International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Crowd Density Mapping and Anomaly Detection using YOLOv8 and DEEPSORT

Dr. Deepali Ujlambkar[1*], Gaurav Shirke[2], Suraj Pawar[3], Tejas Pawar[4], and Raviraj Shingate[5]

[1]Assistant Professor, Department of Computer Engineering, AISSMS College of Engineering, Pune, India.
[2,3,4,5] Student, Department of Computer Engineering, AISSMS College of Engineering, Pune, India.
{gauravshirke895, surajpawar0216, tejasanandapawar2003, ravirajshingate173}@gmail.com

Abstract:
Effective crowd management in high-density public spaces remains a critical challenge, especially during large-scale events or emergencies. This study introduces a real-time system that integrates crowd density estimation and behavioral anomaly detection using deep learning and video surveillance. The proposed framework leverages YOLOv8 for high-precision person detection and DeepSORT for continuous multi-object tracking. Anomalies—such as panic movements, physical altercations, and prolonged immobility—are identified using motion trajectory analysis and rule-based behavioral classification. Crowd density is categorized into low, medium, or high based on per-frame detections, and visualized through a dynamic Google Maps interface. The system also issues immediate alerts for critical events using SMTP-based notifications. Evaluated on benchmark datasets including ShanghaiTech, UCF-QNRF, and a custom-labeled Roboflow dataset, the model achieves a detection precision of 86.2% and supports near real-time processing with optimized latency on both CPU and GPU platforms. This approach provides scalable, location-aware crowd analytics, making it highly applicable for smart surveillance, urban safety, and emergency response systems.

Keywords: Crowd density estimation, Anomaly detection, YOLOv8, LSTM, CNN (Convolutional Neural Network), Real-time monitoring, Crowd detection, Interactive heatmaps, Public safety, CCTV surveillance.

Introduction:

Crowd density mapping and anomaly detection play a vital role in modern crowd management and public safety. They have become increasingly important across various settings such as large events, transit hubs, public areas, and in city planning. Being able to accurately assess how densely people are packed in an area and identify unusual or risky behavior can significantly enhance safety protocols, emergency response times, and overall management strategies. As urban areas grow and gatherings become more frequent and larger, the challenge of monitoring and managing crowds in real time becomes more complex. However, advancements in computer vision, machine learning, and deep learning have transformed how we approach these challenges—making it possible to automate and streamline both density estimation and anomaly detection with greater speed and accuracy.

Understanding how people are distributed within a space is key to identifying potential overcrowding, which could lead to safety hazards. Traditionally, estimating crowd density involved manual observation or basic image processing techniques, which were often too slow or unreliable in dense and dynamic environments. Today, more advanced approaches—especially those using deep learning—have greatly improved the accuracy and efficiency

of this task. Convolutional Neural Networks (CNNs), for instance, have shown great promise in interpreting surveillance footage to estimate how crowded an area is. These models have proven effective even in complex scenes filled with people. Research, such as that by [1], has highlighted how CNN-based systems can handle high-density environments and still provide reliable estimates of crowd size and distribution.

In addition to density mapping, the ability to detect unusual or abnormal behavior in crowds has become a major focus. Whether it's sudden movements, altercations, or other irregular activities, identifying anomalies early is crucial to preventing dangerous situations like stampedes or security threats. Older methods based on statistical rules or hand-crafted features often fell short when dealing with the unpredictable nature of large crowds. In contrast, newer deep learning techniques—like Recurrent Neural Networks (RNNs) and autoencoders—can learn patterns of normal crowd behavior and flag anything that deviates from those patterns. One study, [2], introduced an autoencoder-based system that learns to recreate expected crowd flow and detects anomalies by measuring how far the reconstruction strays from the actual input.

By combining both crowd density mapping and anomaly detection, a more complete and powerful system can be developed for crowd oversight. Systems that track crowd levels while also watching for irregularities can deliver deeper, real-time insights and help prevent problems before they arise. This dual-function approach is especially critical in high-risk venues such as stadiums, airports, and concerts, where safety is paramount. For example, [3] presented a unified system that used CNNs and LSTMs to simultaneously estimate crowd density and detect anomalies, offering live updates in the form of heatmaps and alerts about suspicious activity. Other researchers, like those in [4] and [5], have explored unsupervised learning approaches that don't rely on labeled data, making it easier to deploy such systems in real-world scenarios where annotations are difficult to gather. These models use clustering and generative techniques to learn typical patterns and then highlight anything that doesn't fit.

Another recent improvement in this field is the use of attention mechanisms, which help models focus on the most important areas in a scene. This is particularly helpful in crowded settings where subtle movements might otherwise go unnoticed. The study by [6], for instance, showed how attention-based networks could improve detection by emphasizing key areas of interest, leading to better results and greater interpretability. For real-time use—such as in airports or during live events—speed is just as important as accuracy. Systems need to process data quickly without compromising performance. Researchers, including those cited in [7], have looked at ways to make deep learning models run faster while still being accurate, enabling real-time responses from security teams or organizers if something seems off.

In summary, the fusion of crowd density analysis and anomaly detection holds great potential for boosting public safety, refining crowd management strategies, and supporting the development of smarter, more responsive cities. As technology evolves, we can expect even more capable systems that offer real-time insights and proactive alerts in dynamic, crowded environments. Deep learning and computer vision will remain at the heart of this progress, pushing forward the capabilities of modern crowd monitoring solutions.

Literature Survey:

Sayantan Roy et al. [1] presented a comprehensive survey focused on crowd anomaly detection using deep learning models to enhance public safety in complex urban scenarios. The paper thoroughly analyzes both traditional techniques like SVM, KNN, and HMM, and modern deep learning models such as CNN, RNN, LSTM, and GAN. It discusses how each approach is suited to address specific aspects of crowd behavior, from motion tracking to behavior classification. The study emphasizes the role of benchmark datasets such as UCSD, Avenue, PETS2009,

and ShanghaiTech in training and evaluating these models. Hybrid models like CNN-LSTM and Autoencoders are highlighted for their superior accuracy in complex environments. Real-time anomaly detection, occlusion handling, and variability in environmental conditions remain significant challenges. The paper calls for integrating multi-modal inputs such as audio and IoT data to improve system robustness. It also emphasizes the need for explainable AI to enhance interpretability and trust in automated systems. Application areas include disaster prevention, security at public gatherings, and smart city surveillance.

