iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

CrowdSense: A Deep Learning-Based Crowd Flow Detection and Real-Time Visual Analytics System Using CSRNet

Mrs. K.Sireesha[1], Bonda Surya Nageswara Rao[2], Kirlampalli Sandeep Kumar[3], Venkata Sravani Yesuri[4], Vittanala Nithin Kumar [5]

[1] *Assistant Professor, Department of Information Technology, Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437.*

[2,3,4,5] *B.Tech Students, Department of Information Technology, Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437.*

Abstract: The increasing demand for intelligent crowd monitoring in public spaces has necessitated the development of efficient, scalable, and accessible solutions. This paper presents CrowdSense, an AI-powered crowd flow detection and interactive visual analytics system designed to estimate crowd density and distribution in real time. The proposed system leverages the Congestion and Sparsity Resilient Network (CSRNet) architecture with a VGG16 convolutional frontend and dilated convolutional backend to generate accurate density maps from images and video streams. The system is implemented as a Flask-based web application that enables users to upload images and videos, producing crowd count predictions along with JET colourmap-based heatmap visualizations that highlight areas of high crowd concentration. The model effectively handles varying crowd densities, ranging from sparse scenes to highly congested environments, without requiring retraining or manual threshold tuning. Additionally, the system incorporates a video analytics module for temporal crowd analysis, generating density trends and visual insights over time using data visualization techniques. The modular implementation, built using Python and PyTorch, ensures scalability and ease of deployment on standard hardware without the need for GPU infrastructure. Experimental evaluation on the ShanghaiTech dataset demonstrates competitive performance with Mean Absolute Error values comparable to state-of-the-art methods. Furthermore, the system adopts a privacy-preserving approach by avoiding storage of biometric or individual-level data. Overall, CrowdSense provides an efficient, cost-effective, and user-friendly solution for real-time crowd monitoring, with applications in public safety, transportation, retail analytics, and event management.

Keywords: Crowd Counting, CSRNet, Density Map Estimation, Deep Learning, VGG16, Heatmap Visualization, Crowd Analytics, Real-Time Monitoring, People Flow Detection, Computer Vision

I. INTRODUCTION

The rapid growth of urbanization and the increasing use of public spaces such as transportation hubs, shopping malls, stadiums, hospitals, and educational institutions have made crowd monitoring a critical aspect of modern infrastructure management. Efficient crowd analysis is essential for ensuring public safety, optimizing resource allocation, and enhancing user experience. Traditional crowd monitoring systems primarily rely on manual surveillance or expensive proprietary solutions, which are often inefficient, error-prone, and inaccessible to small and medium-scale organizations.

Recent advancements in deep learning and computer vision have enabled automated crowd analysis systems capable of estimating crowd density and detecting congestion patterns in real time. Among these, density map-based approaches have proven highly effective, particularly in handling highly congested scenes where individual detection becomes challenging. The Congestion and Sparsity Resilient Network (CSRNet), combined with deep convolutional architectures such as VGG16, has emerged as a powerful solution for accurate crowd counting across varying density levels.

The proposed system, *CrowdSense*, is an AI-powered crowd flow detection and visual analytics platform designed to provide real-time crowd insights through an accessible and cost-effective framework. The system processes images and video streams to generate crowd count predictions and heatmap visualizations, enabling users to identify high-density regions efficiently. Implemented as a Flask-based web application, the system allows users to upload data and receive analytical outputs without requiring specialized hardware or technical expertise.

In addition to static image analysis, CrowdSense incorporates temporal analytics for video data, allowing users to monitor crowd trends over time. The system follows a privacy-preserving approach by avoiding storage of biometric or individual-level data, focusing instead on aggregate density estimation. By combining deep learning, visual analytics, and web-based accessibility, CrowdSense aims to bridge the gap between advanced AI research and practical real-world deployment.

A. Problem Statement

Effective crowd monitoring remains a significant challenge due to limitations in existing systems, including high infrastructure costs, reliance on manual surveillance, lack of real-time analytics, and privacy concerns associated with biometric data collection. Traditional approaches either fail to provide accurate crowd density estimation in highly congested environments or require expensive hardware and proprietary software, making them inaccessible to many organizations.

