A YOLO-Driven Pedestrian Detection Framework for Vehicle-to-Everything Networks with Enhanced Accuracy

Syahid Anuar
Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
syahid.anuar@utm.my

Raad Hmmood Afiet
Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
afiet@graduate.utm.my (corresponding author)

Received: 20 February 2026 | Revised: 13 March 2026, 2 April 2026, 5 April 2026, and 6 April 2026 and | Accepted: 10 April 2026

 DOI: https://doi.org/10.48084/etasr.18286

ABSTRACT

Accurate pedestrian detection is a critical requirement for autonomous vehicles to enhance road safety and support intelligent transportation systems. This study presents a pedestrian detection framework designed for Vehicle-to-Everything (V2X) environments, where real-time perception and reliable decision-making are essential. The proposed framework integrates a YOLOv8-based deep learning model with temporal behavior modeling to improve detection performance in dynamic urban scenarios. The model is evaluated using the CityPersons benchmark dataset and standard object detection metrics, including precision, recall, Intersection over Union (IoU), and mean Average Precision (mAP). The experimental results demonstrate that the proposed approach achieves an mAP@0.5 of 0.838, indicating a strong balance between detection accuracy and reliability. Further analysis using precision-confidence and recall-confidence curves shows that the model maintains stable detection performance while reducing false-positive predictions. In addition, a comparative evaluation with established object detection models confirms the effectiveness of the proposed framework for pedestrian detection tasks. The results indicate that the proposed approach improves real-time pedestrian recognition in V2X environments and contributes to enhancing road safety and traffic management in autonomous driving systems.

Keywords-Vehicle-to-Everything (V2X); autonomous vehicles; pedestrian detection; deep learning; YOLOv8; precision-confidence; recall-confidence

I. INTRODUCTION

Pedestrian detection is a fundamental capability for autonomous vehicles to prevent collisions and enhance road safety. Reliable detection enables vehicles to identify pedestrians in complex urban environments, thereby reducing accident risks and improving situational awareness. Modern perception systems utilize multiple sensing technologies, including radar, cameras, and Light Detection and Ranging (LiDAR), to detect surrounding objects. These sensing modalities are commonly integrated into Advanced Driver Assistance Systems (ADAS), which aim to improve safety for drivers, passengers, pedestrians, and other road users. Among various ADAS functionalities, pedestrian recognition has become a critical component and is widely adopted in autonomous driving systems. Despite significant advances in perception technologies, pedestrian detection remains a challenging task. Real-world traffic environments involve varying illumination conditions, occlusions, low-resolution imagery, and complex pedestrian behaviors [1-4]. Traditional computer vision approaches often struggle to address these challenges due to their reliance on handcrafted features. In contrast, Deep Learning (DL) techniques have demonstrated superior performance in object detection and recognition by learning hierarchical feature representations from large-scale datasets. Accurate pedestrian detection is particularly essential for Intelligent Transportation Systems (ITS), where reliable identification of pedestrians, whether walking, standing, or crossing, is critical to safe vehicle operation and efficient traffic management [5, 6].

Vehicle-to-Everything (V2X) communication has emerged as a key enabling technology for modern Intelligent Transportation Systems (ITS), allowing vehicles to exchange information with surrounding entities such as other vehicles, infrastructure, and pedestrians [7]. In conventional V2X systems, pedestrian safety often relies on communication between pedestrian mobile devices and vehicular networks to prevent potential collisions [8]. However, vision-based detection methods provide an additional safety layer by directly analyzing visual data captured from vehicle-mounted sensors to

identify pedestrian locations and movements in real time [5]. Recent advances in Artificial Intelligence (AI), Machine Learning (ML), and DL have significantly enhanced perception systems, enabling accurate real-time object detection in complex traffic environments [9]. Recent studies have demonstrated significant improvements in pedestrian detection and behavior analysis using DL techniques in intelligent transportation systems [10, 11]

Various DL-based object detection frameworks have been developed for pedestrian detection, including R-CNN, Fast R-CNN, Faster R-CNN, RetinaNet, and YOLO [12-18]. These models have substantially improved detection performance through Convolutional Neural Networks (CNNs) and advanced feature extraction techniques [15-19]. Among these approaches, the YOLO (You Only Look Once) family has gained significant attention due to its ability to perform object detection in a single forward pass, enabling high detection speed while maintaining competitive accuracy. In addition, several studies have explored enhancements in backbone architectures and detection frameworks to improve pedestrian detection in complex environments [14-19]. YOLO-based models are particularly suitable for V2X applications because they support real-time processing while maintaining reliable detection performance under varying lighting conditions, occlusions, and complex urban geometries [6, 19-20].