Suraj Shukla et al. [2] proposed the use of deep learning-based smart surveillance systems to detect abnormal behavior in crowd settings. The study focuses on CNNs as the core architecture for analyzing video footage in real-time to improve the accuracy of anomaly detection. The paper outlines the limitations of manual CCTV monitoring, citing human error and the inability to interpret large-scale, dynamic crowd behavior. Various dimensions of crowd analysis are reviewed, including motion estimation, density evaluation, and behavioral prediction. A taxonomy of traditional versus deep learning methods is presented, demonstrating the transition from manual feature extraction to automated learning systems. The authors emphasize automation as essential for scalable and efficient monitoring in public areas like airports, malls, and religious gatherings. The study covers major challenges such as occlusion, noisy data, and variable lighting conditions that impact detection performance. Several algorithms and techniques, including LK Optical Flow, GBM, and social force models, are discussed in the context of behavior segmentation. It stresses the need for systems capable of interpreting real-time behavioral shifts to prevent stampedes and terrorist attacks.

Dharmesh Tank et al. [3] conducted a comprehensive review of recent developments in deep learning for crowd anomaly detection, particularly post-2019 innovations. The paper categorizes crowd analysis into statistical and behavioral aspects, focusing on density estimation, counting, tracking, and anomaly recognition. A taxonomy of techniques—microscopic and macroscopic—is introduced to distinguish individual-focused from crowd-level models. It evaluates the effectiveness of various deep models like CNNs, Transformers, and Attention-based systems in detecting subtle and complex anomalies in video footage. Key stages in the detection pipeline are outlined: detection, tracking, feature extraction, behavior classification, and final anomaly recognition. The paper highlights the superiority of deep learning in overcoming challenges posed by occlusion, posture variation, and environmental changes. It addresses the need for automated systems in high-risk scenarios like festivals, sporting events, and pandemic control, where human supervision is impractical. Feature extraction techniques including direction, velocity, density, and emotional states (valence and arousal) are analyzed to deepen behavior understanding. The study identifies gaps in model adaptability and interpretability and stresses the importance of integrating emotion-based metrics for improved accuracy.

Md. Haidar Sharif et al. [4] provided a deep and structured survey focusing on crowd anomaly detection through state-of-the-art deep learning techniques. The paper evaluates models developed between 2020 and 2022, presenting a taxonomy and statistical analysis of their performance. Key methods reviewed include CNNs, Autoencoders, GANs, Transformers, and DenseNets, with a comparison of their suitability for different crowd scenarios. A significant contribution is the demonstration that the architecture heterogeneity of pre-trained convolutional models has minimal effect on anomaly detection outcomes. The authors emphasize unsupervised learning approaches due to the scarcity of labeled anomaly data, often using reconstruction loss as an anomaly score. Benchmark datasets like CUHK Avenue, UCF-Crime, and Minnesota2022 are discussed in terms of complexity and relevance. Feature extraction components such as motion pattern, trajectory, collectiveness, and crowd arousal are explored in detail. The study also outlines performance metrics, including AUC, PSNR, F1-score, and event-level accuracy, to standardize model evaluation.

Hrishikesh Gaikwad et al. [5] proposed a real-time crowd monitoring system using deep learning and video-based surveillance for accurate tracking and counting in high-density public areas. The study emphasizes the inefficiency of manual monitoring and advocates for automated detection systems using CNN-based models. The proposed framework includes modules for background subtraction, pixel filtering, and individual tracking to determine real-time population counts. Python and libraries like OpenCV and TensorFlow are used to implement the system, with features extracted from images for model training and testing. The system processes frame-by-frame video to identify the number of individuals in crowded spaces such as malls, universities, and railway stations. CNNs are employed to extract hierarchical features from video frames, learning to distinguish between normal flow and dense crowd formations. Pooling layers reduce noise and computational load while retaining relevant spatial features. The paper explains how the model classifies frames based on crowd presence using learned features. Applications include optimizing resource allocation, preventing congestion, and enhancing emergency response systems. It highlights the model's effectiveness in structured settings but acknowledges limitations in generalizing to complex, chaotic environments.

Muhammad Haris Kaka Khel et al. [6] propose a real-time crowd monitoring framework that extends YOLOv4 through a hybridized approach, aimed at estimating not just the count of people but also their speed and direction of movement. The study addresses challenges of large-scale crowd tracking, particularly in religious and public gathering scenarios like Hajj and sports events, where overcrowding poses significant safety risks.The framework leverages a combination of YOLOv4, model pruning, and convolutional block attention modules (CBAM) to enable effective people detection on low-resource devices. The pruning strategy reduces computational overhead, making the system suitable for edge deployment. Meanwhile, the CBAM enhances feature extraction, focusing on relevant spatial and channel information. Training was conducted on the JHU dataset, which contains highly congested crowd images. The hybrid model achieved an accuracy improvement of 33% over standard YOLOv4 and reached a mean Average Precision (mAP) of 92.1%. The system effectively detects individuals, tracks their movement, and calculates direction and speed—all vital for real-time crowd analytics and emergency preparedness.

Sarah Altowairqi et al. [7] present a comprehensive review of the current advancements in crowd anomaly detection (CAD) with a focus on machine learning and video surveillance. Their study organizes the field based on a taxonomy that includes labeling availability (supervised vs. unsupervised), anomaly types (global vs. local), and components like density estimation, object tracking, and behavior analysis.The authors highlight how CAD plays a crucial role in scenarios such as crowd stampedes, riots, and terrorist threats. The paper discusses both traditional image processing techniques (e.g., background subtraction) and modern deep learning approaches that include CNNs and LSTMs for real-time behavior prediction.Publicly available datasets like UCSD, ShanghaiTech, and UCF-Crime are discussed along with their strengths and weaknesses. Key challenges identified include handling dense, unstructured crowds, dataset imbalance, and achieving reliable anomaly detection without excessive false positives. The review also outlines future research directions such as multi-modal data integration (e.g., combining video, thermal, and audio inputs), privacy-preserving surveillance, and the ethical implications of widespread crowd monitoring. The paper serves as a foundational reference for researchers exploring intelligent surveillance systems in public safety domains.

This paper by Guangshuai Gao et al. [8] provides an extensive survey of over 300 research works focused on deep learning-based density estimation and crowd counting. The authors explore several categories of approaches, including detection-based, regression-based, density-map-based, and CNN-based models, highlighting their evolution and application.The review emphasizes CNN-based density estimation as the most effective technique in dense crowd scenarios, where traditional detection-based models underperform due to occlusion and scale variation. Benchmark models such as CSRNet, MCNN, and SANet are examined for their network architectures, performance

metrics, and dataset adaptability. A key contribution of this study is the evaluation of top-performing models on datasets like NWPU and JHU-CROWD. The authors also provide open-source tools for density map generation and result benchmarking, thus enhancing reproducibility in research.

