized, which means scaling their pixel intensity values in the range of 0 and 1. This is a critical normalization that will ensure consistency in the input and increase the capability of the model to learn and identify patterns. This reduces lighting, contrast, and color intensity variations, which are the natural challenges for surveillance video streams in various environments.

The output frame will be a kind of output sequence (V′), which is a sequence of normalized frames. This continues cycling through this process for each frame in the video stream until all frames have been resized and normalized.

In the end, it gives output a set of normalized video frames (V) that can be input to methods like object detection as well as tracking.

The high-level overview of the algorithm is: since each video frame needs to undergo the same pre-processing before ICA can be performed on it, ensuring that the pre-processing for each frame is as uniform as possible boosts the overall strength and performance of the ICA system. Filtering a lot of the noise and prepping the input data sets the base to achieve precise and effective analysis of a crowd.

Algorithm 3: Human Detection Using Modified YOLOv4 (HDUMY)
Input: normalized Frame from V.'
Output: Detected objects D

1. Begin
2. D←Detect Objects (normalized Frame)
3. B = {B1, B2, ..., Bn} //B denotes bounding boxes corresponding to D
4. For each b in B
5. Class Label←Extract Class Label (b)
6. Conf Score←Extract Confidence Score (b, class Label)
7. (Xi, Yi, Wi, Hi)←Extract Bounding Box Coordinates (D, B)
8. End For
9. Return D (D along with bounding box B_i, class Label, confScore)
10. End

A core component of the Intelligent Crowd Analysis (ICA) system, the Human Detection Using Modified YOLOv4 (HDUMY) algorithm searches for normalized video frames to detect people. This Algorithm takes the pre-processed frames and detects the objects and outputs detected Objects(D) along with coordinates of the bounding box, class label, and confidence scores using the Modified YOLOv4 model. Extracting the geometrical structure that allows us to understand where every object in a scene will be is essential for further analysis, such as tracking, behavior, or anomaly detection. To do so, the first step of the algorithm is to feed it into a modified YOLOv4 model for object detection. Through this, the model detects all objects present in a frame and outputs a bounding box list = (B = {B1, B2,..., Bn}) where n Keypoints refers to the detected object, which is called D.

Each box relates to a region within the frame and specifies where an object is located. The algorithm extracts relevant information for each bounding box (b) in (B). In step one, extract the class label corresponding to (b), which indicates what type of object was detected, e.g., "person." Then, the confidence score of detection is calculated, indicating how likely it is that the detected object corresponds to a particular class. A higher score means that you can trust the result more. The algorithm provides the coordinates of a bounding box as well ((Xi, Yi, Wi, Hi)) for each detection. The coordinates define where the bounding box is located, and the relative size compared to a frame, indicating that this area contains an object. This high-resolution data allows ICA to track people and their movements in the next phase of the (Institute for Creative Automation) system. After traversing the bounding boxes, it will append the number of detected objects (D), all the class labels, the confidence scores, and the x and y coordinates of each bounding box. Now we have the overall detection output, and downstream components such as NMS and Deep SORT object tracking can work on the detection output. With the speed as well as accuracy of this customized YOLOv4 model, the actual algorithm retains more attractive human detection throughout a challenging situation like a densely populated or quickly moving location. Once you assimilate into the ICA structure, it provides you with the underlying base required to hold the advanced surveillance functionality to enable real-time, accurate crowd detection and analysis.

Algorithm 4: Non-Maximum Suppression (NMS)
Input: Detected objects D (with bounding boxes B), IoU threshold th
Output: Selected objects with bounding boxes D'

1. Begin
2. Sort D
3. For each B_i in D
4.   iou ← ComputeIoU(B_i, B_j) for all
5.   IF iou > th Then
6.   Suppress B_i
7. Else
8.   Add B_i to D'
9. End If
10. Return D'
11. End

Notice that Non-Maximum Suppression (NMS) is an inescapable post-processing procedure at one of the Intelligent Crowd Analysis (ICA) systems [3] to refine D obtained from the human detection stage. The purpose of this is to minimize false bounding boxes with a high intersection over union, which have been chosen from high-confidence detections for each object.

This will maintain the process results to be accurate and well-arranged, which will further aid in processes like object tracking and behavior analysis.

Inputs: The set of detected objects (D), represented by bounding boxes (B), confidence scores, class labels, and a specific Intersection Over Union (IoU) threshold (th). IoU threshold: This is a critical parameter for how much overlap two bounding boxes need to have to suppress one of the boxes.

It starts with sorting all the detected bounding boxes (B) in decreasing order of their confidence scores. This guarantees that the highest confidence bounding box is evaluated first to give priority to more reliable detections. It then loops over all bounding boxes (Bi) in (D). It computes the IoU with all remaining bounding boxes (Bj) in the list for every (Bi). The IoU is the intersection over the union of the two bounding boxes. For any bounding box (Bj), if the IoU between (Bi) and (Bj)>th, suppress (Bj) as it is redundant. At this stage, overlapping boxes are removed to remove boxes that probably refer to the same object and preserve, as the final detection, only the most confident one per object. If IoU < threshold between (Bi) and (Bj), append (Bi) to list D' of detections which are confirmed.

The above steps are repeated for all the bounding boxes in (D), and finally, get (D') a set of selected detections consisting of nonoverlapping bounding boxes.

After the algorithm performs this sequence for all the bounding boxes, it outputs (D′). These detections are refined and do not have redundancy, which leads to cleaner and more accurate output while maintaining the required information for further processes like Deep SORT object tracking.

The Non-Maximum Suppression (NMS) algorithm reduces overlapping detections, increasing the ICA's efficacy and accuracy.

This guarantee ensures the system always processes the most relevant information when tracking and analyzing behavior in crowded and dynamic surveillance scenarios.

Algorithm 5: Deep Sort-Object Tracking (DSOT)
Input: Selected objects with bounding boxes D'
Output: Tracker with matched objects T

1. Begin
2. For each frame f and selected objects D'
3. P←KalmanFilter(f) //predict object states
4. Assign a unique ID to each tracked object using motion and appearance similarity
5. Update T using tracked objects
6. End For
7. Return T
8. End

The DSOT algorithm plays an integral part in the Intelligent Crowd Analysis (ICA) system by maintaining the persistent identity and trajectory of objects detected from one video frame to another. Combining motion prediction with appearance similarity provides a unique ID for tracked objects, making the framework robust in dynamic and cluttered environments [17].

What this algorithm takes as input is the filtered set of targeted objects with bounding boxes (page index) (D′) after passing through Non-Maximum Suppression (NMS). The output is a Tracker (T) consisting of trajectories of matched objects and the object IDs. Frame-by-frame, the algorithm iterates through the video sequence (f) and the Detected Objects (D').

The Kalman filter estimates the states of previously tracked objects for each frame. It recognizes the movement of things, using their past position and speed to determine where they might be. This step provides a motion-based Prediction (P) for associating detections across frames. This algorithm allows both jaws to track by combining motion prediction and appearance similarity.

In (D'), each detected object obtains a unique ID by comparing its motion trajectory with the predicted states via the Kalman filter and its visual features. Deep feature embeddings from a Convolutional Neural Network (CNN) were used to compute the similarity of appearance.

The algorithm can use this characteristic of each object to tell the difference between similar objects and deal with occlusions. The matched objects get assigned unique IDs - this information is fed into the Tracker (T).

This update step is to associate the current detections with the predicted states for continuity in the tracking process. This allows the tracker to retain objects that have been matched successfully, adding unmatched detections if they can be identified as new tracks and even removing some tracks that did not get matched recently to deal with cases of disappearing objects.

This process is iteratively repeated over all frames in the video sequence, and an updated tracker keeps a list of the trajectory, IDs, and current status of all tracked objects. This output (T) gives a basis for further analysis, including behavior monitoring, anomaly detection, etc., because T provides information based on tracking data consistently and accurately.

By enabling reliable tracking of individuals even in complex scenarios, the DSOT algorithm significantly improves the support offered by the ICA system for real-time crowd behavior analysis. Incorporating similar motion prediction coupled with a person's appearance can achieve high accuracy, making it most applicable for surveillance, security, and monitoring public safety.

Algorithm 6: Crowd Monitoring and Behaviour Analysis (CMBA)
Input: Objects in tracker T, distance threshold th1, abnormality threshold t2
Output: Abnormal behaviour detection results R

1. Begin
2. For each object t_i in T

Violation of Social Distance

3. For each object, pair t_i and t_j
4. Di, j = sqrt ((Xi-Xj) ^2 + (Yi-Yj) ^2) //compute distance
5. If Di, j > th1 Then
6. flag violation
7. Add flag to R

End For
Monitor Restricted Area
If t_i is in a restricted zone, Z
flag restricted zone entry
Add flag to R
End If
Abnormal Crowd Activity
If movement energy Ei > th2, then
flag abnormal activity
Add flag to R
End For
Return R
End

The CMBA (Crowd Monitoring and Behaviour Analysis) algorithm is a crucial component of the ICA (Intelligent Crowd Analysis) system for performing real-time violation/abnormal behaviour detection in crowds. The events are analyzed through the Tracker (T), which processes all the objects that it tracks and detects any kind of violation, such as social distance violations, entry into a restricted area, and abnormal crowd activities. The final results (R) produced at the end are the events that indicate a breach and offer meaningful and actionable guidance on how to maintain safety in high-density areas. The following algorithm iterates over each tracked object (ti) in (T). The initial component being analyzed is the violation of social distance rules through mathematical guidelines in calculating the distance between every two groups (ti, tj) to measure Euclidean distance(D{i,j}) based on the coordinates of objects ((Xi, Yi)). If the role distance is less than a distance threshold th1, we say there is an intrusion and a defined violation occurs where object i is too close to object j, with R, where each indicated violation is appended for inspection or as a warning.

