International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05317
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Real-Time Crowd Density Monitoring Using YOLOv8 and Deterministic Logic for Public Safety Applications

Bathala Harsha[1], Deepankar Ambarish[2], Chethan G[3], K Raghava Aryan [4] & Dr.G. Nagendra babu[5]

[1 2 3 4 5 6] *Department of Computer Science and Engineering, JAIN (Deemed-to-be-University)*

Abstract - Monitoring crowd density in real time is critical for preventing congestion-related hazards in high-traffic public environments such as streets, transit hubs, and large gathering areas. Conventional surveillance systems largely depend on manual observation, which is prone to delayed responses and subjective judgment. This paper presents the **Crowd Density Monitoring System (CDMS)**, an autonomous and edge-optimized framework designed for real-time situational awareness and proactive crowd safety management. The proposed system employs **YOLOv8-based person detection** for accurate human localization and integrates **optical flow analysis** to capture motion intensity within the scene. A **deterministic logic engine** combines crowd density and motion cues to compute a normalized **congestion index on a 0–100 scale**, enabling transparent and interpretable risk assessment without reliance on black-box prediction models. The framework further identifies spatial hotspots and classifies crowd conditions into normal, warning, and critical states, triggering automated operational directives and email alerts under hazardous conditions. Experimental evaluation using controlled surveillance video scenarios demonstrates that CDMS operates reliably in real time with low latency on GPU-enabled edge hardware, effectively detecting critical crowding events and supporting timely intervention. The proposed system highlights the practicality of deterministic, explainable analytics for real-world crowd monitoring and public safety applications.

Key Words: Crowd Density Monitoring, YOLOv8, Deterministic Logic Engine, Optical Flow, Real-Time Surveillance, Edge Computing

1.INTRODUCTION

Rapid urbanization, large-scale public gatherings, and increasing concerns regarding public safety have made crowd density monitoring a critical research problem in modern intelligent surveillance systems. Accurate estimation of crowd density and real-time people counting play a vital role in preventing stampedes, managing congestion, optimizing infrastructure usage, and ensuring safety during large events. Traditional crowd monitoring methods based on manual observation or simple sensor-based systems are often inefficient, error-prone, and incapable of handling complex, dynamic environments, particularly in high-density scenarios [1].

With the advancement of computer vision and deep learning, automated crowd analysis systems have gained significant attention. Early approaches relied on handcrafted features and classical techniques such as background subtraction and optical flow to estimate crowd movement and density [5]. While these methods performed reasonably well in controlled environments, they struggled in real-world scenarios involving occlusions, scale variations, illumination changes, and highly dense crowds. These limitations motivated the adoption of Convolutional Neural Networks (CNNs), which demonstrated superior capability in learning robust visual representations directly from data [4].

Recent research has explored two dominant paradigms for crowd analysis: detection-based methods and density-estimation-based methods. Detection-based approaches, often built upon object detection frameworks such as YOLO, perform well in sparse to medium-density crowds but degrade significantly in highly congested scenes due to severe occlusions [9]. Conversely, density estimation methods predict crowd density maps and are more effective in dense environments; however, they often lack spatial precision and struggle with real-time deployment, especially on edge or embedded systems [6], [8]. Hybrid frameworks that combine detection and density estimation have been proposed to address these issues; however, many remain computationally expensive or require complex architectures, limiting their suitability for real-time and edge-based applications [1].

Another critical limitation observed in existing literature is the lack of temporal consistency in crowd counting systems. Frame-wise processing ignores motion patterns and temporal correlations, leading to unstable predictions across video frames. Several studies have introduced recurrent or attention-based temporal modeling to mitigate this problem, but such models significantly increase computational complexity and memory requirements, making them impractical for deployment in real-world surveillance systems [2], [3], [13]. Additionally, many state-of-the-art methods rely heavily on deep and complex architectures or extensive pretraining, which further restricts their usability in low-resource environments [7]. In contrast, lightweight motion cues such as optical flow can provide temporal awareness without the overhead of recurrent or attention-based architectures.

