

Real-Time Pedestrian Detection and Passive Signal Monitoring Using YOLOv12

Mohd Nur Aiman Sulaiman¹, Nik Shahidah Afifi Md Taujuddin^{1*}, Suhaila Sari¹, and Ahmad Raqib Ab Ghani²

¹Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Johor, Malaysia.

²Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Johor, Malaysia.

KEYWORDS

*Pedestrian Crossing
Image Processing
Image Detection and
Counting*

ARTICLE HISTORY

*Received 23 July 2026
Accepted 6 October 2026
Available online 20 October
2026*

ABSTRACT

Pedestrian crossings are vital for urban safety and traffic management. However, fixed-time signal controls ignore real-time pedestrian demand, leading to inefficient traffic flow and underutilized infrastructure. To address this, this project introduces a Real-Time Pedestrian Detection and Passive Signal Monitoring Using YOLOv12. The core objectives were to develop a pedestrian detection and counting system, implement analytics functionalities, and evaluate overall performance using recorded footage. The proposed system was validated using real-world video and publicly available datasets. Experimental results demonstrated that the system successfully performs real-time pedestrian counting, demand classification, and adaptive traffic light control. Notably, the adaptive control strategy optimized traffic flow by dynamically reducing the pedestrian green-light duration from 8 seconds to 4 seconds when no active crossing demand was detected. This intelligent adjustment resulted in a cumulative time savings of approximately 2.15 minutes across 26 crossing cycles. The successful development of this system highlights the immense capability of integrating computer vision and deep learning for pedestrian monitoring and analytics applications.

© 2026 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Safe and efficient citizen mobility forms the foundation of a functional urban society. As global urbanization accelerates, managing pedestrian flow in high-density areas has emerged as a critical societal challenge. Globally iconic intersections, such as Tokyo's Shibuya Scramble Crossing—which accommodates over 2,500 pedestrians per green light [1]—demonstrate the immense scale of human mobility. Rather than serving merely as transportation hubs, these locations function as economic and cultural centers reliant on seamless pedestrian movement.

This challenge is equally pronounced in Malaysia, particularly within Kuala Lumpur's vibrant Bukit Bintang district. As the nation's primary hub for retail, tourism, and entertainment, Bukit Bintang experiences massive daily foot traffic. Recent government initiatives aimed at revitalizing the area [2] project a significant future increase in pedestrian volume, placing unprecedented strain on existing infrastructure.

Supporting this growth requires pedestrian crossings to evolve from conventional, fixed-time systems into intelligent, adaptive solutions. This transition represents far more than a technological upgrade; it is a critical urban intervention designed to enhance public safety, mitigate congestion, and cultivate a superior urban environment for all citizens.

Pedestrian safety remains a critical global issue, evidenced by thousands of annual fatalities worldwide, including 7,388 traffic-related pedestrian deaths in the United States in 2021 [3] and hundreds of annual casualties in Malaysia. Despite the widespread study of adaptive traffic signals, many pedestrian crossings still rely on fixed-duration timings that fail to account for actual pedestrian presence. These rigid systems cannot adapt to dynamic real-world conditions, such as extending crossing times for slow-moving elderly individuals or detecting pedestrians remaining on the crosswalk. Furthermore, they fail to adjust to heightened risk profiles, such as nighttime

*Corresponding author:

E-mail address: Nik Shahidah Afifi Md Taujuddin <shahidah@uthm.edu.my>.
<https://doi.org/10.56532/mjsat.v6iS2.808>

2785-8901/ © 2026 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

conditions where low visibility and altered pedestrian behaviors introduce distinct hazards. Consequently, fixed-timer systems often allocate crossing time unnecessarily after pedestrians have already crossed, resulting in avoidable vehicular delays and traffic inefficiencies. To mitigate these challenges, there is a clear need for a real-time pedestrian monitoring system that can dynamically adjust signal timing based on actual crossing utilization.

Many existing smart pedestrian crossing systems rely heavily on single-sensor technologies, such as infrared or motion sensors, to detect pedestrian presence. While these sensors are inexpensive and simple to deploy, they cannot provide detailed insights regarding pedestrian behavior, movement patterns, or crowd density. Furthermore, their detection reliability can be severely compromised by environmental factors, including adverse weather, fluctuating lighting, and physical obstructions [4].

