LiDAR-Net: A Real-scanned 3D Point Cloud Dataset for Indoor Scenes

Supplementary Material

A. Semantic Categories

Our LiDAR-Net encompasses 24 semantic categories across learning, working and living scenes. All of the semantic categories are showcased in Fig. S1.

B. Detailed Experimental Results

Section 4 in the main paper introduces eight benchmark methods employed across three important tasks to test our LiDAR-Net.

B.1. Results on LiDAR-Net

Tables 2 and 3 in the main paper display the mean evaluation metrics, and IoU of KPConv [6] and RandLA-Net [3] for each semantic categories. Additionally, Table S1 lists IoU of PointNet [5] and PointNet++ [6]. Qualitative comparison is displayed in Fig. S2. Significant variations exist among algorithms within these categories. For instance, KPConv achieves 38.49% for pipe, surpassing other methods remarkably, while RandLA-Net achieves 51.22% for door, demonstrating superior performance. Moreover, Table S1 reveals that certain categories such as pillar and reflection noise present substantial challenges in semantic segmentation research. Pillars may be confused with walls because the two parts have similar appearance, which is shown in the top row of Fig. S2. But the segmentation of pillars and walls plays an important role in tasks like smart construction sites and vectorized indoor modeling. Reflection noise, which is mainly caused by specular reflections on glass or metal, typically suspends in the air. Current point cloud understanding methods do not focus on these anomalies due to dataset limitations. Our LiDAR-Net offers real-scanned data to the academic community, aiming to address these challenges.

Qualitative comparisons for instance segmentation and object detection are displayed in Figs. S3 and S4, respectively. Statistics of each categories are listed in Tables S2 and S3. All 24 semantic categories in LiDAR-Net contain instance labels (except for the 'Unknown' category, not included among the 24 categories). In the object detection task, following [4, 7], floors, ceilings, pipes, walls, pillars, reflection noises and ghosts, which lack definite boundaries, are excluded. Point clouds input to VoteNet [7] and GroupFree3D [4] are downsampled to 50000 points.

Table S2 reveals that ASIS [9] and 3D-BoNet [10] demonstrate good performance for certain building structures, such as floors and ceilings. But they lack adequate segmentation for walls and pillars. Additionally, both methods do not segment pipes and stairs. In the object detection task, (Table S3), VoteNet [7] and GroupFree3D [4] do not achieve satisfactory performance for stairs, wardrobes and tabletop others, including both large and small objects. We anticipate that our challenging dataset will stimulate further progress in these areas.

B.2. Generalization Test

In Section 4.3 of the main paper, we train the benchmark methods on existing popular datasets, including S3DIS [1], ScanNet [2], and ScanNet++ [11]. Then we test the deep models on 5 real-scanned rooms. We list the results of each communal categories in Tables S4, S5, and S6. ScanNet++ employs an open vocabulary annotation system encompassing more than 1000 categories. To train the deep models, we extract and combine keywords for each category in ScanNet++. For example, 'Office table' is treated as 'Table'. Tables S4, S5, and S6 demonstrate that deep models trained on LiDAR-Net exhibit superior performances across all three tasks. The realistic distribution of raw points in LiDAR-Net, together with the comprehensive annotations of scanning anomalies, significantly enhances the generalization of deep models for real-world applications.