Lijia Deng et al. [9] present a data-centric review of deep learning models for crowd counting, focusing on architectural designs, dataset classification, and performance benchmarks. They propose a novel Three-Tier Standardized Dataset Taxonomy (TSDT), which categorizes datasets into small-, large-, and hyper-scale based on annotation density.This survey also introduces a new metric, Average Pixel Occupied (APO) per object, offering a more refined assessment of dataset clarity compared to traditional image resolution. The authors categorize crowd counting models into six types: multi-scale, single-column, multi-column, attention-based, multi-task, and weakly supervised networks. Representative models such as CSRNet and SANet are evaluated for their architecture and use cases.The authors observe a research shift from small-scale datasets to more complex, real-world datasets. They highlight critical challenges such as handling occlusion, perspective distortion, and variations in head sizes due to crowd depth. Transfer learning and synthetic data generation are suggested as promising solutions to data scarcity.

Muhammad Jawad Babar et al. [10] present an extensive and structured survey on the state-of-the-art methodologies for crowd counting and density estimation using deep neural networks, primarily Convolutional Neural Networks (CNNs). The survey not only covers the latest deep learning-based approaches but also evaluates conventional methods, categorizing and comparing them across multiple dimensions such as dataset use, feature extraction techniques, and performance evaluation metrics.The paper begins by defining the core objectives of crowd counting (estimating the number of people in a scene) and density estimation (generating a pixel-level density map). Traditional methods, such as regression-based and detection-based algorithms, are discussed for their historical significance and limitations, particularly in handling occlusion, perspective distortion, and scale variation in crowded scenes.A significant contribution of the paper is the detailed classification of CNN-based crowd counting techniques, which are further broken down by their architectural structures (e.g., single-column, multi-column, attention-based, encoder-decoder, generative adversarial models). It evaluates notable models like MCNN, CSRNet, SaCNN, SANet, and CP-CNN, highlighting their strengths in extracting contextual and multi-scale features to achieve high-quality density maps.

Riddhi Sonkar et al. [11] propose a deep learning-based system designed to identify and mitigate abnormal crowd behavior in real-time surveillance footage. The study aims to enhance public safety by detecting potential threats such as riots, theft, and terrorist activities that are more likely to occur in crowded environments like shopping malls, public events, and religious sitesThe system architecture incorporates a combination of Convolutional Neural Networks (CNNs), K-Nearest Neighbors (KNN), and ViBe background subtraction algorithms. Initially, video surveillance input is processed into image frames. ViBe is employed to subtract the background and isolate moving objects. CNNs are used for deep object extraction to detect human subjects, while KNN calculates position differences between sequential frames using Euclidean distance to determine motion.The proposed system evaluates three key motion parameters—speed, direction, and angle—to determine behavioral anomalies. If any object's movement surpasses predefined thresholds, it is classified as abnormal, and an alarm is triggered. The system is designed to be adaptive, where thresholds can be modified depending on the application (e.g., high-security zones or public transit).

Modi Harshadkumar S. et al. [12] presented a dual-model system for anomaly detection and multi-label classification in crowded scenes using deep learning techniques. The system is structured in two major phases: supervised and unsupervised learning. In the supervised phase, a custom model combining Convolutional Neural Networks (CNNs) and an Enhanced Recurrent Neural Network (E-RNN) is deployed. The CNN is responsible for anomaly detection and is optimized using the Elephant Herding-Grey Wolf Optimization (EH-GWO) algorithm to

fine-tune thresholds and network parameters. E-RNN handles multi-label anomaly classification, providing improved accuracy over traditional models.In the unsupervised phase, the authors utilize an Inception Capsule Autoencoder (Inception-CAE) to extract spatio-temporal features from video frames. Anomalous activities are detected by calculating the reconstruction error, and the Coyote Threshold Optimization Algorithm (CTOA) determines threshold boundaries for classification.

Khan, Muhammad Asif et al. [13] This paper addresses a critical challenge in crowd density estimation—how to train effective crowd counting models with imperfect or noisy labels. Traditional deep learning models rely on accurate, dot-annotated datasets for supervision, but acquiring such datasets is labor-intensive and often impractical in real-time scenarios. The authors propose an innovative two-stage pipeline: a secondary deep model, called the annotator, is first trained on accurate annotations to generate noisy or imperfect labels for the same dataset. These imperfect labels are then used to train a lightweight target model, demonstrating that effective density estimation is possible even in the absence of ground-truth annotations.The paper introduces a lightweight variation of CSRNet, termed CSRNet_lite, with significantly fewer parameters (3.9M vs. 16.2M). This model serves as the target model, optimized for faster inference and lower computational cost. The method is tested on benchmark datasets such as DroneRGBT, showing that the model trained on imperfect labels achieves nearly comparable performance to models trained with accurate labels.

Anjali Gupta. et.al [14] , has suggested a system for managing large crowds using deep learning techniques. This system aims to improve safety and security in urban environments by analyzing crowd behavior. The system utilizes the YOLO algorithm for object detection, which processes video frames to detect and categorize individuals into classes such as "Abnormal Violation" or "Serious Violation," based on their proximity to others. The model also employs Non-Maximum Suppression (NMS) to refine bounding boxes. The study highlights the importance of real-time monitoring for applications such as festivals, tourist hotspots, and pandemic-related crowd control. Despite the system's effectiveness in counting and classifying individuals, challenges remain in handling complex crowd scenarios, which limits predictive accuracy.

Mary Jane C. Samonte. et.al [15] , has suggested to develop CrowdSurge, a smart crowd density monitoring solution using YOLOv4 and Closed-Circuit Television (CCTV) to detect and manage crowd density, integrated with a mobile and web application for real-time monitoring and alerts. Key findings reveal that the YOLOv4 model effectively detects and counts people, with an accuracy rate of 91.81%, while the browser and mobile platforms enable administrators and personnel to manage crowd density effectively. Additionally, the system showed minor vulnerabilities, rated as low to medium in severity, according to the CVSS vulnerability assessment. However, gaps remain in the scalability of the system, as it was tested in a limited environment with only two cameras, and improvements are suggested for testing in larger areas with more diverse crowd scenarios.

