

Article

SimoSet: A 3D Object Detection Dataset Collected from Vehicle Hybrid Solid-State LiDAR

Xinyu Sun ¹, Lisheng Jin ^{1,2,*}, Yang He ¹, Huanhuan Wang ¹, Zhen Huo ¹ and Yewei Shi ¹

¹ School of Vehicle and Energy, Yanshan University, Qinhuangdao 066004, China; sunxinyu@stumail.ysu.edu.cn (X.S.); ysuvechy@stumail.ysu.edu.cn (Y.H.); vehicle_whh@stumail.ysu.edu.cn (H.W.); vehicle_huo@stumail.ysu.edu.cn (Z.H.); shiyw@stumail.ysu.edu.cn (Y.S.)

² Hebei Key Laboratory of Special Delivery Equipment, Yanshan University, Qinhuangdao 066004, China

* Correspondence: jinls@ysu.edu.cn

Abstract: Three-dimensional (3D) object detection based on point cloud data plays a critical role in the perception system of autonomous driving. However, this task presents a significant challenge in terms of its practical implementation due to the absence of point cloud data from automotive-grade hybrid solid-state LiDAR, as well as the limitations regarding the generalization ability of data-driven deep learning methods. In this paper, we introduce SimoSet, the first vehicle view 3D object detection dataset composed of automotive-grade hybrid solid-state LiDAR data. The dataset was collected from a university campus, contains 52 scenes, each of which are 8 s long, and provides three types of labels for typical traffic participants. We analyze the impact of the installation height and angle of the LiDAR on scanning effect and provide a reference process for the collection, annotation, and format conversion of LiDAR data. Finally, we provide baselines for LiDAR-only 3D object detection.

Keywords: autonomous driving; hybrid solid-state LiDAR; point cloud dataset; 3D object detection; deep learning



Citation: Sun, X.; Jin, L.; He, Y.; Wang, H.; Huo, Z.; Shi, Y. SimoSet: A 3D Object Detection Dataset Collected from Vehicle Hybrid Solid-State LiDAR. *Electronics* **2023**, *12*, 2424. <https://doi.org/10.3390/electronics12112424>

Academic Editor: Silvia Liberata Ullo

Received: 4 May 2023
Revised: 24 May 2023
Accepted: 25 May 2023
Published: 26 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Autonomous driving technology has attracted widespread attention in recent years due to its potential to free drivers from tiring driving activities and improve travel safety and traffic efficiency [1,2]. As a key component of environment perception technology, the task of 3D object detection forms the basis and premise for intelligent vehicles to obtain information regarding the surrounding area to ensure safe autonomous driving. To achieve this, LiDAR is considered as the core sensor in strong perception schemes and has been deployed in autonomous vehicles [3,4].

Deep learning is an end-to-end method that does not require manual feature engineering and can uncover potential features of data [5,6]. Therefore, it is favored by more and more researchers in 3D point cloud object detection algorithms, driving the need for benchmark datasets. Existing open-source datasets collect point cloud data using mechanical spinning LiDAR, which differs from the hybrid solid-state LiDAR channel assembled by mass-produced vehicles, resulting in significant differences in the distribution of presented point cloud data. The opposite results may be exhibited on mechanical spinning LiDAR and hybrid solid-state LiDAR due to the different point cloud resolutions [7]. However, due to the regulatory protection of data by autonomous driving companies, there are no publicly available hybrid solid-state LiDAR point cloud datasets—inhibiting algorithm development and deployment.

To help fill the gap of automotive-grade hybrid solid-state LiDAR point cloud data and facilitate the application testing of 3D object detection algorithms, we launched SimoSet, an open-source dataset for training 3D point cloud object detection deep learning models, to further accelerate the application of 3D perception technology and provide baseline results

for two classic 3D object detection methods. At least to our knowledge, SimoSet is the first 3D object detection dataset that uses automotive-grade hybrid solid-state LiDAR. We hope that SimoSet will serve as a valuable resource to promote the research and application of deep learning-based 3D object detection algorithms.

The main contributions of this paper are listed as follows:

- (1) We present a single modal point cloud dataset named SimoSet. It is the world's first open-source dataset dedicated to the task of 3D object detection and uses a hybrid solid-state LiDAR to collect point cloud data.
- (2) Data for SimoSet were collected in a university campus, including complex traffic environments, varied time periods and lighting conditions, and major traffic participant annotation classes. Based on SimoSet, we provide baselines for LiDAR-only 3D object detection.
- (3) The SimoSet dataset is aligned to the KITTI format for direct use by researchers. We share the procedure of data collection, annotation, and format conversion for LiDAR, which can be used as a reference for researchers to process custom data.

2. Related Work

2.1. Related Datasets

Over the past few years, data-driven approaches to deep learning have become increasingly popular, and there have been an increasing number of datasets being released. These datasets have greatly promoted the research and development of 3D Perception technology in the field of autonomous driving [8,9]. Numerous datasets for 3D object detection using point cloud data have been released by organizations globally [10–20]. We list some of the open-source datasets collected with LiDAR and annotated by 3D bounding boxes in Table 1.

Table 1. Autonomous driving dataset for 3D point cloud object detection.

Dataset	Year	LiDARs	Scenes	Ann. Frames	Classes	Night	Locations
KITTI	2012	1 × MS	22	7481	8	No	Germany
ApolloScape	2019	2 × MS	-	144k	6	Yes	China
H3D	2019	1 × MS	160	27k	8	No	USA
Lyft L5	2019	3 × MS	366	46k	9	No	USA
Argoverse	2019	2 × MS	113	22k	15	Yes	USA
A*3D	2019	1 × MS	-	39k	7	Yes	SG
A2D2	2020	5 × MS	-	12k	14	Yes	Germany
nuScenes	2020	1 × MS	1k	40k	23	Yes	SG, USA
Waymo Open	2020	5 × MS	1150	200k	4	Yes	USA
Cirrus	2020	2 × FF	12	6285	8	Yes	USA
Pandaset	2021	1 × MS	103	8240	28	Yes	USA
		1 × FF					
Simoset	2023	1 × FF	52	4160	3	Yes	China

Notes: (-) indicates that no information is provided; MS: Mechanical spinning; FF: Forward-facing.

