

Human Counting And Tracking In Dense Population

Dr. Lubna Kauser¹, MD Furqaan Ul Haque², Mohammed Anzar Hussain³, Nigewan Noormohammad Hussain⁴, Vishnu⁵

¹ Assistant Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Khaja Bandanawaz University, Kalaburagi, India.

^{2,3,4,5} Student, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Khaja Bandanawaz University, Kalaburagi, India.

ABSTRACT

With the rapid growth of urban populations and increasing use of surveillance systems, monitoring crowded environments has become an important requirement for ensuring public safety and efficient crowd management. Traditional surveillance systems rely on manual monitoring of CCTV cameras, which is time-consuming and prone to human error. Therefore, there is a need for intelligent systems that can automatically detect and analyze crowd behavior in real time.

This project presents an AI-based system for human counting and tracking in dense populations using computer vision and deep learning techniques. The system utilizes the YOLOv8 object detection model to detect people present in video frames captured from surveillance cameras. The model processes each frame, identifies individuals by drawing bounding boxes around them, and calculates the total number of people present in the monitored area.

The proposed system extracts frames from the video stream, applies deep learning-based object detection, and displays the detection results visually along with the crowd count. The system is capable of processing video data in real time, making it suitable for applications such as public surveillance, crowd management at events, transportation hubs, and smart city monitoring.

The experimental results demonstrate that the system can effectively detect and count people in crowded environments with good accuracy and performance. The proposed approach reduces the need for continuous manual monitoring and provides an efficient automated solution for crowd analysis and surveillance.

Future improvements may include integration with smart city infrastructure, advanced crowd behavior analysis, and deployment on cloud-based monitoring platforms for large-scale applications.

Keywords – AI, Tracking, Python, YOLO

I. INTRODUCTION

1.1 General introduction

Public gatherings including malls, train stations, airports, stadiums, and public events have seen a dramatic uptick in the number of attendees in recent years, thanks to fast urbanisation and population development. For the sake of public safety and effective crowd control, monitoring these congested areas has grown into a crucial need. CCTV cameras, which are traditionally overseen by security guards by hand, are the backbone of traditional surveillance systems. Watching several video feeds by hand, however, is laborious and error-prone.

The development of AI and computer vision has allowed for the real-time analysis of video feeds by automated systems, allowing them to extract valuable information. Object detection systems have been greatly enhanced by deep learning approaches. YOLO, is an excellent object identification technique that can quickly and accurately identify several things in a picture.

An intelligent system for monitoring crowds using deep learning and computer vision methods is the subject of the "Human Counting and Tracking in Dense Population" research. The YOLOv8 object detection model is used to evaluate video feeds from security cameras and identify any humans present in the scene. The number of persons in the monitored region is counted by the system after the individuals are spotted.

In addition to crowd detection and counting, the system can also analyze demographic information such as

gender distribution and age estimation. The collected information is displayed on a monitoring dashboard that allows users to observe crowd conditions in real time. This automated system helps reduce manual workload and improves the efficiency of surveillance operations.

The proposed system can be applied in several real-world scenarios including crowd monitoring in public places, event management, transportation hubs, and smart city surveillance systems.

II. LITERATURE SURVEY

2.1 Research Paper -1

Author:

Zhiheng Ma, et.al

Title:

“Bayesian Loss for Crowd Count Estimation with Point Supervision”

Year:

2019 (Presented at ICCV 2019)

Overview:

This paper introduces a Bayesian Loss function designed for improving precision & heftiness of crowd enumeration models trained using point-level annotations. Traditional crowd counting methods generate density maps by convolving annotated head points with fixed or adaptive Gaussian kernels. However, this approach introduces bias because the chosen kernel width is often arbitrary and may not accurately represent true crowd distributions, especially under varying scene perspectives or densities. For addressing this, researchers projected probabilistic, Bayesian formulation as directly models the uncertainty of point annotations and optimizes the expected crowd count in a principled manner. This method eliminates the dependence on handcrafted Gaussian density maps and provides a mathematically consistent way to train crowd counting networks using only point-level supervision.

