Rath Wishart et al 2024 - Evaluating Safety Metrics for VRUs at Urban Traffic Intersections using Infrastructure LIDAR

Conference Paper *in* SAE Technical Papers · April 2024

DOI: 10.4271/2024-01-2641

CITATIONS
0

READS
311

2 authors:

Prabin Kumar Rath
Arizona State University
11 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Jeffrey Wishart
Arizona State University
67 PUBLICATIONS **781** CITATIONS
SEE PROFILE

2024-01-2641 Published 09 Apr 2024

Evaluating Safety Metrics for Vulnerable Road Users at Urban Traffic Intersections Using High-Density Infrastructure LiDAR System

Prabin Kumar Rath and Blake Harrison Arizona State University

Duo Lu Rider University

Yezhou Yang Arizona State University

Jeffrey Wishart Science Foundation AZ/AZ Comm Authority

Hongbin Yu Arizona State University

Citation: Rath, P.K., Harrison, B., Lu, D., Yang, Y. et al., "Evaluating Safety Metrics for Vulnerable Road Users at Urban Traffic Intersections Using High-Density Infrastructure LiDAR System," SAE Technical Paper 2024-01-2641, 2024, doi:10.4271/2024-01-2641.

Received: 24 Oct 2023 Revised: 06 Jan 2024 Accepted: 08 Jan 2024

Abstract

Ensuring the safety of vulnerable road users (VRUs) such as pedestrians, users of micro-mobility vehicles, and cyclists is imperative for the commercialization of automated vehicles (AVs) in urban traffic scenarios. City traffic intersections are of particular concern due to the precarious situations VRUs often encounter when navigating these locations, primarily because of the unpredictable nature of urban traffic. Earlier work from the Institute of Automated Vehicles (IAM) has developed and evaluated Driving Assessment (DA) metrics for analyzing car following scenarios. In this work, we extend those evaluations to an urban traffic intersection testbed located in downtown Tempe, Arizona. A multimodal infrastructure sensor setup, comprising a high-density, 128-channel LiDAR and a 720p RGB camera, was employed to collect data during the dusk period, with the objective of capturing data during the transition from daylight to night. In this study, we present and empirically assess the benefits of high-density LiDAR in low-light and dark conditions—a persistent challenge in VRU detection when compared to traditional RGB traffic cameras. Robust detection and tracking algorithms were utilized for analyzing VRU-to-vehicle and vehicle-to-vehicle interactions using the LiDAR data. The analysis explores the effectiveness of two DA metrics based on the i.e. Post Encroachment Time (PET) and Minimum Distance Safety Envelope (MDSE) formulations in identifying potentially unsafe scenarios for VRUs at the Tempe intersection. The codebase for the data pipeline, along with the high-density LiDAR dataset, has been open-sourced with the goal of benefiting the AV research community in the development of new methods for ensuring safety at urban traffic intersections.

Introduction

Smart infrastructure technology is indispensable for the advancement of automated mobility, playing a vital role in ensuring the safety of vulnerable road users (VRUs) and human-driven vehicles (HDVs). Infrastructure-based sensors, when combined with vehicle-to-infrastructure (V2I) connectivity, could potentially offer crucial situational awareness to Automated Driving Systems (ADS)-controlled vehicles (AVs), especially in complex urban driving scenarios, such as at traffic intersections.

Conventionally, cameras have been used as the primary infrastructure sensors for real-time traffic monitoring and surveillance [1]. These sensors aid safety and security in smart cities; however, the quality of data obtained from these sensors is insufficient for monitoring AV-related safety metrics. Prior work from IAM by Wishart et al. [2] introduced DA metrics for quantifying the safety performance of AVs. Reliability of these metrics hinges on precise estimation of object odometry within the infrastructure's reference frame [3]. While image data from calibrated infrastructure cameras can be used for detection and tracking at traffic intersections [4], it is important to note that such data are susceptible to environmental factors, with detection being critically affected by varying lighting and weather conditions [5, 6].

Many of the limitations of infrastructure-based cameras can be mitigated with the use of LiDAR units.

These time-of-flight sensors remain unaffected to changes in lighting conditions, generating consistent data during both day and night time scenarios. In an earlier work at IAM, Anshuman et al.[7] showed the efficacy of infrastructure-based LiDAR units in measurement of DA metrics for vehicles. A follow-on work from Siddarth et al. [8] compared odometry fidelity between infrastructure-based and vehicle-top-based LiDAR setups. While previous studies have established LiDAR units as being reliable for calculating AV metrics, these analyses have primarily focused on vehicles at freeway intersections. In this work, we expand the analysis to encompass urban city intersections, with a particular emphasis on VRUs. We intend to begin answering the following key questions: (1) Are infrastructure-based LiDARs suitable for analyzing urban traffic scenarios? (2) How effective are AV safety metrics in analyzing interactions between vehicles and VRUs?

In the following sections, we delve into the literature on 3D detection-tracking and AV metric calculation. We provide details on the pre-processing techniques, deep learning model, and tracking algorithm that facilitate real-time processing of high-density LiDAR data frames. Finally, we outline the experimental setup used for data collection and present both qualitative and quantitative results obtained from our analysis.

Related Work

LiDAR-based object detection and tracking has been researched extensively in recent years, primarily to address the failure modes encountered with cameras. With the advent of deep learning, there has been a shift from clustering-based approaches [9, 10] to end-to-end learning-based pipelines. Today, the state-of-the-art is based on representation learning on perception datasets such as KITTI [11], Waymo [12], and nuScenes [13] along with many others that have been made publicly available by the AV industry. Although deep learning allows for multi-modal fusion to incorporate data from multiple sensors such as LiDAR units and cameras, LiDAR-only methods are more attractive options due to their data consistency and invariance properties [14].

Most of the approaches for deep learning on LiDAR point clouds fall into two categories: point-based and voxel-based methods. Point-based methods treat point clouds as a permutation-invariant set of points and learn spatial features for each point using a point-encoder such as PointNet [15]. In contrast, voxel-based methods discretize 3D space into a finite grid of small cubes known as voxels. Points within each voxel are averaged, and the model learns the spatial relations within the voxels using Convolutional Neural Networks (CNNs). Both methodologies involve a trade-off between precision and time complexity. Point-based methods are typically more precise but slower, while voxel-based methods are faster but less precise due to data loss during the discretization process [14]. More recent methods utilize a hybrid point-voxel approach where points within the discrete voxels are processed using a point encoder and the learned voxel representations are further passed to CNNs, thus incorporating the best of both worlds [16]. Another popular methodology involves using Bird's Eye View (BEV) data for detection, in which the point cloud is flattened along the z-axis to produce a 2D set of points. While these methods often deliver excellent runtime performance, they are prone to substantial data loss, which can constrain detection accuracy [17].