Next, the algorithm looks for break-ins to a no-go area → For each tracked object (ti), we determine whether its current position intersects with a no-go area (Z). If an object passes through this area, it would be flagged as an entry-exit violation in a restricted zone and logged in (R). This feature automatically detects and handles any movement in vulnerable areas. Detection of abnormal crowd activity is the final step in the algorithm.

For each object (ti), we calculate the movement energy (Ei), which describes the intensity of use combining speed and route. Abnormality threshold (th2) If (Ei)> th2 → abnormality in crowd behavior,→ people in the crowd are either dispersing rapidly or gathering up at a denser state. This abnormal activity is triggered and recorded on (R). This happens with all objects through The Algorithm. Besides monitoring and analyzing for social disturbance, it simultaneously detects entry to restricted zones and abnormal activity. After all of the objects have been traversed, R, containing the flagged events, is returned as output by the time these results can be converted to near-real-time alerts that will either trigger safety mechanisms or provide information that can help assess what went wrong on an operational level.

The CMBA algorithm provides a holistic analysis of crowd monitoring by integrating the spatial analysis, positional analysis, and behavioral traits of crowds. It can capture and identify violations and mark them in real-time, so that deep learning technology can help manage crowds, maintain public road safety, and quickly control risks in monitoring scenarios. We focus on the first two steps of the proposed framework, where the main novelty is the Crowd Monitoring and Behavior Analysis (CMBA) module that incorporates some form of structured behavioral reasoning in the surveillance pipeline. While traditional systems merely detect and track individuals over time, CMBA takes spatial and temporal relationships between individuals and translates them into actionable insights.

Lapts, ensure that the social distance between two individuals is measured by the Euclidean distance:

$$D(i,j) = \sqrt{[(Xi - Xj)^2 + (Yi - Yj)^2]}$$

A violation is detected when $D(i,j) < \tau d$, where τd represents a predefined or adaptive distance threshold.

Movement energy is defined to capture abnormal behavior patterns:

$$Ei = ||Vi|| + \lambda \Delta Vi$$

where Vi represents velocity, and ΔVi denotes acceleration. High energy values indicate sudden or abnormal motion.

Violation of restricted areas is also determined based on an overlap between the tracked positions and the established boundaries of restricted zones. It allows the system to conduct real-time behavioral reasoning, which is different from detection-based systems.

4. Experimental Results

The proposed Intelligent Crowd Analysis (ICA) algorithm has been tested in this data set as an example of effectiveness and robustness evaluation under the real-world surveillance data incorporated with synthetic situations. The experimental design principle was to capture video streams of different crowd densities, weather conditions, and camera angles to provide realistic surveillance scenarios. This system was then implemented on a regular computer utilizing GPU acceleration to achieve the high-performance capability required for real-time applications. The testing outcomes help a trustworthy human detection skill of modified YOLOv4-tiny whilst entering into challenging scenarios and handling problems like occlusions, overlapping objects, scale change, etc.

This was ensured using a Deep SORT tracker, which helped keep an individual identity of objects tracked even in dynamic environments. The Crowd Monitoring and Behavior Analysis (CMBA) module detected social distancing violations, restricted area entries, and abnormal activities in a crowd while giving actionable insights for better management. This system proved its potential for real-time crowd monitoring by highlighting a safer public and being able to supplement the areas of effective surveillance in complex environments. To ensure robust results in real environments, the Intelligent Crowd Analysis (ICA) system performance was evaluated on a range of benchmark datasets and collaborating on real video surveillance data. PETS 2009: This dataset contains multi-camera pedestrian sequences of different crowd density (low, medium, high), and precise ground-truth annotations with a total of 4200 annotated frames are used for different density level.

The UCF-Crowd dataset consists of 1535 images with density maps and head-point annotations over diverse real-world scenarios. In addition, a bespoke surveillance dataset was built using CCTV footage recorded for 30 days over various public environments (shopping complex and transit hubs), resulting in 8,400 annotated frames with bounding boxes, zone boundaries, and activity event labels across three violation classes, viz., social distancing violations, restricted area accesses, and abnormal crowd behavior.

This resulted in a total of 14,135 annotated frames that were randomly divided into training, validation, and test splits with a 70:15:15 ratio (9,894/2,120/2,121 frames, respectively). Augmentation strategies, including mosaic transformation, random horizontal flipping, HSV color-space adjustment (hue ±0.015, saturation ±0.7, value ±0.4), and random scaling, were performed only on the training set to deal with class imbalance and overfitting [18]. We ran all experiments on a workstation with an NVIDIA RTX 3080 GPU (10 GB VRAM), Intel Core i9-10900K CPU, and 64 GB DDR4 RAM running Ubuntu 20.04 LTS. We trained the Modified YOLOv4-tiny model using the Darknet framework, and imported it into a Python 3.8 / PyTorch 1.11 pipeline (via ONNX export). Video preprocessing and extraction of frames were performed using OpenCV 4.5. It was trained for 100 epochs using a batch size of 16 with an input resolution of 416×416 pixels.

We used Adam as an optimizer (with a base learning rate of 0.001 that decayed in learning rate by a factor of 0.1 at epochs 60 and 90). To avoid overfitting, we used Early stopping with a patience of 15 epochs. We used a Basic configuration by using the Deep SORT appearance feature extractor, which is pre-trained on the Market-1501 re-identification dataset, and fine-tuned it on the custom surveillance dataset for 20 epochs. NMS IoU threshold=0.5, confidence threshold=0.4 (found using grid search based on the validation set as mentioned in Section 3.2).

Furthermore, thresholds of the CMBA module, including the threshold of social distance (τd) and the threshold of abnormality (τa), were empirically defined by validation data. We performed sensitivity analysis by changing these thresholds to assess system stability; it shows that the chosen values provide the best compromise between detection and false rate (Table 4). The baseline models were retrained on identical data splits, augmentation pipelines, and hardware configurations. Full training details and model weights are available at [repository URL]

Precision, Recall, F1-score, and Accuracy have been computed to assess the overall quality of detection, and Mean Average Precision at IoU 0.5 (mAP@0. 5) localization accuracy, and FPS for assessing real-time capabilities. The evaluation protocol of MOTChallenge is referenced to report MOTA, MOTP, Identity Switches (IDS), and Mostly Tracked (MT) trajectories for multiobject tracking. The performance of the behavioral analysis was measured in terms of Precision, Recall, F1-score, and average detection delay (in milliseconds) based on the per-violation category.

Specifically, we compared our ICA framework against YOLOv3, Faster R-CNN, SSD, and FairMOT, which represent single-stage, two-stage, and joint detection-tracking paradigms under the same experimental settings. Robustness: All experiments were performed multiple times independently, and the average performance, together with the standard deviation and 95% confidence intervals, was reported.

The small variance across runs validates the robustness and consistency of the proposed ICA framework to perturbations. All implementation details, hyperparameter settings, dataset splits, and evaluation protocols are written out for further reproducibility. Source code and trained model weights are provided publicly to allow validation and extension of the proposed framework by independent researchers.

A crowd detection and anomaly detection system is illustrated in Figure 5. It uses object detection algorithms to detect people in a crowd and draw them with bounding boxes. Every sub-image shows how many people are found in particular scenes; these can be a public place like a shopping area, a square open space, a lawn, or an indoor public place. Panel Top row left: Street crowd (18 total, 16 violation count.) The yellow boxes look like they are marking some of the violators of a rule set, hence the disclaimer in social distancing. Green boxes likely represent individuals not violating the rules, top right panel: A less crowded area with a crowd count of 14.

The scene appears to be an open urban space, and all individuals are enclosed in green bounding boxes, suggesting no rule violations are being detected here. Bottom left panel: The image shows a gathering of fifteen people in a grassy setting, surrounded by green bounding boxes. It is most likely used to monitor crowd density or identify activity in public spaces.

The bottom right panel shows an indoor area, perhaps a building entrance, with a crowd of 10 people. Again, each individual is surrounded by a green bounding box, a sign of normal crowd behavior.

According to predetermined criteria, such as violations of social distance or hazardous crowding in public places, the system seems to primarily focus on counting the number of people present and maybe recognizing abnormalities. A computer vision system most likely powered by deep learning techniques for object detection, such as SSD or YOLO, is suggested by bounding boxes and real-time counting.

Figure 6 showcases optical flow analysis applied to crowd movement, a common technique used in abnormality detection for crowd behavior. Optical flow refers to the pattern of apparent motion of objects, surfaces, or edges in a visual scene caused by the relative movement between an observer (e.g., a camera) and the scene.

