Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

An Intelligent Person Detection System for Public Spaces Using YOLOv11

P. Paul Aditya

Department of Computer Science and Engineering, University College of Engineering Kakinada (A), Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India

Abstract: The increasing use of camera-based monitoring in public spaces requires person detection systems that can operate across changes in illumination, viewpoint, background complexity, crowd density, and object scale. This paper presents an intelligent person detection framework based on YOLOv11m for open-area surveillance. A custom dataset containing more than 6,000 annotated outdoor images was prepared to represent diverse scene conditions and person distributions. The images were used to train the YOLOv11m detector, which performs person localization through bounding boxes and confidence scores in a single-stage detection pipeline. The trained model was integrated with a Streamlit application to provide image upload, preprocessing, detection visualization, performance reporting, and optional person-count and occupancy information. Evaluation used Precision, Recall, mAP@50, Box Loss, and inference time. The reported final evaluation values were 89.41% Precision, 87.22% Recall, 88.21% mAP@50, and 0.499 Box Loss. A representative application inference reported demonstration environment. Training curves showed progressive improvement in Precision, Recall, and mAP@50 with a corresponding reduction in Box Loss. The resulting framework provides an integrated approach for person detection and performance analysis in varied outdoor public-space scenes.

Keywords: YOLOv11m, Person Detection, Object Detection, Deep Learning, Public-Space Surveillance, Real-Time Detection, Outdoor Surveillance, Computer Vision.

I. INTRODUCTION

The rapid development of Artificial Intelligence (AI) has brought significant improvements to computer vision, particularly in the analysis of images and video streams. Deep learning techniques have made it possible to develop systems that can automatically recognize and locate objects within visual data. Among the different approaches used for this purpose, Convolutional Neural Network (CNN)-based detectors have become widely adopted because they can learn relevant visual patterns directly from training data. The YOLO family of object detection models is particularly useful when detection results are required with low processing latency.

Automatic identification of people is an important capability in intelligent monitoring applications. Such systems can support activities including public-space observation, crowd management, campus protection, traffic supervision, smart-city services, and emergency situations. The availability of high-resolution cameras and increasingly capable GPUs has also made it practical to perform person detection with shorter inference times. In addition, transfer learning and larger annotated datasets allow detection models to learn features that remain useful across different scenes and environmental conditions.

Despite these developments, detecting people in open and outdoor environments remains challenging. A model trained primarily on restricted or controlled scenes may not perform consistently when exposed to changes in illumination, weather, camera viewpoint, background complexity, object scale, or the degree of occlusion. Dense groups of people can further complicate detection because individuals may partially overlap or appear at different distances from the camera. These factors can lead to missed detections or inaccurate localization, reducing the reliability of conventional surveillance solutions in practical outdoor settings.

Artificial Intelligence and computer vision have enabled automated analysis of images and video for applications that previously depended heavily on manual observation. Deep learning based object detectors can learn visual representations directly from annotated examples and can localize multiple objects within a scene. The YOLO family is particularly suitable for applications in which detection and processing speed are important because localization and classification are performed within a unified detection pipeline [11], [18].

Person detection is a fundamental capability for intelligent public-space monitoring. It can support crowd observation,

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

campus security, transportation-hub monitoring, smart-city services, emergency response, and other situations in which the presence and location of people must be identified from visual data. However, open outdoor scenes introduce substantial variation in illumination, weather, camera viewpoint, background appearance, person scale, and occlusion. Dense groups can further increase the difficulty because individuals may overlap or appear only partially visible [4], [5].

Vision-based occupancy and human-detection research has therefore developed along several directions, including object detection, crowd counting, occupancy estimation, thermal sensing, and multimodal approaches [1], [6], [8], [10]. Although these approaches demonstrate the usefulness of visual sensing, practical deployment still requires a detector that can identify individual persons while maintaining suitable computational behaviour. Privacy is also relevant in camera-based monitoring, making it desirable to process visual information in a controlled application environment and to avoid unnecessary storage or exposure of personal information [9], [20].