Recent surveys and benchmarking studies highlight the growing demand for lightweight, real-time, and deployment-ready crowd monitoring systems, particularly for smart cities, transportation hubs, and edge-based surveillance applications [4], [10], [15]. Despite notable progress, achieving a balance between accuracy, computational efficiency, and real-time performance remains an open research challenge. Most existing solutions either prioritize accuracy at the expense of speed or sacrifice robustness for efficiency, leaving a clear gap for practical, scalable crowd density monitoring systems [6], [8].

Motivated by these limitations, this project focuses on developing a real-time crowd density monitoring and people counting system that emphasizes computational efficiency, robustness under varying crowd densities, and practical deployability. By leveraging modern deep learning-based detection architectures and optimizing them for real-time performance, the proposed system aims to overcome the shortcomings of traditional and existing deep learning approaches. The system is designed to function effectively across diverse crowd scenarios, including sparse, medium, and high-density environments, while maintaining stable and reliable performance suitable for real-world surveillance applications. Unlike end-to-end black-box learning approaches, the proposed system emphasizes a deterministic decision-making pipeline that

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05317
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

combines learned perception with transparent rule-based analytics.

The contributions of this project align with current research directions by addressing key gaps identified in recent IEEE and high-impact journal publications. Specifically, the proposed approach focuses on reducing computational overhead, improving counting stability, and enabling real-time execution without reliance on overly complex architectures or computationally intensive temporal models. As a result, this work contributes a practical and research-driven solution that advances the applicability of crowd density monitoring systems in real-world safety-critical environments.

2. Literature Survey

J. Liu et al. propose DecideNet, a hybrid crowd counting framework that integrates detection-based and density-estimation-based approaches using an attention-guided mechanism to handle varying crowd densities effectively. The model employs a detection branch for sparse regions and a density estimation branch for congested areas, with an attention module dynamically balancing their contributions. Evaluated on benchmark datasets such as ShanghaiTech and UCF_CC_50, DecideNet demonstrated improved robustness across diverse crowd scenarios compared to single-branch models. However, the architecture is computationally heavy and not optimized for real-time deployment on edge or surveillance systems, limiting its applicability in continuous monitoring environments. This highlights the need for lightweight and real-time crowd density solutions suitable for practical deployments[1].

M. Ling et al. present a motional foreground attention-based video crowd counting method that leverages motion cues to distinguish dynamic crowd regions from static backgrounds. The approach integrates motion-aware attention modules to enhance feature representation for crowd regions, leading to improved counting accuracy in complex scenes. Tested on large-scale video datasets, the method achieves notable gains over baseline CNN models. However, the dependence on motion estimation introduces additional computational overhead and sensitivity to camera motion and noise, which can degrade performance in real-world surveillance scenarios. This emphasizes the need for robust yet computationally efficient crowd density estimation frameworks[3].

Y. Hou et al. introduce a Frame-Recurrent Video Crowd Counting (FRVCC) framework that explicitly models temporal correlations across consecutive video frames. By incorporating a recurrent structure, the model effectively reduces temporal inconsistency and improves stability in video-based crowd counting tasks. Experiments conducted on video crowd datasets show that FRVCC outperforms frame-independent methods in terms of accuracy and temporal smoothness. Despite its strong performance, the reliance on recurrent processing increases computational complexity and inference latency, making it less suitable for real-time applications where low-latency responses are critical. These limitations motivate the exploration of more efficient temporal modeling strategies for real-time crowd monitoring systems[2].

Z. Fan et al. provide a comprehensive survey of CNN-based crowd counting and density estimation methods, categorizing approaches into detection-based, regression-based, and density-map-based techniques. The survey critically analyzes challenges such as severe occlusion, scale variation, perspective distortion, and real-time constraints. While CNN-based methods significantly outperform traditional approaches, the survey identifies a persistent trade-off between accuracy and computational efficiency, especially for real-time and edge-based deployments. The authors highlight the lack of unified solutions that balance precision, speed, and scalability, establishing a clear research gap addressed by modern lightweight crowd monitoring systems[4].

