Journal of Science and Technology in Civil Engineering, HUCE, 2026, 20 (1): 100–112

AI-BASED ESTIMATION OF VEHICLE DWELL TIME AT SIGNALIZED INTERSECTIONS IN MOTORCYCLE-DOMINATED MIXED TRAFFIC ENVIRONMENT: A CASE STUDY IN HANOI

Tam Vu[a,*], Ngoc Viet Pham[a]

[a] *Faculty of Transportation Engineering, Hanoi University of Civil Engineering, No. 55 Giai Phong road, Bach Mai ward, Hanoi, Vietnam*

Article history:
Received 27/11/2025, Revised 04/02/2026, Accepted 09/3/2026

Abstract

This study addresses the critical challenge of accurately measuring vehicle dwell times at signalized intersections characterized by motorcycle-dominated traffic flows, a common yet understudied scenario in many Southeast Asian urban centers. We propose a novel computer vision framework that integrates YOLOv8-based vehicle detection with an innovative tracking approach utilizing the H20 reference point to replace conventional centroid-based methods. This strategic reference point selection demonstrates enhanced stability against common challenges in mixed traffic environments, including partial occlusions, and perspective distortions inherent in surveillance camera setups. Applied to a comprehensive case study at the Nguyen Trai – Nguyen Van Loc intersection in Hanoi, Vietnam. Our area-based dwell time measurement algorithm successfully captured stopping durations by combining velocity thresholding with geometric analysis of vehicle trajectories within a precisely defined monitoring zone. Experimental results demonstrate that the proposed system achieves significantly higher accuracy compared to conventional centroid-based approaches, with the mean absolute error reduced to approximately 2–3 seconds across all vehicle classes. These findings offer transportation authorities in developing countries an automated, scalable solution for intersection performance analysis, enabling data-driven traffic management decisions and supporting the optimization of signal timing plans for heterogeneous traffic conditions characterized by high motorcycle dominance. The method compatibility with existing surveillance infrastructure further enhances its potential for practical implementation in urban traffic monitoring systems.

Keywords: computer vision; YOLO; intersections; dwell time; mixed traffic.

https://doi.org/10.31814/stce.huce2026-20(1)-07

1. Introduction

Urban traffic congestion has emerged as a critical challenge in rapidly developing cities worldwide, generating substantial economic losses, environmental pollution, and compromising public safety through extended exposure times in transportation networks. In many Southeast Asian countries, this issue is particularly acute due to exceptional population densities, accelerated motorization rates, and infrastructure that struggles to accommodate mixed traffic flows where motorcycles constitute the dominant mode [1, 2]. The unique characteristics of motorcycle-dominated traffic – including smaller vehicle dimensions, higher maneuverability, and complex stopping patterns at intersections – present distinct challenges for conventional traffic analysis methodologies [3].

Although vehicular traffic management has received considerable research attention, the precise measurement of vehicle dwell times at signalized intersections remains an under-explored yet critical component for understanding intersection efficiency and capacity. Dwell time – defined as the

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

duration vehicles remain stationary within designated intersection zones – serves as a fundamental parameter for evaluating traffic flow characteristics, signal timing effectiveness, and overall intersection performance [4]. Accurate dwell time data provides essential inputs for optimizing signal phasing, assessing intersection level of service, and identifying potential safety improvements in complex traffic environments.

Traditional approaches for measuring vehicle dwell times, including manual observation, and inductive loop detectors, face significant limitations in motorcycle-dominated contexts [5, 6]. Manual observation methods are prone to human error, especially during peak traffic periods, with documented accuracy reductions when monitoring high-volume motorcycle flows [7]. Inductive loop detectors systematically undercount motorcycles due to their smaller metallic mass and unconventional stopping positions, while tube-based systems struggle to distinguish between individual vehicles in dense motorcycle clusters [5, 8]. These limitations are compounded by the characteristic behaviors of motorcycle traffic, including filtering between lanes, angled stopping positions, and partial occlusion scenarios that challenge both human observers and conventional sensors.

To address these measurement challenges, this study proposes an integrated computer vision framework that leverages deep learning-based detection with YOLO architecture and introduces an innovative tracking mechanism utilizing the H20 reference point [9] – the midpoint of the lower bounding box edge – for enhanced stability in vehicle trajectory analysis. The proposed methodology specifically addresses the limitations of conventional centroid-based tracking in mixed traffic environments, where motorcycle occlusion, and irregular stopping patterns traditionally compromise measurement accuracy. The framework is validated through extensive case studies in Vietnam, where motorcycle-dominated intersections present ideal testing conditions for evaluating algorithm performance under real-world complexities.

The remainder of this paper is organized as follows: Section 2 reviews existing literature on vision-based vehicle tracking and dwell time measurement; Section 3 details the proposed methodology including the YOLO-based detection and H20-enhanced tracking system; Section 4 presents experimental results and performance evaluation; and Section 5 discusses conclusions, practical implications, and potential directions for future research.