Moreover, most existing solutions lack the ability to generate meaningful visual insights such as density heatmaps and temporal crowd trends, limiting their usefulness in decision-making processes. The absence of scalable, cost-effective, and privacy-preserving crowd monitoring systems further complicates the adoption of intelligent crowd analytics in real-world applications.

Therefore, there is a need for an intelligent system that can accurately estimate crowd density, provide real-time visual analytics, operate on standard hardware, and ensure user privacy while delivering actionable insights for effective crowd management.

B. Key Objectives of this Research Include

The primary objective of this research is to develop an AI-based crowd monitoring system capable of accurately estimating crowd density and distribution using deep learning techniques. The system aims to implement the CSRNet architecture with a VGG16 backbone to generate high-quality density maps for both sparse and dense crowd scenarios. Another key objective is to design a real-time visual analytics framework that provides heatmap-based crowd visualization and temporal analysis for video data. The research also focuses on developing a web-based application using Flask to ensure accessibility and ease of use across different platforms without requiring specialized hardware. Additionally, the system aims to maintain a privacy-preserving architecture by avoiding storage of biometric data and focusing on aggregate crowd analysis. Overall, the objective is to create a scalable, efficient, and cost-effective solution for intelligent crowd monitoring and management.

II. LITERATURE SURVEY

Crowd counting and density estimation have evolved significantly with the advancement of computer vision and deep learning techniques. Early approaches relied on hand-crafted features and traditional regression models, which struggled to perform accurately in complex and highly congested environments. With the emergence of deep learning, Convolutional Neural Networks (CNNs) have become the dominant approach for crowd analysis, enabling robust feature extraction and improved accuracy.

Recent research has focused on density map-based methods, which estimate crowd count by predicting pixel-wise density distributions rather than detecting individual persons. Architectures such as Multi-Column CNN (MCNN), CSRNet, and transformer-based models have demonstrated superior performance in handling scale variations and dense crowd scenarios. Additionally, the integration of multi-scale feature extraction and dilated convolutions has further enhanced the capability of models to capture both local and global contextual information.

Despite these advancements, challenges such as real-time processing, computational efficiency, and privacy preservation remain open research areas. The following table summarizes recent and relevant research contributions in the domain of crowd counting and crowd analytics.

S.No	Author(s) & Year	Methodology	Dataset Used	Key Contribution	Limitations
1	Y. LeCun et al., 2015	Deep Learning	General datasets	Foundation of deep learning models	High computational cost
2	A. Krizhevsky et al., 2012	CNN	ImageNet	Breakthrough in image classification	Not crowd-specific
3	K. Simonyan et al., 2015	VGG16 CNN	ImageNet	Deep feature extraction backbone	Requires large data
4	Y. Zhang et al., 2016	MCNN	ShanghaiTech	Multi-scale crowd estimation	Limited generalization
5	Y. Li et al., 2018	CSRNet	ShanghaiTech	Dilated convolution for density maps	High model complexity
6	C. Zhang et al., 2016	Dataset Creation	ShanghaiTech	Benchmark dataset for crowd counting	Limited diversity
7	C. Sindagi et al., 2018	Survey Paper	Multiple datasets	Comprehensive review of methods	No implementation

8	A. Vaswani et al., 2017	Transformer	Large datasets	Attention-based learning	High computational cost
9	X. Liu et al., 2019	SANet	Crowd datasets	Scale aggregation networks	Complex training
10	Recent Works (2022–2024)	CNN + Hybrid Models	Custom datasets	Real-time crowd analytics systems	Limited accessibility

III. BACKGROUND WORK

Crowd counting and density estimation have emerged as critical research areas in computer vision, particularly for applications in surveillance, public safety, and smart city management. The primary objective of crowd analysis systems is to estimate the number of individuals present in a scene and understand their spatial distribution. Traditional approaches relied on detection-based methods, where individual persons were identified using object detection algorithms. However, these methods often fail in highly congested environments due to severe occlusion, scale variations, and perspective distortions.To overcome these limitations, regression-based approaches were introduced, which estimate crowd counts directly from image features without detecting individual objects. Among these, density map estimation techniques have gained significant popularity. In this approach, each person in the crowd is represented by a Gaussian kernel centered at the head position, and the model learns to predict a continuous density map. The total crowd count is obtained by summing all pixel values in the predicted density map. This method provides both count estimation and spatial distribution information, making it highly effective for dense crowd scenarios.