2.2 Research Paper-2

Author:

Xinkun Cao et.al

Title:

“SANet for Precise and Effective Crowd Counting”

Year:

2018 (Presented at ECCV 2018)

Overview:

SANet was proposed to improve both accuracy and efficiency in crowd counting tasks with focus on problem of scale discrepancy—a common challenge in crowd scenes due to perspective distortion and varying distances. Unlike earlier multi-column approaches (e.g., MCNN, Switch-CNN), *SANet* integrates multi-scale feature fusion within a single, unified encoder–decoder architecture. The model aggregates multi-scale contextual features using Scale Aggregation Modules (SAMs), allowing it to adaptively learn scale-aware representations without relying on multiple separate CNN branches. Furthermore, *SANet* introduces a novel Local Pattern Consistency (LPC) loss alongside the traditional Euclidean loss to ensure that generated density maps not only predict accurate counts but also preserve local structural patterns, leading to visually coherent and high-quality density maps.

2.3 Research Paper-3

Author:

Yuhong Li, et.al

Title:

"CSRNet: Dilated CNN's for Understanding Highly Congested Scenes"

Year:

2018 (Presented at CVPR 2018)

Overview:

The CSRNet model represents a significant advancement in crowd counting using deep CNN's. It focuses upon understanding highly congested scenes where traditional CNNs often fail due to limited receptive fields and loss of spatial information. In order to avoid downsampling and increase model's ability to acquire long-range contextual information, CSRNet employs the idea of dilated (atrous) convolutions to expand receptive field while preserving spatial resolution.

2.4 Research Paper-4

Author:

Deepak Babu et.al

Title:

"Switch-CNN for Crowd Counting"

Year:

2017 (Presented at CVPR 2017)

Overview:

The *Switch-CNN* model addresses as toughest challenges in crowd counting — intra-image density and scale variation. Traditional CNN-based models, such as MCNN, employ fixed receptive fields, which limits their ability to accurately handle scenes where different regions exhibit drastically varying densities (e.g., sparse areas and extremely congested zones within the same image). To overcome this, Switch-CNN introduces a dynamic routing mechanism through a *switch classifier* that intelligently assigns local image patches to the most suitable CNN regressor, depending on the estimated local crowd density. This allows the network to adapt to heterogeneous crowd distributions within a single image effectively.

2.5 Research Paper-5

Author:

Yingying Zhang et.al

Title:

"Single-Image Crowd Counting via MCNN"

Year:

2016 (Presented at CVPR 2016)

Overview:

The paper introduces a pioneering approach for crowd counting using DL. Whereas conventional approaches which relied on hand-crafted features and perspective normalization, MCNN directly learns to estimate crowd density from a single image using convolutional neural networks. This approach effectively addresses challenges caused by variations in crowd density, perspective distortion, and scale differences within an image. By employing multiple columns with different receptive fields, the network captures information across varying scales, making it more robust in diverse crowd scenes.

Each column processes the same input image independently to extract multi-scale features. The outputs of all three columns are then concatenated and passed through a final convolutional layer to generate the density map, which represents the spatial distribution of people within the image.

Across all datasets, MCNN achieved significant improvements in terms of MAE and MSE compared to traditional feature-based methods. It became a foundational baseline for density-map-based counting tasks, setting new state-of-the-art results at time of publication.

III. OBJECTIVES

1.1 Objectives of the Project

1. Developing an automated crowd monitoring system

Project aims for designing system which can automatically monitor crowded environments using computer vision techniques.

2. To implement real-time video processing

The system processes video input from surveillance cameras and analyzes each frame in real time.

3. Detecting individuals using the YOLOv8 object detection model

The system uses YOLOv8 deep learning algorithm to accurately detect people present in the scene.

4. To count the number of people present in a monitored area

System calculates total number of detected individuals to estimate crowd density.

5. To visualize results using a monitoring dashboard

The system displays detection results and crowd statistics through a user-friendly dashboard interface.

IV. METHODOLOGY

4.1 Existing System

Traditional crowd monitoring systems mainly rely on CCTV cameras and human operators to observe video streams. Security personnel manually monitor the surveillance footage to identify people present in specific region or detect suspicious activities.

