A Multi-Scene Roadside Lidar Benchmark towards Digital Twins of Road Intersections

Miao Tang[1], Dianyu Yu[1], Peiguang Li[1], Chengwen Song[1], Pu Zhao[1], Wen Xiao[1,2], Nengcheng Chen[2]

[1] China University of Geosciences, School of Geography and Information Engineering, 430074 Wuhan, China
- (miao.tang, yudianyu, lpg, cwsong, pu.zhao, wen.xiao)@cug.edu.cn
[2] China University of Geosciences, National Engineering Research Center of Geographic Information System, 430074 Wuhan, China
- chennengcheng@cug.edu.cn

Keywords: Intelligent Intersection, Dataset, 3D Object Detection and Tracking, Dual Lidar.

Abstract

In recent years, the evolution of digital twin technology has paved the way for the construction of intelligent holographic intersections. This can be facilitated by utilizing precise point clouds from roadside lidar. With its capability of real-time monitoring, lidar plays a crucial role in enhancing intersection perception, enabling precise detection and tracking of road objects, as well as providing accurate speed estimates. Despite the introduction of few roadside lidar datasets aimed at enhancing supervised learning algorithms, their applicability to intelligent intersection monitoring remains limited. To address this, this paper presents an Intelligent Intersections (Int2sec) dataset, which exhibits several salient features: 1) it encompasses a broad array of urban intersection scenarios accompanied by a substantial quantity of object annotations; 2) the deployment of dual lidar stations facilitates a thorough scanning of scenes, thereby ensuring expansive scene coverage and mitigating the mutual occlusion phenomenon amongst objects; and 3) the dataset not only catalogues the coordinates, dimensions, and orientations of objects but also encompasses additional attributes such as tracking IDs and real-time motion statuses. Furthermore, the paper evaluates the efficacy of various prominent benchmarking networks, providing a critical analysis and prospective for future research.

1. Introduction

With urbanization rapidly advancing worldwide, road intersections face escalating challenges, including increased complexity in traffic conditions, frequent congestion, and heightened accident rates. The coexistence of motor vehicles and non-motorized vehicles at intersections often leads to intricate traffic signal configurations and markings, exacerbating congestion and safety risks. European Road Safety Observatory (European Commission, 2018) reports in 2018 that 39.6% of traffic accidents in European cities occur at intersections, underscoring the pressing need for effective interventions in intersection management.

Efforts to address these challenges have seen the emergence of intelligent holographic intersections (Kušić et al., 2023; Thonhofer et al., 2023), which leverage multiple sensors for real-time detection of vehicles and pedestrians, enabling precise monitoring of intersection activity. Additionally, digital twin technology, as highlighted by Zio and Miqueles (2024), offers promising avenues for enhancing risk assessment and emergency management at intersections. By integrating real-time data from holographic intersections with digital twin technology, it becomes feasible to achieve accurate perception and prediction of intersection traffic conditions, mitigating visibility blind spots for both vehicles and pedestrians and reducing the occurrence of accidents.

Lidar, renowned for its precise 3D perception capability (Li and Ibanez-Guzman, 2020), stands as a key technology in the development of digital twins to facilitate intelligent intersections. By leveraging its advantages of precise sensing and immunity to lighting conditions (Zhang et al., 2022a), lidar is pivotal for continuous data acquisition and intersection monitoring, crucial components of digital twin frameworks. Bai et al. (2022) demonstrated that employing roadside lidar deployed at intersections can accurately capture the position and velocity of objects, thereby enhancing the precision of the digital representation of the intersection. By strategically optimizing the placement of roadside lidar units at the intersection, Jiang et al. (2023) enables them to achieve more comprehensive monitoring of the intersection. Therefore, roadside lidar can serves as a robust and accurate perception component integrated into intelligent intersection systems, contributing to the fidelity and effectiveness of digital twins in urban traffic management.

In light of the diminished detection efficacy observed in a single roadside lidar when confronted with object occlusion (Wu et al., 2020), the integration of multiple roadside lidars emerges as a viable strategy to enable comprehensive and precise perception for intersection monitoring. Placing several lidar units at distinct locations covering road intersections significantly alleviates the detrimental impact of object occlusions on performance (Wang et al., 2021; Busch et al., 2022). This strategic deployment not only addresses the shortcomings inherent in relying solely on a single lidar unit but also ensures comprehensive coverage crucial for intersection safety and traffic management.

Currently, several roadside lidar datasets have emerged, such as LUMPI (Busch et al., 2022), IPS300+ (Wang et al., 2021), and TUMTraf (Zimmer et al., 2023). While LUMPI and IPS300+ datasets deploy multiple lidars at road intersections, they only support a single object detection or object tracking task. In the TUMTraf dataset, the placement of two lidar sensors side-by-side does not allow for monitoring the intersection from multiple angles, thus failing to effectively mitigate occlusion issues between objects. The latest HoloVIC (Ma et al., 2024) supports object detection and tracking tasks by placing multiple lidars and cameras aligned to each other without specifying the lidar data range. The number and type of scenarios are also limited. To address the shortcomings of existing roadside multi-lidar datasets, we have created and benchmarked a new dataset

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-341-2024

for intelligent intersections, named Int2sec[1], which is a large-scale, multi-task, multi-lidar dataset for roadside real-time perception. Our main contributions in this work are as follows:

- A new Int2sec roadside lidar dataset for both single- and dual-sensor 3D object detection and tracking learning tasks is released, with annotations of 10 different intersection scenes and distances up to 220 m.
- To showcase the challenge of the data, a benchmark analysis of point-based, voxel-based, and point-voxel fusion-based methods for multi-class object detection are conducted, based on which multi-object tracking are also benchmarked using the tracking-by-detection strategy.

2. Related work

2.1 Perception Based On Roadside Lidar

Roadside lidar has emerged as a useful tool for enhancing perception levels at road intersections (Sun et al., 2022). Existing research demonstrates various methods and frameworks utilizing roadside lidar for precise detection and tracking of pedestrians and vehicles, as well as for vehicle classification and speed estimation. For instance, Zhao et al. (2019) utilized roadside lidar to enhance intersection perception levels through precise detection and tracking of pedestrians and vehicles. Similarly, Zhang et al. (2020) proposed a tracking framework based on roadside lidar, enabling accurate vehicle speed estimates and enhancing perception of vehicle speeds at intersections or on highways. Moreover, Wu et al. (2019) introduced a method for vehicle classification using roadside lidar technology. Additionally, recent advancements include the construction of a deep learning-based real-world object perception platform using roadside lidar, as demonstrated by Bai et al. (2023). These studies collectively underscore the significance of roadside lidar in advancing perception capabilities at intersections.

2.2 Existing Roadside Lidar Datasets

In recent years, the number of vehicle-based lidar datasets has grown rapidly, effectively contributing to the development of autonomous driving technology. Compared to autonomous driving datasets, the number of roadside lidar datasets is still insufficient. DAIR-V2X (Yu et al., 2022) dataset focuses on the vehicle-infrastructure cooperative task, and does not implement the multiple roadside lidars perception. Datasets such as TUMTraf (Zimmer et al., 2023), IPS300+ (Wang et al., 2021), and LUMPI (Busch et al., 2022) all use two or more roadside lidars to collect the data. The TUMTraf has two lidars installed side-by-side on the road gantries for collection. IPS300+ (Wang et al., 2021) and LUMPI (Busch et al., 2022) both chose to use intersections as their acquisition scenarios. However, they only collected data in a single scenario, without considering the diversity of traffic intersections. Furthermore, IPS300+ dataset has less representation of the truck and bus categories. LUMPI dataset employs five random combinations of placed lidars, the initial labeling results are obtained through traditional operations such as background filtering, but the labeling information and benchmark results for this dataset are not publicly available. HoloVIC (Ma et al., 2024) present a new data supporting for infrastructure-based roadside perception and vehicle-infrastructure cooperative perception. It benchmarks detection and tracking tasks using a combination of multiple cameras and two lidars. It has four intersection and one straight road scenes. While each dataset mentioned above offers unique advantages, they also have limitations. This highlights the need for further development and exploration of roadside lidar datasets to address these limitations.

[1] https://github.com/Geo3DSmart/Int2sec

2.3 Lidar-based 3D Object Detection

3D object detection predicts the 3D bounding box from a sparse point cloud containing information such as the center coordinates, dimensions, and angles of the detected objects. Some representative methods utilize different feature expressions for 3D object detection. IA-SSD (Zhang et al., 2022b) directly learns fine-grained features of points and utilizes semantic information of the points for key point sampling, which leads to better single-stage 3D object detection. While PV-RCNN (Shi et al., 2020) utilizes the point and voxel fusion strategy for two-stage 3D object detection, which are the generation of the initial detection frame in the first stage and the key point feature extraction and refinement operation of the initial detection frame in the second stage. CenterPoint (Yin et al., 2021) method represents the 3D object as a centroid and uses a heatmap detector to detect the center of the object, then it uses the centroid features to regress the 3D bounding box of the object. In this paper, we have selected some of the above methods to be trained on our dataset as the baseline.

2.4 Lidar-based 3D Object Tracking

Multi-objective tracking (MOT) plays an important role in autonomous driving and digital twins. Existing 3D MOT approaches are mainly based on the “tracking by detection” strategy, which consists of trajectory prediction, data association, trajectory state update, and trajectory life management (Weng et al., 2020). Results of the tracking depend on powerful detectors (Yin et al., 2021; Wu et al., 2022) and effective data association strategies. For the data associate process, some methods utilize deterministic trajectories and detection results for association. AB3DMOT (Weng et al., 2020) utilizes 3D Kalman filtering and Hungarian algorithms for trajectory prediction and data association based on 3D detection results from lidar point cloud data. PC-TCNN (Wu et al., 2022) utilizes the spatial-temporal features of tracklet proposal for refinement. Some methods consider the uncertainty of trajectory and detection results, and UG3DMOT (He et al., 2023) realizes data correlation between detection and trajectory results based on Jensen-Shannon divergence. In this paper, we provide a 3D multi-object tracking task based on a 3D object detection benchmark models.