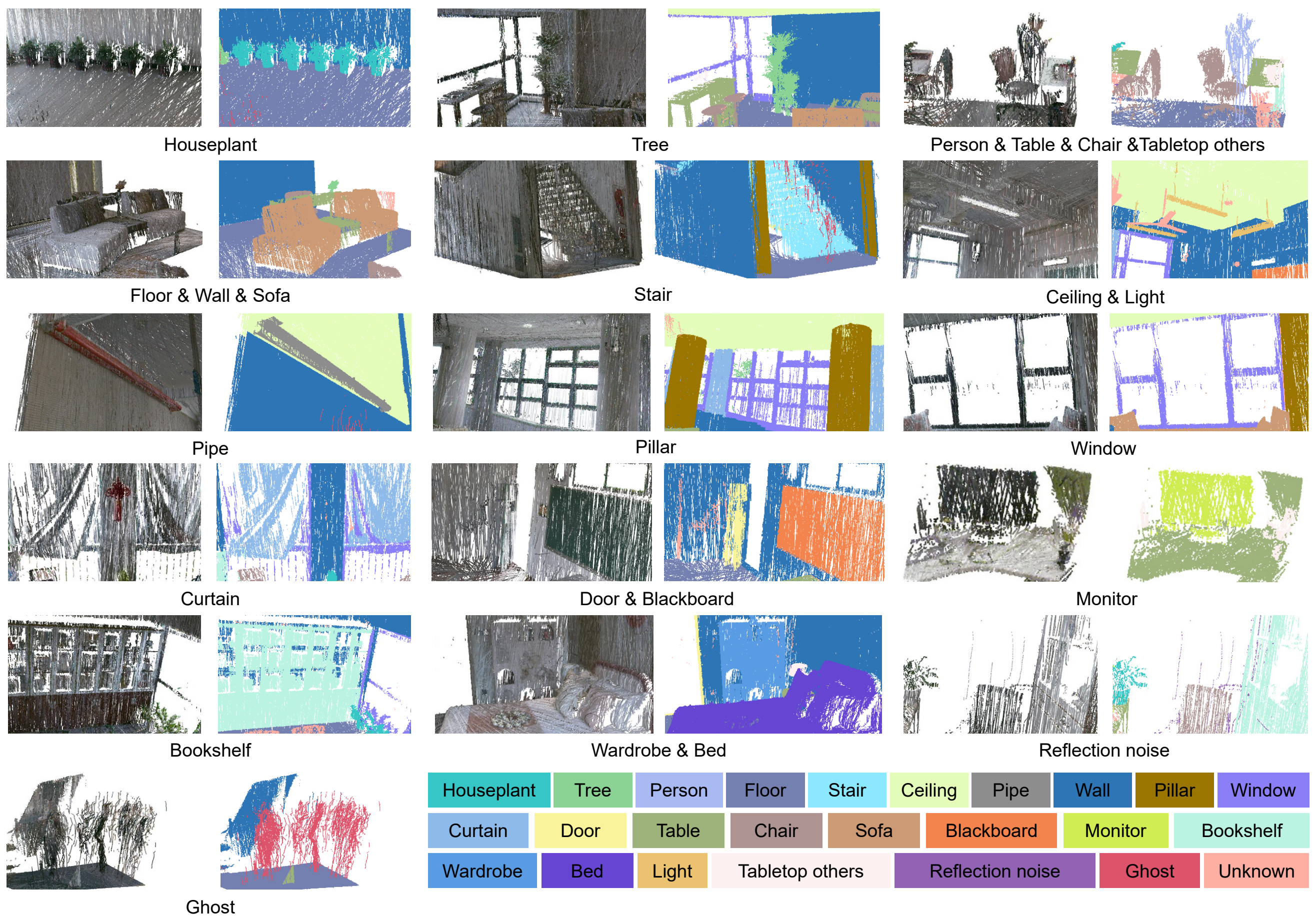

Figure S1. Original points and color-coded labels of each semantic category. Some figures contain several categories. The color map is illustrated at the bottom right.

Methods	Houseplant	Tree	Person	Floor	Stair	Ceiling	Pipe	Wall	Pillar
PointNet [5]	0.00	0.07	4.89	77.96	1.41	76.34	10.08	61.05	0.00
PointNet++ [6]	2.35	3.82	33.64	61.48	1.53	76.46	16.58	62.31	4.23
KPConv [8]	26.98	7.94	45.22	85.43	25.70	86.51	38.49	74.92	5.68
RandLA-Net [3]	3.56	42.02	40.75	93.17	0.42	86.12	19.43	73.81	0.91
	Window	Curtain	Door	Table	Chair	Sofa	Blackboard	Monitor	Bookshelf
PointNet [5]	1.93	7.34	0.00	42.39	16.76	6.50	26.15	8.37	0.22
PointNet++ [6]	10.41	37.60	19.13	63.05	54.00	28.58	50.36	30.27	4.74
KPConv [8]	9.48	22.21	34.44	79.88	75.45	58.02	74.79	55.23	26.76
RandLA-Net [3]	18.00	41.33	51.22	61.69	42.28	0.00	74.77	6.06	7.73
	Wardrobe	Bed	Light	Tabletop others	Reflection noise	Ghost	Unknown	mIoU	
PointNet [5]	0.00	28.11	4.54	1.87	0.22	26.52	4.13	16.27	
PointNet++ [6]	0.00	34.33	40.02	22.46	0.23	52.89	9.28	28.79	
KPConv [8]	34.05	39.51	65.21	40.58	19.22	56.82	18.80	44.29	
RandLA-Net [3]	0.00	0.00	54.13	14.91	0.00	67.53	15.09	32.60	

