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*— Crowd density analysis is crucial for applications such as smart cities, surveillance and public safety. Traditional counting does not work in more complicated situations like obscured or high density. This document introduces a precise crowd density estimation method based on deep learning and the CSRNet architecture. Our model creates correct density maps from photos and video frames, taking advantage of dilated convolutions to capture context and space features. The experiments conducted with the Shanghai Tech and Mall datasets show that the proposed method is more accurate than current solutions, with lower Mean Absolute Error and Mean Squared Error.**

***Keywords* – Crowd Density Estimation, Deep Learning, CSRNet, Convolutional Neural Network (CCN), Head Map, Public Safety, Computer Vision,**

I. INTRODUCTION

Crowd density estimation is a fundamental problem in computer vision, which has broad applications, including intelligent surveillance, public safety, urban management, and transportation control. The more accurate the people count is, the easier the authorities will have to decide on events, emergencies, and crowded situations. Traditional methods, performed on handcrafted features or detection algorithms like HOG and Haar cascades, often fail in high-density scenarios due to severe occlusions, perspective distortion, and varied crowd scale.

Recent advancements in deep learning have strengthened crowd analysis performance. The CNNs automatically learn hierarchical features from raw images for robust crowd density estimation. These include models based on density map regression, like MCNN and CSRNet, which have demonstrated higher accuracy than conventional counting or detection techniques.

This paper provides a deep learning-based framework for the accurate estimation of crowd density from photos and videos using the CSRNet architecture. Due to the fact that the prediction performance of crowd count maps requires large receptive fields, the model relies on dilated convolutions to extract multiscale contextual information without losing spatial resolution. Experimental on benchmark datasets such as Shanghai Tech and Mall show that the proposed method outperforms the current approaches. techniques concerning the mean absolute error and mean squared error.

II. LITERATURE SURVEY

The task of estimating crowd density is a new and exciting domain in computer vision due to its implications for public safety, transportation management, and event monitoring. Early techniques using handcrafted features such as HOG and SIFT, followed by regression models, often failed in complex situations when occlusions or scale changed. With the advent of deep learning and primarily Convolutional Neural Networks (CNNs), a complete transformation took place, which allowed feature extraction directly from raw images from end-to-end. These models improved accuracy and robustness, especially in noisy and/or transformed dynamic scenes (large and challenging crowds).

In order to estimate crowd density in the real-world context, Alotaibi et al. [1] proposed the merging of Explainable Artificial Intelligence with informed deep-learning models and architectures. Their findings demonstrated that interpretability of AI-based surveillance generated a higher degree of operational trust and transparency of the system. The explainable framework enabled operators to visualize AI-based decisions, thereby increasing the attribution of accountability for AI-based predictions. The incorporation of explainability in a deep-learning model achieved a high level of accuracy while maintaining interpretability. Inspired by this idea, we also focused on developing density maps, which present visual information about crowd distribution patterns for a better understanding by humans, in conjunction with estimating the level of crowd density.

Menevse et al. [2] introduced an adaptive image analysis of crowds that have various densities and complexities. Their approach improves the estimation accuracy both in sparse and highly congested settings through dynamic adaptation to local scene variations. In fact, the flexibility of this system represents how important context-aware analysis is for crowd monitoring. This work lays a sound foundation for models that can learn multiscale representations to ensure dependable performance under different environmental and perspective conditions. Our framework integrates this concept by leveraging dilated convolutions that preserve spatial details necessary for dense crowd estimation while capturing global context.

Kamra et al. [3] proposed a hybrid machine learning and deep learning framework for density estimation and crowd analysis. Their model is designed to trade off accuracy for computational efficiency by combining CNN-based representations with handcrafted features. The work demonstrated that such hybrid models could be pushed to low-resource devices and yet deliver high performance while being lightweight. This strategy illustrates the possibility of enhancing flexibility of models by combining conventional techniques with more contemporary ones. Our proposed model shares a similar objective of striving for computational efficiency and scalability toward applications pertaining to real-time crowd analysis, albeit using a fully convolutional structure.

Ibrahim and Turan [4] presented an extended survey on recent advances of deep learning-based approaches to crowd analysis. They categorized the current techniques into three groups, namely, detection-based, regression-based, and density map-based. Their study shed light on the problems of model generalization, occlusion, fluctuating illumination, and scene complexity. The authors emphasized that multi-scale and context-aware feature extraction methods are necessary for accurate estimation. These observations influenced our choice to implement CSRNet, a dilated convolutional network that enhances performance in dense and unconstrained scenes by modelling spatial context without sacrificing resolution.

Alsubai et al. [5] presented an AI-driven framework for sustainable smart cities, including real-time crowd density estimation. Their work focused on scalability, energy efficiency, and continuous monitoring for public safety applications. Rajendran and Shankaran [6] proposed a big data-enabled deep learning system for crowd surveillance in real time. Both works illustrate how AI-based crowd analysis can be integrated into extensive infrastructures. Our model, inspired by these works, will be tailored for edge deployment and real-time inference to support the expanding concept of intelligent and sustainable urban surveillance systems.

