

# PEDESTRIAN FLOW COUNTING AT URBAN INTERSECTIONS USING A RETRAINED YOLOV8 MODEL

Tam Vu<sup>a,\*</sup>

<sup>a</sup>*Faculty of Transportation Engineering, Hanoi University of Civil Engineering,  
55 Giai Phong road, Hai Ba Trung district, Hanoi, Vietnam*

## Article history:

*Received 28/5/2025, Revised 17/6/2025, Accepted 20/6/2025*

---

## Abstract

This study addresses the limited application of computer vision techniques for counting pedestrians at urban intersections by developing an integrated deep learning-based framework. Three models are proposed: a detection model trained on a large, diverse pedestrian dataset; a tracking model incorporating a novel reference point on bounding boxes with an enhanced identity-switch handling algorithm; and a counting model tailored to pedestrian crossing behaviors with movement-specific algorithms. The framework was applied to a case study in Vietnam, where pedestrian flow is often complex due to mixed traffic environments. A comprehensive dataset comprising 22 video footages under both daytime and nighttime conditions yielded over 120,000 labeled pedestrian instances. The tracking model effectively captures pedestrian trajectories across crosswalks, while the counting model introduces a multi-line crossing technique to enhance accuracy at signalized intersections. Evaluation results show the counting model achieves over 98% accuracy compared to manual annotations across various time frames and pedestrian densities. These findings offer valuable tools for urban transport planners and policymakers in Vietnam and similar countries, enabling automated pedestrian data collection, improving intersection safety assessment, and supporting infrastructure design.

**Keywords:** computer vision; intersections; pedestrian flow counting.

[https://doi.org/10.31814/stce.huce2025-19\(2\)-08](https://doi.org/10.31814/stce.huce2025-19(2)-08) © 2025 Hanoi University of Civil Engineering (HUCE)

---

## 1. Introduction

Urban traffic congestion has become a serious and widespread problem across many countries in recent decades, causing a variety of negative impacts such as air and noise pollution, deteriorating public health, and a general reduction in urban quality of life. In many Asian countries this issue is even more severe due to high population density and overburdened road infrastructure [1]. The rapid urbanization and motorization in Asian countries, particularly in megacities, have led to severe traffic congestion, air pollution, and unsustainable energy consumption [2]. Although vehicle congestion has received significant research and policy attention, pedestrian movement - particularly at intersections - remains under-researched, despite being a vital component of urban mobility and safety.

Despite the extensive focus on vehicular traffic, however, intersections also play a critical role in pedestrian mobility and safety. Intersections are key conflict points within any traffic network, where movements from multiple directions converge. Efficient design and control of intersections are essential not only for vehicle throughput but also for pedestrian safety and mobility. Accurate data on pedestrian volume and movement patterns at intersections serves as a foundational input for planning crosswalks, signal timing, and safety measures.

Traditional methods for counting pedestrians, such as manual observation, pressure-sensitive mats, and infrared sensors, face numerous challenges in dense urban environments. Manual counting

---

\*Corresponding author. E-mail address: [tamvm@huce.edu.vn](mailto:tamvm@huce.edu.vn) (Vu, T.)

techniques using sheets or clickers tend to underestimate pedestrian volumes by 8-25%, with higher error rates at the beginning and end of observation periods [3]. Infrared-based counters systematically undercount actual pedestrian traffic, with performance varying across different sites [4]. These methods often struggle with occlusion, varying pedestrian behavior, and the lack of clear walking lanes. Additionally, pedestrians may not always follow designated paths, especially in developing countries where informal crossings are common. Such complexity makes it difficult for both human observers and automated systems to reliably detect and count pedestrians, particularly during peak hours or in crowded, mixed-use junctions.

To address these challenges, this study proposes an integrated framework consisting of a detection model, a tracking model, and specialized counting algorithms designed for complex pedestrian movements at intersections. The framework is applied at intersections in Vietnam and Singapore, where high pedestrian activity often intersects with chaotic mixed traffic, posing significant risks and measurement difficulties. The remainder of this paper is structured as follows: Section 2 reviews existing literature on vision-based pedestrian detection and counting; Section 3 outlines the proposed methodology; Section 4 presents the case studies; and Section 5 discusses the results, conclusions, and avenues for future research.

## 2. Literature Review

Traditional methods for pedestrian counting, such as manual observation, infrared sensors, and inductive loops, are often limited in complex and crowded urban environments. To overcome these constraints, a variety of automated approaches using sensor technologies and advanced modeling techniques have been explored.

The performance of a dual-sensor passive infrared device was evaluated and found effective under moderate conditions but less reliable in high-density environments [5]. A 2D LiDAR-based system was developed for crowded scenarios, achieving over 97% accuracy at the disaggregate level [6]. A statistical method using feature-based regression in the spatiotemporal domain was proposed, attaining 97.2% accuracy without requiring individual detection and tracking - an advantage in occluded scenes [7]. While these technologies demonstrate strong potential, common limitations remain, including high hardware cost, reliance on fixed sensor placement, sensitivity to occlusion, and challenges in scaling to complex, mixed-use intersections.

