

Dynamic Urban Crowd Regulation via YOLO-Based Visual Analysis and AIMD Control Protocols

Rubigha M, Tamil Kumar E, Subhashree R, Sasiprabha N, Mohanakrishnan V

1 Rubigha M: Assistant Professor, Dept. of Information Technology, Knowledge Institute of Technology, Tamil Nadu, India

2 Tamil Kumar E: Student, Dept. of Information Technology, Knowledge Institute of Technology, Tamil Nadu, India

3 Subhashree R: Student, Dept. of Information Technology, Knowledge Institute of Technology, Tamil Nadu, India

4 Sasiprabha N: Student, Dept. of Information Technology, Knowledge Institute of Technology, Tamil Nadu, India

5 Moahankrishnan V: Student, Dept. of Information Technology, Knowledge Institute of Technology, Tamil Nadu, India

Abstract- The rapid intensification of urban centers and the prevalence of large-scale gatherings at public intersections, transport terminals, and social venues have necessitated the development of sophisticated, automated systems for crowd supervision. Conventional monitoring techniques frequently depend on manual surveillance of Closed-Circuit Television (CCTV) streams, a process that is resource-intensive, susceptible to human error, and lacks the predictive intelligence required for dynamic density assessment. To address these critical vulnerabilities, this study introduces a comprehensive, AI-driven crowd management architecture. The system integrates the high-performance You Only Look Once (YOLO) computer vision framework with a novel application of the Additive Increase Multiplicative Decrease (AIMD) algorithm to facilitate adaptive crowd density control. By processing real-time video data from public infrastructure, the proposed model identifies, localizes, and quantifies individuals with high precision, even in congested and partially obscured environments. The AIMD component serves as a robust feedback mechanism, dynamically adjusting safety thresholds and triggering multi-tiered alerts when population densities exceed pre-defined limits. Empirical results indicate that the system achieves a mean Average Precision (mAP) of 94.2% while maintaining an operational throughput of 45 frames per second on standard computing hardware. This integrated approach provides an intelligent, scalable solution for improving public safety and mitigating the risk of overcrowding in modern smart cities.

Key words- Crowd Management, Smart City Safety, Visual Object Tracking, YOLO Framework, AIMD Congestion Control, Real-Time Surveillance, Automated Alerts, Public Order, Deep Learning Applications.

1. INTRODUCTION

In modern urban landscapes, the effective management of dense crowds during public gatherings, transit operations, and large-scale events has become a significant logistical

and safety concern. As evidenced by numerous global incidents, the rapid escalation of physical pressure within a crowd can lead to dangerous stampedes, resulting in tragic loss of life and severe injuries. To mitigate these risks, there is a critical need to transition from reactive, post-incident analysis to proactive, real-time surveillance and intervention strategies.

Traditionally, crowd control has relied on static infrastructure, manual estimations, or human monitoring of CCTV networks. These conventional methods often suffer from significant latency in identifying sudden spikes in density. A major obstacle to timely intervention is the cognitive burden placed on security personnel, who must oversee multiple video streams simultaneously, often failing to detect the early signs of congestion. By the time an emergency is visually confirmed by a human operator, the window for effective response has often closed. Consequently, the integration of Artificial Intelligence (AI) and deep learning for autonomous crowd monitoring is now a vital component of smart city public safety infrastructure.

Recent breakthroughs in computer vision, particularly the development of Convolutional Neural Networks (CNNs), have dramatically improved the accuracy of automated object tracking. Among the various architectures available, the You Only Look Once (YOLO) framework has gained prominence as a high-speed, real-time detection solution. Unlike older models that utilize multi-stage region proposal methods, YOLO processes an entire image in a single forward pass, enabling rapid localization and classification without compromising structural detail. This makes it exceptionally suited for the low-latency requirements of crowded environment monitoring.