The work presented in this paper focuses on a YOLOv11m-based person detection system designed for outdoor and open public spaces. The system uses a custom dataset containing more than 6,000 annotated images, integrates model inference with a Streamlit interface, and reports both visual detections and quantitative performance. The main contributions are: (1) preparation of a diverse outdoor person-detection dataset; (2) development of a YOLOv11m training and inference pipeline; (3) integration of detection, visualization, and performance reporting in an interactive application; and (4) evaluation using Precision, Recall, mAP@50, Box Loss, and inference time.

II. RELATED WORK

Occupancy detection and people-counting systems have been investigated using computer vision, deep learning, sensors, and multimodal sensing. Reviews of occupancy detection describe camera-based methods as an important non-contact sensing option, while also noting the effects of illumination, viewpoint, occlusion, and scene configuration [1], [6], [13]. Crowd-counting research has similarly shown that density, scale variation, and overlapping persons make reliable visual estimation challenging [4], [5].

Recent work on deep learning has expanded the use of learned visual representations for object and occupancy detection. Survey studies covering 3D occupancy prediction and deep-learning-based object and occupancy detection highlight the importance of robust feature extraction and scene understanding when people occur at different depths and scales [3], [10]. Advances in model optimization and deployment have also increased the practicality of deep neural networks for real-world visual applications [17].

Privacy remains an important consideration for camera-based occupancy systems. Thermal imaging, low-resolution approaches, labeling-free methods, and privacy-aware learning have been investigated to reduce exposure of identifiable visual information [9], [20]. These methods demonstrate that detection accuracy should be considered together with the operational and privacy requirements of the application.

The smart room occupancy detection using neural network based approaches and transfer learning techniques [8] and also examines the use of transfer learning with ResNet, VGG and MobileNet. Privacy-preserving approaches for occupancy counting [10], with emphasis on methods that minimize the amount of personally identifiable visual information captured during detection.[11] deep learning approaches for three-dimensional object detection and occupancy are discussed .

YOLO-based approaches have received particular attention for real-time human detection and counting because they combine localization and classification in a single forward pass [18]. The present work follows this direction but concentrates on person detection in outdoor public-space scenes using YOLOv11m and a custom dataset. Rather than relying only on model output, the proposed framework also provides an interactive interface and a structured performance-analysis component.

2.1 Research Gap

Existing occupancy and person-detection studies demonstrate strong progress, but a practical outdoor monitoring system must address several requirements simultaneously: variation in scene conditions, multiple-person localization, measurable detection quality, manageable inference behaviour, and accessible result presentation. Methods designed for controlled environments may not represent the variability found in open public spaces. The proposed framework addresses this gap by combining a diverse outdoor dataset, YOLOv11m detection, post-inference filtering, visual result generation, and application-level performance reporting in one workflow.