The innovative KITTI dataset [10], released in 2012, is considered the first benchmark dataset collected by an autonomous driving platform. KITTI uses a 64-channel mechanical spinning LiDAR to collect data from urban areas, rural areas, and highways during the daytime in Karlsruhe, Germany. In the KITTI dataset, an object is annotated when it appears in the field of view of the vehicle's front-view camera. The ApolloScape [11], H3D [12], Lyft L5 [13], Argoverse [14], and A*3D [15] datasets launched in 2019 expanded the quantity and quality of open-source datasets for 3D object detection. ApolloScape utilizes two survey-grade LiDAR to capture dense point cloud data in complex traffic environments. H3D focuses on congested and highly interactive urban traffic scenes. Lyft L5 and Argoverse comprise abundant high-definition semantic maps and extend to tasks such as motion prediction and motion planning. A*3D increases the diversity of scenes in terms of time periods, lighting conditions, and weather circumstances, including a large

number of night and heavily occluded scenes. The 2020 A2D2 [16] employs five LiDAR units to acquire point cloud data with precise timestamps. nuScenes [17] provides point cloud data within 70 m using a 32-channel mechanical spinning LiDAR, annotates 23 object classes, introduces new detection metrics that balance all aspects of detection performance, and demonstrates how the amount of data affects the excitation of algorithm performance potential. Waymo Open [18] equips one medium-range LiDAR with a scanning range of 75 m and four short-range LiDAR with a scanning range of 20 m to acquire various weather data in multiple regions. Two difficulty levels are set based on the number of points in the 3D bounding box, and the detection performance is divided into three levels according to the object distance. Currently, nuScenes and Waymo Open are the most commonly used open-source datasets for evaluating 3D object detection algorithms. Cirrus [19] adopts a long-range bi-pattern LiDAR to obtain point cloud data within a range of 250 m. PandaSet [20] provides point cloud data within a range of 250 m via a mechanical spinning LiDAR and a forward-facing LiDAR. The larger the data scale, the more diverse time and weather conditions and the more challenging scenes of open-source datasets promote the progress of 3D object detection technology. Vehicle long-range master LiDAR has evolved from being of the mechanical spinning type to the hybrid solid-state type. However, at present, there is no 3D object detection dataset composed of hybrid solid-state LiDAR point cloud data, and the generalization ability of data-driven deep learning algorithms is weak. These factors hinder the development and deployment of 3D object detection algorithms.

2.2. LiDAR-Only 3D Object Detection Methods

Three-dimensional (3D) object detection with point cloud data alone can mainly be summarized into two categories: voxel-based and point-based methods, depending on how one wishes to convert point cloud data to 3D representations for localizing objects.

The voxel-based methods convert irregular point cloud data into ordered grid representations and typically use PointNet to extract features. VoxelNet [21] is an innovative dataset that voxelizes the sparse point cloud and then uses Voxel Feature Extractor (VFE) and 3D convolutions to generate geometrical representations; however, the huge computational burden of the 3D convolutions results in low computational efficiency. To save the computational cost of empty voxels, SECOND [22] introduces 3D sparse convolutions and 3D submanifold sparse convolutions to reduce memory consumption. However, 3D sparse convolution is not user-friendly. To this end, PointPillars [23] converts 3D point cloud data to a 2D pseudo-image and uses highly optimized 2D convolution to achieve excellent performance. The user-friendly advantages of PointPillars mean that it is a mainstream approach suitable for industrial deployment. In order to further improve detection accuracy, CIA-SSD [24] makes use of spatial semantic features. Part-A2 [25] performs semantic segmentation on foreground points. SA-SSD [26] advises an auxiliary network that is only used during the training stage. The auxiliary network is used to guide the backbone to learn the structural information of 3D objects. This structure improves the detection accuracy without increasing the inference time. Voxel R-CNN [27] aggregates K-nearest neighbor voxel features in the second stage to refine 3D bounding boxes. FastPillars [28] proposes a Max-and-Attention Pillar Encoding (MAPE) module that can be used to minimize the loss of local fine-grained information in feature dimension reduction and applies structural reparameterization to make the network more compact and efficient. CenterPoint [29] presents using point representations to describe objects, which solves the constraints imposed by anchor-based methods on angle and size and diminishes the search space for objects. VoxelNeXt [30] proposes an efficient structure that predicts objects directly from sparse voxel features rather than relying on hand-crafted proxies. The voxel-based methods can achieve decent detection performance with promising efficiency. However, voxelization inevitably introduces quantization loss, leading to the loss of fine-grained 3D structural information. Moreover, the localization performance largely depends on the size of voxel grids. Smaller voxel grids can obtain more fine-grained feature representations but at the cost of longer running time.

The point-based methods directly learn geometry from raw point cloud data without additional preprocessing steps and typically use PointNet++ to extract features. PointRCNN [31] suggests a point-based two-stage 3D region proposal paradigm. In the first stage, proposals are generated by segmented foreground points. In the second stage, high-quality 3D bounding boxes are regressed, exploiting semantic features and local spatial cues. However, it is inefficient to extract features directly from the original point cloud data. Therefore, 3DSSD [32] recommends farthest point sampling in feature space and Euclidean space and applies fusion strategy to remove part of the background points. To reduce memory usage and computational cost, IA-SSD [33] extracts semantic information from points, keeping as many foreground points as possible. The point-based methods adopt a two-stage pipeline. Specifically, they estimate 3D object proposals in the first stage and refine the object proposals in the second stage. The point-based methods have higher detection accuracy, but they spend 90% of the runtime organizing irregular point data [34], which is inefficient.

The point-voxel methods combine the advantages of point-based and voxel-based methods. PV-RCNN [35] aggregates point features into the voxel-based framework through a voxel-to-keypoint encoding technique. HVPR [36] designs a memory module to simplify the interaction between voxel features and point features. To avoid the computational burden of point-based methods while preserving the precise geometric shape of the object in the original point cloud, LiDAR R-CNN [37] generates 3D region proposals using voxel features in the first stage, and refines the geometric information of the 3D bounding boxes utilizing the raw point cloud coordinates in the second stage. Overall, different detection pipelines have their own advantages in terms of detection accuracy and/or operational efficiency.

3. SimoSet Dataset

Here, we introduce sensor specification, sensor placement, scene selection, data annotation, and data format conversion, then provide a brief analysis of our dataset.

3.1. Sensor Specification and Layout

The data collection uses a forward-facing hybrid solid-state LiDAR (RS-LiDAR-M1, 150 m range at 10% reflectivity). RS-LiDAR-M1 employs a micro-electro-mechanical system (MEMS) solution that covers a $120^\circ \times 45^\circ$ spatial area through a zigzag scan pattern. Table 2 presents detailed specifications for the LiDAR.

Table 2. LiDAR specifications.

LiDAR	Details
1 × hybrid solid-state LiDAR	MEMS mirror-based scanning, 120° horizontal FOV, 25° Vertical FOV, equivalent to 125 channels @ 10 Hz, 150 m range @ 10% reflectivity (Robosense RS-LiDAR-M1)

Notes: FOV: Field of view.

In mass-produced vehicles, the installation locations for forward-facing hybrid solid-state LiDAR include above the front windshield, on both sides of the headlights, and in the intake grille. Due to the curved structure of the windshield, the signal of the LiDAR will be attenuated, resulting in the inability to meet the requirements for ranging and resolution. Therefore, the plan of installing LiDAR inside the front windshield has not been promoted yet. There are obvious differences in size between different brands or series of vehicles in the same category. In order to measure the scanning effect of hybrid solid-state LiDAR with different heights, the positions above the front windshield, the sides of the headlights, and the intake grille of the sedan are set heights of 1.4 m, 0.7 m, and 0.6 m, respectively. Whereas, the positions above the front windshield, the sides of the headlights, and the intake grille of the SUV are set at heights of 1.6 m, 0.8 m, and 0.7 m, respectively. The location and height of hybrid solid-state LiDAR on sedan and SUV are shown in Figure 1.

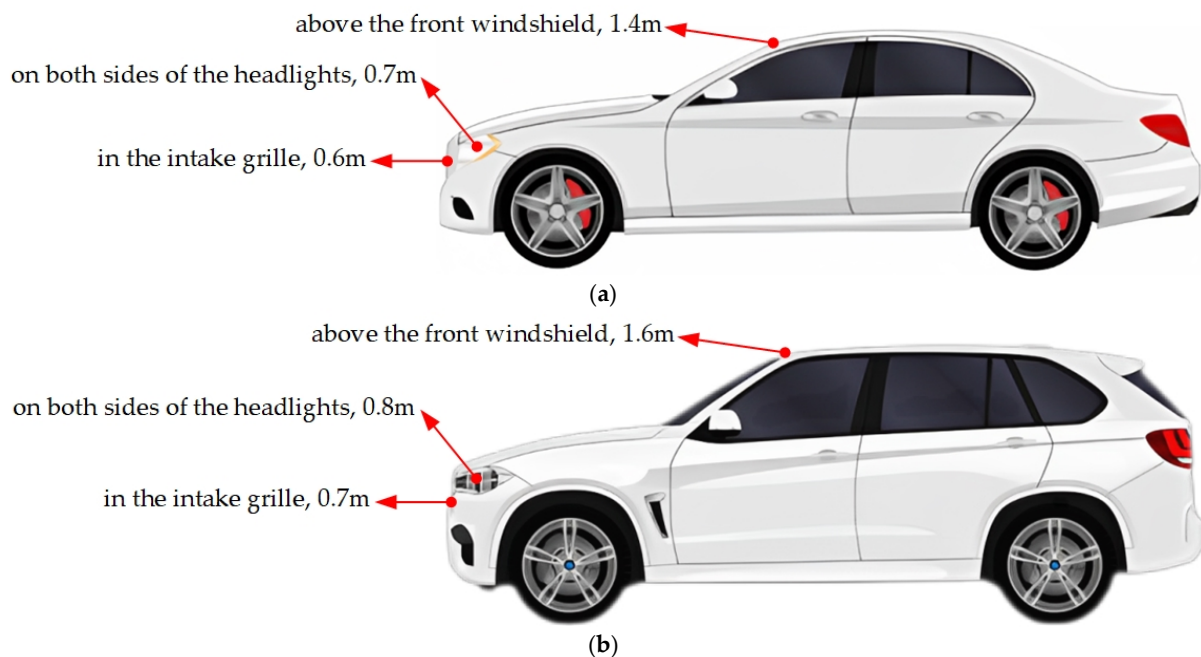


Figure 1. (a) The location and height of hybrid solid-state LiDAR on sedan. (b) The location and height of hybrid solid-state LiDAR on SUV.

To determine the installation position of the hybrid solid-state LiDAR on the vehicle, a LiDAR mounting bracket with two degrees of freedom is designed. The fixation plate can rotate along the pitch direction and slide vertically and is used to place the hybrid solid-state LiDAR. The spatial coordinate system is established according to the right-hand rule. The direction from the origin O of the coordinate system to the positive direction of the X -axis is defined as the initial 0° rotation around the Y -axis. When the view from the positive direction of the Y -axis is towards the origin O of the coordinate system, the counterclockwise direction is defined as the positive direction of rotation around the Y -axis. The angle range of the fixation plate rotating around the Y -axis is from 0° to 15° . The height range of the fixation plate sliding along the Z -axis is from 0.1 m to 1.7 m. The LiDAR mounting bracket and its coordinate system are shown in Figure 2.

The installation height directly affects the scanning effect of the hybrid solid-state LiDAR. The hybrid solid-state LiDAR is placed at 0° . The blind spot, the farthest ground line distance, and the effective detection distance are measured at the height corresponding to above the front windshield, the sides of the headlights, and the intake grille of both a sedan and an SUV. The effective detection distance is defined as the farthest distance at which the number of point cloud data representing pedestrian objects is no less than 5. It is observed that, although there is fluctuation in the point cloud from the same laser beam between adjacent frames, when the angle remains unchanged, the blind spot, the farthest ground line distance, and the effective detection distance tend to shorten in terms of height. In the current 360° full coverage perception scheme, hybrid solid-state LiDAR is used as the main long-range LiDAR for forward detection on the vehicle, focusing on long distance. Therefore, the installation height of the hybrid solid-state LiDAR is set at 1.6 m.

The installation angle is an important factor affecting the scanning effect of the hybrid solid-state LiDAR. For the task of 3D object detection, more attention is paid to traffic participants walking or driving on the ground, such as cars, cyclists, and pedestrians. Therefore, the hybrid solid-state LiDAR is placed at 1.6 m. Subsequently, the blind spot, the farthest ground line distance, and the effective detection distance are measured every 5° along the overlooking direction from 0° to 15° , starting from 0° horizontally. When the height remains unchanged, as the angle increases, the blind spot, farthest ground line distance, and effective detection distance all decrease. This change is most significant at

15°. The installation angle of hybrid solid-state LiDAR is set at 0° for ease of installation and maintenance.

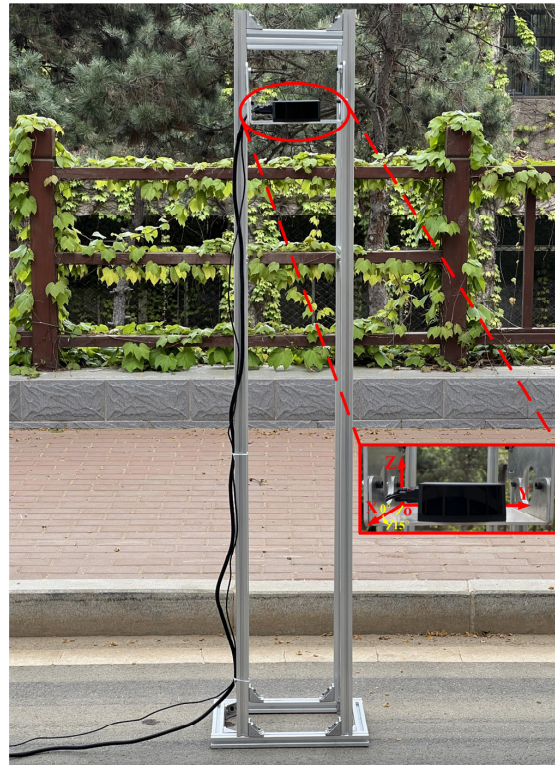


Figure 2. LiDAR mounting bracket and its coordinate system.

The placement position of the hybrid solid-state LiDAR (above the front windshield of the SUV) is determined with respect to the position of the mass-produced vehicles and the scanning effect of the test at typical heights and angles. The hybrid solid-state LiDAR is installed on the autonomous vehicle test platform according to the simulated position and pose, as shown in Figure 3.

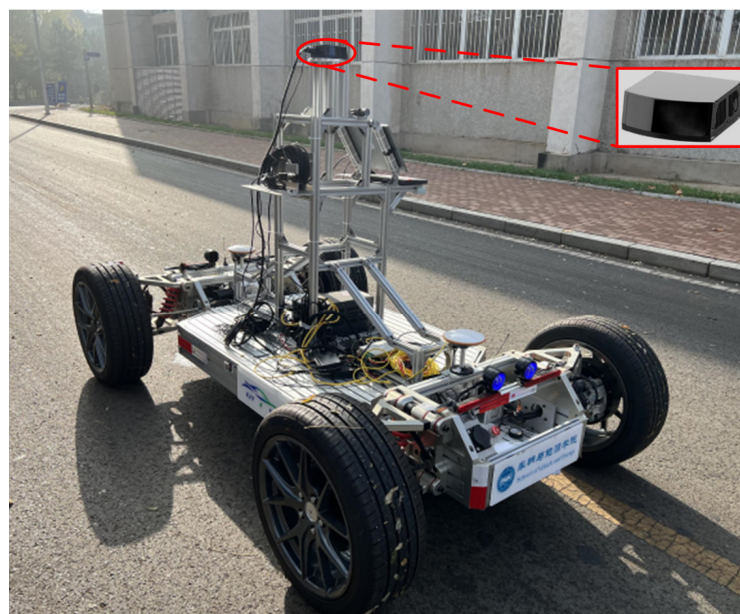


Figure 3. The autonomous vehicle test platform with hybrid solid-state LiDAR installed according to the simulated position and pose.

3.2. Scene Selection and Data Annotation

The raw data packets were collected through an autonomous vehicle test platform equipped with a hybrid solid-state LiDAR at Yanshan University campus. After obtaining the raw sensor data, 52 scenes were carefully selected, each lasting 8 s. There were a total of 4160 point cloud frames of the forward-facing hybrid solid-state LiDAR. These scenes covered different driving conditions, including complex traffic environments (e.g., intersections, construction), important traffic participants (e.g., cars, cyclists, pedestrians), and different lighting conditions, throughout the day and at night. The diversity of the scenes helps to capture the complex scenarios found in real-world driving.

SimoSet provides high-quality ground truth annotations of the hybrid solid-state LiDAR data, including 3D bounding box labels for all objects in the scenes. The annotation frequency remains 10 Hz. For the 3D object detection task, cars, cyclists, and pedestrians are exhaustively annotated in the LiDAR sensor readings. Each object is labeled as a 3D upright bounding box $(x, y, z, l, w, h, \theta)$ with 7 degrees of freedom (DOF), where x, y, z represent the center coordinates; l, w, h , are the length, width, height; and θ denotes the heading angle of the bounding box in radians. All cuboids contain at least five LiDAR points [20], while cuboids with less than five object points are discarded. All ground truth labels of point cloud data are created by human annotators using SUSTechPOINTS [38]. Multiple phases of label verification are performed to ensure high precision as well as quality annotations. An example of a labeled hybrid solid-state LiDAR point cloud is shown in Figure 4, where the annotated bounding boxes are displayed in blue, and the object point cloud data within the 3D bounding boxes are colored in red.

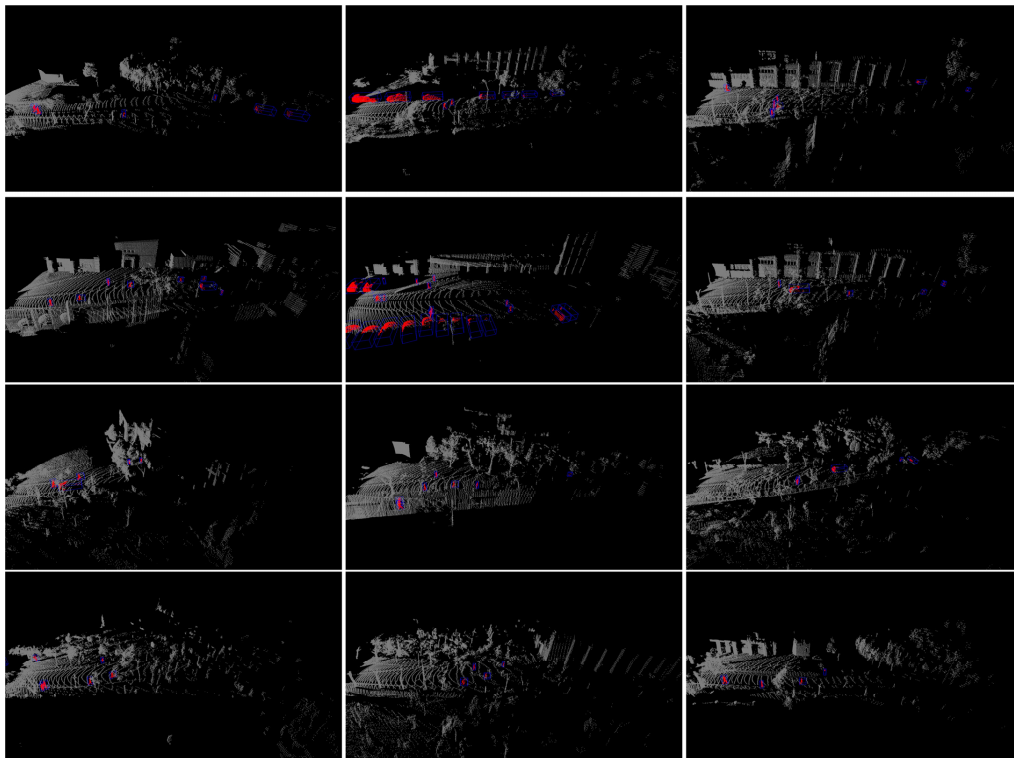


Figure 4. Annotated point cloud of hybrid solid-state LiDAR.