Motivated by these developments, this study proposes a YOLOv8-based pedestrian detection framework specifically designed for V2X environments. The proposed framework aims to improve detection accuracy and robustness while maintaining low false-positive rates. The model is evaluated using the CityPersons benchmark dataset and integrated within a simulated V2X environment to assess its effectiveness for real-time pedestrian detection in intelligent transportation systems.

II. CONTRIBUTIONS OF THIS STUDY

The main contributions of this study can be summarized as follows:

1. A YOLOv8-based pedestrian detection framework specifically designed for V2X environments, enabling reliable real-time pedestrian recognition in dynamic urban traffic scenarios.
2. The framework is designed to support future integration of temporal modeling techniques for analyzing pedestrian behavior and motion patterns in dynamic environments.
3. The framework is evaluated on the CityPersons benchmark dataset, employing standard object detection metrics, including precision, recall, Intersection over Union (IoU), and mean Average Precision (mAP).
4. The proposed model is compared with widely used object detection models, including Faster R-CNN, SSD, YOLOv3, and YOLOv5, to validate its effectiveness for pedestrian detection tasks.

III. RESEARCH METHODOLOGY

A. YOLO-Based Model

This study adopts a YOLO-based object detection architecture to develop an effective pedestrian detection system for V2X environments. YOLO performs object detection in a single forward pass by simultaneously predicting bounding box coordinates and class probabilities, enabling real-time detection with high computational efficiency [18]. This characteristic makes YOLO suitable for safety-critical applications, such as autonomous driving, where rapid perception and decision-making are required. Despite its advantages, detecting small or distant pedestrians remains a challenge for many object detection models. To address this limitation, a Feature Pyramid Network (FPN) enables multi-scale feature extraction by combining hierarchical feature representations at different spatial resolutions. This approach improves the model's ability to detect pedestrians of varying sizes and distances in complex urban environments [5, 18].

The integration of the FPN facilitates the fusion of high-level semantic information with fine-grained spatial details extracted from different layers of the network. This multi-scale representation enhances detection performance under challenging conditions, including crowded scenes, occlusions, and varying distances between pedestrians and vehicles. Thus, the proposed model achieves more reliable pedestrian detection in dynamic V2X scenarios where timely and accurate perception is essential for traffic safety [5]. Furthermore, the YOLO architecture operates as a unified end-to-end detection framework, eliminating the need for multi-stage processing typically required in traditional object detection approaches. This design improves computational efficiency and supports real-time inference, which is essential for V2X-enabled intelligent transportation systems where rapid response to pedestrian movements is required [18].

B. Backbone Network

The backbone network is responsible for extracting hierarchical feature representations from input images, which are subsequently utilized by detection layers to predict bounding box coordinates and class probabilities. CNN-based architectures, such as ResNet and Darknet, are commonly employed as backbone networks due to their ability to learn discriminative visual features from complex scenes [20]. In ResNet, residual connections mitigate the problem of vanishing gradients, enabling the training of deeper networks, while Darknet architectures provide an effective balance between computational efficiency and detection accuracy.

In YOLO-based detection systems, Darknet-based backbones are frequently adopted because they offer high processing speed while maintaining reliable feature extraction capabilities. However, alternative backbone architectures such as ResNet can also be employed to improve feature learning and network depth, thereby enhancing detection performance in complex scenarios [18, 20]. These architectures generate multi-level feature maps that encode spatial and semantic information from the input image.

The extracted feature maps are forwarded to subsequent detection components to estimate object locations, objectness scores, and class probabilities [18]. This hierarchical feature representation enables the detection model to identify objects across varying scales and visual conditions.

In the proposed framework, CSPDarknet is employed as the backbone network, combined with Spatial Pyramid Pooling (SPPNet) to enlarge the receptive field and Path Aggregation Network (PANet) to enhance multi-scale feature fusion. In addition, three YOLO detection heads are utilized to perform multi-scale pedestrian detection by processing feature maps at different resolutions. This architecture improves the detection of pedestrians with varying sizes and distances while maintaining computational efficiency in dynamic V2X environments.