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

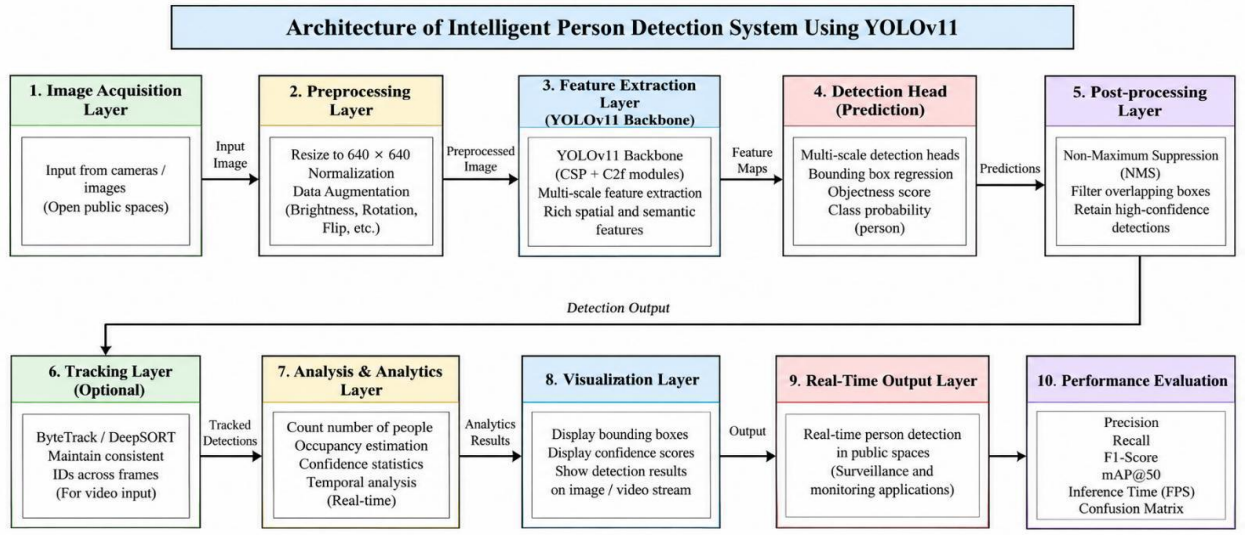


Figure 1: Architecture of the proposed YOLOv11m-based person detection system

III. PROPOSED METHODOLOGY

The proposed system is organized as an end-to-end image-based detection pipeline. The main stages are image acquisition and preprocessing, YOLOv11m inference, detection filtering and bounding-box generation, result visualization, and quantitative performance evaluation. The architecture in Fig. 1 separates these functions while maintaining a direct flow from user input to the final detection report.

At the input stage, the user supplies an outdoor image through the Streamlit interface. The image is validated and prepared for model inference. Preprocessing ensures that the input is compatible with the trained YOLOv11m configuration and supports consistent processing across different images. The prepared image is then passed to the trained detector.

During inference, YOLOv11m extracts hierarchical visual features and produces object predictions containing bounding-box coordinates, confidence values, and class information. Since the application is intended for person detection, predictions associated with the person class are retained. Confidence and Intersection over Union (IoU) settings are used during prediction to control the quality of accepted detections. The final detections are rendered as bounding boxes with confidence values so that the spatial locations of people can be interpreted directly from the image.

After detection, the application generates a visual result and supporting statistics. The original and annotated images can be presented together, while the application can report the number of detected persons, average confidence, inference time, and occupancy percentage when the corresponding functionality is enabled. The framework also records model-performance information so that detection quality can be examined beyond a single visual example.

3.1 Dataset Preparation

The training data consist of more than 6,000 annotated outdoor images representing different illumination levels, backgrounds, camera viewpoints, person scales, and crowd distributions. Each image contains annotations for individual persons so that the detector can learn both the visual characteristics of people and their spatial extent. The dataset was prepared to provide greater environmental diversity than a narrowly controlled collection.

Data preprocessing and augmentation are used to improve the variability of training examples. The objective is to expose the model to plausible changes in appearance and scene conditions so that learned features are not tied to one particular background, lighting condition, or viewpoint. The prepared data are supplied to YOLOv11m for supervised object-detection training.

Table 1: Dataset and training configuration

Parameter	Configuration
Dataset type	Custom outdoor person-detection dataset
Annotated images	More than 6,000
Target class	Person

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

Model	YOLOv11m
Training duration	120 epochs shown in reported training curves
Evaluation metrics	Precision, Recall, mAP@50, Box Loss
Application	Streamlit-based interactive interface

3.2 YOLOv11m Training and Inference

YOLOv11m is used as the primary detection network because the proposed task requires simultaneous localization and classification of multiple persons. The training stage uses the annotated dataset to learn person-related visual features. During inference, the trained weights are loaded and the model processes each input image in a single detection pipeline. The output contains bounding boxes, class information, and confidence values for the detected instances.

The methodology maintains a fixed inference configuration during evaluation so that performance measurements are comparable across test images. Predictions are filtered using the configured confidence and IoU thresholds, after which the retained boxes are used for visualization and application-level statistics. This controlled sequence separates model inference from result presentation and supports repeatable analysis.

3.3 Performance Evaluation

The proposed model is evaluated using Precision, Recall, mAP@50, and Box Loss. Precision indicates the proportion of reported detections that are correct, whereas Recall reflects the proportion of actual persons that are successfully detected. mAP@50 summarizes detection performance using an IoU threshold of 0.50. Box Loss provides an indication of localization error during training, with lower values generally corresponding to improved bounding-box alignment.