3. Int2sec Dataset

Table 1 has listed related roadside lidar datasets and Int2sec to ease comparison. Although there are some data targeting at the emerging need of infrastructure-based road monitoring, they have not be able to cover a variety of scenes where lidars are placed accordingly. In contrast, the Int2sec dataset includes up to 10 scenes and 8 classes, and supports both 3D object detection and object tracking tasks at a long distance. In addition, labels for single lidar data (Int2sec-S) and that for the integrated dual-lidar data (Int2sec-D) are both annotated to allow for a flexible usage of the data.

Dataset	Year	Beam	Frame	3D boxes	Tracks	Range	Classes	Scenes	Task	
									Detection	MOT
DAIR-V2X-I	2021	300	71k	1.2M	-	100	10	-	✓	✗
IPS300+	2022	80	14.1k	4.5M	-	150	7	1	✓	✗
LUMPI	2022	16/64	90k	-	-	200	6	1	✗	✓
TUMTraf	2023	64	5.3k	71.9k	506	140	10	1	✓	✓
HoloVIC-I*	2024	150	63k	1.3M	-	-	3	5	✓	✓
Int2sec-S (Our)	2024	80	12k	338k	1988	220	8	10	✓	✓
Int2sec-D (Our)	2024	80	6k	198k	1034	220	8	10	✓	✓

Note: * Represents the infrastructure part of HoloVIC, four of scenes are at intersections, and one is along a straight road.

Table 1. Information of our Int2sec dataset and related roadside lidar datasets.

3.1 Data acquisition

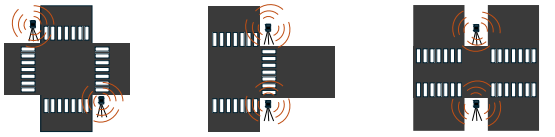
A roadside data acquisition device was assembled, comprising an 80-beam lidar, a Jetson Orin Nano minicomputer, a mobile power source, a tripod, a GNSS receiver, and a monitor. Detailed specifications of each component are outlined in Table 2. By leveraging lidar and the diagonal arrangement of the dual acquisition device at intersections (Figure 1), our dataset achieves higher point density and an extended sensing range. This arrangement optimizes data collection coverage and enhances the dataset's utility for intersection monitoring. Both data collection devices utilize GNSS timing and record timestamps into the collected point clouds, which are used to synchronize the point clouds during the process of integrating multiple lidar datasets.

The dataset was acquired in an urban area of China, encompassing 19 point cloud sequences from 10 intersections with diverse characteristics. Covering a total area of 49 square kilometers, including 4 square kilometers of campus roads and 45 square kilometers of urban roads, it ensures a variety of traffic scenarios and conditions. The dataset spans three weather conditions (sunny, rainy and cloudy), two lighting scenarios (daytime and nighttime) and four-time slots (morning, noon, afternoon and evening) to cover different traffic conditions, ensuring data diversity and representativeness.

To comply with local laws and regulations, the dataset will not include any location information, such as road names, GNSS coordinates, or map data. Each sequence folder contains metadata to describe scene category, weather condition, detection IDs and track IDs.

Equipment	**Detail**
Lidar Sensor	80 beams, 10Hz frame rate, 905 nm laser wavelength 360°horizontal FOV, $-40°$ to $0°$ vertical FOV, $\leq$ 230 m range, GNSS time transfer
Computing Unit	8 GB 128-bit LPDDR4x , 6 core NVIDIA Car-mel 64-bit CPU

Table 2. Key specifications of roadside data collection units.



(a) Crossroad (b) T-junction (c) H-intersection

Figure 1. Dual-lidar placement at three typical junctions.

3.2 Registration

The registration process of the two lidar point clouds from Int2sec-D is divided into two main phases: coarse and fine alignment. In the coarse alignment phase, using the feature of the elevation formed by the roadside and the road, first, the road surface in the scene is extracted and the elevation points in the point cloud are projected onto the road surface, which enhances the features of the road edge lines. Next, the projected point cloud is processed using the RANSAC (Fischler and Bolles, 1981) algorithm to identify the main line features in the scene. Finally, these line features are used to perform a preliminary alignment, a step that involves picking out pairs of straight lines in the source and object point clouds whose angle exceeds a certain threshold, and applying the RANSAC algorithm to remove those incorrect matches and obtain the correct matching pairs. The purpose of this step is to provide a good initial orientation for the fine alignment. In the fine alignment stage, the classical ICP algorithm is used to further optimize the alignment results, which ensures that the point cloud datasets can be accurately aligned with each other (Figure 2).

3.3 Data annotation

From the 11,076 frames of lidar point cloud data, we identified eight types of objects commonly found in road scenes: cars (including small cars and SUVs), vans (including minivans and box trucks), pedestrians, cyclists, motorcyclists, tricycles, trucks (including lorries, vans, and small construction trucks), and buses (including buses, coaches, fire trucks, and large construction trucks). The SUSTechPOINTS open-source annotation tool (Li et al., 2020) was utilized for data annotation. It enables efficient annotation of continuous sequences of point clouds semi-automatically, allowing objects to be tracked across all frames of the point cloud sequence, followed by

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-341-2024 | © Author(s) 2024. CC BY 4.0 License.

manual fine-tuning. Furthermore, to minimize human errors, the data underwent multiple cross-checks.

All objects containing five or more points were annotated, reaching a maximum range of 220 meters. The object annotation includes the 3D bounding box, object category, instance ID, and motion state, as detailed in Table 3. Each 3D bounding box comprises seven parameters: the coordinates of the object's geometric center in the lidar coordinate system, its length, width, height, and rotation angle. During the creation of the dual lidar dataset labels, the distances between the same object and the two individual lidars are calculated, and the label from the lidar closer to the object is adopted as the reference, followed by manual fine-tuning. Additionally, our dataset annotates the motion status of objects, distinguishing between global static, local static, and moving states, indicating its potential to support the moving object segmentation (MOS) task in the future. Figure 2 visualizes the dataset annotation.

Parameter	Detail
center	Bounding box location in meters in the global frame: center x, center y, center z.
size	Bounding box size in meters: width, length, height.
rotation	Bounding box orientation in radians in the global frame: angle.
id	ID that identify the same object in a sequence.
state	Identify the state of motion of an object.

Table 3. Annotation information.

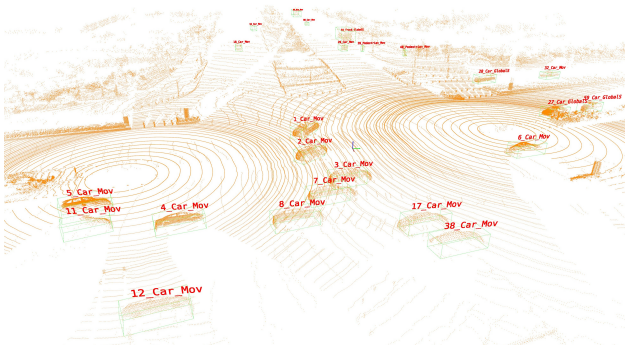


Figure 2. Visualisation of data annotation.

3.4 Data statistics

The number of object instances of the Int2sec-S and Int2sec-D are depicted in Figures 3, exhibiting similar distributions in terms of numbers. The dataset comprises 1988 single lidar object trajectories spanning a total length of 60.1 km. The average trajectory length is 30.23 m, with the maximum trajectory extending to 178 m. Figure 4 and Figure 5 illustrates the distribution of labels across eight categories in Int2sec-S and Int2sec-D respectively. The primary road users are cars and pedestrians, with motorcycles following closely behind. This trend can be attributed to the inclusion of scenarios encompassing campuses and suburbs. The counts for bus and bicycle categories are slightly lower in comparison.

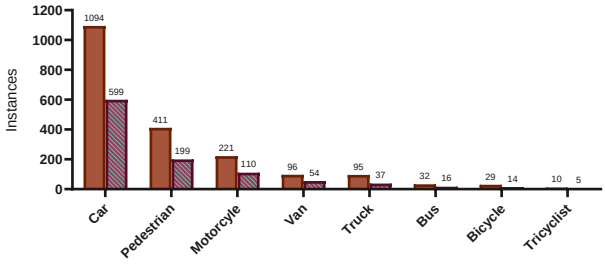


Figure 3. Number of instances of Int2sec-S (left, empty column) and Int2sec-D (right, dashed column).

Figure 6 and Figure 7 illustrates the distribution of various classes across different distances in Int2sec-S and Int2sec-D respectively, with annotated objects reaching distances as far as 220 meters. Across all categories, a consistent trend emerges where the frequency of occurrences decreases with distance, with greater numbers observed at closer proximities. Notably, classes such as cars, pedestrians, and motorcycles exhibit their highest frequencies within the 30 to 60 m range, whereas trucks demonstrate peak occurrences between 60 to 90 m. Furthermore, our annotations encompass eight categories beyond the 120 m mark, thus introducing a novel challenge to long-distance 3D object perception.

4. Multi-task Benchmarking

4.1 Benchmarking for 3D Object Detection

4.1.1 Baseline Methods In this paper, point-based method, voxel-based method and point-voxel fusion-based method are all tested to benchmark their performance on the data. Three representative methods from each of the category are IA-SSD (Zhang et al., 2022b), CenterPoint (Yin et al., 2021) and PV-RCNN (Shi et al., 2020), respectively.