In recent years, deep learning models based on convolutional neural networks have been increasingly adopted for pedestrian detection and counting tasks. The YOLO (You Only Look Once) family of object detectors has shown promising results. The effectiveness of pre-trained YOLOv8 models for pedestrian detection at intersections using existing surveillance infrastructure has been demonstrated [8]. A method for detecting directional pedestrian flows on crosswalks using both synthetic and real-world datasets was introduced to address the spatial complexity of intersection environments [9].

However, most existing studies focus on generalized or Western urban contexts, utilize limited datasets, or concentrate on a single pedestrian class. Few works consider the unique, lane-free, and mixed-traffic conditions in Southeast Asian countries, particularly Vietnam, where vulnerable road users such as cyclists, wheelchair users, and scooter riders frequently share intersection spaces. Moreover, few prior studies provide a full pipeline encompassing detection, tracking, and counting - customized for the complexity of real-world intersections.

To address these gaps, a comprehensive pedestrian counting framework is proposed in this study, especially tailored to Vietnamese intersection conditions. This framework consists of (i) a large-scale, labeled dataset representative of local environments, (ii) five distinct traffic participant classes

- Pedestrian (PED), Bicycle, Personal Mobility Devices (PMD), Personal Mobility Aids (PMA), and Others [10] - for YOLOv8 pre-training, and (iii) a counting algorithm capable of accurately identifying and quantifying movements for each class. The proposed approach is intended to overcome the limitations of existing methods and provide a scalable, accurate, and context-aware solution for traffic monitoring and urban planning at mixed-use intersections.

### 3. Methodology

The methodological framework of this research is presented in the first subsection, outlining the overall pipeline for detecting, tracking, and counting pedestrians at intersections. The second subsection proposes an extended counting method - referred to as the Maximum four-lines counting method - that can simultaneously distinguish and count five different classes: Pedestrian (PED), Bicycle, Personal Mobility Devices (PMD), Personal Mobility Aids (PMA), and Others [10].

#### 3.1. Methodological framework

The methodology of this research that is shown in Figure 1 includes eight following steps.

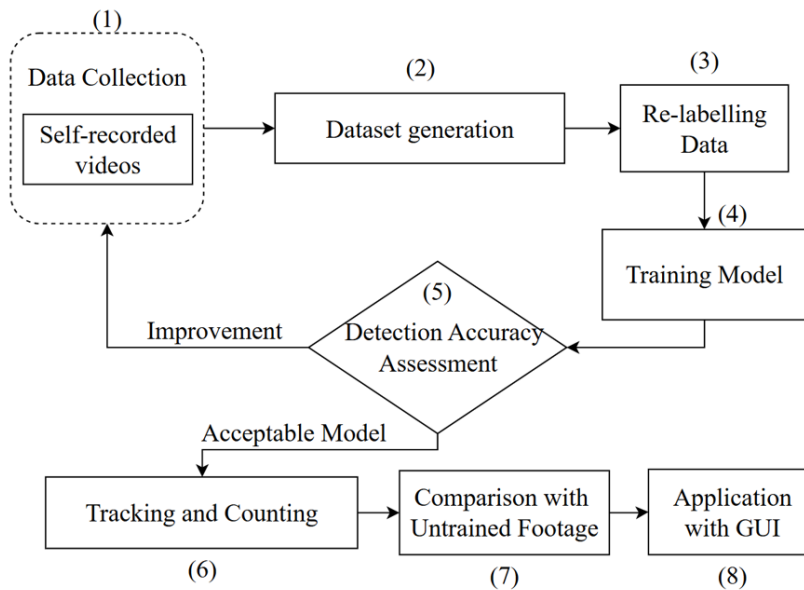


Figure 1. Methodological framework of this study

**Step 1 – Data Collection.** Video footage was collected from fixed surveillance cameras installed at urban intersections with high pedestrian and vulnerable road user activity. Recordings were conducted under diverse environmental conditions, including daytime, nighttime, rain, and sunlight, to ensure dataset variability and represent real-world complexity. These videos reflect typical Vietnamese intersection scenes with mixed flows of pedestrians, cyclists, wheelchair users, and personal mobility device users. High-resolution cameras were employed to capture fine-grained details, which are critical for detecting small and often occluded objects such as children or assistive mobility devices.

**Step 2 – Dataset Generation.** From the recorded videos, representative frames were extracted at regular intervals to construct the training dataset. These frames were carefully selected to ensure coverage of different pedestrian densities, movement directions, lighting conditions, and object classes. The dataset was categorized into five specific classes: pedestrians, cyclists, personal mobility device

users (e.g., e-scooters), personal mobility aids users (e.g., wheelchair), and others users. This multi-class structure enables the model to learn discriminative features for each class, which is particularly important in crowded and unstructured intersection environments.

**Step 3 – Re-labelling Data.** To ensure high-quality annotations, all extracted frames underwent manual re-labelling using tools LabelImg [11]. During this step, bounding boxes were corrected for position, size, and class consistency to minimize labeling noise. Mislabelled or occluded objects were corrected or annotated using standardized class definitions to support consistent model training. This process significantly improved the dataset quality, which is essential for training a reliable deep learning model.