In addition to detection-quality metrics, inference time is considered because the system is intended for interactive use. The combination of accuracy-related measures, localization loss, and processing time provides a broader view of the behaviour of the trained detector than any single metric alone.

IV. SYSTEM IMPLEMENTATION

The system implementation as a real-time person detection application based on YOLOv11m, with the objective of identifying and localizing people in outdoor and open-area environments. The implementation integrates dataset preparation, image preprocessing, model training, inference, detection visualization, and performance analysis into a unified workflow. A custom dataset containing more than 6,000 annotated outdoor images is used to train the detection model. The images represent different environmental conditions, including variations in illumination, weather, background complexity, camera viewpoints, person scale, and crowd density.

The implementation combines the trained YOLOv11m model with a Streamlit. The interface accepts an image upload, forwards the image to the preprocessing and inference modules, and presents the resulting annotated image and performance information. This design keeps the user interaction layer separate from the model-processing layer and allows the same trained weights to be reused without exposing the underlying implementation details.

The detection module loads the trained model, performs prediction at the configured image size, applies confidence and IoU filtering, and extracts person detections. For each retained prediction, the system records the bounding-box coordinates and confidence score. The annotated output is then generated by drawing the detected regions and their labels on the input image. The result can be stored for subsequent inspection and documentation.

During application execution, an image uploaded by the user is temporarily retained within the application workflow and passed to the preprocessing and detection modules. After inference is completed, the processed image containing the detected bounding boxes and corresponding confidence scores can be stored in the designated output directory for subsequent examination and analysis. The trained YOLOv11m weights are maintained in the best.pt file, which is loaded when performing detection, while the model-training statistics are maintained in results.csv for subsequent performance analysis and visualization.

After detection, the generated bounding boxes and confidence values are overlaid on the original image to produce an annotated output. The processed image is then presented through the Stream lit interface along with a detection summary containing the detected-person count, confidence information, inference time, and occupancy level. The system additionally estimates the occupancy percentage using the predefined capacity of the monitored area. This complete process enables the application to provide an interactive and efficient mechanism for real-time person detection in open-area surveillance scenarios.

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

The application also includes performance reporting. Training data are represented using metric curves for Precision, Recall, mAP@50, and Box Loss. These plots make it possible to observe learning progression across epochs and identify whether detection quality improves while localization error decreases. The reported final metrics are displayed together with the visual detection output so that quantitative and qualitative evidence can be interpreted together.



Figure 2: Example person detection output generated by the proposed system

4.1 End-to-End Workflow

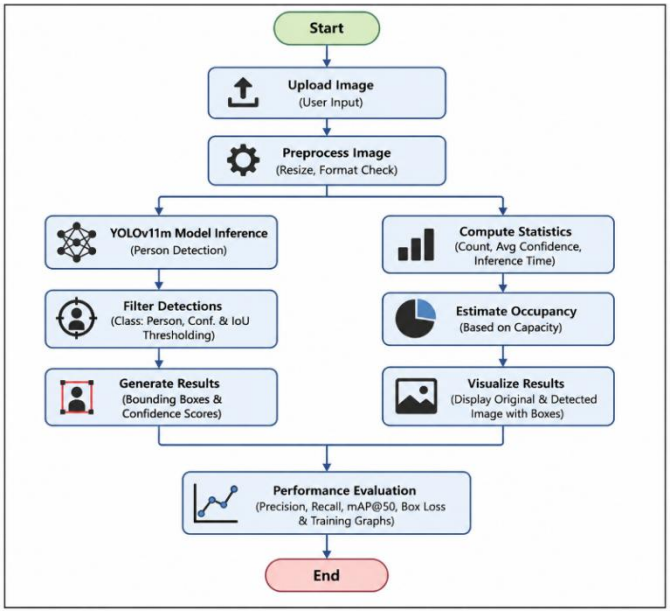


Figure 3: End-to-End workflow of the model

The end-to-end operation of the proposed system as illustrated in Figure 3 starts when the user provides an outdoor image through the Stream lit-based web application. The uploaded file is first checked to confirm that its format is supported by the application. Once the validation is completed, the image is forwarded to the preprocessing stage, where it is resized to meet the input requirements of the trained YOLOv11m model.