Although CCTV surveillance systems are widely used in public places such as airports, railway stations, shopping malls, and stadiums, they have several limitations. Continuous monitoring of multiple video feeds is difficult for human operators and may lead to fatigue and reduced attention over time.

Another major limitation of traditional systems is that they do not provide automated crowd analysis. Human operators must manually estimate crowd density and identify overcrowding situations, which can result in errors and delayed responses.

Furthermore, traditional surveillance systems do not provide additional analytics such as demographic information, crowd statistics, or real-time data logging. As a result, these systems are inefficient when dealing with large-scale crowd monitoring.

4.2 Proposed System

An artificial intelligence (AI) crowd monitoring and analytics system is presented in the suggested system. This system use deep learning methods to identify and quantify individuals in densely populated areas.

In real time, system examines and analyses video feeds that are recorded by security cameras. In order to identify people in scene, YOLOv8 object detection model is used to every frame of video.

System determines the overall population in monitored region when each person is spotted. Gender breakdown, age approximation, and data from crowd surveillance are some of the supplementary analytics offered by system.

The results of the detection process are displayed on a monitoring dashboard that allows users to observe crowd conditions and access recorded data.

The proposed system offers several advantages over traditional surveillance systems:

- Automated detection and counting of individuals
- Real-time crowd monitoring
- Improved detection accuracy using deep learning
- Reduced workload for security personnel
- Availability of crowd analytics and demographic insights

Suggested system offers a smart way to keep an eye on busy areas by integrating computer vision, DL, and

data analytics.

4.3 System Architecture

Projected method follows structured architecture that includes multiple phases of data processing and analysis. The main components of the system are:

1. Video Input Module
2. Frame Processing Module
3. YOLOv8 Object Detection Model
4. Crowd Counting Module
5. Output Visualization Module

The video input module receives video streams from a camera or recorded video file. The frame processing module extracts individual frames from the video and prepares them for analysis. The YOLOv8 model processes each frame and detects people present in the image. The crowd counting module calculates the total number of detected individuals. Finally, the output visualization module displays the detection results with bounding boxes and the total crowd count.

AI-Based Crowd Monitoring System

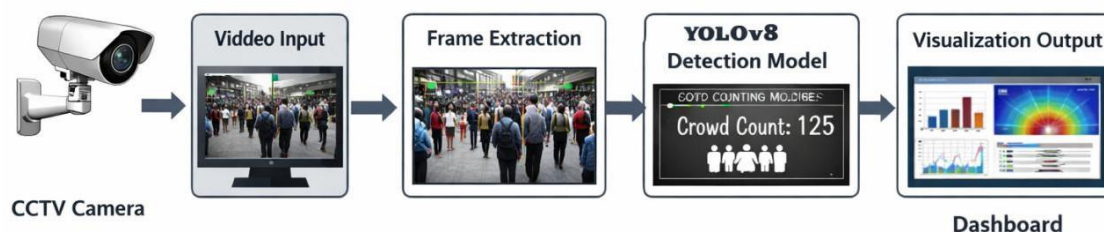


Figure 1 System Architecture of Crowd Monitoring System

4.4 Data Collection

Data collection is an important step in developing a computer vision-based system. In this project, video data containing people in crowded environments is used as the primary input.

The data used for the system could be attained via many sources such as:

- Public surveillance database
- Online crowd monitoring datasets
- Recorded video footage
- Sample videos used for research purposes

These datasets contain images and video frames of people present in different environments such as streets, shopping malls, public events, and transportation hubs.

Using diverse datasets helps the deep learning model learn to detect individuals under numerous conditions like:

- Different lighting conditions
- Varying crowd densities

- Different camera angles and perspectives

4.5 Data Preprocessing

To ensure that input data to DL module is of high quality and consistency, preprocessing processes are used before it is fed in model. Data preprocessing ensures that the model receives properly formatted images for accurate detection.

The preprocessing steps include:

Frame Extraction:

Video streams are divided into individual frames so that each frame can be analyzed separately.

Image Resizing:

Extracted frames are resized into size suitable for the YOLOv8 model.

Normalization:

Pixel values of images are normalized for improving efficacy of DL module.

