International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-17

Crowd Counting and Density Hotspot Detection Using YOLOv11 and K-Means Clustering

Riya Abhyankar, Arundhati Melinkeri, Trupti Mahajan, Dr. Sinu Nambiar
Department of Artificial Intelligence and Data Science
Marathwada Mitra Mandal College of Engineering Pune, Maharashtra, India

Abstract. Urban populations are rapidly growing with large scale public events , hence monitoring the crowd behaviours and count has become a necessity for modern surveillance systems. An accurate crowd count helps to estimate number of people in each area while anomaly detection helps identifying situations such as overcrowding, abnormal patterns. The rpaper will present the YOLOv11 object detection algorithm and apply it with K-means clustering to count crowds and detect anomalies. The proposed system aims to provide a simple yet effective mechanism for real-time crowd analysis by leveraging detection, clustering, and visualization techniques. The experiment results demonstrate the ability of the system to measure crowd size and identify crowded zones, making it a useful tool for surveillance purposes in urban environments.

Keywords: Crowd counting, Anomaly detection, Density estimation, Heatmaps, K-Means clustering, YOLOv11.

I. Introduction

In the modern world, where cities are rapidly expanding, large numbers of people gather in places such as festivals, shopping centres, sports games, and music concerts. The number of people attending these events could range from thousands to millions. The latest version of the YOLO series, YOLOv11, offers accurate object detection capabilities along with enhanced feature extraction and extremely fast inference processing. Paired with K-means clustering, it identifies individuals by analysing their spatial coordinates and grouping them into distinct clusters.

Cameras capture people at varying distances, causing drastic differences in apparent size. We tackle this with Dynamic Region Division, an adaptive method that handles both nearby and distant subjects seamlessly.

II. Literature Survey

In this paper, we investigate the possibility of detecting anomalous individuals in crowds based on a neural network YOLOv11 architecture.The authors have developed a method called TransCAM, combining Gradient-weighted Class Activation Mapping and Transformer-based attention maps. It allows for improving interpretability by highlighting key areas of a crowded scene by making it possible to predict anomalies based on a learning algorithm. The model performs decently when identifying bikers and skaters in crowds. Its efficiency is confirmed by 98.5% score on the tests. Testing was done using the UCSD Ped2 dataset and certain tricks like data augmentation, such as image flips and changes in brightness.

This paper describes the use of a novel approach involving the use of a network based on two types of data input – visual and thermal – in order to count individuals in crowds. For this purpose, the authors have designed an attention module allowing the network to pay attention to the most relevant areas. In the end, the network successfully performs the task, regardless of how difficult the conditions may be. Tests of the algorithm were carried out on several datasets; results have been compared with similar works. The network showed better results than other approaches to the problem.

This research is about using a kind of model to count people in stores and adjust the temperature to save energy. The researchers used a tuned YOLOv11 model to detect and count people in different parts of the store. The model can tell if the crowd is small, medium or big. Then it adjusts the temperature to make people comfortable and save energy. The researchers made a web application using Dash. It saved a lot of energy. The study shows that deep learning can help make stores more efficient and sustainable.

This paper is about a kind of model that is lightweight and can count people in crowds in real-time. The model uses a kind of backbone that is good at capturing the big picture. The researchers added a module that combines techniques to handle crowds that are very dense. The model is fast and accurate. The researchers tested the model on some datasets. It did better than other models.

This review is about the history of using learning to count people in crowds. The researchers looked at over 300 studies. Categorized them. They found that some models are better than others. The researchers also talked about the challenges of counting people in crowds. They said that it is hard to handle scales and perspectives. They also said that there is a gap between real-world data. The researchers provided some tools and codes to help researchers.

This research is about a kind of model that can count people in crowds without using a lot of labeled data. The researchers used a -supervised approach that combines attention and information enhancement. The model can tell the difference between people and the background. The researchers tested the model on some datasets. It did well. The model is good at handling low-quality images.