R. Wang et al. introduce an efficient crowd counting framework using dual knowledge distillation, where a compact student network learns from both density and feature representations of a larger teacher model. This approach significantly reduces model size while retaining competitive accuracy. Experiments on standard datasets confirm improved efficiency compared to baseline CNNs. However, the dependency on pre-trained teacher networks and complex training pipelines increases system complexity, posing challenges for rapid deployment and scalability in real-time surveillance systems[6].

A. Alhawwary *et al.* propose **PatchFlow**, a lightweight optical flow estimation framework designed to reduce computational cost while maintaining reasonable motion estimation accuracy. The two-stage patch-based design enables faster inference compared to traditional dense optical flow algorithms. Although PatchFlow demonstrates efficiency improvements, optical-flow-based crowd analysis methods remain sensitive to illumination changes and complex motion patterns. Moreover, such methods struggle in highly congested scenes where individual motion becomes ambiguous, limiting their effectiveness for dense crowd monitoring applications[5].

M. A. Khan et al. focus on crowd counting at the edge using weighted knowledge distillation, aiming to deploy accurate models on resource-constrained devices. The proposed method balances accuracy and efficiency by selectively distilling important features. While the results show promise for edge-based crowd analysis, the approach still relies on sophisticated training strategies and does not fully address challenges such as real-time responsiveness and integration with complete surveillance pipelines. This reinforces the need for end-to-end lightweight systems for practical crowd monitoring[7].

R. Chavan *et al.* present **CrowdDCNN**, a deep convolutional neural network optimized for real-time crowd counting on IoT edge devices. The model achieves competitive accuracy while significantly reducing inference time, making it suitable for embedded environments. However, the approach primarily focuses on count estimation and does not incorporate advanced spatial or temporal analysis, such as crowd density heatmaps or movement patterns, which are crucial for proactive crowd management and safety applications[8].

G. Gao et al. provide an extensive survey on CNN-based density estimation and crowd counting, summarizing datasets, evaluation metrics, and methodological advancements. The survey highlights the superiority of density-map-based methods in dense crowds but also notes their high computational cost and limited real-time applicability. The authors emphasize the need for efficient architectures capable of maintaining accuracy while supporting real-time deployment, particularly in large-scale surveillance systems[10].

Z. Gündüz et al. investigate YOLO-based crowd detection methods, analyzing the performance of different YOLO variants for real-time crowd detection from video streams. The study demonstrates that YOLO models offer high processing speeds and are effective in sparse to moderately dense crowds. However, detection-based approaches tend to underperform in highly congested scenes due to severe occlusion and overlapping individuals. This limitation motivates hybrid or density-based approaches for accurate crowd analysis in dense environments[9].

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05317
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Q. Wang et al. present a comprehensive survey and benchmark on frame-level temporal modeling for video crowd counting, analyzing how temporal information improves robustness and consistency. While temporal models enhance performance, the survey concludes that many existing approaches are unsuitable for real-time systems due to high computational demands. This observation underscores the importance of designing efficient crowd monitoring systems that balance temporal awareness and computational feasibility[11].

L. Lyu et al. propose a temporal fusion framework with motional foreground attention to enhance video crowd counting performance. The method improves density estimation by fusing spatial and temporal features. Despite improved accuracy, the increased architectural complexity and processing requirements limit scalability and real-time performance in continuous monitoring scenarios, reinforcing the demand for simpler yet effective solutions[12].

Recent review studies on lightweight crowd counting for embedded and edge systems highlight that although many approaches attempt to reduce model size, most still compromise either accuracy or robustness. The surveys conclude that there is a lack of unified systems that integrate real-time processing, spatial density visualization, and scalable deployment, particularly for smart city and public safety applications[14].

R. Ma *et al.* further extend the FRVCC framework by emphasizing explicit temporal correlation modeling to improve robustness under challenging conditions such as camera shake and illumination changes. While the model achieves strong performance, its recurrent design increases latency, making it less suitable for real-time crowd monitoring applications deployed on edge devices[13].