2. Literature review

Traditional methods for vehicle dwell time measurement, including manual observation, and inductive loop detectors, have demonstrated significant limitations in complex urban intersections characterized by mixed traffic flows. To overcome these constraints, various automated approaches utilizing sensor technologies and computer vision techniques have been extensively explored in recent decades.

A Kalman filter-based tracking approach integrated with morphological operations was developed for highway surveillance applications, achieving over 90% tracking consistency in sparse traffic conditions [10]. Statistical methods utilizing feature-based correlation in the spatiotemporal domain were proposed for vehicle counting, attaining 93% accuracy without requiring complex vehicle models – an advantage in fixed-camera scenarios [11, 12]. While these computer vision approaches demonstrated considerable potential over traditional sensors, common limitations remained, including sensitivity to illumination changes, limited robustness in crowded environments, and challenges in distinguishing overlapping vehicles in dense traffic conditions.

The advent of deep learning architectures, particularly convolutional neural networks, has revolutionized vehicle detection and tracking capabilities. The YOLO (You Only Look Once) family of single-stage object detectors has emerged as a prominent solution for real-time traffic monitoring applications. The effectiveness of YOLOv5 models for vehicle detection at intersections using existing traffic camera infrastructure has been demonstrated, achieving mean average precision exceeding 93.5% across multiple vehicle classes [13]. Meng Cui et al., 2024 similarly proposed an improved algorithm that achieved an average accuracy improvement of 6 percentage points, resolving vehicle re-identification challenges after occlusion [14]. Furthermore, recent research has explored specialized reference points for enhanced tracking stability; notably, the H20 reference point—defined as the midpoint of the lower bounding box edge – has shown superior performance compared to conventional centroid-based methods, particularly for vehicles with irregular orientations and partial occlusions [9].

However, the majority of existing studies focus on standardized traffic conditions in developed countries, utilize limited vehicle categorization schemes, or concentrate primarily on automotive traffic. Few works adequately address the unique challenges posed by motorcycle-dominated traffic environments prevalent in Southeast Asian countries, particularly Vietnam. Moreover, existing dwell time measurement methodologies predominantly rely on centroid-based tracking approaches, which prove suboptimal for motorcycles due to their susceptibility to orientation changes during stopping phases. Additionally, few prior studies provide comprehensive frameworks that integrate specialized reference points with robust tracking algorithms specifically optimized for dwell time analysis in heterogeneous traffic conditions.

To address these research gaps, this study proposes a novel vehicle dwell time measurement framework specifically tailored to motorcycle-dominated intersection environments. The framework comprises three key components: (i) a YOLO-based detection model trained on an extensive dataset representative of Southeast Asian traffic conditions; (ii) an innovative tracking mechanism utilizing the H20 reference point for enhanced stability in trajectory analysis; and (iii) a dwell time measurement algorithm incorporating velocity-based state transition logic to accurately distinguish between complete stops and rolling decelerations. The proposed approach aims to overcome the limitations of conventional tracking methods and provide transportation authorities with a scalable, accurate solution for intersection performance analysis in developing urban areas and especially in optimizing traffic lights.

3. Methodology

The methodological framework of this study is organized into two main components. The first subsection describes the complete processing pipeline for vehicle detection, tracking, and dwell time measurement at signalized intersections, detailing the sequential stages from data acquisition to the final dwell time estimation. The second subsection presents an area-based dwell time measurement method designed to address the specific challenges of motorcycle-dominated traffic environments. This method integrates the H20 reference point to improve tracking stability across 15 vehicle classes, including motorcycles, cars, buses, trucks, etc. The overall framework combines computer vision techniques with velocity-based state transition logic to accurately capture vehicle stopping durations under the complex traffic conditions commonly observed at urban intersections in Southeast Asia.

3.1. Methodological framework

The methodology of this research, as shown in Fig. 1, includes six sequential steps designed to accurately measure vehicle dwell time at signalized intersections.

Step 1 – Vehicle Detection and Classification. The YOLOv8 model was implemented for vehicle detection and classification. The model was trained on a comprehensive dataset comprising fifteen vehicle classes [9]. Transfer learning techniques were employed by fine-tuning pre-trained weights on intersection-specific data, with customized hyperparameter tuning and data augmentation strategies to enhance model robustness in dense traffic conditions [9].