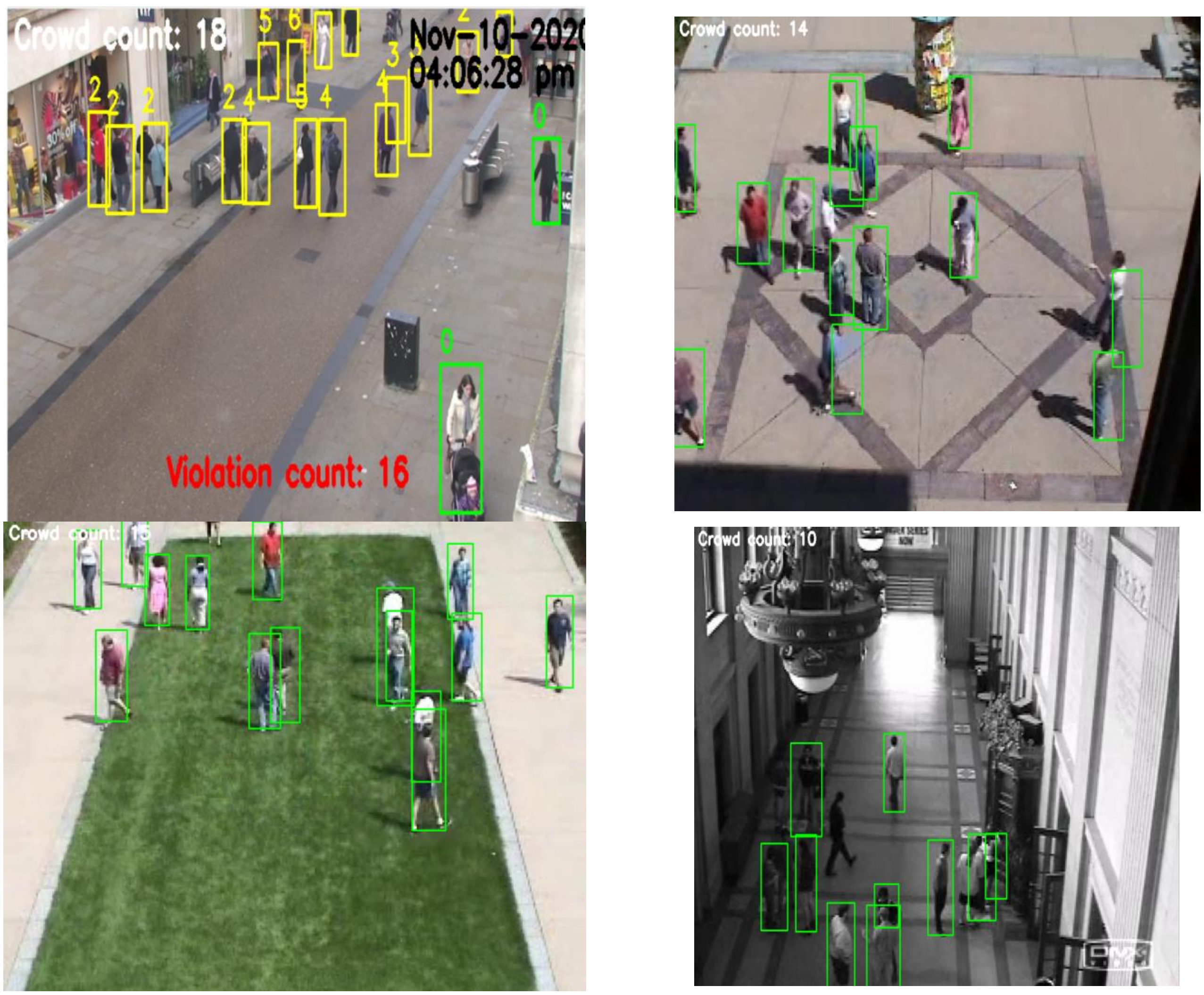

Fig. 5 Results of crowd behavior analysis

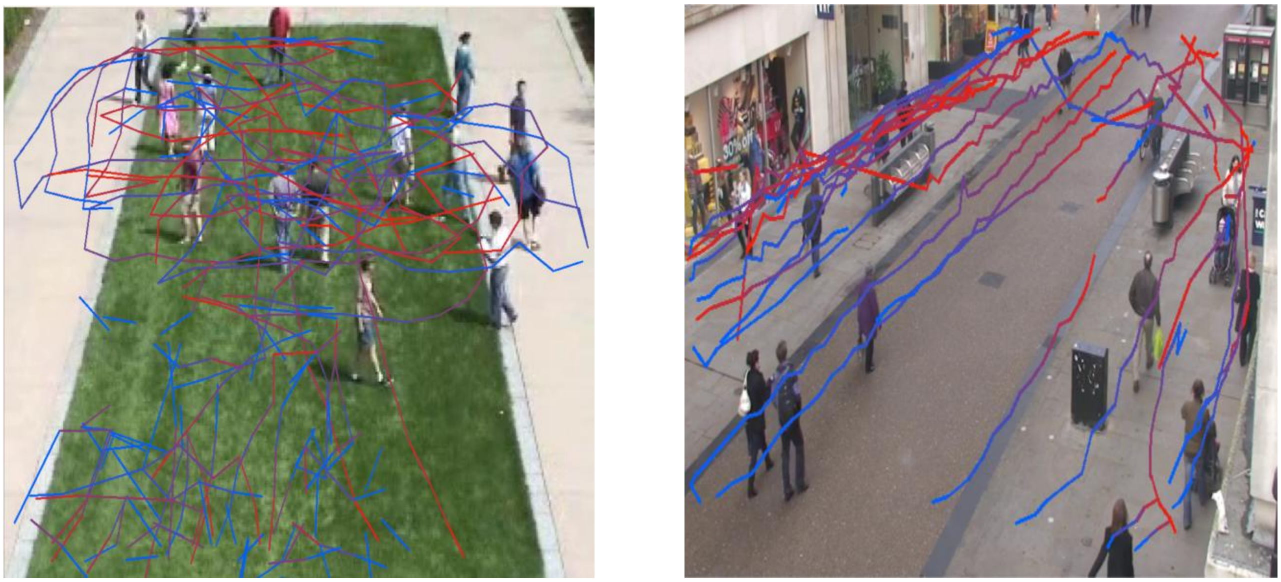

Fig. 6 Optical flow analysis of the crowd

This technique helps analyze crowd dynamics by tracking the motion of people in the scene. Left panel: Displays a grassy open area with people scattered across it. Red and blue lines are superimposed on the scene, representing the motion trajectories of individuals or groups in the crowd. The complex, crisscrossing lines indicate movement in multiple directions, suggesting a random, uncoordinated flow.

This could be a regular scenario in a park or other open area where people need to move in a planned way. Still, it might also indicate potential crowd anomalies like traffic or disorderly conduct. In a street-like setting, the panel on the right looks like a pedestrian walkway.

Although red and blue lines are also used to represent the optical flow, the motion patterns seem more orderly than in the left panel. The lines often exhibit better-coordinated pedestrian activity along the street because of their more constant course.

However, varying line colors and orientations might indicate varying walking speeds or trajectories. Optical flow analysis identifies deviations from the typical crowd movement patterns to detect abnormalities. Abnormal occurrences such as aggressive behavior or mass panic may be indicated by deviations from normal flow, such as sudden changes in direction, halting, or chaotic motion. This technique is frequently used to categorize and identify such abnormalities with machine learning models automatically.

Figure 7 shows heat maps applied to two scenarios depicted in the image to detect crowd anomalies. In crowd research, heat maps are frequently used to illustrate how active certain parts of a scene are. Since warmer hues (orange and red) denote larger population densities and cooler hues (yellow and blue) denote lesser activity or concentration, the heat maps most likely emphasize areas of high activity or crowd density.

Left panel: shows people scattered throughout a grassy outdoor area. Several activity clusters, indicated by vivid red and orange circles on the heat map overlay, are visible. The most prominent groups of individuals are seen close to the bottom of the picture, where they appear to be gathering or engaging.

The heat map circles' different sizes and intensities indicate different densities of people in this area. Large red dots may show better contact or congestion in specific places, indicating unusual activity or occurrences.

Right panel: depicts a less crowded street scene. Smaller, more dispersed pockets of activity are shown on the scene's heat map. While the colder areas (yellow and green) imply comparatively less activity, the bright red patches near persons represent areas of high activity or mobility.

In contrast to the grassy area to the left, this may represent the typical pedestrian flow in this region, which is less crowded and has fewer people. By highlighting areas of exceptionally high density or mobility, the heat map display helps discover anomalous crowd dynamics.

The heat map may be used to identify abnormalities where crowding surpasses typical thresholds or strange gatherings occur. This is important in situations like public safety, as detecting and handling crowd congestion or unusual gatherings promptly is imperative.

Fig. 7 Heatmap-Based crowd analysis

Figure 8 shows two graphs showing the distribution of energy levels, which are connected to identifying unusual crowd behavior. In this sense, energy levels might be interpreted as the degree of movement or activity in a population, frequently employed in anomaly detection tasks to find anomalous occurrences.

