Dual Adaptive Data Augmentation for 3D Object Detection

Joohyun Lee[1], Jin-Hee Lee[2], Jae-Keun Lee[3], Je-Seok Kim[2], Soon Kwon[2,3,*], and Sangdong Kim[1,2,*]
[1]*Department of Interdisciplinary Engineering, DGIST, Daegu, Republic of Korea*
[2]*Division of Automotive Technology, DGIST, Daegu, Republic of Korea*
[3]*FutureDrive, Daegu, Republic of Korea*

***Abstract*—This paper proposes Dual Adaptive Data Augmentation (DADA) method for 3D object detection. Training deep learning models requires large amounts of data, which is time-consuming and expensive. To address this challenge, data augmentation methods have been proposed to generate augmented objects. However, conventional methods rely on fixed parameters and ignore scene and object characteristics. To address these limitations, we propose DADA, which consists of two modules: Scene-based ADA and Density-based ADA. Scene-based ADA adjusts augmented objects based on the distribution of Ground Truth (GT) objects in each scene, allowing augmentation to focus on sparse scenes with fewer GT objects while keeping overall data volume. Density-based ADA utilizes LiDAR characteristics to apply different sampling methods, generating diverse augmented objects based on object density. Experiment results show considerable improvement in performance on the KITTI and ONCE datasets.**

***Index Terms*—3D Object Detection, Data Augmentation, LiDAR**

I. Introduction

LiDAR-based 3D object detection models have demonstrated remarkable performance across various real-world applications such as autonomous driving, security, and the overall ICT industry. These object detection models typically require large amounts of annotated data for effective training. Especially in 3D object detection for the real-world domain, there are various real-world datasets. Most of these datasets are autonomous driving datasets [1], [2], [3] that cover a variety of driving scenarios. However, the process of generating these datasets can be time-consuming and expensive to collect and process data. To solve this problem, data augmentation techniques have been widely used to generate new training data by changing the existing annotated GT data without significant costs.

Data augmentation techniques for 3D point clouds are still less studied than those for 2D images. The general data augmentation techniques for point clouds include the global augmentation, which generates data with changes such as translation and rotation for all points, and the local augmentation, which generates augmented GT-based objects with translation and rotation for GT objects. In addition, there are data augmentation techniques that generate data by pasting objects from other point clouds, such as GT Sampling. These data augmentation techniques have the effect of increasing the training data through additional data generated from GT objects and improving the performance of 3D object detection. However, existing data augmentation techniques utilize invariable GT-based objects and rely on fixed parameters to apply the same technique to all scenes. Existing techniques focus on generating diverse scenes rather than diverse objects. We have observed that generating diverse objects and focusing on sparse scenes to augment diverse objects are more effective than training diverse scenes. In this paper, we propose DADA, a data augmentation technique that can generate diverse augmented objects for each scene by considering two key factors: scene and point density.

DADA is composed of Scene-based Adaptive Data Aug-

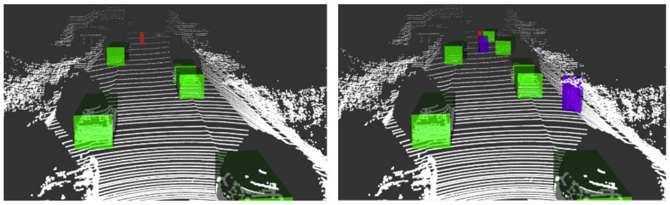

Fig. 1: The left and right figures show the KITTI training dataset before and after applying Scene-based ADA, respectively. The green, blue, and red boxes are the bounding boxes of the car, pedestrian, and cyclist classes, respectively. In the right figure, two cars and two pedestrians were augmented.

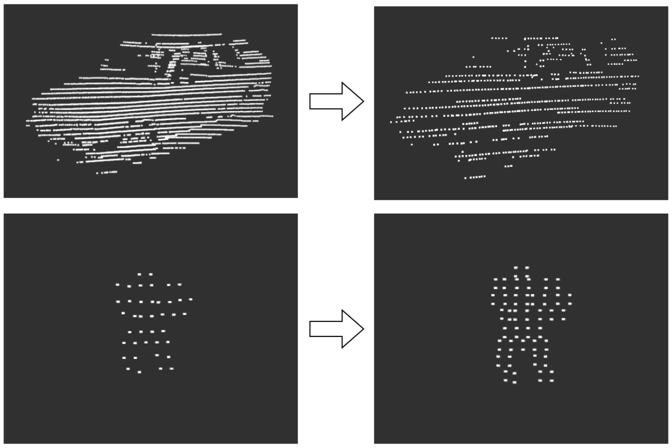

Fig. 2: The two figures at the top show the results of downsampling with Density-based ADA for the car class of the KITTI training dataset, and the two figures at the bottom show the results of upsampling with Density-based ADA for the pedestrian class of the KITTI training dataset.

[*] Corresponding authors: Soon Kwon `soonyk@dgist.ac.kr`, Sangdong Kim `kimsd728@dgist.ac.kr`

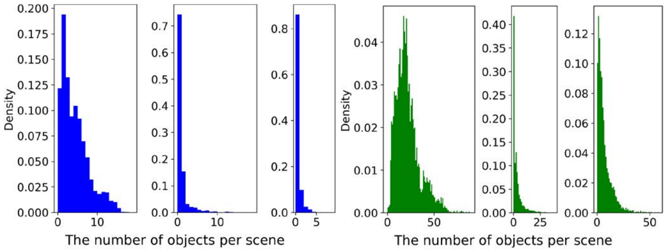

Fig. 3: Histogram for the number of objects which are car (vehicle in ONCE), pedestrian, and cyclist per scene in KITTI (left, blue) and ONCE (right, green) training datasets.

mentation (Scene-based ADA) and Density-based Adaptive Data Augmentation (Density-based ADA). Scene-based ADA is a technique that adaptively adjusts the augmentation according to the distribution of the objects in each scene. As shown in Fig. 3, the KITTI and ONCE datasets have unbalanced distributions in the number of objects in each scene. Existing techniques do not effectively respond to the varying number of objects in different scenes, and naively apply fixed parameters for object augmentations for each object class to all scenes. In contrast, as shown in Fig. 1, the Scene-based ADA approach determines the number of augmented data by considering the number of GT objects in the scene, so that data with fewer GT objects can be intensively augmented, and every scene contains a variety of objects.

While Scene-based ADA is about how many augmented objects to utilize, Density-based ADA is about how to generate augmented objects. Unlike 2D images from cameras, 3D point clouds from LiDAR are scale-invariant, depth-enabled, and characterized by sparse data. These features make them easier to distinguish between the background and the objects, which makes it possible to apply data augmentation to individual objects. In order to generate augmented objects, we propose Density-based ADA, a technique that diversifies data by integrating downsampling and upsampling strategies that utilize the characteristics of LiDAR according to the point density of the object. Conventional GT Sampling is not efficient for data augmentation because GT objects are reused without any changes. On the contrary, as shown in Fig. 2, our technique leverages the LiDAR characteristics to perform more adaptive sampling according to the object density, which can generate new sparse objects from dense ones and vice versa while preserving their shapes. This approach enables efficient data augmentation by generating new and diverse data from existing data and utilizing it for training.

We can summarize our contributions as follows:

- We have observed that focusing augmentation on sparse scenes by generating diverse objects from GT objects and adaptively adjusting the data augmentation per scene leads to significant performance gains.
- Our method can be applied to a variety of 3D object detection models, all of which performed well on the KITTI and ONCE validation datasets.
- In particular, it is noteworthy that our DADA can replace GT Sampling, which is used by most models in OpenPCDet, a huge open-source project for LiDAR-based 3D object detection, and our experiment results are superior in the evaluation of all classes and various conditions (KITTI metric, ONCE metric) compared to GT Sampling.

II. RELATED WORK

A. *Data Augmentation*

Data augmentation is a technique that generates additional training data from existing one to prevent overfitting and improve performance when training on a limited dataset. For instance, in the 2D image domain, there are Cutout [4], which augments an image patch by cutting out a portion of it, and Mixup [5], which mixes two images and labels. There is also CutMix [6], which combines these two techniques to augment a portion of an image by replacing it with a patch from another image.