**Step 4 – Training Model.** The pre-trained YOLOv8 model was fine-tuned on the re-labeled dataset using transfer learning techniques [12]. YOLOv8, known for its high accuracy and real-time inference capability, is particularly well-suited for dense and dynamic traffic scenes. Training involved customized hyperparameter tuning and data augmentation techniques, such as horizontal flipping, random brightness adjustment, and rotation, to enhance generalization. The training process aimed to adapt the model to Vietnamese intersection scenarios, especially where object occlusion, overlapping, and diverse movement directions are common.

**Step 5 – Detection Accuracy Assessment.** The performance of the trained detection model was evaluated using standard metrics, including Precision, Recall, mAP@50, and mAP@50–95 [13]. A dedicated validation set, separate from the training data, was used to conduct quantitative assessments. If the detection performance was deemed insufficient (e.g., due to misclassification of wheelchair users or tracking failures in crowded scenes), the dataset and labeling process were revised iteratively. This step ensured that only a model with satisfactory detection accuracy would proceed to the next stages.

**Step 6 – Tracking and Counting.** Upon achieving acceptable detection accuracy, object tracking was performed using ByteTrack [14], which associates detection results across video frames while preserving unique object identities. ByteTrack is effective in crowded environments and during temporary occlusion, making it suitable for unstructured pedestrian flows [14]. ByteTrack is selected because this appears to be one of the fastest tracking methods, capable of processing real-time video at high frame rates (fps) [14]. Based on the generated trajectories, a Maximum Four-Lines Counting Method was employed (in section 3.2), in which up to four virtual lines were drawn for each movement direction (e.g., straight, left-turn, right-turn). The method ensures that each object is counted only once when crossing a predefined line, enhancing accuracy in multi-directional flow conditions.

**Step 7 – Comparison with Untrained Footage.** To evaluate the model's generalization ability, the trained system was applied to untrained video footage from different locations and environmental conditions. The system's counting results were compared with manual ground truth to assess its performance in unseen scenarios.

**Step 8 – Application with GUI.** Finally, the complete system was integrated into a user-friendly Graphical User Interface (GUI) designed for non-technical users such as urban planners and traffic engineers. The GUI allows users to upload video footage, visualize real-time detection and counting, and export analytical data in tabular format.

### 3.2. *Maximum Four-lines counting method*

To accurately count pedestrians and other road users such as cyclists, wheelchair users, and personal mobility device riders at complex intersections, this study proposes a novel “Maximum Four-Lines Counting Method”. This approach enhances robustness by addressing challenges posed by multi-directional flows and non-lane-based movements typically seen in dense urban environments.

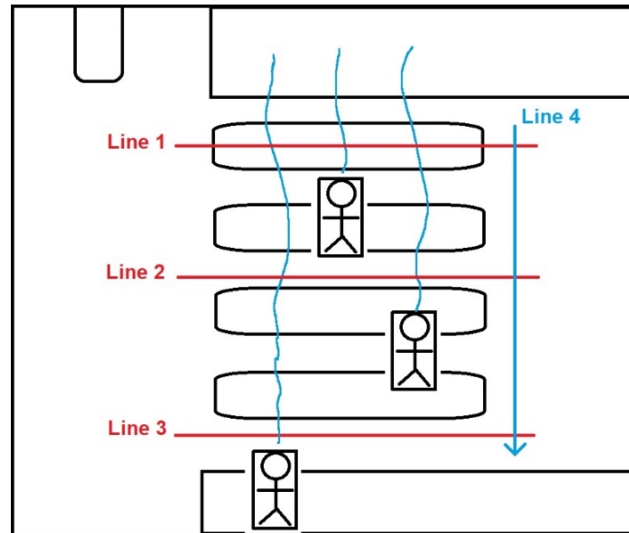


Figure 2. Illustration Four-lines counting method

In this method (shown in Fig. 2), four lines are drawn on the road surface near the region of interest (e.g., crosswalk or intersection exit). Each detected and tracked object is evaluated to determine whether it crosses any of these three lines. For each object, intersections between its movement path and all four counting lines are computed simultaneously. If a pedestrian crosses any line, a counter associated with that line is incremented.

To avoid false from objects moving in the opposite direction, an additional directional guidance line is defined. This orientation line serves as a reference to determine the moving direction of each object. Only objects whose movement vectors align with the direction of the line are considered valid for counting. This directional filtering is essential in uncontrolled intersections where bi-directional movement is common.