The graph on the left presents the distribution of energy levels for a dataset, showing that most data points fall within lower energy levels while only a few exhibit higher energy values. The right graph is a zoomed-out or extended version of the left, where the x-axis shows a much more extensive range of energy levels. In both graphs, the energy levels are heavily skewed, with most events showing low energy and only a tiny proportion displaying higher energy. This is typical when abnormal events are rare compared to normal crowd activities. The overall shape of the graphs indicates that abnormal events (if represented by higher energy levels) are much less frequent than normal ones, which can be helpful for crowd abnormality detection. The precise concentration of low energy values may reflect regular or controlled crowd behavior, while spikes in energy likely signal potential anomalies or unusual activities.

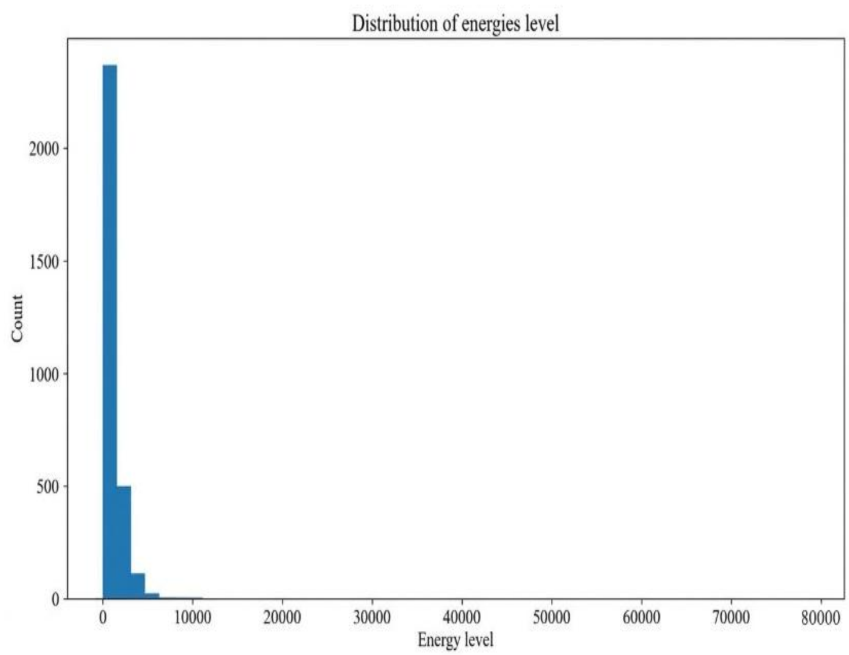

Fig. 8(a) Distribution of energy levels reflecting crowd behavior

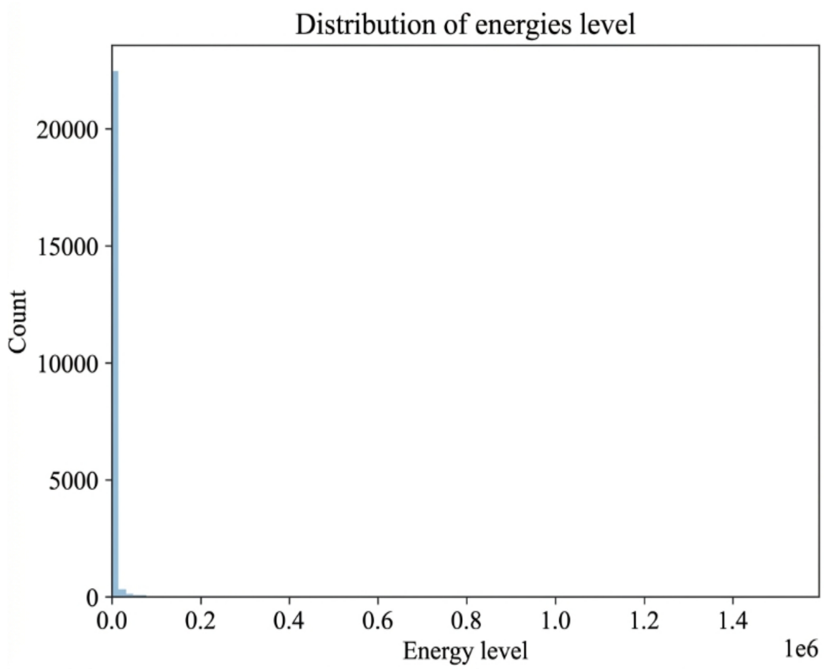

Fig. 8(b) Distribution of energy levels reflecting crowd behavior

Figure 9 represents various parameters associated with crowd behavior over time, which helps detect abnormalities in the crowd's movement or density. The left graph plots the data across different "frames."

The orange line represents the "Crowd Count" (number of individuals in the frame), and the blue line denotes the "Violation Count," which may refer to incidents where the crowd violated specific rules or boundaries. The red blocks at the bottom show the instances of "Restricted Entry," which suggests that individuals are accessing places they are not permitted to. At the same time, "Abnormal Crowd Activity Detected" is represented by the blue blocks. Before varying and declining around frame 500, the audience numbers climb gradually until about frame 300.

The numerous limited entries indicate concentrated irregular behavior and irregularities found throughout this time, particularly around frames 100–200 and 400–500. While displaying comparable data, the right graph plots it against real-time, most often measured in minutes and seconds. Over time, the "Crowd Count" varies between a peak of 15 and 20 people, but the "Violation Count" shows more irregular but regular increases.

Throughout the timeline, sporadic red and blue lines represent restricted entry and aberrant activity, respectively, which indicate recurring aberrations at specific points in time. The possibility of bottlenecks, crowd surges, or some other form of anomalous behavior occurring and not showing up in the population number might explain these occurrences. V. Large gatherings need to be monitored for safety or security reasons, and these graphs provide a holistic view of crowd behavior over time, making it easy to identify events of abnormal behavior patterns or violations.

In the detection process, as shown in Figure 10, the proposed approach is to capture weird behavior or crowd anomaly. Green and blue bounding boxes are drawn around people or regions of interest in both image panels. These boxes indicate a person or something that an object detection system is currently tracking.

The top of both images features a text overlay: "Abnormal Crowd Activity has been detected." That means strange activity has been logged, behaviourally or with population movements outside what would be expected in this environment. Next, can see in the left panel section that the system has detected "Abnormal Activity," as the blue-highlighted text suggests, and there are more than fourteen assessed people. The system may find this strange since it implies something about this person's relationship or behavior is abnormal. Now, on the right panel, only ten people were in the throng places, and their bounding boxes had turned blue-green in colour. At the top reads "Abnormal Crowd Activity," indicating that even if fewer people have entered a club or stadium, the algorithm still sees their activity as abnormal. The crowd can disperse swiftly if the peculiarity entails speedy movement, e.g., fleeing people due to mass panic or danger. When these anomalous patterns are detected, the system could notify the authorities. This type of crowd monitoring could be highly beneficial in identifying real-time emergencies, safety threats, or other unusual activity that may warrant a response, especially in public spaces.