Noise Reduction:

Unnecessary noise or distortions in the images are minimized to enhance detection accuracy.

These preprocessing steps help in improving efficiency and accuracy of object detection process.

4.6 YOLOv8 Object Detection Model

Suggested method for person detection in video frames relies on YOLOv8 model. Acronym YOLO refers to fact that model only needs to run neural network on whole picture once.

By dividing the picture into a grid and predicting bounding boxes and class likelihoods for each grid cell, YOLO differs from typical object identification algorithms which scan several parts of an image independently. The model is able to accurately and quickly recognise several objects at once because of this. The advantages of using YOLOv8 include:

- High detection accuracy
- Real-time processing capability
- Ability to detect multiple objects simultaneously
- Efficient deep learning architecture

The YOLOv8 model has been trained on large datasets that include images of various objects, including people. During detection, the model identifies individuals and draws bounding boxes around them in the video frames.

4.7 Crowd Detection and Counting

After detecting people in the video frames, the next step is counting number of individuals present in monitored area. The crowd counting module calculates total number of detected people based on the bounding boxes generated by the YOLOv8 model.

Each bounding box represents a detected person. The system counts these detections and density in real time. If the number of detected individuals exceeds a certain threshold, the system can alert authorities about potential overcrowding situations.

V. RESULTS AND DISCUSSION

5.1 Introduction

This section provides thorough evaluation of work carried out in the project “**Human Counting and Tracking in Dense Population.**” Performance of the proposed system was analyzed by testing it with different video inputs containing varying crowd densities. The YOLOv8 object detection model was used to detect individuals in video frames and calculate count of people present.

Results obtained from experiments demonstrate effectiveness of the system in performing real-time crowd detection and counting.

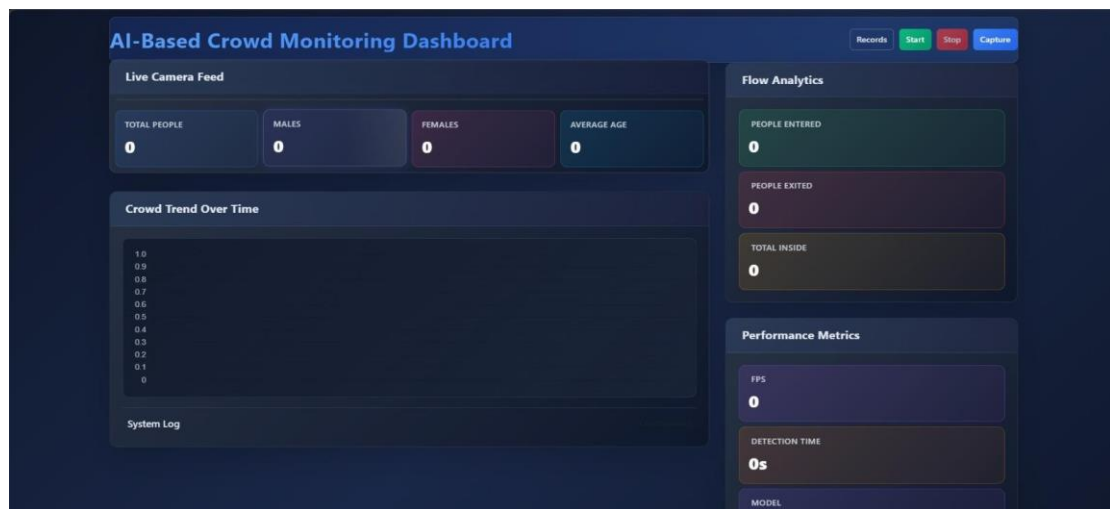


Figure 2 Crowd Monitoring Dashboard

5.2 SYSTEM EXECUTION OUTPUT

The system was executed using Python and OpenCV libraries in the development environment. When the program is executed, the system captures video frames from the camera or video file and processes them using the YOLOv8 object detection model.

During execution, the system continuously analyzes each frame and detects individuals present in the scene. Once a person is detected, system draws bounding box around detected individual.

The detection results are displayed on the output window along with performance information such as detection confidence and processing speed.