In high-density, mixed pedestrian environments such as urban intersections, conventional single-line counting method often suffer from inaccuracies due to frequent occlusions, overlapping trajectories, and meandering pedestrian movements. For example, in the peak periods, large groups of pedestrians enter the scene simultaneously and travel across varying directions. To address these issues, the Maximum Four-Lines Counting Method is introduced as a more robust alternative that significantly reduces missed detections and identity switches caused by occlusions. Furthermore, this approach improves temporal resolution, as movement of pedestrians can be tracked across successive lines, enabling more accurate classification of complete versus partial crossings. In contrast, the single-line method relies on a singular frame of reference, which can easily be disrupted in crowded conditions, leading to undercounting or duplication. Rather than relying on a single counting line - which may miss detections due to occlusion, imperfect alignment, or erratic trajectories - this method takes the maximum count from the three lines as the final result for each class of object. This strategy ensures that any legitimate crossing is recorded while minimizing the risk of undercounting.

The determination of whether an object has crossed a line is based on the geometric analysis of its trajectory. As an object moves, it generates a series of positions across consecutive frames. Between each pair of adjacent positions, a short segment is formed. The algorithm checks whether any of these segments intersect with any of the predefined counting lines. An object is deemed to have “passed” a line if: (1) The trajectory segment intersects the counting line; (2) The intersection point lies within both the segment and the line bounds.



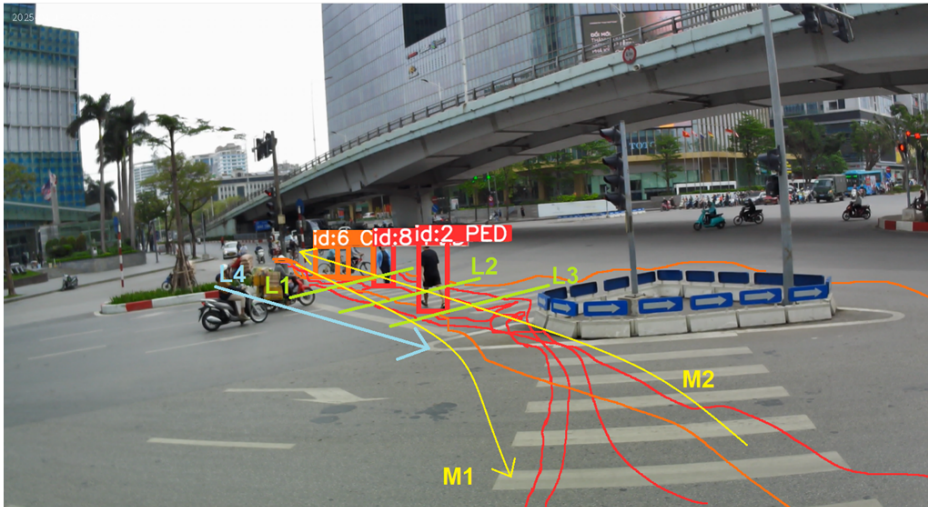


Figure 3. Maximum four-lines counting method at the Dao Tan – Nguyen Chi Thanh intersection

This fine-grained, frame-to-frame checking method significantly improves counting reliability, particularly in scenes where individuals may not cross precisely at the same location or may appear intermittently due to occlusion. The integration of directional filtering and multiple candidate lines provides redundancy and adaptability, yielding more accurate pedestrian counts in mixed traffic conditions.

#### 4. Case study

This case research, data collection was conducted over multiple sessions during both daytime and nighttime to capture diverse lighting and congestion scenarios. A high-resolution fixed surveillance camera was installed at an elevated vantage point to ensure a clear, unobstructed view of the intersection and all pedestrian crossing points. The recorded footage was used to generate the custom dataset for model training and to evaluate the system's performance across different traffic conditions.





##### 4.1. Visual dataset

A total of 22 video footages were collected from various cities and provinces across Vietnam to construct a comprehensive dataset tailored for pedestrian detection. The dataset includes 13 daytime and 9 nighttime recordings, allowing the model to be trained and evaluated under diverse lighting and traffic conditions. Cameras used in data collection had varying resolutions, including 1080p, 720p, and 480p, with all recordings captured at a frame rate of 30 frames per second or higher to ensure sufficient temporal granularity for motion analysis. To build the training dataset, two frames per second were extracted from these segments, resulting in a total of 37,207 images. Each image was manually annotated with bounding boxes corresponding to five distinct pedestrian-related classes: pedestrian, cyclist, personal mobility device, personal mobility aids, and other.

Table 1 summarizes the distribution of annotated objects in the custom pedestrian dataset, categorized into five distinct classes. These categories were defined to reflect the diversity of pedestrians typically observed in urban intersections in Vietnam. The Pedestrian (PED) class, representing individuals walking on foot, constitutes the majority of the dataset, with 92,827 labelled instances, accounting for approximately 77.04% of all annotations. This dominance aligns with the typical composition of crosswalk traffic in urban Vietnamese settings. The Cyclist class, comprising individuals riding bicycles, is the second most common, with 13,741 annotations (11.4%), followed by

users of Personal Mobility Devices (PMD) such as e-scooters and hoverboards, with 6,632 instances (5.5%). The dataset also includes Personal Mobility Aids (PMA) users, which covers individuals using wheelchairs or motorized mobility scooters with 3,725 labelled objects (3.09%). Finally, the Other category includes less frequent but contextually relevant objects such as people pushing strollers, carrying large loads, or using atypical devices. This group accounts for 3,564 annotations (2.96%).