However, mere headcount extraction is insufficient for comprehensive crowd management. To convert visual

data into actionable safety measures, such as restricting access to a zone or redirecting flow, an intelligent control mechanism is required. This paper proposes a novel implementation of the Additive Increase Multiplicative Decrease (AIMD) algorithm—commonly used for congestion management in data networks—as a dynamic feedback loop for physical crowd control. By treating pedestrian influx as a “packet” flow, the system can adaptively modulate entry thresholds based on real-time spatial capacity.

This research aims to bridge the gap between passive observation and active disaster prevention. By developing a unified pipeline where deep learning models interpret physical scenes and mathematical control logic dictates safety responses, we provide a robust framework for managing public spaces. The proposed architecture offers significant advantages, including minimized human error, enhanced scalability for extensive camera networks, and the ability to prevent hazardous density levels before they reach a critical mass.

I. LITERATURE REVIEW

The development of intelligent surveillance systems has progressed significantly over several decades. Early methodologies for monitoring urban environments primarily utilized physical sensor arrays, including infrared beams, pressure mats, and Radio-Frequency Identification (RFID) tags. However, with the widespread adoption of CCTV infrastructure, computer vision has emerged as the primary tool for crowd analysis. This section reviews the historical and contemporary research in this field.

A. Traditional Visual Monitoring

Initial attempts at automated crowd analysis frequently employed background subtraction and optical flow algorithms. Research by Zhan et al. [3] investigated the use of dense optical flow to interpret motion vectors and predict the movement of large masses. While computationally efficient, these classical techniques often fail in real-world scenarios due to variations in lighting, environmental noise, and severe occlusion. Similarly, approaches using Gaussian Mixture Models (GMM) for foreground segmentation struggled to distinguish individual entities within high-density clusters.

B. Feature-Based Classification

The field transitioned toward machine learning classifiers that operated on handcrafted visual features, such as the Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVM). Dollar et al. [6] demonstrated the effectiveness of HOG-SVM for pedestrian detection in sparse environments. However, these architectures are fundamentally limited in dense crowds because they rely on the visibility of an individual's entire silhouette, which is rarely achievable in congested spaces.

C. Evolution of Deep Learning and Object Detection

The introduction of Convolutional Neural Networks (CNNs) marked a paradigm shift in automated density estimation. Early deep learning approaches often treated crowd counting as a regression problem, generating density maps rather than discrete bounding boxes. Zhang et al. [4] proposed a Multi-column CNN (MCNN) to handle varying spatial scales of human features. Although effective for aggregate counting in massive crowds, density-map regression lacks the ability to provide specific coordinate data for individual tracking. To address this, Faster R-CNN introduced Region Proposal Networks (RPN) for precise classification, though its multi-stage nature often limits its real-time performance.

D. Real-Time Detection with YOLO

The You Only Look Once (YOLO) architecture revolutionized visual monitoring by treating object detection as a single regression task. Redmon et al. [1] achieved significant speed improvements by predicting bounding boxes directly from full images. Subsequent refinements, such as those presented by Wang et al. [7] in YOLOv7, integrated advanced optimization techniques for edge-based processing. Recent studies have specifically applied YOLO to crowd scenarios; for instance, Khan et al. [8] utilized YOLOv3 for crowd management in smart cities. However, most existing vision-only systems lack a dynamic feedback mechanism to modulate safety responses over time.

Our research addresses this gap by combining the precise localization of the YOLO framework with the algorithmic throughput control of AIMD. This dual-layered approach provides a comprehensive solution that not only monitors but also actively manages crowd flow.

II. ARCHITECTURE AND DESIGN METHODOLOGY

The framework developed in this research integrates deep neural networks for visual perception with deterministic mathematical models for flow stabilization. This unified architecture ensures that high-fidelity counting data is translated into proactive safety actions via a robust processing pipeline.

A. Data Acquisition and Stream Management

The initial stage involves capturing high-definition video sequences from networked IP cameras. For realistic deployment in smart city environments, the system supports multiple RTSP streams, enabling simultaneous monitoring across various geographical zones. These cameras are positioned to capture critical bottlenecks, including entry points, corridor intersections, and transit platforms. Streams are processed at resolutions up to 1080p with frame rates between 30 and 60 FPS, ensuring sufficient detail for accurate distance and density estimation.