Another critical limitation is the industry's primary focus on basic detection rather than comprehensive pedestrian analytics. Most current systems merely determine whether a pedestrian is present within the crossing zone to trigger warning signals or traffic control mechanisms. However, they lack advanced analytical capabilities, such as pedestrian counting, occupancy duration tracking, demand analysis, peak usage identification, and crossing behaviour trends. Such high-level data is vital for traffic authorities to evaluate infrastructure performance and plan future urban improvements [5].

Additionally, many smart crossing systems operate on rigid, predefined control rules rather than adaptive decision-making frameworks. Even when a pedestrian triggers the signal, the allocated crossing duration usually remains fixed regardless of actual real-time demand. Prior research has demonstrated that adaptive traffic control strategies significantly enhance traffic efficiency by dynamically responding to fluctuating conditions, thereby exposing the shortcomings of conventional fixed-duration methods [5].

Finally, while recent vision-based pedestrian monitoring systems offer superior detection and can extract trajectories, behaviors, and safety metrics from video streams, gaps still remain. Many of these implementations are restricted to isolated safety monitoring or academic research; they lack integrated analytics dashboards that provide real-time decision support and long-term performance evaluation for municipal authorities [6].

As the core methodology of this research, object detection and recognition are advanced computer vision tasks that integrate image classification and localisation. Unlike simple classification models that merely identify the overall content of an image, object detection models locate specific items of interest within the frame [7] (see Figure 1). The model defines the spatial coordinates of each target object by outputting a precise bounding box paired with a corresponding class label. Consequently, the model's output consists of a structured list of identified items, each with its exact spatial coordinates. In the context of this study, this capability is vital; it enables the system to not only detect the presence of pedestrians but also accurately count them and analyze their specific positions relative to the crosswalk.



Fig. 1. The result of the Object Detection System [7]

The "You Only Look Once" (YOLO) architecture revolutionised real-time object detection by framing the task as a single regression problem. Instead of using complex multi-stage pipelines, YOLO directly predicts bounding boxes and class probabilities from an entire image in a single evaluation pass. This design yields incredible processing speeds, making it ideal for live traffic management systems that require split-second decision-making. Since its inception, the YOLO series has evolved significantly, with each new iteration introducing key improvements in detection efficiency and performance [8-14]. Figure 2 shows an example of human detection using YOLOv5.

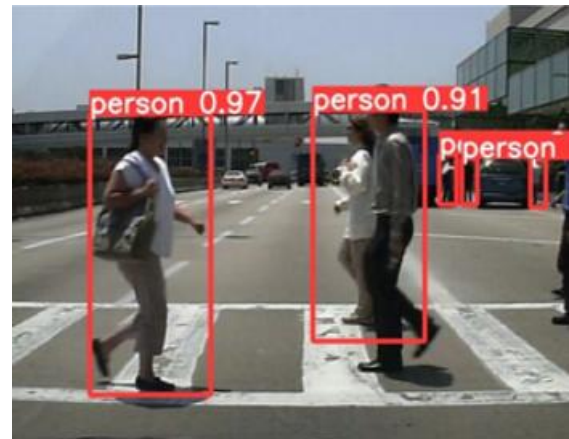


Fig. 2. YOLOv5 Training Detection Result [9]

2. METHODOLOGY

This project follows a systematic three-phase development process. First, the Data Preparation phase involves aggregating raw data from various sources. Next, the Model Development phase utilises Google Colab and the YOLOv12 framework to train the object detection model on the prepared dataset. Concluding the workflow, the Analysis phase evaluates the model's performance through accuracy graphs to ensure it aligns with the project's target benchmarks.

2.1 Data Preparation

Data collection was conducted via digital sources, with a strict focus on high-quality datasets that accurately simulate the target deployment environment. Rather than relying solely on manual on-site recordings, a total of 3,576 images were

aggregated from two primary digital sources: public computer vision repositories and pre-recorded traffic footage.

