

Diversified and Realistic 3D Augmentation via Iterative Construction, Random Placement, and HPR Occlusion

Jungwook Shin^{1, 2}, Jaeill Kim¹, Kyungeun Lee¹, Hyunghun Cho¹, Wonjong Rhee^{1, 3, 4}

¹ Department of Intelligence and Information, Seoul National University, Seoul, 08826, South Korea

² SK Telecom Co., Ltd, Seoul, 04539, South Korea

³ Interdisciplinary Program in Artificial Intelligence (IPAI), Seoul National University, Seoul, 08826, South Korea

⁴ Artificial Intelligence Institute, Seoul National University, Seoul, 08826, South Korea
{jungwook.shin, jaeill0704, ruddms0415, webofthink, wrhee}@snu.ac.kr

Abstract

In autonomous driving, data augmentation is commonly used for improving 3D object detection. The most basic methods include insertion of copied objects and rotation and scaling of the entire training frame. Numerous variants have been developed as well. The existing methods, however, are considerably limited when compared to the variety of the real world possibilities. In this work, we develop a diversified and realistic augmentation method that can flexibly construct a whole-body object, freely locate and rotate the object, and apply self-occlusion and external-occlusion accordingly. To improve the diversity of the whole-body object construction, we develop an iterative method that stochastically combines multiple objects observed from the real world into a single object. Unlike the existing augmentation methods, the constructed objects can be randomly located and rotated in the training frame because proper occlusions can be reflected to the whole-body objects in the final step. Finally, proper self-occlusion at each local object level and external-occlusion at the global frame level are applied using the Hidden Point Removal (HPR) algorithm that is computationally efficient. HPR is also used for adaptively controlling the point density of each object according to the object's distance from the LiDAR. Experiment results show that the proposed DR.CPO algorithm is data-efficient and model-agnostic without incurring any computational overhead. Also, DR.CPO can improve mAP performance by 2.08% when compared to the best 3D detection result known for KITTI dataset.

1 Introduction

3D object detection using point cloud has become an important research topic (Zhou and Tuzel 2018; He et al. 2020; Shi and Rajkumar 2020; Yang, Luo, and Urtasun 2018; Pan et al. 2021), especially in the field of autonomous driving (Geiger et al. 2013). While there has been a large improvement in designing network backbone (Zhou and Tuzel 2018; Lang et al. 2019; Yan, Mao, and Li 2018; Shi et al. 2019, 2020, 2021) and detection head (Yin, Zhou, and Krahenbuhl 2021; Zheng et al. 2021; Hu et al. 2022), there has been relatively less effort on devising augmentation methods specific to the point cloud (Fang et al. 2020, 2021; Hu et al. 2020).

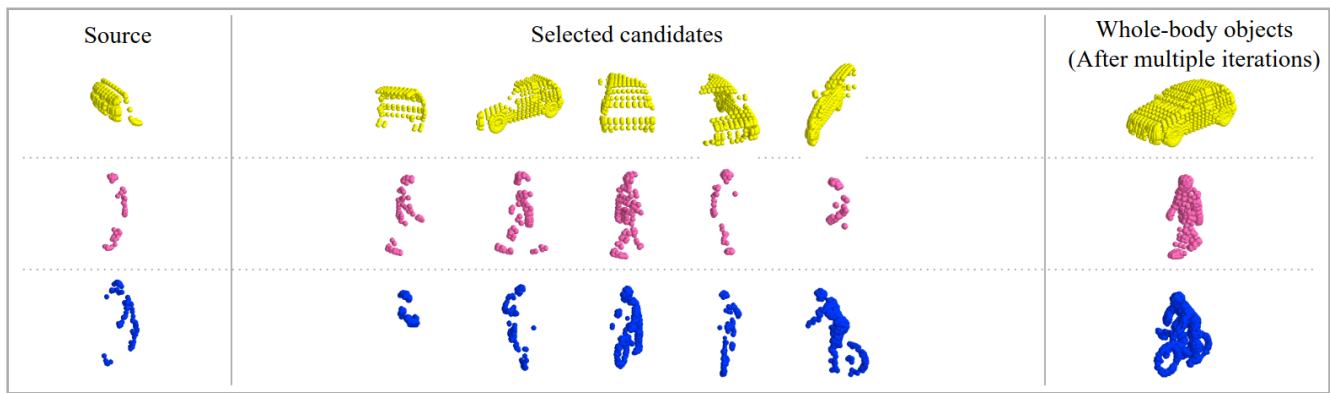
In 3D object detection, most of the models utilize global data augmentation that is a conventional method applied to

all points in a valid point cloud range. Oversampling (Yan, Mao, and Li 2018) is another conventional method that has been suggested to alleviate the extreme imbalance between the foreground and background classes. Although global augmentation and oversampling can improve the detection performance, they are straightforward methods that do not consider enhancing object-level diversity.

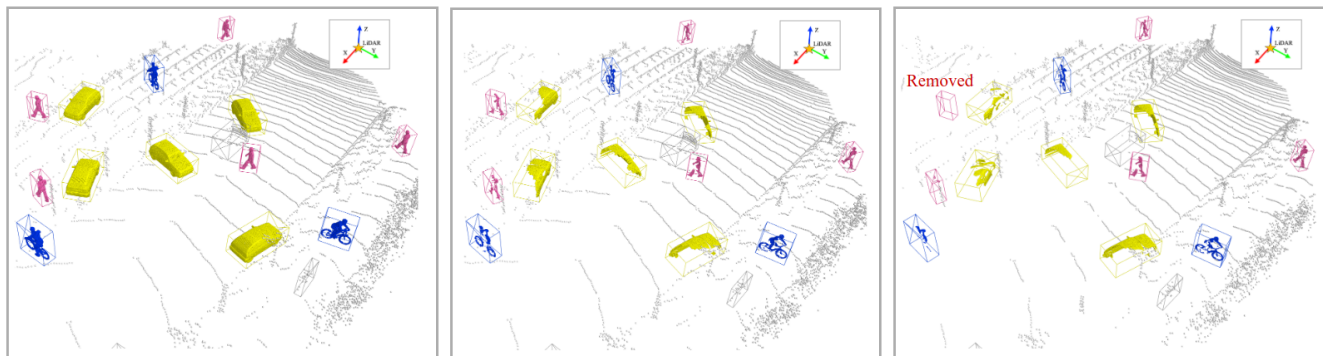
There have been efforts to provide object-level diversity by adopting 2D image augmentation methods such as random rotation and translation (Zhou and Tuzel 2018; Lang et al. 2019; Engelcke et al. 2017; Wang et al. 2019; Yang et al. 2019). The domain gap between 2D image and 3D point cloud, however, is significant and the 2D augmentation methods do not fully utilize the properties specific to 3D (Choi, Song, and Kwak 2021). Therefore, it is crucial to focus on the distinct properties of point-cloud and develop specialized augmentation methods.