This paper is about a kind of model that is specifically designed for video-based crowd counting. The researchers used a kind of convolution that captures the movement of people. The model can handle the noise in the annotations. The researchers tested the model on some datasets. It did better than other models.

This research is about a kind of algorithm that can handle the problem of perspective distortion in subway surveillance. The researchers used a dynamic region division algorithm that can tell the difference between far regions. The model uses techniques for each region. The researchers tested the model on some datasets. It did well.

This survey is about the approaches to crowd analysis. The researchers looked at vision-based and sensor-based methods. They said that each approach has its limitations. The researchers talked about the potential of using sensors to improve crowd monitoring. They also talked about the challenges of counting people in crowds.

This research is about a kind of model that uses a block-based approach to count people in crowds. The researchers used a kind of module that can handle scale variations. The model can capture the contextual density distributions. The researchers tested the model on some datasets. It did well.

This research is about counting the number of people in a crowd using object detection algorithms such as YOLO. The researchers compared models and found that YOLO is the best. The researchers tested the models on some datasets. Yolo did the best. The study shows that YOLO can be used in real-world applications.

This paper is about a kind of model that can generate high-quality density maps. The researchers used a kind of module that can capture the local spatiotemporal relationships. The model can handle movement and background changes. The researchers tested the model on some datasets. It did well.

This research is, about a kind of framework that can detect anomalies in crowds. The researchers used a kind of module that can handle congested and occluded scenes. The model can identify object-level irregularities. The researchers tested the model on some datasets. It did well. The study shows that the framework can be used to support sustainability goals.

III. Proposed Alogorithm

Figure 1 shows integration of YOLOv11 model along with K-means clustering to analyse the crowd density, the count of crowd and the spatial distribution of the crowd.

The system architecture is divided into four modules i.e. the Input Handling, Object Detection and Data Preparation, Crowd Clustering and Data Analysis, And Visualizing with Output Generation.

1. Input Handling :
The system begins the initialization process by loading the required libraries , weights and the model used for training i.e. the YOLOv11 model. The user provides a singular input in the form of image or video. It detects the type of input and processes the data accordingly.

2. Object Detection and Data Preparation :
Each frame is processed using the YOLOv11 model and the people within the crowd are detected. Using K-means clustering we calculate the coordinates of the people within the crowd scene. These coordinates are organized into a data matrix which forms the input for clustering analysis.

3. Crowd Clustering and Data Analysis :
The coordinates calculated from the detected individuals are processed. They are then grouped into clusters based on their proximity.

For each cluster, the centroid is calculated. The crowd density for each cluster is calculated based on the number of individuals within each other. This helps in identifying the concentration of people within the density zones.

4. Visualization and Output Generation

A density heatmap is generated based on the spatial distribution of the clusters. High - density areas are represented by warm colors like red or yellow. Low density areas are represented by cooler tones such as blue, green, violet etc. The number of people within each cluster are also displayed in the density heatmap as output.

