Paper ID: 916

DA-CIL: Towards Domain Adaptive Class-Incremental 3D Object Detection

Ziyuan Zhao[1,2], Mingxi Xu[1,3], Peisheng Qian[1], Ramanpreet Singh Pahwa[1,2], Richard Chang[1]

[1] Institute for Infocomm Research (I2R), A*STAR, Singapore
[2] Artificial Intelligence Analytics & Informatics (AI3), A*STAR, Singapore
[3] Nanyang Technological University, Singapore

Introduction

- **Background**

  Deep learning-based 3D object detection has received considerable attention in 3D point clouds.

- **Challenges**

  Catastrophic Forgetting: The performance of DL models on old classes tends to decrease substantially when trained on novel classes.

  Domain Shift: DL models trained on one domain (i.e., source domain) always suffer tremendous performance degradation when evaluated on another domain (i.e., target domain).

We identify a new CIL scenario where domain shift occurs when adapting new classes across domains and formulate a new CIL paradigm to enable domain adaptive class-incremental learning for 3D object detection.

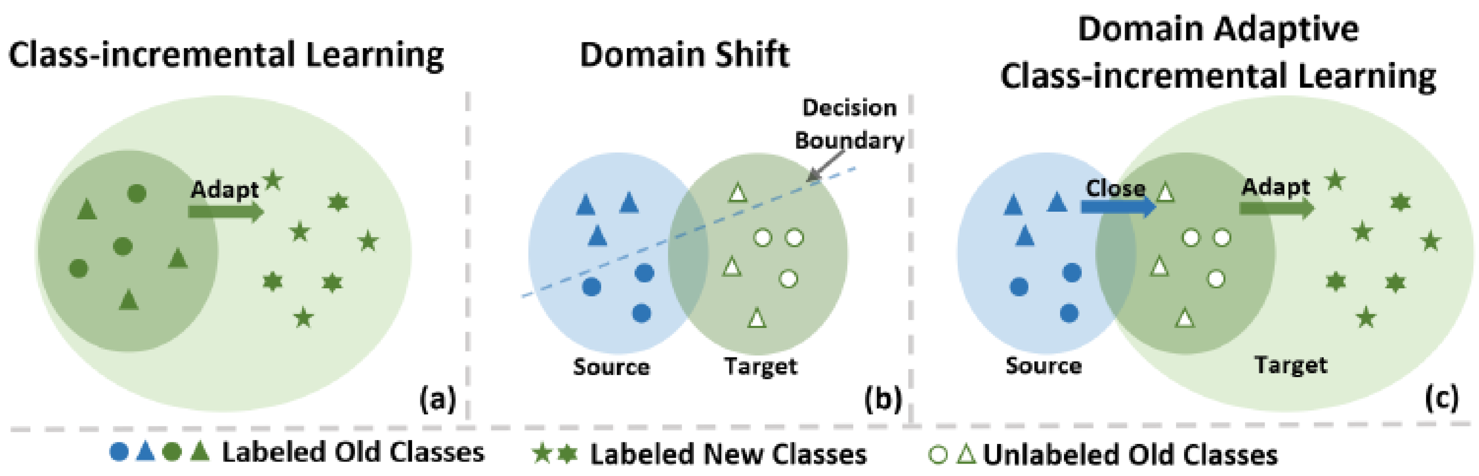
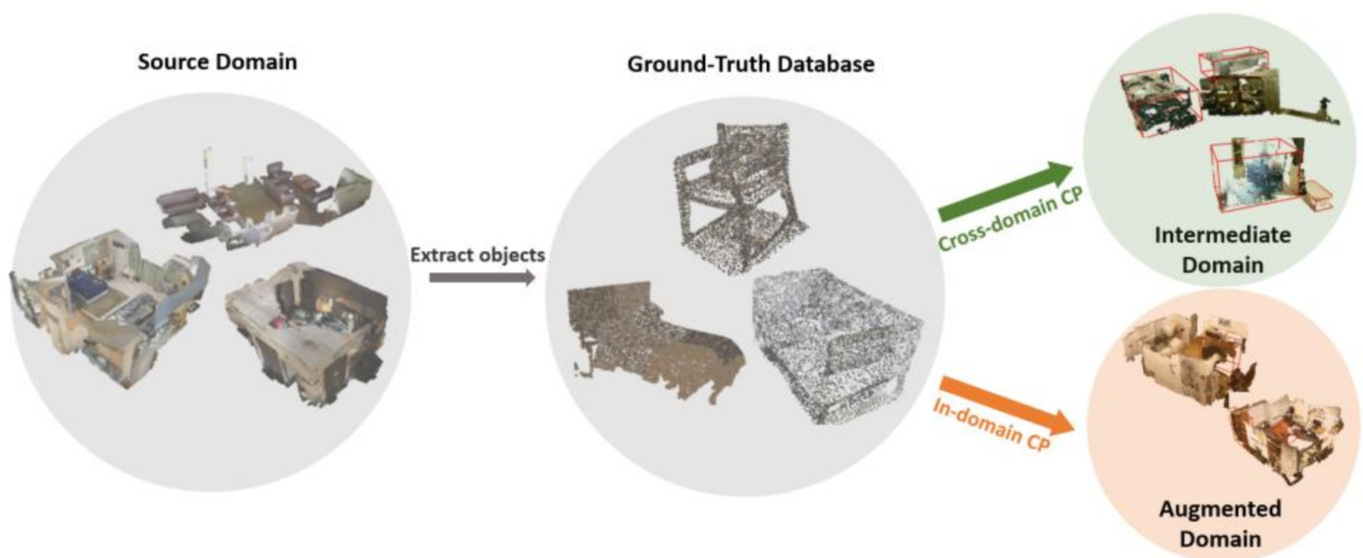