Our work focuses on the detection of VRUs, such as pedestrians, cyclists, and individuals using micro-mobility vehicles, which presents a unique set of challenges. LiDAR point clouds exhibit an inherent property of increasing sparsity as the radial distance increases. VRUs are relatively small entities for LiDAR, resulting in extremely sparse point returns, which lack the necessary features for object detection. Voxel-based and BEV-based approaches perform poorly in VRU detection as they compromise on information to reduce runtime complexity, whereas, point-based methods are too slow and do not meet the minimum real-time requirements for AV applications [16]. Hence, point-voxel based hybrid approach are an ideal choice for our task. One research group [7] used PointPillars [18] as it stands out to be one of the fastest hybrid algorithms, which discretizes 3D space into pillars along the z-axis. Each pillar is encoded using PointNet to generate a pseudo image, which is further processed by a 2D CNN to extract multi-scale features. These features are then stacked and fed into an SSD [19] head for getting the detection results. Another research group [8] utilized Complex-YOLO [20] which is an extension of the single-stage YOLO object detection algorithm to 3D point clouds. It uses Eular Region Proposal Network (E-RPN) for accurate heading prediction. In this work, we utilize PV-RCNN [21], an abbreviation for Point-Voxel RCNN, which employs voxel set abstraction and keypoint sampling for precise object detection. We found that PV-RCNN is robust to point cloud sparsity, and the ROI grid pooling algorithm enables successful identification of VRU features even with 15-20 point sparse returns.

Tracking algorithms consume the detection results and fill in any missing detections by maintaining a belief state on individual objects in the scene. Bayesian algorithms, such as the Kalman filter [22] and its variants, are the most widely used methods for such multi-object tracking. Typically, the 7-DOF bounding box state is tracked over a short time horizon, and objects are removed from the tracking list if no associations can be found for a certain threshold number of data frames. Recently, deep-representational features from object-centric models have also been used as state vectors in Kalman filters [23]. Tracking algorithms assign a consistent and unique track identifier to individual entities, preventing identity switches during partial occlusion [24]. An inherent advantage in 3D tracking compared to 2D is that, since two entities cannot occupy the same 3D space, it is easier for tracking algorithms to associate bounding boxes over time.

The position, velocity, and acceleration information obtained from tracking algorithms can be used to evaluate safety metrics between pairs of entities within a traffic intersection [25]. Elli et al. [26] utilized the CARLA simulator [27] to extract this information and validated the DA

metrics proposed in [2], thereby demonstrating their robustness and relevance over conventional counterparts. Jammula et al. [4] collected camera videos mounted on the infrastructure to track vehicles from a BEV perspective and calculated real-world DA metrics. Similarly, researchers in [7, 8] utilized LiDAR units to perform the DA metrics calculations. However, earlier research has not explored scalability aspects for metrics calculation. Previous work [28, 7, 29] either assume the availability of vehicle pairs or brute-force evaluation on all possible pairs with $O(N^2)$ complexity. In this study, we address such run-time scalability issues and assess the effectiveness of the conventional PET and the MDSE DA metric for VRU safety analysis.

Data Collection Setup

Our multi-modal sensor setup included a Hesai Pander 128 LiDAR unit along with a Hikvision HD surveillance camera as shown in Figure 1 and Figure 6. The Hesai LiDAR provided 128 rings with a 200m range at the Tempe intersection where the diagonally opposite corner was at a distance of 65m from the sensor setup. The high range allowed us to capture dense point clouds, thus alleviating the sparsity problem with the LiDAR. Both the LiDAR unit and camera were tilted at 45 degrees, facing towards the center of the intersection. Data were collected at four different intervals on a single day Sept 11, 2022, MST between 6:20 pm and 7:30 pm, with each interval lasting from 30 seconds to 10 minutes. Duration details of our dataset are shown in Table 1. The variation in interval lengths is due to some scenarios being specific to VRU-vehicle interactions, while others are standard vehicle-vehicle scenarios. Figure 2 shows the transition from daylight to night. An important feature to observe

FIGURE 1 Data collection setup at downtown Tempe intersection.

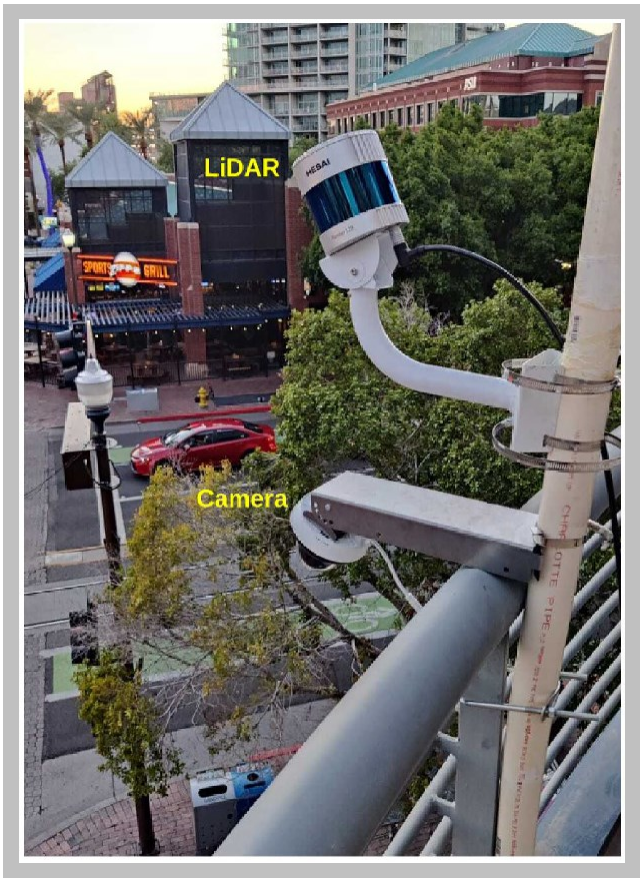
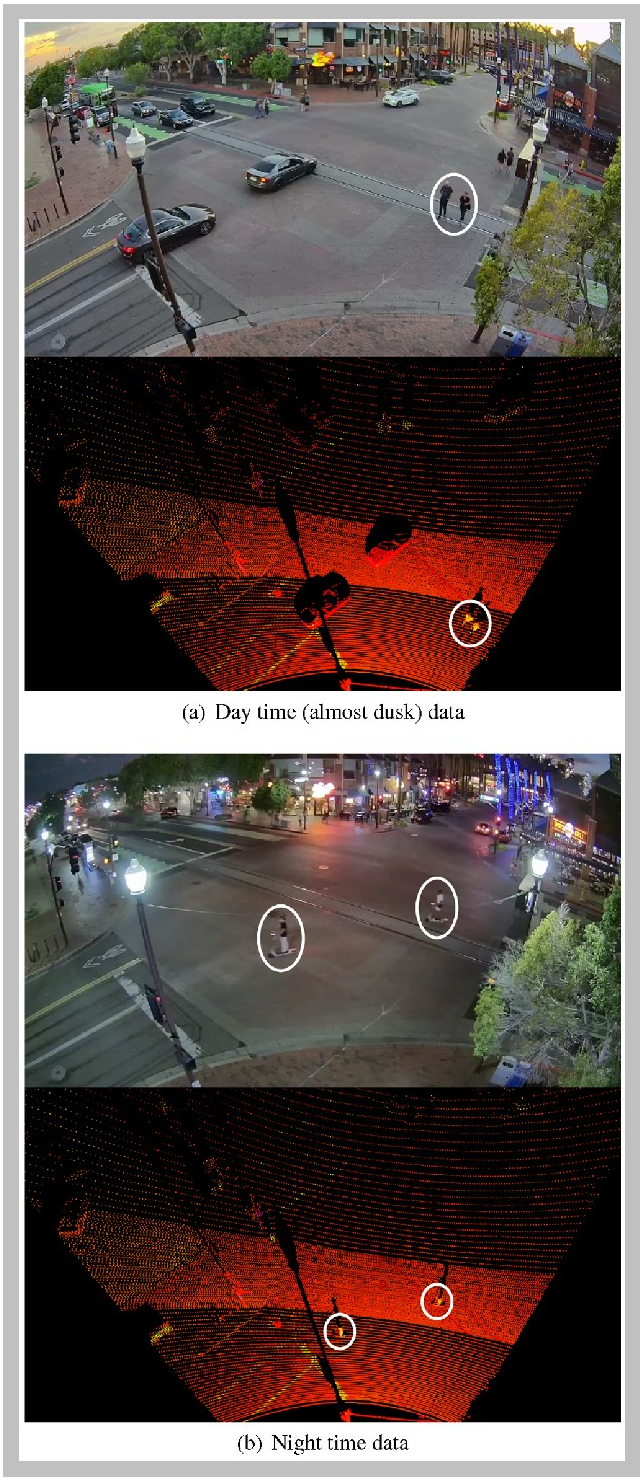