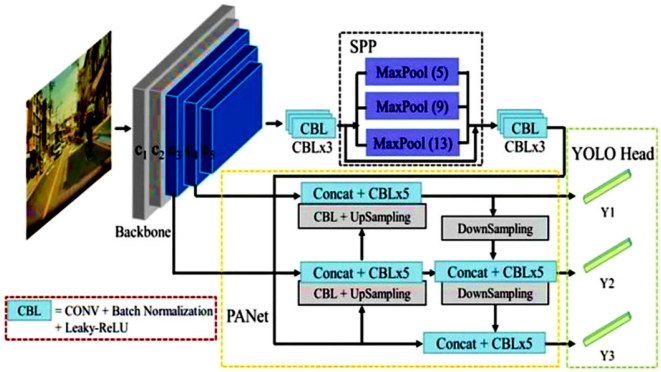

Fig. 1. The main parts of YOLO-based models with the backbone part.

To further improve detection performance across different scales, the proposed framework incorporates an FPN, which enables multi-scale feature extraction by combining hierarchical feature maps generated at different levels of the backbone network. This approach enhances the model's ability to detect pedestrians at varying distances and scales in complex urban environments [5, 18]. In the proposed architecture, the feature pyramid consists of multiple levels (P3, P4, P5, P6, and P7), each corresponding to feature maps at different spatial resolutions. A key component of the FPN structure is the lateral connection, which integrates high-level semantic features from deeper layers with fine-grained spatial features from earlier layers. This fusion is achieved through up-sampling and convolution operations to align feature dimensions and refine feature representations. By combining semantic and spatial information across multiple scales, the FPN enhances contextual feature representation and improves pedestrian detection performance in complex V2X traffic scenarios.

The fusion process can be expressed as:

$$P_i = Upsample(P_i + 1) + Conv(P_i) \quad (1)$$

where P_i represents the feature map at pyramid level i, and $P_i + 1$ denotes the feature map at the next higher semantic level. The up-sampling operation ensures spatial alignment, while the convolution operation adjusts channel dimensions and refines feature representation.

This multi-scale feature fusion improves the detection of small and distant pedestrians by enhancing feature representation across different spatial resolutions. Such capability is particularly important in dynamic V2X environments, where object scale varies significantly due to changes in distance and motion [18].

C. Detection Head

The detection head consists of convolutional layers followed by YOLO detection layers that predict bounding box coordinates, objectness scores, and class probabilities [20]. This component performs object localization and classification by mapping feature representations into detection outputs. The detection process operates on multiple feature scales, enabling the model to detect objects of varying sizes and distances within the scene. Multi-scale detection improves robustness in complex environments where pedestrians appear at different spatial resolutions due to perspective variations and distance from the camera [18]. By utilizing feature maps from different levels of the network, the detection head can accurately localize both small and large objects.

This multi-scale prediction mechanism is essential for applications such as autonomous driving and intelligent transportation systems, where real-time and reliable object detection is required for safe operation. Detecting pedestrians across multiple scales enhances system performance in dynamic V2X environments characterized by varying object sizes and motion patterns [18].

The detection process can be mathematically expressed as:

$$\Omega_i = Conv(P_i) \quad (2)$$

where P_i represents the feature map at pyramid level i, and Ω_i denotes the output tensor containing bounding box coordinates, confidence scores, and class predictions. The convolution operation transforms extracted features into prediction vectors associated with predefined anchor boxes [18]. This multi-scale detection strategy improves localization accuracy and robustness in complex V2X environments, where pedestrian sizes and positions vary dynamically.

D. On-Maximum Suppression (NMS)

After the object detection stage, NMS is applied to eliminate redundant bounding box predictions and retain the most confident detections. NMS is a post-processing technique commonly used in YOLO-based object detection frameworks to remove overlapping bounding boxes corresponding to the same object [18]. This process ensures that each detected object is represented by a single bounding box with the highest confidence score.

Let $B = \{b_1, b_2, \ldots, b_n\}$ denote the set of predicted bounding boxes and $S = \{s_1, s_2, \ldots, s_n\}$ represent their corresponding confidence scores. NMS selects the optimal subset of detections based on a predefined IoU threshold T. The operation can be formally expressed as:

$$\eta = NMS(B, S, T) \quad (3)$$

where B denotes the set of predicted bounding boxes and S represents their associated confidence scores.

During the NMS process, the bounding box with the highest confidence score is selected first. All remaining bounding boxes with an IoU greater than the threshold T relative to the selected box are suppressed. This procedure is repeated iteratively until all candidate detections are processed. By eliminating redundant predictions, NMS improves localization accuracy and enhances detection reliability in dynamic V2X environments [18].