During the training process, the size of the voxel is set to [0.2,0.2, 0.2], the detection distance of x is from -100m to 100m, the detection distance of y is from -75m to 75m, and z is from -5m to 3 m. The other training parameters are kept the same as the original model, and six Nvidia RTX3090 GPUs are used for training. The results of the experiments on the test set and the training set are shown in Table 4 and Table 5.

4.1.2 Evaluation Evaluation metrics are crucial for characterizing different 3D object detection methods. This paper adopts the same evaluation metric of AP3D based on 3D IoU with orientation as used in the ONCE dataset (Mao et al., 2021) for the 3D object detection benchmark. The predictions are initially sorted based on their prediction scores. Prediction boxes with lower 3D IoU values compared to all ground truth boxes of the same category are labeled as false positives. Additionally, if the orientation of a prediction result falls outside the range of $\pm 90^{\circ}$ from the orientation of the matched ground truth box, it is also classified as a false positive. The remaining matched predictions are considered true positives. Subsequently, we define 50 score thresholds ranging from a recall rate of 0.02 to 1.00 in steps of 0.02 and calculate the corresponding precision rates to construct the precision-recall curve $p(r)$. The formula for the direction-aware AP3D is:

$$AP_{3D} = 100 \int_{0}^{1} max\{p(rc'|rc' \geq rc)\}drc. \quad (1)$$

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-341-2024 | © Author(s) 2024. CC BY 4.0 License.

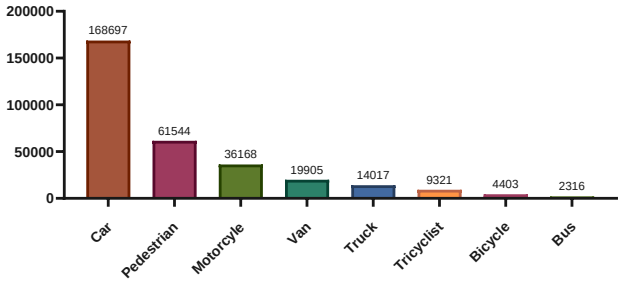


Figure 4. Number of objects in different categories in Int2sec-S.

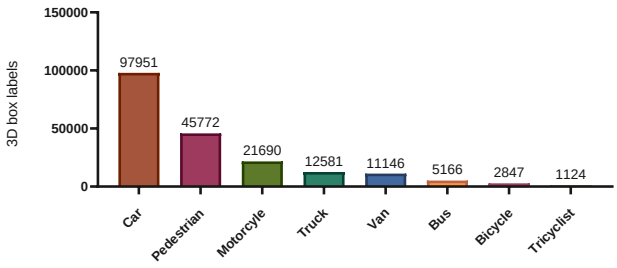


Figure 5. Number of objects in different categories in Int2sec-D.

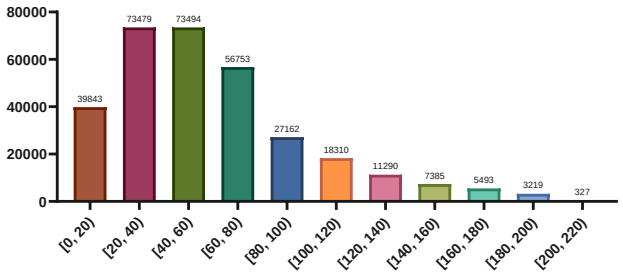


Figure 6. Distribution of objects at different distances in Int2sec-S.

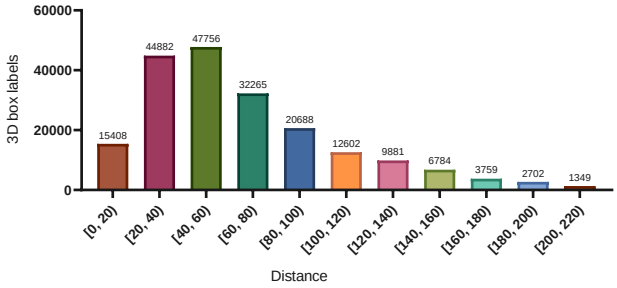


Figure 7. Distribution of objects at different distances in Int2sec-D.

Mean AP (mAP) is thus obtained by averaging the scores of categories detected. The evaluation metric divides each category into three parts by distance: 0-30 meters, 30-50 meters, and 50 meters and beyond, and calculates the mAP separately to examine the performance of the algorithms over different distance ranges.

To avoid duplication, only the results on Int2sec-D are reported. Table 4 and Table 5 show the 3D object detection model results on the Int2sec-D dataset. The three detection methods have good performance on close range 0-30 m, but the highest is only 27.84% overall AP on pedestrians and cyclists, and the overall effect of detection accuracy at long range is not high, and PV-RCNN is the lowest at 2.91% AP in the interval of 50-100m. Figure 8 shows the results of the IA-SSD detection.