3.3. Format Conversion and Dataset Statistics

We create a virtual camera coordinate system, add fake images, and convert the data format to the widely known KITTI dataset format for researchers to use. In the pre-processing stage of point cloud data, we discard the filtering of point cloud range by image borders. We measured that the number of point clouds of the pedestrian objects fluctuate near the annotation threshold at a distance of 75 m. Considering the location

and number of the annotated objects, two levels of difficulty are designed based on the horizontal distance of objects. LEVEL_1 and LEVEL_2 correspond to the object ranges of (0 m, 35 m] and (35 m, 70 m], respectively. The object horizontal distance, LEVEL_1 and LEVEL_2, are defined as follows:

$$\text{range}_{\text{object}} = \sqrt{x^2 + y^2} \tag{1}$$

$$\text{LEVEL}_1 = \text{range}_{\text{object}} \in (0 \text{ m}, 35 \text{ m}] \tag{2}$$

$$\text{LEVEL}_2 = \text{range}_{\text{object}} \in (35 \text{ m}, 70 \text{ m}] \tag{3}$$

where $\text{range}_{\text{object}}$ is the horizontal distance of the object, x and y are the horizontal center coordinates of the object, LEVEL_1 represents the easy level, and LEVEL_2 denotes the hard level.

The evaluation metric for 3D object detection adopts Average Precision (AP) [39]. The AP is calculated as:

$$\text{AP}|_R = \frac{1}{|R|} \sum_{r \in R} \rho_{\text{interp}}(r) \tag{4}$$

where R is the equally spaced recall level, r is the recall, and $\rho_{\text{interp}}(r)$ is the interpolation function. Specifically, $\text{AP}|_{R_{40}}$ is used, $R_{40} = \{1/40, 2/40, 3/40, \dots, 1\}$, that is, averaging precision results on 40 recall positions.

SimoSet provides three types of class annotations. The 3D annotation cuboids for cars, cyclists, and pedestrians are 8884, 4063, and 9611, respectively. See Figure 5 for statistics regarding annotated cuboids by classes in SimoSet. The basic classes of traffic participant annotations can meet the requirements for developing 3D object detection algorithms while avoiding the problem of class imbalance.

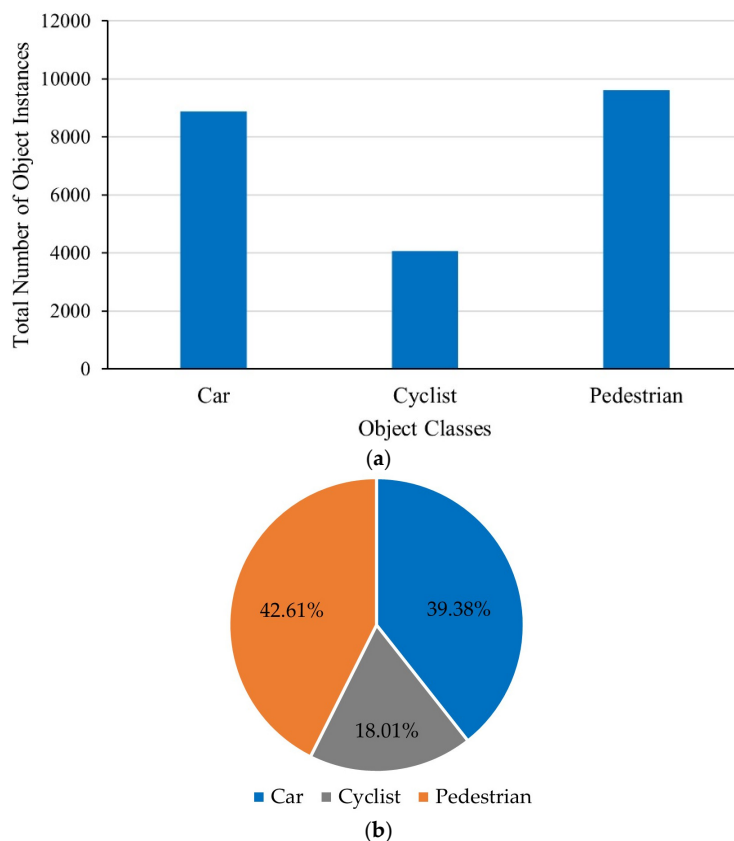


Figure 5. (a) Statistics regarding annotated cuboids by classes in SimoSet. (b) Proportion statistics of annotated cuboids by classes in SimoSet.

SimoSet provides pre-defined training (32 scenes) and tests (20 scenes) set splits. The training set contains 2560 frames with 14,249 3D object annotations. The test set consists of 1600 frames with 8309 3D object annotations. The proportion of 3D object annotations in the training and test sets is shown in Figure 6. SimoSet defines 3D bounding boxes for 3D object detection, and we anticipate the need to extend it to perform 3D multi-object tracking tasks in the future.

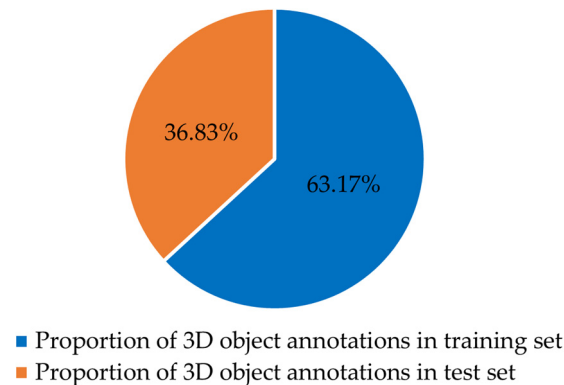


Figure 6. Proportion of 3D object annotations in training and test sets.

4. Baseline Experiments

We established baselines on our dataset with methods for LiDAR-only 3D object detection. Training and test sets were created according to the SimoSet pre-defined dataset split. The AP of seven DOF 3D boxes in 40 recall positions is adopted as the evaluation benchmark. Cars, cyclists, and pedestrians are chosen as detection objects, with 0.7 Intersection-over-Union (IoU) used for vehicles and 0.5 IoU used for cyclists and pedestrians. The class imbalance problem is an inherent attribute of autonomous driving scenarios. We tested the performance metrics of baseline algorithms under the Det3D [40] framework with a class-balanced sampling and augmentation strategy.

