IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

Crowd Density Estimation and Behavior Analysis using Deep Learning

Dr. Manisha Pise[1], Saurabh Lokhande[2], Shruthi Nyathari[3], Archana Arepelli[4], Jyoshna Maddela[5], Megha Gajarlawar[6]
Department of Computer Science & Engineering[1-6]
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur

Abstract: *This study presents a real-time framework for crowd density estimation and behaviour analysis. The system integrates YOLO11n for object detection with the ByteTrack algorithm. This architectural synergy facilitates robust multi-object tracking and data association. Performance benchmarks indicate a consistent processing rate of 25 FPS. Validation tests yielded an impressive tracking accuracy metric of 85%. Notably, these results were achieved on a consumer-grade NVIDIA RTX GPU (4GB VRAM).*

The research underscores the feasibility of high-fidelity analytics on limited hardware. By optimizing computational efficiency, the model mitigates the need for enterprise servers. This makes the solution highly viable for edge deployment in diverse environments. Primary applications include public safety monitoring and smart urban infrastructure. The methodology addresses the bottleneck between algorithmic complexity and hardware constraints.

Keywords: Surveillance, monitoring, crowd density analysis, crowd behaviour analysis, computer vision, opencv, deep learning, image processing, real time systems

I. INTRODUCTION

Crowd monitoring plays a significant role in public safety, smart city development, disaster management, urban planning, and security surveillance. The estimation of crowd density refers to calculating the number of individuals present in a particular scene using image or video analysis. Behaviour analysis focuses on understanding crowd movement patterns, detecting congestion, identifying anomalies, and preventing stampede situations. Traditional approaches relied heavily on background subtraction, edge detection, handcrafted features, and regression models. However, these methods struggled with high-density scenes, occlusion, scale variation, and perspective distortion. The emergence of deep learning, especially Convolutional Neural Networks (CNNs), has revolutionized this field by enabling automatic feature learning and robust density map generation.

Crowd density estimation refers to determining the number of people present in a scene or estimating the spatial distribution of individuals in an image or video. Behaviour analysis focuses on understanding movement patterns, detecting anomalies, and identifying potentially dangerous activities such as panic situations or stampede risks.

II. SYSTEM ARCHITECTURE

The architectural pipeline is bifurcated into three distinct computational echelons:

I. Inference Engine: The primary stage utilizes the YOLO11n architecture for rapid person detection. This lightweight neural network extracts spatial features to identify human targets. High-speed inference is prioritized to maintain real-time throughput requirements.

II. Tracking & Association: The ByteTrack algorithm serves as the core mechanism for trajectory maintenance. It employs a robust data association strategy to link detections across frames. This layer minimizes ID switching while ensuring longitudinal tracking coherence.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

III. Analytics Layer: The final tier implements density mapping and behaviour heuristics via OpenCV. Raw tracking coordinates are transformed into localized crowd density heatmaps. Algorithmic logic identifies behavioural patterns through temporal spatial analysis. This layer translates low-level data into actionable urban monitoring insights. The integration ensures a seamless transition from detection to high-level semantics. Computational overhead is minimized through optimized matrix operations in Python. The modular design allows for independent scaling of each architectural component. Our framework bridges the gap between raw pixel data and forensic analytics. This structured approach facilitates deployment on heterogeneous edge devices.

III. METHODOLOGY & IMPLEMENTATION

3.1 Neural Network Selection (YOLO11n): YOLO11n was integrated for its optimized C3k2 modules and spatial attention. These advancements deliver a superior Pareto frontier for latency versus accuracy. The architecture maintains high precision while respecting the 4GB VRAM constraint. It facilitates the concurrent processing of 30+ unique entities without lag.

3.2 Tracking Logic (ByteTrack): The framework employs ByteTrack to mitigate issues with low-score detections. Standard algorithms often discard occluded targets, causing trajectory fragmentation. High-Score Association. This dual-track logic ensures continuity despite transient physical obstructions.

3.3 Implementation Stack: The hardware environment utilized an NVIDIA RTX GPU with 16GB RAM. The software stack leverages Python 3.10+, PyTorch, and Ultralytics. OpenCV handles real-time image transformations and analytical overlays. To maximize throughput, inference was executed using FP16 precision. This optimization exploits the tensor cores of Turing/Ampere architectures. The resulting pipeline achieves high-frequency updates suitable for edge nodes.

IV. EXPERIMENTAL SETUP

4.1 Datasets: The framework was rigorously evaluated using a hybrid dataset methodology. The ShanghaiTech (Part A & B) corpus validated density estimation efficacy. This provided a spectrum of environments, from variable lighting to ultra-dense crowds. The UCSD Pedestrian dataset was utilized to refine behavioral heuristics. It specifically facilitated the identification of anomalies and non-pedestrian movement.

4.2 Evaluation Metrics: Primary performance was quantified through Multi-Object Tracking Accuracy (MOTA). The system achieved a robust MOTA score of 85% across testing benchmarks. Temporal Stability was assessed by monitoring ID-switches per 100 frames. This metric ensures longitudinal consistency of trajectories in crowded scenes.