Sudharson D. et.al [16] , has suggested an AI-based monitoring system for real-time crowd management and security enhancement. The study aims to predict crowd densities and detect abnormal activities such as weapons, fire, falls, and smoke using YOLOv8, a high-performance object detection model. The system sends instant alerts to security personnel through Twilio for rapid response. The model has achieve over 95% accuracy in tracking individuals across various public environments. The dataset used for training contains 500 images across multiple classes, with augmentation techniques applied via Roboflow. While the proposed model demonstrates effectiveness, limitations include constrained real-world testing, potential false positives, and the challenge of generalizing behavior across cultural contexts. Future research must address these gaps to enhance adaptability, optimize real-time performance, and manage costs.

Dr. Shailender Kumar. et.al [17] , has suggested on implementing real-time multiple object tracking (MOT) by leveraging deep learning techniques. Its goal is to accurately detect, identify, and track objects within a specific zone. For detection, the YOLOv4 model is employed, known for its speed and efficiency in real-time environments. Tracking is handled using the DeepSORT algorithm, which extends the Simple Online and Real-time Tracking (SORT) method by incorporating appearance-based features to improve accuracy and reduce identity switches. Kalman filters are applied for motion prediction and to handle occlusions effectively. The MOT17 dataset, containing thousands of frames and bounding boxes, is used for evaluation, achieving a tracking accuracy of about 60% and an FPS of up to 31. Despite these promising results, limitations such as environmental factors and a smaller dataset led to reduced performance, highlighting the need for larger datasets and refined models for future work.

The reviewed literature highlights significant advancements in real-time crowd monitoring, especially in crowd density estimation and anomaly detection. Building on these developments, our project employs YOLOv8 for accurate detection of individuals and abnormal behaviors, combined with DeepSORT for reliable multi-object tracking across video frames. This combination enables the system to monitor individual movements and behavioral trends over time without relying on recurrent models like LSTM. The system is designed to detect key anomalies such as panic, physical fights, and stagnation—the condition where a crowd or group remains immobile in a specific area for an extended period. DeepSORT facilitates the identification of these temporal patterns by maintaining identity tracking, while YOLOv8 provides high-precision detections of people and event-specific behaviors.

In addition to detection, the project features an interactive map-based visualization of real-time crowd density levels and anomaly alerts using heatmaps and geolocation markers. This empowers public safety officials and event organizers to make quick, informed decisions for crowd management and risk mitigation. By focusing on real-time responsiveness and handling high-density scenarios, our system addresses critical challenges identified in existing research and offers a practical solution for smart surveillance and crowd safety.

Methodology:

The proposed system aims to achieve real-time crowd density estimation and anomaly detection—specifically identifying panic, fights, and people standing still for unusually long durations—using surveillance video input. The methodology integrates computer vision, deep learning-based object detection, and behavioral analysis for robust monitoring. The complete process is structured into the following key stages:

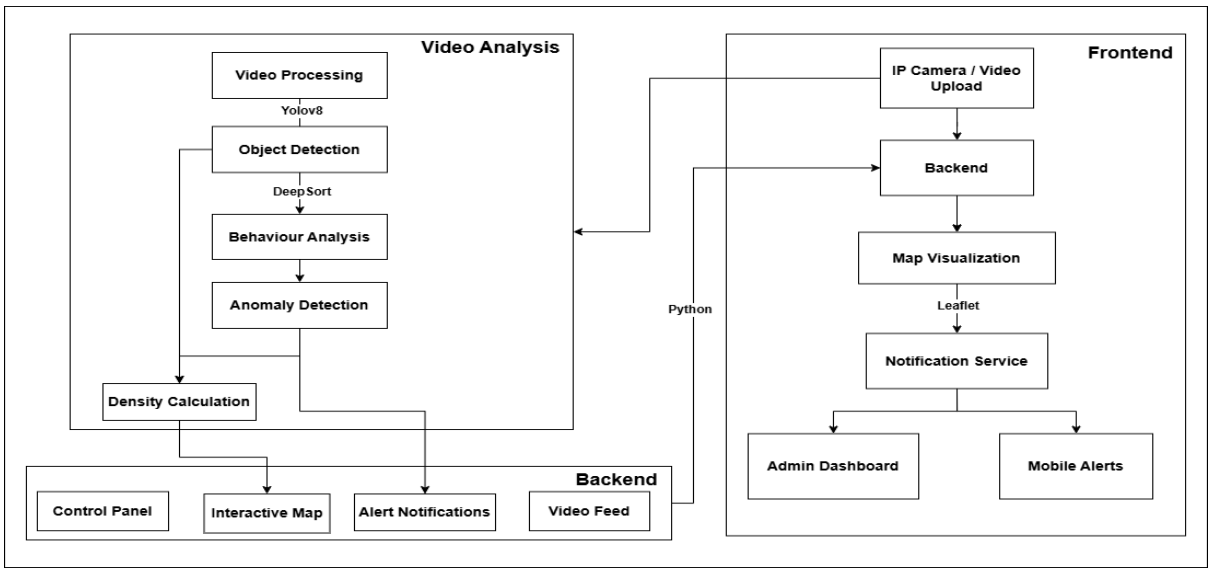

Figure 1: System Architecture

1 Data Collection and Preprocessing

The system utilizes a combination of publicly available crowd datasets, including UCF-QNRF, ShanghaiTech Parts A and B, and the Francisco Mena dataset, to train and evaluate the crowd density estimation model. These datasets offer a wide range of crowd sizes, densities, and perspectives, helping the model generalize across different real-world scenarios. For anomaly detection, a custom dataset from Roboflow is employed, annotated with behavioral classes: panic, fight, and people not moving. This dataset includes labeled frames representing diverse abnormal crowd behaviors from various angles and scenes, helping improve anomaly classification performance.

A custom dataset was also utilized for training the model. To build this dataset, video footage and still images were gathered from various sources, including direct video capture and publicly available recordings. These visual samples were then carefully processed and manually annotated to create accurate ground truth labels required for supervised learning. The annotation process involved identifying and labeling relevant objects and behaviors within the images to ensure high-quality data for model training. An illustration of the dataset being annotated is provided in Figure 2.