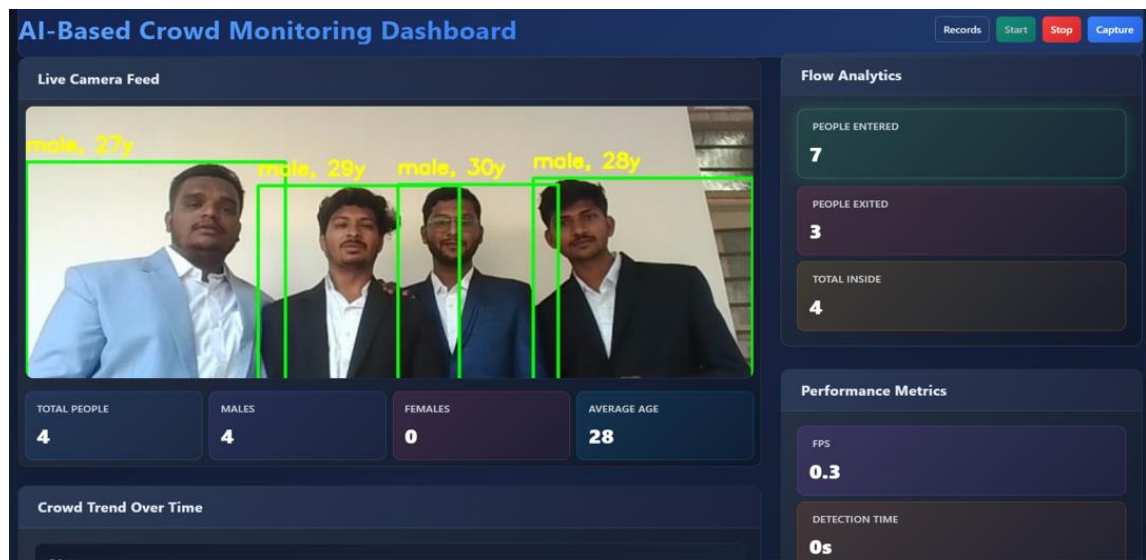


Figure 3 System Execution Output

5.3 Crowd Detection Result

The YOLOv8 object detection model identifies people present in the video frames and highlights them using

bounding boxes. Each bounding box corresponds to a detected individual.

This visualization allows users to easily observe the people present in the monitored environment. The detection results demonstrate that the system can accurately identify individuals even in moderately crowded scenes.

The bounding boxes generated by the model indicate the location of each detected person along with the confidence score of the detection.

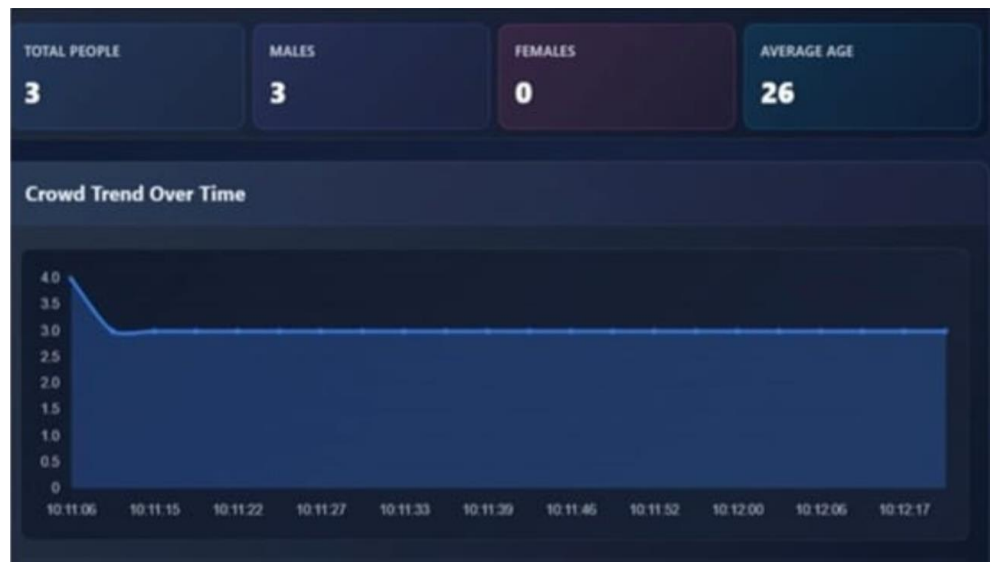


Figure 4 Crowd Detection Output

5.4 Crowd Counting Result

After detecting individuals in the frame, module calculate total number of people present in monitored area. The crowd count is displayed on the screen along with the detection output.