The raw images were uploaded to Roboflow for annotation (see Figure 3). To ensure accuracy, bounding boxes were generated using Roboflow's auto-labeling tool and then manually reviewed and refined.

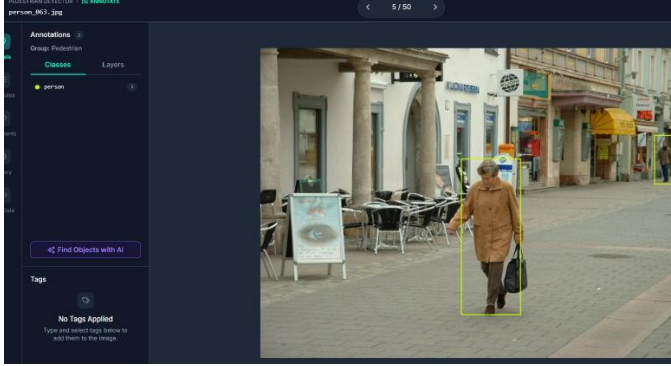


Fig. 3. Annotation Process in Roboflow

2.2 Augmentation

Augmentations were implemented in Roboflow to improve model generalisation across diverse environments. All images were resized to a standard resolution of 614 x 614 pixels to align with the YOLOv12 architecture's input specifications. The training set then underwent several augmentations, including horizontal flipping and brightness to mimic fluctuating weather and lighting conditions. Additionally, mosaic augmentation was used to improve the system's performance in detecting small objects in complex urban scenes.

2.3 YOLOv12 Model Architecture

The performance of a real-time pedestrian detection system depends heavily on the quality of its training data combined with the efficiency of its neural network architecture. The YOLOv12 architecture optimises information flow and captures global contextual details within an image, enabling the model to effectively isolate target objects from complex backgrounds. YOLOv12 is an attention-centric model with a Transformer-CNN hybrid framework designed to maximise accuracy without compromising real-time inference speed. It's also benchmarked on the COCO dataset at 640 x 640 resolution. Figure 4 shows the YOLO v12 architecture.

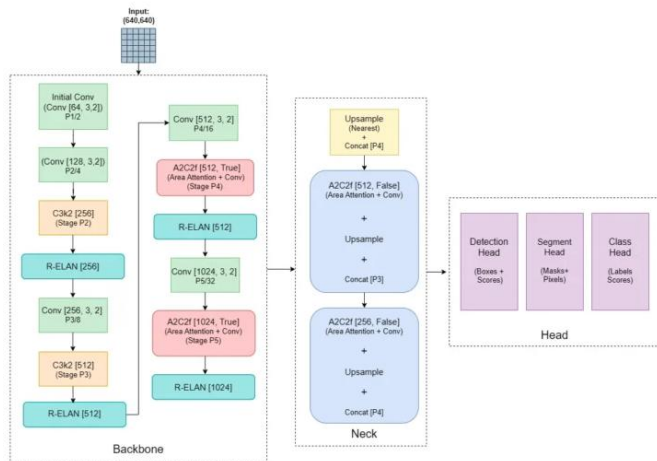


Fig. 4. YOLO v12 architecture [15]

This design is well-suited to challenging environments characterised by dense crowds and frequent occlusions, offering more precise localisation than its purely convolutional predecessors.

Training was conducted on a standard Google Colab instance equipped with an NVIDIA Tesla T4 GPU. To accelerate convergence and maximise detection accuracy, transfer learning was employed by initialising the network with pre-trained weights from the COCO dataset. This strategy allowed the model to leverage satisfactory, pre-learned feature extractors rather than training from scratch.

The training framework was defined by the following hyperparameter configurations:

- i. Epochs: Set to 50, a duration selected to provide sufficient learning time while mitigating overfitting risks, given the rapid convergence facilitated by transfer learning.
- ii. Image Resolution: Fixed at 640 x 640 pixels to match the native input specifications of the YOLOv12 architecture, thereby preserving granular pedestrian features without excessive down-sampling.
- iii. Batch Size: Configured to 16 to balance GPU memory utilisation with gradient estimation stability.