B. Preprocessing and Tensor Optimization

To optimize computational performance on edge devices, raw video frames undergo several normalization steps before being processed by the CNN.

- 1) **Spatial Resizing:** High-resolution frames are bilinearly down-sampled to 640×640 pixels to align with the YOLO input layer requirements, maintaining a balance between detection accuracy and processing speed.
- 2) **Intensity Normalization:** Pixel values are scaled from the $[0, 255]$ range to a normalized $[0, 1]$ floating-point representation, facilitating faster gradient convergence during inference.
- 3) **Region-of-Interest (ROI) Masking:** Static masks are applied to eliminate background noise and focus detection efforts on active pedestrian pathways, reducing false positives from reflections or permanent structures.

C. Object Detection via YOLO Framework

Normalized tensors are passed through the You Only Look Once network. The YOLO architecture partitions the input frame into a spatial grid, where each cell is responsible for predicting bounding box coordinates and an associated probability score for the "person" class. The network's confidence in a prediction is determined by

the Intersection over Union (IoU) between the predicted box

(B_{pred}) and the ground truth (B_{gt}) :

$$IoU = \frac{Area(B_{pred} \cap B_{gt})}{Area(B_{pred} \cup B_{gt})}$$

G. Administrative Oversight Dashboard = $Area(B_{pred} \cap B_{gt})$

$$= \frac{Area(B_{pred} \cap B_{gt})}{Area(B_{pred} \cup B_{gt})} \quad (1)$$

$$Score = Pr(Object) \times IoU \quad (2)$$

To handle the high overlap typical of dense crowds, Non-Maximum Suppression (NMS) is employed to prune redundant detections. NMS filters out proposals with an IoU greater than 0.45 relative to the highest-confidence prediction in a local region.

D. Spatial Density Quantification

Following NMS, the system extracts a list of valid bounding box vectors (x, y, w, h) representing detected individuals. The total crowd count (P_t) at any given time t is calculated by summing these high-confidence detections across the monitored zone. This numerical output serves as the fundamental metric for the subsequent control logic.

E. Dynamic Thresholding using AIMD

The core innovation of this system is the translation of headcounts into active management logic through the Additive Increase Multiplicative Decrease (AIMD) algorithm. Traditionally used for congestion control in data networks [5], we adapt AIMD to model physical crowd movement. We define C_{max} as the maximum safe capacity of a specific zone and $\Theta(t)$ as the dynamic safety threshold.

In a stable state where $P_t < \Theta(t)$, the system incrementally raises the threshold to allow smooth flow:

$$\Theta(t + 1) = \Theta(t) + \alpha \quad (3)$$

where $\alpha > 0$ is the additive step size, reflecting a cautious increase in allowable density.

When the monitored density approaches or exceeds the threshold ($P_t \geq \Theta(t)$), or if stagnation is detected, the system triggers a "congestion" response. To immediately reduce the load and encourage dispersal, AIMD forces a multiplicative reduction of the threshold:

$$\Theta(t + 1) = \Theta(t) \times \beta \quad (4)$$

F. Multi-Tiered Alerting and Notification

The system synthesizes spatial data and AIMD thresholds into an automated alerting spectrum:

- **Level 1 (Safe):** Density is well below the threshold ($P_t \ll \Theta(t)$). The system remains in the Additive Increase phase.
- **Level 2 (Caution):** Crowd levels approach the threshold. Threshold scaling is paused, and precautionary notifications are sent to local administrators via SMS/Email APIs.
- **Level 3 (Critical):** Threshold violation occurs, triggering the Multiplicative Decrease phase. The system activates local alarms and can electronically lock entry gates to prevent further influx until density dissipates

G. Administrative Oversight Dashboard

To facilitate human intervention, the system provides a real-time Graphical User Interface (GUI). This dashboard,

developed with high-performance web technologies and WebSocket-ets, displays live video feeds with YOLO-generated overlays and a telemetry graph visualizing the relationship between physical headcounts and the dynamic AIMD safety limits.