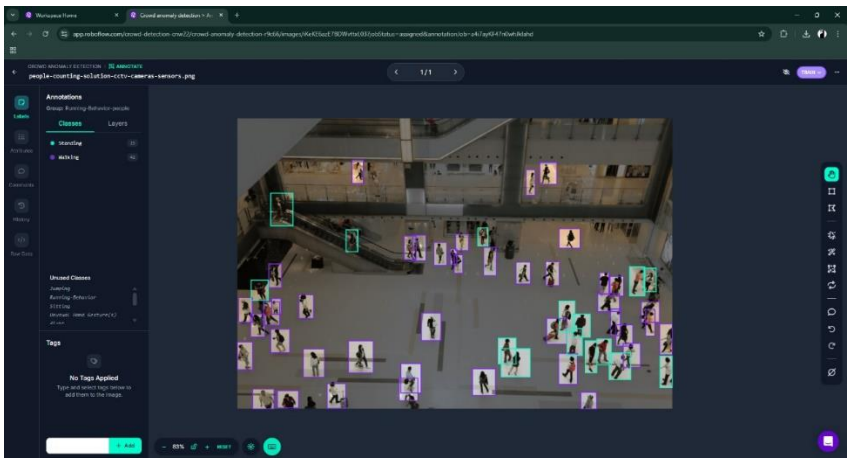

Figure 2: Manual Annotations

The collected videos are preprocessed through frame extraction at a standard rate (e.g., 10 FPS) to ensure consistent input during training and inference. Each frame is scaled to 640×640 pixels and normalized to meet the input requirements of the YOLOv8 model. Data augmentation techniques such as rotation, brightness adjustment, and flipping are applied to enhance robustness against environmental variability and camera perspectives, as suggested in related works [1][4][12].

2 Crowd Detection Using YOLOv8

YOLOv8, a state-of-the-art real-time object detection algorithm, is adopted for detecting individuals in video frames due to its enhanced speed, accuracy, and support for custom training workflows [3][7]. The model is fine-tuned on annotated crowd datasets using the 'person' class. YOLOv8's anchor-free architecture allows it to detect human figures across varying scales and dense regions more efficiently than earlier versions. This feature makes it highly suitable for crowd analysis, especially in scenes with heavy occlusion and perspective distortion [2][4].

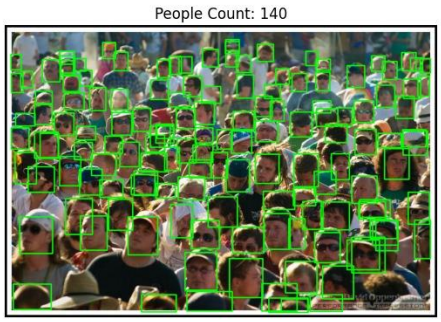

Figure 3: Crowd Detection & Counting

For every frame, the model outputs bounding boxes along with confidence values for each person it detects. These detections are logged and passed to the tracking module to maintain temporal consistency and enable behavior monitoring. The improved backbone and detection head of YOLOv8 ensures higher precision and fewer false positives, which is essential for both density estimation and behavior classification [5][11].

3 Real-Time Tracking Using DeepSORT

To track individuals across consecutive frames, DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric) is used in conjunction with YOLOv8. While YOLOv8 handles object detection, DeepSORT maintains a consistent identity for each person across frames, enabling trajectory mapping and motion pattern recognition [9][15].

DeepSORT enhances the basic SORT algorithm by incorporating appearance descriptors through a deep neural network, improving its ability to re-identify individuals even after occlusion or sudden movements. It uses Kalman filtering for motion prediction and Hungarian matching for data association. The system relies on the bounding boxes and embeddings from YOLOv8 to update tracks, which is vital for anomaly classification, particularly when detecting prolonged immobility or sudden chaotic motion indicative of panic or fights [8].

4 Anomaly Detection through Behavioral Analysis

Anomaly detection is performed by analyzing spatio-temporal features of individuals tracked over time. Three types of anomalies are considered: panic, fight, and standing still for long durations. A rule-based and threshold-driven approach is used to classify these behaviors based on motion dynamics and bounding box overlap.

- Panic Detection: Individuals moving rapidly in disorganized patterns or showing abrupt acceleration and directional shifts are flagged as panic behavior. These features are extracted from velocity vectors computed using DeepSORT trajectories over short time windows. Literature confirms that rapid irregular motion correlates with panic scenarios [13][14].

- Fight Detection: Fights are characterized by multiple individuals engaging in repetitive, aggressive movements with significant bounding box overlap. The system monitors proximity-based interactions and jerky motion patterns. If multiple individuals remain in tight clusters with high motion variance, a fight event is detected [6][10].

- Immobility Detection: When a person remains stationary beyond a predefined temporal threshold (e.g., 30 seconds), they are marked as potentially in distress or exhibiting suspicious behavior. The tracker logs continuous low-movement patterns and correlates them with crowd flow to distinguish between normal resting and anomalous inactivity [16].

Each detected anomaly is assigned a severity level based on the number of individuals involved, location density, and duration. These anomalies are time-stamped and sent to the alert module.

Figure 4: Anomaly Detection

Recommended Camera Coverage:

To ensure reliable person and behavior detection:

- Optimal coverage area: 10m x 10m (100 square meters) per camera.
- This range ensures that each individual is captured with sufficient pixel resolution for YOLOv8 to detect postures, motion, and proximity.
- Higher coverage (e.g., 20x20m) risks reduced per-person resolution, which may lower precision, especially for behavior classification (e.g., mistaking gestures).

Optimal Recommended Crowd Size for Reliable Detection

Anomaly Type	Optimal Crowd Size in Frame	Explanation
Panic	**8–15 people**	This range allows the system to detect group-based fast, erratic motion patterns and crowd direction changes more reliably. Smaller groups may resemble normal walking behavior.
Fight	**3–6 people**	Sufficient for detecting aggressive motion patterns between individuals and verifying interactions like pushing or hitting.
Immobility (standing too long)	**15-20 person**	Can be detected even using DeepSORT temporal tracking and spatial anchoring.

Table 1: Optimal Recommended crowd size

5 Crowd Density Estimation and Classification

Crowd density is determined by measuring how many individuals are detected within a specific area in each video frame. Based on the bounding box count and frame area, the density is classified into three categories:

- Low: Fewer than 10 people per frame
- Medium: 10–30 people per frame
- High: More than 30 people per frame

This classification helps emergency responders and users quickly assess crowding risk and understand the severity of anomalies when viewed on the map. The adaptive thresholding is inspired by density-aware methods discussed in [1][3][6].

6 Real-Time Map Visualization and Alerting System

The processed results are visualized on an interactive Google Map interface, where each video input is mapped to a geolocation marker. The marker's color represents the crowd density level: green for low, orange for medium, and red for high. Anomalies are also displayed as overlay icons with real-time labels and timestamps.

For critical events like panic or fights, the system sends automated email alerts using an SMTP integration, attaching snapshot evidence and location metadata. This provides real-time situational awareness for emergency responders, security personnel, and the public. The live dashboard updates dynamically with video feeds, density scores, and anomaly types [7][14][15].