E. CityPersons Dataset

The CityPersons dataset [19] was used to evaluate the performance of the proposed YOLO-based pedestrian detection framework in V2X environments. This is a widely used benchmark for pedestrian detection and has been extensively adopted in DL-based object detection research due to its challenging urban scenarios. It is derived from the Cityscapes dataset and is specifically designed to evaluate pedestrian detection performance in complex real-world environments [19]. The dataset consists of high-resolution images captured from urban street scenes under varying environmental conditions, including different illumination levels, crowded environments, and partial occlusions. Such diversity enables comprehensive evaluation of detection models under realistic traffic conditions, which is essential for intelligent transportation systems and autonomous driving applications. Each image is annotated with detailed pedestrian bounding boxes, including information on visibility and occlusion levels, allowing for accurate assessment of detection performance in challenging scenarios.

The dataset was divided into training and validation subsets following standardized benchmarking protocols, ensuring consistent and reproducible evaluation across different detection approaches. Figure 2 shows example images from the CityPersons dataset, illustrating typical urban scenes with multiple pedestrians and varying background complexity. The standardized evaluation framework provided by CityPersons enables fair comparison with existing pedestrian detection methods and supports reproducible experimental validation in V2X-based intelligent transportation systems.

Fig. 2. Sample images from the CityPersons dataset.

F. Simulation Design and Implementation

The proposed V2X simulation architecture was designed to model communication and interaction among vehicles, infrastructure, and pedestrians within a controlled environment. The simulation framework enables the evaluation of pedestrian detection performance under representative traffic conditions while maintaining controllable experimental parameters. The architecture integrates virtual On-Board Units (OBUs), Roadside Units (RSUs), pedestrian agents, and communication protocols within a centralized simulation environment. This design supports sensing, object detection, and message exchange among simulated entities, allowing validation of traffic management logic, pedestrian detection algorithms, and V2X communication mechanisms. This enables systematic performance evaluation before real-world deployment.

IV. IMPLEMENTATION FRAMEWORK

The simulation platform is implemented using technologies that support system integration, DL inference, and computer vision processing:

- Java was used as the primary development environment for coordinating system components and managing simulation control logic.
- Deeplearning4j (DL4J) was utilized to implement and deploy the YOLO-based pedestrian detection model within the Java ecosystem.
- OpenCV was employed for image preprocessing, feature extraction, and real-time computer vision operations.

The implementation process consists of the following stages.

1. Environment setup: A Java-based simulation environment developed to coordinate vehicle nodes, infrastructure units, pedestrian agents, and communication modules. This setup enables controlled interaction among V2X entities within the simulation framework.
2. DL4J integration: The Deeplearning4j library is integrated into the Java environment to support model loading, inference, and training operations for the pedestrian detection model.
3. YOLO-based model deployment: The YOLO-based pedestrian detection framework, including the backbone network, feature pyramid structure, detection head, and NMS, was deployed within the DL4J environment to perform real-time pedestrian detection within the simulation.

This integrated architecture enables end-to-end simulation of pedestrian detection and V2X communication processes. The framework provides a controlled platform for evaluating detection accuracy, processing latency, and system-level interaction performance in intelligent transportation scenarios.

A. Evaluation Criteria

The performance of the proposed YOLO-based pedestrian detection framework was evaluated using standard object detection metrics that provide quantitative measures of detection accuracy, localization performance, and classification reliability under varying traffic conditions.

- Precision measures the proportion of correctly predicted pedestrian detections among all predicted detections. It reflects the model's ability to minimize false-positive predictions:

  $$Precision = TP / (TP + FP) \quad (4)$$

  where TP denotes true positives and FP denotes false positives.

- Recall measures the proportion of correctly detected pedestrians among all ground-truth instances. It reflects the model's ability to identify all relevant pedestrian objects:

  $$Recall = TP / (TP + FN) \quad (5)$$

  where FN denotes false negatives.

- IoU measures the overlap between the predicted bounding box and the corresponding ground-truth bounding box. It is defined as:

  $$IoU = Area(B\_pred \cap B\_gt)/Area(B\_pred \cup B\_gt) \quad (6)$$

  A detection is considered correct when the IoU exceeds a predefined threshold.

- Mean Average Precision (mAP) is a standard evaluation metric in object detection that summarizes detection performance across different recall levels. In this study, mAP@0.5 is used, where detections are considered correct when IoU ≥ 0.5.

B. Training Configuration and Implementation Details

The YOLOv8 model was trained using the CityPersons dataset, with all input images resized to 640×640 pixels. The dataset was partitioned according to the official training and validation splits provided by the benchmark.