Table S1. Benchmark semantic segmentation results on LiDAR-Net for each semantic category. Each percent sign (%) of IoU is omitted.

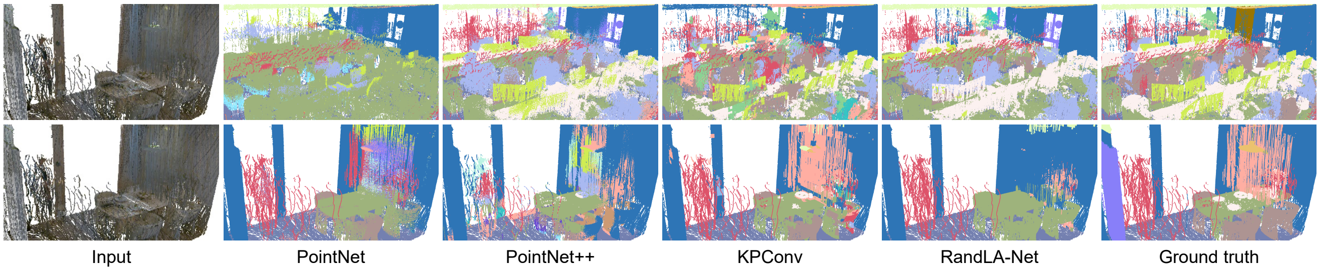

Figure S2. Results of semantic segmentation on LiDAR-Net.

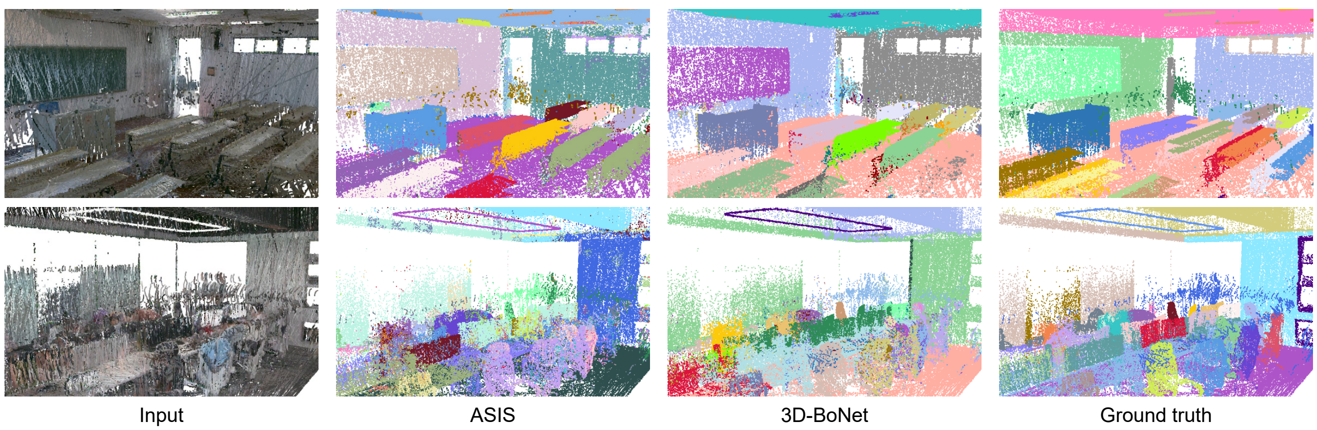

Figure S3. Results of instance segmentation on LiDAR-Net.

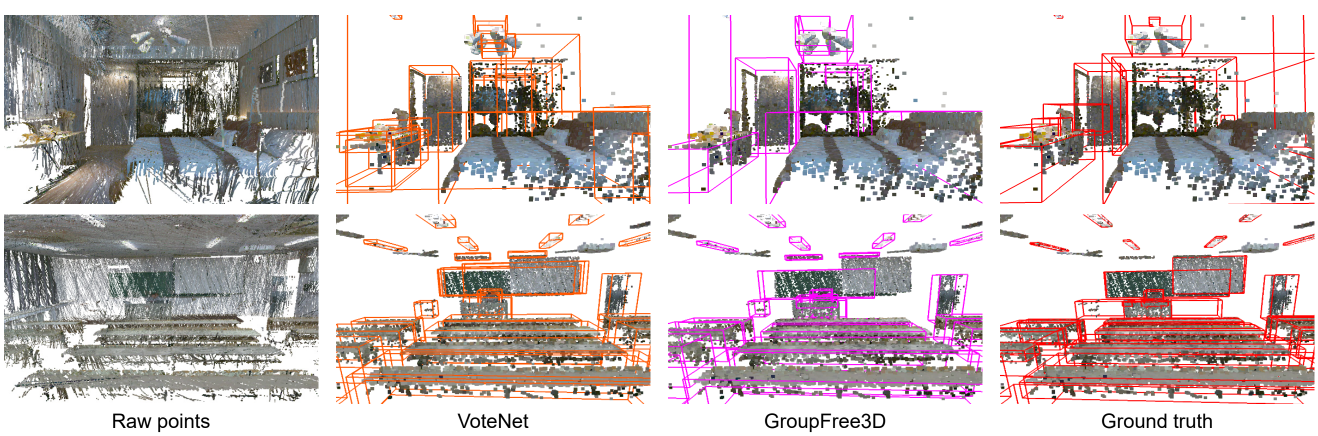

Figure S4. Results of object detection on LiDAR-Net.

Metrics	Methods	Houseplant	Tree	Person	Floor	Stair	Ceiling	Pipe	Wall	Pillar
Pre	ASIS [9]	6.82	57.14	40.91	91.84	0.00	80.43	0.00	40.45	14.29
	3D-BoNet [10]	7.89	50.00	22.03	95.92	0.00	96.15	0.00	50.63	10.53
Rec	ASIS [9]	8.33	18.18	45.00	53.57	0.00	41.11	0.00	22.57	3.03
	3D-BoNet [10]	8.33	18.18	21.67	55.95	0.00	55.56	0.00	38.91	3.03
		Window	Curtain	Door	Table	Chair	Sofa	Blackboard	Monitor	Bookshelf
Pre	ASIS [9]	21.21	47.76	82.24	60.09	38.37	23.81	84.38	38.89	11.11
	3D-BoNet [10]	18.75	43.75	85.71	47.65	30.88	0.00	83.87	18.87	5.56
Rec	ASIS [9]	9.52	35.16	61.97	35.50	15.93	26.32	75.00	13.59	15.38
	3D-BoNet [10]	4.08	23.08	59.15	29.54	9.26	0.00	72.22	9.71	7.69
		Wardrobe	Bed	Light	Tabletop others	Reflection noise	Ghost	mPre / mRec		
Pre	ASIS [9]	0.00	66.67	53.27	12.35	0.00	23.64	37.32		
	3D-BoNet [10]	100.00	100.00	69.40	0.00	0.00	18.00	39.82		
Rec	ASIS [9]	0.00	50.00	45.89	4.90	0.00	20.97	25.08		
	3D-BoNet [10]	10.00	100.00	47.72	0.00	0.00	16.98	24.63		

Table S2. Benchmark instance segmentation results on LiDAR-Net for each semantic category. Each percent sign (%) of precision (Pre) and recall rate (Rec) is omitted.

Metrics	Methods	Houseplant	Tree	Person	Stair	Window	Curtain	Door	Table	Chair
AP@0.25	VoteNet [7]	33.32	20.89	64.12	2.86	38.48	73.54	81.49	44.15	52.10
	GroupFree3D [4]	45.69	47.02	69.71	0.00	39.83	75.77	76.85	36.09	43.44
AP@0.50	VoteNet [7]	2.78	8.87	13.57	0.00	4.75	29.44	51.64	6.68	20.74
	GroupFree3D [4]	19.09	24.91	15.34	0.00	4.26	24.15	33.52	5.35	23.60
		Sofa	Blackboard	Monitor	Bookshelf	Wardrobe	Bed	Light	Tabletop others	mPre / mRec
AP@0.25	VoteNet [7]	85.56	60.04	25.10	27.31	28.46	100.00	22.02	2.62	44.83
	GroupFree3D [4]	70.73	27.99	45.01	33.90	18.27	100.00	60.65	4.63	46.80
AP@0.50	VoteNet [7]	29.27	1.10	1.23	0.01	10.94	100.00	2.95	0.08	16.71
	GroupFree3D [4]	28.13	1.34	2.37	5.89	10.29	100.00	22.30	0.06	18.86

Table S3. Benchmark object detection results on LiDAR-Net for each object category. Each percent sign (%) of AP@0.25 and AP@0.5 is omitted.

Methods	Datasets	Floor	Wall	Window	Door	Table	Chair	Sofa	Bookshelf	mIoU
PointNet [5]	S3DIS [1]	58.53	22.50	0.00	1.91	24.45	0.00	0.00	0.00	13.42
	ScanNet [2]	60.52	0.00	0.00	0.00	0.00	0.09	0.00	0.00	7.58
	ScanNet++ [11]	0.00	47.48	0.00	0.00	0.00	0.00	0.00	0.00	5.94
	LiDAR-Net	64.60	37.63	7.37	0.00	27.96	2.33	0.00	0.00	17.49
PointNet++ [6]	S3DIS [1]	52.94	2.74	2.07	1.44	0.11	8.95	0.00	0.00	8.53
	ScanNet [2]	68.26	0.00	9.64	0.00	0.53	2.56	0.00	4.65	10.71
	ScanNet++ [11]	37.65	8.30	0.00	0.00	0.00	21.16	0.00	0.09	8.40
	LiDAR-Net	70.99	75.12	5.97	5.74	40.06	30.52	0.00	8.54	29.62
KPConv [8]	S3DIS [1]	78.89	27.77	0.22	0.96	31.28	18.11	2.08	20.02	22.42
	ScanNet [2]	38.92	52.07	2.69	5.99	11.41	52.24	0.20	0.00	20.44
	ScanNet++ [11]	22.86	40.52	0.00	0.73	0.01	0.35	0.00	0.05	8.07
	LiDAR-Net	58.63	72.10	4.07	5.76	46.72	46.88	30.85	0.14	33.14
RandLA-Net [3]	S3DIS [1]	87.57	23.76	1.93	40.71	21.87	54.97	31.14	22.66	35.58
	ScanNet [2]	64.06	65.66	12.01	24.98	38.85	59.64	0.00	0.21	33.18
	ScanNet++ [11]	70.04	24.30	2.49	0.14	36.44	47.20	8.15	0.04	23.60
	LiDAR-Net	91.64	79.87	12.56	8.45	45.86	55.26	0.00	24.34	39.75

Table S4. Quantitative comparisons of generalization to real-scanned data for semantic segmentation task. The percent sign (%) is omitted.