The crowd count is continuously updated as new frames are processed. This allows the system to monitor crowd density in real time and provide accurate information about the number of people present.

This feature is particularly useful for monitoring crowded places like transportation hubs, and public events.

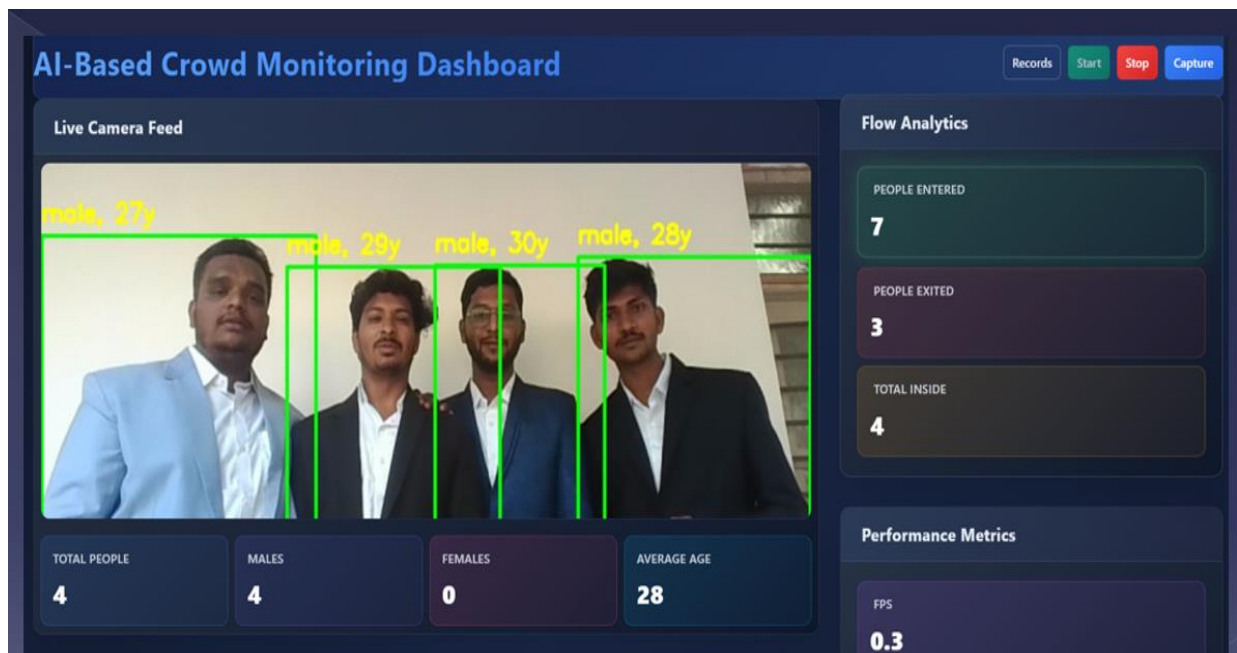


Figure 5 Crowd Counting Output

5.5 Performance Analysis

Various video inputs with varying crowd densities were used to test the proposed system's performance.

When tested on video frames, the YOLOv8 detection algorithm accurately identified persons. Consistent detection findings were maintained as system processed frames in real time.

System can reliably monitor congested areas and offer accurate crowd counting findings, according to the trial results.

The performance of the system depends on factors such as lighting conditions, camera resolution, and crowd density. However, the model performs well in most practical surveillance scenarios.



Figure 6 Detection Performance

5.6 Crowd Monitoring Dashboard and Record Analytics

The developed system also includes a Crowd Monitoring Dashboard that stores and displays detailed analytics of the detected individuals in real time. The dashboard records information about the detected crowd and provides useful insights into crowd behavior and movement.