III. System Architecture and Design

The Crowd Density Analysis System proposal uses deep learning techniques for automatic estimation and tracking of the number of people in a given area. As shown by the block diagram, the system works along a sequential workflow where each module carries out a specific function within the pipeline for data processing as a whole. The top-to-bottom structure provides a clear visual of how the raw visual data is processed through several steps to create insightful crowd analytics and notifications.

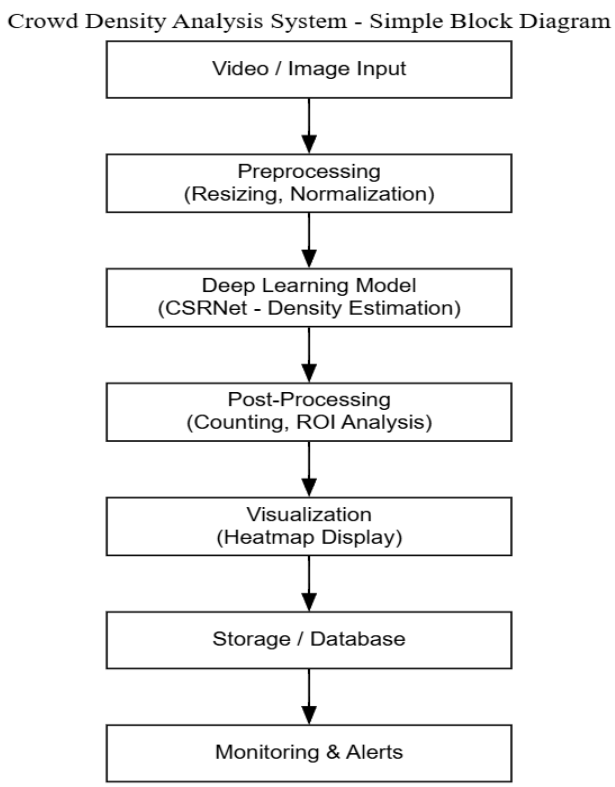

Figure-1. Block Diagram of Crowd density Analysis System

The process initiates with the Video/Image Input, whereby the live video streams or still images are recorded through CCTV cameras, drones, or another form of surveillance equipment. Subsequent to that, these frames are fed into the Preprocessing module, where required operations such as noise reduction, normalization, and resizing are performed. The goal of preprocessing is to make the data standardized and optimized for precise feature extraction in the following step.

The pre-processed frames are fed into the Deep Learning Model, which, in this context, is the Congested Scene Recognition Network. This model leverages convolutional and dilated layers to extract multi-level spatial features and subsequently yield a density map in which the number of people in a region corresponds to the intensity of a pixel. These density maps form the basis on which the post-processing unit identifies high-density regions of interest for further analysis and determines total crowd counts.

The results are then used by the Visualization and Dashboard module to present the results as overlays or heatmaps on the input image. Following this, these findings are kept safe in a database/storage system for future use and trend analysis. The continuous monitoring of crowd trends is performed by the Monitoring and Alerts block, as it automatically sends out alerts in cases where congestion or unusually high densities are found. This integrated workflow guarantees that effective, real-time crowd monitoring will be performed appropriate for event management, smart cities, and public safety applications.

IV. IMPLEMENTAION SETUP

A comprehensive software-based framework for real-time crowd density monitoring, visualization, and estimation constitutes the implementation of the proposed Crowd Density Analysis System. Its architecture integrates modules on data acquisition, preprocessing, deep learning inference, and visualization. The high scalability and accuracy of the implementation enable effective crowd estimation in highly crowded environments. For model training, testing, and visualization, standard deep learning tools and publicly available benchmark datasets were used to ensure the reproducibility and consistency of performance. The details of the software implementation of the system are addressed in the next subsection.

A. Software Implementation

Crowd Density Analysis System was developed solely on the Python programming language because it has powerful deep learning and image processing capabilities. The main framework is PyTorch, which offers flexibility in neural network construction and training. Based on the Congested Scene Recognition Network architecture, the system includes the following: a dilated convolutional backend for density map generation and a truncated VGG-16 frontend for feature extraction. Other supporting libraries used are OpenCV for tasks involving preprocessing, NumPy for numerical operations, and Matplotlib for visualization.

Preprocessing created ground-truth density maps using Gaussian kernels, normalized pixel intensities, and resized images to 512×512 pixels in preparation for training. Images from Shanghai Tech Part A & Part B and the Mall Dataset are used in this work, which contain numerous scenes with different crowd densities and environmental conditions for training. For the evaluation of prediction accuracy, the Mean Squared Error loss function was utilized, while the Adam optimizer, which has a learning rate of 1×10−5, can ensure stable convergence in the training process. The experiments were carried out with Google Colab Pro, an NVIDIA T4 GPU (16 GB), to accelerate computation.