FIGURE 2 Sample data from the Tempe intersection. VRUs (pedestrians and e-scooter riders) are highlighted for visualization. Point clouds shown here are the raw data obtained from the LiDAR without any pre-processing.



(a) Day time (almost dusk) data

(b) Night time data

TABLE 1 Time intervals for the dataset collected on Sept 11, 2022 MST at Downtown Tempe. In total, our dataset comprised 18 minutes of data collection.

Interval #	Start Time	End Time
1	6:28 pm	6:30 pm
2	6:38 pm	6:41 pm
3	7:08 pm	7:11 pm
4	7:13 pm	7:23 pm

here is the consistency in the LiDAR data across the two figures, in contrast to the significant change in lighting conditions observed in the camera data. Our dataset consists solely of data from vehicles driven by humans. However, if we assume that future autonomous vehicles (AVs) will mimic human driving patterns, the methods described in this study should be effective for integrating and adapting AVs, treating them as ego vehicles for analysis.

Methodology

The high-density point cloud data obtained from the infrastructure-based LiDAR unit is processed in three steps to get robust odometry estimates of intersection entities. These estimations are then used to evaluate metrics formulations for real-time traffic safety analysis.

Preprocessing

The 128-channel LiDAR captures data at a rate of 20 frames per second, with each frame containing approximately 690,000 points. To cover the intersection adequately, only 110 degrees of the horizontal field of view are needed, accounting for approximately 250,000 points. Many of these points are redundant data from the surfaces of nearby objects, which do not significantly contribute to detection accuracy. Therefore, we down-sample the point cloud using a high-resolution voxel grid with a leaf size of 10cm. This down-sampling process leads to a more uniform density distribution compared to the original dense point cloud, with minimal data loss. The resulting point cloud consists of around 30,000 points, making it suitable for real-time 3D object detection.

Object Detection

We employ the PV-RCNN [21] model from the open-source MMDetection3D [30] library as it provides a simple API with efficient CUDA implementations. We specifically selected the model pre-trained on the KITTI [11] dataset as we empirically observed that the scan patterns of Hesai LiDAR closely resemble those of Velodyne LiDARs. The KITTI point cloud data is collected from a vehicle top perspective; therefore, we transform the point cloud by aligning the ground parallel to the XY plane while positioning the LiDAR unit 2 meters above the ground plane. This step is essential because deep-learning models do not perform well on out-of-distribution data. Nonetheless, it's worth noting that infrastructure-based LiDAR units capture data from an isometric viewpoint. As a result, the infrastructure-based LiDAR units collect a relatively higher amount of information compared to a vehicle-top setting. For instance, infrastructure-based LiDAR units can observe not only the vehicle's sides but also the roofs, even for large-scale vehicles like trucks, buses, etc. The down-sampling and perspective transformation steps bring the point cloud closer to the KITTI data distribution, thereby assisting the model in producing high-quality detection results.

Tracking and Data Association

Tracking is accomplished by applying a 7-DOF Kalman filter to the bounding box results obtained from the object detection model. We utilize the implementation from [31] as it has an intuitive interface for parameter tuning and readily fits into our detection framework. The detection model provides confidence scores for the bounding box predictions, which are used to tune the sensitivity [10] of the tracking algorithm. We set a confidence threshold of 0.4 for vehicles and 0.0 for VRUs. In safety systems, emphasizing recall is essential to minimize false negatives. Our threshold parameters achieve a higher recall in VRU detection while maintaining the number of false negatives within acceptable limits. In general, we found that the PV-RCNN model exhibits impressive recall in VRU detection, and it retains high precision even at low confidence thresholds [21], thereby enabling accurate tracking of traffic entities. The association is done using the Hungarian algorithm on Generalized Intersection Over Union (GIOU) scores between pairs of bounding boxes from consecutive detection frames. We start tracking as soon as we associate an object in ζ number of consecutive frames and remove the track if we fail to associate for β frames. We use $\zeta = 3$ and $\beta = 4$ for vehicles, whereas for VRUs we use a more relaxed set of parameters with $\zeta = 1$ and $\beta = 5$. In both cases, we set the GIOU association threshold at 1.5.

Safety Metrics

The evaluation of safety metrics traditionally involves the examination of pairs of entities in a scene that meet specific preconditions. For instance, metrics like PET and MDSE are designed to assess interactions between entities that have overlapping trajectories. Detecting such interactions with vehicles is straightforward because conflict points are predetermined based on lane design [32] or pre-specified at road intersections. However, in urban intersections involving VRUs, the situation is more complex. Pedestrians may follow erratic trajectories, sometimes crossing without a walk signal or deviating from designated crosswalk zones. In such scenarios, it becomes crucial to identify potentially unsafe situations and determine conflict points on the fly for online metrics

calculations. Arguably any pair of entities that could run into an accident while moving in different directions must have a common intersection point before the crash. This intuition motivates us to define real-time conflict points by checking for intersections between scaled velocity vectors of the traffic entities. Specifically, we utilize the following three-step algorithm to identify potentially unsafe entity pairs:

1. For every entity i in the scene, we find the instantaneous velocity vector $v^i(t)$ and scale it along the vector direction $\tau = 5$ seconds into the future to obtain $v^i_{scaled}(t)$.
2. Using the Bentley–Ottmann algorithm [33] we find all the intersecting scaled vector pairs in the XY plane and consider the corresponding entity pairs to be in an unsafe situation.
3. Aligning with the terminology from [4], we assign leader-follower relationships such that VRUs are always designated as leaders in VRU-vehicle pairs, while in vehicle-vehicle pairs, any vehicle can assume the role of the leader.

Safety metrics calculations are computationally inexpensive, typically involving simple arithmetic operations. However, computing them for every possible entity pair is sub-optimal, as many pairs will not interact due to differences in their locations and directions. Our work introduces a novel method to identify potentially unsafe pairs with a definable conflict point. The time complexity for our algorithm is $O(Nlog(N))$ thus rendering it suitable for real-time application. The scaling parameter τ can be viewed as a horizon for assessing potential unsafe interactions. Setting $\tau = 5$ implies checking for possible interactions within a 5-second window into the future. Once the potentially unsafe pairs are determined, we calculate PET (Eqn 1) and MDSE (Eqn 2) metrics with respect to the instantaneous conflict points detected by our algorithm. Illustrations of two such unsafe situations are shown in Figure 3.

$$PET = \left| t_2 - t_1 \right| \tag{1}$$

where

t_1: Arrival time of the leader to conflict point
t_2: Arrival time of the follower to conflict point

A PET metric violation (PETV) occurs when the value drops below a pre-determined threshold. For our analysis, we set this threshold at 1.5s.

$$MDSEI = \begin{cases} 0, & \text{if } d^{long,intersect} < d^{long,intersect}_{min} \\ 1, & \text{otherwise} \end{cases}$$

$$d^{long,intersect}_{min} = v^{long}_{l}\rho_l + \frac{1}{2}a^{long}_{l,max,accel}\rho_l^2 + \frac{\left(v^{long}_{l} + \rho_l a^{long}_{l,max,accel}\right)^2}{2a^{long}_{l,min,deccel}} \tag{2}$$

where

$d^{long,intersect}$: Instantaneous distance of the leader from the conflict point

v^{long}_{l}: Norm of the instantaneous velocity of the leader ρ_l: Reaction time of the leader

$a^{long}_{l,max,accel}$, $a^{long}_{l,min,deccel}$: Responsibility-Sensitive Safety (RSS) parameters for MDSE formulation

The MDSE Infringement (MDSEI) [2] is a binary value that indicates the MDSE metric violation. It occurs when the conflict point lies within the safety envelope of the leader.

The parameter values for our metric calculations are derived from the analysis presented by Elli et al. in [26]. For both vehicle-to-vehicle and VRU-to-vehicle interactions, we assume Naturalistic Driving Study (NDS) parameters: $\rho_l = 0.2s$, $a^{long}_{l,max,accel} = 1.8m/s^2$, and $a^{long}_{l,min,deccel} = 3.6m/s^2$. It must be noted that these parameters were originally presented for the analysis of vehicle-to-vehicle interactions. We retain these parameters unchanged and evaluate their behavior and performance in the context of VRU-to-vehicle interactions. The results of our experiments are discussed in the following section.

Results

A comprehensive set of results from the detection model is available in the Appendix Figure 7 and Figure 8 where vehicles and VRUs are highlighted with green and red bounding boxes respectively. We found that density uniformization using down-sampling improves the detection quality with pre-trained models while allowing an average 10Hz real-time loop frequency. Table 2 presents a comparative analysis of object detection performance between camera and LiDAR technology. For the camera-based detection, we employed a pre-trained YOLOv8 model. This comparison was conducted without applying any threshold on the Intersection Over Union (IOU) of bounding boxes, ensuring that the results are unbiased estimates of the detection performance. Consistent with our initial expectations about the superiority of LiDAR in night time detection, the data reveals that LiDAR outperforms cameras in both day time and night time scenarios.

Figure 6 shows the tracklets obtained from a 10-minute dataset (Interval 4) captured at the Tempe

TABLE 2 Camera vs LiDAR quantitative comparison for object detection. P (Precision) and R (Recall) [34] were estimated from our dataset at 0% IOU threshold. The third sub-column (All) is the mean over both classes.

Day Time - Camera				**Day Time - LiDAR**			
	Ped	**Veh**	**All**		**Ped**	**Veh**	**All**
P	0.434	0.656	0.545	P	0.876	0.991	**0.933**
R	0.380	0.698	0.539	R	0.532	0.798	**0.665**
Night Time - Camera				**Night Time - LiDAR**			
	Ped	**Veh**	**All**		**Ped**	**Veh**	**All**
P	0.143	0.494	0.377	P	0.827	0.944	**0.885**
R	0.384	0.756	0.632	R	0.761	0.791	**0.776**

TABLE 3 Safety metrics violations for VRU-vehicle interactions. Column descriptions from left to right: interval number, number of tracked VRUs, count of unsafe situations identified, columns three to five delineate metrics violations within the identified unsafe situations, while the sixth column indicates the number of situations where both PET and MDSE metrics were violated. The number of vehicle tracklets is shown in Table 4 second column.

Interval #	VRU track-lets	Unsafe situations	PETV	MDSEV	Common violations
1	23	5	2	0	0
2	12	5	1	0	0
3	27	12	8	4	3
4	147	37	27	15	10
Total	**232**	**59**	**38**	**19**	**13**

TABLE 4 Safety metrics violations for vehicle-vehicle interactions. Column descriptions from left to right: interval number, number of tracked vehicles, count of unsafe situations identified, columns three to five delineate metrics violations within the identified unsafe situations, while the sixth column indicates the number of situations where both PET and MDSE metrics were violated.

Interval #	Vehicle track-lets	Unsafe situations	PETV	MDSEV	Common violations
1	29	19	6	7	3
2	42	28	10	14	4
3	66	42	14	29	11
4	193	148	73	80	32
Total	**330**	**237**	**103**	**130**	**50**

intersection. The patterns observed in the tracked trajectory plots offer insights into the areas at the intersection where the density of VRUs is notably high, particularly at the corners where pedestrians typically wait for the walk signal.

The quantitative results derived from our collected dataset are showcased in Table 3 for VRU-to-vehicle interactions and in Table 4 for vehicle-to-vehicle interactions. In unsafe situations where no metric violations occur, we observed that only the tips of the scaled vectors intersect. As the entities are distant from each other in these cases, there are no metric violations for such interactions. We evaluated both the conventional PET and the more recent MDSE metrics for numerous identified unsafe situations in our dataset, illustrating two such examples in Figure 3. The MDSE metric in Figure 4 effectively identifies an unsafe VRU-to-vehicle interaction, particularly when a vehicle makes a sharp turn close to a pedestrian crossing the road. On the other hand, Figure 5 illustrates the PET metric violation for an unsafe interaction between two vehicles at the intersection, when one of the vehicles makes a near-miss turn. Both situations were labeled as unsafe due to the close proximity between the two entities while moving through the intersection. These observations were consistent across various analyses, reaffirming the robustness of our proposed system in handling intersection dynamics. Table 5 demonstrates the impact of the velocity scaling parameter τ on safety metrics calculations, revealing that higher scaling leads to a more conservative analysis of the intersection. While both metrics proved effective in our analysis, we empirically found MDSE to be more versatile for practical applications due to its customizable parameters, which allow for fine-tuning the metric to meet traffic-specific requirements. For example, using conservative parameters [26] for MDSE calculations identified both situations shown in Figure 3 as unsafe, whereas setting an appropriate PET threshold to classify both situations as unsafe was challenging.