4.2 Benchmarking for 3D Object Tracking

4.2.1 Baseline Methods The common strategy of tracking-by-detection relies on the results of detection, thus the 3D multi-object tracking benchmark is using the 3D object detection results as input. In this paper, PC-TCNN (Wu et al., 2022) is adopted as the tracking algorithm to associate the same object in the sequence and generate the trajectory information of this object in the scene. The input detection results of IA-SSD, PV-RCNN, and CenterPoint are evaluated respectively.

4.2.2 Evaluation Given a point cloud sequence and the corresponding detection results, the goal of 3D object tracking is to correlate the same object in different frames to obtain a tracking ID that uniquely identifies each object in the sequence. We use AMOTA (Weng et al., 2020), AMOTP, MOTA, and IDS metrics to measure the tracking task.

$$AMOTA = \frac{1}{N} \sum_{i \in \{\frac{1}{N}, \frac{2}{N}, \ldots 1\}} (1 - \frac{fn_i + fp_i + ids_i}{num_{gt}}) \quad (2)$$

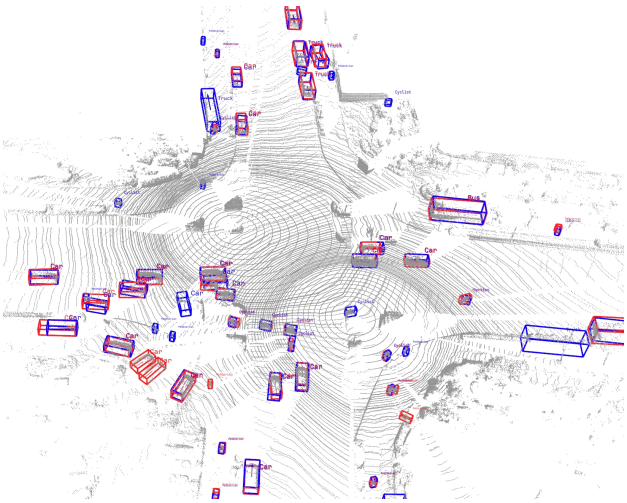


Figure 8. The detection results of IA-SSD. The red boxes represent the ground truth while the blue ones show the predicted bounding boxes

where fn_i,fp_i,ids_i are the number of false negatives, false positives, and tracking id changes at recall value r. num_{gt} is the number of all truth objects.

Similarly, only the results on Int2sec-D are reported. Table 6 and Tabel 7 shows that different methods achieved different results in different metrics, with IA-SSD achieving the best 65.6% AMOTA and 66.8% MOTA and CenterPoint scoring the best 79.7% AMOTP and 4 IDS. Figure 9 shows the results of IA-SSD based tracking visualization.

5. Discussion

Due to the rotational scanning mechanism of the lidar device itself, there is a slight time difference between the synchron-

Method	Vehicle (AP@50)(%)				Pedestrian (AP@50)(%)				Cyclist (AP@50)(%)				mAP(%)
	Overall	0–30 m	30–50 m	50 -100m	Overall	0–30 m	30–50 m	50 -100m	Overall	0–30 m	30–50 m	50 -100m	
IA-SSD	28.34	78.22	31.61	14.67	17.94	24.53	13.90	11.66	20.11	31.30	11.30	13.33	22.13
PV-RCNN	26.84	75.64	30.12	13.40	16.07	32.38	8.15	10.94	21.84	50.89	15.40	2.91	21.58
CenterPoint	22.05	67.23	21.58	12.29	27.84	44.38	13.91	25.64	14.45	30.11	14.34	2.32	21.45

Table 4. Evaluation results of our model and baseline models on the Int2sec-D test set.

Method	Vehicle (AP@50)(%)				Pedestrian (AP@50)(%)				Cyclist (AP@50)(%)				mAP(%)
	Overall	0–30 m	30–50 m	50 -100	Overall	0–30 m	30–50 m	50 -100m	Overall	0–30 m	30–50 m	50 -100m	
IA-SSD	34.38	80.95	39.43	16.59	14.66	27.28	15.43	4.66	24.82	39.62	23.31	11.83	24.62
PV-RCNN	31.17	74.18	34.77	15.77	8.92	13.88	10.72	4.73	24.84	42.30	20.91	11.76	21.64
CenterPoint	26.93	70.16	20.62	17.86	12.36	23.03	11.85	4.29	25.31	41.80	22.64	12.34	21.54

Table 5. Evaluation results of our model and baseline models on the Int2sec-D validation set.