Point clouds have special properties that are distinct from those of images. Most importantly, point clouds exhibit complex *occlusion* because the viewpoint from LiDAR is fixed. For images, occluded pixels can be easily found by correspondence. For point clouds, there are two obstacles for manipulating occlusions (Xu, Zhong, and Neumann 2022). At a local object level, one part of the object is occluded by another part of the same object (*self-occlusion*). At the global level, the limitation can be extended to the entire frame where a given object can be occluded by another object (*external-occlusion*). Other properties of point cloud include *point density* and *intensity*. In point cloud, the number of points per a unit volume is dependent on the distance to LiDAR and each point has its own intensity which is not easy to simulate (Yue et al. 2018).

To deal with the properties of 3D point cloud, we propose DR.CPO (**D**iversified and **R**ealistic 3D augmentation via iterative **C**onstruction, random **P**lacement, and **H**PR **O**cclusion). It can generate infinitely diverse frames with diversified objects, placements, and occlusions without requiring the expense of additional data collection or computational overhead. DR.CPO consists of three steps: (1) iterative construction, (2) random placement, and (3) HPR (Hidden Point Removal) occlusion. First, we generate a whole-body object in a stochastic manner using the objects from the real point cloud's ground-truth database. Second, we randomly rotate and place objects to anywhere inside the Li-



(a) Iterative Construction



(b) Random Placement

(c) HPR Occlusion (s-HPR)

(d) HPR Occlusion (e-HPR)

Figure 1: Illustration of DR.CPO with 3D modeling objects: (a) Iterative construction: DR.CPO constructs a whole-body 3D object via an iterative process where each source’s completion candidates are pre-selected and indexed before the model training starts. (b) Random placement: The constructed whole-body object can be randomly located anywhere in the frame with a random rotation as long as its bounding box does not overlap with an existing object. (c, d) HPR occlusion: (c) s-HPR: For each individual object, a self-occlusion is applied via HPR where the distance and visibility from the fixed LiDAR viewpoint are reflected for the randomly chosen location and rotation. (d) e-HPR: For the entire frame, an external-occlusion is applied via HPR where the inter-object spatial dependency is reflected. Some of the objects can be completely removed.

DAR detection range. Thirdly, we apply self-occlusion and external-occlusion to generate realistic views based on HPR algorithm (Katz, Tal, and Basri 2007; Mehra et al. 2010). Applying self-occlusion not only removes the invisible parts of an object from the LiDAR, but also controls the point density depending on the distance from the LiDAR. Then, applying external-occlusion affects the entire frame considering the inter-object spatial dependency. The process of DR.CPO is illustrated in Figure 1, and our contributions can be summarized as below.

- We introduce DR.CPO, a data augmentation algorithm specifically developed for 3D point clouds, to provide diversified and realistic views at the level of individual objects and at the level of entire frame.
- We develop an iterative method for constructing realistic whole-body objects where only real observations are used.
- We present a computationally light occlusion method

based on HPR. Our method can implement self-occlusion, external-occlusion, and density adjustment.

- We show that DR.CPO is data efficient, model-agnostic, and computationally efficient.
- We achieve state-of-the-art performance on the KITTI dataset. DR.CPO improves mAP performance by 2.08%.

2 Related Works

3D Object Detection: In voxel-based methods, VoxelNet (Zhou and Tuzel 2018) divides point clouds into voxels where linear network, such as PointNet (Qi et al. 2017), is then applied to convert the voxels into 3D tensors. By adopting a method of sparse convolution, SECOND (Yan, Mao, and Li 2018) alleviated the computational challenges in VoxelNet, and hence improved the speed. In point-based methods, Point-RCNN (Shi, Wang, and Li 2019) directly generates 3D proposals from raw points in a bottom-up manner

and, in the second stage, the proposals are refined by combining local spatial features and global semantic features learned in the first stage. In voxel+point based methods, PV-RCNN (Shi et al. 2020) implements a method incorporating both voxel-based operation to generate 3D proposals and point-based operation to refine the proposals. By implementing an advanced detection head on top of PV-RCNN, PV-RCNN++ (Shi et al. 2021) has become an improved framework that is more than 2x faster and requires less memory than PV-RCNN while achieving a comparable or better performance. Considering the advantages of PV-RCNN++, we use PV-RCNN++ as the default backbone in this work.

Conventional 3D Augmentations: Among the 3D data augmentation methods that are mostly extended from 2D augmentations (e.g., translation, rotation, scaling, flipping, and removal (Qi et al. 2018; Wang and Jia 2019; Shi et al. 2020; Chen et al. 2020; Zhu et al. 2019; Hahner et al. 2020; Reuse, Simon, and Sick 2021)), ground-truth sampling (GTS) (Yan, Mao, and Li 2018; Lehner et al. 2019) is a simple copy-and-paste augmentation strategy and has become popular despite its lack of consideration of occlusion between objects. Global data augmentation (GDA) (Zhou and Tuzel 2018; Hu et al. 2022) manipulates the entire frame globally and is another effective augmentation strategy (Reuse, Simon, and Sick 2021). After the combination of GTS and GDA was proposed as a default data augmentation set in (Yan, Mao, and Li 2018), the combination was widely employed in the subsequent studies (Lang et al. 2019; Shi, Wang, and Li 2019; Shi et al. 2020; He et al. 2020). In our work, we refer to the augmentation that consists of GTS and GDA as the Conventional Data Augmentation (CDA).

Constructing Whole-body Object: Learning-based approaches (Tchapmi et al. 2019; Yang et al. 2018; Xie et al. 2020) used a 3D modeling dataset including a full shape, such as Shapenet (Chang et al. 2015). However, all objects in autonomous driving dataset have incomplete shapes, including occlusions and signal misses, making a learning-based approach challenging. To use learning-based approaches, a new dataset is required for training. Rendering-based approaches (Fang et al. 2020) attempted to augment 3D objects using CAD models; however, the intensity attribute was ignored in their experiments. Fang et al. (2021) constructed shapes using raycasting and rendering based on the depth map of 3D CAD model. In these rendering-based approaches, as with data labeling, CAD models must be manually prepared. Geometry-based approaches (Xu, Zhong, and Neumann 2022) constructed an object using a heuristic score that was calculated by bounding box similarity, Chamfer distance, and overlapped voxel quantity. Inspired by this, we construct a whole-body using only geometric information without the use of external data.

Occlusion-based Augmentation: To consider the shape miss in LiDAR data caused by external-occlusion, signal miss, and self-occlusion (Xu, Zhong, and Neumann 2022), a few works have proposed occlusion-aware augmentation methods. In (Hu et al. 2020), the destroyed information

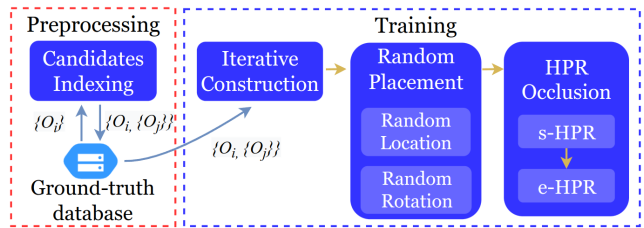


Figure 2: Pipeline of DR.CPO: In the pre-processing phase, candidates for completion are selected for each source 3D object. The chosen candidates are indexed and additionally stored in the annotated ground-truth database such that the iterative construction step can be computed efficiently. Random placement and HPR occlusion follow the iterative construction step.

about visibility in LiDAR data was recovered by raycasting algorithm. When inserting virtual objects for augmentation, a visibility constraint was used to ensure that the objects are placed where they should not be occluded. The LiDAR rendering technique automatically generates virtual objects that satisfy the occlusion constraint (Abu Alhaija et al. 2018; Fang et al. 2020). Recently, (Fang et al. 2021) rendered virtual objects with occlusion constraint and placed them in reasonable locations of real background frames using ValidMap. In our work, because we enforce self and external occlusion via HPR after placing virtual objects, they can be located anywhere without the need of concerning visibility or ValidMap.

3 Methods

DR.CPO aims to generate diversified and realistic augmentations for improving 3D object detection. The pipeline of DR.CPO is shown in Figure 2 where it consists of a pre-processing step and three training-time steps. Before presenting the details, we briefly address the notations. A point cloud (PC) of a frame can be considered as the union of foreground points (FP) and background points (BP), and it can be formulated as below:

$$PC = FP \cup BP, FP = \cup_{i=1}^N O_i. \quad (1)$$

In the equation, O_i is the i 'th object in the frame and its information can be expressed as (P_i, s_i, B_i) .

- P_i refers to a set of points, $P_i = \{p_{i,1}, \dots, p_{i,T_i}\}$, where each point $p_{i,t}$ is a 4-dimensional vector (x, y, z, r) with the first three as the 3D coordinates and r as the point's reflection intensity.
- s_i is the ground-truth class label with three possible classes: $s_i \in \{\text{Car, Pedestrian, Cyclist}\}$.
- B_i contains the annotation of O_i 's 3D bounding box. Specifically, $B_i = (cx_i, cy_i, cz_i, l_i, w_i, h_i, \theta_i)$ with the first three as the center's coordinates, the following three as the length, width, and height of the bounding box, and θ_i as the heading angle of the bounding box.

3.1 Iterative Construction

The goal of this step is to sample a source object O_i from the ground-truth database and to construct it back to a whole-body object. To do so, we select another object O_j that has the same class label ($s_i = s_j$) and can help fill up the missing points of O_i . Then, P_i is complemented by additionally including the points P_j into P_i . This complementing operation is iteratively repeated until O_i is sufficiently complemented to have P_i represent a whole-body object. An illustration of the iterative construction can be found in Figure 1(a) and visualizations for KITTI can be found in the supplementary¹ (Figure 8, 9, and 10). For a given source object O_i , we maximize the diversity of the constructed object by making the iterative construction stochastic and by performing the iterative construction on the fly during the training. The computational overhead is minimized by performing an one-time pre-processing of the ground-truth database where the set of desirable candidates for complementing, $\{O_j\}$, are indexed for each source object O_i .

Indexing Candidate Objects (Pre-processing): In this pre-processing step, we identify and index K objects as the candidates for complementing each object O_i in the database. An object O_j is considered to be a candidate for complementing O_i if it satisfies two criteria: (1) high bounding box size similarity, and (2) high partition density for the low density partitions of O_i . First, we convert all the objects into a canonical pose such that they have the same heading angle (i.e., $\theta_i = 0$ for all i). Then, we calculate the bounding box similarity with O_i for all the objects O_j ($i \neq j$) in the ground-truth database to select the best $2K$ candidates. Among the $2K$ candidates, the best K candidates that have high density partitions for O_i 's low density partitions (due to occlusion, signal miss, sparsity, etc.) are selected as the final candidates. Note that we define the partitions as the non-overlapping subdivided regions of the object O_i with a fixed number of partitions for each class. Also, we evaluate the density as the number of points per partition. Further details can be found in Supplementary A.1. $K = 400$ was used for our experiments. Partition can be considered as an extension of voxelization. In this study, we adopt the idea of partition to make it easy to analyze the density of each partition. This allows an effective method for combining multiple objects. Previously, partition was used in (Choi, Song, and Kwak 2021) for the purpose of applying a different augmentation technique to each partition.

Constructing a Whole-body Object (Training Time):

For a randomly selected source object O_i , we iteratively and stochastically construct its whole-body using the indexed K objects. The details can be summarized as the following where all steps are performed in the canonical pose. Since the car and cyclist objects are (roughly) symmetric on the x-axis in the canonical pose, (1) we first mirror their points as an additional operation before the iterative complementing. This step is skipped for the pedestrian objects. Then, (2) for each partition, we uniform randomly sample

¹Refer to our arXiv version for the Supplementary materials.