The prepared image is subsequently supplied to the YOLOv11m detection module for inference. The model analyzes the image and identifies person instances by generating their corresponding bounding-box coordinates and confidence scores. As the application is intended for person detection, only predictions associated with the Person class are retained. Confidence and Intersection over Union (IoU) thresholds are applied to remove detections that do not satisfy the predefined criteria.

Once valid detections have been obtained, the system determines the total number of persons present in the image. It also derives supporting information such as the average confidence score, inference time, and occupancy percentage, with the latter being determined using the predefined capacity of the monitored area. These values provide additional information for interpreting the detection results and assessing the processing efficiency of the system.

The detected image is annotated by placing bounding boxes and their corresponding confidence values around the identified persons. The resulting annotated image and detection statistics are displayed through the Streamlit interface, enabling the user to examine the original input and processed output conveniently.

V. RESULTS AND DISCUSSION

The developed YOLOv11m-based person detection system produces the detection results through the Stream lit web application after an input image is uploaded by the user. The trained model analyses the image and identifies persons present in the scene by generating a separate bounding box for each detected individual. A corresponding confidence score is associated with every detection, providing an indication of the model's prediction reliability.

For result interpretation, the application presents both the original input image and the processed image containing the detected persons. Along with the visual output, the system displays key detection statistics, including the total number of detected persons, average confidence score, inference time, and occupancy percentage. These values allow the detection performance and processing efficiency of the proposed system to be examined for different outdoor surveillance scenes.

The annotated detection image is also automatically stored in the designated output directory, providing a persistent record

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

of the generated results for subsequent analysis and documentation. Thus, the person detection output combines visual localization with quantitative detection information, making the results easier to interpret and evaluate within the proposed application.

The reported experiments show that the trained YOLOv11m detector can identify multiple persons in outdoor scenes with corresponding bounding boxes and confidence scores. The representative detection shown in Fig. 2 demonstrates that the system can localize several people even when the scene contains construction-like backgrounds, varying person scales, and partial occlusion.

The final model-performance values reported by the developed application are 89.41% Precision, 87.22% Recall, and 88.21% mAP@50, with a Box Loss of 0.499. These values indicate that the trained detector learned useful person-related features and achieved a balanced level of detection and localization performance on the evaluated data. The reported demonstration also recorded an inference time of 7.954 s for the processed image in the application environment.

The training curves provide additional evidence about the learning behaviour. Precision and Recall generally increase over the training sequence, while mAP@50 shows an overall upward trend despite intermediate fluctuations. Box Loss decreases from approximately 1.35 at the beginning of training to about 0.50 toward the later epochs. The combined behaviour suggests progressive improvement in both detection quality and bounding-box localization.

Performance should nevertheless be interpreted in relation to the evaluation setup. Inference time depends on hardware, software configuration, image dimensions, model execution mode, and application overhead. Likewise, the reported metrics are specific to the prepared dataset and evaluation configuration. Broader deployment would therefore require testing on additional environments, camera viewpoints, and crowd conditions.

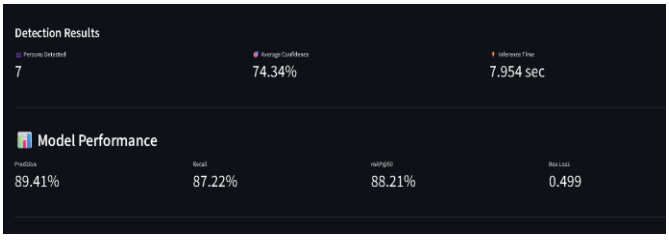


Figure 4: Reported detection and model-performance values in the Streamlit application

Table 2: Reported performance values of the proposed system

Metric	Reported value
Precision	89.41%
Recall	87.22%
mAP@50	88.21%
Box Loss	0.499
Representative inference time	7.954 s

5.1 Training Behaviour

Figure 4 shows the Precision and Recall curves across the training epochs. Precision begins near 0.70 and reaches approximately 0.89 toward the end of training, while Recall increases from approximately 0.68 to 0.87. The curves contain local fluctuations, but the overall trend is upward. Such fluctuations are expected during optimization because batch composition and difficult examples can influence epoch-level performance.