The modular architecture of the system permits seamless transitioning between deployment and training. A Streamlit-based user interface was built with features for image and video stream uploads to get real-time outputs of crowd counts and density maps. It also provides a crowd analysis-by-region dashboard with an alerting system on estimated density higher than safety thresholds. Simplicity, modularity, and extensibility are considered in software design to ensure the system is easily integrable with current surveillance systems or installed on local servers or cloud environments. On the whole, the implementation of software demonstrates the practical way in which the deep learning technique can be applied to solve crowd analysis problems. The proposed system provides accurate, real-time crowd estimation, which may help with crowd management for public safety, smart city infrastructure, and event scenarios by fusing interactive visualization tools with CSRNet's powerful feature extraction capabilities.

V. RESULTS AND ANALYSIS

A. Density Map Output.

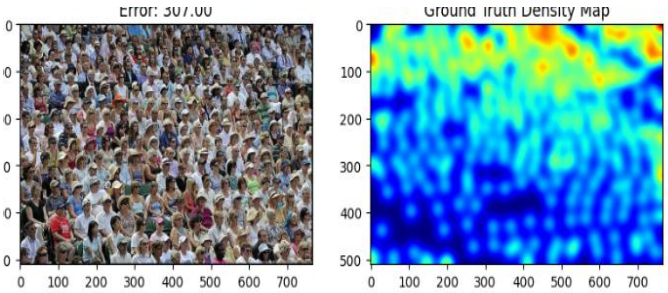

Figure-2 Displays the Predicted Density Map as a result of Crowd Density Analysis with Deep learning.

An output from the CSRNet (Congested Scene Recognition Network) model can be observed in figure 2. High density groups are illustrated in the left input crowd image and the corresponding ground truth density map is illustrated on the right. Areas on the density map that are colored red provide evidence of higher crowd density. This is achieved by using color intensity (blue to red) to show the density of people. While areas that indicate density are not visible to head detection, the total count (even for high density areas) in essence is correct due to the unity of integration over the density map. This prepares the context for future head-count analysis and proves the functionality of CSRNet density estimation module.

B. Density Map Output.

Figure-3 Crowd Counting Result

In Figure 3, you can see the real-time head detection and counting being performed from surveillance footage captured in a shopping center. Each head detected is represented by a red dot, while the total count representing the number of people detected is displayed in the top left. The algorithm finds heads at different scales and distances so it can work in many lighting and occlusion contexts. This real-time implementation effectively enables the live monitoring of crowd density and flow, which is crucial for public safety personnel, crowd control, and for use in smart surveillance systems.

C. Live Camera Implementation (Laptop Camera)

Figure-4 Head Counting using Laptop camera

In Figure 4, it depicts the system's live deployment with a laptop camera. The above model displays the total detected heads detected (e.g. "Head Count: 4"), while overlaying the red dot markers on each of the detected heads in real time when multiple individuals are detected. This test demonstrates the versatility of trained model in uncontrolled, real-world contexts, where dynamic changes in lighting and orientations of poses are present. The model demonstrates excellent efficiency and response for indoor crowd monitoring applications like in offices, labs, and classrooms.

VI. CONCLUSION AND FUTURE SCOPE

It effectively demonstrates how deep learning-based methods can be integrated into a crowd analysis-based head counting system. To achieve individual counting without full-body detection, the suggested method combines a CSRNet model for creating crowd density maps with a YOLO-based model for head detection. Even in complex scenarios like changing lighting, background noise, and partial occlusions, the provided method is capable of estimating crowd size in real-time video streams and camera inputs. The system is ideal for applications involving safety monitoring, crowd control, and surveillance because it functions both indoors and outdoors, according to the experimental results.

The system can be updated in the future to use multiple cameras to handle highly dense situations where there is considerable overlap crowds. Employing temporal tracking algorithms can enable real-time motion analysis and anomaly detection. Further, this enables deployment on edge/embedded devices, such as drones or smart surveillance units, optimized with lightweight architecture models for real-time monitoring in public areas. Defence, public safety, and intelligent surveillance are just a few fields where the system can be applied by making the dataset more varied in crowd conditions and enhancing rotated or partially visible head detection.

REFERENCES

[1]Alotaibi, Sultan Refa, et al. "Integrating Explainable Artificial Intelligence with Advanced Deep Learning Model for Crowd Density Estimation in Real-world Surveillance Systems." IEEE Access (2025).

[2] M. M. Menevse, et al , "An Analysis Approach for Crowds with Varying Density and Complexity Using Image Data," 2025 29th International Conference on Information Technology (IT),
Zabljak, Montenegro, 2025, pp. 1-4, Doi:
10.1109/IT64745.2025.10930264.

[3] Kamra, Vikas, et al. "A Novel Approach for Crowd Analysis and Density Estimation by Using Machine Learning Techniques." 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS). IEEE, 2024.

[4] M. Ibrahim, et al , "Deep Learning in Crowd Analysis: A Survey of Current Research Trends," 2023 17th International Conference 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," in IEEE Access, vol. 12, pp. 121983-121993, 2024.
doi:10.1109/ACCESS.2024.3390049

[6] L. Rajendran, et al, "Bigdata Enabled Realtime Crowd Surveillance Using Artificial Intelligence And Deep Learning," 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea:
(South), 2021, pp.129-132, doi
10.1109/BigComp51126.2021.00032.