To establish the baseline for LiDAR-only 3D object detection, the typical voxel-based (SECOND, PointPillars, SA-SSD) and point-based (PointRCNN) methods were retrained, taking the widely deployed PointPillars algorithm for point cloud object detection as an example. As shown in Figure 7, the PointPillars network structure consists of three parts: a pillar encoder module that converts a point cloud to a sparse pseudo-image, a 2D convolutional backbone for feature extraction, a detection head for 3D boxes regression. PointPillars first divides the 3D space point cloud into pillars, extracts features using PointNet, and converts them into a sparse pseudo-image. Then, multiple 2D convolutions are utilized for downsampling to generate feature maps of different resolutions, which are aligned to the same size through multiple 2D deconvolution upsampling before being concatenated. Finally, the multi-scale features are fed into a region proposal network composed of 2D convolutional neural networks to regress the class, location, orientation, and scale.

Considering the different coverage between the forward hybrid solid-state LiDAR used by SimoSet and the mechanical spinning LiDAR used in existing open-source datasets, it is unsurprising that the network configuration is slightly different. The model is trained on single frame point cloud data of hybrid solid-state LiDAR. The detection range along the x -axis is set to [0 m, 76.8 m], and the detection range along the y -axis is set to [\times 40.8 m, 40.8 m] and [−3 m, 2 m] along the z -axis. The pillar size is set to (0.1 m, 0.1 m, 5 m). LiDAR point cloud frame data are based on the ego vehicle frame, whose x -axis is positive in the forward direction, y -axis is positive to the left, and z -axis is positive in the upward direction. In the 3D proposal generation module, the anchor sizes (l , w , h) were 4.5 m, 1.93 m, 1.51 m for cars; 1.58 m, 0.7 m, 1.49 m for cyclists; and 0.7 m, 0.66 m, 1.65 m for pedestrians. All object classes have anchors oriented to 0 and $\pi/2$ radians. The backbone, neck, head, and framework details do not be modified.

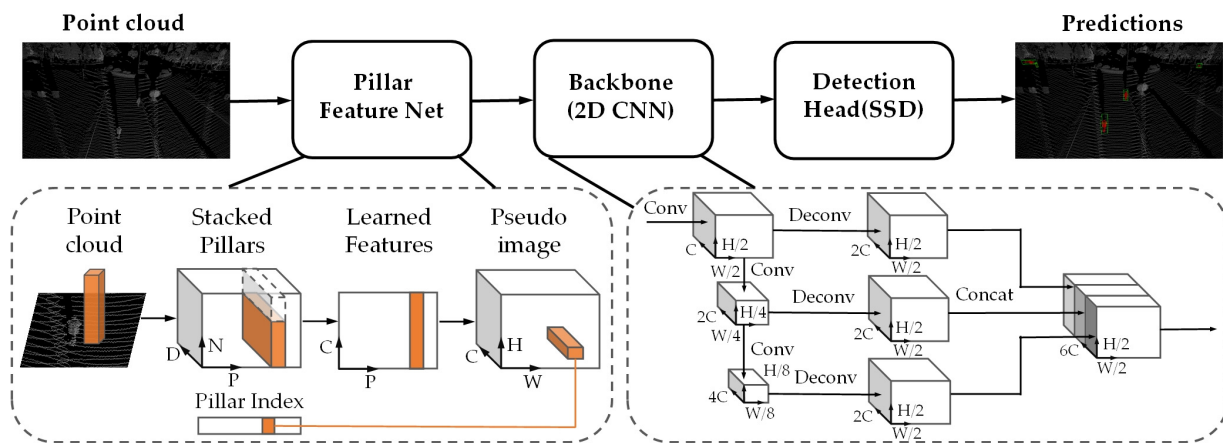


Figure 7. The overall network structure of PointPillars.

The AP results of 3D proposals at different levels on the test set for the trained PointPillars and PointRCNN models can be seen in Table 3.

Table 3. Baselines 3D AP for LiDAR-only 3D object detection.

Method	Type	Stage	GPU	Class	AP _{3D} (%)	
					LEVEL_1	LEVEL_2
SECOND	Voxel-based	One	GTX 3070	Car	80.23	65.71
				Cyclist	81.52	45.08
				Pedestrian	78.16	38.92
PointPillars	Voxel-based	One	GTX 3070	Car	78.73	63.15
				Cyclist	84.05	47.13
				Pedestrian	81.50	42.55
SA-SSD	Voxel-based	One	GTX 3070	Car	82.79	68.93
				Cyclist	86.36	47.87
				Pedestrian	85.28	45.29
PointRCNN	Point-based	Two	GTX 3070	Car	80.65	67.00
				Cyclist	85.27	47.53
				Pedestrian	85.45	21.81

From Table 3, it can be seen that the point-based method has better performance than the voxel-based methods, and that the auxiliary network can aid feature representation. The PointRCNN has lower AP for pedestrian object at LEVEL_2, which may be due to the sparsity of the point cloud for distant pedestrians, resulting in insufficient neighbor point features being relied upon by the PointRCNN model. As the distance increases, the object point cloud gradually becomes sparse, leading to a decline in the detection performance of the object. We found that the AP of cars in the test results is lower than that of cyclists and pedestrians. This may be due to the fact that, in some scenes, vehicles are parked along the roadside sequentially, with the preceding vehicle occluding the shape features of the following vehicle’s point cloud. The horizontal FOV of the hybrid solid-state LiDAR is only 120°. When the object vehicle enters or exits the blind spot of the ego vehicle, no point cloud data are in the parts of the object vehicle outside of the coverage area of the LiDAR. However, as the distance between the object vehicle and the ego vehicle is relatively close, the object still requires special attention; thus, it is annotated. This is another reason why the AP of cars is lower than that of cyclists and pedestrians. The trained PointPillars model is applied to the test set for inference. A visualization of the results is shown in Figure 8. The predicted and ground truth bounding boxes are shown in green and blue, respectively. The object point cloud within the 3D bounding boxes are colored in red.