Figure 2: The difference between CIL and DA-CIL. (a) CIL is designed for adapting new classes without forgetting old classes in the same domain. (b) Domain shift always exists between source and target domains. (c) DA-CIL is designed to close the domain shift and adapt new classes across different domains.

Method

- **Dual-Domain Copy-Paste Augmentation**

- To relieve data scarcity and reduce the domain gap at the data level, we extensively leverage copy-paste (CP) augmentation techniques for creating cross-domain and in-domain point clouds.



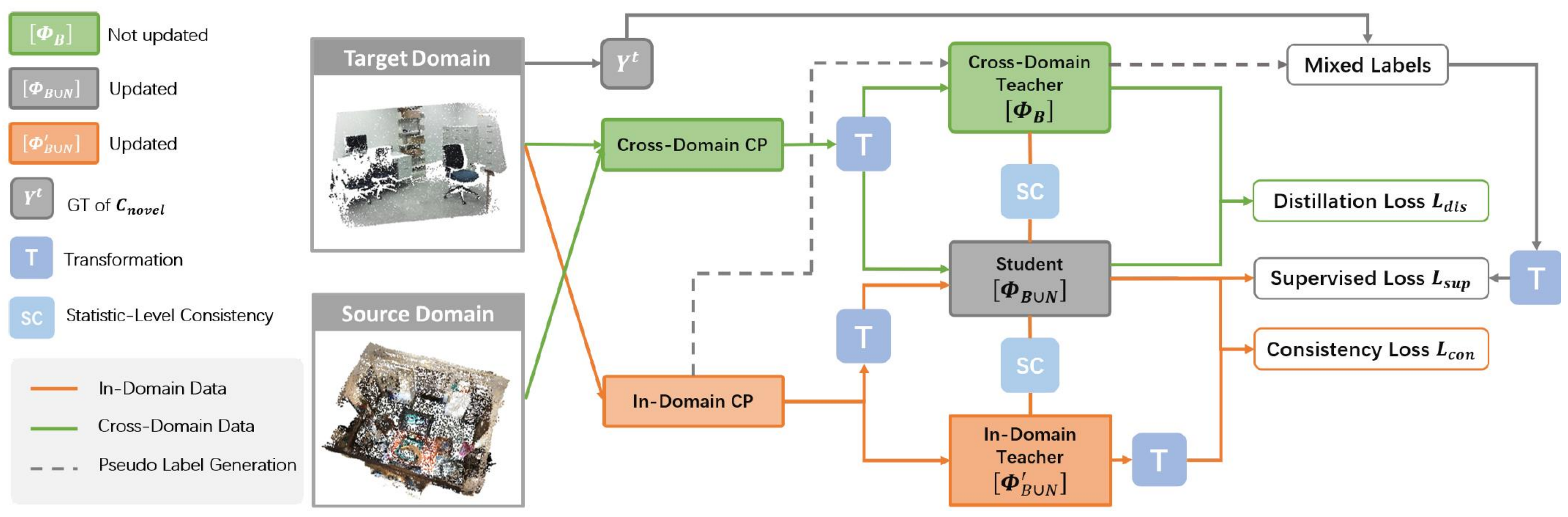
- **Multi-Level Consistency**

- To facilitate the dual-teacher knowledge transfer, we propose multi-level consistency regularization from two aspects.
  - Statistics-Level (SC) Consistency (Batch Normalization parameters alignment)
  - Bounding Box-Level Consistency (center-, class-, and size-level consistency loss)

$$L_{center} = dist(Pair_N) + dist(Pair_M),$$

$$L_{class} = \frac{1}{M}\sum_{i=1}^{M} D_{KL}(p_s^i, p_d^j), \quad (b_s^i, b_d^j) \subset Pair_M,$$

$$L_{size} = \frac{1}{M}\sum_{i=1}^{M} MSE(s_s^i, s_d^j),$$



- **Dual-Teacher Training**

- **Cross-domain Teacher:** The student model learns the underlying knowledge in base classes from a cross-domain teacher via distillation loss (square of Euclidean distance between the classification logits of different bounding boxes).
- **In-domain Teacher:** Meanwhile, the EMA in-domain teacher helps student model capture structure and semantic invariant information in objects with consistency loss.

$$L_{con} = L_{center} + \lambda_{class} L_{class} + \lambda_{size} L_{size},$$

- The mixed labels (pseudo base + real novel) are transformed by the same augmentation step that is applied on the augmented source domain to compute a supervised loss with the backbone VoteNet.

$$L = \lambda_{sup} L_{sup} + \lambda_{dis} L_{dis} + \lambda_{con} L_{con},$$

Experimental Results

- **Dataset**

- **ScanNet and SUN RGB-D**. We set 5 categories (bathtub, bed, bookshelf, chair, desk) as base classes and 5 additional categories (dresser, nightstand, sofa, table, toilet) in SUN RGB-D as novel classes in the target domain.