Conclusion and Future Work

In this paper, a high-density, infrastructure-based LiDAR unit was utilized to collect traffic data at a city intersection. We discussed pre-processing techniques and the data pipeline in-depth, providing detailed explanations for design choices and parameters. We introduced an online methodology for identifying unsafe situations at traffic intersections and evaluated metrics for real-time safety analysis. The results confirmed the viability of LiDARs for analyzing urban traffic scenarios. Our experiments also showed that DA safety metrics are effective for analyzing vehicle-to-vehicle and VRU-to-vehicle interactions at intersections. While our results are promising, there remains significant room for improvement. For instance, the Hungarian association algorithm for tracking exhibits a complexity of $O(N^3)$, potentially becoming a bottleneck when analyzing larger intersections. Therefore, more efficient algorithms are necessary for tracking. Our selection of parameters for safety metrics was guided by previous literature. Nevertheless, more suitable parameters can be established by considering the unique characteristics of the intersection. Moreover, the run-time loop frequency of 10Hz may not suffice for analyzing fast-moving vehicles. Although such scenarios are infrequent in urban traffic intersections, having a higher observation frequency for more detailed safety analysis could be advantageous. Lastly, our metric formulations did not take the

FIGURE 3 Unsafe situation identification followed by real-time PET metrics evaluation. The blue arrow indicates the velocity vector, the dotted lines represent the 5s scaled projection, and the pink dot shows the instantaneous conflict point calculated by our system. The tailing lines represent the tracked trajectory of the entities over time. Note: Only the unsafe entities are emphasized in the images to direct the reader's attention. Other detected entities are not considered in this visualization.

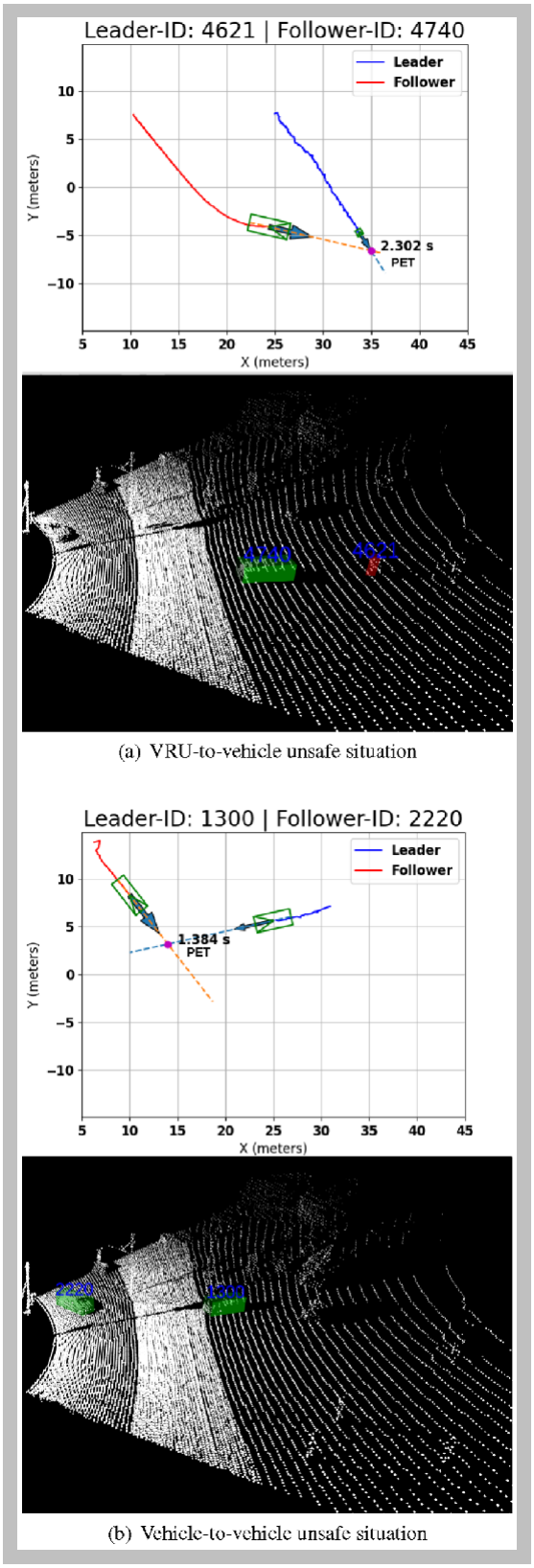


FIGURE 4 VRU-to-vehicle metrics evaluation for the scenario shown in Figure 3(a). Here the MDSEI infringement occurs at 0.325s when the distance to the conflict point (CP) becomes less than the distance to the safety envelope (SE). Time step 0 in these plots is relative to the time when the situation was identified as unsafe by our system.

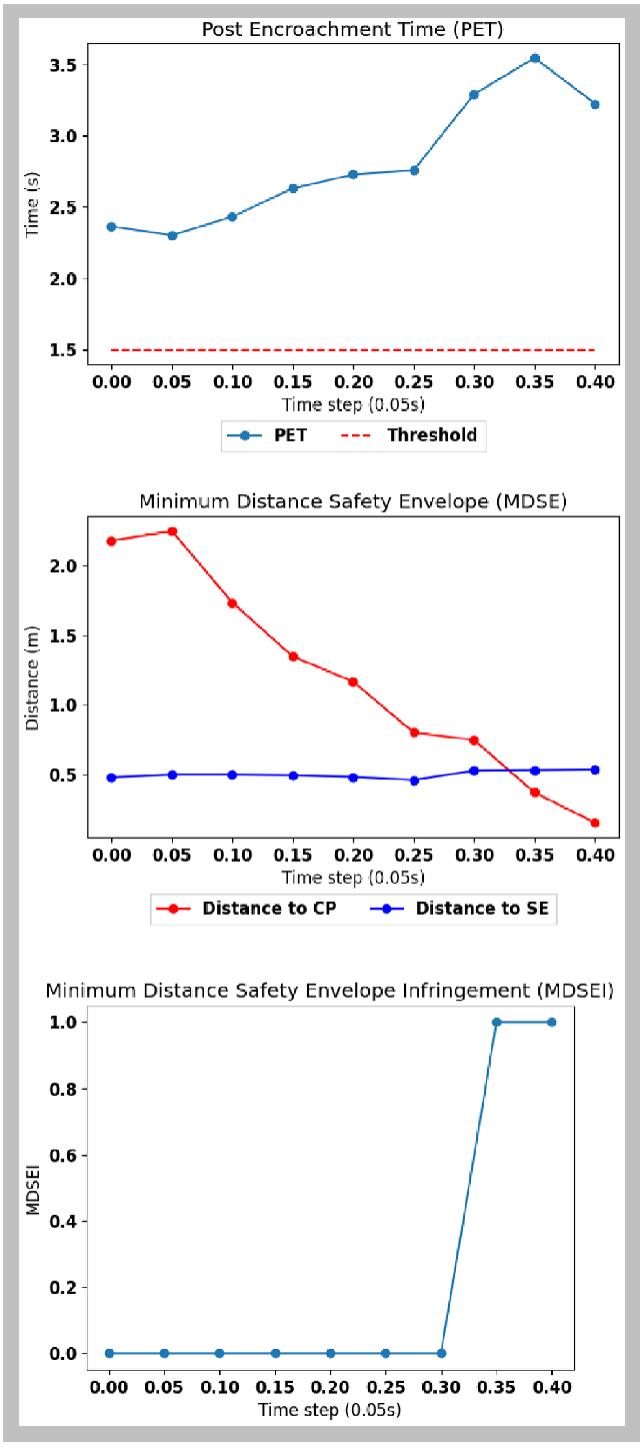


FIGURE 5 Vehicle-to-vehicle metrics evaluation for the scenario shown in Figure 3(b). Here PET violation occurs multiple times after 0.5s. Time step 0 in these plots is relative to the time when the situation was identified as unsafe by our system.