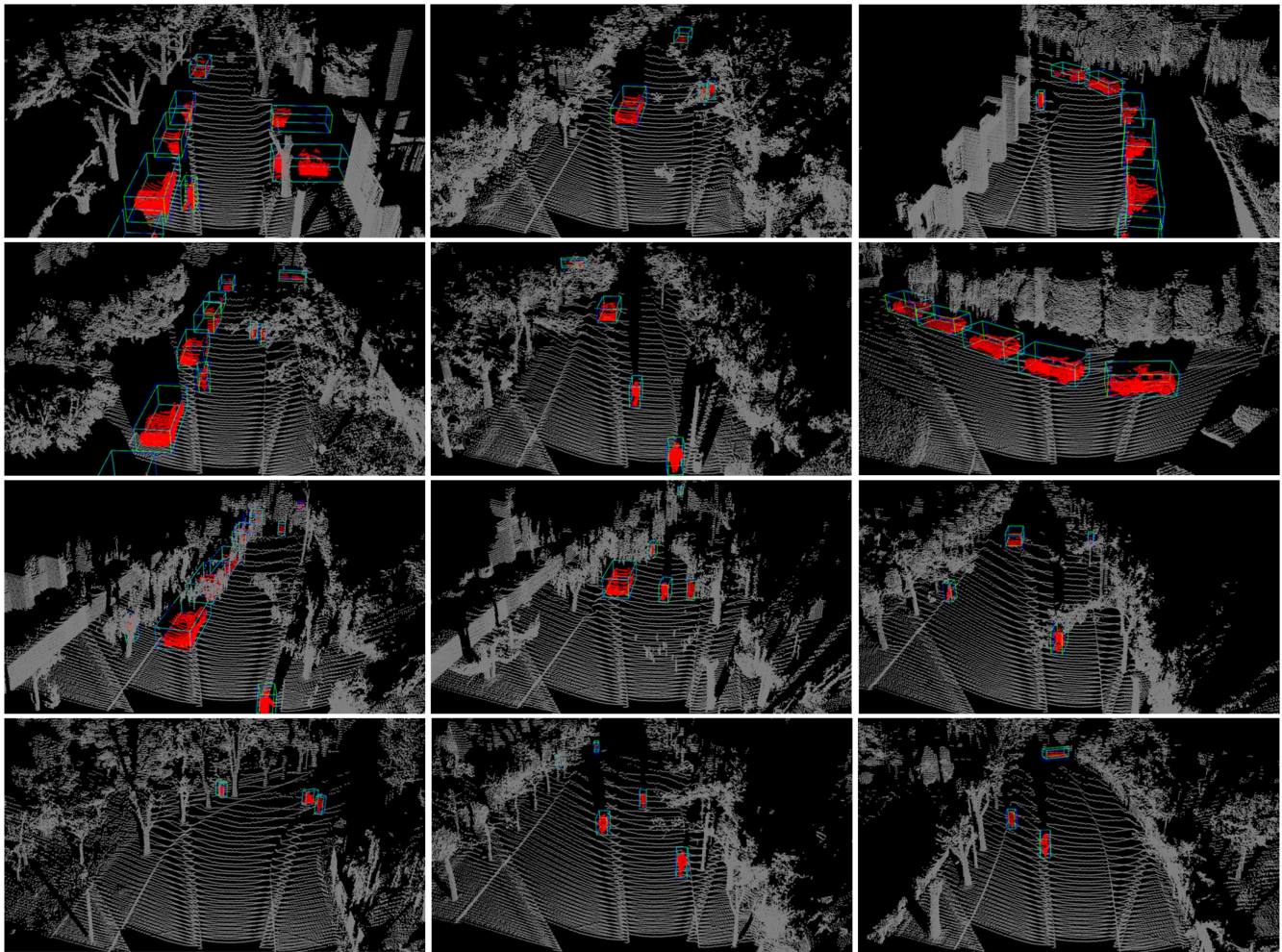


Figure 8. The inference results of the trained PointPillars model on the test set.

5. Conclusions

In this paper, we introduced SimoSet, the world's first open-source dataset collected from automotive-grade hybrid solid-state LiDAR for 3D object detection. We collected 52 scenes in a university campus, annotated typical traffic participants of 3 types, counted the number of objects for each type, and provided pre-defined training and test set splits and statistical proportion. Two levels of difficulty are designed based on the distance of the objects, and the AP of 3D bounding boxes is used as the evaluation metric. The performance of the LiDAR-only detectors on the SimoSet were chosen as the baselines, and the sample quality of SimoSet was demonstrated by utilizing the PointPillars algorithm. We also presented details on the installation height and angle of the hybrid solid-state LiDAR, data collection, 3D object annotation, and format conversion. By introducing SimoSet, we hope to help researchers accelerate the development and deployment of 3D point cloud object detection in the field of autonomous driving. We acknowledge that the number of data labels included in SimoSet is currently limited. In the future, we plan to expand the dataset with more diverse weather conditions, such as rain, snow, and fog. Further, we will extend SimoSet to include 3D point cloud object tracking.

Author Contributions: Conceptualization, X.S. and Y.H.; methodology, X.S.; software, Y.H.; validation, H.W. and Z.H.; formal analysis, H.W.; investigation, Z.H.; resources, L.J.; data curation, Y.S.; writing—original draft, X.S.; writing—review and editing, Y.H.; visualization, Y.S.; supervision, L.J.; project administration, L.J.; funding acquisition, L.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China (grant No. 2021YFB3202200), and National Natural Science Foundation of China (grant No. 52072333).