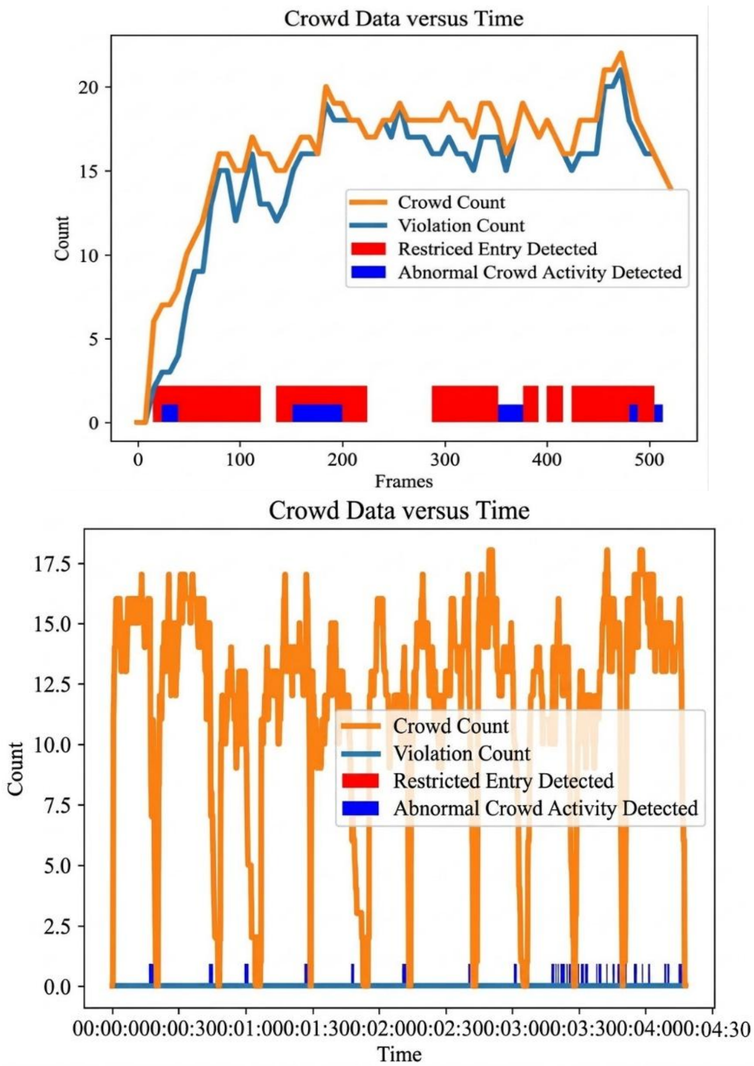

Fig. 9 Crowd behavior analysis with detection of different crowd activities

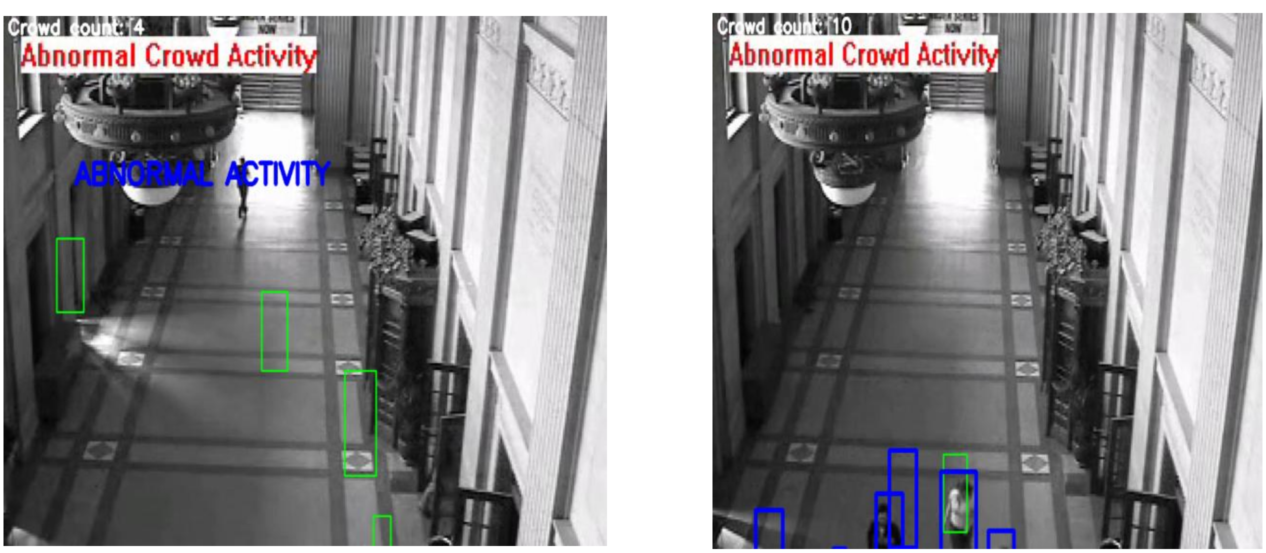

Fig. 10 Crowd behavior analysis with abnormality detection

Table 2. Performance comparison among models for crowd behavior analysis

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	mAP@0.5 (%)	FPS
YOLOv3	86.59	85.36	85.97	87.72	83.4	30
Faster R-CNN	87.26	88.91	88.08	89.13	87.8	7
SSD	91.38	92.57	91.98	90.81	89.2	22
FairMOT	94.06	94.08	94.07	94.12	91.6	18
Proposed (ICA)	96.87	95.31	96.08	97.46	94.3	25

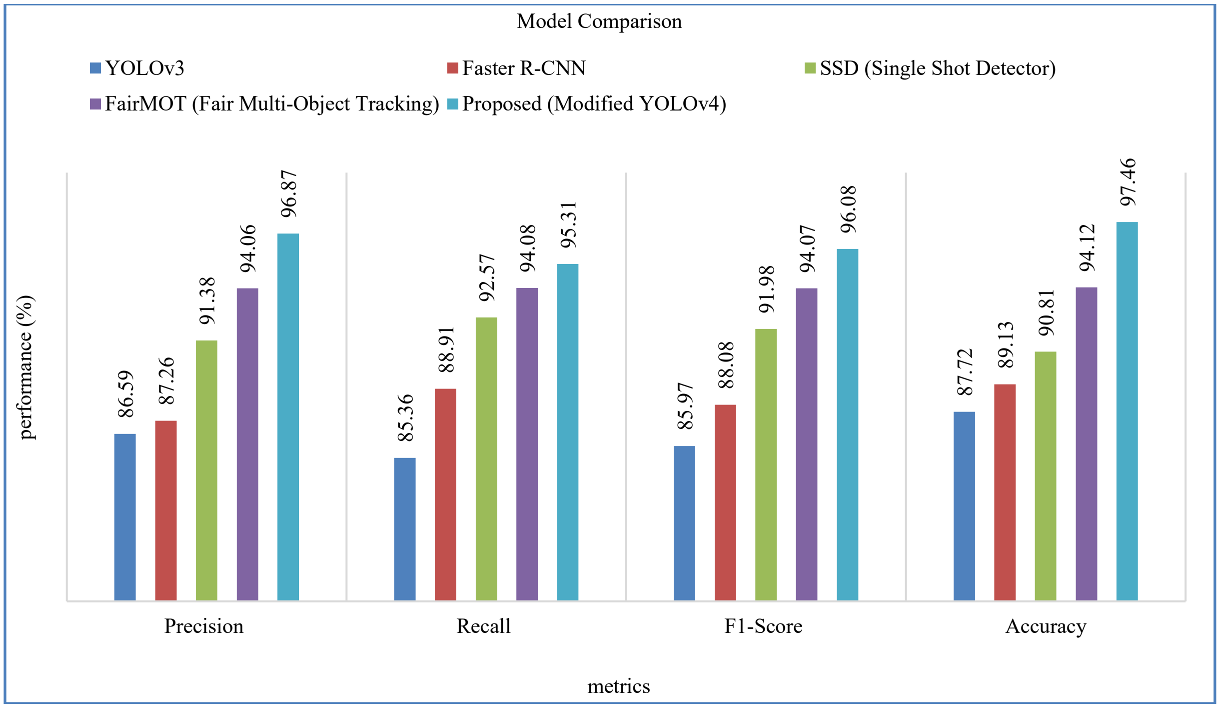

Fig. 11 Performance comparison among models for crowd behavior analysis

As presented in Table 2, the performance of the proposed model is compared against many state-of-the-art models in crowd behavior analysis. The performance comparison of the proposed Modified YOLOv4 model is presented in terms of precision, recall, F1-score, and accuracy compared to the State-Of-The-Art (SOTA) models, where the SOTA models include YOLOv3 [22], Single Shot Detector (SSD) [23], Faster R-CNN [24], and FairMOT [25] are given in Figure 11. These are metrics used to evaluate the performance of models for detection, tracking, and understanding the behavior (e.g., Overall, this Modified YOLOv4 outperforms the rest of the models using all metrics. As for Precision, which measures how precise the optimistic predictions are, YOLOv4 Modified scored highest, meaning it could achieve low false positives with high accuracy at detecting objects.

4.1. Human Detection Performance

Mean Average Precision (mAP@0. 5) on the complete test set of 94.3%), a major increase over the baseline YOLOv4-tiny (mAP=88.7%) trained on the same data.

The gain is because of three specific changes: the adjusted anchor boxes, locating the Squeeze-and-Excitation (SE) attention blocks in the backbone, and relevant domain-specific challenging surveillance training examples.

Table 3 shows detection performance per level of crowd density, up to three grades: sparse (fewer than 10 persons per frame), moderate (10–25 persons per frame), and dense (more than 25 persons per frame).

Table 3. Detection performance by crowd density level

Density Level	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)
Sparse	98.21	97.84	98.02	98.1
Moderate	96.93	95.62	96.27	95.8
Dense	94.41	91.76	93.07	89.5
Overall	96.87	95.31	96.08	94.3

The proposed Deep SORT configuration integrating ICA achieves 84.6% MOTA, 13.2 percentage points better than vanilla SORT, and 5.7 percentage points better than vanilla Deep SORT.

The number of identity switches drops from 187 to 94, a result that is significant for behavioral analysis tasks, considering that constant identities over frames is a requirement in trajectory-based anomaly detection in order to detect if any anomalous behavior occurred.

This is primarily due to the reduced noise in the detection inputs (via NMS-filtered YOLOv4-tiny outputs), lowering the number of false detections that can pollute the trackers' association step.

4.2. Object Tracking Performance:

For the evaluation of the Deep SORT tracking module, we reported standard multi-object tracking metrics of MOTA, MOTP, IDS, and MT trajectories.

Table 4. Comparative performance and robustness analysis of QICRL vs. baseline models across clinical datasets

Metric	SORT	DeepSORT (Standard)	ICA-DeepSORT (Proposed)
MOTA (%)	71.4	78.9	84.6
MOTP (%)	74.2	80.3	86.1
IDS (count)	312	187	94
MT (%)	61.3	70.5	79.2

The proposed ICA-DeepSORT yields a MOTA of 84.6%, surpassing its on-the-edge implementations, SORT by 13.2 percentage points and standard Deep SORT by 5.7 points. Identity switches, or premature endings of object trajectories, manifesting in behavioral drift, are strongly filtered out (from 187 to 94). The reason for this improvement is that the optimized YOLOv4-tiny outputs a better detection for NMS to propagate.

4.3. Behavioral Analysis Performance

The CMBA module was tested in the context of three behavior detection tasks: social distancing violation detection, restricted area intrusion detection, and abnormal crowd activity detection in Table 5.