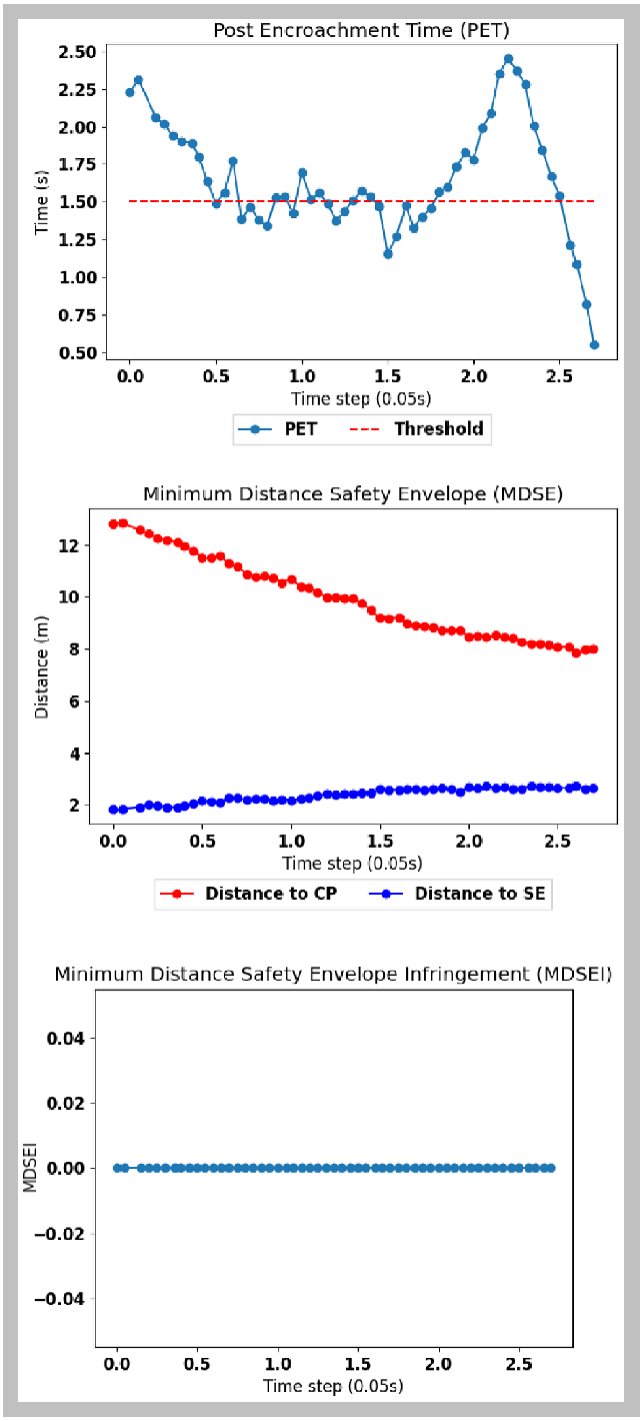
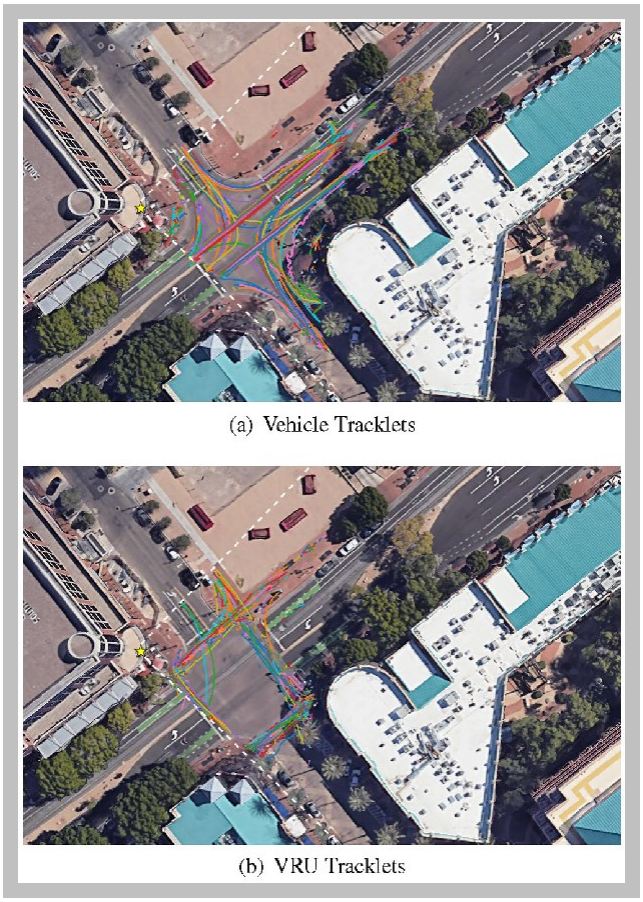


TABLE 5 Effect of projection scaling parameter τ on the results. Results for τ = 5s are shown in the last rows of Table 3 and Table 4.

3s scaling				
	Unsafe situations	**PETV**	**MDSEV**	**Common violations**
VRU-to-vehicle	18	11	5	2
Vehicle-to-vehicle	107	56	57	26
7s scaling				
	Unsafe situations	**PETV**	**MDSEV**	**Common violations**
VRU-to-vehicle	105	62	34	19
Vehicle-to-vehicle	306	131	154	64

FIGURE 6 Trajectory of tracked objects (Sept 11, 2022 MST, 7.13 pm to 7.23 pm) superimposed on a satellite image of the Tempe intersection. The sensor setup is located at the star shown in the images. The range of the LiDAR sensor has been represented using white-dotted lines. Each color represents a different object tracklet.



(a) Vehicle Tracklets

(b) VRU Tracklets

dimensions of entities into consideration, which can have a significant impact, especially when dealing with vehicles of different sizes. Future research might incorporate novel formulations that consider entity dimensions. The scripts for metrics analysis, LiDAR data, tracking results, and video demonstrations are available at this repository.