B. D. Sunil et al. present a recent overview of density estimation and crowd counting techniques, emphasizing emerging challenges such as real-world deployment, scalability, and adaptability to dynamic environments. The study stresses the need for integrated systems that go beyond counting accuracy to support actionable insights such as congestion detection and crowd behavior analysis, aligning closely with modern crowd density monitoring objectives[15].

3. Proposed Work

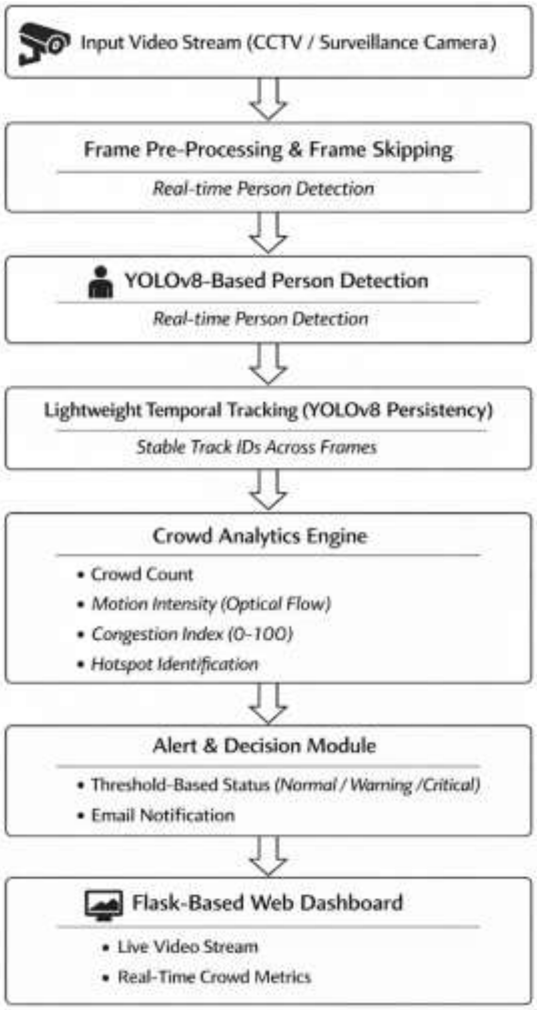


Fig -1: Architecture of the proposed crowd density monitoring system

The proposed crowd density monitoring system follows a structured and end-to-end pipeline designed to support real-time surveillance, accurate people counting, and actionable crowd analysis in practical environments such as public spaces, transportation hubs, and large gatherings. The workflow begins with continuous video acquisition from fixed surveillance cameras, where each frame is processed in a streaming manner with lightweight temporal cues to support low-latency performance. Unlike computationally intensive density-estimation-only methods or heavily recurrent architectures, the proposed approach emphasizes efficiency, robustness, and deployability on standard surveillance hardware.

For person detection, the system employs the YOLOv8 object detection model, selected for its strong balance between detection accuracy and real-time inference speed. YOLOv8 processes each incoming video frame and produces precise bounding boxes for detected individuals, even under moderate occlusion and varying illumination conditions. This detection-based approach enables accurate localization of people in sparse to moderately dense scenes while maintaining high frame rates, which is essential for continuous monitoring applications.

To ensure consistent identification of individuals across consecutive frames and to avoid duplicate counting, the detected bounding boxes are passed to the YOLOv8 integrated tracking module, which associates detections across consecutive frames

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05317
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

using motion consistency to maintain persistent identities. This enables stable person localization across consecutive frames and reduces duplicate detections during continuous monitoring. Short-lived or noisy tracks are filtered to improve tracking stability and counting reliability, particularly in challenging surveillance scenarios.

Crowd density estimation is performed using a detection-driven analytical strategy rather than pixel-wise density regression networks. The system computes crowd concentration by aggregating detected person counts and motion intensity over time, enabling efficient identification of congested conditions without the overhead of dense regression-based models.

In addition to detection-based crowd counting, the system computes real-time crowd statistics including total occupancy, congestion index, crowd status, and spatial hotspot identification. These metrics are continuously updated and streamed to a web-based dashboard built using the Flask framework.

The overall architecture of the proposed crowd density monitoring system, including detection, tracking, and deterministic analytics integration, is illustrated in Fig. 1. By combining fast object detection with efficient tracking and spatial density analysis, the proposed method addresses several limitations identified in prior work, such as high computational overhead, lack of real-time responsiveness, and absence of integrated visualization and analytics. The resulting system offers a practical, scalable, and research-driven solution for real-world crowd density monitoring and public safety applications.

4. Results and Discussions

The proposed Crowd Density Monitoring System (CDMS) was evaluated using two controlled video scenarios designed to represent distinct real-world crowd conditions: a normal-flow scenario (sample.mp4) and a critical congestion scenario (sample1.mp4). Both videos were processed under identical system configurations to ensure consistency in evaluation. The assessment focuses on real-time occupancy estimation, congestion index behavior, system status classification, and alert triggering performance, which collectively reflect the system's effectiveness in practical surveillance environments.

In the normal crowd scenario, the system demonstrated stable and non-intrusive behavior. An average occupancy of 31.87 individuals was recorded, with the peak occupancy reaching 40 individuals. Despite the relatively high number of people present, the average congestion index remained at 28.87, while the peak congestion index reached 38.9, both remaining below the predefined warning threshold. Consequently, the system consistently classified the situation as NORMAL, and no alerts were triggered. This outcome confirms that the proposed system does not rely solely on absolute crowd size but instead evaluates congestion based on a balanced combination of density and motion cues, thereby avoiding unnecessary false alarms in orderly crowd scenarios.

In contrast, the critical crowd scenario exhibited significantly different system behavior. Although the average occupancy was lower at 10.31 individuals, spatial constraints and increased motion intensity led to a substantially higher average congestion index of 43.02, with a peak congestion index of 58.0. Upon exceeding the critical threshold, the system correctly classified the situation as CRITICAL and triggered a single automated alert. This result highlights a key advantage of the proposed approach: hazardous crowd conditions are detected based on crowd dynamics rather than raw population count alone. Such behavior is essential in real-world safety scenarios, where dangerous congestion can arise even with relatively fewer individuals due to bottlenecks or sudden movement restrictions.

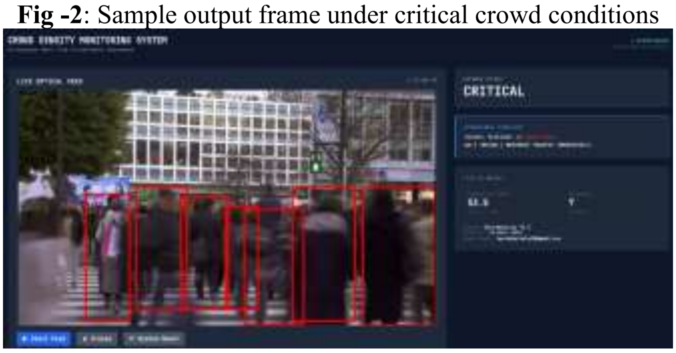
Scenario	Average Occupancy	Peak Occupancy	Average Congestion Index	Peak Congestion Index	Final Status	Alerts Triggered
Normal (sample.mp4)	31.87	40	28.87	38.9	NORMAL	0
Critical (sample1.mp4)	10.31	20	43.02	58.0	CRITICAL	1

Table 1: Quantitative crowd analytics for normal and critical scenarios

These results demonstrate that the proposed congestion index effectively captures hazardous crowd dynamics that are not reflected by occupancy count alone.

Visual validation further supports the numerical findings. Fig. 2 illustrates a representative output frame from the critical scenario, showing real-time person detection and tracking under congested conditions. The bounding boxes and system status overlay confirm that the detection and tracking pipeline operates reliably even in partially occluded environments. Fig. 3 presents the operational dashboard during runtime, displaying live telemetry such as congestion index, occupancy, system status, and recommended actions. This integration of visual feedback with quantitative analytics enhances situational awareness and supports timely decision-making by operators.

Fig -2: Sample output frame under critical crowd conditions



To further contextualize the proposed system, Table 2 provides a qualitative comparison between CDMS and representative categories of existing crowd monitoring approaches discussed in the literature.

Table 2: Qualitative comparison of the proposed CDMS with existing crowd monitoring approaches

Feature	Detection-based Methods [9]	Density-estimation Methods [1], [3]	Edge-based CNN Systems [8]	Proposed CDMS
Real-time processing	✓	✗	✓	✓
Edge deployment support	✗	✗	✓	✓
High-density robustness	✗	✓	◐	✓
Motion-aware analysis	✗	◐	✗	✓
Deterministic decision logic	✗	✗	✗	✓
False alert suppression	✗	✗	◐	✓
Integrated alert system	✗	✗	◐	✓
Operational dashboard	✗	✗	◐	✓

Legend: ✓ Supported, ◐ Partially Supported, ✗ Not Supported

The comparative analysis demonstrates that while existing approaches focus primarily on detection accuracy or density estimation, they often lack deployability, deterministic decision-making, and integrated alert mechanisms. In contrast, the proposed CDMS delivers an end-to-end solution that combines real-time perception, motion-aware analytics, deterministic logic, and operational visualization. These characteristics make the system particularly suitable for real-world crowd safety applications where reliability, interpretability, and low latency are critical.

5. CONCLUSIONS

This work presented the design and implementation of an autonomous Crowd Density Monitoring System (CDMS) aimed at improving public safety in high-traffic environments through real-time visual analytics. The proposed system integrates YOLOv8-based person detection with a deterministic logic engine and motion-aware analysis to quantify crowd congestion and identify hazardous conditions proactively. Unlike traditional surveillance approaches that rely on manual observation or static thresholds, CDMS evaluates crowd dynamics using a congestion index derived from both spatial density and motion intensity, enabling reliable detection of unsafe crowd behavior in real-world scenarios.

Experimental evaluation using controlled video scenarios demonstrated the effectiveness of the proposed approach. In the normal-flow scenario, the system maintained a stable classification without triggering unnecessary alerts, despite higher absolute occupancy. In contrast, the critical scenario was correctly identified as hazardous based on elevated congestion index values, even with a lower number of individuals present. This behavior confirms that the system does not depend solely on raw crowd counts, but instead captures underlying crowd dynamics that are more indicative of safety risks. The successful triggering of alerts only under genuinely critical conditions highlights the robustness of the deterministic decision logic and its ability to suppress false positives.

Furthermore, the system's edge-optimized design enables real-time performance on GPU-enabled hardware while maintaining low latency and operational reliability. The integration of a live dashboard and automated email alert mechanism enhances situational awareness and supports timely decision-making for crowd management personnel. Overall, the results validate that CDMS provides a practical, interpretable, and deployment-ready solution for real-time crowd density monitoring, addressing key limitations identified in existing detection-based and density-estimation-based approaches.

6. FUTURE ENHANCEMENTS

While the proposed CDMS demonstrates strong performance in controlled scenarios, several enhancements can be explored to further improve its applicability and robustness. One immediate extension involves incorporating multi-camera support with cross-view tracking to handle larger environments and reduce blind spots caused by occlusions or camera placement limitations. This would allow more accurate crowd flow analysis across interconnected spaces such as transit hubs or large public venues.

Future work may also focus on advanced spatial visualization techniques, such as adaptive heatmap generation or region-based congestion forecasting, to provide more intuitive representations of crowd behavior. Enhancing the motion analysis component with more robust optical flow techniques or learning-based motion models could further improve sensitivity to subtle crowd movements in highly congested environments.

Additionally, extending the alerting mechanism to support multi-channel notifications, such as SMS, control room dashboards, or integration with traffic and access control systems, could improve response coordination during critical events. Long-term deployment and validation using real-world surveillance data from diverse environments would also help assess system scalability, robustness under varying lighting conditions, and resilience to environmental noise.

Ultimately, incorporating predictive analytics to anticipate congestion trends before critical thresholds are reached could transform CDMS from a reactive monitoring system into a proactive crowd safety management platform. These enhancements would further strengthen the system's role in smart city infrastructure and large-scale public safety applications.