Table 1. Pedestrians types and number of labelled objects

| No | Pedestrian types               | Images   | Number of labelled objects | Percentage (%) |
|----|--------------------------------|--|----------------------------|----------------|
| 1  | Pedestrian (PED)               |   | 92,827                     | 77.04          |
| 2  | Cyclist                        |   | 13,741                     | 11.4           |
| 3  | Personal mobility device (PMD) |   | 6,632                      | 5.5            |
| 4  | Personal mobility aids (PMA)   |  | 3,725                      | 3.09           |
| 5  | Other                          |  | 3,564                      | 2.96           |

#### 4.2. Model evaluation

In the trained model in the Step 4 was run with 100 epochs. Fig. 4 shows performance indexes of the trained model after training progress in this study. The top five images illustrate the performance of the training set while the bottom five images indicate the performance of the validation set. Fig. 4 illustrates confusion matrix normalized that indicate the results of the pedestrian classification model evaluated on the validation set.

In Fig. 4, the  $x$ -axis shows the number of epochs, while the  $y$ -axis represents the corresponding values at each epoch. These charts illustrate the changes in loss values, Precision, Recall and mAP associated with different IoU thresholds (mAP50 and mAP50-95) during the training progress. The loss values decrease gradually, and the performance metrics increase steadily. For the training set, the Precision values are higher than 0.91 after 50 epochs. For the validation set, the mAP50 value reaches a peak of approximately 0.82 and the mAP50-90 value is around 0.51. In other words, the model starts converging at around epoch 53. To conclude, the performance of the trained model is acceptable.

Fig. 5 shows that the model demonstrates good recognition performance because correct prediction of most types is higher than 0.72 except PMA with the proportion of 0.65. Of these, the correction prediction of PED and Cyclist strong with the value of higher 0.8. In addition, misclassification rates for other vehicle types are minor, all below 0.1.