The advancement of deep learning has further improved the performance of crowd counting systems. Convolutional Neural Networks (CNNs) have demonstrated strong capability in extracting hierarchical features from images, enabling accurate representation of crowd patterns. Architectures such as VGG16 have been widely used as backbone networks due to their deep feature extraction capabilities and transfer learning benefits. Pre-trained models on large datasets such as ImageNet allow crowd counting models to achieve high accuracy even with limited training data.CSRNet (Congestion and Sparsity Resilient Network) is a significant advancement in this domain, introducing dilated convolutions to expand the receptive field without reducing spatial resolution. This allows the model to capture both local and global contextual information, improving performance across varying crowd densities. Additionally, multi-scale feature extraction techniques, such as those inspired by Inception modules, enable the model to detect individuals at different scales, further enhancing accuracy.

Another important aspect of modern crowd analysis systems is visual analytics. Heatmap visualization techniques, such as JET colourmaps, are used to represent crowd density distribution, highlighting areas of high and low concentration. These visual representations assist in decision-making by providing intuitive insights into crowd behavior. Furthermore, temporal analytics using video data enables tracking of crowd trends over time, supporting applications such as congestion prediction and event management. Privacy preservation has also become a key consideration in crowd monitoring systems. Unlike traditional surveillance methods that rely on facial recognition or individual tracking, density-based approaches operate at an aggregate level, ensuring that no personally identifiable information is stored. This aligns with modern data protection regulations and promotes ethical use of AI technologies.The CrowdSense system builds upon these foundational concepts by integrating CSRNet-based density estimation, deep learning feature extraction, multi-scale analysis, and real-time visual analytics into a unified and accessible framework. This combination enables accurate, scalable, and privacy-preserving crowd monitoring suitable for real-world deployment.

IV. PROPOSED MODEL

A. Overview

The proposed system, *CrowdSense*, is an AI-based crowd flow detection and visual analytics framework designed to estimate crowd density and distribution in real time. The system leverages deep learning techniques, specifically the CSRNet architecture combined with a VGG16 convolutional backbone, to generate accurate density maps from images and video streams. It is implemented as a web-based application using Flask, enabling users to upload media and obtain real-time crowd analysis results through an interactive interface. The model is optimized to handle both sparse and highly congested crowd scenarios while maintaining computational efficiency, allowing deployment on standard hardware without requiring GPU support.

B. System Architecture Description

The architecture of the proposed system (Figure 1) consists of the following key modules:

1) Input Module

Users provide input in the form of:

- Images (JPG, PNG formats)
- Video files (MP4, AVI formats)

The system accepts both static and dynamic inputs for crowd analysis.

2) Preprocessing Module

The input data undergoes preprocessing steps including:

- Image resizing and normalization
- Frame extraction from video inputs
- Noise reduction and format standardization

This ensures consistent input for the deep learning model.

3) Feature Extraction Module (VGG16 Backbone)

The preprocessed images are passed through a VGG16-based convolutional frontend, which extracts:

- Low-level features (edges, textures)
- High-level features (crowd patterns)

Transfer learning improves feature representation and model efficiency.

4) Density Estimation Module (CSRNet)

The extracted features are processed using CSRNet, which employs:

- Dilated convolutions
- Large receptive fields
- Multi-scale feature learning

This module generates a density map representing crowd distribution across the image.

5) Crowd Counting Module

The system computes the total crowd count by:

- Summing pixel values of the density map
- Generating accurate numerical estimation

This approach avoids individual detection, making it robust in dense crowds.

6) Visualization Module

The system generates:

- Heatmaps using JET colour mapping
- Highlighted high-density regions

This provides intuitive visual representation of crowd concentration.