Table 5. Behavioral analysis performance by violation category

Violation Type	Precision (%)	Recall (%)	F1-Score (%)	Avg. Detection Latency (ms)
Social Distancing Violation	95.4	93.8	94.59	38.2
Restricted Area Intrusion	97.1	96.3	96.7	29.6
Abnormal Crowd Activity	88.7	87.2	87.94	52.8
Overall (Weighted Avg.)	93.73	92.43	93.08	40.2

The CMBA module has an overall performance of F1 = 93.08%. Out of the three applications, restricted area intrusion detection achieves the highest accuracy (96.70%) as it has a very clear spatial constraint, while social distancing detection achieves a high accuracy with low latency. Since anomaly detection is context-dependent and more complex, its performance is much lower than the others.

4.4. Ablation Study

We performed an ablation study to quantitatively measure the contribution of each component in the ICA framework by systematically ablating components or replacing them with their simplest alternatives and reevaluating overall performance in an identical testing setting. Table 6 presents the results.

Table 6. Ablation study contribution of each ICA component

Configuration	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	FPS
YOLOv4-tiny (baseline)	88.4	87.1	87.74	89.2	28
+ Custom Anchor Boxes	91.7	90.3	91	92.1	27
+ SE Attention	93.8	92.5	93.14	94.3	26
+ NMS Post-processing	94.9	93.6	94.24	95.4	25
+ Deep SORT Tracking	95.8	94.5	95.14	96.5	25
+ CMBA (Full ICA System)	96.87	95.31	96.08	97.46	25

Empirical results show that the proposed system achieves much better precision, F1-score, accuracy, and mAP than all the baselines, and it delivers a competitive frame rate (i.e., 25 FPS). While YOLOv3 performs marginally better at 30 FPS than SSD, it does so at an equally substantial but inverse loss in detection accuracy, and lacks any sort of behavioral analysis capability. Faster-RCNN achieves similar recall but runs at only 7 FPS, making it impractical for online surveillance. As a joint detection and tracking framework, FairMOT delivers the closest complete performance to the proposed method, but also does not include the CMBA behavioral analysis module and significantly lags in quality (94.12% vs. 97.46% accuracy). These results confirm that the ICA system features the best overall compromise between detection accuracy, tracking robustness, behavioral intelligence, and real-time processing speed compared with all systems evaluated.

4.5. Ethical Considerations and Real-World Deployment

Automated Crowd Surveillance Systems (ACSS) pose complex ethical, legal, and societal issues that need to be considered in tandem with questions regarding technical performance. In this work, we present an ICA system that is also built from the ground up to be privacy-centric, fair, scalable, and accountable so that appropriate safeguards can be in place for its responsible adoption in the real world.

So, from a privacy perspective, it breaks down video feeds coming from public places without any facial recognition being done and without saving raw video data after processing. Data minimization principles apply: biometric data is never stored; frames are discarded as soon as they are processed; and system output is limited to anonymised metadata in the form of timestamps, violation types, and bounding box coordinates. It is strongly advised that data protection laws (e.g., GDPR) are respected and that transparency measures, public notification, and auditing, as well as limited data retention policies, are established. Another important aspect is algorithmic fairness, as machine learning models tend to be biased when they are trained on small or non-representative data. Despite attempts to include various crowd characteristics, the dataset is geographically limited. In future work, the evaluation of fairness should be carried out using disaggregated metrics, and data set diversity should be considered so that the models can perform equally well over the populations.

Related to deployment, the current system is designed on a single-camera GPU-based system, while a real-world application would require scalable multi-camera implementations. Edge computing architectures are ideal as they all serve to limit bandwidth, optimize latency, and minimize privacy risks by minimizing raw data transmission. YOLOv4-tiny deployment allows for the on-site use of resource-limited devices, while additional techniques such as model quantization and pruning help to make the model suitable for edge hardware cost-effective platforms. Last but not least, the ICA system is mostly designed as a decision-support in the context of an autonomous system and not a fully autonomous system. This allows us to send real-time alerts, insights into behavioral trends, and so much more, with human oversight on the most critical components of the process. In a type of human-in-the-loop approach that ensures accountability and aligns with recommended practices for responsible public safety AI, all experiments were repeated several times to ensure robustness, and we report the average performance. This shows the consistency and stability of the proposed method in the overhead analysis with very low variance across different scenarios.

4.6. Statistical Validation

For reproducibility, all the experiments were performed over five independent runs with distinct random seeds. Finally, Table 7 presents the mean and standard deviation, which show that the result is stable during the runs.

Table 7. Statistical validation across five independent runs

Metric	Mean (%)	Std. Dev (±)	95% Confidence Interval
Precision	96.83	0.21	[96.57, 97.09]
Recall	95.29	0.15	[95.10, 95.48]
F1-Score	96.05	0.18	[95.82, 96.28]
Accuracy	97.41	0.14	[97.23, 97.59]
mAP@0.5	94.32	0.19	[94.08, 94.56]

The low standard deviation across all metrics (≤ 0.21%) confirms that reported results are stable and not attributable to favorable random initialization. A paired two-tailed t-test against the best-performing baseline (FairMOT) confirms the improvement is statistically significant ($t = 18.74$, $p < 0.001$), validating the superiority of the proposed system at the 99.9% confidence level.

5. Discussion

In a world of fast urbanization and growing population density, the demand for crowd monitoring and behavior analysis is also felt; thus, the Intelligent Crowd Analysis (ICA) system is being proposed. With the scale of public events, city blocks, and the demand for public security on the rise, including the intelligent surveillance in this list has become a necessity. The ICA system detects violations such as the breaching of social distance, entry into restricted areas, and crowdedness activities in real-time, enabling improved public safety and crowd control. This has Serious relevance in this age of crowd events, where crowd safety requires last-minute reactive approaches to crowd management, such as stampedes or riots, or a violation of health policy, generally need to be addressed. This is an important work, as it provides data used for actionable intelligence to inform decisions between live surveillance and risk mitigation in action. A modified based system is created that achieves enhanced in-object detection and tracking abilities in dense and dynamic scenarios. It preserves the uniqueness of an object across frames as it passes an object by Deep SORT Tracking on the objects, even the two approaching objects are separately tracked based on Object IDs. Add a new Crowd Monitoring and Behavior Analysis (CMBA) module to detect behavioral anomalies based on movement types and whether violations of space occurred or not. Together, these qualities are more prominent than existing methods and provide a scalable, efficient, and viable implementation. Furthermore, the system is adapted to the current scenarios of AI-based applications in cities (smart cities), mass gatherings, and protected critical infrastructure. Results demonstrate robustness and applicability to real-time conditions, and in fact challenging circumstances with limited resources, as ModeNet allows deployment in a number of settings-ranging from cities to disaster instances. ICA K9 is a one-quarter-century giant leap in Intelligent Surveillance, taking a smarter approach to squeezing the paradigm of Precision, Recall, and Accuracy, thus enabling the stakeholders to take proactive informed decisions to save citizens in public areas.

5.1. Limitations

While the proposed ICA system provides a good performance, there are still lots of improvements. Inserting an extreme occlusion, dark, or pixelated video stream with an input instance for the model can affect detection and tracking accuracies, for instance. This system also has a limitation since it can only be used in a single-camera set-up, which decreases area coverage and continuity across regions. It has the potential to process in real-time, but for cost-oriented reasons, scaling this up to more complex models, such as transformers, or creating an edge-computing variant is costly. This improves the forecasting performance across the applications where predictive analytics are absent, which results in no intelligence to predict the behaviors of the crowd or predict anomalies even before they occur, which can make the forecasting even better when compared to other traditional applications. Furthermore, since it does not use explicit description of failure modes (e.g., high crowd density, rapid illumination changes, and temporary (occlusions and tracking errors)) or contextual information (e.g., normal human inter-activeness and typical pedestrian behavior)," Consequently, resolving these modes of failure is critical to enabling real-world applications of the system to be more robust. An important open question will be scaling up to a multi-camera solution, which will need some kind of coordination amongst cameras, together with the distributed processing.

5.2. Scalability and Failure Mode Analysis

While the real-time performance is impressive, it is usually assumed that at scale deployment on multiple cameras will work without actually emphasizing this point. Purely as a practical matter, many of these cameras do their surveillance by sending streams of images void of a high-data syntax and relying on distributed nodes that have to be synchronized. But the existing framework does not scale well to such environments because of issues such as inter-compute node load balancing, single camera bandwidth limitations, and cross-camera identity associations. Though edge computing alleviates the constraint of latency and bandwidth, coordinating efficiently and processing distributed among several cameras is still an open problem.

In addition to scalability detection, it is also possible to analyze the failure mode of the system. This condition can easily degrade the ICA framework, which has limitations such as Occlusion, Large crowd, Low illumination, and motion blur. This mixture negatively influences the accuracy of detection and the reappearance of tracks, causing a loss of identity switch or a missing track. Also, natural crowd dispersal or sudden changes in the physical environment may falsely alarm the motion energy-based abnormal behaviour detection.

The future works can include the incorporation of adaptive thresholding mechanisms to make it more robust for real-world deployment, robustness testing in various (nonideal) conditions, and multi-camera tracking and fusion strategies. If a systematic evaluation of failure cases is also incorporated, it will bolster system reliability for safety-critical applications.