Training was conducted for 100 epochs with a batch size of 16 and an initial learning rate of 0.001. The Adam optimizer was employed to update the model parameters. Data augmentation techniques, including random horizontal flipping, scaling, and color jittering, were applied during training to improve model generalization and reduce overfitting.

The experiments were performed on a workstation equipped with an NVIDIA RTX 3060 GPU, 16 GB RAM, and an Intel Core i7 processor, implemented using the Deeplearning4j (DL4J) framework within a Java environment, while OpenCV was utilized for image preprocessing and computer vision operations.

V. RESULTS AND DISCUSSION

Figure 3 shows the Precision-Recall (PR) curve, demonstrating the detection performance of the model at varying recall levels. The proposed model achieved an mAP@0.5 of 0.838, indicating a balanced trade-off between precision and recall in pedestrian detection tasks. This curve demonstrates that the model maintains high precision across most recall levels, reflecting its ability to accurately detect pedestrians while minimizing false-positive predictions. As recall approaches 1.0, a slight reduction in precision is observed, which is consistent with the inherent trade-off in object detection systems, where maximizing recall may introduce additional false-positive detections. Overall, the results indicate that the proposed model provides stable and reliable detection performance, making it suitable for real-time pedestrian recognition in V2X-enabled intelligent transportation environments.

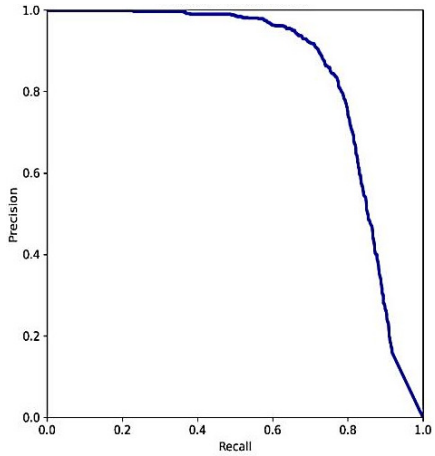

Fig. 3. PR curve of the proposed pedestrian detection model.

A. Experimental Comparison Setup

To evaluate the effectiveness of the proposed YOLOv8-based pedestrian detection framework, a comparative analysis was conducted with several widely used object detection models, including Faster R-CNN, SSD, YOLOv3, and YOLOv5. The performance values of these baseline models are indicative results derived from the literature and are used for comparative discussion purposes. In contrast, the proposed YOLOv8-based model was trained and evaluated using the CityPersons dataset under the experimental configuration described. Table I presents the comparative results.

TABLE I. COMPARATIVE PERFORMANCE OF PEDESTRIAN DETECTION MODELS

Model	Backbone	mAP@0.5	Precision	Recall
Faster R-CNN	ResNet-50	0.801	0.83	0.77
SSD	VGG-16	0.784	0.80	0.75
YOLOv3	Darknet-53	0.812	0.85	0.79
YOLOv5	CSPDarknet	0.825	0.87	0.81
Proposed YOLOv8 model	Enhanced Multi-scale	0.838	0.89	0.83

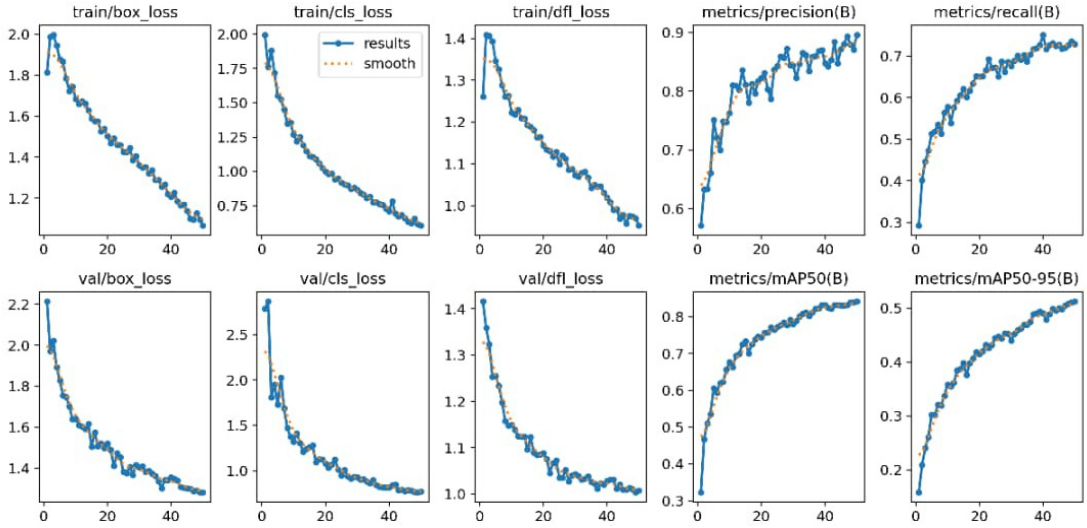

Fig. 4. Training and validation of the proposed pedestrian detection model.