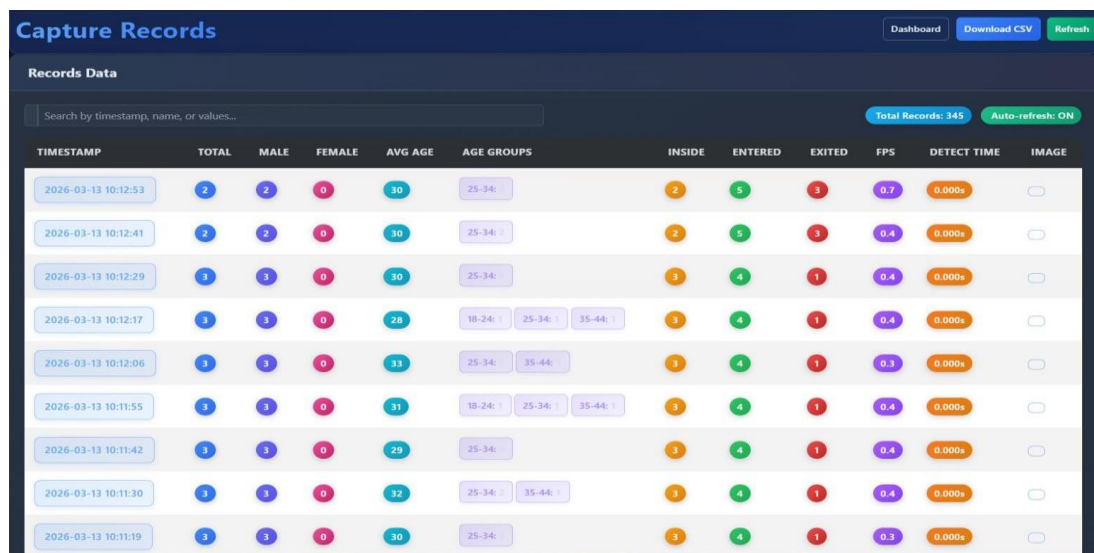
The dashboard maintains a table of captured records that includes information such as timestamp, total number of detected people, gender distribution, average age of detected individuals, age group classification, and movement statistics including the number of people entering and exiting the monitored area.

The system automatically updates the dashboard whenever new detections occur. This feature enables continuous monitoring and helps authorities analyze crowd patterns effectively.

In addition, the dashboard provides performance metrics such as frames per second (FPS) and detection time, which indicate the efficiency of the detection system.

The dashboard also allows users to download the recorded data in CSV format for further analysis and record keeping.

The implementation of this dashboard improves the usability of the system by transforming raw detection outputs into meaningful crowd analytics.



TIMESTAMP	TOTAL	MALE	FEMALE	AVG AGE	AGE GROUPS	INSIDE	ENTERED	EXITED	FPS	DETECT TIME	IMAGE
2026-03-13 10:12:53	2	2	0	30	25-34	2	5	3	0.7	0.000s	
2026-03-13 10:12:41	2	2	0	30	25-34	2	5	3	0.4	0.000s	
2026-03-13 10:12:29	3	3	0	30	25-34	3	4	1	0.4	0.000s	
2026-03-13 10:12:17	3	3	0	28	18-24; 25-34; 35-44	3	4	1	0.4	0.000s	
2026-03-13 10:12:06	3	3	0	33	25-34; 35-44	3	4	1	0.3	0.000s	
2026-03-13 10:11:55	3	3	0	31	18-24; 25-34; 35-44	3	4	1	0.4	0.000s	
2026-03-13 10:11:42	3	3	0	29	25-34	3	4	1	0.4	0.000s	
2026-03-13 10:11:30	3	3	0	32	25-34; 35-44	3	4	1	0.4	0.000s	
2026-03-13 10:11:19	3	3	0	30	25-34	3	4	1	0.3	0.000s	

Figure 7 Crowd Monitoring Dashboard with Capture Records

5.7 Comparison with Existing Systems

Traditional surveillance systems mainly rely on CCTV cameras that are monitored manually by security personnel. These systems only provide video recordings and do not perform automated analysis of crowd conditions. Human operators must continuously observe the video feeds to estimate crowd density and identify suspicious activities.