Figure 5 presents mAP@50 and Box Loss. mAP@50 rises from approximately 0.55 to 0.88, while Box Loss decreases from approximately 1.35 to 0.50. The reduction in localization loss alongside increasing mAP@50 indicates that the model progressively improves its ability to place bounding boxes around persons. The late-stage curves also suggest that the model continues to refine its learned representation during the later training epochs.

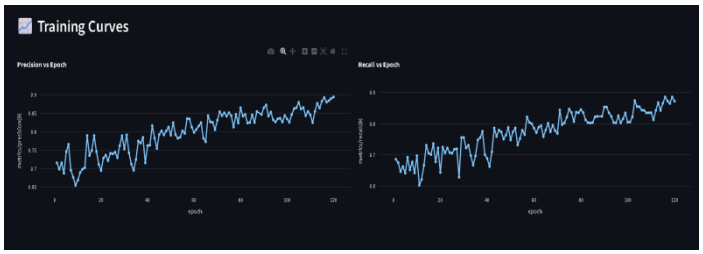


Figure 5: Precision and Recall training curves

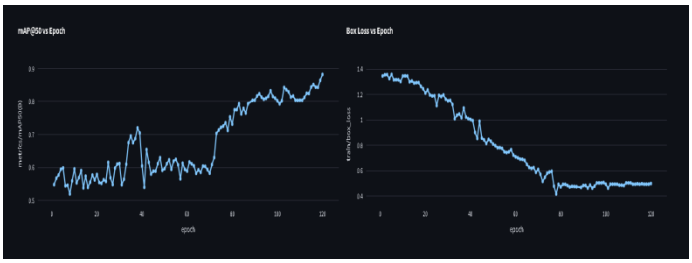


Figure 6: mAP@50 and Box Loss training curves

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

5.2 Discussion

The results indicate that dataset diversity and a dedicated person-detection training process can support reliable detection across varied outdoor scenes. The model is not evaluated solely through a single accuracy value; instead, Precision, Recall, mAP@50, and Box Loss are considered together. This is important because a detector may achieve high precision while missing some people, or detect many people while producing more false positives. Considering several measures gives a more complete description of model behaviour.

The integration of the detector with Streamlit also provides a practical layer around the trained model. Users can upload an image and immediately inspect the localized detections and performance information. The framework therefore connects model development with an accessible application workflow, which is useful for demonstrations, testing, and further system development.

VI. CONCLUSION

This paper presented an intelligent person detection system for public spaces using YOLOv11m. The proposed framework combines a custom outdoor dataset of more than 6,000 annotated images, YOLOv11m-based object detection, confidence and IoU filtering, visual result generation, and Streamlit-based interaction. The design addresses the practical need to identify multiple persons under variations in lighting, background, viewpoint, scale, and crowd distribution.

The implementation demonstrates that the YOLOv11m model is capable of accurately detecting multiple persons in diverse outdoor scenarios by generating precise bounding boxes with high confidence scores. The use of a large and diverse outdoor dataset enables the model to effectively learn person-related features under varying lighting conditions, backgrounds, viewing angles, and crowd densities, thereby improving its robustness and generalization capability.

The integration of an interactive web interface, automated performance evaluation, and visualization dashboard further enhances the usability of the system. Users can easily upload outdoor images, perform person detection, and analyze model performance through graphical representations and quantitative evaluation metrics. This makes the proposed system both a practical surveillance application and a valuable research platform for studying deep learning-based object detection.

The reported results show 89.41% Precision, 87.22% Recall, 88.21% mAP@50, and 0.499 Box Loss. The training curves show an overall increase in detection metrics and a decrease in localization loss across the training process. These findings support the use of the developed framework as a foundation for outdoor person-detection applications.

Future development can investigate larger and more diverse datasets, additional challenging crowd scenarios, faster inference configurations, video-stream processing, multi-object tracking, privacy-aware processing, and edge deployment. Evaluation on independent datasets and different camera systems would also help determine how well the trained model transfers to unseen public-space environments.