Data Availability Statement: SimoSet dataset and baselines code are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Koopman, P.; Wagner, M. Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 90–96. [[CrossRef](#)]
2. Khan, M.A.; El Sayed, H.; Malik, S.; Zia, T.; Khan, J.; Alkaabi, N.; Ignatious, H. Level-5 Autonomous Driving-Are We There Yet? A Review of Research Literature. *ACM Comput. Surv.* **2023**, *55*, 27. [[CrossRef](#)]
3. Li, Y.; Ibanez-Guzman, J. Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [[CrossRef](#)]
4. Roriz, R.; Cabral, J.; Gomes, T. Automotive LiDAR Technology: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 6282–6297. [[CrossRef](#)]
5. Qi, C.R.; Su, H.; Mo, K.C.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 77–85.
6. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet plus plus: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
7. Theodose, R.; Denis, D.; Chateau, T.; Fremont, V.; Checchin, P. A Deep Learning Approach for LiDAR Resolution-Agnostic Object Detection. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 14582–14593. [[CrossRef](#)]
8. Arnold, E.; Al-Jarrah, O.Y.; Dianati, M.; Fallah, S.; Oxtoby, D.; Mouzakitis, A. A Survey on 3D Object Detection Methods for Autonomous Driving Applications. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3782–3795. [[CrossRef](#)]
9. Li, Y.; Ma, L.F.; Zhong, Z.L.; Liu, F.; Chapman, M.A.; Cao, D.P.; Li, J.T. Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 3412–3432. [[CrossRef](#)] [[PubMed](#)]
10. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
11. Huang, X.Y.; Wang, P.; Cheng, X.J.; Zhou, D.F.; Geng, Q.C.; Yang, R.G. The ApolloScape Open Dataset for Autonomous Driving and Its Application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2702–2719. [[CrossRef](#)] [[PubMed](#)]
12. Patil, A.; Malla, S.; Gang, H.M.; Chen, Y.T. The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes. In Proceedings of the International Conference on Robotics and Automation, Montreal, QC, Canada, 20–24 May 2019; pp. 9552–9557.
13. Houston, J.; Zuidhof, G.; Bergamini, L.; Ye, Y.W.; Chen, L.; Jain, A.; Omari, S.; Iglovikov, V.; Ondruska, P. One Thousand and One Hours: Self-driving Motion Prediction Dataset. *arXiv* **2020**, arXiv:2006.14480.
14. Chang, M.F.; Lambert, J.; Sangkloy, P.; Singh, J.; Bak, S.; Hartnett, A.; Wang, D.; Carr, P.; Lucey, S.; Ramanan, D. Argoverse: 3D Tracking and Forecasting with Rich Maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8740–8749.
15. Pham, Q.H.; Sevestre, P.; Pahwa, R.S.; Zhan, H.J.; Pang, C.H.; Chen, Y.D.; Mustafa, A.; Chandrasekhar, V.; Lin, J. A*3D Dataset: Towards Autonomous Driving in Challenging Environments. In Proceedings of the IEEE International Conference on Robotics and Automation, Paris, France, 31 May–15 June 2020; pp. 2267–2273.
16. Geyer, J.; Kassahun, Y.; Mahmudi, M.; Ricou, X.; Durgesh, R.; Chung, A.S.; Hauswald, L.; Pham, V.H.; Mühlegg, M.; Dorn, S.; et al. A2D2: Audi Autonomous Driving Dataset. *arXiv* **2020**, arXiv:2004.06320.
17. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A Multimodal Dataset for Autonomous Driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11618–11628.
18. Sun, P.; Kretzschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.N.; Caine, B.; et al. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 2443–2451.
19. Wang, Z.; Ding, S.H.; Li, Y.; Fenn, J.; Roychowdhury, S.; Wallin, A.; Martin, L.; Ryvola, S.; Sapiro, G.; Qiu, Q. Cirrus: A Long-range Bi-pattern LiDAR Dataset. In Proceedings of the IEEE International Conference on Robotics and Automation, Xi'an, China, 30 May–5 June 2021; pp. 5744–5750.
20. Xiao, P.C.; Shao, Z.L.; Hao, S.; Zhang, Z.S.; Chai, X.L.; Jiao, J.; Li, Z.S.; Wu, J.; Sun, K.; Jiang, K.; et al. PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving. In Proceedings of the IEEE International Transportation Systems Conference, Indianapolis, IN, USA, 19–22 September 2021; pp. 3095–3101.
21. Zhou, Y.; Tuzel, O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4490–4499.
22. Yan, Y.; Mao, Y.X.; Li, B. SECOND: Sparsely Embedded Convolutional Detection. *Sensors* **2018**, *18*, 3337. [[CrossRef](#)] [[PubMed](#)]

23. Lang, A.H.; Vora, S.; Caesar, H.; Zhou, J.B.; Yang, J.O.; Beijbom, O. PointPillars: Fast Encoders for Object Detection from Point Clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 12689–12697.
24. Zheng, W.; Tang, W.L.; Chen, S.J.; Jiang, L.; Fu, C.W. CIA-SSD: Confident IoU-Aware Single-Stage Object Detector from Point Cloud. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; pp. 3555–3562.
25. Shi, S.S.; Wang, Z.; Shi, J.P.; Wang, X.G.; Li, H.S. From Points to Parts: 3D Object Detection from Point Cloud with Part-Aware and Part-Aggregation Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2647–2664. [[CrossRef](#)] [[PubMed](#)]
26. He, C.H.; Zeng, H.; Huang, J.Q.; Hua, X.S.; Zhang, L. Structure Aware Single-Stage 3D Object Detection from Point Cloud. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11870–11879.
27. Deng, J.J.; Shi, S.S.; Li, P.W.; Zhou, W.G.; Zhang, Y.Y.; Li, H.Q. Voxel R-CNN: Towards High Performance Voxel-based 3D Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; pp. 1201–1209.
28. Zhou, S.F.; Tian, Z.; Chu, X.X.; Zhang, X.Y.; Zhang, B.; Lu, X.B.; Feng, C.J.; Jie, Z.Q.; Chiang, P.K.; Ma, L. FastPillars: A Deployment-friendly Pillar-based 3D Detector. *arXiv* **2023**, arXiv:2302.02367.
29. Yin, T.W.; Zhou, X.Y.; Krahenbuhl, P. Center-based 3D Object Detection and Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 11779–11788.
30. Chen, Y.K.; Liu, J.H.; Zhang, X.Y.; Qi, X.J.; Jia, J.Y. VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking. *arXiv* **2023**, arXiv:2303.11301.
31. Shi, S.S.; Wang, X.G.; Li, H.S. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 770–779.
32. Yang, Z.T.; Sun, Y.N.; Liu, S.; Jia, J.Y. 3DSSD: Point-based 3D Single Stage Object Detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2019; pp. 11037–11045.
33. Zhang, Y.F.; Hu, Q.Y.; Xu, G.Q.; Ma, Y.X.; Wan, J.W.; Guo, Y.L. Not All Points Are Equal: Learning Highly Efficient Point-based Detectors for 3D LiDAR Point Clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18931–18940.
34. Liu, Z.J.; Tang, H.T.; Lin, Y.J.; Han, S. Point-Voxel CNN for Efficient 3D Deep Learning. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.
35. Shi, S.S.; Guo, C.X.; Jiang, L.; Wang, Z.; Shi, J.P.; Wang, X.G. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10526–10535.
36. Noh, J.; Lee, S.; Ham, B. HVPR: Hybrid Voxel-Point Representation for Single-stage 3D Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 14600–14609.
37. Li, Z.C.; Wang, F.; Wang, N.F. LiDAR R-CNN: An Efficient and Universal 3D Object Detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 7542–7551.
38. Li, E.; Wang, S.J.; Li, C.Y.; Li, D.C.; Wu, X.B.; Hao, Q. SUSTech POINTS: A PorTable 3D Point Cloud Interactive Annotation Platform System. In Proceedings of the IEEE Intelligent Vehicles Symposium, Las Vegas, NV, USA, 23–16 June 2020; pp. 1108–1115.
39. Simonelli, A.; Bulò, S.R.; Porzi, L.; Lopez-Antequera, M.; Kotschieder, P. Disentangling Monocular 3D Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1991–1999.
40. Zhu, B.J.; Jiang, Z.K.; Zhou, X.X.; Li, Z.M.; Yu, G. Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection. *arXiv* **2019**, arXiv:1908.09492.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.