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: *Crowd density estimation and behaviour analysis are critical components of modern intelligent surveillance systems. With increasing urbanization, large-scale public gatherings, transportation hubs, and religious events require automated monitoring systems capable of accurately estimating crowd size and detecting abnormal behaviour. This review paper presents a comprehensive analysis of five major research works ranging from classical optical flow-based tracking to modern deep learning-based convolutional neural network (CNN) models. The paper discusses detectionbased methods, regression approaches, density map estimation techniques, Bayesian loss models, and hybrid architectures. Furthermore, comparative analysis of datasets, evaluation metrics, strengths, limitations, and real-world applications are presented. The study highlights the evolution of crowd analysis from handcrafted feature methods to robust deep neural architectures suitable for real-time monitoring. Finally, open challenges and future research directions are discussed.*

Keywords: *Crowd density*

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. Together, these systems form intelligent surveillance frameworks capable of realtime monitoring.

Earlier approaches relied on traditional computer vision techniques such as background subtraction, optical flow, and handcrafted feature extraction. However, these methods faced challenges in handling occlusion, scale variation, illumination changes, and extremely dense crowds. The advancement of deep learning, particularly Convolutional

II. LITERATURE REVIEW

Neural Networks (CNNs), significantly improved the performance of crowd counting systems by enabling automatic feature learning and density-map regression.

This review integrates five major research works covering classical crowd density analysis, CNN-based real-time monitoring, advanced hybrid deep learning frameworks, Bayesian loss optimization, and critical deep learningbased

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

survey analysis. The goal is to provide a structured and detailed understanding of crowd density estimation and behaviour analysis methodologies.

Title	Author(s)	Methodology	Advantages	Disadvantages
Crowd Density Prediction Using Deep Learning (2025)	Dhananjaya Kumar, Ankush K, Tasmiya	Built using Python and YOLOv5 for real-time human detection. The architecture used is CSPDarknet (Cross-Stage Partial Network).	• Real-time responsiveness delivers immediate processing. • High robustness and accuracy maintain reliable detection precision. • Flexible and resource-efficient.	• Limited real-time performance. • Accuracy degrades in low-light environments. • Risk of crowd-induced misclassification.
Density Estimation and Crowd Counting (2025)	Shantanu Todmal, Rakshith Venkatesh, Balachandra Devarangadi Sunil	Diffusion-based model for density maps with Gaussian kernels and regression. Uses event-driven sampling (optical flow) for video. Evaluated using MAE.	• Accurate results. • Handles dense crowds effectively. • Reduces computation. • Suitable for real-time use.	• High computational cost. • Possible accuracy loss. • Complex model requiring tuning.
CNN Approach for Real-time Monitoring (2024)	Khushi Kawade, Jagriti Singh	Custom CNN with convolutional, pooling, and fully connected layers. Dataset split into 80% training and 20% testing. Adam optimizer (learning rate = 0.001) with MSE loss. Evaluated using MAE and MSE across epochs.	• Rapid convergence (~10 epochs). • Avoids overfitting and generalizes well. • Robust to occlusion and perspective variation. • Potential for real-time applications.	• Requires large datasets. • Needs GPU/TPU for scalability. • Limited exploration of advanced CNN architectures. • Practical deployment issues such as lighting, camera angle, and extreme crowd density.
Critical Review on Crowd Counting (2023)	Akshita Patwal, Manoj Diwakara, Vikas Tripathi, Prabhishek Singh	Survey of classical and deep learning methods including CBL, Bayesian Loss, U-ASD Net, SRNet, FSCNet, ADCrowdNet, SFCN, RSANet, CRNet, DSSINet, and MBTTBF. Evaluated on datasets such as ShanghaiTech, UCSD, UCF-CC-50, UCF-QNRF, MALL, and WorldExpo'10.	• Covers both classical and deep learning methods. • Identifies benchmark datasets. • Highlights progress in handling occlusion, scale, and perspective issues. • Provides structured comparisons.	• Deep learning approaches require large datasets. • Domain shift from synthetic to real-world data. • High computational cost. • Existing datasets lack realism due to mixed objects.
Advanced Approach with Deep Learning	Mr. Bharath B, Mr. Ashish L	Survey of traditional and hybrid deep learning models including Extended Bayesian Loss,	• Combines traditional and hybrid methods. • Handles occlusion, scale variation, and	• Requires extensive labeled datasets. • Computationally expensive. • Real-time deployment

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

(2025)		Bayesian Loss, U-ASD Net, SRNet, FSCNet, ADCrowdNet, GTA5 Crowd Counting, and VGG-16 spatial reconstruction.	viewpoint distortion. • Tested on both dense and sparse datasets.	remains challenging. • Limited adaptability and domain adaptation issues.

III. ANALYSIS OF FIVE RESEARCH PAPERS

The five reviewed papers demonstrate a clear evolution from classical optical flow tracking methods to advanced deep learning architectures. Early methods focused on handcrafted features and regression models, while modern approaches utilize CNN-based density estimation and Bayesian loss optimization for improved performance.