References

1. Datondji, S.R.E., Dupuis, Y., Subirats, P., and Vasseur, P., "A Survey of Vision-Based Traffic Monitoring of Road Intersections," *IEEE Transactions on Intelligent Transportation Systems* 17, no. 10 (2016): 2681-2698.
2. Wishart, J., Como, S., Elli, M., Russo, B. et al., "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles," SAE Technical Paper 2020-01-1206 (2020).
3. Kidambi, N., Wishart, J., Elli, M., and Como, S., "Sensitivity of Automated Vehicle Operational Safety Assessment (OSA) Metrics to Measurement and Parameter Uncertainty," SAE Technical Paper 2022-01-0815 (2022), https://doi.org/10.4271/2022-01-0815.
4. Jammula, V.C., Wishart, J., and Yang, Y., "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated Vehicles Using Real-World Data," SAE Technical Paper 2022-01-0062 (2022), https://doi.org/10.4271/2022-01-0062.
5. Medina, J.C., Chitturi, M., Benekohal, R.F., and Board, T., "Illumination and Wind Effects on Video Detection Performance at Signalized Intersections," Urbana, 2008.
6. Medina, J., Chitturi, M., and Benekohal, R., "Effects of Fog, Snow, and Rain on Video Detection Systems at Intersections," *Transportation Letters* 2, no. 1 (2010): 1-12.
7. Srinivasan, A., Mahartayasa, Y., Jammula, V., Lu, D. et al., "Infrastucture-Based LIDAR Monitoring for Assessing Automated Driving Safety," SAE Technical Paper 2022-01-0081 (2022).
8. Das, S., Rath, P., Lu, D., Smith, T. et al., "Comparison of Infrastructure-and Onboard Vehicle-Based Sensor Systems in Measuring Operational Safety Assessment (OSA) Metrics," SAE Technical Paper 2023-01-0858 (2023), https://doi.org/10.4271/2023-01-0858.
9. Zhao, J., Xu, H., Liu, H., Wu, J. et al., "Detection and Tracking of Pedestrians and Vehicles Using Roadside LiDAR Sensors," *Transportation Research Part C: Emerging Technologies* 100 (2019): 68-87.
10. Kumar Rath, P., Ramirez-Serrano, A., and Kumar Pratihar, D., "Real-Time Moving Object Detection and Removal from 3D Pointcloud Data for Humanoid Navigation in Dense GPS-Denied Environments," *Engineering Reports* 2, no. 12 (2020): e12275.
11. Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., "Vision Meets Robotics: The Kitti Dataset," *The International Journal of Robotics Research* 32, no. 11 (2013): 1231-1237.
12. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A. et al., "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2446-2454, 2020.
13. Caesar, H., Bankiti, V., Lang, A.H., Vora, S. et al., "Nuscenes: A Multimodal Dataset for Autonomous Driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11621-11631, 2020.
14. Zamanakos, G., Tsochatzidis, L., Amanatiadis, A., and Pratikakis, I., "A Comprehensive Survey of LIDAR-Based 3D Object Detection Methods with Deep Learning for Autonomous Driving," *Computers & Graphics* 99 (2021): 153-181.
15. Qi, C.R., Su, H., Mo, K., and Guibas, L.J., "Pointnet: Deep Learning on Point Sets for 3d Classification and Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652-660, 2017.
16. Qian, R., Lai, X., and Li, X., "3D Object Detection for Autonomous Driving: A Survey," *Pattern Recognition* 130 (2022): 108796.
17. Zhu, Z., Zhang, Y., Chen, H., Dong, Y. et al., "Understanding the Robustness of 3D Object Detection With Bird's-Eye-View Representations in Autonomous Driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21600-21610, 2023.
18. Lang, A.H., Vora, S., Caesar, H., Zhou, L., Yang, J., and Beijbom, O., "Pointpillars: Fast Encoders for Object Detection from Point Clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12697-12705, 2019.
19. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A.C., "SSD: Single Shot Multibox Detector," in *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, 21-37, Springer, 2016.
20. Simon, M., Milz, S., Amende, K., and Gross, H.-M., "Complex-YOLO: Real-time 3D Object Detection on Point Clouds," 2018.
21. Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., and Li, H., "Pv-Rcnn: Point-Voxel Feature Set Abstraction for 3d Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10529-10538, 2020.
22. Salih, Y., and Malik, A.S., "3D Object Tracking Using Three Kalman Filters," in *2011 IEEE Symposium on Computers & Informatics*, 501-505, IEEE, 2011.
23. Bewley, A., Ge, Z., Ott, L., Ramos, F. et al., "Simple Online and Realtime Tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, 3464-3468, 2016.
24. Luo, W., Xing, J., Milan, A., Zhang, X. et al., "Multiple Object Tracking: A Literature Review," *Artificial Intelligence* 293 (2021): 103448.
25. Wang, C., Xie, Y., Huang, H., and Liu, P., "A Review of Surrogate Safety Measures and Their Applications in Connected and Automated Vehicles Safety Modeling," *Accident Analysis & Prevention* 157 (2021): 106157.
26. Elli, M.S., Wishart, J., Como, S., Dhakshinamoorthy, S. et al., "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated Vehicles in Simulation,"

SAE Technical Paper 2021-01-0868 (2021), https://doi.org/10.4271/2021-01-0868.

27. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. et al., "CARLA: An Open Urban Driving Simulator," in *Conference on Robot Learning*, 1-16, PMLR, 2017.
28. Wu, J., Xu, H., Zheng, Y., and Tian, Z., "A Novel Method of Vehicle-Pedestrian Near-Crash Identification with Roadside LiDAR Data," *Accident Analysis & Prevention* 121 (2018): 238-249.
29. Bonela, S.R. and Kadali, B.R., "Review of Traffic Safety Evaluation at T-Intersections Using Surrogate Safety Measures in Developing Countries Context," *IATSS Research* 46, no. 3 (2022): 307-321.
30. MMDetection3D Contributors, "MMDetection3D: OpenMMLab Next-Generation Platform for General 3D Object Detection," 2020, https://github.com/open-mmlab/mmdetection3d.
31. Wang, Q., Chen, Y., Pang, Z., Wang, N. et al., "Immortal Tracker: Tracklet Never Dies," arXiv preprint arXiv:2111.13672, 2021.
32. Levin, M.W. and Rey, D., "Conflict-Point Formulation of Intersection Control for Autonomous Vehicles," *Transportation Research Part C: Emerging Technologies* 85 (2017): 528-547.
33. Bentley and Ottmann, "Algorithms for Reporting and Counting Geometric Intersections," *IEEE Transactions on Computers* 100, no. 9 (1979): 643-647.
34. Powers, D.M., "Evaluation: From Precision, Recall and f-Measure to Roc, Informedness, Markedness and Correlation," arXiv preprint arXiv:2010.16061, 2020.