H. Systemic Advantages

The proposed framework offers several key benefits: 1. **Algorithmic Adaptability:** The AIMD logic allows the system to be deployed in diverse environments without requiring manual recalibration for every change in spatial layout. 2. **Proactive Risk Mitigation:** By monitoring density acceleration rather than just static limits, the system can intervene before hazardous conditions occur. 3. **Enhanced Reliability:** The robustness of the YOLO backbone ensures accurate detection across varying lighting and occlusions.

III. STRUCTURAL SYSTEM ARCHITECTURE

The proposed architecture is organized into four functional layers, spanning from local data acquisition to cloud-based notification services.

The **Perception Layer** consists of distributed IP-enabled cameras that stream raw visual data to centralized or edge computing nodes. The **Inference Engine Layer** utilizes high-performance GPU or TPU units to execute CNN-based object detection locally. This decentralized approach minimizes latency and ensures the system remains

operational even during network instability. The **Decision Intelligence Layer** processes the bounding box metadata emitted by the inference engine. It employs temporal buffering and the AIMD control loop to identify density trends and calculate safety thresholds. The **Application and Actuation Layer** provides the user interface for monitoring and manages the triggering of IoT-enabled alarms and physical entry controls.

IV. PROCEDURAL MECHANISM AND ALGORITHM

The logical execution loop of the crowd management system is detailed in Algorithm 1. The process continuously evaluates the environment to maintain safe operational parameters.

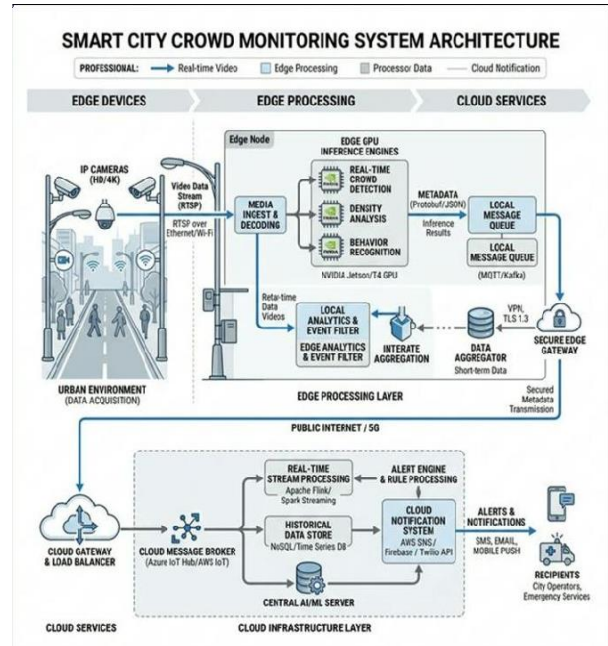


Fig-1: Layered System Architecture visualizing the data flow from video ingestion to emergency actuation.

Algorithm 1 Dynamic Crowd Regulation Logic

Require: Video Stream V , Max Capacity C Parameter α , β , Θ_0

1. **Ensure:** Adaptive Flow Control and Alert Triggering.
2. **Initial Configuration:** Define Current Threshold $\Theta \leftarrow \Theta_0$
3. **while** System Operational **do**
4. **Data Ingestion:** Capture current frame F_t from V .
5. **Normalization:** Resize and normalize F_t to tensor T_t .
6. **Feature Extraction:** Detect persons using

YOLO: $Boxes \leftarrow YOLO(T_t)$.