In addition, there are also alternatives that apply techniques from the 2D domain to the 3D domain, such as PointMixup [7], which applies Mixup to a point cloud. Another proposed technique [8] involved upsampling from lower resolution points to increase resolution. SECOND [9] proposed a new data augmentation technique called GT Sampling. The technique involved cropping individual objects from each scene to create a GT database. During training, each scene is augmented by pasting objects from other scenes in the database. On the other hand, [10] introduced a technique that performs a global data augmentation on all points in a scene and subsequently a local data augmentation on each object. Real3D-Aug [11] proposed a technique to properly localize and handle the occlusion when performing the data augmentation. [12] proposed contextual GT Sampling that utilizes the semantic information to address data imbalances. LiDAR-AUG [13] proposed a technique for adding a CAD model to a scene and generating an augmented LiDAR point cloud through a rendering module. PA-AUG [14] divided the object into multiple partitions and probabilistically applied the existing local data augmentation to each partition region, which not only improved the accuracy of the given dataset but also performed well on corrupted data. Similar to PA-AUG, SE-SSD [15] divided the object into six pyramid shapes and performed dropout, swap, and sparsify operations on each of them to augment the data. PPBA [16] found that previous data augmentation techniques were manually designed. PPBA automated data augmentation techniques to find the optimal parameters. PointAugment [17] also utilized an adversarial learning strategy to automatically optimize and augment point clouds.

We found that existing data augmentation studies are divided into two categories, creating various objects from GT objects and augmenting GT objects with other scenes. In response, we propose DADA, which adaptively improves and integrates these two categories.

B. *3D Object Detection*

3D object detection is the task of classifying and localizing objects by taking a point cloud as input, which is a set of points

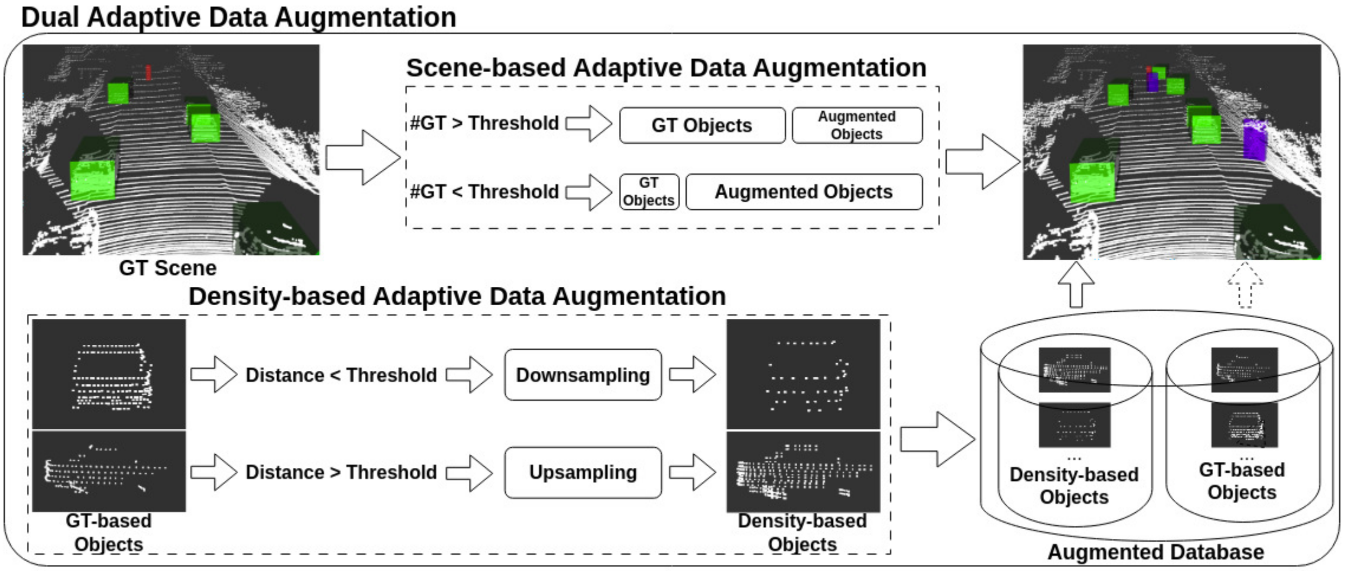

Fig. 4: The overall architecture of Dual Adaptive Data Augmentation method.

obtained by a LiDAR laser scan. As in 2D object detection, it is divided into one-stage and two-stage detectors depending on whether the region proposal is separated or not. VoxelNet [18] is a one-stage baseline model that performs end-to-end 3D object detection. Starting from VoxelNet, various models have been proposed. SECOND used sparse convolution instead of dense convolution to improve speed according to the sparse nature of LiDAR point clouds. In addition, SECOND addressed the issue in VoxelNet where large loss values are obtained when in the same bounding box but the direction is opposite. SECOND created an angle regression loss that can solve this problem and proposed GT Sampling as a data augmentation technique to improve performance. PV-RCNN [19] is a two-stage model that utilizes both voxel and point methods. The voxel-based method proposes high-quality 3D objects, and the point-based method reduces information loss to maximize the preservation of location information and improve detection performance. Finally, Voxel R-CNN [20], like PV-RCNN, is a two-stage model and utilizes voxels to reduce the computational cost, with the finding that accurate localization of points is not essential for 3D object detection, and achieves performance beyond that of point-based models by utilizing voxels.

We utilized the existing novel models PV-RCNN, SECOND, and Voxel R-CNN to verify the performance of the proposed DADA method.

III. DUAL ADAPTIVE DATA AUGMENTATION

A. Overview

As opposed to existing techniques based on fixed parameter values, our goal is to enhance the data augmentation to adaptively augment the training dataset, maximizing its efficacy in training 3D object detection models and improving the detection performance. As shown in Fig. 4, our proposed Dual Adaptive Data Augmentation includes two adaptive data augmentation techniques. Each adaptive data augmentation consists of Density-based ADA for generating augmented objects and Scene-based ADA for utilizing augmented objects with a database of augmented objects created by Density-based ADA. These two techniques form one efficient pipeline for data augmentation, DADA.

B. Density-based Adaptive Data Augmentation

We propose Density-based ADA, a downsampled and upsampled point generation technique that utilizes the characteristics of LiDAR to generate various data. For the points of the existing GT object, we simulate the points according to the LiDAR characteristics while maintaining the shape, and proceed with downsampling or upsampling according to the point density. This generates difficult objects from dense points and easy objects from sparse points, forming an augmented database including Density-based objects.

In order to describe the distribution of points in a LiDAR laser scan, our method performs a coordinate transformation from the orthogonal coordinate system, which is the coordinate system traditionally used by point clouds, to a spherical coordinate system to easily express the angles of the points. After that, we perform sampling according to the LiDAR characteristics. For example, the Velodyne HDL-64E LiDAR sensor used to acquire the KITTI dataset [1] has a vertical resolution of 0.4 degrees and a horizontal resolution of 0.08 to 0.35 degrees (5Hz to 20Hz). Also, the 40-beam LiDAR sensor used to acquire the ONCE dataset [3] has a vertical resolution of 0.33 degrees and a horizontal resolution of 0.2 to 0.4 degrees (10Hz to 20Hz). According to the LiDAR characteristics of the acquired data, the object is evenly sliced and sampled with the corresponding LiDAR resolution. Since our proposed technique describes a LiDAR laser scan, it can maintain its shape even after sampling, and its performance can be improved by generating robust data for training.

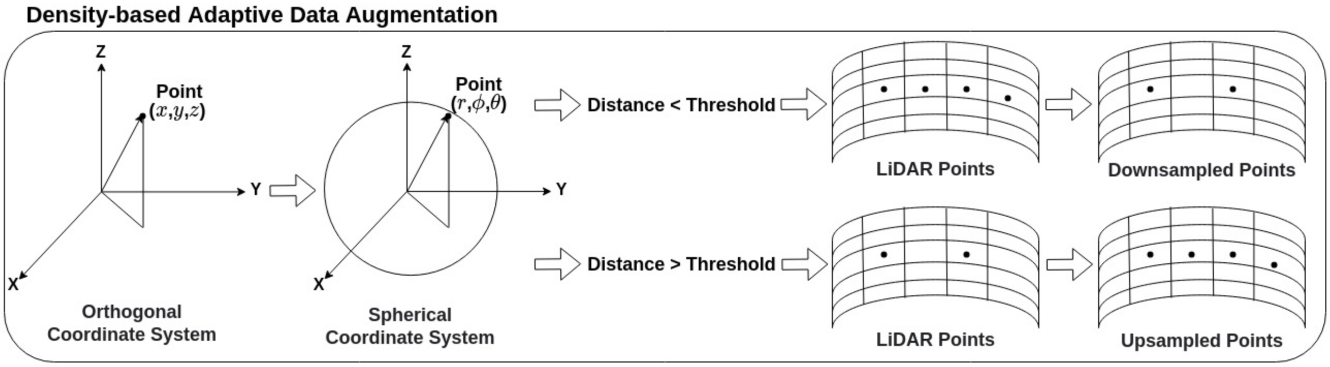

Fig. 5: The detailed process of Density-based Adaptive Data Augmentation.

Specifically, as shown in Fig. 5, a point *p* of the point cloud is first represented as (x, y, z) in the orthogonal coordinate system. To apply vertical and horizontal angle resolution, we convert point *p* to (r, ϕ, θ) in the spherical coordinate system so that it can be expressed in terms of distance and angle. The (r, ϕ, θ) of point *p* using the LiDAR as the origin are calculated as follows:

$$r = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

$$\theta = \arctan\left(\frac{z}{\sqrt{x^2 + y^2}}\right) \tag{2}$$

$$\phi = \arctan\left(\frac{y}{x}\right) \tag{3}$$

After converting to (r, ϕ, θ), the points of the object can be represented by simulating them at the resolution of the LiDAR. Then we denote which horizontal and vertical scans include each point on the grid, and the grid is horizontally and vertically sliced to remove or generate points of the object. Through this sampling, the natural point features generated by LiDAR can be represented as they are, and the points can be adjusted while maintaining the shape of the object in contrast to conventional sampling techniques. By training the sampled points of these objects, Density-based ADA can generate various samples rather than just reusing the GT objects for training in conventional techniques.

C. *Scene-based Adaptive Data Augmentation*

We propose Scene-based Adaptive Data Augmentation, which generates augmented objects through Density-based ADA to create a database, then adjusts their number according to the number of GT objects by considering the situation of each scene.

First, we adaptively adjust the number of augmented objects based on the number of GT objects for each scene. The model is trained through Scene-based ADA by adding more augmented objects for scenes with fewer objects and, conversely, by adding fewer augmented objects for scenes with more objects. As shown in (4), the sum of the number of GT objects and the number of augmented objects for a scene can be set as a parameter called *K* to adaptively adjust the data augmentation for sparse and dense scenes.

$$\Sigma GT.Obj + \Sigma Aug.Obj = K \tag{4}$$

Second, Scene-based ADA is performed by utilizing the augmented database created by the Density-based ADA technique. The augmented database is created by downsampling and upsampling the points divided by distance for dense and sparse GT objects, respectively. Then we determine the number of augmented objects through the parameter *K* and proceed to add them to the existing GT data, augmenting the training scene. For each scene, Density-based objects that do not overlap with GT objects are extracted from the augmented database and placed so that there are a total of *K* objects, by combining the number of GT and augmented objects. We created training data augmented with Scene-based ADA and utilized data for training to improve performance.

IV. EXPERIMENTS

A. *Implementation detail*

Our technique is trained and evaluated on PV-RCNN, SECOND, and Voxel R-CNN as 3D object detection models. Our implementation is based on OpenPCDet [21], an open-source platform for LiDAR-based 3D object detection that supports all of these models. All models were trained and evaluated on 8 NVIDIA TITAN X machines with 80 epochs. The IoU thresholds in the KITTI's evaluation were set to 0.7, and 0.5 for car, and pedestrian classes, respectively, and the evaluation was divided into easy, moderate, and hard according to the KITTI metric. On the other hand, the IoU thresholds in the ONCE's evaluation were set to 0.7, 0.3, and 0.5 for vehicle, pedestrian, and cyclist classes, respectively. In the case of Voxel R-CNN, only the car class was considered for training and evaluation because the original architecture based on OpenPCDet was utilized. For the cyclist class in the KITTI dataset, we considered only car and pedestrian classes because it does not guarantee the consistency of performance due to class imbalance.

To compare the performance of Scene-based ADA with the existing data augmentation without Scene-based ADA, we need to keep the total number of objects augmented for training similar in both methods. As shown in Fig. 3, we

TABLE I: 3D object detection performance results for the car and pedestrian classes on the KITTI validation dataset for a model trained by augmenting only certain scenes. The best performance values per model are bolded.

Model	Data Augmentation	AP_{car}			AP_{ped}			Avg.	
	Scene Distribution	Easy	Moderate	Hard	Easy	Moderate	Hard	Car	Pedestrian
PV-RCNN [19]	w/o GT Sampling	88.83	78.98	78.40	65.66	59.64	57.27	82.07	60.86
	Dense Scenes	89.26	79.12	78.55	**67.95**	61.57	56.93	82.31	62.15
	Sparse Scenes	**89.35**	**79.33**	**78.79**	67.28	**62.05**	**57.91**	**82.49**	**62.41**
SECOND [9]	w/o GT Sampling	86.23	75.53	72.74	51.06	47.01	44.36	78.17	47.48
	Dense Scenes	87.85	77.28	74.70	56.20	53.04	48.68	79.94	52.64
	Sparse Scenes	**88.45**	**77.88**	**76.59**	**57.54**	**53.60**	**48.95**	**80.97**	**53.36**

TABLE II: 3D object detection performance results evaluated on the KITTI validation dataset for the car and pedestrian classes. The best performance values per model are bolded.

Model	Data Augmentation		AP_{car}			AP_{ped}			Avg.	
	Scene	Density	Easy	Moderate	Hard	Easy	Moderate	Hard	Car	Pedestrian
PV-RCNN [19]	w/o GT Sampling		88.83	78.98	78.40	65.66	59.64	57.27	82.07	60.86
	w/ GT Sampling		89.01	79.18	78.57	63.65	57.31	53.14	82.25	58.03
	✓	✗	**89.43**	79.25	78.67	65.55	59.31	54.42	82.45	59.76
	✗	✓	89.39	79.24	78.64	67.51	59.41	56.41	82.42	61.11
	✓	✓	89.30	**79.39**	**78.84**	**68.58**	**62.57**	**57.32**	**82.51**	**62.82**
SECOND [9]	w/o GT Sampling		86.23	75.53	72.74	51.06	47.01	44.36	78.17	47.48
	w/ GT Sampling		88.09	**78.28**	76.81	56.73	**53.53**	47.78	81.06	52.68
	✓	✓	**88.61**	78.23	**76.98**	**58.75**	53.40	**49.12**	**81.27**	**53.76**
Voxel R-CNN [20]	w/o GT Sampling		89.02	78.95	78.19	-	-	-	82.05	-
	w/ GT Sampling		89.38	79.26	78.53	-	-	-	82.39	-
	✓	✓	**89.49**	**79.31**	**78.66**	-	-	-	**82.49**	-

identified the distribution of objects in each class in the KITTI training dataset of 3,712 frames to similarize the number of augmented objects. The *K* value for training with Scene-based ADA was set to 19 and 11 for car and pedestrian classes, respectively, while training without Scene-based ADA was performed with a fixed number of augmentations for each scene, 15 and 10 for each class. Same as KITTI, for the ONCE training dataset of 4,961 frames, the *K* value was set to 26, 4, 4, 13, and 16 for car, bus, truck, pedestrian, and cyclist classes, respectively, while the GT Sampling's fixed parameter value was set to 9, 3, 3, 10, and 10. In addition, for the performance evaluation of Density-based ADA, we experimented with various distance thresholds, defining a threshold based on 30 meters as a reference for distance-specific evaluation in the Waymo Open Dataset [2].

B. Overall results

Before evaluating the performance of Scene-based ADA, we first verified the effectiveness of data augmentation in dense and sparse scenes. Table I shows the results for the car and pedestrian classes, performed under three different setups: without GT Sampling, with dense scenes, and with sparse scenes augmentation, respectively. As in Scene-based ADA, we kept the number of augmented objects in dense and sparse scenes similar. Moreover, in the dense scenes setup, we have performed data augmentation only in dense scenes and none in sparse scenes, and vice versa for the sparse scenes setup, as listed on Table I. From the results in Table I, we can observe that both PV-RCNN and SECOND have better augmentation effects for sparse scenes except for the easy pedestrian class in PV-RCNN. We have confirmed the significance of applying adaptive augmentation according to the object distribution in the scenes and hence proposed a Scene-based ADA.

In Table II, we conducted a performance evaluation for three distinct cases: when GT Sampling was not applied, when existing GT Sampling was applied, and when DADA was applied. In the case of the car class, all of them achieved better performance than conventional GT Sampling except for SECOND's moderate car class. In particular, SECOND's easy car class improved by 0.52 AP compared to the existing GT Sampling. In addition, PV-RCNN's easy car class improved by 0.18 AP from 88.83 AP without GT Sampling to 89.01 AP with GT Sampling, and by 0.29 AP from 89.30 AP with DADA, which is a remarkable result. For the pedestrian class, the results are even more notable. Specifically, in the case of the pedestrian class, our proposed DADA method improves the performance of the easy pedestrian class by 4.93 AP compared to the traditional GT Sampling, while the traditional fixed GT Sampling in PV-RCNN has even degraded the performance compared to the one which is not applied GT Sampling.

PV-RCNN in Table II shows additional results to prove the effectiveness of the combination of Scene-based ADA and Density-based ADA in the DADA technique. Each row lists experiment results with the different setups in the following order: without GT Sampling, with conventional GT Sampling, Scene-based ADA only, Density-based ADA only, and full DADA in the PV-RCNN model. It can be observed that the performance is still excellent when only one of Scene-based or Density-based ADA is applied, and even greater when both are applied, showing the superiority of DADA.

Table III and Table IV show that our proposed DADA generalizes the excellent performance on another dataset, ONCE.

TABLE III: 3D object detection performance results evaluated on the ONCE validation dataset for the vehicle, pedestrian, and cyclist classes. The best performance values per model are bolded.

Model	Data Augmentation	0-30m			30-50m			50m-		
	Method	Vehicle	Pedestrian	Cyclist	Vehicle	Pedestrian	Cyclist	Vehicle	Pedestrian	Cyclist
PV-RCNN [19]	w/ GT Sampling	**87.75**	39.12	72.15	71.66	29.28	54.16	56.61	15.99	36.67
	w/ DADA	**87.75**	**39.62**	**73.82**	**72.15**	**29.77**	**56.52**	**58.90**	**17.25**	**37.73**
SECOND [9]	w/ GT Sampling	83.87	37.71	70.55	64.25	**27.79**	51.59	47.26	15.11	35.18
	w/ DADA	**84.02**	**38.80**	**71.08**	**65.30**	26.16	**53.59**	**52.44**	**16.93**	**35.23**

TABLE IV: Overall 3D object detection performance results about Table III. The best performance values per model are bolded.

Model	Data Augmentation	Overall				0-30m	30-50m	50m-
	Method	mAP	Vehicle	Pedestrian	Cyclist	mAP	mAP	mAP
PV-RCNN [19]	w/ GT Sampling	56.09	77.23	30.96	60.08	66.34	51.70	36.42
	w/ DADA	**57.77**	**77.64**	**33.90**	**61.76**	**67.06**	**52.81**	**37.96**
SECOND [9]	w/ GT Sampling	53.68	71.01	32.17	57.86	64.04	47.88	32.52
	w/ DADA	**54.73**	**72.79**	**32.42**	**58.99**	**64.63**	**48.35**	**34.87**

As shown in Fig. 3, the ONCE dataset is larger than KITTI, with more scenes and a wider variety of objects per scene. For this reason, DADA performs well on ONCE datasets and generalizes reliably across different classes and distances.