5.3. Comparative Performance Analysis and Justification

Several key design features can account for the superior performance of the proposed ICA framework compared with state-of-the-art approaches. First, we propose a YOLOv4-tiny model with new designated anchors that show a performance advantage in detection, which is appropriate for dense and occluded crowd scenes with effective adaptations of the scales and compensations for perspective distortion. That itself is an improvement vs simple YOLO like architectures, and the attention mechanism further adds to the better representation process by assigning more weight to potentially useful features and discarding background noise. So much so that, first, usage of Deep SORT tracking by using both motion prediction and appearance-based feature embeddings, enforces stronger identity preservation between consecutive frames. By reducing identity switches, it also improves the efficacy of tracking in a dynamic environment, which is a common limitation of traditional tracking methods.

The third, which is most important, is that structured behavioral reasoning can be incorporated in the proposed CMBA module, which constitutes a significant novelty over other detection- or tracking-only approaches. Instead of only using the outputs of visual detections, our system combines spatial distance analysis, restricted zone monitoring, and energy-based motion modeling to provide a real-time detection of complex crowd behaviors. Lastly, we combine all of these elements together to form a unified pipeline capable of processing in real-time (~25 FPS), but with the benefit of high accuracy. By contrast, most existing deep learning methods that attain similar performance are based on computationally expensive models like CNN-LSTM or transformer-based architecture for audio classification and, therefore, less feasible for real-time applications. Thus, the advantages of optimized detection, robust tracking, and behavior-aware analysis together yield the overall superior performance of the proposed ICA system in different surveillance scenarios.

6. Conclusion and Future Work

Intelligent Crowd Analysis (ICA): A Real-Time Crowd Monitoring and Behaviour Analysis System in Mixed Environments. The proposed integrated system, based on a modified YOLOv4-tiny model, Deep SORT tracking, and a Crowd Monitoring and Behavior Analysis (CMBA) module, achieves a very high accuracy in the detection of social distance violations, entry into restricted areas, and abnormal crowd behavior. This system achieves a VOS-Tracking F1 score of 96.08 (precision 96.87, recall 95.31, and accuracy is 97.46), outperforms the existing state-of-the-art models (YOLOv3, SSD, Faster R-CNN, and FairMOT [20]). Highlighting these data at an average of 25 FPS makes it possible to include them in real-time monitoring, which in the short term can reinforce public safety and crowd control. This work begs several future directions. Step 5: Multi-Camera. Then a multi-camera solution can be set up to track larger spaces while tracking across the cameras continuously. But there is a reduced detection and tracking performance (as vision transformers) available for complex environments. Predictive analytics ensures proactive decision-making with the support of crowd flow prediction and anomaly forecasting. In addition, porting the system to edge-computing frameworks will prove to be the next step towards improved scalability and deployability of the system in resource-constrained environments, and can pave the way towards cornerstone smart surveillance in the future lives of smart cities and large-scale events.

