CSOT: Cross-Scan Object Transfer for Semi-Supervised LiDAR Object Detection (Supplementary Materials)

Jinglin Zhan[1], Tiejun Liu[1], Rengang Li[1], Zhaoxiang Zhang[2], and Yuntao Chen[3⋆]

[1] IEIT Systems
[2] Institute of Automation, CAS
[3] Centre for Artificial Intelligence and Robotics, HKISI, CAS
{zhanjinglin,liutj,lirg}@ieisystem.com,
zhaoxiang.zhang@ia.ac.cn, chenyuntao08@gmail.com

A Additional Informative Ablations.

A.1 The orthogonality of CSOT with self-training

To eliminate the potential gains of iterative self-training, we perform an additional cycle of self-training on both self-trained and our partially-supervised detectors. Results in Tab. A validate that the benefits of pseudo-labeling methods and our proposed CSOT approach are independent and complementary. Self-Training 2× schd (Tab. A, Line 3) refers to utilizing the pretrained Self-Training 1× schd (Tab. A, Line 2) detector as the teacher model to provide pseudo-labels for unlabeled scans, followed by an additional cycle of self-training. As shown in the 2th and 4th Lines of Tab. A, generalizing detectors on unlabeled point clouds via pseudo-labeling SSOD paradigm (self-training) or our CSOT SSOD paradigm bring an improvement of 4.5 NDS and approximate 7 mAP over the baseline. As shown in the 3th and 5th Lines of Tab. A, iterative employment of self-training techniques does not lead to sustained gains in detection performance, while incorporating just one cycle of self-training on CSOT is able to further optimize the NDS from 61.0 to 64.8 and mAP from 50.4 to 56.4.

A.2 Ablate spatial-aware classification loss

As described in Sec. 3.3 of the main text, we propose a spatial-aware classification loss for our partial supervision to handle false negative issues caused by treating all unlabeled objects as backgrounds. The proposed spatial-aware classification loss and the corresponding randomly sampling strategy are ablated in Tab. B. Masking regions without supervision signals for partially-labeled scans boost mAP by 1.2. Randomly sampling 1% negative samples in heatmap further increase mAP from 55.7 to 56.4.

⋆ Corresponding author.

Table A: Ablate the orthogonality of CSOT with Self-Training technique. Detectors are trained with 5% annotations in nuScenes dataset and all results are evaluated on nuScenes validation split.

	Annos	Scans	Training Details	mAP	NDS
1	5%	5%	Fully-Supervised (Baseline)	43.8	56.5
2	5%	100%	1 + Self-Training 1x schd	51.0	61.0
3	5%	100%	1 + Self-Training 2x schd	50.7	60.7
4	5%	100%	CSOT (Ours)	50.4	61.0
5	5%	100%	4 + Self-Training 1x schd	56.4	64.8

Table B: Ablate the proposed spatial-aware classification loss. Detectors are trained with 5% annotations in nuScenes dataset and all results are evaluated on nuScenes validation split.

	L_{fully}	L_{local} w/o sample	L_{local} w/ sample
mAP / NDS	54.5 / 63.8	55.7 / 64.8	56.4 / 64.8

A.3 Run-to-Run variance

As shown in Tab. C, the performance of semi-supervised detectors trained with CSOT SSOD paradigm has low run-to-run variance. Repeating experiments on nuScenes 3 times yields mAPs of 56.4/56.3/56.3, and NDS of 64.8/64.9/64.9.

Table C: Ablate Run-to-run variance. Detectors are trained with 5% annotations in nuScenes dataset for three times. All results are evaluated on nuScenes validation split.

	1st exp	2nd exp	3rd exp
mAP / NDS	56.4 / 64.8	56.3 / 64.9	56.3 / 64.9

B Visualization

B.1 Qualitative comparison of object transfer approaches

Generally, object recognition can be categorized into two types: using the object's appearance itself for identification, or leveraging surrounding semantic information to make inferences. In this work, we identifies the importance of contextual consistency maintenance for using copy-paste in LiDAR SSOD, and proposes a novel object transfer method sustaining both local and global contextual consistency. Comparing with previous object transfer approaches, we are the only one that considers global semantic consistency and uses learned-based methods for object transfer, and thus set various new SoTA results across detectors and datasets. Detailed results are listed in Tab. 5 in the main text and visualized in Fig. A in this supplementary material.

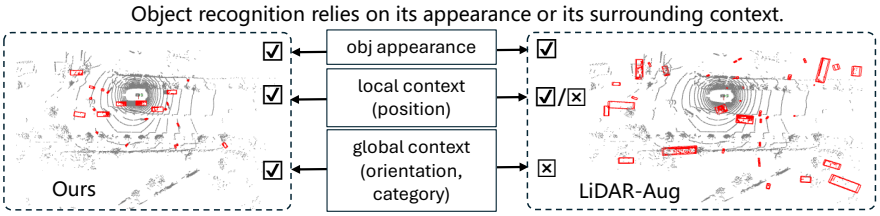

Fig. A: Visualize different object transfer approaches for providing qualitative justification and analysis of our methods.

