A. Appendix

A.1. CoIn++

Our CoIn framework provides better initial pseudo labels, and its performance can be further enhanced by a self-training framework. Specifically, we propose a CoIn-based instance-level pseudo-label mining method, named CoIn++. As illustrated in Fig. 5, CoIn++ contains two key parts: (1) a Global Transformation module, which generates different point cloud representations of the same scene; and (2) an IoU-Guided Fusion module, which mines more reliable labels.

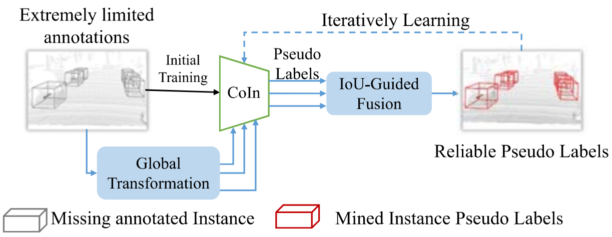

Figure 5. Our proposed CoIn++ pipeline, which mainly consists of a Global Transformation module and IoU-Guided Fusion module.

Global Transformation. The successful practice on the fully supervised methods [32, 45, 22] proves that the global transformation operation can enhance the robustness and accuracy of the detection performance. Inspired by this, we use global transformation on $train$ split to obtain robust pseudo labels. Specially, we adopt two types of common global transformation operations: (i) flipping the whole scene along X-axis; and (ii) rotating the whole scene around Z-axis with angles of $-\frac{\pi}{8}$ and $\frac{\pi}{8}$. We obtain six different scene representations of the same scene (denoted as P_{1-6}).

IoU-Guided Fusion. Under extremely limited annotations, we obtain a well-performed initial CoIn. The CoIn takes P_{1-6} as inputs and generates six corresponding prediction results. Inspired by SS3D[14], we propose an IoU-Guided fusion module to fuse the results of six transformed scenes to obtain more reliable pseudo labels. Because the adopted global transformation is reversible, after we get the detection results, we first perform the inverse transformation on the prediction boundary box according to the transformation rules. Then, we calculate the IoU matrix between predicted bounding boxes of the same instance in different scenes. Finally, we filter out outmatched predictions according to the threshold τ_{IoU} and use WBF [25] strategy to fuse boxes with high overlap.

A.2. Indistinguishable Features

As shown in Fig. 6, we compared indistinguishable features with discriminative features. The characteristic of indistinguishable features is that the distribution of features between different categories is relatively chaotic. When mapped to a one-dimensional space, it is not possible to differentiate between different categories. In contrast to indistinguishable features, discriminative features can be easily differentiated in a one-dimensional space.

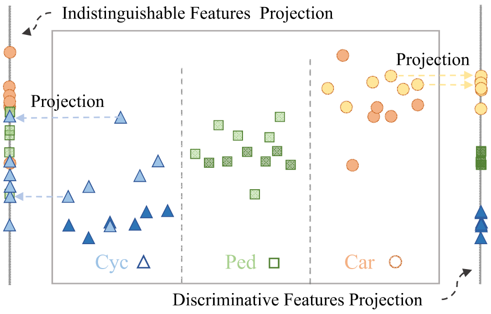

Figure 6. The comparison between indistinguishable features and discriminative features.

A.3. Visualization

Comparison of qualitative results. Fig. 7 shows the comparison of qualitative results between CenterPoint and CoIn with 2% annotations on the val split of KITTI[7] dataset. The yellow, blue and red bounding boxes represent the ground truth, CenterPoint[40] and CoIn, respectively. Our CoIn can solve most of CenterPoint's missed detections. But unfortunately, for distant objects (very sparse point cloud), CoIn cannot detect successfully due to significant similarity differences.

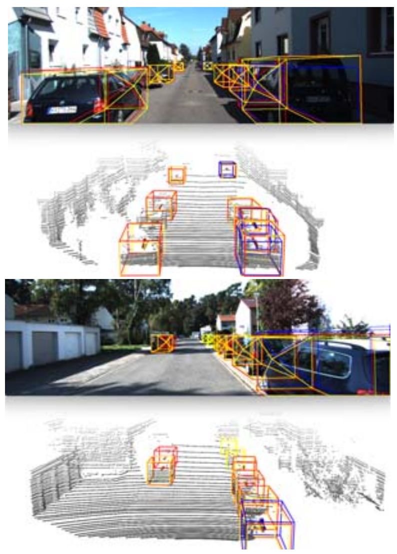

Figure 7. Comparison of qualitative results between CenterPoint and CoIn with 2% annotations on the val split of KITTI [7] dataset. The yellow, blue and red bounding boxes represent the ground truth, CenterPoint [40], and CoIn, respectively.

References

[1] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 7

[3] Qi Cai, Yingwei Pan, Ting Yao, and Tao Mei. 3d cascade rcnn: High quality object detection in point clouds. *IEEE Transactions on Image Processing*, 31:5706–5719, 2022. 2

[4] Yukang Chen, Yanwei Li, X. Zhang, Jian Sun, and Jiaya Jia. Focal sparse convolutional networks for 3d object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[5] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wen gang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. 2, 6, 7

[6] Jian Ding, Enze Xie, Hang Xu, Chenhan Jiang, Zhenguo Li, Ping Luo, and Gui-Song Xia. Unsupervised pretraining for object detection by patch reidentification. *arXiv preprint arXiv:2103.04814*, 2021. 3

[7] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012. 2, 6, 9

[8] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1735–1742. IEEE, 2006. 1

[9] Chenhang He, Hui Zeng, Jianqiang Huang, Xian-Sheng Hua, and Lei Zhang. Structure aware single-stage 3d object detection from point cloud. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11873–11882, 2020. 2