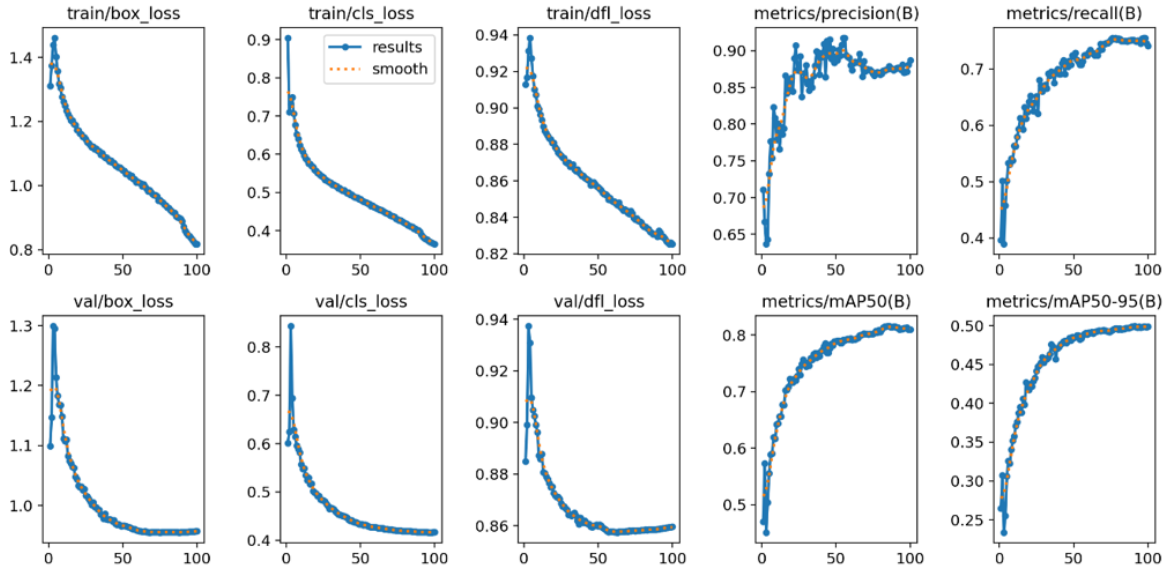


Figure 4. The performance of the trained model

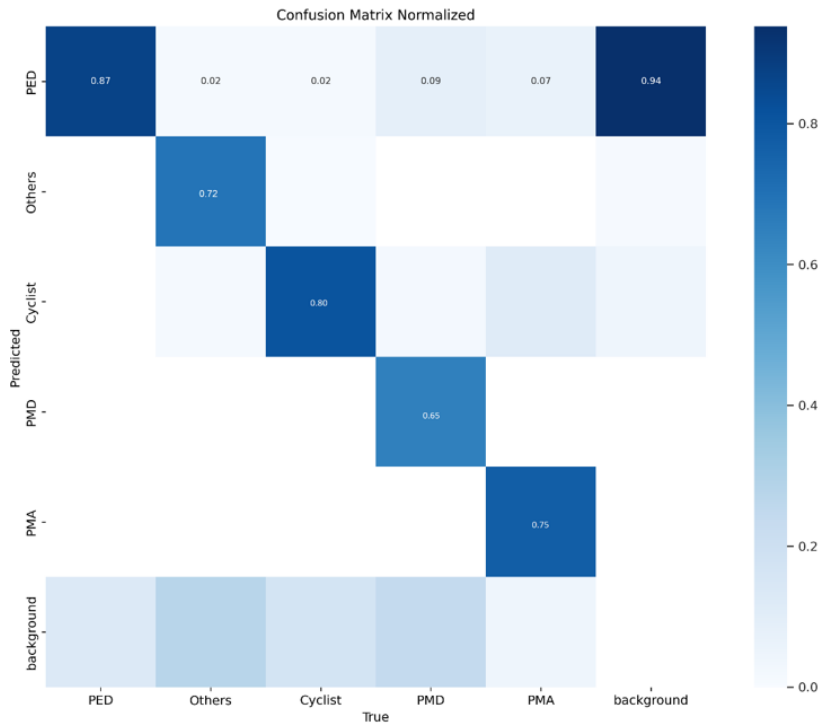


Figure 5. Confusion matrix normalized

#### 4.3. The results of counting model

Case study: Dao Tan – Nguyen Chi Thanh intersection, Vietnam (Fig. 3). To validate the effectiveness of the proposed pedestrian detection and counting framework in a real-world context, a case study was conducted from 07:00 to 07:15 on April 15, 2025 at the Dao Tan – Nguyen Chi Thanh intersection in Hanoi, Vietnam. The area is situated near commercial centers, residential zones, and public transportation hubs, making it a critical location for evaluating pedestrian flow.



The intersection was selected due to its dynamic and complex traffic conditions, including frequent pedestrian crossings, cyclists navigating through vehicle flows, such as wheelchair users and e-scooter riders. These factors provide a robust test environment for assessing the performance of the detection, tracking, and counting algorithms under real-world challenges such as occlusion, varying movement directions, and inconsistent pedestrian behavior.

The counting results by Max four-line customised model (trained with our dataset) and manual counting are compared to evaluate the performance of these models. The lines drawn on the road surface to intersect trajectories for counting purpose are the same for the YOLOv8 and max four-line customised model.

Table 2. Results of pedestrians counting at the Dao Tan – Nguyen Chi Thanh intersection (07:00-07:15 - April 15, 2025)

| M1 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
|-------------------------------------|------|---------|-----|-----|-------|-------|
| Manual counting                     | 32   | 6       | 0   | 0   | 1     | 39    |
| Maximum four-lines                  | 31   | 7       | 0   | 0   | 0     | 38    |
| Maximum four-lines/Manual Ratio (%) | 97%  | 116%    | 0%  | 0%  | 0%    | 97%   |
| M2 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
| Manual counting                     | 14   | 3       | 0   | 0   | 0     | 17    |
| Maximum four-lines                  | 14   | 3       | 0   | 0   | 0     | 17    |
| Maximum four-lines/Manual Ratio (%) | 100% | 100%    | 0%  | 0%  | 0%    | 100%  |

Table 2 presents the comparison between manual pedestrian counts and the results produced by the proposed “Maximum Four-Lines Counting Method” for two distinct traffic movements (M1 and M2) at the Dao Tan – Nguyen Chi Thanh intersection during the morning peak period (07:00–07:15, April 15, 2025). For M1 movement, the total manual count recorded 39, while the automated method detected 38, achieving an overall accuracy of 97%. Specifically, the model detected 31 out of 32 pedestrians (97%), 7 out of 6 cyclists (116%), and accurately recorded zero counts for PMD, PMA, and other categories. In the M2 movement, both manual and automatic counts yielded identical totals of 17, resulting in 100% accuracy across all detected classes. This includes perfect precision for pedestrians and cyclists, demonstrating the method’s reliability in less crowded or structurally simpler movement scenarios.

To further demonstrate the applicability and robustness of the proposed counting methodology beyond the Vietnamese context, a short test was conducted using publicly available footage from a pedestrian-heavy intersection in a foreign urban setting (Figs. 6 and 7). The video illustrates how the model performs in detecting and counting different classes of road users in a non-trained environment. Despite differences in infrastructure, lighting conditions, and pedestrian behavior, the model maintained a high level of detection accuracy and classification consistency. Two video segments were extracted from the original footage available at the following YouTube (source: <https://www.youtube.com/watch?v=MifILi54DPE>).

Two representative footage samples from Singapore, shown in Figs. 6 and 7, were selected to illustrate the model’s performance in unfamiliar urban environments. Each clip captures approximately one complete traffic light cycle, including a full pedestrian crossing phase at a signalized intersection. These samples serve as examples of out-of-distribution scenarios, where the footage characteristics - such as urban layout, pedestrian behavior, and camera placement - differ significantly from the Vietnamese training data. Despite these differences, the model successfully detects and tracks pedestrian

movements, demonstrating its robustness.