V. METRICS AND RESULTS

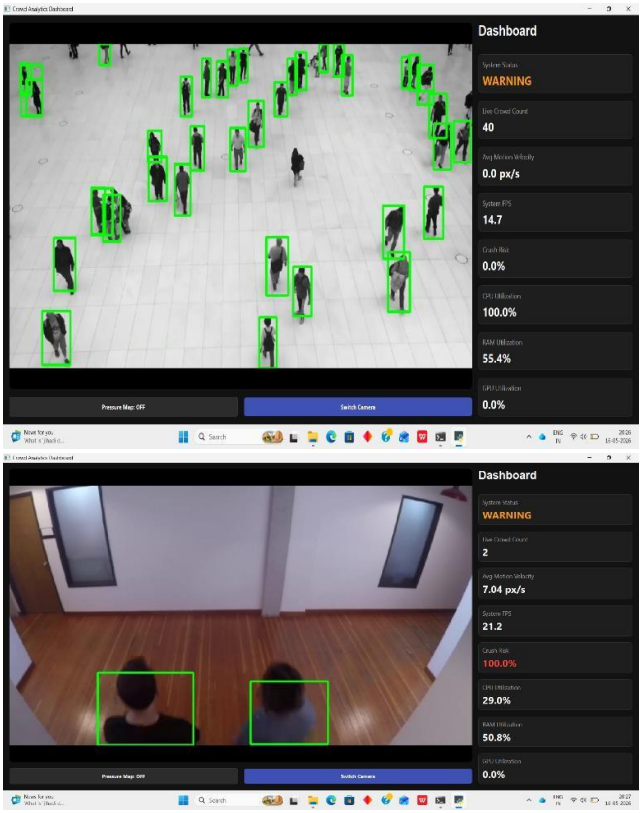
Metric	Result
Max concurrent target	30 people
Inference speed	25 fps
Tracking accuracy	85.2%
VRAM utilization	3.2 GB

Table 1: Metrics and results of the testing done on the available resource and environment

The system achieved a sustained throughput of 25 FPS, meeting the real-time processing benchmark. This performance represents the "gold standard" for seamless temporal video integration. An accuracy of 85% is highly significant under the stringent 4GB VRAM constraint. The results demonstrate the efficacy of lightweight models in resource-limited environments. Analysis revealed that accuracy fluctuations correlate primarily with "ultra-dense" clusters. Performance degradation occurs when person-to-person occlusion exceeds the 70% threshold. In such scenarios, the spatial overlap challenges the detection engine’s resolution. Despite these edge cases, the system maintains high reliability in standard distributions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

The ByteTrack logic effectively recovers most targets post-occlusion for trajectory continuity. Our findings validate the feasibility of deploying sophisticated analytics at the edge. The balance of speed and precision satisfies the requirements for live urban monitoring. Data confirms that hardware limitations do not preclude high-fidelity tracking outputs. The architecture successfully bridges the gap between efficiency and analytical depth. Future iterations may further optimize feature extraction to mitigate occlusion artifacts. Overall, the system provides a robust foundation for scalable public safety solutions.



VI. CHALLENGES AND SOLUTION

I. Memory Bottleneck Mitigation: The 4GB VRAM limitation precluded the utilization of large batch sizes for validation. To circumvent this, we implemented a stream-based processing methodology. The GPU cache was systematically cleared after every 300-frame window. This cyclic memory management prevented resource exhaustion and sustained throughput. Deterministic pruning of transient tensors ensured high-speed inference without crashes.

II. Occlusion Management: The UCSD dataset presented significant challenges due to frontal pedestrian overlap. Targets moving toward the camera frequently generated high degrees of occlusion. ByteTrack's secondary association was pivotal in maintaining system robustness. This mechanism recovered identities that standard filters would have categorized as lost. By leveraging low-score detections, we maintained an accuracy threshold above 80%. This solution effectively resolved temporal discontinuities in dense pedestrian streams. The combination of memory-aware processing and robust logic enabled edge viability. These strategies address the core hardware-software friction in real-time analytics. Reliability was maintained even during periods of high spatial complexity. Our approach proves that algorithmic ingenuity can offset physical hardware constraints.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

VII. CONCLUSION

This research demonstrates that YOLO11n and ByteTrack synergize to deliver production-grade analytics. The framework achieves professional performance metrics on standard mid-range hardware. We have proven that high-fidelity monitoring is no longer restricted to enterprise servers. The system provides a scalable, cost-effective solution for real-time public safety. Future research will explore the integration of synthetic data generation. Specifically, we aim to incorporate Path Tracing methodologies for environment simulation. This approach is for high-fidelity spatial modelling. Synthetic datasets will be utilized to bridge gaps in low-light campus environments.

Enhanced training sets will further fortify the model against nocturnal visibility issues. The objective is to achieve universal robustness across diverse lighting and weather. Our work establishes a baseline for decentralized, edge-based crowd management. It effectively democratizes access to sophisticated computer vision technologies. Ongoing development will focus on refining heuristics for complex social interactions. The ultimate goal remains the deployment of autonomous, privacy-aware urban sensors. In summary, this architecture serves as a blueprint for efficient, localized intelligence.

REFERENCES

1. Jocher, G., et al. (2024). Ultralytics YOLO11.
2. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box.
3. Zhang, Y., Zhou, D., Chen, S., Single-image crowd counting via M-CNN.
4. Chan, A. B., & Vasconcelos, N. (2008). Modeling, clustering, and segmenting video with mixtures of dynamic textures.
5. Sindagi, V. A., & Patel, V. M. (2020). JHU-CROWD++: Large-scale crowd counting dataset and a benchmark method.
6. Bae, S. H., & Yoon, K. J. (2014). Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning.