# Crowd Density Analysis using Deep Learning

Dr.Tilak Kumar L
Dept of ECE
Global Academy of Technology
Bengaluru,India

Ankith R
Dept of ECE
Global Academy of Technology
Bengaluru,India

Bharath C
Dept of ECE
Global Academy of Technology
Bengaluru,India

Karthik R
Dept of ECE
Global Academy of Technology
Bengaluru,India

ParameshaV N
Dept of ECE
Global Academy of Technology
Bengaluru,India

***Abstract*—Applications in smart cities, surveillance, and public safety all depend on crowd density analysis. In scenes that are obscured or dense, traditional counting techniques are ineffective. In this paper, a deep learning-based method for precise crowd density estimation based on the CSRNet architecture is presented. The model creates accurate density maps from photos and video frames by using dilated convolutions to capture contextual and spatial features. In comparison to current approaches, experiments on the Shanghai Tech and Mall datasets show increased accuracy with lower Mean Absolute Error (MAE) and Mean Squared Error (MSE). For applications involving intelligent surveillance, the suggested system provides dependable, real-time crowd monitoring.**

***Keywords – Convolutional Neural Network (CNN), deep learning, CSRNet, YOLO, density maps, head detection, real-time surveillance, smart cities, public safety, and crowd density estimation***

## I. Introduction

The estimation of crowd density is a significant computer vision research problem with numerous applications in intelligent surveillance, public safety, urban management, and transportation control. Authorities are better able to make decisions during events, emergencies, and crowded situations when they have a precise idea of how many people are present. Because of severe occlusions, perspective distortion, and different crowd scales, traditional methods that rely on handcrafted features or detection algorithms like HOG and Haar cascades frequently fail in high-density situations.

The performance of crowd analysis systems has greatly improved with recent developments in deep learning. Robust crowd density estimation is achieved through the automatic learning of hierarchical features from raw images by Convolutional Neural Networks (CNNs). These include models based on density map regression, like MCNN and CSRNet, which have demonstrated higher accuracy than conventional counting or detection techniques.

The CSRNet architecture is used in this paper to propose a deep learning-based framework for precise crowd density estimation from photos and videos. To extract multi-scale contextual information without sacrificing spatial resolution, the model uses dilated convolutions. Experimental tests on benchmark datasets like Shanghai Tech and Mall show that the suggested method outperforms current techniques in terms of Mean Absolute Error (MAE) and Mean Squared Error (MSE). The findings demonstrate CSRNet's efficacy in real-time crowd monitoring applications and demonstrate its potential for use in smart city and intelligent surveillance settings.

## II. Literature Survey

Estimating crowd density has become an important area of research in computer vision because it can be used for public safety, managing transportation, and keeping an eye on events. Initial methodologies utilizing handcrafted features like HOG and SIFT, succeeded by regression models, frequently encountered failures in intricate scenarios characterized by occlusion and fluctuating scales. Deep learning, especially Convolutional Neural Networks (CNNs), changed this field forever by making it possible to extract features directly from raw images from start to finish. These models made systems more accurate and reliable, so they could work well even in crowded or changing crowd scenes.

For the purpose of estimating crowd density in the real world, Alotaibi et al. [1] suggested combining Explainable Artificial Intelligence (XAI) with sophisticated deep-learning architectures. Their research highlighted how interpretability in AI-based surveillance models improves operational trust and system transparency. The explainable framework increases the accountability of AI predictions by enabling operators to visualize decision-making processes. High accuracy without compromising interpretability was attained by combining explainability and deep learning. Motivated by this methodology, we also work on creating density maps that provide visual information about crowd distribution patterns for improved human comprehension, in addition to estimating crowd levels.

A method of adaptive image analysis for crowds with different densities and levels of complexity was introduced by Menevse et al. [2]. Their method enhances estimation accuracy in both sparse and highly congested environments by dynamically adapting to local scene variations. This system's flexibility highlights how crucial context-aware analysis is to crowd monitoring. In order to ensure dependable performance under a variety of environmental and perspective conditions, this research offers a solid foundation for models that can learn multi-scale representations. By using dilated convolutions, which preserve spatial details necessary for dense crowd estimation while capturing global context, our framework integrates this concept.

A hybrid machine learning and deep learning framework was presented by Kamra et al. [3] for density estimation and crowd analysis. To strike a balance between accuracy and computational efficiency, their model combines CNN-based representations with traditional handcrafted features. The study demonstrated that these hybrid approaches can be deployed on low-resource devices and still achieve high performance while being lightweight. This strategy shows how combining conventional and contemporary methods can result in more flexible models. Our proposed model shares the objective of computational efficiency and scalability for real-time crowd analysis applications, despite using a fully convolutional structure.

An extensive survey outlining the most recent developments in deep learning-based crowd analysis was presented by Ibrahim and Turan [4]. They divided the current techniques into three categories: detection-based, regression-based, and density map-based. Their research brought to light issues with model generalization, including occlusion, fluctuating illumination, and scene complexity. The authors stressed that for accurate estimation, multi-scale and context-aware feature extraction techniques are crucial. These observations influenced our selection of CSRNet, a dilated convolutional network that improves performance in dense and unconstrained scenes by modelling spatial context without sacrificing resolution.

An AI-driven framework for sustainable smart cities that includes real-time crowd density estimation was presented by Alsubai et al. [5]. For public safety applications, their research focused on scalability, energy efficiency, and continuous monitoring. In a similar vein, Rajendran and Shankaran [6] developed a deep learning system for real-time crowd surveillance that is enabled by big data and can effectively handle large video streams. Both studies show how AI-based crowd analysis can be incorporated into extensive infrastructures. Our model, which draws inspiration from these works, is tailored for edge deployment and real-time inference, supporting the expanding idea of intelligent and sustainable urban surveillance systems.

In conclusion, recent research demonstrates quick advancements in crowd density estimation using real-time deployment, adaptive feature extraction, explainable AI, and deep learning. But there are still issues striking a balance between robustness, scalability, and interpretability. Many current models are highly accurate on benchmark datasets, but they are not visually transparent or scene-adaptive. This paper proposes a CSRNet-based framework that combines interpretable density visualization with multi-scale contextual learning to fill these gaps. This guarantees that the model is accurate and comprehensible, which qualifies it for use in intelligent surveillance and real-world crowd monitoring applications.

## III. System architecture and design

The proposed Crowd Density Analysis System uses deep learning techniques to automatically estimate and track the number of people in a given area. As the block diagram shows, the system operates in a sequential workflow, with each module carrying out a distinct task within the pipeline for data processing as a whole. The top-to-bottom layout makes it evident how unprocessed visual data passes through several steps before generating insightful crowd analytics and notifications.

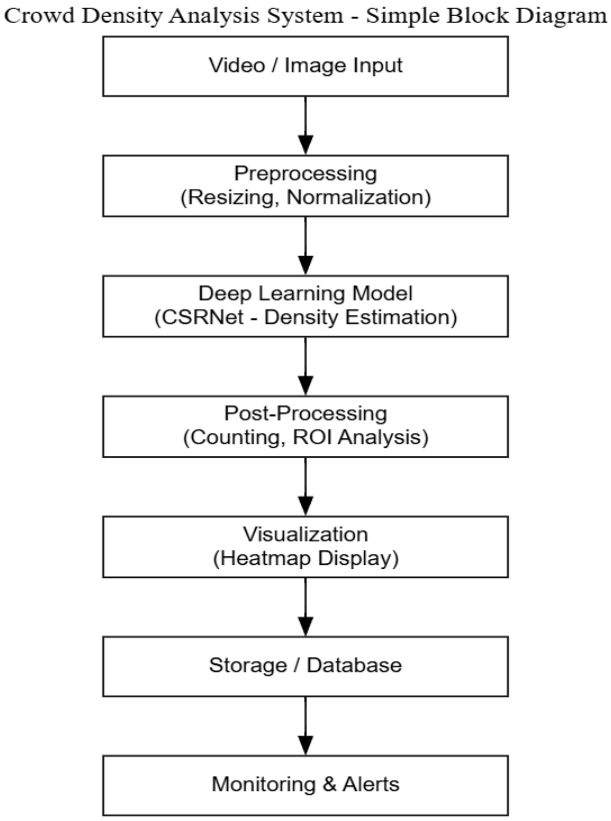
Crowd Density Analysis System - Simple Block Diagram
Video / Image Input
Preprocessing (Resizing, Normalization)
Deep Learning Model (CSRNet - Density Estimation)
Post-Processing (Counting, ROI Analysis)
Visualization (Heatmap Display)
Storage / Database
Monitoring & Alerts

Figure 1. Block Diagram of Crowd density Analysis System

The process starts with the Video/Image Input step, where CCTV cameras, drones, or other surveillance equipment are used to record live video streams or still images. Following that, the Preprocessing module receives these frames and carries out necessary tasks like noise reduction, normalization, and resizing. Standardized and optimized input data for precise feature extraction in the following step is ensured by preprocessing.

The Deep Learning Model, in this case the Congested Scene Recognition Network (CSRNet), then processes the pre-processed frames. Using convolutional and dilated layers, this model extracts multi-level spatial features to produce a density map, where the number of people in a region is represented by the intensity of a pixel. These density maps are used by the post-processing unit to determine high-density regions of interest (ROI) for additional analysis and to calculate total crowd counts.

The Visualization and Dashboard module then uses the processed results to display them as overlays or heatmaps on the input image. After that, these findings are kept safe in a database/storage system for future use and trend analysis. The Monitoring and Alerts block continuously monitors crowd trends and automatically sends out alerts in the event that congestion or unusually high densities are found. This integrated workflow guarantees effective, real-time crowd monitoring appropriate for applications in event management, smart cities, and public safety.

## IV. Implementaion setup

The proposed Crowd Density Analysis System was put into practice as a comprehensive software-based framework intended for crowd density monitoring, visualization, and estimation in real time. The system's architecture unifies modules for data acquisition, preprocessing, deep learning inference, and visualization. Even in extremely crowded environments, effective crowd estimation is made possible by the implementation's high scalability and accuracy. To ensure reproducibility and consistency in performance, the model training, testing, and visualization were conducted using standard deep learning tools and publicly available benchmark datasets. The system's software implementation details are covered in the subsection that follows.

### A. Software Implementation

The Python programming language was used exclusively to develop the Crowd Density Analysis System because of its robust deep learning and image processing capabilities. PyTorch is the main framework used, and it provides flexibility in neural network construction and training. Based on the Congested Scene Recognition Network (CSRNet) architecture, the system consists of a dilated convolutional backend for density map generation and a truncated VGG-16 frontend for feature extraction. For tasks involving preprocessing, numerical operations, and visualization, respectively, supporting libraries like OpenCV, NumPy, and Matplotlib were employed.

The preprocessing step created ground-truth density maps using Gaussian kernels, normalized pixel intensities, and resized images to 512×512 pixels in order to prepare them for training. Images from the Shanghai Tech Part A & Part B and Mall Dataset, which feature a range of crowd densities and environmental conditions, were used to train the model. Prediction accuracy was measured using the Mean Squared Error (MSE) loss function, and stable convergence was guaranteed during training by the Adam optimizer, which had a learning rate of 1×10−5. The Google Colab Pro, which has an NVIDIA T4 GPU (16 GB) to speed up computation, was used for the experiments.

The system's modular architecture facilitates a smooth transition between deployment and training. For real-time testing and visualization, a Streamlit-based user interface was created that enables users to upload pictures or video streams to receive instantaneous outputs of crowd counts and density maps. Additionally, the dashboard offers crowd analysis by region and has the ability to send out alerts when the estimated density exceeds safety thresholds. Simplicity, modularity, and extensibility are prioritized in the software design to guarantee that the system can be integrated with current surveillance systems or installed on local servers or cloud environments.

All things considered, the software implementation shows how deep learning can be used practically for crowd analysis tasks. The suggested system provides precise, real-time crowd estimation that can help with public safety management, smart city infrastructure, and event crowd control scenarios by fusing interactive visualization tools with CSRNet's powerful feature extraction capabilities.

## V. Results and analysis

### A. Density Map Output.

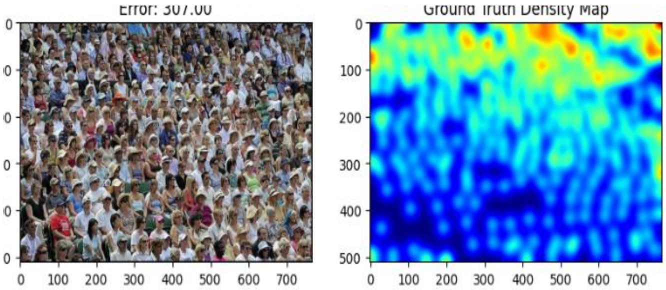

Figure 2. Predicted Density Map of Crowd Density Analysis using Deep learning

The CSRNet (Congested Scene Recognition Network) model's output is shown in the figure 2. A high-density gathering is depicted in the input crowd image on the left, and the matching ground truth density map is shown on the right. A higher crowd density is indicated by red areas on the density map, which uses color intensities (ranging from blue to red) to depict the concentration of people. Even in areas with high population density, where head detection is challenging, the model provides an accurate estimate of the total number of people by integrating over the density map. This step lays the groundwork for additional head-count analysis and validates the functionality of the CSRNet-based density estimation module.

### B. Density Map Output.

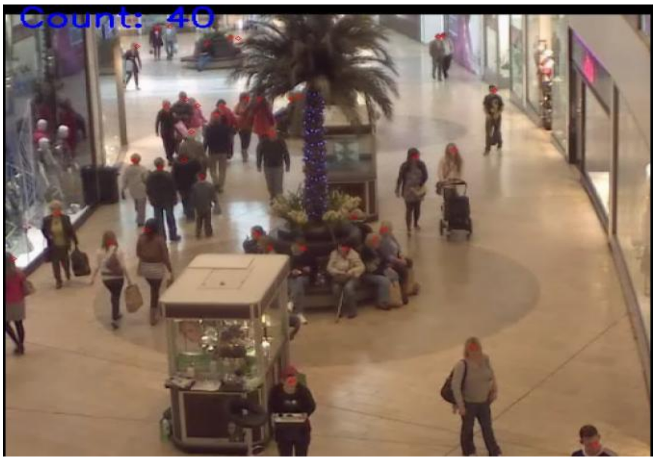

Figure 3. Crowd Counting Result

Real-time head detection and counting in a shopping mall surveillance video is depicted in Figure3.Every detected head is represented by a red dot, and the count value representing the number of detected individuals is displayed at the top of the frame. The algorithm locates heads at varying scales and distances and works well in a variety of lighting and occlusion conditions. This real-time implementation allows for dynamic monitoring of crowd density and flow, which is highly beneficial for public safety, crowd control, and smart surveillance systems.