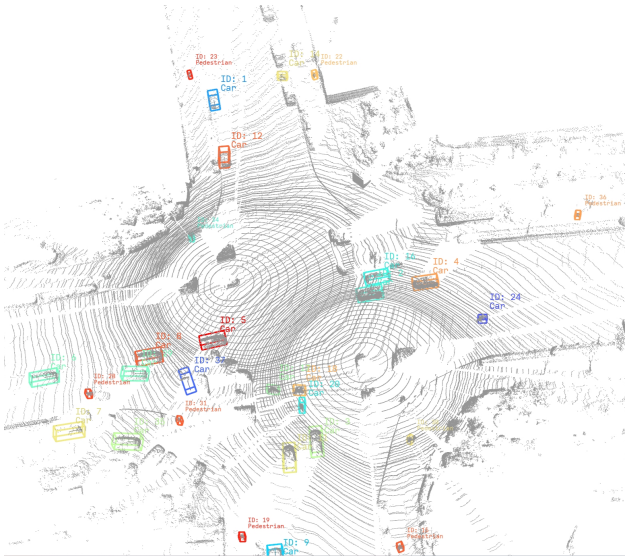


Figure 9. The tracking results are based on IA-SSD and PC-TCNN.

Tracking	Detection	AMOTA%	AMOTP%	MOTA%	IDS
PC-TCNN	IA-SSD	65.6	66.1	66.8	94
	PV-RCNN	54.7	71.4	56.7	6
	CenterPoint	43.8	79.7	46.4	4

Table 6. Tracking results of baseline models on the Int2sec-D test set.

Tracking	Detection	AMOTA%	AMOTP%	MOTA%	IDS
PC-TCNN	IA-SSD	56.4	82.5	58.3	48
	PV-RCNN	50.9	80.1	53.6	1
	CenterPoint	41.3	90.0	43.5	2

Table 7. Tracking results of baseline models on the Int2sec-D validation set.

ized dual lidar frames, which may lead to ghosting phenomenon when the vehicle speed is too fast in the road scene. For this phenomenon, the labeling of the objects is taking the closer lidar as the reference and the annotations are manually refined. In addition, this paper also provides individual lidar labeling to facilitate the needs for feature fusion detection from the two lidars.

Table 4 and Table 5 indicate that the object detection models perform better at close range compared to the distant range on the Int2sec-D dataset. This is attributed to the diminishing geometric features of objects with increasing distance, thereby compromising the accuracy of detecting distant objects. In addition, some object instances are consistently situated in the distant range across the entire scene sequence. Although the detection models exhibit satisfactory performance in detecting large objects such as vehicles in the close range, they struggle with smaller objects like pedestrians and cyclists due to their less pronounced geometric features, making them prone to background confusion. These factors collectively contribute to the overall subpar detection accuracy of the object detection models on the Int2sec-D dataset. This demonstrates the challenge of our data and the difficulty in real-world applications. Despite the underperformance of the object detection models, Table 6 and Table 7 demonstrate that the object tracking model performs well. This is attributed to the majority of object instances being concentrated in the close range within a scene sequence, with a significant proportion belonging to the category of vehicles. Consequently, the performance of close-range tracking significantly influences the overall tracking results.

For moving objects at high speeds, long-range object detection and tracking is required to ensure adequate braking distance. The number of point clouds of long-distance objects is low due to the small number of lidar scanning beams and the occlusion of objects, it is difficult to detect 3D objects at a long distance. The longest detection distance of 150 m was achieved on the IPS300+ dataset, but only 600 frames of label data were publicly available on this dataset and tracking labels were not provided. Our proposed Int2sec is labeled up to 220 m according to the number of points on objects and contains both detection and tracking labels, providing a database for subsequent exploration of long-range detection and tracking tasks.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-341-2024 | © Author(s) 2024. CC BY 4.0 License.

6. Conclusion

This paper proposes the Int2sec roadside lidar point cloud dataset, including annotations of both single and dual lidars. Compared with existing roadside lidar datasets, Int2sec covers 10 intersection scenarios and provides multi-category labeling ranging up to 220 m. Further benchmark analyses of dual-lidar 3D object detection and tracking tasks are conducted using representative networks in different categories. Building upon existing lidar datasets, this roadside lidar dataset offers a more diverse range of intersection scenarios for roadside lidar perception and digital twin holographic intersection studies. This enhancement presents new opportunities and potential for in-depth investigation within this research field.

Acknowledgements

The authors would like to thank all the reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China (Grant 42201485).

References

Bai, Z., Nayak, S. P., Zhao, X., Wu, G., Barth, M. J., Qi, X., Liu, Y., Sisbot, E. A., Oguchi, K., 2023. Cyber Mobility Mirror: A Deep Learning-Based Real-World Object Perception Platform Using Roadside LiDAR. *IEEE Transactions on Intelligent Transportation Systems*, 24(9), 9476–9489.

Bai, Z., Wu, G., Qi, X., Liu, Y., Oguchi, K., Barth, M. J., 2022. Infrastructure-based object detection and tracking for cooperative driving automation: A survey. *IEEE Intelligent Vehicles Symposium (IV)*, 1366–1373.