B. *Confidence Analysis*

Further evaluation of the model performance was conducted using Precision-Confidence and Recall-Confidence curves, as illustrated in Figure 4. The Precision-Confidence curve shows a rapid increase in precision as the confidence threshold increases, reaching a maximum value of 1.00 at a confidence level of 0.829. This indicates that high-confidence predictions correspond to highly reliable pedestrian detections with minimal FP rates. In contrast, the Recall-Confidence curve exhibits a decreasing trend as the confidence threshold increases. This behavior reflects the inherent trade-off between precision and recall in object detection systems, where higher confidence thresholds improve prediction reliability, but may exclude some TP detections, resulting in reduced recall. These results indicate that the proposed YOLOv8-based pedestrian detection model maintains a balanced trade-off between detection accuracy and reliability. Such a balance is essential for real-time pedestrian detection in V2X environments, where both accurate recognition and dependable predictions are required to support safe decision-making.

Figure 5 presents the normalized confusion matrix of the proposed YOLO-based pedestrian detection model, illustrating classification performance across two classes: pedestrian and background. The confusion matrix provides insight into the model's capacity to correctly identify pedestrian instances while minimizing misclassification. As shown in the matrix, 77% of pedestrian instances are correctly classified, indicating effective detection of pedestrian objects in the dataset. However, 23% of pedestrian instances are misclassified as background, representing FN detections. These errors are primarily associated with challenging scenarios, such as partial occlusion, long-distance visibility, and complex urban backgrounds. In contrast, the background class is classified with 100% accuracy, indicating that the model effectively avoids FP detections where background regions are incorrectly identified as pedestrians. This result demonstrates the model's strong capability in distinguishing pedestrians from non-pedestrian objects.

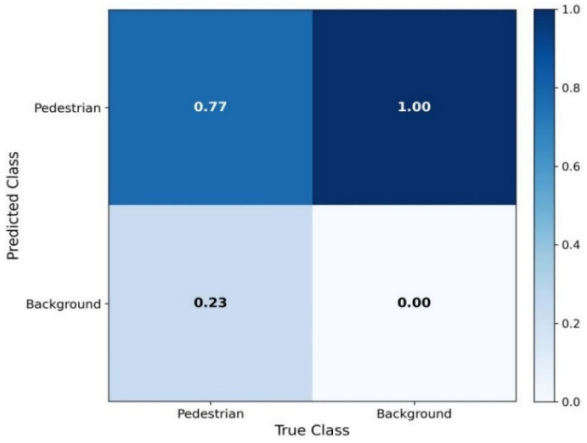

Fig. 5. Confusion matrix of the proposed pedestrian detection model.

Figures 6 and 7 illustrate the Precision-Confidence and Recall-Confidence curves of the YOLO-based pedestrian detection model, respectively. The Precision-Confidence curve shows a rapid increase in precision as the confidence threshold increases, reaching a maximum value of 1.00 at a confidence level of 0.829. This indicates that high-confidence predictions correspond to highly reliable detections with minimal FP rates. In contrast, the Recall-Confidence curve shows a decreasing trend as the confidence threshold increases. The model achieves a recall of approximately 0.90 at lower confidence levels; however, recall decreases as the threshold becomes more restrictive. This behavior reflects the inherent trade-off between precision and recall in object detection, where increasing the confidence threshold improves prediction reliability while reducing the number of detected TP.

These results demonstrate the importance of selecting an appropriate confidence threshold to balance detection reliability and coverage. Such a balance is critical for real-time pedestrian detection in V2X environments, where both accurate recognition and comprehensive detection are required to support safe and effective decision-making.