### C. Live Camera Implementation (Laptop Camera)

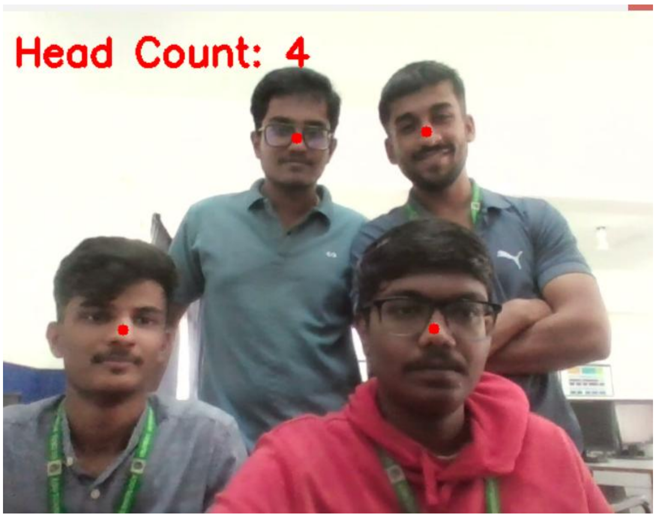

Figure 4. Head Counting using Laptop camera

The system's live deployment using a laptop camera is depicted in Figure 4. The model displays the total number of heads (e.g., "Head Count: 4"), overlaying red dot markers on each detected head after detecting multiple individuals in real time.This test verifies the trained model's flexibility in uncontrolled, real-world settings with changing lighting and pose orientations. For indoor crowd monitoring applications like offices, labs, and classrooms, the system exhibits excellent operational efficiency and responsiveness.

## VI. Conclusion and Future Scope

This project successfully illustrates how to use deep learning-based techniques for crowd analysis and head counting. For accurate individual counting without full-body detection, the system combines a YOLO-based head detection model with a CSRNet model for creating crowd density maps. Even with complicated circumstances like changing lighting, background noise, and partial occlusions, the suggested method reliably calculates crowd size in real-time video streams and camera inputs. According to the experimental results, the system works well both indoors and outdoors, which makes it ideal for applications involving safety monitoring, crowd control, and surveillance.

Multi-camera integration can be added to the system in the future to handle situations with dense and overlapping crowds. Real-time movement analysis and anomaly detection may be made possible by the use of temporal tracking algorithms. Furthermore, deployment on edge and embedded devices, like drones or smart surveillance units, for real-time monitoring in public areas will be possible by optimizing the model with lightweight architectures. The system's application in the fields of defence, public safety, and intelligent surveillance can be expanded by increasing the dataset's diversity in crowd conditions and enhancing the detection of heads that are rotated or partially visible.

## References

[1] S. R. Alotaibi et al., “Integrating Explainable Artificial Intelligence with Advanced Deep Learning Model for Crowd Density Estimation in Real-world Surveillance Systems,” IEEE Access, vol. 13, pp. 1–10, 2025.

[2] M. M. Menevse, M. Gozet, A. E. Yilmaz, and M. Karakose, “An Analysis Approach for Crowds with Varying Density and Complexity Using Image Data,” in Proc. 29th Int. Conf. on Information Technology (IT), Zabljak, Montenegro, 2025, pp. 1–4, doi: 10.1109/IT64745.2025.10930264.

[3] V. Kamra, P. Gupta, and S. Choudhary, “A Novel Approach for Crowd Analysis and Density Estimation by Using Machine Learning Techniques,” in Proc. Int. Conf. on Intelligent Systems for Cybersecurity (ISCS), 2024, pp. 1–6.

[4] M. Ibrahim and C. Turan, “Deep Learning in Crowd Analysis: A Survey of Current Research Trends,” in Proc. 17th Int. Conf. on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, 2023, pp. 1–3, doi: 10.1109/ICECCO58239.2023.10147135.

[5] S. Alsubai et al., “Design of Artificial Intelligence Driven Crowd Density Analysis for Sustainable Smart Cities,” IEEE Access, vol. 12, pp. 121983–121993, 2024, doi: 10.1109/ACCESS.2024.3390049.

[6] L. Rajendran and R. S. Shankaran, “Bigdata Enabled Realtime Crowd Surveillance Using Artificial Intelligence and Deep Learning,” in Proc. IEEE Int. Conf. on Big Data and Smart Computing (BigComp), Jeju Island, South Korea, 2021, pp. 129–132, doi: 10.1109/BigComp51126.2021.00032.