- **Quantitative Results**

Table 1: 3D object detection performance mAP@0.25 on the SUN RGB-D validation set.

	Method	C_s^b	C_t^b	C_t^n	Base	Novel	All
Base only	Base train	✓	×	×	26.60	-	-
	Base train	×	✓	×	57.63	-	-
Class-Incremental Learning (CIL)	Freeze-and-add	×	✓	✓	54.24	10.61	32.42
	Fine-tune	×	✓	✓	3.48	54.10	28.79
	MT [27]	×	✓	✓	49.63	**61.21**	55.42
	SDCoT [32]	×	✓	✓	52.04	59.48	55.76
	Ours	×	✓	✓	**53.81**	59.50	**56.66**
CIL under Domain Shift	Freeze-and-add	✓	×	✓	23.31	2.04	12.68
	Fine-tune	✓	×	✓	1.73	58.24	29.99
	MT [27]	✓	×	✓	33.84	**59.03**	46.43
	SDCoT [32]	✓	×	✓	29.41	57.74	43.57
	Ours	✓	×	✓	**36.41**	58.80	**47.60**
Base & Novel	Joint train	✓	×	✓	6.80	52.52	29.66
	Joint train	×	✓	✓	58.11	59.27	58.69

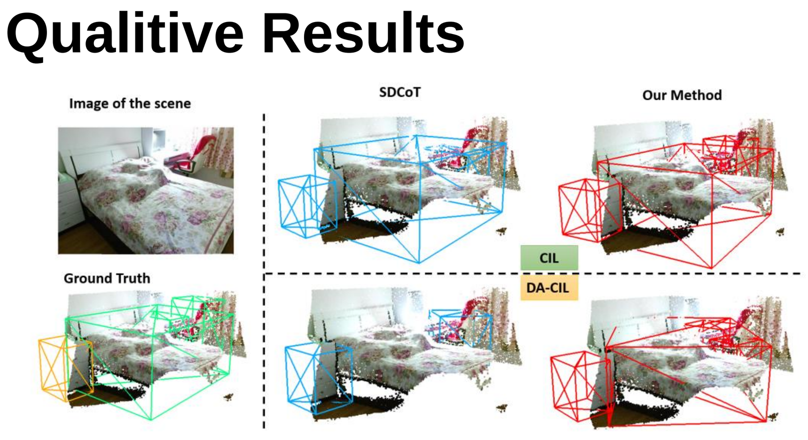
Table 3: Object detection performance mAP@0.25 of different augmentation techniques.

Method	Base	Novel	All
Mix3D [illegible]	33.48	57.89	45.69
CutMix [illegible]	31.71	57.15	44.43
Ours	**36.41**	**58.80**	**47.60**

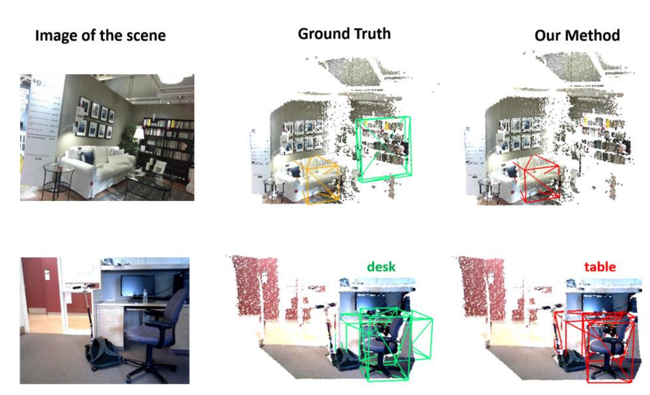
Table 4: Object detection performance mAP@0.25 with exclusion of components in our method.

Method	Base	Novel	All
No cross-domain CP	33.67	**59.95**	46.81
No in-domain CP	35.70	58.39	47.05
No BN consistency	**36.52**	57.20	46.86
Ours	36.41	58.80	**47.60**

- **Qualitive Results**



Good



Failure

Acknowledgment

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AI3 HTPO Seed Fund no. C211118008.

For more information, please contact: zhaoz@i2r.a-star.edu.sg or friend me via LinkedIn or ResearchGate.
https://jacobzhaoziyuan.github.io/

BMVC 2022 | NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | Institute for Infocomm Research I²R | MEMORIES FOR THE FUTURE 2002 - 2022

CREATING GROWTH, ENHANCING LIVES