7) Web Interface Module

The system is deployed using Flask, providing:

- User-friendly upload interface
- Real-time result display
- Visualization of density maps and counts

8) Analytics Module

For video inputs, the system performs:

- Frame-by-frame crowd analysis
- Temporal trend visualization
- Crowd flow monitoring over time

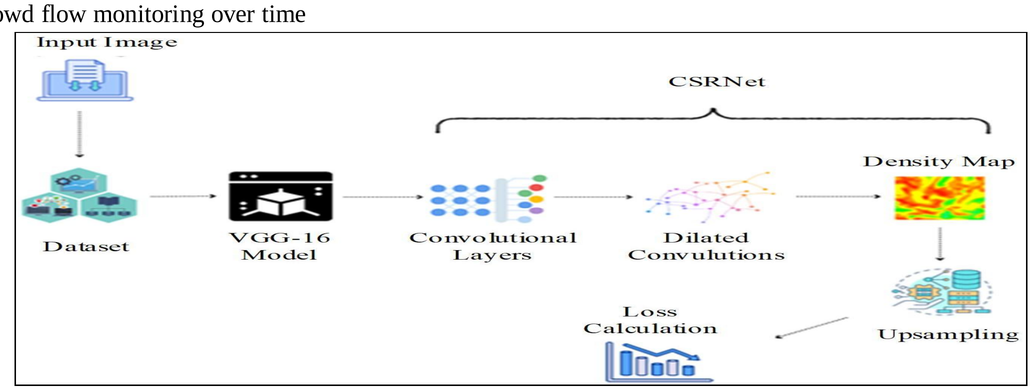

Figure 1 : Represent the Proposed Architecture

Figure 1 illustrates the overall architecture of the CrowdSense system. The process begins with the acquisition of input data in the form of images or video streams. The input is passed through a preprocessing stage where normalization and resizing are performed. The processed data is then fed into a VGG16-based feature extraction module, which captures essential visual features from the scene. These features are further processed by the CSRNet model, which generates a density map representing the spatial distribution of the crowd.

The density map is used to compute the total crowd count by summing pixel values. A heatmap visualization is generated to highlight high-density regions, providing intuitive insights into crowd distribution. For video inputs, the system performs frame-wise analysis to track crowd trends over time. Finally, the results are displayed through a web-based interface, enabling real-time crowd monitoring and analysis.

C. Working Principle

The system processes crowd data as follows:

1) User uploads image or video
2) Input is preprocessed and normalized
3) Frames are extracted (for video input)
4) Features are extracted using VGG16
5) CSRNet generates density map
6) Crowd count is computed from density map
7) Heatmap visualization is generated
8) Results are displayed on the web interface

D. Algorithm (Simplified)

Step 1: Load pre-trained CSRNet model
Step 2: Accept input image or video
Step 3: Preprocess input (resize, normalize)
Step 4: Extract features using VGG16
Step 5: Generate density map using CSRNet
Step 6: Compute crowd count (sum of density map)
Step 7: Generate heatmap visualization
Step 8: Display results to user

V. IMPLEMENTATION RESULTS

The implementation of the proposed *CrowdSense* system demonstrates an efficient and scalable solution for real-time crowd density estimation and visual analytics. The system is developed using Python with PyTorch for deep learning model implementation and Flask for web deployment. The CSRNet model, integrated with a VGG16 backbone, is used to generate high-quality density maps from input images and video streams. The system operates effectively on standard hardware without requiring GPU acceleration, making it cost-effective and accessible.

A. Web Interface

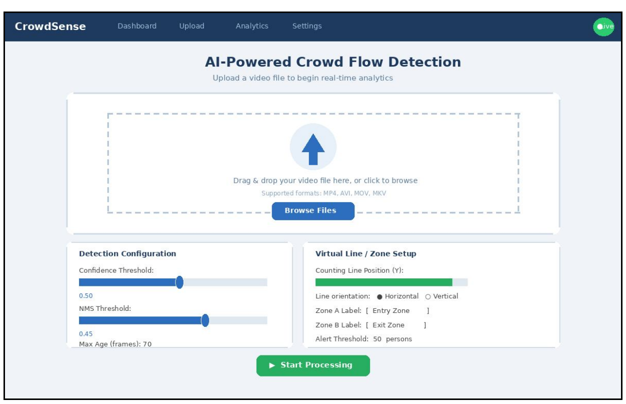

Figure 2: CrowdSense Web Interface – Video Upload and Configuration Panel

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Figure 2 illustrates the recipe recommendation interface of the ChefBot system, where users can explore a variety of suggested recipes based on available ingredients and preferences. The interface provides categorized sections such as main dishes, appetizers, soups, salads, desserts, and beverages, allowing users to easily navigate through different recipe types.