Methods	Metrics	Datasets	Window	Door	Table	Chair	Sofa	Bookshelf	mPre/mRec
ASIS [9]	Pre	S3DIS [1]	0.00	75.00	0.00	19.30	0.00	0.00	15.72
		ScanNet [2]	0.00	21.43	40.00	21.37	0.00	12.50	15.88
		ScanNet++ [11]	0.00	100.00	16.67	42.19	0.00	0.00	26.48
		LiDAR-Net	57.14	50.00	34.78	35.62	0.00	20.00	32.92
	Rec	S3DIS [1]	0.00	42.86	0.00	4.17	0.00	0.00	7.84
		ScanNet [2]	0.00	42.86	14.29	9.47	0.00	20.00	14.44
		ScanNet++ [11]	0.00	14.29	3.57	10.23	0.00	0.00	4.68
		LiDAR-Net	21.05	42.86	28.57	9.85	0.00	20.00	20.39
3D-BoNet [10]	Pre	S3DIS [1]	0.00	45.45	0.00	0.00	0.00	0.00	7.58
		ScanNet [2]	11.76	40.00	22.22	20.00	0.00	50.00	24.00
		ScanNet++ [11]	0.00	100.00	0.00	39.74	0.00	0.00	23.29
		LiDAR-Net	37.50	33.33	33.33	42.74	0.00	0.00	24.48
	Rec	S3DIS [1]	0.00	71.43	0.00	0.00	0.00	0.00	11.91
		ScanNet [2]	10.53	28.75	14.29	5.73	0.00	20.00	13.22
		ScanNet++ [11]	0.00	14.29	0.00	11.83	0.00	0.00	4.35
		LiDAR-Net	47.37	85.71	32.14	19.08	0.00	0.00	30.72

Table S5. Quantitative comparisons of generalization to real-scanned data for instance segmentation task. The percent sign (%) is omitted.

Methods	Metrics	Datasets	Window	Door	Table	Chair	Sofa	Bookshelf	mAP
VoteNet [7]	AP@0.25	S3DIS [1]	4.95	5.71	9.75	13.36	10.74	0.00	7.42
		ScanNet [2]	6.77	50.37	20.71	10.53	20.36	4.20	18.83
		ScanNet++ [11]	10.33	30.69	10.84	9.93	1.34	1.17	10.72
		LiDAR-Net	45.55	69.05	51.13	29.55	7.21	5.39	34.65
	AP@0.5	S3DIS [1]	0.07	1.12	3.17	0.32	0.00	0.00	0.78
		ScanNet [2]	0.00	7.14	1.91	0.52	1.05	0.00	1.77
		ScanNet++ [11]	0.00	14.29	0.09	0.21	0.00	0.00	2.43
		LiDAR-Net	5.65	14.29	19.90	5.61	1.67	5.00	8.69
GroupFree3D [4]	AP@0.25	S3DIS [1]	2.90	27.81	3.37	4.71	8.75	0.04	7.93
		ScanNet [2]	5.77	6.86	13.89	37.71	36.32	24.92	20.91
		ScanNet++ [11]	4.66	4.69	5.07	31.44	5.99	28.00	13.31
		LiDAR-Net	46.93	41.28	61.68	38.30	2.75	6.84	32.96
	AP@0.5	S3DIS [1]	0.00	0.00	0.13	1.45	0.10	0.00	0.28
		ScanNet [2]	0.00	0.00	1.84	2.89	2.56	0.00	1.21
		ScanNet++ [11]	0.00	2.11	0.01	7.85	0.02	0.00	1.67
		LiDAR-Net	6.89	35.78	4.76	7.88	0.03	0.00	9.22

Table S6. Quantitative comparisons of generalization to real-scanned data for object detection task. The percent sign (%) is omitted.

References

[1] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017. 1, 4, 5

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1, 4, 5

[3] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11108–11117, 2020. 1, 2, 4

[4] Ze Liu, Zheng Zhang, Yue Cao, Han Hu, and Xin Tong. Group-free 3d object detection via transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2949–2958, 2021. 1, 4, 5

[5] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1, 2, 4

[6] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, 2017. 1, 2, 4

[7] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9277–9286, 2019. 1, 4, 5

[8] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019. 2, 4

[9] Xinlong Wang, Shu Liu, Xiaoyong Shen, Chunhua Shen, and Jiaya Jia. Associatively segmenting instances and semantics in point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4096–4105, 2019. 1, 4, 5

[10] Bo Yang, Jianan Wang, Ronald Clark, Qingyong Hu, Sen Wang, Andrew Markham, and Niki Trigoni. Learning object bounding boxes for 3d instance segmentation on point clouds. *Advances in neural information processing systems*, 32, 2019. 1, 4, 5

[11] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12–22, 2023. 1, 4, 5