Busch, S., Koetsier, C., Axmann, J., Brenner, C., 2022. LUMPI: The Leibniz University Multi-Perspective Intersection Dataset. *IEEE Intelligent Vehicles Symposium (IV)*, 1127–1134.

European Commission, 2018. Annual accident report 2018. `https://road-safety.transport.ec.europa.eu` (1 May 2024).

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

He, J., yan Fu, C., Wang, X., 2023. 3D Multi-Object Tracking Based on Uncertainty-Guided Data Association. *arXiv preprint arXiv:2303.01786*.

Jiang, W., Xiang, H., Cai, X., Xu, R., Ma, J., Li, Y., Lee, G. H., Liu, S., 2023. Optimizing the placement of roadside lidars for autonomous driving. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 18381–18390.

Kušić, K., Schumann, R., Ivanjko, E., 2023. A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Advanced Engineering Informatics*, 55.

Li, E., Wang, S., Li, C., Li, D., Wu, X., Hao, Q., 2020. Sustech points: A portable 3d point cloud interactive annotation platform system. *IEEE Intelligent Vehicles Symposium (IV)*, 1108–1115.

Li, Y., Ibanez-Guzman, J., 2020. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4), 50–61.

Ma, C., Qiao, L., Zhu, C., Liu, K., Kong, Z., Li, Q., Zhou, X., Kan, Y., Wu, W., 2024. HoloVIC: Large-scale Dataset and Benchmark for Multi-Sensor Holographic Intersection and Vehicle-Infrastructure Cooperative. *arXiv preprint arXiv:2403.02640*.

Mao, J., Niu, M., Jiang, C., Liang, H., Chen, J., Liang, X., Li, Y., Ye, C., Zhang, W., Li, Z. et al., 2021. One million scenes for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*.

Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., Li, H., 2020. PV-RCNN: Point-voxel feature set abstraction for 3D object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10526–10535.

Sun, P., Sun, C., Wang, R., Zhao, X., 2022. Object Detection Based on Roadside LiDAR for Cooperative Driving Automation: A Review. *Sensors*, 22(23).

Thonhofer, E., Sigl, S., Fischer, M., Heuer, F., Kuhn, A., Erhart, J., Harrer, M., Schildorfer, W., 2023. Infrastructure-Based Digital Twins for Cooperative, Connected, Automated Driving and Smart Road Services. *IEEE Open Journal of Intelligent Transportation Systems*, 4, 311–324.

Wang, H., Zhang, X., Li, J., Li, Z., Yang, L., Pan, S., Deng, Y., 2021. IPS300+: a Challenging Multimodal Dataset for Intersection Perception System. *International Conference on Robotics and Automation (ICRA)*, 2539–2545.

Weng, X., Wang, J., Held, D., Kitani, K., 2020. 3D multi-object tracking: A baseline and new evaluation metrics. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10359–10366.

Wu, H., Deng, J., Wen, C., Li, X., Wang, C., Li, J., 2022. CasA: A Cascade Attention Network for 3-D Object Detection From LiDAR Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11.

Wu, J., Xu, H., Zheng, Y., Zhang, Y., Lv, B., Tian, Z., 2019. Automatic vehicle classification using roadside LiDAR data. *Transportation Research Record*, 2673(6), 153–164.

Wu, Y., Wang, Y., Zhang, S., Ogai, H., 2020. Deep 3D object detection networks using LiDAR data: A review. *IEEE Sensors Journal*, 21(2), 1152–1171.

Yin, T., Zhou, X., Krahenbuhl, P., 2021. Center-based 3D Object Detection and Tracking. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11779–11788.

Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., Guo, Z., Li, H., Hu, X., Yuan, J., Nie, Z., 2022. DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21329–21338.

Zhang, J., Xiao, W., Coifman, B., Mills, J. P., 2020. Vehicle tracking and speed estimation from roadside lidar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5597–5608.

Zhang, J., Xiao, W., Mills, J. P., 2022a. Optimizing Moving Object Trajectories from Roadside Lidar Data by Joint Detection and Tracking. *Remote Sensing*, 14(9).

Zhang, Y., Hu, Q., Xu, G., Ma, Y., Wan, J., Guo, Y., 2022b. Not all points are equal: Learning highly efficient point-based detectors for 3d lidar point clouds. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18931–18940.

Zhao, J., Xu, H., Liu, H., Wu, J., Zheng, Y., Wu, D., 2019. Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors. *Transportation research part C: emerging technologies*, 100, 68–87.

Zimmer, W., Creß, C., Nguyen, H. T., Knoll, A. C., 2023. Tumtraf intersection dataset: All you need for urban 3d camera-lidar roadside perception. *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 1030–1037.

Zio, E., Miqueles, L., 2024. Digital Twins in safety analysis, risk assessment and emergency management. *Reliability Engineering & System Safety*, 110040.