2.4 Performance Evaluation Metrics

To quantitatively measure the model's ability to detect and classify pedestrians, the proposed YOLOv12 framework was evaluated using key performance metrics: Precision, Recall, (mAP@50), and a Normalised Confusion Matrix. Precision as in (1) quantifies the ratio of correctly identified pedestrians to all positive detections made by the model. A high-precision value indicates that the system minimises false positives, ensuring that detected objects are highly likely to be actual pedestrians.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Where:

TP = True Positive

FP = False Positive

Recall as in (2) measures the ratio of successfully detected pedestrians to the total number of actual pedestrians present. A high recall value signifies that the model minimizes false negatives, ensuring that very few pedestrians are missed during real-time monitoring.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Where:

TP = True Positive

FN = False Negative

Meanwhile, Mean Average Precision (mAP@50) assesses overall detection performance by summarising the precision-recall curve at an Intersection over Union (IoU) threshold of 0.50, where higher values signify superior localisation and accuracy. Complementing this, a normalized confusion matrix visually compares actual versus predicted labels, highlighting

classification errors and verifying the model's ability to isolate pedestrians from background noise. Finally, the model's learning progress and convergence behaviour are monitored via Box Loss, Classification Loss, and Distribution Focal Loss (DFL). Progressively decreasing values across these loss functions indicate successful optimisation and heightened detection accuracy.

2.5 System Application Development

The proposed system utilises a dedicated Region of Interest (ROI) Setup module, allowing users to interactively define polygon-based monitoring zones directly on the video frame. During this initialisation process, users manually select a series of coordinate points corresponding to designated waiting zones and crossing areas. These spatial coordinates are subsequently stored in JavaScript Object Notation (JSON) format for persistent retrieval.

Because accurate counting is a critical factor in the system's decision-making, it assigns a unique, persistent ID to each detected pedestrian. The application logic maintains active tracking trajectories and increments the crowd counter only when a new, unique ID enters the predefined ROI. This satisfactory tracking mechanism prevents double-counting caused by intermittent occlusion or frame-to-frame jitter, thereby providing a stable estimate of the waiting crowd size.

After successful detection, tracking, and counting, the analytics module processes the raw data to extract meaningful insights into pedestrian behaviour. By leveraging metrics from the tracking stage—such as pedestrian counts, waiting, crossing volumes, occupancy durations, and spatial movement paths—the system continuously evaluates density within the waiting zones. Based on these real-time counts, the system classifies pedestrian demand into three discrete categories: Low, Medium, and High. This demand classification framework simplifies raw crowd metrics into actionable insights to support adaptive traffic control strategies. Figure 4 shows the developed Pedestrian Analytics Dashboard.



Fig. 5. Pedestrian Analytics Dashboard

3. RESULT AND DISCUSSION

3.1 Training Performance Analysis

The YOLOv12 model was trained for 50 epochs on the custom dataset, with the stability of the optimisation process reflected in the loss trajectories and validation metrics. As illustrated in Figure 6, the model exhibits a consistent learning curve throughout training. The Box Loss decreases steadily to approximately 1.08, indicating that the framework achieved precise localisation for pedestrian bounding boxes. Furthermore, the Mean Average Precision (mAP@50), which quantifies overall detection performance, reached 80.5%.

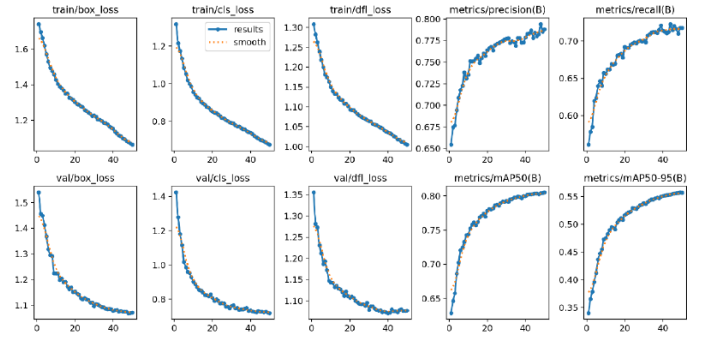


Fig. 6. Result of 50 epochs using YOLOv12

To evaluate classification accuracy, a normalised confusion matrix was analysed. As shown in Figure 7, the model yields a True Positive rate of 0.80 (80%) for the "person" class, successfully identifying 80% of all pedestrians in the validation dataset. Crucially, the model recorded a 0% false detection rate against the background, demonstrating high reliability and resilience against false triggers—a vital requirement for real-world deployment in adaptive traffic systems.