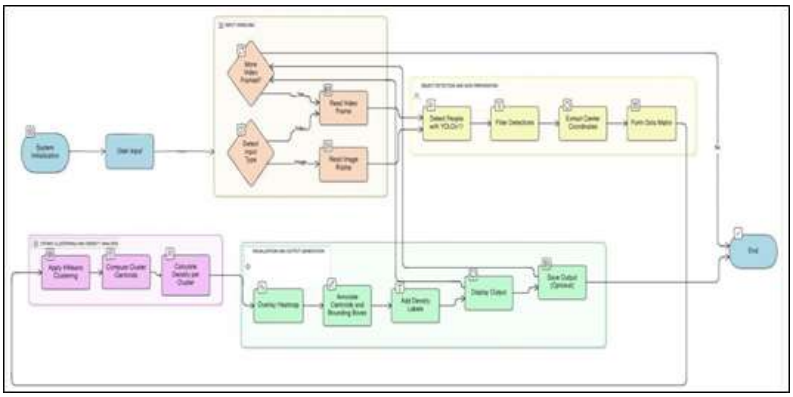

Figure 1 – System Architecture

IV. Implementation

- This work presents an inference and evaluation pipeline for a detection-based counting task. The system takes labeled input samples, runs them through a trained model, and evaluates the predictions using a mix of regression and classification metrics.
- The pipeline begins with data preprocessing, where inputs are formatted for the model. A trained model (likely built using frameworks such as PyTorch or Tensor-Flow) then performs inference to generate predictions, including detected objects and their counts. These outputs are handled again, often with thresholding or filtering, to make final predictions.
- Standard metrics are used to do the evaluation. MAE and RMSE tell you how far off predicted counts are from the real counts. Precision, recall, and F1-score tell you how good the detection is. Exact match accuracy is another parameter that checks whether if the predicted counts are the same as the true counts.
- The system is implemented in Python, using tools such as NumPy, Pandas, and Scikit-learn, Ultralytics and developed in a Jupyter Notebook environment.

Figure 2 – Frontpage of Crowd Analysis site

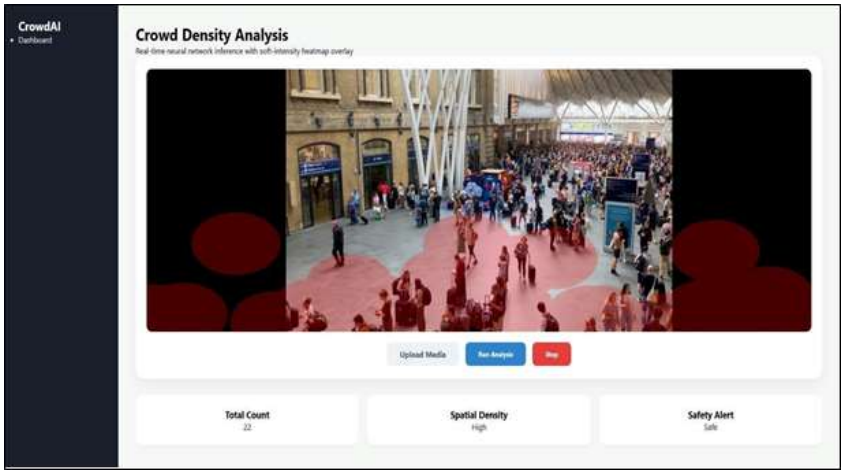

Figure 3 - Crowd count and detection

V. Gap Analysis

Table 1 – Gap Analysis table

Metric	Value	Interpretation
Mean Absolute Error (MAE)	3.00	Predictions off by 3 units on average
Root Mean Square Error (RMSE)	3.00	Consistent error (no large outliers)
Average Accuracy %	88.00%	High overall performance
Precision (Detection)	1.0000	No false positives
Recall (Detection)	1.0000	No missed detections
F1-Score (Detection)	1.0000	Perfect detection balance
Exact Count Match Accuracy	0.0000	Counts never exactly correct
Total Samples	2	Extremely small dataset

Merits :

- Strong detection performance
- Precision, Recall, and F1-score of 1.0 indicate the model reliably identifies objects.
- Consistent predictions
- Equal MAE and RMSE suggest stable, predictable errors (no extreme outliers).
- Well-defined evaluation pipeline
- Uses a combination of regression and classification metrics, giving a multi-dimensional view of performance.

Limitations:

- Poor counting accuracy
- Exact match accuracy = 0 and MAE = 3 show the model cannot count objects correctly.

- Very small dataset (n = 2)
- Results are not statistically reliable or generalizable.
- Misleading overall accuracy (88%)

High accuracy does not reflect true performance due to limited data.

- Rigid evaluation metric
- Exact match accuracy fails to account for near-correct predictions.
- Systematic bias in predictions
- Consistent error indicates the model may always overcount or undercount.
- Lack of detailed error analysis
- No per-sample insights, making debugging and improvement difficult.

VI. Future Scope

- Use additional data for obtaining more accurate outcomes:

As mentioned before, currently, there are only two images used for evaluation, which is not enough for receiving reliable information about the accuracy of the algorithm.