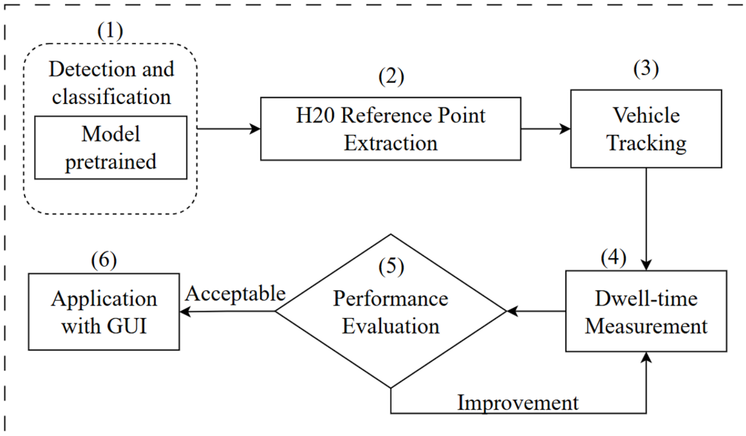

Figure 1. Methodological framework of this study

Step 2 – H20 Reference Point Extraction. Following vehicle detection, the H20 reference point – defined as the midpoint of the horizon within the bounding box and at a height equal to 1/20 of the total height of bounding box [9]. This innovative approach replaces conventional centroid-based tracking methods, providing enhanced stability particularly for motorcycles. The H20 reference points offers reduced sensitivity to vehicle orientation changes and improved accuracy in partial occlusion scenarios.

Step 3 – Vehicle Tracking. The ByteTrack algorithm was employed to associate detection results across video frames while preserving unique vehicle identities. Based on the generated trajectories using H20 reference points, dwell time measurement was performed through velocity-based state transition logic. ByteTrack demonstrates strong performance in crowded scenes and under temporary occlusions, making it well-suited for unstructured [15]. It is chosen as it is among the fastest tracking algorithms, capable of handling real-time video processing at high frame rates (fps) [15].

Step 4 – Dwell-time Measurement. The dwell-time is determined by detecting the state of the object based on changes in its velocity. A vehicle is considered to be in a stopped state when its velocity falls below a predefined threshold across consecutive frames. The dwell time is recorded as the duration from the moment the vehicle enters the stopped state until it leaves the designated dwell-time zone. This zone corresponds to the signal stop area, starting from the stop line.

Step 5 – Performance Evaluation and Validation. The system performance was rigorously evaluated using accuracy compared to manual ground truth annotations. A dedicated validation set, separated from training data, was used for quantitative assessment. If performance was deemed unacceptable, iterative improvements were implemented through parameter optimization and algorithm refinement. The manual ground truth was established by having video footage reviewed frame-by-frame, during which the precise timestamps were recorded for each vehicle when a complete stop was first reached and when the stop line was crossed by the vehicle front to exit the measurement area. The dwell time was then calculated as the difference between these two timestamps.

Step 6 – Application with GUI. The complete system was integrated into an intuitive Graphical User Interface designed for transportation engineers and urban planners. The GUI enables users to upload intersection footage, visualize vehicle detection and dwell time measurements, configure Area of Interest parameters, and export analytical reports for traffic management decisions.

3.2. Area-based dwell time measurement method

To accurately measure vehicle dwell time at signalized intersections, this study proposes a novel "Area-Based Dwell Time Measurement Method" that enhances robustness in motorcycle-dominated traffic environments by addressing challenges posed by irregular stopping patterns and frequent occlusions.

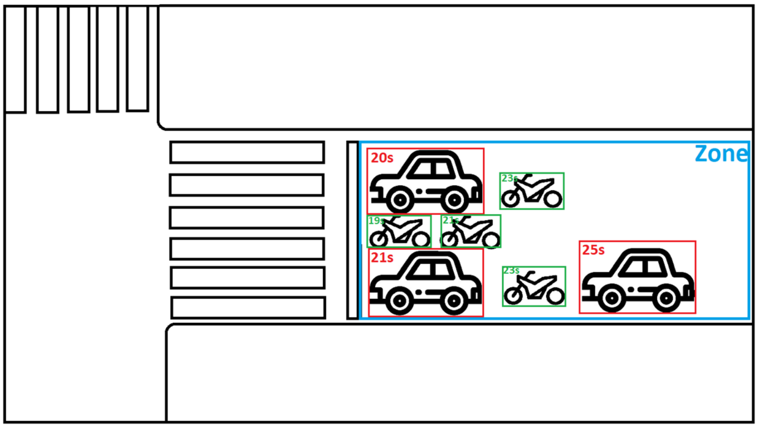

Figure 2. Area-Based Dwell Time Measurement Method

In this method (shown in Fig. 2), a designated measurement area is defined extending from the stop line to the maximum detection range of the surveillance system. This area encompasses the complete stopping zone where vehicles typically queue during red signal phases. Each detected and tracked vehicle is continuously monitored while positioned within this predefined area.

The H20 reference point – defined as the midpoint of the horizontal axis at a height equal to 1/20 of the total bounding box height from the bottom edge – serves as the primary tracking anchor for movement analysis. This specialized reference point provides enhanced stability over conventional centroid-based methods, particularly for motorcycles that frequently exhibit leaning behaviors during stopping phases.