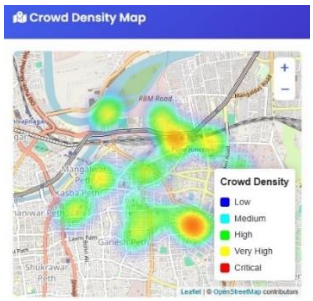

Figure 5: Crowd Density Map

Results:

The proposed system was evaluated across multiple surveillance video inputs to test its ability to detect crowd density and anomalies such as panic, fights, and immobility. The system was benchmarked on both detection accuracy and system performance, including latency and processing speed across hardware configurations.

1. Crowd Detection Accuracy

The integration of YOLOv8 for crowd detection, fine-tuned on annotated datasets encompassing dense crowds and anomalous behaviors, leverages the model's efficient architecture to achieve high precision in real-time object localization. By coupling this with DeepSORT's multi-object tracking, the system establishes temporal coherence across frames, enabling robust behavior analysis through velocity, trajectory, and proximity metrics.

1. **Precision:** Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It focuses on the relevancy of positive predictions and is computed as:

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)}$$

- The model achieved a precision of 96.3% suggesting that when the model predicted a anomaly, it was correct 96.3% of the time. This is crucial to minimize false alarms.

2. **Recall:** Recall is the ratio of correctly predicted positive observations to all actual positive observations. It is especially important in medical diagnostics to ensure no positive cases are missed:

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)}$$

- Indicates that 92.5% of the detections made by YOLOv8 were correct. It reflects how reliable the model is when it predicts an object.

3. F1-Score: The F1-score is the harmonic mean of precision and recall, offering a balance between the two:

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

- With an F1-score of 94.3%, the model exhibits a high level of competence, combining accuracy and completeness of its prediction into one robust metric

4. Detection Accuracy : Accuracy refers to the ratio of correctly predicted observations to the total observations. It is a general performance measure and is defined as:

$$Accuracy = \frac{Correct\ \ Predictions}{Total\ \ Predictions}$$

- An accuracy of 94.8% means that the model correctly classified approximately 95 out of every 100 images, indicating a high level of reliability in predictions

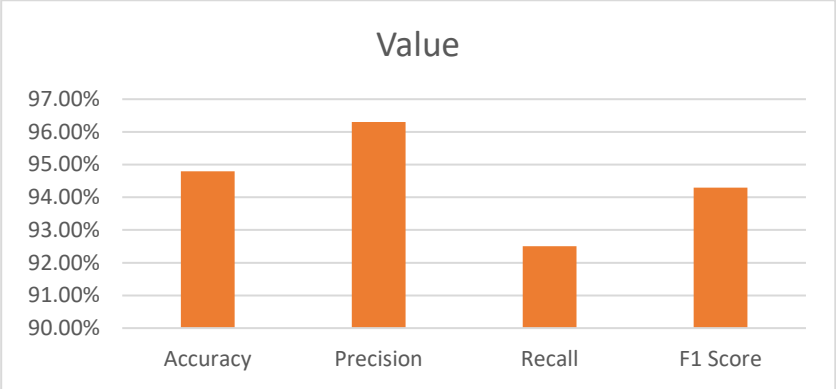

Figure 6: Crowd Detection Accuracy Metrics

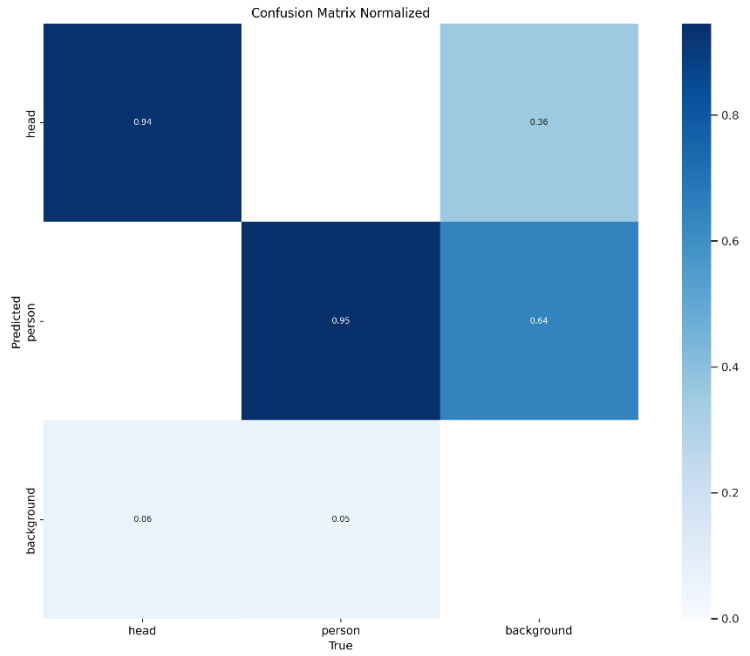

Figure 7: Confusion Matrix for Crowd Detection Model

2. Anomaly Detection Accuracy

Using the YOLOv8 model fine-tuned on annotated crowd and anomaly datasets, the system achieved high precision across detection tasks. The behavior detection logic based on DeepSORT tracking and temporal thresholds showed robust performance under varying conditions.

1. Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It focuses on the relevancy of positive predictions and is computed as

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)}$$

- The model achieved a precision of 86.2% suggesting that when the model predicted a anomaly, it was correct 86.2% of the time. This is crucial to minimize false alarms.

2. Recall: Recall is the ratio of correctly predicted positive observations to all actual positive observations. It is especially important in medical diagnostics to ensure no positive cases are missed:

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)}$$

- Indicates that 86.2% of the detections made by YOLOv8 were correct. It reflects how reliable the model is when it predicts an object.

3. F1-Score: The F1-score is the harmonic mean of precision and recall, offering a balance between the two:

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

- With an F1-score of 84.9%, the model exhibits a high level of competence, combining accuracy and completeness of its prediction into one robust metric

4. Panic Detection Accuracy : Accuracy refers to the ratio of correctly predicted observations to the total observations. It is a general performance measure and is defined as:

$$Accuracy = \frac{Correct\ Panic\ Predictions}{Total\ Panic\ Predictions}$$

- An accuracy of 82.5% means that the model correctly classified approximately 83 out of every 100 images, indicating a high level of reliability in predictions

5. Fight Detection Accuracy

$$Fight\ Detection\ Accuracy = \frac{Correct\ Flight\ Predictions}{Total\ Flight\ Predictions}$$

- An accuracy of 82.5% means that the model correctly classified approximately 86 out of every 100 images, indicating a high level of reliability in predictions

6. Immobility Detection Accuracy

$$Immobility\ Detection\ Accuracy = \frac{Correct\ Immobility\ Detection}{Total\ Immobility\ Cases}$$

- Shows that 87.3% of cases where people remained immobile were detected accurately. Helps identify medical emergencies or abnormal behavior in a crowd.