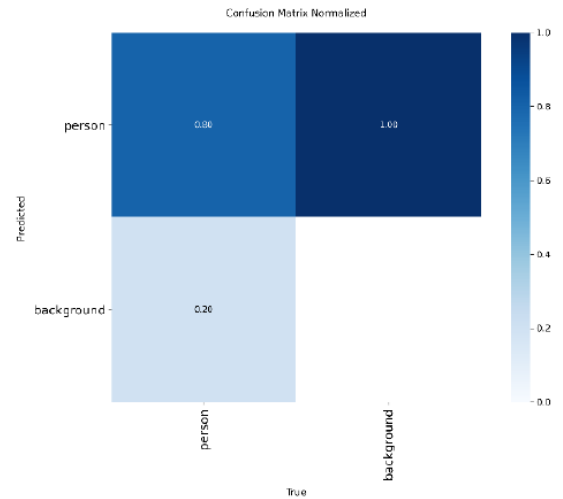


Fig. 7. Normalized Confusion Matrix using YOLOv12

3.2 Pedestrian Detection Results Using Pexels Video Dataset

The performance of the pedestrian detection was evaluated using Detection Success Rate (DSR) as define in Equation (3).

$$DSR = \frac{\text{Detected Pedestrian}}{\text{Actual Pedestrian}} \times 100\% \quad (3)$$

The selected Pexels videos feature pedestrians navigating urban environments under relatively stable lighting conditions and moderate crowd density. Figure 8 illustrates a representative example of the pedestrian detection output generated by the proposed system, in which the YOLOv12 model successfully identifies individuals in the video frames and bounds them with precise bounding boxes.



(a)



(b)

Fig. 8. Pedestrian detection result using pexels video on frame (a) 192, (b) 252

To assess the practical efficacy of the developed system, the automated detection outputs were benchmarked against manual human observations. As detailed in Table 1, the proposed framework maintained a high success rate across the evaluated video sequences. Specifically, the system achieved an average detection success rate of 82.65% on the Pexels dataset, validating the capability of the YOLOv12 model to accurately localize pedestrians under controlled environmental settings. This satisfactory performance is primarily attributed to optimal scene characteristics—namely, stable illumination, moderate crowd density, and minimal physical occlusion—which enabled consistent feature extraction and highly dependable detection results.

Table 1. Detection Performance Using Pexels Video

Frame	Human Observation	YOLO Detection	Detection accuracy (%)
36	8	6	75.00
120	9	7	77.78
192	4	4	100.00
228	4	3	75.00
252	9	7	77.78
276	4	3	75.00
324	3	3	100.00
360	5	4	80.00
420	6	5	83.33
Average detection accuracy			82.65

The second experiment was conducted using empirical video footage recorded at the pedestrian crossing on Jalan Rahmat, Batu Pahat, Johor, to evaluate the proposed monitoring system under actual deployment conditions. Compared to the controlled Pexels benchmark, this real-world footage presents a significantly more strenuous testing environment characterized by fluctuating lighting intensities, frequent pedestrian occlusions, dynamic backgrounds, and highly irregular movement patterns.

Figure 9 illustrates a representative detection result from the Jalan Rahmat dataset. The YOLOv12 model successfully localized pedestrians within the target zone, encapsulating identified individuals within precise bounding boxes. Concurrently, the object tracking module assigned unique, persistent tracking identifiers to each pedestrian, facilitating continuous trajectory monitoring across successive frames.



(a)



(b)

Fig. 9. Pedestrian detection result using recorded footage at Jalan Rahmat, Batu Pahat, Johor video on frame (a) 6635, (b) 8676

To quantify system efficacy, automated outputs were benchmarked against manual human tallies, as summarised in Table 2. Despite the environmental complexities inherent to the real-time field footage, the system maintained satisfactory detection accuracy (64.67%). Marginal tracking anomalies were primarily observed when pedestrians were partially occluded, positioned at a significant distance from the camera, or navigating dense crowds.