To determine vehicle movement states, displacement analysis is conducted by calculating the Euclidean distance traveled by the H20 reference point across consecutive frames. This frame-to-frame displacement measurement is transformed to real-world velocity (m/s) using camera calibration parameters and perspective transformation techniques. The velocity calculation follows the formula:

$$v = \frac{\Delta d}{\Delta t} * C \tag{1}$$

where Δd represents the pixel displacement of the H20 reference points between frames; Δt is the inter-frame time interval; and C is the conversion factor from pixels to meters derived from camera calibration.

A vehicle is classified as being in "dwell state" when its calculated velocity remains below the 1 m/s threshold for a minimum of five consecutive frames [16]. This multi-frame validation ensures robustness against momentary pauses or tracking fluctuations. The dwell time measurement initiates when this stationary condition is first detected and terminates when the vehicle front exceeds the stop line boundary while exiting the measurement area.

The "Area-Based Dwell Time Measurement Method" provides comprehensive spatial coverage that accommodates diverse stopping patterns. This approach significantly reduces measurement errors caused by partial occlusions and vehicle orientation changes. Furthermore, the method improves temporal accuracy, as vehicle movements can be continuously monitored throughout their presence in the measurement zone, enabling precise differentiation between complete stops and slow rolling movements. The determination of dwell state transitions is based on continuous velocity monitoring and positional analysis. The algorithm continuously evaluates whether the vehicle current position remains within the predefined measurement area. The instantaneous velocity falls below the 1 m/s threshold and the low-velocity condition persists for the required minimum duration.

This fine-grained, frame-to-frame velocity monitoring significantly improves measurement reliability, particularly in complex scenarios where vehicles may exhibit stop-and-go behavior or temporary positional adjustments while waiting. The integration of area-based monitoring and velocity thresholding provides adaptive measurement capabilities, yielding more accurate dwell time data in mixed traffic conditions characterized by motorcycle dominance.

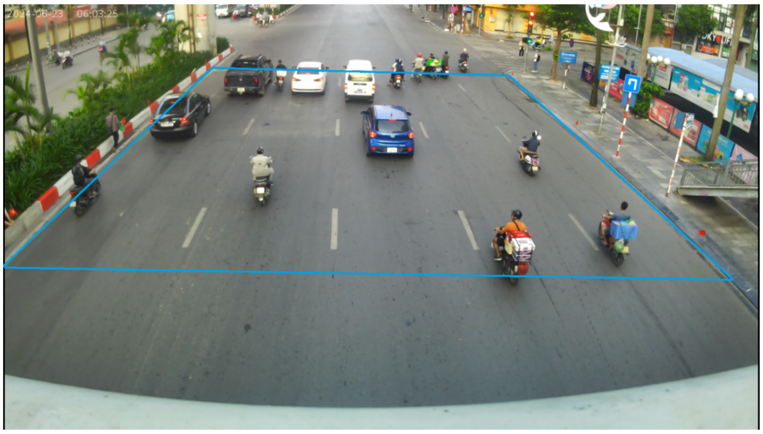

Figure 3. The quadrilateral using the Area-Based Dwell Time Measurement Method at the Nguyen Trai – Nguyen Van Loc intersection

Fig. 3 illustrates the practical implementation of the Area-Based Dwell Time Measurement Method at the intersection of Nguyen Trai – Nguyen Van Loc Street. The quadrilateral area demarcates the precise measurement zone, extending from the stop line to the maximum detectable range of the surveillance system. This carefully defined polygon encompasses the critical stopping region where vehicles accumulate during red signal phases, ensuring comprehensive coverage of diverse queuing patterns characteristic of motorcycle-dominated intersections.

The proposed method effectively captures the complete stopping duration from initial deceleration below the velocity threshold to final departure across the stop line, providing transportation engineers with precise data for intersection performance analysis and signal timing optimization.

4. Case study

The proposed methodology was implemented and evaluated at the signalized intersection of Nguyen Trai – Nguyen Van Loc Street in Hanoi, Vietnam. This intersection represents a typical urban traffic node characterized by high motorcycle dominance. The intersection features complex traffic patterns including straight-moving flows, left-turning movements, and significant pedestrian crossings, creating challenging conditions for accurate vehicle tracking and dwell time measurement. Video data was collected during the morning (6:00–6:30), a high-resolution surveillance camera (1920×1080 pixels, 30 fps) was positioned at an elevated location providing comprehensive coverage of the northwest approach of the intersection.

Figure 4. Vehicle trajectory H20 reference point at the Nguyen Trai – Nguyen Van Loc intersection

Fig. 4 provides a visual of vehicle trajectories generated by the proposed H20 reference point (lines). The trajectories are overlaid on a sample frame from the video footage at the study intersection. As evidenced in the figure, the trajectories derived from the H20 point demonstrate significantly higher stability and smoothness, particularly for motorcycles. The H20 point, the vehicle contact point with the road, produces a more consistent and physically grounded path. This enhanced trajectory stability is fundamental to achieving accurate velocity calculation and reliable dwell time measurement, as it reduces noise in the displacement data used to determine a vehicle stop-and-go states.