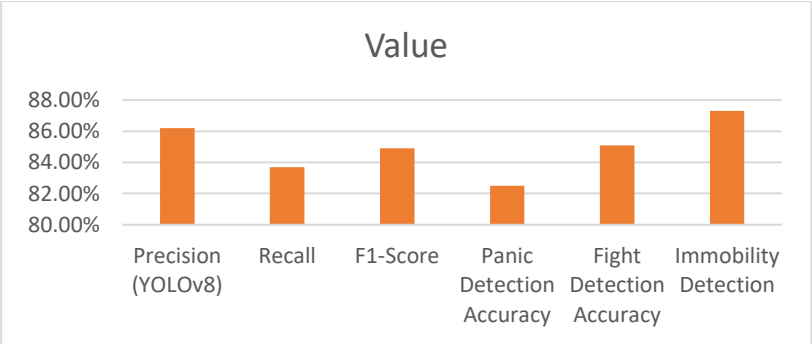

Figure 8: Anomaly Detection Accuracy Metrics

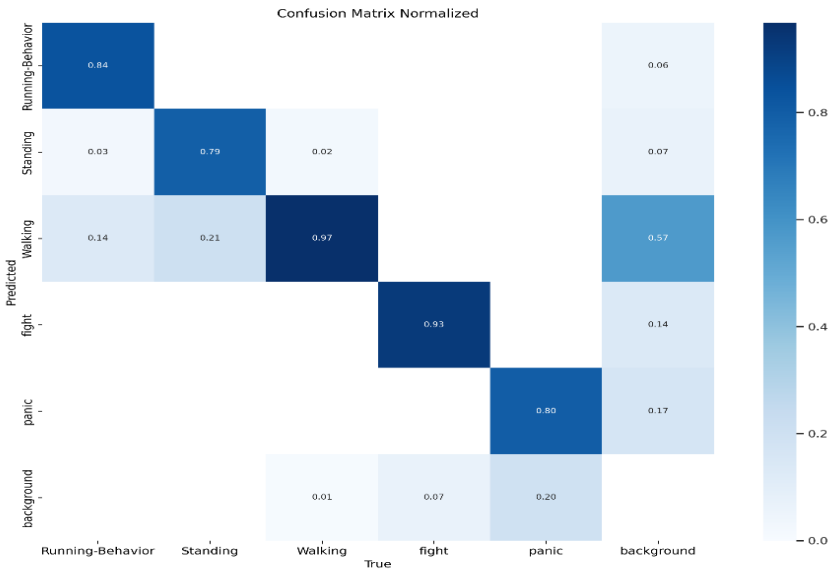

Figure 9: Confusion Matrix for Anomaly Detection Model

These results validate the model's capability to differentiate between normal and abnormal crowd behaviors effectively.

2 Component-Wise Latency and FPS Analysis

Latency analysis was performed for individual processing components of the system. The optimizations include model pruning, reduced frame resolution, and batch inference strategies, which significantly improve overall responsiveness.

Component	Original Latency	Optimized Latency
Frame Processing	150–300 ms	50–100 ms
Person Detection	80–150 ms	30–60 ms
Behavior Detection	100–200 ms	15–30 ms (amortized)
Tracking	50–100 ms	20–40 ms
Total per Frame	300–600 ms	100–200 ms
Effective FPS	1–3 FPS	3–5 FPS

Table 2: Latency Metrics

With these optimizations, the system achieves real-time performance in GPU environments and near-real-time performance on high-end CPUs. The drop in latency enhances the system's capability to trigger instant alerts, especially in time-sensitive scenarios like crowd panic or violent behavior.

3 Hardware-Based Performance Comparison

The system was benchmarked on two different configurations:

- **GPU (Google Colab Pro, NVIDIA Tesla T4):** Achieves around 24 to 28 frames per second for raw model processing, with an overall latency of approximately 100 to 200 milliseconds per frame when including visualization and alert components.
- **CPU (Intel Core i7-10750H):** Delivers a frame rate of about 5 to 8 FPS, with total per-frame latency ranging from 300 to 600 milliseconds.

These results show a substantial improvement in responsiveness and efficiency when using GPU resources, making the system suitable for real-time surveillance applications.

4 Visualization and Alerts

The visual dashboard presented real-time crowd density and behavioral anomalies overlaid on Google Maps. Visual cues like density color codes (green/yellow/red) and anomaly icons (panic/fight/immobile) updated dynamically. The email alerting system triggered notifications within 3–5 seconds after anomaly detection, ensuring minimal response delay.