In contrast, the proposed AI-based crowd monitoring system provides automated crowd detection and analysis using deep learning techniques.

Projected module also offers additional features such as demographic analysis, dashboard visualization, and data logging. These features allow users to monitor crowd conditions more efficiently and obtain useful insights about crowd behavior.

Compared to traditional systems, the proposed system improves detection accuracy, reduces manual workload, and provides real-time analytics for better decision making.

Table 1 Comparison Between Existing and Proposed Systems

Feature	Existing System	Proposed System
Monitoring	Manual monitoring by operators	Automated detection using AI
Crowd Counting	Estimated manually	Automatically calculated
Accuracy	Depends on human observation	High accuracy using YOLOv8
Real-time Analysis	Limited	Real-time processing
Analytics	Not available	Dashboard analytics available

VI. CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In order to automate the process of crowd monitoring, the "Human Counting and Tracking in Dense Population" project used AI and computer vision methods. The main goal of the research was to create a system that could use deep learning algorithms to identify and count individuals in congested areas.

If surveillance footage contains people, the suggested system can identify them using the YOLOv8 object detection paradigm. Through model processing of each frame, bounding boxes are placed around identified persons, allowing for precise people identification and counting by the system.

Through a visualisation dashboard, the technology also offers supplementary crowd analytics, including gender distribution, age estimate, and monitoring data. Authorities may benefit from these capabilities since they help transform raw video footage into actionable information for controlling congested settings.

The results of the experiments prove that the system is capable of accurately and efficiently counting people in real time. This deployment proves that, in comparison to manual monitoring, deep learning approaches greatly enhance the efficiency of surveillance systems.

When it comes to crowd analysis and monitoring, the suggested technology offers a smart and automatic approach. Public spaces including stadiums, airports, train stations, malls, and public events may greatly benefit from the system's ability to enhance crowd control and safety.

6.2 Future Scope

Although the proposed system performs well in detecting and monitoring crowds, there are several areas where the system can be further improved and expanded in the future.

1. Integration with Automatic Alert Systems

The system can be enhanced by adding automatic alert mechanisms that notify authorities when crowd density exceeds a predefined limit. This will help in taking immediate action to prevent overcrowding and ensure public safety.

2. Development of Crowd Density Heatmaps

Future versions of the system can include heatmap visualization to represent areas with high crowd concentration. This will make it easier for authorities to identify congested zones quickly.

3. Multi-Camera Monitoring System

The system can be extended to support multiple surveillance cameras simultaneously. This will allow monitoring of large areas such as entire buildings, campuses, or city zones from a centralized system.

4. Advanced Crowd Behavior Analysis

The system can be improved by incorporating behavior analysis to detect unusual activities or abnormal crowd movements. This can help in identifying potential threats and improving security measures.

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788.
- [2] Y. Zhang, D. Zhou, S. Chen, S. Gao and Y. Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 589–597.
- [3] Y. Li, X. Zhang and D. Chen, "CSRNet: Dilated Convolutional Neural Networks for Understanding Highly Congested Scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 1091–1100.
- [4] X. Cao, Z. Wang, Y. Zhao and F. Su, "Scale Aggregation Network for Accurate and Efficient Crowd Counting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 734–750.
- [5] D. B. Sam, S. Surya and R. Venkatesh Babu, "Switching Convolutional Neural Network for Crowd Counting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 4031–4039.
- [6] G. Bradski, "The OpenCV Library," *Dr. Dobbs's Journal of Software Tools*, vol. 25, no. 11, pp. 120–126, 2000.
- [7] Python Software Foundation, "Python Documentation," [Online]. Available: <https://www.python.org/doc/>.
- [8] Ultralytics, "YOLOv8 Object Detection Model Documentation," [Online]. Available: <https://docs.ultralytics.com>.
- [9] OpenCV, "Open Source Computer Vision Library Documentation," [Online]. Available: <https://docs.opencv.org>.
- [10] Z. Ma, X. Wei, X. Hong and Y. Gong, "Bayesian Loss for Crowd Count Estimation with Point Supervision," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019.