4.1. Experimental set up and implementation

The proposed framework was implemented and evaluated at the signalized intersection of the Nguyen Trai and Nguyen Van Loc street in Hanoi, Vietnam. This location represents a typical motorcycle-dominated urban intersection with complex traffic patterns.

The YOLOv8x model developed by Vu et al. [9] was employed for vehicle detection and classification, leveraging its specialized training on Vietnamese traffic datasets. This model demonstrated exceptional performance in the motorcycle-dominated environment of the Nguyen Trai – Nguyen Van Loc intersection, achieving near-comprehensive vehicle detection across all videos. The model prior exposure to similar traffic patterns and vehicle characteristics in Vietnamese urban contexts contributed significantly to its robust detection capabilities in this challenging scenario.

The results in Table 1 indicate that the Vu et al. pre-trained model [9] provides broader class coverage and stronger detection performance than YOLO11 [17] in mixed traffic. In particular, the pre-trained model detects additional vehicle categories that YOLO11 [17] fails to capture (e.g., bus and other vehicles) and supports a more fine-grained taxonomy by distinguishing sub-classes by size and type (e.g., small/medium/large passenger vehicles and light/medium trucks). Quantitatively, it also produces substantially higher detection counts for key classes such as cars (xe_con) and motorcycles (xe_may), suggesting improved robustness and coverage under heterogeneous, real-world conditions.

Table 1 shows the Vu et al. pre-trained model [9] successfully identified a total of 26,860 vehicle instances across 11 distinct classes, confirming its capability to handle the diverse vehicle types present in Vietnamese traffic environments. Particularly noteworthy is the model effectiveness in detecting motorcycles, which accounted for 15,450 detections (approximately 57.5% of total vehicles), accurately reflecting the motorcycle-dominated nature of the intersection. The detection coverage

Table 1. Vehicle detection statistics from video of the Nguyen Trai – Nguyen Van Loc intersection

Vehicle class	Total detections of Yolo11	Total detections of model pretrained	Average confidence score of model pretrained
xe_con	7646	6234	0.80
buyt	1135	312	0.58
khach_nho		1058	0.65
khach_vua		639	0.51
khach_lon		795	0.77
tai_nho	2115	1238	0.67
tai_trung_1		636	0.62
tai_trung_2		333	0.50
xe_may	13481	15450	0.70
xe_dap	203	110	0.57
xe_khac	0	55	0.67
Overall	24580	26860	0.64

across all vehicle types, from personal cars (xe_con) to specialized transport vehicles (khach_lon, tai_trung), indicates the model comprehensive understanding of local traffic characteristics. The observed variation in confidence scores across vehicle classes can be attributed to several factors inherent to the traffic surveillance setup. The rear-mounted CCTV perspective, which is characteristic of traffic monitoring systems in Vietnam, presents unique challenges for vehicle recognition. This specific camera angle affects different vehicle types unequally – while passenger cars (xe_con) maintained the highest confidence score (0.80), larger vehicles like buses (0.58) and certain truck categories (tai_trung_2: 0.50) showed reduced confidence, likely due to their more complex visual appearances and varying orientations when viewed from behind.

Despite these variations, the overall detection performance remains sufficient for subsequent tracking and dwell-time analysis. The model ability to maintain reasonable accuracy across all vehicle types, while particularly effective for motorcycles that dominate the traffic flow, provides a solid foundation for this study as shown in Fig. 5. This performance characteristic also suggests good potential for future integration with existing traffic surveillance infrastructure, where similar camera angles and perspectives are commonly employed.

Following vehicle detection, the H20 reference point was extracted for each identified vehicle. The H20 point is mathematically defined as the midpoint positioned along the lower edge of the bounding box at a height equal to 1/20 of the total bounding box height from the bottom [9].

The rear-mounted CCTV angle, typical of Vietnamese traffic monitoring infrastructure, creates unique challenges for vehicle positioning analysis. In this configuration, the H20 reference point demonstrates superior performance for several reasons. First, for larger vehicles such as buses and trucks, the H20 point maintains closer proximity to the road surface, providing enhanced stability against bounding box fluctuations caused by vehicle suspension movements or detection variances. This ground-hugging characteristic ensures more consistent positioning data compared to the centroid, which may shift significantly with changes in vehicle orientation or detection boundaries.