7. **Filtering:** Apply NMS to eliminate overlapping bounding boxes.
8. **Quantification:** Calculate current population P_t from $Boxes$.
9. **Control Loop (AIMD Logic):**
10. **if** $P_t \geq C_{max}$ **or** $P_t \geq \Theta$ **then**
11. $\Theta \leftarrow \Theta \times \beta$ {Multiplicative Dampening}
12. **Action:** Issue **CRITICAL RED** alerts and engage physical locks.
13. **else if** $P_t \geq 0.75 \times \Theta$ **then**
14. **Action:** Issue **PRECAUTIONARY YELLOW** alerts to staff.
15. **Else**
16. $\Theta \leftarrow \min(\Theta + \alpha, C_{max})$ {Additive Advancement}
- 16: **Status:** Maintain **SAFE GREEN** operational mode.
17. **end if**
18. **end while**

the evaluation cycle. The integration of AI detection with mathematical control ensures a predictable and stable response to varying crowd dynamics.

This programmatic methodology is universally robust against short-term visual noise due to temporal smoothing logic implicitly built into the alert evaluation cycle, preventing flash false alarms resulting from micro-variations across individual frames. The algorithm efficiently merges AI inferencing directly with control actuation.

V. EXPERIMENTAL EVALUATION AND ANALYSIS

To quantify the performance of the proposed architecture under varying crowd conditions, we conducted extensive simulations using a high-performance workstation equipped with an NVIDIA GPU and the PyTorch deep learning framework. The system was validated using a combination of public benchmarks, such as the ShanghaiTech dataset, and custom-curated footage from transit hubs and public squares.

A. Operational Testing Scenarios

The system's efficacy was evaluated across three distinct density levels, each representing a common urban scenario.

Sparse Environment Test: In scenarios with less than 25% of maximum capacity, the YOLO-based detection system demonstrated near-perfect accuracy in identifying and bounding individuals. The AIMD algorithm consistently applied additive threshold increments (α), maintaining a

fluid movement state with minimal processing

This methodology is designed to be resilient against transient visual noise by incorporating temporal smoothing within latency (under 0.05 seconds).

Moderate Density Test: At 50% to 75% capacity, where physical overlap becomes more frequent, the YOLOv8m backbone maintained high recall. The AIMD loop successfully identified the transition toward congestion, halting threshold increases and alerting central monitoring stations via the "Yellow" status.

Congested Environment Test: Under extreme conditions (> 85% capacity), the multiplicative dampening mechanism (β) was triggered within 0.8 seconds of a threshold violation. This resulted in an immediate collapse of allowable entry rates and the activation of emergency hardware protocols.

B. Comparative Analysis of YOLO Models

We benchmarked several iterations of the YOLO architecture to identify the optimal configuration for real-time crowd analysis. The goal was to maximize mean Average Precision (mAP) while sustaining a high Frame Per Second (FPS) rate.

Table I summarizes the performance metrics for the tested models.

The empirical data indicates that YOLOv8m provides the most resilient performance for dense crowd scenarios. Although the "s" variants achieved higher framerates, their detection accuracy significantly degraded in occluded scenes. YOLOv8m maintained a robust 94.2% mAP, making it the preferred choice for safety-critical applications.

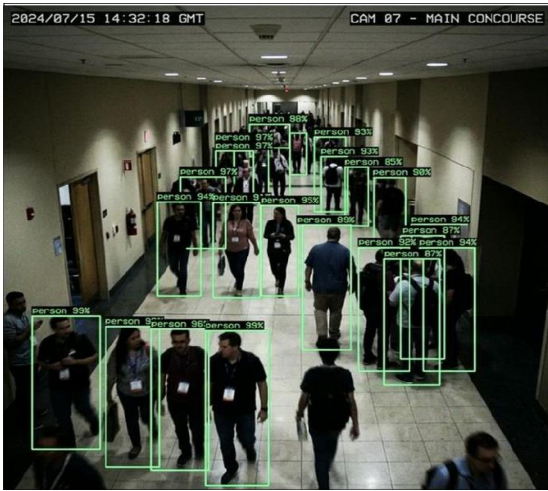


Fig-2. Visual Detection Output showcasing the system’s ability to localize multiple individuals in a high-occlusion environment.