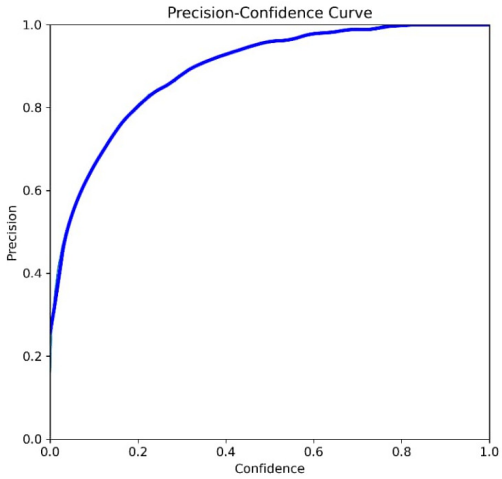

Fig. 6. The Precision-Confidence curve of the pedestrian detection model.

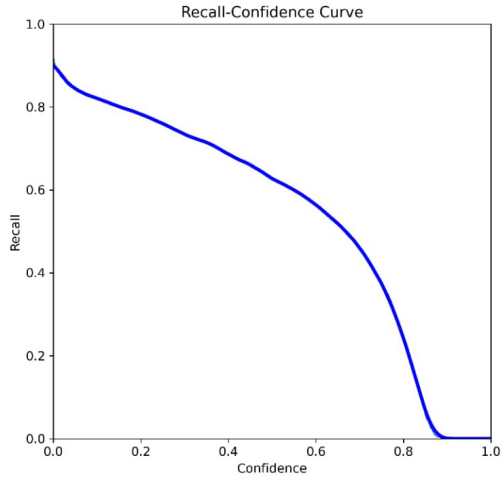

Fig. 7. The Recall-Confidence curve of the pedestrian detection model.

VI. CONCLUSION

This study presented a YOLOv8-based pedestrian detection framework for V2X environments, aimed at improving road safety and traffic awareness in urban transportation systems. The proposed framework leverages real-time detection capability, multi-scale feature extraction, and optimized detection mechanisms to achieve accurate pedestrian recognition under challenging conditions, including varying illumination, occlusions, and complex urban scenes. Experimental evaluation on the CityPersons dataset demonstrates that the proposed model achieves an mAP@0.5 of 0.838, indicating a balanced trade-off between precision and recall. Comparative analysis with established object detection models, including Faster R-CNN, SSD, YOLOv3, and YOLOv5, shows that the proposed approach provides improved detection performance while maintaining robustness across diverse scenarios. In addition, Precision-Confidence and Recall-Confidence analyses confirm that the model produces reliable predictions at higher confidence thresholds while preserving substantial recall.

These results demonstrate the effectiveness of the proposed framework for real-time pedestrian detection in V2X-enabled intelligent transportation systems, supporting accurate and timely perception for autonomous vehicles. By combining high detection accuracy with computational efficiency, the framework contributes to enhancing traffic safety and reducing the risk of pedestrian-related incidents.

Future work will focus on integrating pedestrian behavior prediction models, adaptive confidence thresholding under varying traffic densities, and deployment in real-world V2X environments to further improve predictive safety and system robustness.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

The authors would like to acknowledge Universiti Teknologi Malaysia (UTM) for providing the research environment and support. No external funding was received for this research.

DATA AVAILABILITY

The CityPersons dataset used in this study can be accessed from the official Cityscapes repository [21].

GENERATIVE AI USE DECLARATION

During the preparation of this work, the authors used ChatGPT, Jenni AI, and SCISPACE to assist in improving the clarity and language of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] S. Pang, J. Xue, Q. Tian, and N. Zheng, "Exploiting local linear geometric structure for identifying correct matches," *Computer Vision and Image Understanding*, vol. 128, pp. 51–64, Nov. 2014, https://doi.org/10.1016/j.cviu.2014.06.006.

[2] X. Cao, S. Guo, J. Lin, W. Zhang, and M. Liao, "Online tracking of ants based on deep association metrics: method, dataset and evaluation," *Pattern Recognition*, vol. 103, July 2020, Art. no. 107233, https://doi.org/10.1016/j.patcog.2020.107233.

[3] Y. Zhang, Y. Jin, J. Chen, S. Kan, Y. Cen, and Q. Cao, "PGAN: Part-Based Nondirect Coupling Embedded GAN for Person Reidentification," *IEEE MultiMedia*, vol. 27, no. 3, pp. 23–33, July 2020, https://doi.org/10.1109/MMUL.2020.2999445.

[4] C. Han *et al.*, "Re-ID Driven Localization Refinement for Person Search," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 9813–9822, https://doi.org/10.1109/ICCV.2019.00991.