Second, in scenarios where vehicles are positioned immediately at the stop line, the centroid

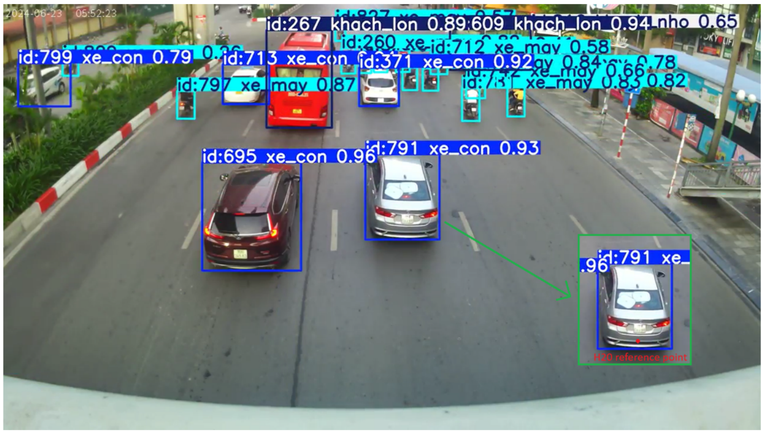

Figure 5. The recognition and classification model was implemented at the Nguyen Trai – Nguyen Van Loc intersection

often falls outside the designated measurement area due to its placement higher in the vehicle body, particularly for tall vehicles. In contrast, the H20 point, being anchored to the vehicle base, remains reliably within the area of interest, ensuring accurate spatial registration for dwell time computation.

Third, and particularly crucial for one-way traffic analysis, the H20 point minimizes misclassification risks between opposing lanes. When applied to large vehicles extending across lane boundaries, the centroid may erroneously position itself in adjacent lanes, potentially leading to incorrect lane assignment and dwell time attribution. The H20 point road-level positioning maintains consistent lane alignment, as it closely follows the vehicle actual road contact point rather than its geometric center.

This robustness against perspective distortions and lane assignment errors makes the H20 reference point particularly well-suited for integration with existing traffic surveillance systems, where camera angles and positions are often constrained by practical installation considerations. The method consistency across diverse vehicle types and its resilience to the specific challenges posed by rear-mounted camera perspectives validate its selection for accurate trajectory analysis in motorcycle-dominated mixed traffic environments.

5. Results and analysis

The dwell times of vehicles within the defined area were successfully measured using the proposed method. The performance of the area-based dwell time measurement method is validated against manual ground truth measurements. A sample of the first 10 recorded entries is presented in Table 2, illustrating the system direct output.

This table further highlights the limitation of the YOLO11 box-center–point implementation even though it nominally follows the same Area-Based Dwell Time Measurement Method. In several cases, YOLO11 fails to return any dwell-time estimate (“None”), which is consistent with missed or unstable detections when vehicles are partially occluded or visually fragmented within the queue. Even when YOLO11 does produce a value, the estimated dwell times are often biased downward relative to both the proposed AI method and the manual reference.

The results demonstrate the system core capability to track and measure dwell times for multiple vehicles of different classes simultaneously. For instance, as shown at timestamp 06:00:50, the system seamlessly recorded the dwell times for “khach_nho”, “xe_con”, and “tai_trung_1” concurrently. This parallel processing capability represents a significant advancement, overcoming the fundamental limitation of traditional manual methods, which struggle to monitor more than a few vehicles at a time.

The comparative results reveal generally close alignment between automated and manual measurements, with several instances showing perfect matches (e.g., Vehicle ID 7060) and minor variations within acceptable limits. The discrepancies observed in some cases, such as vehicle ID 6934 (7s vs 8s) and 7044 (60s vs 63s), can be attributed to differences in the precise definition of stopping and human observation. The system demonstrates particular strength in consistently tracking multiple vehicles simultaneously, a task that poses challenges for manual observation, especially during high-density traffic conditions. The overall pattern indicates that the automated method provides reliable dwell time measurements while offering significant advantages in processing speed and scalability compared to traditional manual approaches.

Table 2. Sample Area-Based Dwell Time Measurement Method for vehicles at Nguyen Trai – Nguyen Van Loc intersection

Vehicle ID	Vehicle Class	Start Time by AI method (s)	Exit Time by AI method (s)	Dwell Time by AI method (s)	Dwell Time by Yolo11 using center point (s)	Dwell Time by Manual (s)
6934	xe_con	06:00:42	06:00:49	7	7	8
7044	khach_nho	06:00:50	06:01:50	60	59	63
7060	xe_con	06:00:50	06:01:46	56	56	56
7061	tai_trung_1	06:00:50	06:01:47	57	55	54
7122	xe_con	06:00:56	06:01:50	54	50	58
7470	xe_con	06:01:14	06:01:52	37	None	35
7922	xe_may	06:01:30	06:01:41	11	None	15
7967	tai_trung_2	06:01:38	06:01:54	16	12	17
8800	xe_con	06:02:55	06:04:08	72	67	70
8943	xe_con	06:03:11	06:03:34	23	20	26

The dwell time characteristics across different vehicle classes at the Nguyen Trai – Nguyen Van Loc intersection are summarized in Table 3, revealing significant variations in stopping behavior patterns. The analysis of 200 vehicle stopping events provides valuable insights into intersection performance under mixed traffic conditions.