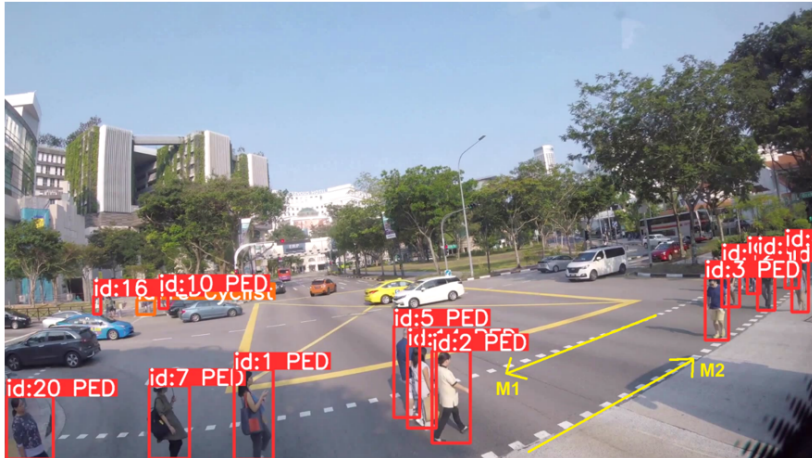


Figure 6. The result detects pedestrian within “15 seconds” from the Singapore footage

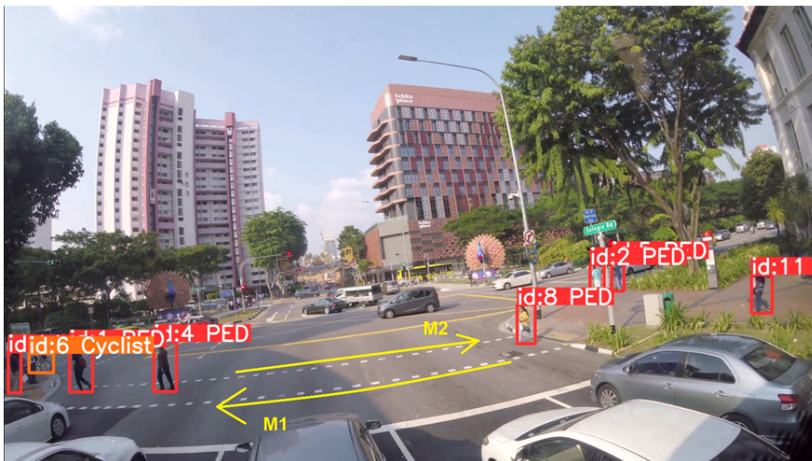


Figure 7. The result detects pedestrian within “1 minute” from the Singapore footage

Table 3. Results of pedestrians counting “15 seconds” at the Singapore intersection

| M1 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
|-------------------------------------|------|---------|-----|-----|-------|-------|
| Manual counting                     | 14   | 0       | 0   | 0   | 0     | 14    |
| Maximum four-lines                  | 13   | 0       | 0   | 0   | 0     | 13    |
| Maximum four-lines/Manual Ratio (%) | 93%  | 0%      | 0%  | 0%  | 0%    | 93%   |
| M2 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
| Manual counting                     | 6    | 0       | 0   | 0   | 0     | 6     |
| Maximum four-lines                  | 6    | 0       | 0   | 0   | 0     | 6     |
| Maximum four-lines/Manual Ratio (%) | 100% | 0%      | 0%  | 0%  | 0%    | 100%  |

Tables 3 and 4 show the counting results at a Singapore intersection under 15-second and 1-minute intervals, representing out-of-distribution conditions. Despite being trained on Vietnamese data, the model achieved high accuracy. In the 15-second test, it reached 93% and 100% accuracy

Table 4. Results of pedestrians counting “1 minute” at the Singapore intersection

| M1 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
|-------------------------------------|------|---------|-----|-----|-------|-------|
| Manual counting                     | 6    | 1       | 0   | 0   | 0     | 7     |
| Maximum four-lines                  | 6    | 1       | 0   | 0   | 0     | 7     |
| Maximum four-lines/Manual Ratio (%) | 100% | 100%    | 0%  | 0%  | 0%    | 100%  |
| M2 movement                         | PED  | Cyclist | PMD | PMA | Other | TOTAL |
| Manual counting                     | 0    | 1       | 0   | 0   | 0     | 1     |
| Maximum four-lines                  | 0    | 1       | 0   | 0   | 0     | 1     |
| Maximum four-lines/Manual Ratio (%) | 0%   | 100%    | 0%  | 0%  | 0%    | 100%  |

for M1 and M2 pedestrian movements, respectively. In the 1-minute test, it achieved perfect results for all detected classes. These outcomes indicate strong generalizability and reliability of the model in unfamiliar environments. The results indicate that the “Maximum Four-Lines Counting Method” performs with high reliability in mixed pedestrian environments. These outcomes confirm the model’s applicability for real-world pedestrian monitoring in complex urban intersections.

To evaluate the real-time capability of the proposed framework, all experiments were conducted using a workstation equipped with an AMD Ryzen Threadripper 3960X 24-Core Processor, 128GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24GB of memory. Under this hardware configuration, the system consistently achieved an average inference speed of approximately 27 frames per second (FPS) on 1080p video inputs, even in high-density and mixed traffic scenarios. This performance might demonstrate the model’s suitability for real-time deployment in practical urban monitoring applications.