Table 2. Detection Performance Using Recorded Footage

<i>Frame</i>	<i>Human Observation</i>	<i>YOLO Detection</i>	<i>Detection accuracy (%)</i>
45	2	2	100.00
61	3	3	100.00
106	3	2	66.67
115	3	3	100.00
151	3	0	0
199	1	0	0
218	3	1	33.33
287	5	4	80.00
289	5	5	100.00
<i>Average detection accuracy</i>			64.67

Ultimately, these findings validate the system's operational viability under real-world conditions, demonstrating that it provides a sufficiently satisfactory foundation for downstream applications such as pedestrian counting, occupancy monitoring, and analytics generation.

3.3 Pedestrian Analytics Result

A primary objective of this research is to evaluate crosswalk utilisation and quantify the potential time savings achievable through adaptive signal management. Unlike conventional infrastructure that relies on rigid, fixed signal cycles, the proposed framework analyses empirical pedestrian crossing behaviour to generate operational recommendations based on real-time activity.

Fig 5 illustrates the occupancy profiles and time-saving analytics generated by the system dashboard. Using field footage from Jalan Rahmat, the system captured 26 discrete crossing cycles during the observation window. In this dataset, the mean pedestrian crossing duration was 3.36 seconds, with a maximum of 13.02 seconds.

These empirical findings reveal that pedestrians generally require minimal time to clear the crosswalk. In numerous instances, the actual crossing duration was substantially shorter than the static block of time typically allocated by traditional traffic controllers, leaving a significant portion of the signal cycle underutilised. Based on this continuous occupancy monitoring, the system projected a cumulative time savings of approximately 2.15 minutes across the analysed observation interval. While the current iteration of the system operates as a passive analytics platform rather than an active traffic light actuator, it successfully generates data-driven optimisation recommendations based on live occupancy trends.

Ultimately, these analytics demonstrate the system's capacity to extract high-fidelity insights regarding infrastructure utilisation. This predictive data provides a viable foundation for future deployments in adaptive traffic signal optimisation, intelligent transportation networks, and broader smart-city applications.

4. CONCLUSION

This project successfully developed and evaluated an intelligent, vision-based pedestrian monitoring and adaptive signal control framework powered by YOLOv12. By combining advanced object localization, persistent tracking pipelines, and digital ROI mapping, the proposed system effectively overcomes the rigid inefficiencies inherent to traditional fixed-time traffic signals.

Experimental results across controlled benchmarks and real-world urban field tests validated the system's capacity to accurately classify pedestrian demand profiles in real time. The integration of an adaptive controller dynamically optimized green-light intervals according to live crosswalk utilization, yielding substantial, measurable cumulative time savings. Complemented by a centralized analytics dashboard that converts raw coordinate streams into high-fidelity visual insights, this project fully achieved its target objectives. It establishes a satisfactory, data-driven methodology that provides a viable blueprint for future innovations in intelligent transportation networks and autonomous smart-city infrastructure.

5. FUTURE WORK

Although the proposed framework met its target objectives, several strategic enhancements could expand its operational capabilities. First, future developments should focus on integrating the system directly with actual roadside traffic light hardware. This transition from a passive monitoring tool to an active actuator will enable full-scale field testing to validate its capacity to reduce signal underutilization without compromising pedestrian safety. Second, while the current adaptive logic relies on predefined thresholds and duration rules, integrating intelligent decision-making paradigms—such as reinforcement learning—could allow the controller to autonomously optimise signal timing based on real-time behaviour.