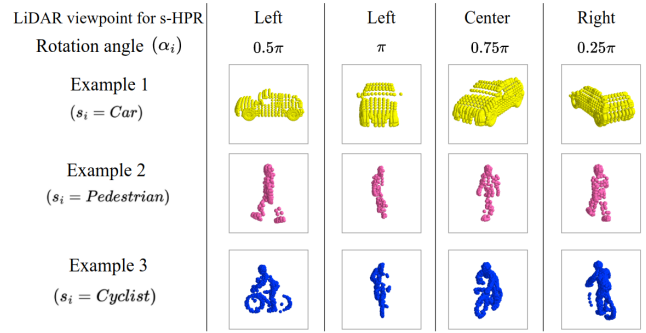


Figure 3: Diversification via placement: An iteratively constructed whole-body object can be placed at a random location, resulting in a random distance and a random viewpoint from the LiDAR. Therefore, the randomness in location provides a dimension of diversification. Also, the object can be randomly rotated and provide another dimension of diversification. Subsequently, s-HPR is applied according to the location and rotation to make the augmented point-cloud objects look realistic. The three objects constructed in Figure 1(b) were used for this illustration.

a candidate from the K candidates and add its points to P_i . (3) We iteratively repeat (2) until the object O_i can be regarded as a whole-body object. We determine the object to be a whole-body when the proportion of partitions whose density is higher than the mean is at least 85%. For a formal description and examples, see Supplementary A.1. Due to the stochastic nature of candidate sampling in (2), our method can generate *combinatorially diverse* whole-body objects for a given source object O_i .

3.2 Random Placement

In the conventional data augmentation, an object from the ground-truth database must be placed at the original location with the original angle. If not, the occlusion and density of the object does not match the placement and thus causing a negative effect for the learning. DR.CPO does not suffer from this limitation because an object from ground-truth database is first transformed into a whole-body object and because a proper occlusion and density adjustment for the chosen location and angle are reflected in the HPR occlusion step. Taking full advantage of the flexibility, we place each whole-body object randomly and maximize the diversity of the augmentation as in the example of Figure 1(b). To be specific, we first randomly rotate the object along the z-axis: $\theta_i = \alpha_i$, where α_i is uniformly sampled between -1.00π and $+1.00\pi$. Then, we select the location following a uniform distribution over the (x, y) coordinates. As we will show later, the diversity shown in Figure 3 through the randomness is essential for improving the performance and enhancing the data efficiency.

3.3 HPR Occlusion

To enforce self-occlusion and external-occlusion, we utilize Hidden Point Removal (HPR) (Katz, Tal, and Basri 2007).

We have chosen HPR mainly because of its computational efficiency. To the best of our knowledge, we are the first to adopt HPR for enforcing occlusion in autonomous driving.

Hidden Point Removal (HPR): Given a set of points, P_i , our goal is to determine if each point $p_{i,t}$ is visible from the LiDAR viewpoint C . For this purpose, we use the HPR method illustrated in Figure 11 in the supplementary. To determine whether the point $p_{i,t}$ is visible from C , HPR defines a sphere with a sufficiently large radius R centered at origin C . Spherical flipping is used to reflect every point $p_{i,t}$ internal to the sphere along the ray from C to $p_{i,t}$ to its image outside the sphere, following the equation:

$$\hat{p}_{i,t} = p_{i,t} + 2(R - \|p_{i,t}\|) \frac{p_{i,t}}{\|p_{i,t}\|} \quad (2)$$

where $\|p_{i,t}\|$ is the length of $p_{i,t}$ from C . If we define $\hat{P}_i = \{\hat{p}_{i,t}\}_{t=1}^T$, the visibility of $p_{i,t}$ is determined by whether it lies on the convex hull of $\hat{P}_i \cup \{C\}$. Rigorously speaking, we choose ϵ -visible points by adjusting R for each class. Thanks to the efficient convex hull operation, HPR is computationally outstanding while meeting the precision required for the LiDAR data augmentation. For a full description of HPR, see (Katz, Tal, and Basri 2007).

s-HPR and e-HPR for Occlusion: We first apply HPR to perform self-occlusion (s-HPR) and subsequently to perform external-occlusion (e-HPR) as shown in Figure 1(c) and 1(d). For s-HPR, HPR is applied to each individual object where the object’s location and angle are accounted for. A crucial advantage of using s-HPR is that it automatically adjusts the point density of the object depending on the distance from the LiDAR. Examples are shown in Figure 4. The density adjustment is an important feature to make the augmented object’s occlusion look realistic, and its effect on performance will be addressed in Section 5.3. Another important role of s-HPR is to reduce the number of points such that the computational burden of e-HPR can be lessened. For e-HPR, HPR is applied to the entire frame such that external-occlusion related to the inter-object dependency can be calculated. Compared to the s-HPR, this is a computationally heavy operation and the reduction in the number of points by s-HPR is beneficial for keeping e-HPR’s computation burden low. Computational efficiency of DR.CPO will be addressed in Section 4.4.

4 Experiments

In this section, we show that DR.CPO is a data-efficient and model-agnostic augmentation method that can improve the performance of voxel, point, and voxel+point models, including the state-of-the-art voxel+point model, without incurring an extra computational cost. For empirical evaluations, we use KITTI (Geiger et al. 2013) dataset that is a standard 3D LiDAR point cloud benchmark and evaluate the standard metric AP_{R40} . For a fair comparison, DR.CPO is set to follow the same class distribution as in CDA (details can be found in Table 3). Unless stated otherwise, we use PVRCNN++ (Shi et al. 2021), an open-source and state-of-the-art voxel+point method, as our default backbone model.

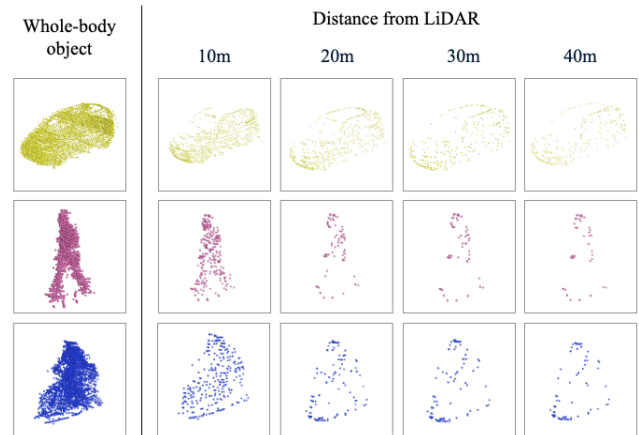


Figure 4: Density adjustment with s-HPR: Depending on the object’s distance from the LiDAR, point density is adjusted by s-HPR. Three examples are shown for the whole-body objects constructed from KITTI dataset.

Method	Car	Pedestrian	Cyclist	Mean	Diff.
SECOND	82.67	52.31	62.69	65.89	-3.64
SECOND with DR.CPO	83.15	51.67	73.75	69.53	
Point-RCNN	82.84	56.92	77.46	72.40	-1.08
Point-RCNN with DR.CPO	82.36	59.86	78.22	73.48	
PV-RCNN	86.70	57.61	75.87	73.39	-2.79
PV-RCNN with DR.CPO	84.75	62.33	81.47	76.18	

Table 1: DR.CPO is a model-agnostic augmentation method. DR.CPO can effectively enhance the performance of existing models that are based on voxel, point, and voxel+point representations. AP_{R40} performance is shown.

For training any model with DR.CPO, we trained the model by following the original protocol and code. Additional details of the implementations and experimental settings are provided in Supplementary B.

4.1 Data Efficiency

To examine the data efficiency of DR.CPO, we trained our model using subsets of KITTI’s training examples. The results are shown in Figure 5. Compared to the baseline (no augmentation) and CDA shown in gray and blue dotted lines, respectively, DR.CPO can match their performance with a much smaller data size. An exception is the class of car for CDA, and this is discussed in Section 6. For the overall performance, DR.CPO can perform as well as the baseline and CDA using only 4.60% and 37.60% of the training dataset, respectively.

4.2 Model-Agnostic Effectiveness

Because DR.CPO is an augmentation method, we have chosen three latest and well-known models and examined if

Method	Car			Pedestrian			Cyclist			Overall			Mean	Diff.
	E	M	D	E	M	D	E	M	D	Car	Ped	Cyc		
ContFuse	83.68	68.78	61.67	-	-	-	-	-	-	71.38	-	-	-	
3D IoU Loss	84.43	76.28	68.22	-	-	-	-	-	-	76.31	-	-	-	
SECOND with LiDAR-Aug	88.65	76.97	70.44	58.53	54.56	51.78	-	-	-	78.69	54.96	-	-	
PointPillars with LiDAR-Aug	87.75	77.83	74.90	59.99	55.15	52.66	-	-	-	80.16	55.93	-	-	
PointRCNN with LiDAR-Aug	89.56	79.51	77.89	67.46	59.06	56.23	-	-	-	82.32	60.92	-	-	
PV-RCNN with LiDAR-Aug	90.18	84.23	78.95	65.05	59.90	55.52	-	-	-	84.45	60.16	-	-	
Part-A2-Net PV-RCNN with STRL	91.70	87.79	84.61	-	-	-	81.91	68.12	61.92	88.03	-	70.65	-	
VoxelNet	81.97	65.46	62.85	57.86	53.42	48.87	67.17	47.65	45.11	70.09	53.38	53.31	58.93	-19.37
StarNet	81.63	73.99	67.07	48.58	41.25	39.66	73.14	58.29	52.58	74.23	43.16	61.34	59.58	-18.72
PointPillars with PA-AUG	83.70	72.48	68.23	57.38	51.85	46.91	70.88	47.58	44.80	74.80	52.05	54.42	60.42	-17.87
PV-RCNN with PA-AUG	89.38	80.90	78.95	67.57	60.61	56.58	86.56	72.21	68.01	83.08	61.59	75.59	73.42	-4.88
StarNet with PPBA	84.16	77.65	71.21	52.65	44.08	41.54	79.42	61.99	55.34	77.67	46.09	65.58	63.12	-15.18
Point-GNN	93.11	89.17	83.90	55.36	47.07	44.61	81.17	67.28	59.67	88.73	49.01	69.37	69.04	-9.26
HotSpotNet	87.60	78.31	73.34	82.59	65.95	59.00	53.10	45.37	41.47	79.75	69.18	46.65	65.19	-13.11
Object as Hotspots (Dense)	91.09	82.20	79.69	85.85	66.45	62.16	68.88	62.82	55.78	84.33	71.49	62.49	72.77	-5.53
STD	94.74	89.19	86.42	60.02	48.72	44.55	81.36	67.23	59.35	90.12	51.10	69.31	70.18	-8.12
3DSSD	92.66	89.02	85.86	60.54	49.94	45.73	85.04	67.62	61.14	89.18	52.07	71.27	70.84	-7.46
PointRCNN	90.10	80.41	78.00	64.18	56.71	49.86	91.72	72.47	68.18	82.84	56.92	77.46	72.40	-5.89
PV-RCNN with Pattern-Aware GT	92.13	84.79	82.56	65.99	58.57	53.66	90.38	72.03	67.96	86.49	59.41	76.79	74.23	-4.07
SECOND	90.97	79.94	77.09	58.01	51.88	47.05	78.50	56.74	52.83	82.67	52.31	62.69	65.89	-12.41
PointPillars	87.75	78.39	75.18	57.30	51.41	46.87	81.57	62.94	58.98	80.44	51.86	67.83	66.71	-11.59
BtcDet	93.15	86.28	83.86	69.39	61.19	55.86	91.45	74.70	70.08	87.76	62.15	78.74	76.22	-2.08
PV-RCNN	92.57	84.83	82.69	64.26	56.67	51.91	88.88	71.95	66.78	86.70	57.61	75.87	73.39	-4.90
PV-RCNN++ (Reproduced)	91.72	84.82	82.03	66.00	60.05	54.70	91.39	70.14	66.92	86.19	60.25	76.15	74.20	-4.10
PV-RCNN++ with DR.CPO (ours)	91.27	81.67	81.28	73.23	67.66	62.96	93.08	78.21	75.32	84.74	67.95	82.20	78.30	

Table 2: Compilation of KITTI 3D LiDAR detection performance. State-of-the-art models are also included. For each column, any model performance that is superior to ours is indicated in bold.

their performance can be enhanced with DR.CPO. The results are shown in Table 1. In the table, SECOND is a voxel-based method, Point-RCNN is a point-based method, and PV-RCNN is a voxel+point-based method. The improvements from DR.CPO are 3.64%, 1.08%, and 2.79%, respectively.

4.3 Comparison With State-of-the-Art Models

Because of the improved detection head and the computationally efficient representative keypoints, we have used PV-RCNN++ as the backbone of DR.CPO and achieved mAP score of 78.30%. We have compiled the previously reported records as much as we can into Table 2, where the latest models such as SECOND, PointPillars, BtcDet, PV-RCNN, PV-RCNN++ are also included. In particular, BtcDet (Xu, Zhong, and Neumann 2022), which integrates multiple techniques including an extra shape occupancy net-

work, is known to be the current state-of-the-art model. Simply by incorporating DR.CPO into the augmentation pipeline, the computationally efficient PV-RCNN++ can perform better than all the other models shown in Table 2 including BtcDet. As in the data efficiency and model-agnostic effectiveness studies, the gain comes from the extraordinary improvements in pedestrian and cyclist categories. The improvements are large enough to overcome the loss in car category. We discuss the loss of car category in Section 6.

4.4 Computational Efficiency

For all of our experiments, DR.CPO was set to follow the same class distribution as in CDA for a fair comparison. This can be confirmed from the average object counts per frame in Table 3. Even with similar object counts, DR.CPO needs to perform additional computations including iterative construction and HPR occlusion. The computational efficiency

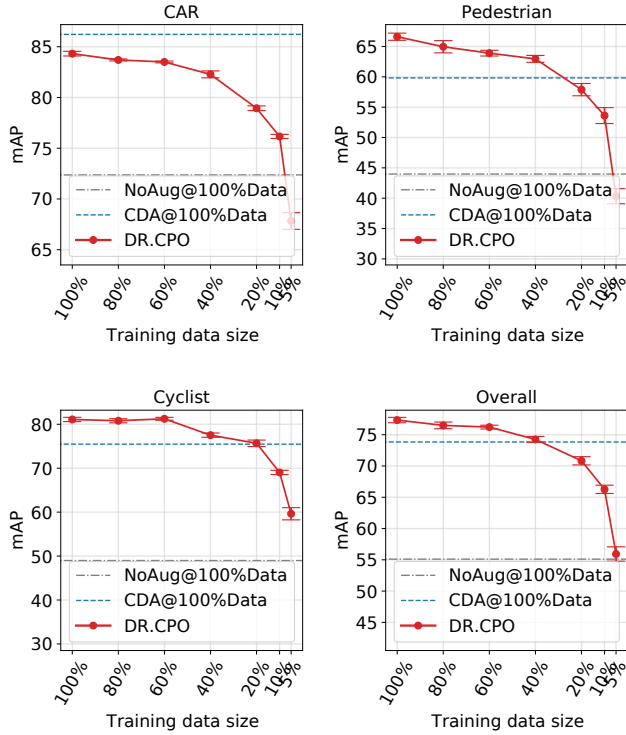


Figure 5: Data efficiency of DR.CPO: The training data size versus mAP score is plotted for KITTI. Except for car, DR.CPO achieves significant improvements over no-augmentation and CDA.

in terms of training time per epoch, however, is not compromised as can be seen in the same table. In fact, DR.CPO is slightly faster than CDA. Despite the overhead of our augmentation method, the training time is more important than the data preparation time and DR.CPO requires less training time thanks to the reduced number of points after s-HPR and e-HPR. Overall, DR.CPO does not incur any additional burden when compared to the ubiquitously used CDA.

5 Analysis of DR.CPO

In this section, we analyze DR.CPO’s three elements. When analyzing an element, the other two are kept the same as in the default setup of DR.CPO.

5.1 Iterative Construction

We have examined if constructing a whole-body object can indeed positively influence the performance, and the results are shown in Table 4. By looking at the mean performance column, it can be confirmed that self-mirroring and iteration for complementing are both helpful. The gain from mirroring is only 0.11% (the difference between the first and second rows). The performance gain is mostly due to the iteration process, and it is 95% of the total gain. From the car column, it can be seen that multiple iterations can actually hurt the performance of the car detection. For the cyclist

		No Aug.	CDA	Ours
Average counts per frame	Car	4.00	13.70	13.40
	Pedestrian	0.60	11.80	12.90
	Cyclist	0.20	10.30	11.00
Total points per frame		18912	23170	17203
Training time per epoch (sec)	Data prep.	2.10	4.11	9.55
	Computation	124.10	134.26	117.21
	Total	126.10	138.37	129.40
Performance (mAP)		55.06	73.48	78.30

Table 3: Object counts, the number of points, and training time per epoch. Our method increases objects per frame without increasing the total points due to s-HPR and e-HPR operations. Overall training time does not increase because the delay in time on data preparation (Data prep.) is offset by reduced time on processing model (Computation).

Mirroring	Iteration	Car	Pedestrian	Cyclist	Mean	Diff.
No	None	85.08	63.01	79.75	75.95	-2.35
Yes	None	85.18	63.31	79.65	76.06	-2.24
Yes	Single	85.25	67.85	80.57	77.89	-0.41
Yes	Multiple	84.74	67.95	82.20	78.30	

Table 4: Iterative construction: the effect of mirroring and iteration.

Random Location	Random Rotation	Car	Pedestrian	Cyclist	Mean	Diff.
		84.31	65.58	79.85	76.58	-1.72
✓		84.28	67.01	79.32	76.87	-1.43
	✓	85.33	66.00	83.16	78.16	-0.14
✓	✓	84.74	67.95	82.20	78.30	

Table 5: Random placement: ablation study of location and rotation.

column, however, the gain from single iteration to multiple iteration is large (1.63%).

5.2 Random Placement

To analyze the effects of random location and random rotation, we have performed an ablation study and the results can be found in Table 5. While both are helpful as expected, the random location has improved the mean performance by 0.22% while random rotation has improved it by 1.51%. Apparently, the additional diversity from rotation is more effective than the additional diversity from random location. To further investigate random rotation, we have analyzed the effect of rotation range and the results are shown in Table 6. It can be observed that generally a larger rotation range is better. It is interesting to note that the largest gain is obtained when the rotation range is increased from $(-0.75\pi, +0.75\pi)$ to $(-1.00\pi, +1.00\pi)$, where the gain is 1.02%. Rotation of 1.00π corresponds to switching the front and the back sides of an object, and thus inverting the direction of movement.

Rotation range	Car	Pedestrian	Cyclist	Mean	Diff.
$(-0.25\pi, +0.25\pi)$	85.13	64.41	80.87	76.81	-1.49
$(-0.50\pi, +0.50\pi)$	85.36	64.99	80.46	76.94	-1.36
$(-0.75\pi, +0.75\pi)$	84.56	65.67	81.61	77.28	-1.02
$(-1.00\pi, +1.00\pi)$	84.74	67.95	82.20	78.30	

Table 6: Random placement: the effect of rotation range.

s-HPR	e-HPR	Car	Pedestrian	Cyclist	Mean	Diff.
		84.64	52.33	78.94	71.97	-6.33
	✓	84.98	64.83	80.22	76.68	-1.62
✓		85.47	63.58	83.05	77.37	-0.93
✓	✓	84.74	67.95	82.20	78.30	

Table 7: HPR occlusion: ablation study of s-HPR and e-HPR. s-HPR is responsible for density adjustment in addition to self-occlusion.

5.3 HPR Occlusion

To understand the effects of s-HPR and e-HPR, an ablation study was performed and the results are shown in Table 7. While the effectiveness of both can be confirmed, the impact is larger when s-HPR is removed (-1.62%) than when e-HPR is removed (-0.93%). This can be explained by s-HPR’s additional functionality of adjusting density as explained in Figure 4. When both s-HPR and e-HPR are turned off, the mean performance becomes 71.97% and it is even worse than the CDA baseline of 74.20% that can be found in Table 2 (mean performance of PV-RCNN++). This indicates that it is indeed undesirable to augment with whole-body objects and do not reflect the effects of occlusion. An augmented object becomes unrealistic without a proper occlusion, and it is better to avoid iterative construction and random placement when the occlusion step does not exist.

6 Discussion

Diversified and Realistic Augmentation by DR.CPO: The conventional data augmentation method is constrained in terms of the augmentation diversity it can provide - it relies on insertion of copied objects and rotation and scaling of the entire training frame. Compared to CDA, our method can provide diversified augmentations mainly via combinatorially stochastic nature of whole-body construction, random location, and random rotation. Their effectiveness for enhancing performance has been shown in Table 4, 5, and 6. Illustrations of diversification via placement can be found in Figure 3 and Figure 6. Our method is also realistic in a few ways. Unlike the synthesis or rendering approaches, we construct the whole-body objects using only the real LiDAR observations collected with the real LiDAR system. Our approach can be relatively noisy but the constructed objects are guaranteed to be realistic. Another factor that makes our augmentation realistic is the adoption of HPR. It enables application of occlusion without any concern on the computational burden. In consequence, random placements of whole-body objects can become realistic by applying self-occlusion and external-occlusion with HPR. Despite

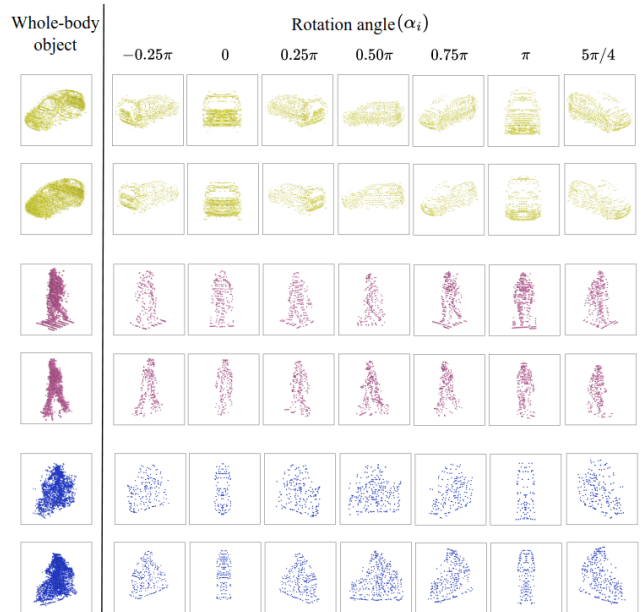


Figure 6: Diversity of the occluded shapes: Applying s-HPR to the randomly rotated objects results in a wide range of shapes depending on the object’s rotation angle. For the above examples, distance is fixed at 10m and LiDAR viewpoint is always at the center.

the advantages listed above, we were not able to consider the environment dependent distributions of the object location, heading angle, object class, object shape, etc. We have simply employed uniform random strategies, and this remains as a limitation of our work.

Performance of Car Detection: Our method provides an excellent performance for pedestrian and cyclist detection as can be seen in Table 2. The performance for car detection, however, is relatively less desirable. From the analysis tables in Section 5, it can be observed that the car detection performance can be easily improved by sacrificing the performance of the other two categories. For instance, the car performance can be improved from 84.74% to 85.47% by disabling e-HPR (see the last two rows of Table 7). Similar observations can be made from Table 4, 5, and 6 where customizing the strategies for car detection can be beneficial for improving car detection. In fact, we can even achieve 86.43% of car detection performance by applying a simple yet effective modification. In Table 14 of Supplementary E, we are showing the results for including three additional car objects into each scene. With this modification, the car performance is improved at the cost of the pedestrian performance, where the mean performance remains almost the same (slightly improved from 78.30% to 78.34%) and the computational overhead also remains almost the same (the total points per frame only marginally increases from 17,203 to 17,628).

Despite the obvious solutions for improving the car detection performance, we have focused on improving the mean

performance only because our goal of this work is to develop a general augmentation scheme for any class of object. We didn't want to focus on tuning or developing class-dependent techniques. Adjusting the object numbers in each training scene (e.g., three more cars) is a way of controlling the performance trade-off over the three object categories, and thus it is presented only for the purpose of discussion. In this work, we have considered the mean performance as the only genuine metric of performance. As for the general augmentation schemes, it certainly remains open to investigate if a general modification, such as manipulating intensity of augmented points, can improve the car detection performance without harming the pedestrian and cyclist detection performance.

Waymo dataset: In the stage of iterative construction, we fill up the missing points of an object by adding points of other objects of the same class. It means that the objects of the same class should have similar shapes and characteristics for a successful construction. Unfortunately, the objects of Waymo dataset are labeled in a quite coarse manner. For instance, motorcycles and motorcyclists share the label of vehicle. Because of the need for additional labeling, we did not evaluate Waymo dataset. Nonetheless, DR.CPO should be effective for any dataset with decently refined categorization.

7 Conclusion

In this study, we proposed an augmentation method that was developed specifically for LiDAR datasets. Thanks to its diversified and realistic nature, DR.CPO can perform on par with no-augmentation and CDA augmentation using only 4.60% and 37.60% of the training data, respectively. DR.CPO is model-agnostic and can enhance the performance for voxel-based, point-based, and voxel+point-based models. DR.CPO also provides a superior performance without incurring an additional burden in computation. Experiment and analysis results show that DR.CPO is an excellent solution for pedestrian and cyclist detection but that there is a room for improvement for car detection.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A2C2007139) and in part by IITP grant funded by the Korea government (MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)].