Table 3. Dwell time statistics by vehicle class at Nguyen Trai – Nguyen Van Loc intersection

Vehicle class	Sample size	Average dwell time (s)	Standard deviation (Std)	Min (s)	Max (s)
khach_lon	8	46.3	26.2	14.0	85.0
khach_nho	8	45.3	22.4	10.0	74.0
khach_vua	4	51.8	17.1	32.0	68.0
tai_nho	3	31.7	21.5	8.0	50.0
tai_trung_1	7	49.3	22.0	7.0	77.0
tai_trung_2	1	16.0	16.0	16.0	16.0
xe_con	91	38.2	21.9	5.0	76.0
xe_may	78	15.6	14.0	5.0	66.0
Overall	200	36.7	20.2	12.1	64.0

The data reveals a clear dichotomy in dwell time patterns between different vehicle categories. Medium passenger vehicles (khach_vua) exhibited the longest average dwell time at 51.8 seconds, followed closely by medium trucks (tai_trung_1) at 49.3 seconds and large passenger vehicles (khach_lon) at 46.3 seconds. These extended stopping durations likely reflect the slower acceleration characteristics and more cautious driving behavior associated with larger vehicles. In stark contrast, motorcycles (xe_may) demonstrated the shortest average dwell time of 15.6 seconds, less than one-third of the longest durations observed. This substantial difference can be attributed to motorcycles' superior maneuverability, smaller size, and ability to filter through traffic, allowing them to reach stop lines later and depart earlier than larger vehicles. The small sample size for tai_trung_2 ($n = 1$) limits meaningful interpretation for this vehicle class.

The considerable standard deviations across all vehicle classes, particularly for khach_lon (26.2 seconds) and tai_trung_1 (22.0 seconds), indicate high variability in individual stopping behaviors within each category. This variability reflects the complex interplay of factors including driver behavior, traffic signal timing, queue position, and intersection-specific conditions.

The overall average dwell time of 36.7 seconds provides a benchmark for intersection performance assessment, while the range of 12.1 to 64.0 seconds highlights the diverse operational characteristics present in the mixed traffic stream. These findings underscore the importance of considering vehicle-specific dwell time patterns when designing traffic signal timings and intersection management strategies, particularly in motorcycle-dominated environments where traditional homogeneous approaches may prove inadequate.

The accuracy of the proposed dwell time measurement system was quantitatively evaluated using three standard metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). MAE measures the average magnitude of errors without considering direction, calculated as the mean of absolute differences between predicted and actual values. RMSE gives more weight to larger errors through squaring before averaging, making it more sensitive to outliers. MAPE expresses the error as a percentage of actual values, providing a relative measure of accuracy.

Table 4. Performance evaluation metrics by vehicle class

Vehicle	MAE	RMSE	MAPE
khach_lon	3.3	3.8	8.79%
khach_nho	3.9	4.7	15.01%
khach_vua	2.0	2.4	4.55%
tai_nho	1.7	1.9	21.53%
tai_trung_1	1.9	2.7	5.78%
tai_trung_2	1.0	1.0	5.88%
xe_con	2.0	2.9	7.84%
xe_may	2.5	3.2	29.86%
Overall	2.3	3.1	16.66%

The evaluation results presented in Table 4 reveal significant variations in performance across vehicle classes. The overall system achieved MAE of 2.3 seconds, RMSE of 3.1 seconds, and MAPE of 16.66%, indicating generally acceptable accuracy for traffic engineering applications. However, motorcycles (xe_may) exhibited the highest MAPE at 29.86%, significantly exceeding other vehicle classes. This elevated error rate can be attributed to the characteristic behaviors of motorcycles in Vietnamese traffic conditions, including frequent lane filtering, forward creeping while waiting, and stopping positions that often extend beyond the designated stop line. These behaviors create challenges for consistent dwell time measurement, as motorcycles may partially enter and exit the measurement area multiple times during a single stopping event, leading to difficulties in precisely determining the exact start and end points of dwell periods. The variation in performance across vehicle classes underscores the importance of context-specific validation for traffic monitoring systems operating in motorcycle-dominated environments. Despite the challenges with motorcycle measurement, the system overall performance remains viable for practical applications.

All experiments were performed on a workstation featuring an AMD Ryzen Threadripper 3960X 24-core CPU, 128 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory. Under this configuration, the system achieved an average processing throughput of 27 frames per second (FPS) on 1080p video, confirming its suitability for large-scale traffic monitoring applications.

6. Conclusions

This study proposes a comprehensive and adaptable computer vision methodology for the automatic detection, tracking, and measurement of vehicle dwell times in motorcycle-dominated intersection environments. The framework comprises three key components: (i) a specialized YOLOv8 detection model trained on large-scale datasets representative of Southeast Asian traffic conditions; (ii) the innovative H20 reference point tracking method that significantly enhances stability over conventional centroid-based approaches; and (iii) a robust area-based dwell time measurement algorithm designed to accurately capture stopping durations across diverse vehicle classes.