Contact Information

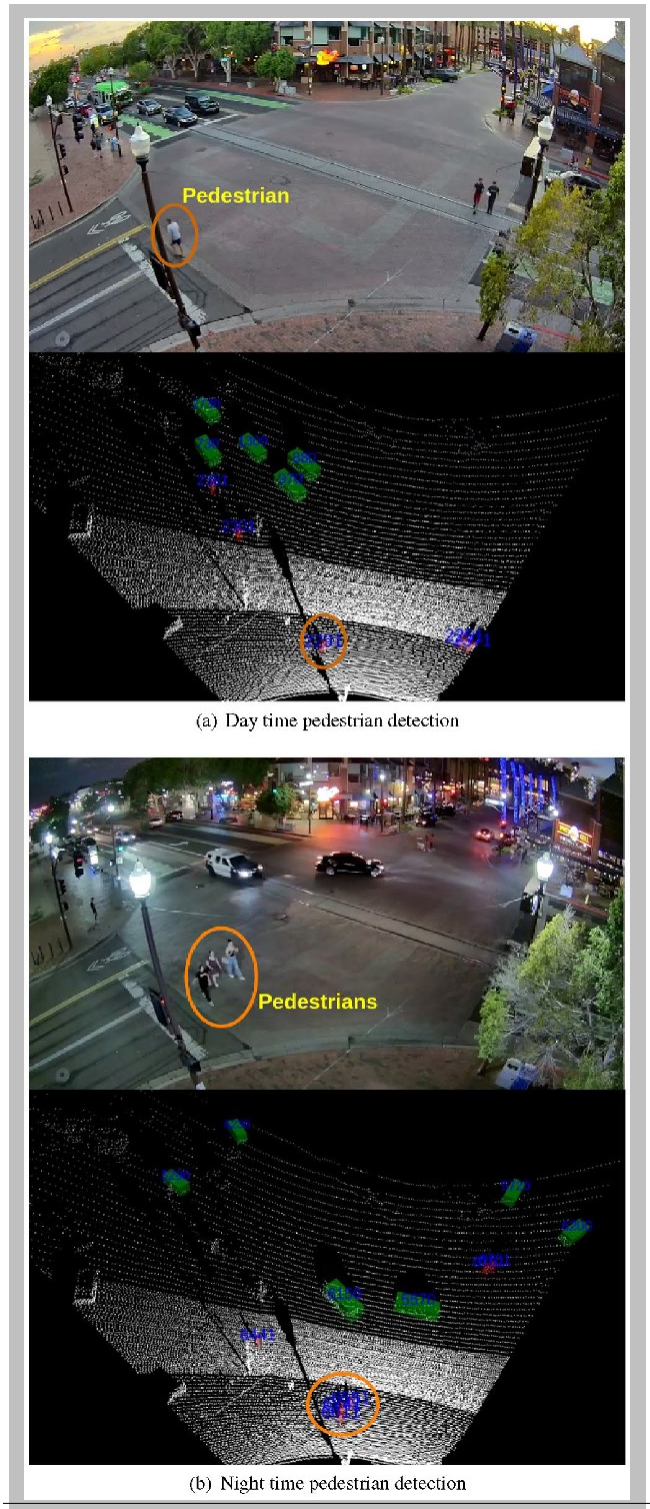
Hongbin Yu, Ph. D.
yuhb@asu.edu
Director, NSF Center for Efficient Vehicles and Sustainable Transportation Systems (EVSTS) https://evsts.asu.edu
Associate Professor, School of Electrical, Computer & Energy
Engineering (ECEE)
Arizona State University
PO Box 875706
Tempe, AZ 80287

Acknowledgments

Authors are grateful to the Institute of Automated Mobility (IAM) for generously funding this work and Maricopa County Department of Transportation (MCDOT) for enabling infra- structural resources required for this work. HY, and PKR acknowledge the support from National Science Foundation (grant nos. 1624842, 1839485, 1933742, 2137295) Industry University Cooperative Research Center (IUCRC) for Efficient Vehicles and Sustainable Transportation Systems (EVSTS) at Arizona State University.

Appendix

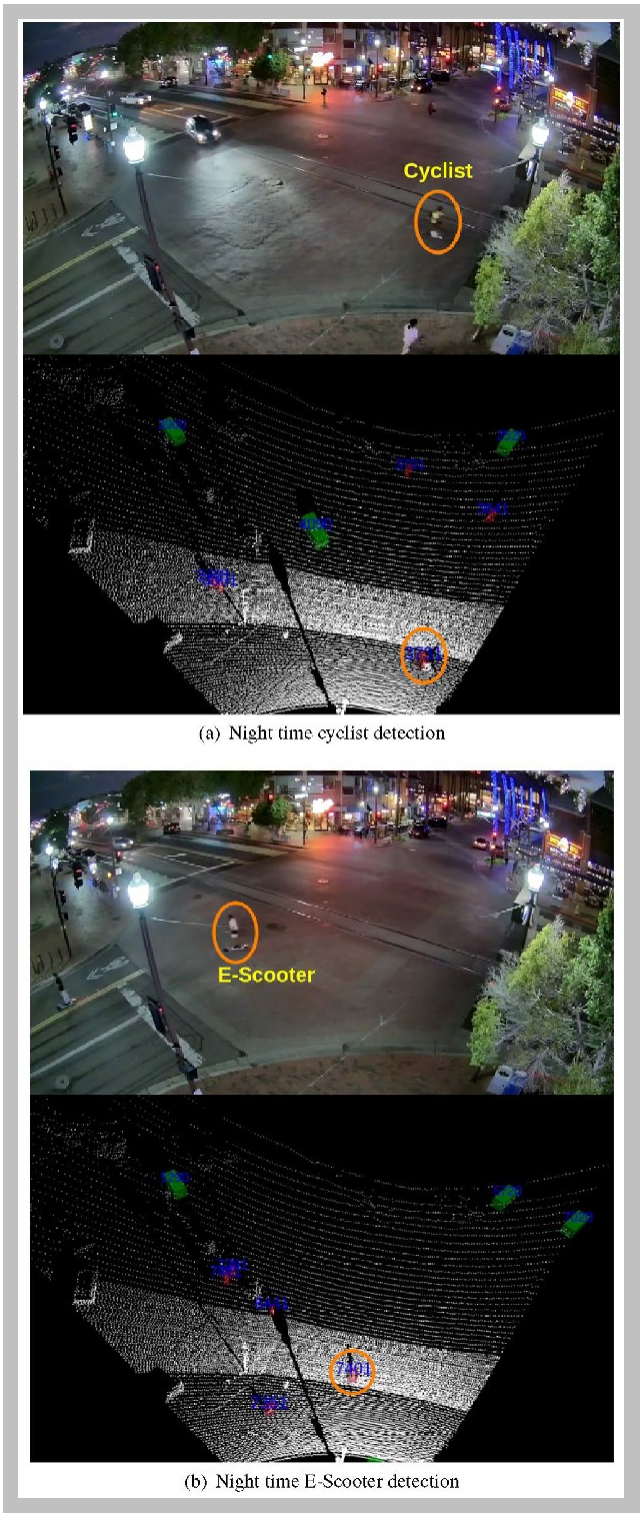
FIGURE 7 Pedestrian detection results (best viewed when zoomed in). Note that distant pedestrians are almost indistinguishable in the camera data; however, they have been consistently and accurately detected in the LiDAR data.



(a) Day time pedestrian detection

(b) Night time pedestrian detection

FIGURE 8 Cyclist and micro-mobility vehicle detection results (best viewed when zoomed in).



(a) Night time cyclist detection

(b) Night time E-Scooter detection

ISSN 0148-7191