REFERENCES

[1] P. Chaudhari, A. Kulkarni, and R. Patil, “Fundamentals, algorithms, and technologies of occupancy detection in smart buildings: A comprehensive review,” *IEEE Access*, vol. 12, pp. 45678–45705, 2024.

[2] H. Xu, Y. Wang, and Z. Li, “A survey on occupancy perception for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 2, pp. 1450–1472, 2024.

[3] Y. Zhang, M. Chen, and T. Liu, “Vision-based 3D occupancy prediction: A review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 1, pp. 112–130, 2025.

[4] L. Deng, J. Li, and S. Zhou, “Deep learning in crowd counting: A survey,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2150–2170, 2024.

[5] M. Alhawsawi and K. Deb, “A comprehensive survey of crowd density estimation and counting approaches,” *IEEE Access*, vol. 13, pp. 22110–22135, 2025.

[6] H. Elkhoukhi, A. Saad, and M. Benammar, “Occupancy sensing technologies in smart buildings: A review,” *IEEE Sensors Journal*, vol. 24, no. 8, pp. 10234–10256, 2024.

[7] E. El-Kenawy, M. Ibrahim, and A. Khodadadi, “Smart room occupancy detection using neural networks and transfer learning: A review,” *IEEE Access*, vol. 13, pp. 33412–33429, 2025.

[8] N. Francis and R. James, “Thermal imaging-based occupancy detection: Methods and privacy considerations,” *IEEE Sensors Journal*, vol. 25, no. 2, pp. 1560–1578, 2025.

[9] J. Jeong, H. Kim, and S. Lee, “Privacy-preserving occupancy counting: A review of labeling-free and low-

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 9, pp 49-56, September-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.309006
Available Online at https://www.jaiet.com

resolution approaches," *IEEE Internet of Things Journal*, vol. 12, no. 3, pp. 1987–2005, 2025.

[10] M. Valverde, A. Ortiz, and F. Garcia, "Deep learning-based 3D object and occupancy detection: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 1, pp. 540–558, 2025.

[11] C. Tian, X. Zhang, and L. Sun, "Recent advances in deep learning: Models, optimization, and deployment," *IEEE Access*, vol. 13, pp. 45000–45025, 2025.

[12] A.Alharthi and M. Hussain, "Advances in crowd counting and density estimation using deep learning," *IEEE Access*, vol. 11, pp. 98765–98790, 2023.

[13] R. Kumar and P. Singh, "Vision-based building occupancy estimation: Current trends and future directions," *IEEE Access*, vol. 12, pp. 60012–60030, 2024.

[14] S. Wong and K. Lim, "Vision-based detection systems for smart infrastructure: A review," *IEEE Sensors Journal*, vol. 23, no. 15, pp. 17045–17063, 2023.

[15] Y. Sun, L. Zhao, and Q. Huang, "Sensor fusion techniques for indoor occupancy estimation: A review," *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 7805–7822, 2024.

[16] T. Rahman and J. Park, "Multi-modal deep learning for occupancy detection in smart buildings," *IEEE Access*, vol. 12, pp. 88221–88240, 2024.

[17] D. Chen, F. Xu, and H. Zhao, "Edge deployment of deep learning models for smart building applications: A survey," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 1, pp. 302–320, 2025.

[18] K. Patel and R. Mehta, "Real-time human detection and counting: A review of YOLO-based approaches," *IEEE Access*, vol. 11, pp. 112340–112360, 2023.

[19] J. Li, W. Wu, and X. Tang, "Vision-based occupancy detection and energy optimization in buildings: A review," *IEEE Transactions on Automation Science and Engineering*, vol. 22, no. 2, pp. 890–905, 2025.

[20] P. Singh and V. Sharma, "Privacy-aware deep learning methods for camera-based indoor occupancy detection," *IEEE Access*, vol. 13, pp. 72110–72128, 2025.

Citation of this Article:

P. Paul Aditya. (2026). An Intelligent Person Detection System for Public Spaces Using YOLOv11. *Journal of Artificial Intelligence and Emerging Technologies (JAIET).* 3(9), 49-56. Article DOI: https://doi.org/10.47001/JAIET/2026.309006

***** End of the Article *****