## 5. Conclusions

This study proposes a comprehensive and adaptable methodology for the automatic detection, tracking, and counting of diverse pedestrian and personal mobility flows at complex intersections. The framework comprises three key components: (i) a large-scale, labeled dataset that accurately reflects the characteristics of local traffic environments; (ii) the classification of road users into five distinct categories - Pedestrian (PED), Cyclist, Personal Mobility Devices (PMD), Personal Mobility Aids (PMA), and Others - for effective pre-training of the YOLOv8 model; and (iii) a robust counting algorithm designed to accurately detect and quantify the movement of each traffic participant class.

Applied to a real-world case study at the Dao Tan – Nguyen Chi Thanh intersection in Hanoi, Vietnam, the methodology demonstrates robust performance across multiple pedestrian classes (pedestrians, cyclists, PMD, PMA, and others). The high accuracy achieved - up to 100% in some movement scenarios - highlights the model’s potential for reliable, real-time data collection in both urban and suburban settings. Furthermore, the extensive visual dataset developed, comprising over 22 video recordings and nearly 37,000 images with more than 120,000 labelled pedestrian-related objects, provides a strong foundation for training transferable AI models. These models can be adapted to similar traffic conditions in Southeast Asia and other regions with mixed, non-lane-based flows.

The results of this research suggest several practical implications. Transport planners and local authorities can apply the proposed framework to improve traffic monitoring, enhance pedestrian safety, and inform signal optimization and infrastructure design. The inclusion of rare but important pedestrian types such as personal mobility devices (PMD) and personal mobility aids (PMA) also extends the method’s relevance to inclusive mobility planning. Nonetheless, several limitations must be acknowledged. Counting accuracy can vary based on camera positioning, occlusion levels, and

the complexity of movements. Future research will focus on establishing optimal camera placement guidelines, expanding the training dataset to better handle occlusions, and developing trajectory-stitching techniques for interrupted tracking.

## Acknowledgments

The author gratefully acknowledges the support provided by Ngoc Viet Pham from Ticipi technology and transport consultancy joint stock company.

## References

- [1] Din, A. U., Ming, J., Rahman, I. U., Han, H., Yoo, S., Alhrahsheh, R. R. (2023). [Green road transportation management and environmental sustainability: The impact of population density](#). *Heliyon*, 9(9):e19771.
- [2] Feng, C.-M., Sun, J. (2012). [Developing Urban Roads and Managing Motorization](#). In Morichi, S., Acharya, S., editors, *Transport Development in Asian Megacities*, Berlin, Heidelberg, Springer Berlin Heidelberg, 77–106.
- [3] Diogenes, M. C., Greene-Roesel, R., Arnold, L. S., Ragland, D. R. (2007). [Pedestrian Counting Methods at Intersections: A Comparative Study](#). *Transportation Research Record: Journal of the Transportation Research Board*, 2002(1):26–30.
- [4] Yang, H., Ozbay, K., Bartin, B. (2010). Investigating the performance of automatic counting sensors for pedestrian traffic data collection. In *Proceedings of the 12th world conference on transport research*, volume 1115, 1–11.
- [5] Greene-Roesel, R., Diogenes, M. C., Ragland, D. R., Lindau, L. A. (2008). *Effectiveness of a commercially available automated pedestrian counting device in urban environments: Comparison with manual counts*.
- [6] Lesani, A., Nateghinia, E., Miranda-Moreno, L. F. (2020). [Development and evaluation of a real-time pedestrian counting system for high-volume conditions based on 2D LiDAR](#). *Transportation Research Part C: Emerging Technologies*, 114:20–35.
- [7] LEE, G.-G., KIM, W.-Y. (2011). [A Statistical Method for Counting Pedestrians in Crowded Environments](#). *IEICE Transactions on Information and Systems*, E94-D(6):1357–1361.
- [8] Patel, M., Elgazzar, H. (2024). [Pedestrian Crosswalk Safety at Intersections using YOLOv8 Detection](#). In *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, 0426–0430.
- [9] Baul, A., Kuang, W., Zhang, J., Yu, H., Wu, L. (2021). [Learning to Detect Pedestrian Flow in Traffic Intersections from Synthetic Data](#). In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2639–2645.
- [10] Litman, T., Blair, R., Mta, L. A. (2006). *Managing personal mobility devices (PMDs) on nonmotorized facilities*.
- [11] TzuTa, L. (2021). [labellmg](#).
- [12] Ultralytics (2023). [Ultralytics YOLOv8](#).
- [13] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., Zisserman, A. (2009). [The Pascal Visual Object Classes \(VOC\) Challenge](#). *International Journal of Computer Vision*, 88(2):303–338.
- [14] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X. (2022). [ByteTrack: Multi-object Tracking by Associating Every Detection Box](#). In *Computer Vision – ECCV 2022*, Cham, Springer Nature Switzerland, 1–21.