This progression reflects a broader paradigm shift in computer vision research, wherein manually engineered feature extraction pipelines have been systematically supplanted by end-to-end trainable neural networks capable of learning hierarchical representations directly from raw data. Classical optical flow techniques, while computationally interpretable and theoretically grounded, were inherently constrained by their reliance on assumptions such as brightness constancy and spatial smoothness — assumptions that frequently fail under real-world conditions involving occlusion, illumination variation, and large interframe motion. The transition toward convolutional neural network architectures marked a significant inflection point, enabling models to implicitly encode spatial context and semantic understanding without explicit feature engineering. In the domain of crowd analysis and density estimation, this shift proved particularly consequential: CNN-based frameworks demonstrated a substantially improved capacity to handle perspective distortion, scale variation, and high-density occlusion — challenges that had persistently limited the effectiveness of regression-based predecessors.

Further refinement came through the incorporation of probabilistic frameworks, most notably Bayesian loss optimization, which introduced principled uncertainty quantification into the learning process. Rather than treating density map generation as a purely deterministic regression task, Bayesian formulations allow models to reason over the distribution of plausible outputs, yielding more calibrated predictions and greater robustness to annotation noise — a pervasive issue in large-scale crowd counting datasets.

Taken together, the reviewed works underscore a consistent methodological trajectory: from rigid, assumption-heavy models toward flexible, data-driven architectures that progressively close the gap between controlled benchmark performance and reliable deployment in unconstrained environments. Future research directions are likely to build upon this foundation by integrating transformer-based attention mechanisms, multi-task learning objectives, and selfsupervised pretraining strategies to further enhance scalability and generalizability across diverse operational contexts.

IV. PROPOSED SYSTEM

The proposed system is an intelligent crowd monitoring framework that combines:

• Crowd Density Estimation — determining the number of people and their spatial distribution in images/video

• Behaviour Analysis — understanding movement patterns and detecting anomalies (panic, stampedes, violent actions)

The core of the proposed system uses deep learning-based density map estimation, where:

1. Annotated head positions are converted into Gaussian-based density maps
2. CNN models learn to map input images → density maps
3. Integrating the density map gives the final crowd count.

The system begins with input acquisition, where images or video frames are captured from surveillance cameras in real-time.

These inputs undergo preprocessing steps such as background subtraction, noise removal, and optical flow analysis to clean and prepare the data. The preprocessed frames are then passed through a CNN-based feature extraction module,

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

where the deep learning model automatically learns spatial features such as head shapes, body outlines, and crowd patterns. This extracted information is used to generate a Gaussian-based density map, where each annotated head position is represented as a probability distribution across the image, effectively highlighting regions of high and low crowd concentration.

V. APPLICATIONS AND REAL- WORLD USE CASES

1. Smart City Surveillance Systems
2. Railway Station Crowd Monitoring
3. Religious Event Management
4. Political Rally Monitoring
5. Stadium and Concert Safety Systems
6. Disaster Prevention and Early Warning Systems
7. Urban Planning and Public Space Optimization

VI. SYSTEM ARCHITECTURE

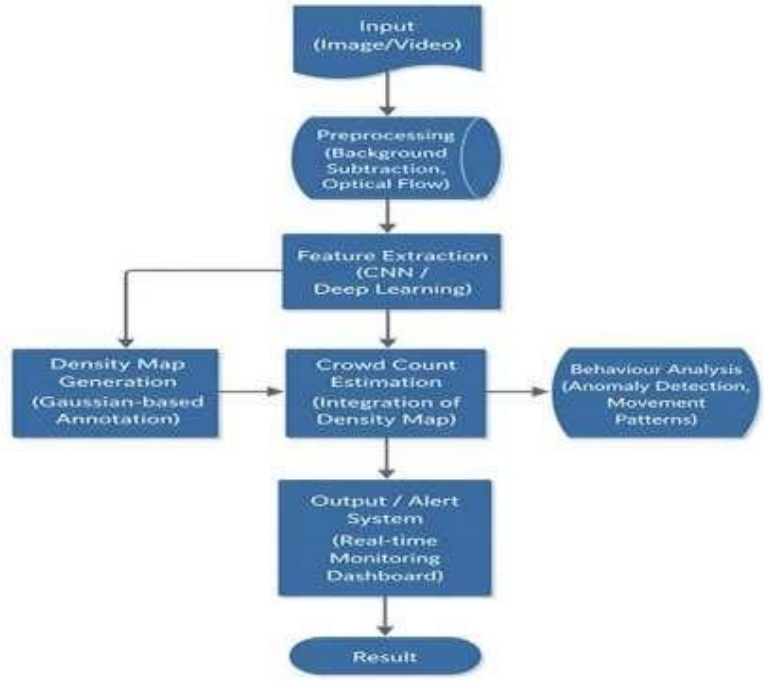

Fig. 6.1:Architecture Flow of Application Module

VII. CONCLUSION

Crowd density estimation and behaviour analysis have undergone significant transformation from classical image processing techniques to sophisticated deep learning-based models. Modern CNN architectures and hybrid approaches provide higher accuracy and robustness in dense and complex environments. However, challenges such as real-time deployment, dataset diversity, and scalability remain open research areas. This review consolidates key contributions from five significant research papers and highlights future research areas.

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

REFERENCES

1. Zhan Zhang, Y., et al. 'Single-Image Crowd Counting via Multi-Column Convolutional Neural Network.' CVPR, 2016.
2. Li, Y., et al. 'CSRNet: Dilated Convolutional Neural Networks for Understanding Highly Congested Scenes.' CVPR, 2018.
3. Cao, X., et al. 'Scale Aggregation Network for Accurate and Efficient Crowd Counting.' ECCV, 2018.