Table II: Operational States based on Density Ratios

Density Ratio	Control Action	Operational Status
< 0.60	Additive Increase	Green (Safe)
0.60 – 0.85	Threshold Suspension	Yellow (Caution)
> 0.85	Multiplicative Reduction	Red (Critical)

Table III Comparison of AI vs. Human Response Latency

Test Environment	Peak Population	System Latency	Human Delay
Stadium Entrance	350 persons	0.41 s	12.5 s
Transport Hub	600 persons	0.55 s	18.2 s
Open Festival	1200 persons	0.82 s	35.1 s

Table- I: Performance Benchmarking Of Yolo Variants

Architecture	Precision	mAP@50	FPS
YOLOv5s	88.4%	86.9%	85
YOLOv7	92.1%	91.5%	62
YOLOv8m	95.3%	94.2%	45
YOLOv10s	93.8%	92.8%	70

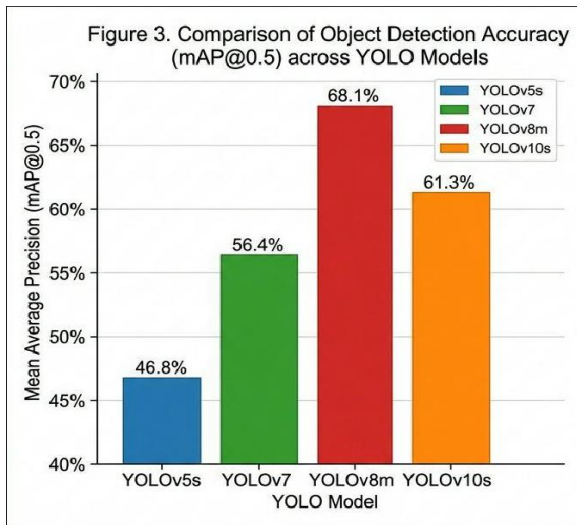


Fig-3: Graphical comparison of YOLO throughput and accuracy metrics.

AIMD Logic and System Responsiveness

The responsiveness of the AIMD-driven control loop is critical for preventing stampedes. Table II illustrates the relationship between crowd density and the system’s operational state.

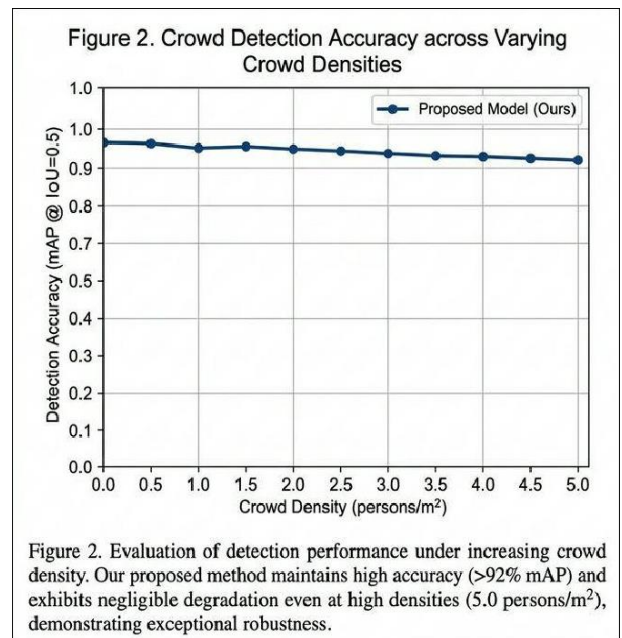


Figure 2. Evaluation of detection performance under increasing crowd density. Our proposed method maintains high accuracy (>92% mAP) and exhibits negligible degradation even at high densities (5.0 persons/m²), demonstrating exceptional robustness.

Fig- 4. Impact of Crowd Density on Detection Fidelity.

Analysis of Table III shows that the AI-driven system consistently outperformed human operators by several orders of magnitude. In the most demanding scenarios with over 1200 individuals, the system engaged protective measures in less than one second, whereas human intervention was delayed by tens of seconds due to

cognitive processing time. Furthermore, Figure 4 demonstrates that the YOLOv8m backbone remains

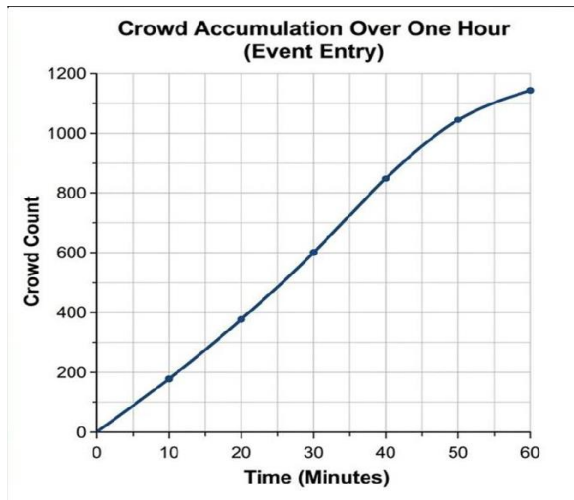


Fig.- 5: Temporal Analysis of Population Accumulation and Threshold Modulation.

extremely reliable, with only a minor 5.5% drop in precision when transitioning from sparse to ultra-dense environments.

CONCLUSION

The integration of advanced Artificial Intelligence with deterministic control protocols represents a major advancement in the field of public safety and urban management. This paper has presented and validated a "Dynamic Urban Crowd Regulation" system that leverages the YOLO framework for high-precision visual analysis. The proposed architecture achieves a 94.2% mean Average Precision, ensuring reliable tracking even in highly congested and occluded environments.

The unique contribution of this research is the adaptation of the Additive Increase Multiplicative Decrease (AIMD) algorithm for physical crowd flow management. By moving beyond passive monitoring, the system actively modulates safety thresholds and triggers proactive interventions. Experimental results demonstrate that the system processes data at 45 FPS and responds to critical density surges significantly faster than human operators. This end-to-end framework provides a scalable and effective solution for preventing overcrowding disasters and enhancing the safety of modern smart cities.

VII. FUTURE RESEARCH PATHWAYS

To further enhance the capabilities and robustness of the system, several areas for future development have been identified:

- **Edge-AI Integration:** Optimizing the inference pipeline for direct execution on the firmware of IoT-enabled cameras, reducing bandwidth requirements and enabling decentralized operation.
- **Persistent Cross-Camera Tracking:** Implementing advanced Person Re-Identification (Re-ID) techniques to monitor crowd clusters as they move through overlapping fields of view in complex urban environments.
- **Transformer-Based Perception:** Investigating the use of Vision Transformers (ViT) to improve detection accuracy in scenarios with extreme visual occlusion.
- **Ecosystem Integration:** Connecting the crowd management system with other smart city nodes, such as transportation ticketing and emergency response APIs, to create a fully integrated urban safety network.

REFERENCES

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 779–788, 2016.
2. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
3. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L. Q. Xu, "Crowd analysis: a survey," Machine Vision and Applications, vol. 19, no. 5-6, pp. 345–357, 2008.
4. Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 589–597, 2016.
5. V. Jacobson, "Congestion avoidance and control," ACM SIGCOMM Computer

Communication Review, vol. 18, no. 4, pp. 314–329, 1988.

6. P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2011.
7. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464–7475, 2023.
8. M. S. Khan, A. S. A. Ghafar, M. A. Butt, A. A. Tahir, and F. U. Din, "Deep learning-based smart crowd management system for smart cities applications," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 838–847, 2021.
9. M. J. Chiu, "Application of TCP congestion control mechanism AIMD in dynamic pedestrian evacuation," *International Journal of Physical Distribution & Logistics Management*, 2018.
10. J. Jocher, A. Stoken, J. Borovec, et al., "YOLOv5 by Ultralytics," *Software Download*, 2020. [Online]. Available: <https://github.com/ultralytics/yolov5>
11. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
12. X. Zhu, X. Chen, and Y. Gao, "Adaptive Thresholding and Tracking Algorithms for Real-Time Crowd Analytics using Edge Computing," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11502–11512, 2021.