Applied to a real-world case study at the Nguyen Trai – Nguyen Van Loc intersection in Hanoi, Vietnam, the methodology demonstrates robust performance across multiple vehicle classes with varying dwell time characteristics. The high accuracy achieved – deviation is only 2–3(s) in dwell time measurement compared to manual ground truth – highlights the model potential for reliable, real-time traffic data collection in complex urban environments.

The results of this research suggest several practical implications for traffic engineering. Transportation authorities and urban planners can apply the proposed framework to improve intersection performance analysis, optimize signal timing plans, and enhance traffic management strategies. The system ability to accurately capture motorcycle behavior – representing over 57% of detected vehicles – provides particularly valuable insights for addressing the unique challenges of motorcycle-dominated intersections. The methodology scalability and compatibility with existing surveillance infrastructure further enhance its practical implementation potential.

Nonetheless, several limitations must be acknowledged. Measurement accuracy can vary based on camera perspective, occlusion intensity, and the complexity of vehicle interactions. Future research will focus on establishing optimal camera placement guidelines, expanding the training dataset to better handle extreme occlusion scenarios, and developing enhanced tracking techniques for maintaining vehicle identity through prolonged interruptions. Additional work will also explore the integration of trajectory prediction models to further improve dwell time measurement precision in high-density traffic conditions.

Acknowledgment

This work was supported by the Vietnamese Ministry of Education and Training under Grant B2025-XDA-03.

References

[1] Bray, D. B., Holyoak, N. (2015). Motorcycles in Developing Asian Cities: A Case Study of Hanoi. In *37th Australasian Transport Research Forum*. Unpublished conference paper.

[2] Feng, C.-M., Sun, J. (2012). Developing Urban Roads and Managing Motorization. In Morichi, S., Acharya, S., editors, *Transport Development in Asian Megacities: A New Perspective*, Berlin, Heidelberg, Springer Berlin Heidelberg, 77–106.

[3] Vu, T., Preston, J. (2023). Assessing the Social Costs of Mixed Transport Systems with a Dominance of Motorcycles. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*, 17(3):80–101.

[4] Dion, F., Ghanim, M. (2007). Impact of Dwell Time Variability on Transit Signal Priority Performance at Intersections with Nearside Bus Stop. In *86th Annual Meeting of the Transportation Research Board*.

[5] Sheik Mohammed Ali, S., George, B., Vanajakshi, L., Venkatraman, J. (2011). A Multiple Inductive Loop Vehicle Detection System for Heterogeneous and Lane-Less Traffic. *IEEE Transactions on Instrumentation and Measurement*, 61(5):1353–1360.

[6] Rajab, S. A., Othman, A. S., Refai, H. H. (2012). Novel vehicle and motorcycle classification using single element piezoelectric sensor. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, IEEE, 496–501.

[7] Majumder, M., Wilmot, C. (2023). Accuracy Assessment and Guidelines for Manual Traffic Counts from Pre-Recorded Video Data. *Journal of Transportation Technologies*, 13(4):497–523.

[8] Kanhere, N. K., Birchfield, S. T., Sarasua, W. A., Khoeini, S. (2010). Traffic Monitoring of Motorcycles during Special Events Using Video Detection. *Transportation Research Record: Journal of the Transportation Research Board*, 2160(1):69–76.

[9] Vu, T., Thai, H. N., Pham, V. N., Vu, H. T., Luong, A. T., Van Luong, T. (2025). Counting Mixed Traffic Volumes at Motorcycle-Dominated Intersections by Using Computer Vision. *International Journal of Intelligent Transportation Systems Research*, 23(1):146–164.

[10] Farag, W. (2021). Kalman-filter-based sensor fusion applied to road-objects detection and tracking for autonomous vehicles. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 235(7):1125–1138.

[11] Meng, Q., Song, H., Zhang, Y., Zhang, X., Li, G., Yang, Y. (2020). Video-Based Vehicle Counting for Expressway: A Novel Approach Based on Vehicle Detection and Correlation-Matched Tracking Using Image Data from PTZ Cameras. *Mathematical Problems in Engineering*, 2020(1):1–16.

[12] Zhang, Y., Zhao, C., Zhang, Q. (2017). Counting vehicles in urban traffic scenes using foreground time-spatial images. *IET Intelligent Transport Systems*, 11(2):61–67.

[13] Mishra, S., Yadav, D. K. (2024). Vehicle Detection in High Density Traffic Surveillance Data Using YOLOv5. *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, 17(2):216–227.

[14] Cui, M., Zhang, Y. (2024). Intersection vehicle detection technology based on DeepSORT algorithm. In Zhang, J., Khan, Z. H., Zeng, P., editors, *Fourth International Conference on Mechanical, Electronics, and Electrical and Automation Control (METMS 2024)*, SPIE, SPIE, page 45.

[15] 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.

[16] Ramezani, M., Geroliminis, N. (2015). Queue Profile Estimation in Congested Urban Networks with Probe Data. *Computer-Aided Civil and Infrastructure Engineering*, 30(6):414–432.

[17] Ultralytics (2024). *Ultralytics YOLO11*.