References

Abu Alhaija, H.; Mustikovela, S. K.; Mescheder, L.; Geiger, A.; and Rother, C. 2018. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 126(9): 961–972.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Chen, Q.; Sun, L.; Wang, Z.; Jia, K.; and Yuille, A. 2020. Object as hotspots: An anchor-free 3d object detection approach via firing of hotspots. In *European conference on computer vision*, 68–84. Springer.

Choi, J.; Song, Y.; and Kwak, N. 2021. Part-aware data augmentation for 3d object detection in point cloud. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3391–3397. IEEE.

Engelcke, M.; Rao, D.; Wang, D. Z.; Tong, C. H.; and Posner, I. 2017. Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 1355–1361. IEEE.

Fang, J.; Zhou, D.; Yan, F.; Zhao, T.; Zhang, F.; Ma, Y.; Wang, L.; and Yang, R. 2020. Augmented lidar simulator for autonomous driving. *IEEE Robotics and Automation Letters*, 5(2): 1931–1938.

Fang, J.; Zuo, X.; Zhou, D.; Jin, S.; Wang, S.; and Zhang, L. 2021. LiDAR-Aug: A General Rendering-based Augmentation Framework for 3D Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4710–4720.

Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11): 1231–1237.

Hahner, M.; Dai, D.; Liniger, A.; and Van Gool, L. 2020. Quantifying data augmentation for lidar based 3d object detection. *arXiv preprint arXiv:2004.01643*.

He, C.; Zeng, H.; Huang, J.; Hua, X.-S.; and Zhang, L. 2020. Structure aware single-stage 3d object detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11873–11882.

Hu, P.; Zigar, J.; Held, D.; and Ramanan, D. 2020. What you see is what you get: Exploiting visibility for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11001–11009.

Hu, Y.; Ding, Z.; Ge, R.; Shao, W.; Huang, L.; Li, K.; and Liu, Q. 2022. Afdetv2: Rethinking the necessity of the second stage for object detection from point clouds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 969–979.

Katz, S.; Tal, A.; and Basri, R. 2007. Direct visibility of point sets. In *ACM SIGGRAPH 2007 papers*, 24–es.

Lang, A. H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; and Beijbom, O. 2019. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12697–12705.

Lehner, J.; Mitterecker, A.; Adler, T.; Hofmarcher, M.; Nessler, B.; and Hochreiter, S. 2019. Patch Refinement–Localized 3D Object Detection. *arXiv preprint arXiv:1910.04093*.

Mehra, R.; Tripathi, P.; Sheffer, A.; and Mitra, N. J. 2010. Visibility of noisy point cloud data. *Computers & Graphics*, 34(3): 219–230.

- Pan, X.; Xia, Z.; Song, S.; Li, L. E.; and Huang, G. 2021. 3d object detection with pointformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7463–7472.
- Qi, C. R.; Liu, W.; Wu, C.; Su, H.; and Guibas, L. J. 2018. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 918–927.
- Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.
- Reuse, M.; Simon, M.; and Sick, B. 2021. About the Ambiguity of Data Augmentation for 3D Object Detection in Autonomous Driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 979–987.
- Shi, S.; Guo, C.; Jiang, L.; Wang, Z.; Shi, J.; Wang, X.; and Li, H. 2020. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10529–10538.
- Shi, S.; Jiang, L.; Deng, J.; Wang, Z.; Guo, C.; Shi, J.; Wang, X.; and Li, H. 2021. PV-RCNN++: Point-voxel feature set abstraction with local vector representation for 3D object detection. *arXiv preprint arXiv:2102.00463*.
- Shi, S.; Wang, X.; and Li, H. 2019. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 770–779.
- Shi, S.; Wang, Z.; Wang, X.; and Li, H. 2019. Part-a² net: 3d part-aware and aggregation neural network for object detection from point cloud. *arXiv preprint arXiv:1907.03670*, 2(3).
- Shi, W.; and Rajkumar, R. 2020. Point-gnn: Graph neural network for 3d object detection in a point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1711–1719.
- Tchapmi, L. P.; Kosaraju, V.; Rezatofighi, H.; Reid, I.; and Savarese, S. 2019. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 383–392.
- Wang, Z.; Fu, H.; Wang, L.; Xiao, L.; and Dai, B. 2019. SC-Net: Subdivision coding network for object detection based on 3D point cloud. *IEEE Access*, 7: 120449–120462.
- Wang, Z.; and Jia, K. 2019. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3d object detection. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1742–1749. IEEE.
- Xie, H.; Yao, H.; Zhou, S.; Mao, J.; Zhang, S.; and Sun, W. 2020. Grnet: Gridding residual network for dense point cloud completion. In *European Conference on Computer Vision*, 365–381. Springer.
- Xu, Q.; Zhong, Y.; and Neumann, U. 2022. Behind the curtain: Learning occluded shapes for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2893–2901.
- Yan, Y.; Mao, Y.; and Li, B. 2018. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10): 3337.
- Yang, B.; Luo, W.; and Urtasun, R. 2018. Pixor: Real-time 3d object detection from point clouds. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 7652–7660.
- Yang, Y.; Feng, C.; Shen, Y.; and Tian, D. 2018. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 206–215.
- Yang, Z.; Sun, Y.; Liu, S.; Shen, X.; and Jia, J. 2019. Std: Sparse-to-dense 3d object detector for point cloud. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1951–1960.
- Yin, T.; Zhou, X.; and Krahenbuhl, P. 2021. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11784–11793.
- Yue, X.; Wu, B.; Seshia, S. A.; Keutzer, K.; and Sangiovanni-Vincentelli, A. L. 2018. A lidar point cloud generator: from a virtual world to autonomous driving. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, 458–464.
- Zheng, W.; Tang, W.; Chen, S.; Jiang, L.; and Fu, C.-W. 2021. Cia-ssd: Confident iou-aware single-stage object detector from point cloud. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 3555–3562.
- Zhou, Y.; and Tuzel, O. 2018. Voxnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4490–4499.
- Zhu, B.; Jiang, Z.; Zhou, X.; Li, Z.; and Yu, G. 2019. Class-balanced grouping and sampling for point cloud 3d object detection. *arXiv preprint arXiv:1908.09492*.