REFERENCES

1. J. Liu, C. Gao, D. Meng and A. G. Hauptmann, "DecideNet: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 5197–5206.
2. Y. Hou, S. Zhang, R. Ma, H. Jia and X. Xie, "Frame-Recurrent Video Crowd Counting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, 2023, doi: 10.1109/TCSVT.2023.3250946.
3. M. Ling, T. Pan, Y. Ren, K. Wang and X. Geng, "Motional foreground attention-based video crowd counting," *Pattern Recognition*, vol. 144, Art. 109891, 2023, doi: 10.1016/j.patcog.2023.109891.
4. Z. Fan, H. Zhang, Z. Zhang, G. Lu and Y. Wang, "A survey of crowd counting and density estimation based on convolutional neural network," *Neurocomputing*, vol. 472, pp. 224–251, 2022, doi: 10.1016/j.neucom.2021.02.103.
5. A. Alhawwary, J. Mustaniemi and J. Heikkilä, "PatchFlow: A Two-Stage Patch-Based Approach for Lightweight Optical Flow Estimation," in *Computer Vision – ACCV 2022*, Lecture Notes in Computer Science, vol. 13847, pp. 535–551, Springer, 2023, doi: 10.1007/978-3-031-26293-7_32.
6. R. Wang, Y. Hao, L. Hu, X. Li, M. Chen, Y. Miao and I. Humar, "Efficient Crowd Counting via Dual Knowledge Distillation," *IEEE Transactions on Image Processing*, vol. 33, pp. 569–583, 2024, doi: 10.1109/TIP.2023.3343609.
7. M. A. Khan, H. Menouar and R. Hamila, "Crowd Counting at the Edge Using Weighted Knowledge Distillation," *Scientific Reports*, vol. 15, Art. 11932, 2025, doi: 10.1038/s41598-025-90750-5.

ISSN: 2583-6129
DOI: 10.55041/ISJEM05317

8. R. Chavan, A. Kanamarlapudi, G. Rani, P. Thakkar and V. S. Dhaka, "CrowdDCNN: Deep Convolution Neural Network for Real-Time Crowd Counting on IoT Edge," *Engineering Applications of Artificial Intelligence*, vol. 126, Art. 107089, 2023, doi: 10.1016/j.engappai.2023.107089.
9. Z. Gündüz and G. Işık, "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," *Journal of Real-Time Image Processing*, 2023, doi: 10.1007/s11554-023-01276-w.
10. G. Gao, J. Gao, Q. Liu, Q. Wang and Y. Wang, "CNN-based Density Estimation and Crowd Counting: A Survey," *arXiv preprint* arXiv:2003.12783, 2020.
11. Q. Wang, J. Gao and Y. Yuan, "Frame-level Temporal Modeling for Video Crowd Counting: A Survey and Benchmark," *IET Intelligent Transport Systems* (comprehensive survey/benchmark), 2024–2025 (online), doi: 10.1049/ipr2.13328. *(survey used for methodology & benchmarks)*
12. L. Lyu, W. Wang and L. Zhang, "Motional foreground attention and temporal fusion for video crowd counting," *Pattern Recognition Letters / Related Conf.* (see Pattern Recognition 2023), 2023, doi: 10.1016/j.patrec.2023.109891.
13. R. Ma, Y. Hou, C. Li, H. Jia and X. Xie, "Frame-Recurrent Video Crowd Counting (FRVCC): explicit temporal correlation modeling for robust video crowd counting," *IEEE Trans. Circuits Syst. Video Technol.*, 2023, doi: 10.1109/TCSVT.2023.3250946.
14. "Efficient and Lightweight Crowd Counting for Embedded and Edge Systems: A Review and Comparative Study," *Applied Soft Computing / Expert Systems with Applications*, 2023–2025 (survey papers on edge deployment and lightweight models), doi: 10.1016/j.eswa.2022.117823.
15. B. D. Sunil, R. Venkatesh and S. Todmal, "Density Estimation and Crowd Counting," *arXiv preprint* arXiv:2511.09723, 2025.