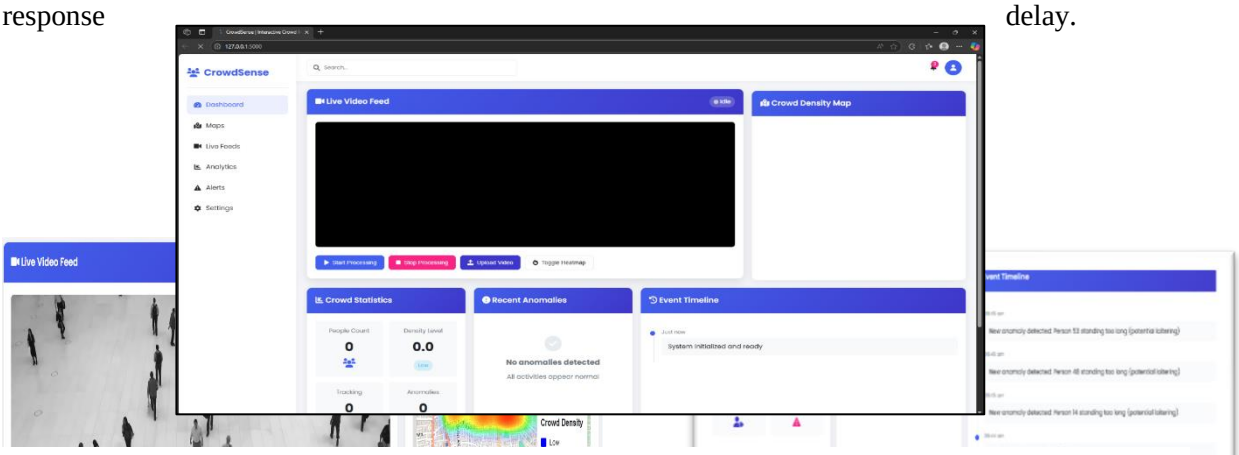

Figure 10: GUI

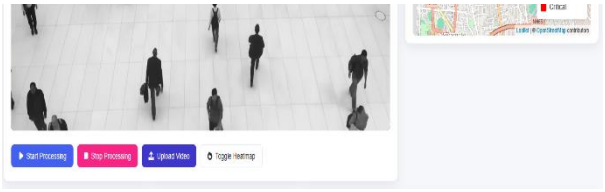

Figure 11: Crowd Density Mapping

Figure 12: Dashboard

Conclusion

This study proposes an effective real-time crowd monitoring framework that integrates YOLOv8 for object detection and DeepSORT for multi-object tracking and behavioral analysis. The system is capable of identifying critical anomalies—such as panic, physical altercations, and prolonged inactivity—while also categorizing crowd density into low, medium, and high levels. The visual output, presented through a Google Maps-based interface, enhances usability and situational awareness for both authorities and the general public. With GPU acceleration, the framework achieves a real-time performance of 10–20 frames per second, ensuring responsive and smooth monitoring suitable for live surveillance applications.

Despite demonstrating a precision rate of approximately 86%, the system has certain limitations, particularly under conditions involving heavy occlusion or extremely dense crowds where detection accuracy may decline.

Additionally, effective anomaly recognition is dependent on a minimum number of individuals within the camera frame to accurately interpret motion and spatial cues. The current approach focuses solely on person-based behaviors, overlooking context- or object-specific anomalies. Future work could address these challenges by incorporating spatio-temporal attention mechanisms, transformer-based models, and multi-camera coordination to enhance robustness, scalability, and contextual awareness in complex environments.

References

[1] S. Author and S. Roy, "A comprehensive survey on crowd anomaly detection using deep learning," *2024*. [Online]. Available: https://doi.org/10.13140/RG.2.2.24390.48961

[2] S. Shukla, B. Kumar, and H. Tiwari, "A comprehensive survey on abnormal crowd behaviour," *Northern Economic Review*, vol. 15, no. 1, 2024. [Online]. Available: https://nerj.org/. DOI: https://doie.org/10.0130/Nerj.2024509272

[3] D. Tank, S. G. Patel, and D. S. Pandya, "Recent advancements in deep learning for crowd anomaly detection: A comprehensive survey," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 21s, pp. 2889–2902, 2024. [Online]. Available: http://www.ijisae.org

[4] M. H. Sharif, L. Jiao, and C. W. Omlin, "Deep crowd anomaly detection: State-of-the-art, challenges, and future research directions," *Artif. Intell. Rev.*, vol. 58, p. 139, 2025. DOI: 10.1007/s10462-024-11092-8

[5] H. Gaikwad, S. Jadhav, N. Gunjal, S. Survase, and K. Shinde, "Real-time crowd monitoring system," *Int. Res. J. Mod. Eng. Technol. Sci.*, Maharashtra, India, 2024. DOI: https://www.doi.org/10.56726/IRJMETS38950

[6] M. H. K. Khel, K. A. Kadir, S. Khan, M. Noor, H. Nasir, N. Waqas, and A. Khan, "Realtime crowd monitoring—estimating count, speed and direction of people using hybridized YOLOv4," *IEEE Access*, 2024.

[7] S. Altowairqi, S. Luo, and P. Greer, "A review of the recent progress on crowd anomaly detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, 2023. DOI: 10.14569/IJACSA.2023.0140472

[8] G. Gao, J. Gao, Q. Liu, Q. Wang, and Y. Wang, "A survey of deep learning methods for density estimation and crowd counting," *Vicinagearth*, vol. 2, no. 2, 2025. DOI: 10.1007/s44336-024-00011-8

[9] L. Deng, Q. Zhou, S. Wang, J. M. Górriz, and Y. Zhang, "Deep learning in crowd counting: A survey," *CAAI Trans. Intell. Technol.*, 2023. DOI: 10.1049/cit2.12241

[10] M. J. Babar, M. Husnain, M. M. S. Missen, et al., "Crowd counting and density estimation using deep network—A comprehensive survey," *TechRxiv*, Jun. 2, 2023. DOI: 10.36227/techrxiv.23256587.v1

[11] R. Sonkar, S. Rathod, R. Jadhav, and D. Patil, "Crowd abnormal behaviour detection using deep learning," in *Proc. ITM Web Conf.*, vol. 32, p. 03040, 2020. DOI: 10.1051/itmconf/20203203040

[12] H. S. Modi, "Multi label anomaly detection for crowd scenes using deep learning," Ph.D. dissertation, Dept. of Computer/IT Eng., Gujarat Technological Univ., India, 2023.

[13] M. A. Khan, H. Menouar, and R. Hamila, "Crowd density estimation using imperfect labels," Qatar Univ., Doha, Qatar, 2023.

[14] A. Gupta, A. Tiwari, R. Singh, and P. Sharma, "Crowd management system using deep learning," *Int. J. Res. Eng. Sci.*, vol. 10, no. 7, pp. 446–449, 2022.

[15] M. J. C. Samonte, D. M. P. Ramos, R. T. Mendoza, and C. A. D. Silva, "CrowdSurge: A crowd density monitoring solution using smart video surveillance with security vulnerability assessment," *J. Adv. Inf. Technol.*, vol. 13, no. 2, pp. 140–147, Apr. 2022.

[16] D. Sudharson, R. M. Joseph, K. P. Anand, and K. U. Srinivasan, "Proactive headcount and suspicious activity detection using YOLOv8," in *Proc. 3rd Int. Conf. Evol. Comput. Mobile Sustainable Networks*, 2023.

[17] S. Kumar, R. Mehta, V. Sharma, and A. Dubey, "Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow," in *Proc. IEEE Conf.*, 2021.

[18] E. B. Varghese et al., "Application of Cognitive Computing for Smart Crowd Management," IEEE Xplore, vol. 22, no. 4, 2021.
[19] D. Garcia-Retuerta et al., "An Efficient Management Platform for Developing Smart Cities: Solution for Real-Time and Future Crowd Detection," Electronics 2021, 10, 765.
[20] X. Ding et al., "Crowd Density Estimation Using Fusion of Multi-Layer Features," IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, pp. 1524-9050, 2020.