V. CONCLUSION

In this paper, we proposed a Dual Adaptive Data Augmentation (DADA) method to improve the performance of deep learning models in 3D object detection tasks. DADA consists of Scene-based ADA and Density-based ADA.

The experiment results show that DADA is an effective data augmentation method in the 3D object detection task. Our proposed DADA method has made a significant contribution by suggesting potential improvements that can be generalized to any 3D object detection model for the real-world applications. Future works will further expand the DADA method and explore its potential applications.

Acknowledgment

This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (23-IT-02). It was also supported by the Technology Commercialization Capacity Building Project of the Ministry of Science and ICT(2023-DG-RD-0041-01).

References

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 32(11):1231–1237, 2013.

[2] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In CVPR, pages 2446–2454, 2020.

[3] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, et al. One million scenes for autonomous driving: Once dataset. arXiv preprint arXiv:2106.11037, 2021.

[4] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017.

[5] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. MixUp: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.

[6] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In ICCV, 2019.

[7] Yunlu Chen, Vincent Tao Hu, Efstratios Gavves, Thomas Mensink, Pascal Mettes, Pengwan Yang, and Cees GM Snoek. Pointmixup: Augmentation for point clouds. In ECCV, pages 330–345, 2020.

[8] Wang Yifan, Shihao Wu, Hui Huang, Daniel Cohen-Or, and Olga Sorkine-Hornung. Patch-based progressive 3d point set upsampling. In CVPR, pages 5958–5967, 2019.

[9] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. Sensors, 18(10):3337, 2018.

[10] Martin Hahner, Dengxin Dai, Alexander Liniger, and Luc Van Gool. Quantifying data augmentation for lidar based 3d object detection. arXiv preprint arXiv:2004.01643, 2020.

[11] Petr Šebek, Šimon Pokorný, Patrik Vacek, and Tomáš Svoboda. Real3d-aug: Point cloud augmentation by placing real objects with occlusion handling for 3d detection and segmentation. arXiv preprint arXiv:2206.07634, 2022.

[12] Daeun Lee and Jinkyu Kim. Resolving class imbalance for lidar-based object detector by dynamic weight average and contextual ground truth sampling. In WACV, pages 682–691, 2023.

[13] Jin Fang, Xinxin Zuo, Dingfu Zhou, Shengze Jin, Sen Wang, and Liangjun Zhang. Lidar-aug: A general rendering-based augmentation framework for 3d object detection. In CVPR, pages 4710–4720, 2021.

[14] Jaeseok Choi, Yeji Song, and Nojun Kwak. Part-aware data augmentation for 3d object detection in point cloud, In IROS, pages 3391–3397, 2021.

[15] Wu Zheng, Weiliang Tang, Li Jiang, and Chi-Wing Fu. Se-ssd: Self-ensembling single-stage object detector from point cloud, In CVPR, pages 14494–14503, 2021.

[16] Shuyang Cheng, Zhaoqi Leng, Ekin Dogus Cubuk, Barret Zoph, Chunyan Bai, Jiquan Ngiam, Yang Song, Benjamin Caine, Vijay Vasudevan, Congcong Li, Quoc V. Le, Jonathon Shlens, and Dragomir Anguelov. Improving 3d object detection through progressive population based augmentation. In ECCV, pages 279-294, 2020

[17] Ruihui Li, Xianzhi Li, Pheng-Ann Heng, and Chi-Wing Fu. Pointaugment: An auto-augmentation framework for point cloud classification, In CVPR, pages 6378–6387, 2020.

[18] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In CVPR, pages 4490–4499, 2018.

[19] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In CVPR, pages 10529–10538, 2020.

[20] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. In AAAI, pages 1201–1209, 2021.

[21] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet, 2020.