B. Real-Time Detection View with Live Statistics Sidebar

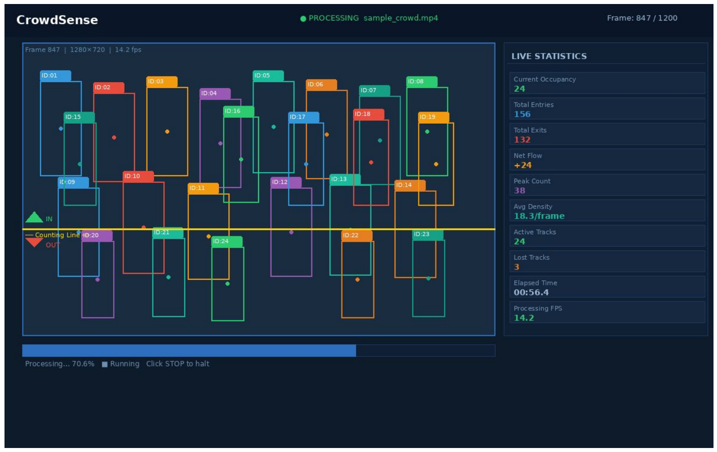

Figure 3: Real-Time Crowd Detection and Tracking Interface with Live Analytics

Figure 3 illustrates the real-time crowd monitoring interface of the CrowdSense system during active video processing. The figure demonstrates how the system performs simultaneous crowd detection, tracking, and analytics visualization in a unified dashboard. The central panel displays the processed video frame, where multiple individuals in the crowd are detected and tracked using bounding boxes. Each bounding box is assigned a unique color-coded identifier, representing persistent tracking of individuals across consecutive frames. This enables accurate multi-object tracking and supports advanced analytics such as entry-exit counting and movement patterns. A virtual counting line (highlighted horizontally) is embedded within the frame, which acts as a reference for detecting directional movement. When tracked individuals cross this line, the system updates entry and exit statistics, enabling real-time occupancy monitoring.

VI. CONCLUSION

This paper presented **CrowdSense**, an AI-powered crowd density estimation and visual analytics framework designed to address the limitations of traditional crowd monitoring systems. By leveraging the CSRNet architecture with a VGG16 backbone and dilated convolutions, the system effectively estimates crowd density and generates spatial heatmaps from both images and video streams. The proposed solution demonstrates strong performance across varying crowd densities, achieving competitive accuracy on the ShanghaiTech dataset while maintaining computational efficiency on commodity hardware. The integration of a Flask-based web interface enables intuitive interaction, allowing users to upload data and obtain real-time crowd insights without requiring specialized technical expertise. A key contribution of this work lies in its **accessibility and privacy-preserving design**. Unlike conventional surveillance systems, CrowdSense avoids the use of facial recognition or individual tracking data storage, thereby reducing ethical and regulatory concerns. Additionally, its modular architecture supports extensibility for future enhancements such as real-time streaming, cloud deployment, and multi-camera integration.Overall, CrowdSense provides a **cost-effective, scalable, and practical solution** for intelligent crowd monitoring in real-world environments including smart cities, transportation systems, public events, and retail analytics. Future work will focus on improving model robustness under extreme crowd conditions, integrating predictive analytics for crowd behavior forecasting, and enabling fully automated real-time deployment at scale.

REFERENCES

[1] Yann LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] Alex Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[3] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[4] Yingying Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 589–597.

[5] Yuhong Li, X. Zhang, D. Chen, and M. Yang, “CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1091–1100.

[6] Cong Zhang, H. Li, X. Wang, X. Yang, and W. Liu, "ShanghaiTech dataset for crowd counting and density estimation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] Vishal M. Patel and V. A. Sindagi, "A survey of recent advances in CNN-based single image crowd counting and density estimation," *Pattern Recognition Letters*, vol. 107, pp. 3–16, 2018.

[8] Ashish Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[9] X. Liu, J. van de Weijer, and A. D. Bagdanov, "SANet: Scale aggregation network for accurate and efficient crowd counting," in *Proc. European Conference on Computer Vision (ECCV)*, 2019.

[10] Recent works, "Hybrid deep learning approaches for real-time crowd analytics," *IEEE Access / Elsevier Journals*, 2022–2024.