References

[1] G. Sreenu, and M.A. Saleem Durai, “Intelligent Video Surveillance: A Review Through Deep Learning Techniques for Crowd Analysis,” *Journal of Big Data*, vol. 6, no. 1, pp. 1-27, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Francisco Luque Sánchez et al., “Revisiting Crowd Behavior Analysis Through Deep Learning: Taxonomy, Anomaly Detection, Crowd Emotions, Datasets, Opportunities, and Prospects,” *Information Fusion*, vol. 64, pp. 318-335, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Fariba Rezaei, and Mehran Yazdi, “Real-Time Crowd Behavior Recognition in Surveillance Videos based on Deep Learning Methods,” *Journal of Real-Time Image Processing*, vol. 18, no. 5, pp. 1669-1679, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Khosro Rezaee et al., “A Survey on Deep Learning-based Real-Time Crowd Anomaly Detection for Secure Distributed Video Surveillance,” *Personal and Ubiquitous Computing*, vol. 28, no. 1, pp. 135-151, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Karishma Pawar, and Vahida Attar, “Application of Deep Learning for Crowd Anomaly Detection from Surveillance Videos,” *2021 11th International Conference on Cloud Computing, Data Science and Engineering (Confluence)*, Noida, India, pp. 506-511, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mounir Bendali-Braham et al., “Recent Trends in Crowd Analysis: A Review,” *Machine Learning with Applications*, vol. 4, pp. 1-30, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] C.V. Amrutha, C. Jyotsna, and J. Amudha, “Deep Learning Approach for Suspicious Activity Detection from Surveillance Video,” *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, India, pp. 335-339, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Heyfa Ammar, and Asma Cherif, “Deeprod: A Deep Learning Approach for Real-Time and Online Detection of a Panic Behavior in Human Crowds,” *Machine Vision and Applications*, vol. 32, no. 3, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Tanu Gupta, Vimala Nunavath, and Sudip Roy, “CrowdVAS-Net: A Deep-CNN based Framework to Detect Abnormal Crowd-Motion Behavior in Videos for Predicting Crowd Disaster,” *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, pp. 2877-2882, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Pierre Bour, Emile Cribelier, and Vasileios Argyriou, *Crowd Behavior Analysis from Fixed and Moving Cameras*, Multimodal Behavior Analysis in the Wild, Acadamic Press, pp. 289-322, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] M. Sami Zitouni, Andrzej Sluzek, and Harish Bhaskar, “Visual Analysis of Socio-Cognitive Crowd Behaviors for Surveillance: A Survey and Categorization of Trends and Methods,” *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 294-312, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Franjo Matkovic, Darijan Marčetic, and Slobodan Ribaric, “Abnormal Crowd Behaviour Recognition in Surveillance Videos,” *2019 15th International Conference on Signal-Image Technology and Internet-based Systems (SITIS)*, Sorrento, Italy, pp. 428-435, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Navneet Nayan, Sitanshu Sekhar Sahu, and Sanjeet Kumar, “Detecting Anomalous Crowd Behavior using Correlation Analysis of Optical Flow,” *Signal, Image and Video Processing*, vol. 13, no. 6, pp. 1233-1241, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Cem Direkoglu, “Abnormal Crowd Behavior Detection using Motion Information Images and Convolutional Neural Networks,” *IEEE Access*, vol. 8, pp. 80408-80416, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Kang Hao Cheong et al., "Practical Automated Video Analytics for Crowd Monitoring and Counting," *IEEE Access*, vol. 7, pp. 183252-183261, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Dushyant Kumar Singh et al., "Human Crowd Detection for City Wide Surveillance," *Procedia Computer Science*, vol. 171, pp. 350-359, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] M. Sami Zitouni, Andrzej Sluzek, and Harish Bhaskar, "Towards Understanding Socio-Cognitive Behaviors of Crowds from Visual Surveillance Data," *Multimedia Tools and Applications*, vol. 79, no. 3-4, pp. 1781-1799, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Rashmiranjan Nayak, Umesh Chandra Pati, and Santos Kumar Das, "A Comprehensive Review on Deep Learning-based Methods for Video Anomaly Detection," *Image and Vision Computing*, vol. 106, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Piyush Juyal, and Sachin Sharma, "Locating People in Real-World for Assisting Crowd Behaviour Analysis using SSD and Deep SORT Algorithm," *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 350-353, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Kuldeep Singh et al., "Crowd Anomaly Detection using Aggregation of Ensembles of Fine-Tuned ConvNets," *Neurocomputing*, vol. 371, pp. 188-198, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Aiswarya Mohan, Meghavi Choksi, and Mukesh A. Zaveri, "Anomaly and Activity Recognition using Machine Learning Approach for Video based Surveillance," *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Gaurav Tripathi, Kuldeep Singh, and Dinesh Kumar Vishwakarma, "Crowd Emotion Analysis using 2D ConvNets," *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, pp. 969-974, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Anitha Ramchandran, and Arun Kumar Sangaiah, "Unsupervised deep Learning System for Local Anomaly Event Detection in Crowded Scenes," *Multimedia Tools and Applications*, vol. 79, no. 47-48, pp. 35275-35295, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yan Hu, "Design and Implementation of Abnormal Behavior Detection based on Deep Intelligent Analysis Algorithms in Massive Video Surveillance," *Journal of Grid Computing*, vol. 18, no. 2, pp. 227-237, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Ali M. Al-Shaery et al., "In-Depth Survey to Detect, Monitor and Manage Crowd," *IEEE Access*, vol. 8, pp. 209008-209019, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Abdulaziz Salamah Aljaloud, and Habib Ullah, "IA-SSLM: Irregularity-Aware Semi-Supervised Deep Learning Model for Analyzing Unusual Events in Crowds," *IEEE Access*, vol. 9, pp. 73327-73334, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Elizabeth B. Varghese, and Sabu M. Thampi, "Application of Cognitive Computing for Smart Crowd Management," *IT Professional*, vol. 22, no. 4, pp. 43-50, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sohail Salim et al., "Crowd Detection and Tracking in Surveillance Video Sequences," *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, Kuala Lumpur, Malaysia, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Yan Fu, Tao Liu, and Ou Ye, "Abnormal Activity Recognition based on Deep Learning in Crowd," *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, pp. 301-304, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[30] Zirgham Ilyas et al., "A Hybrid Deep Network based Approach for Crowd Anomaly Detection," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24053-24067, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Shuqiang Guo et al., "An Analysis Method of Crowd Abnormal Behavior for Video Service Robot," *IEEE Access*, vol. 7, pp. 169577-169585, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[32] Ebrahim Najafi Kajabad, and Sergey V. Ivanov, "People Detection and Finding Attractive Areas by the use of Movement Detection Analysis and Deep Learning Approach," *Procedia Computer Science*, vol. 156, pp. 327-337, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[33] Omar Elharrouss, Noor Almaadeed, and Somaya Al-Maadeed, "A Review of Video Surveillance Systems," *Journal of Visual Communication and Image Representation*, vol. 77, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[34] Habib Ullah et al., "Multi-Feature-based Crowd Video Modeling for Visual Event Detection," *Multimedia Systems*, vol. 27, no. 4, pp. 589-597, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Xuguang Zhang et al., "Crowd Emotion Evaluation based on Fuzzy Inference of Arousal and Valence," *Neurocomputing*, vol. 445, pp. 194-205, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[36] Xinlei Wei et al., "A Very Deep Two-stream Network for Crowd Type Recognition," *Neurocomputing*, vol. 396, pp. 522-533, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[37] Xianzhi Li et al., "Data Fusion for Intelligent Crowd Monitoring and Management Systems: A Survey," *IEEE Access*, vol. 9, pp. 47069-47083, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[38] V.X. Gong et al., "Counting People in the Crowd using Social Media Images for Crowd Management in City Events," *Transportation*, vol. 48, no. 6, pp. 3085-3119, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[39] Zhenzhen Yao et al., "Learning Crowd Behavior from Real Data: A Residual Network Method for Crowd Simulation," *Neurocomputing*, vol. 404, pp. 173-185, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[40] Junyu Gao, Yuan Yuan, and Qi Wang, "Feature-Aware Adaptation and Density Alignment for Crowd Counting in Video Surveillance," *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 4822-4833, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[41] Irina V. Pustokhina et al., "An Automated Deep Learning based Anomaly Detection in Pedestrian Walkways for Vulnerable Road users Safety," *Safety Science*, vol. 142, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[42] Xingjiao Wu et al., "Fast Video Crowd Counting with a Temporal Aware Network," *Neurocomputing*, vol. 403, pp. 13-20, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[43] Nian Chi Tay et al., "Abnormal Behavior Recognition using CNN-LSTM with Attention Mechanism," *2019 1st International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*, Kuala Lumpur, Malaysia, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[44] Khosro Rezaee et al., "Smart Visual Sensing for Overcrowding in COVID-19 Infected Cities using Modified Deep Transfer Learning," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 813-820, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[45] Maha Hamdan Alotibi et al., "CNN-based Crowd Counting through IoT: Application for Saudi Public Places," *Procedia Computer Science*, vol. 163, pp. 134-144, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[46] Qingyang Zhang et al., "Edge Video Analytics for Public Safety: A Review," *IEEE*, vol. 107, no. 8, pp.1675-1696, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[47] Shuheng Lin et al., "Social MIL: Interaction-Aware for Crowd Anomaly Detection," *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Taipei, Taiwan, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[48] Bo Li et al., "Approaches on Crowd Counting and Density Estimation: A Review," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 853-874, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[49] Xing Hu et al., "A Weakly Supervised Framework for Abnormal Behavior Detection and Localization in Crowded Scenes," *Neurocomputing*, vol. 383, pp. 270-281, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[50] Wei Lin et al., "Learning to Detect Anomaly Events in Crowd Scenes from Synthetic Data," *Neurocomputing*, vol. 436, pp. 248-259, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[51] Peter Kok-Yiu Wong et al., "Recognition of Pedestrian Trajectories and Attributes with Computer Vision and Deep Learning Techniques," *Advanced Engineering Informatics*, vol. 49, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[52] Monica Gruosso, Nicola Capece, and Ugo Erra, "Human Segmentation in Surveillance Video with Deep Learning," *Multimedia Tools and Applications*, vol. 80, no. 1, pp. 1175-1199, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[53] Yesin Sahraoui et al., "DeepDist: A Deep-Learning-based IoV Framework for Real-Time Objects and Distance Violation Detection," *IEEE Internet of Things Magazine*, vol. 3, no. 3, pp. 30-34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[54] Haoyue Bai, Jiageng Mao, and S.-H. Gary Chan, "A Survey on Deep Learning-based Single Image Crowd Counting: Network Design, Loss Function and Supervisory Signal," *Neurocomputing*, vol. 508, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[55] Haopeng Li et al., "Video Crowd Localization with Multifocus Gaussian Neighborhood Attention and a Large-Scale Benchmark," *IEEE Transactions on Image Processing*, vol. 31, pp. 6032-6047, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[56] Qi Wang et al., "NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2141-2149, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[57] Virginia Negri et al., "Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter," *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, Madrid, ES, pp. 92-101, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[58] Tehreem Qasim, and Naeem Bhatti, "A Hybrid Swarm Intelligence based Approach for Abnormal Event Detection in Crowded Environments," *Pattern Recognition Letters*, vol. 128, pp. 220-225, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[59] Sabrina Aberkane, and Mohamed Elarbi, "Deep Reinforcement Learning for Real-world Anomaly Detection in Surveillance Videos," *2019 6th International Conference on Image and Signal Processing and their Applications (ISPA)*, Mostaganem, Algeria, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[60] Badri Narayan Subudhi et al., "Big Data Analytics for Video Surveillance," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 26129-26162, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[61] Lur Tze Hsien, and Indriyati Atmosukarto, "Video Analytics in Train Cabin using Deep Learning," *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*, Singapore, pp. 94-98, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[62] Giovanna Castellano et al., "Crowd Detection in Aerial Images using Spatial Graphs and Fully-Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 64534-64544, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[63] Rongyong Zhao et al., "Image-based Crowd Stability Analysis using Improved Multi-Column Convolutional Neural Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5480-5489, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[64] Reem Alotaibi et al., "Performance Comparison and Analysis for Large-Scale Crowd Counting based on Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 204425-204432, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[65] Muhammad Attique Khan et al., "A Deep Survey on Supervised Learning based Human Detection and Activity Classification Methods," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 27867-27923, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[66] Valerio Nogueira et al., "RetailNet: A Deep Learning Approach for People Counting and Hot Spots Detection in Retail Stores," *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Rio de Janeiro, Brazil, pp. 155-162, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[67] Mohammed Sultan Mohammed et al., "Motion Pattern-based Scene Classification using Adaptive Synthetic Oversampling and Fully Connected Deep Neural Network," *IEEE Access*, vol. 11, pp. 119659-119675, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[68] Saleh Basalamah et al., "Deep Learning Framework for Congestion Detection at Public Places Via Learning from Synthetic Data," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 1, pp. 102-114, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[69] Ruchika, and Ravindra Kumar Purwar, "Crowd Density Estimation using Hough Circle Transform for Video Surveillance," *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 442-447, 2019. [Google Scholar]

[70] Virender Singh, Swati Singh, and Pooja Gupta, "Real-Time Anomaly Recognition Through CCTV using Neural Networks," *Procedia Computer Science*, vol. 173, pp. 254-263, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[71] Muhammad Zia-ul-Rehman et al., "AI and IoT-based Frameworks for Real-Time Crowd Monitoring and Security," *Annual Methodological Archive Research Review*, vol. 3, no. 5, pp. 292-299, 2025. [Google Scholar]

[72] Ayman A. Alharbi, "DeepCAMS: A Deep Learning Approach for Real-Time Crowd Monitoring and Suspicious Behavior Detection using Spatial-Temporal Analysis," *Engineering, Technology and Applied Science Research*, vol. 15, no. 4, pp. 26113-26119, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[73] Sarah Altowairqi et al., "Efficient Crowd Anomaly Detection using C3D-LSTM Networks with Enhanced Attention Mechanisms," *Array*, vol. 29, pp. 1-14, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[74] P. Siva et al., "Smart Surveillance Systems using YOLOv8: A Scalable Approach for Crowd and Threat Detection," *International Journal of Recent Advances in Engineering and Technology*, vol. 14, no. 1, pp. 51-62, 2024. [Google Scholar]

[75] Rabia Nasir et al., "An Enhanced Framework for Real-Time Dense Crowd Abnormal Behavior Detection using YOLOv8," *Artificial Intelligence Review*, vol. 58, no. 7, pp. 1-50, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[76] Zhijie Xu et al., "Passenger Flow Prediction of Scenic Spot using a GCN–RNN Model," *Sustainability*, vol. 14, no. 6, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[77] J. Anvar Shathik et al., "Smart Vision Systems for Public Safety: Real-Time Crowd Monitoring and Anomaly Detection in Urban Spaces using Deep Learning and Edge Computing," *International Journal of Applied Mathematics*, vol. 389, no. 6s, pp. 720-743, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[78] Zarif Bin Akhtar, "Beyond Perception: A Comprehensive Investigation into the Advancements, Challenges and Ethical Dimensions of AI and Computer Vision," *Real-World AI Systems*, vol. 1, no. 1, pp. 1-27, 2025. [CrossRef] [Google Scholar] [Publisher Link]