[10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9729–9738, 2020. 1, 3, 4

[11] Jordan S. K. Hu, Tianshu Kuai, and Steven L. Waslander. Point density-aware voxels for lidar 3d object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[12] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12697–12705, 2019. 2

[13] Hanjun Li, Xingjia Pan, Ke Yan, Fan Tang, and Wei-Shi Zheng. Siod: Single instance annotated per category per image for object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14197–14206, 2022. 4, 5

[14] Chuandong Liu, Chenqiang Gao, Fangcen Liu, Jiang Liu, Deyu Meng, and Xinbo Gao. Ss3d: Sparsely-supervised 3d object detection from point cloud. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8428–8437, 2022. 1, 2, 5, 6, 7, 9

[15] Songtao Liu, Zeming Li, and Jian Sun. Self-emd: Self-supervised object detection without imagenet. *arXiv preprint arXiv:2011.13677*, 2020. 3

[16] Jiageng Mao, Minzhe Niu, Haoyue Bai, Xiaodan Liang, Hang Xu, and Chunjing Xu. Pyramid r-cnn: Towards better performance and adaptability for 3d object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021. 1

[17] Qinghao Meng, Wenguan Wang, Tianfei Zhou, Jianbing Shen, Luc Van Gool, and Dengxin Dai. Weakly supervised 3d object detection from lidar point cloud. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 515–531, 2020. 1, 2

[18] Qinghao Meng, Wenguan Wang, Tianfei Zhou, Jianbing Shen, Yunde Jia, and Luc Van Gool. Towards a weakly supervised framework for 3d point cloud object detection and annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4454–4468, 2022. 2

[19] Jinhyung Park, Chenfeng Xu, Yiyang Zhou, Masayoshi Tomizuka, and Wei Zhan. Detmatch: Two teachers are better than one for joint 2d and 3d semi-supervised object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 370–389, 2022. 6

[20] Zengyi Qin, Jinglu Wang, and Yan Lu. Weakly supervised 3d object detection from point clouds. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 4144–4152, 2020. 1, 2

[21] Hualian Sheng, Sijia Cai, Yuan Liu, Bing Deng, Jianqiang Huang, Xiansheng Hua, and Min-Jian Zhao. Improving 3d object detection with channel-wise transformer. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021. 1

[22] 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 *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10526 – 10535, 2020. 2, 6, 9

[23] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–779, 2019. 2, 6

[24] Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 43:2647–2664, 2021. 2

[25] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107, 2021. 9

[26] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, and al. et. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2443 – 2451, 2020. 7

[27] OD Team et al. Openpcdet: An open-source toolbox for 3d object detection from point clouds, 2020. 5

[28] He Wang, Yezhen Cong, Or Litany, Yue Gao, and Leonidas J. Guibas. 3dioumatch: Leveraging iou prediction for semi-supervised 3d object detection. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14610–14619, 2021. 1, 2, 6, 7

[29] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin. Deep metric learning with angular loss. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2612–2620, 2017. 5

[30] Tiancai Wang, Tong Yang, Jiale Cao, and Xiangyu Zhang. Co-mining: Self-supervised learning for sparsely annotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2800–2808, 2021. 1

[31] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, and Lei Li. Dense contrastive learning for self-supervised visual pre-training. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3024–3033, 2021. 3

[32] Hai Wu, Jinhao Deng, Chenglu Wen, Xin Li, and Cheng Wang. Casa: A cascade attention network for 3d object detection from lidar point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 2022. 2, 6, 7, 9

[33] Xiaopei Wu, Liang Peng, Honghui Yang, Liang Xie, Chenxi Huang, Chengqi Deng, Haifeng Liu, and Deng Cai. Sparse fuse dense: Towards high quality 3d detection with depth completion. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2

[34] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3733–3742, 2018. 4, 5

[35] Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. Detco: Unsupervised contrastive learning for object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8392–8401, 2021. 3

[36] Qiangeng Xu, Yiqi Zhong, and Ulrich Neumann. Behind the curtain: Learning occluded shapes for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022. 1

[37] Ceyuan Yang, Zhirong Wu, Bolei Zhou, and Stephen Lin. Instance localization for self-supervised detection pretraining. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3987–3996, 2021. 3

[38] Zetong Yang, Yanan Sun, Shu Liu, and Jiaya Jia. 3dssd: Point-based 3d single stage object detector. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11040–11048, 2020. 1

[39] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Std: Sparse-to-dense 3d object detector for point cloud. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1951–1960, 2019. 1, 2

[40] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 2, 3, 4, 5, 6, 7, 8, 9

[41] Zehan Zhang, Yang Ji, Wei Cui, Yulong Wang, Hao Li, Xian Zhao, Duo Li, Sanli Tang, Ming Yang, Wenming Tan, et al. Atf-3d: Semi-supervised 3d object detection with adaptive thresholds filtering based on confidence and distance. *IEEE Robotics and Automation Letters*, 7(4):10573–10580, 2022. 2

[42] Bing Zhao, Jun Li, and Hong Zhu. Codo: Contrastive learning with downstream background invariance for detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4196–4201, 2022. 3

[43] Na Zhao, Tat-Seng Chua, and Gim Hee Lee. Sess: Self-ensembling semi-supervised 3d object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2

[44] Wu Zheng, Weiliang Tang, Sijin Chen, Li Jiang, and Chi-Wing Fu. Cia-ssd: Confident iou-aware single-stage object detector from point cloud. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. 2

[45] Wu Zheng, Weiliang Tang, Li Jiang, and Chi-Wing Fu. Se-ssd: Self-ensembling single-stage object detector from point cloud. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14494–14503, 2021. 2, 6, 9

[46] Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. Iou loss for 2d/3d object detection. *3DV*, pages 85–94, 2019. 1

[47] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4490–4499, 2018. 2