- Refine the model in terms of object counting:

While the algorithm detects objects without any problems, it fails to provide an adequate count. The refinement of predictions' filtering and merging can significantly improve this process.

- Simplify the criteria used for evaluating outcomes:

Currently, only exact predictions are considered as good ones. In practice, there is no need for such strict evaluation. Instead, small deviations in predictions should be allowed.

- Correct a consistent object counting error:

Apparently, the model provides incorrect object counts constantly (either undercounts or overcounts). In order to overcome this problem, it might be necessary to calibrate the algorithm.

- Add additional error analysis to obtain valuable insights:

The current evaluation method does not provide any detailed information about prediction errors. By analyzing predictions, it will be possible to identify issues with the algorithm.

- Try other models designed for object counting:

Currently, the algorithm relies on object detection techniques, but it might be beneficial to use alternative approaches.

VII. Conclusion

The development in crowd counting and anomaly detection has evolved from simple manual approaches to more sophisticated deep learning algorithms. Recent developments in methods indicate that density estimation methods perform better in case of accurate counting whereas YOLO-based methods provide better performance when detecting anomalies.

The primary improvements are still focused on making these systems faster, more reliable, and simpler to use for people without a lot of expert knowledge.

References

1. Melisa Gozet, Mehmet Karakose, and Asim Egemen Yilmaz, "YOLOv11-Based Explainable Framework for Anomaly Detection in Crowded Scenes Using Attention Fusion", 2025 15th International Conference on Advanced Computer Information Technologies (ACIT)
2. Kunyu Zhou and Hua Yan, "MMFFNet: Multi-Modal Feature Fusion Network for RGB-T Crowd Counting", 2025 IEEE Transactions on Instrumentation and Measurement
3. Santhosh C, Sindhu R, et al., "Real-Time Customer Density Analysis Using Fine-Tuned YOLOv11 for Optimized HVAC Management in Supermarkets", 2025 International Conference on Next Generation Computing Systems (ICNGCS)
4. Mas Nurul Achmadiah, Wen-Kai Kuo, Chi-Chia Sun, and Jun-Wei Hsieh, "RepSFNet: A Single Fusion Network with Structural Reparameterization for Crowd Counting", 2025 IEEE International Conference on Advanced Visual and Signal-Based Systems (AVSS)
5. Guangshuai Gao, Junyu Gao, et al., "A survey of deep learning methods for density estimation and crowd counting", 2025 Vicinagearth
6. Zijian Hu, Yingying Li, et al., "Semi-supervised Crowd Counting Method Based on Attention Mechanism", 2024 17th International Conference on Advanced Computer Theory and Engineering (ICACTE)
7. Yu-Jen Ma, Hong-Han Shuai, and Wen-Huang Cheng, "Spatiotemporal Dilated Convolution with Uncertain Matching for Video-based Crowd Estimation", 2021 IEEE
8. Gaoqi He, Zhenwei Ma, et al., "Dynamic Region Division for Adaptive Learning Pedestrian Counting", 2019 IEEE
9. Mohamed Abdou and Abdelkarim Erradi, "Crowd Counting: A Survey of Machine Learning Approaches", 2020 IEEE
10. Omar Elharrouss, Somaya Al-Maadeed, et al., "Crowd density estimation with a block-based density map generation", 2024 International Conference on Intelligent Systems and Computer Vision (ISCV)
11. Chandradeep Bhatt, Abhay Pratap, et al., "Deep Learning for Crowd Counting: Addressing Crowd Density with Advanced Methods", 2024 Second International Conference on Advances in Information Technology (ICAIT)
12. Li Dong, Haijun Zhang, et al., "CLRNet: A Cross Locality Relation Network for Crowd Counting in Videos", 2024 IEEE Transactions on Neural Networks and Learning Systems.
13. Rabia Nasir, Zakia Jalil, et al., "An enhanced framework for real-time dense crowd abnormal behavior detection using YOLOv8", 2025