[5] S. Anbalagan, P. Srividya, B. Thilaksurya, S. G. Senthivel, G. Suganeshwari, and G. Raja, "Vision-Based Ingenious Lane Departure Warning System for Autonomous Vehicles," *Sustainability*, vol. 15, no. 4, Feb. 2023, https://doi.org/10.3390/su15043535.

[6] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, 2018, https://doi.org/10.48550/ARXIV.1804.02767.

[7] D. R. Kumar and A. Rammohan, "Revolutionizing Intelligent Transportation Systems with Cellular Vehicle-to-Everything (C-V2X) technology: Current trends, use cases, emerging technologies, standardization bodies, industry analytics and future directions,"

Vehicular Communications, vol. 43, Oct. 2023, Art. no. 100638, https://doi.org/10.1016/j.vehcom.2023.100638.

[8] R. Zhang, D. Meng, S. Shen, Z. Zou, H. Li, and H. X. Liu, "MSight: An Edge-Cloud Infrastructure-based Perception System for Connected Automated Vehicles." arXiv, 2023, https://doi.org/10.48550/ARXIV.2310.05290.

[9] L. Huang, W. Huang, H. Gong, C. Yu, and Z. You, "PEFNet: Position Enhancement Faster Network for Object Detection in Roadside Perception System," *IEEE Access*, vol. 11, pp. 73007–73023, 2023, https://doi.org/10.1109/ACCESS.2023.3292881.

[10] B. Ghari, A. Tourani, A. Shahbahrami, and G. Gaydadjiev, "Pedestrian detection in low-light conditions: A comprehensive survey," *Image and Vision Computing*, vol. 148, Aug. 2024, Art. no. 105106, https://doi.org/10.1016/j.imavis.2024.105106.

[11] W. Farhat, O. B. Rhaiem, H. Faiedh, and C. Souani, "Optimized deep learning for pedestrian safety in autonomous vehicles," *International Journal of Transportation Science and Technology*, Apr. 2025, Art. no. S204604302500053X, https://doi.org/10.1016/j.ijtst.2025.04.002.

[12] K. L. Masita, A. N. Hasan, and S. Paul, "Pedestrian Detection Using R-CNN Object Detector," in *2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, Nov. 2018, pp. 1–6, https://doi.org/10.1109/LA-CCI.2018.8625210.

[13] G. Shariha, M. Elmogy, E. El-Daydamony, and A. Atwan, "Multiple Pedestrian Detection Depending on Faster Region-based Convolutional Neural Network (RCNN)," *Mansoura Journal for Computer and Information Sciences*, vol. 15, no. 1, pp. 13–20, June 2019, https://doi.org/10.21608/mjcis.2019.320866.

[14] M. Saeidi and A. Arabsorkhi, "A novel backbone architecture for pedestrian detection based on the human visual system," *The Visual Computer*, vol. 38, no. 6, pp. 2223–2237, June 2022, https://doi.org/10.1007/s00371-021-02280-6.

[15] F. Sultana, A. Sufian, and P. Dutta, "A Review of Object Detection Models Based on Convolutional Neural Network," in *Intelligent Computing: Image Processing Based Applications*, vol. 1157, J. K. Mandal and S. Banerjee, Eds. Springer Singapore, 2020, pp. 1–16.

[16] V. Teju, K. V. Sowmya, S. R. Kandula, A. Stan, and O. P. Stan, "A Hybrid Retina Net Classifier for Thermal Imaging," *Applied Sciences*, vol. 13, no. 14, July 2023, Art. no. 8525, https://doi.org/10.3390/app13148525.

[17] K. Wang and W. Zhou, "Pedestrian and cyclist detection based on deep neural network fast R-CNN," *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, Mar. 2019, Art. no. 1729881419829651, https://doi.org/10.1177/1729881419829651.

[18] Y. Xue, Z. Ju, Y. Li, and W. Zhang, "MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detection," *Infrared Physics & Technology*, vol. 118, Nov. 2021, Art. no. 103906, https://doi.org/10.1016/j.infrared.2021.103906.

[19] S. Zhang, R. Benenson, and B. Schiele, "CityPersons: A Diverse Dataset for Pedestrian Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 4457–4465, https://doi.org/10.1109/CVPR.2017.474.

[20] B. Paranjape, A. Naik, and S. P. Sankar, "A Performance Comparison of Object Detection Algorithms on Traffic Scenes in Indian Roads," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25492–25498, Aug. 2025, https://doi.org/10.48084/etasr.11105.

[21] "Cityscapes Dataset – Semantic Understanding of Urban Street Scenes," Oct. 17, 2020. https://www.cityscapes-dataset.com/.