B.2 Detection results

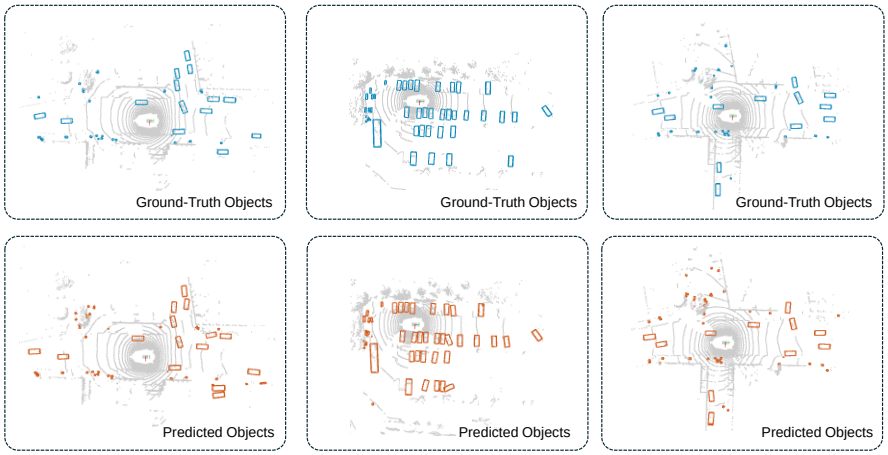

Fig. B: Visualize predictions of our semi-supervised detector trained with 5% annotations in nuScenes dataset. The ground-truth and predicted objects are displayed with blue and red boxes respectively.

The detection results of our semi-supervised detector trained with 5% of annotations from the nuScenes dataset are visualized in Fig. B. The ground truth objects are depicted with blue bounding boxes in the top row, and the predictions from our model are shown with red bounding boxes in the bottom row. Impressively, our semi-supervised detector obtained with a massive 20x reduction in labeled data is still able to achieve relatively accurate predictions, even for more challenging cases that are in the distance or are small in size.

B.3 Precision, recall and P-R Curves

We leverage Precision-Recall/Recall-Confidence/Precision-Confidence curves to give a deeper insight into the performance difference between detectors trained

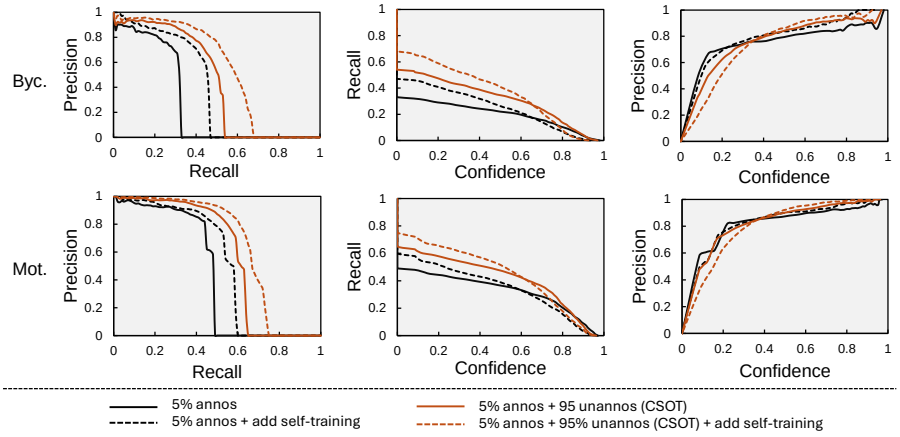

Fig. C: Compare PR-curve, Recall-Confidence curve and Precision-Confidence curve of different detectors trained with 5% annotations in nuScenes dataset. We select two categories with the lowest frequency of occurrence in the nuScenes dataset and complex morphology as representative examples: bicycles and motorcycles.

with self-training and the proposed CSOT SSOD method. We select two categories with the lowest frequency of occurrence in the nuScenes dataset and complex morphology as representative examples: bicycles and motorcycles. As visualized in Fig. C, the proposed CSOT SSOD paradigm demonstrates strong effectiveness at handling these rare and complex objects.

C Limitations and Future Work

We believe the CSOT SSOD paradigm demonstrate significant promise to reduce annotation costs for 3D object detection. In the future, we will conduct a comprehensive analysis of the detector's performance in relation to increased annotations and unlabeled LiDAR scans, which may provide valuable insights into how data quantity and diversity influence model optimization.

There are still many opportunities to build upon and advance the proposed CSOT SSOD paradigm, such as combining with advanced self-training techniques, or exploring effective strategies for expanding the ground-truth object database. For example, the ground-truth database can be enriched by some object-level data augmentation techniques. Meanwhile, the proposed object transfer approach allows for convenient leverage of external well-annotated objects to help mitigate the lack of annotations. The newly generative AI techniques also offer the potential to enable unlimited expansion of annotated objects and unlabeled scans used in our CSOT, which represents an exciting direction for future work to further push the boundaries of scale and cost-effectiveness in 3D object detection. The aforementioned directions will be a focus of our future research.