REFERENCES

- [1] A. Sakurai, N. Kajio, and K. Yamamoto, "ShibuyaSocial: Multi-scale model of pedestrian flows in scramble crossing," arXiv, art. no. 2512.18550, Dec. 2025, doi: <https://doi.org/10.48550/arXiv.2512.18550>.
- [2] "Bukit Bintang set for glow-up to rival Tokyo's Ginza and Singapore's Orchard Road, says housing minister." Accessed: Dec. 11, 2025. url: <https://www.msn.com/en-my/news/national/bukit-bintang-set-for-glow-up-to-rival-tokyo-s-ginza-and-singapore-s-orchard-road-says-housing-minister/ar-AA1PoMRk>.
- [3] M. Patel and H. Elgazzar, "Pedestrian Crosswalk Safety at Intersections using YOLOv8 Detection," 2024 IEEE 15th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 426–430. doi: <https://doi.org/10.1109/UEMCON62879.2024.10754669>.
- [4] Samuel Omeffe, "Assessment of the effectiveness of adaptive traffic signal control systems in reducing pedestrian-vehicle conflicts at high-risk crossings," GSC Advanced Research and Reviews, vol. 24, no. 1, pp. 140–153, Jul. 2025, doi: <https://doi.org/10.30574/gscarr.2025.24.1.0191>.
- [5] M. Pomoni, "Smart Crosswalks for Advancing Road Safety in Urban Roads: Conceptualization and Evidence-Based Insights from Greek Incident Records," Future Transportation, vol. 5, no. 4, Dec. 2025, doi: <https://doi.org/10.3390/futuretransp5040180>.
- [6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey,"

- IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 7, pp. 3523–3542, Jul. 2022, doi: <https://doi.org/10.1109/TPAMI.2021.3059968>.
- [7] C. Ashwini, S. Sharma, A. Srivastava, and S. Sinha, "To identify and recognize the object for traffic analysis system using deep learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 1621–1624, Oct. 2019, doi: <https://doi.org/10.35940/ijitee.L3155.1081219>.
- [8] C. Li and Y. Sun, "Enhancing Pedestrian Detection Algorithm Based on YOLOv3-ES Network," *Proceedings - 2023 International Conference on Image Processing and Computer Vision, IPCV 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 118–122. doi: <https://doi.org/10.1109/IPCv57033.2023.00029>.
- [9] X. Wang, R. Wei, and R. Zou, "Research on Pedestrian Detection Algorithm Based on Improved YOLOv5," *2023 3rd International Conference on Electronic Information Engineering and Computer Science, EIECS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1109–1112. doi: <https://doi.org/10.1109/EIECS59936.2023.10435608>.
- [10] Y. Yue, X. Kong, R. Pan, and W. Shan, "Research on Intelligent Traffic Pedestrian Detection Algorithm Based on Improved YOLOv8," *2024 5th International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 384–388. doi: <https://doi.org/10.1109/ICAICE63571.2024.10864271>.
- [11] X. Li and L. Ma, "DED-YOLOv8:Dense pedestrian detection algorithm based on YOLOv8," *2024 4th International Symposium on Computer Technology and Information Science, ISCTIS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 545–548. doi: <https://doi.org/10.1109/ISCTIS63324.2024.10699141>.
- [12] J. Mao, H. Wang, and D. Li, "Optimized and Improved YOLOv8 Dense Pedestrian Detection Algorithm," *Proceedings - 2024 9th International Conference on Communication, Image and Signal Processing, CCISP 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: <https://doi.org/10.1109/CCISP63826.2024.10765506>.
- [13] C. Hu, Y. Wei, and X. Tao, "Lightweight YOLOv8 Pedestrian Detection Model Based on FasterNet," *2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing, AIIM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 787–790. doi: <https://doi.org/10.1109/AIIM64537.2024.10934361>.
- [14] Li, W. Leong, and H. Zhang, "YOLOv10-Based Real-Time Pedestrian Detection for Autonomous Vehicles," *2024 IEEE 8th International Conference on Signal and Image Processing Applications, ICSIPA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: <https://doi.org/10.1109/ICSIPA62061.2024.10686546>.
- [15] M. Hasan, M. K. M. Rabby, I. Jahan, M. J. A. Soeb, and M. F. Jubayer, "The Evolution and Advancement of YOLO Algorithms in Object Detection: From Real-Time Breakthroughs to Modern Architectures," *Preprints.org*, Oct. 2025, doi: <https://doi.org/10.20944/preprints202510.2019.v1>.