PaSCo: Urban 3D Panoptic Scene Completion with Uncertainty Awareness – Supplementary Material –

Anh-Quan Cao[1] Angela Dai[2] Raoul de Charette[1]
[1]Inria [2]Technical University of Munich

https://astra-vision.github.io/PaSCo

In this document, we begin by reporting implementation details of PaSCo and baselines in Sec. A, and present additional ablations in Sec. B. Finally, we provide additional experiments on uncertainty estimation and panoptic scene completion in Sec. C, showcasing PaSCo performance.

We refer to the **supplementary video**, which is available at https://astra-vision.github.io/PaSCo, for better qualitative judgment of PaSCo performance.

A. Implementation details

PaSCo. Our network employs sparse convolution from the MinkowskiEngine library [3]. The architecture of our Dense 3D CNN is similar to the 3D Completion Subnetwork of SCPNet [11]. Additionally, the implementation of the MLP and voxelization is based on Cylinder3D [14].

Training PaSCo required three days for the Semantic KITTI dataset and five days for the SSCBench-KITTI360 dataset using 2 A100 GPUs (1 item per GPU).

Baselines. We employed the official implementations of **LMSCNet**[1], **JS3CNet**[2], **SCPNet**[3], and **MaskPLS**[4] with their provided parameters.

For SCPNet, despite many email exchanges with the authors we were unable to reproduce their reported performance using their official code as also mentioned by other users [5]. Hence, we put extra effort to reimplement their method following authors' recommendation, which resulted in SCPNet*. Note that the latter is several points better than the official implementation.

B. Additional ablations

Method ablation on SSCBench-KITTI360. We provide additional ablations of our method on SSCBench-KITTI360 validation set in Tab. 6, which align with the results reported in Tab. 5. The first two rows of Tab. 6 ablate the augmentations used during inference (*i.e.* rotation + translation). The subsequent rows present the performance of PaSCo, retrained w/o our proposed components.

	All PQ†↑	All PQ↑	mIoU↑	ins ece↓	ins nll↓	voxel ece↓	voxel nll↓
w/o augmentation	26.42	8.09	20.40	0.6114	4.7203	0.1296	2.0639
w/o rotation augmentation	26.48	7.90	21.20	0.6028	4.6321	0.1293	1.9261
w/o voxel-query sem. loss	24.96	7.59	20.90	0.6381	4.4526	0.1307	1.8124
w/o sem. pruning	22.38	6.50	19.59	0.6244	4.5468	0.1364	2.0997
PaSCo (Ours)	**27.20**	**8.36**	**21.63**	**0.6022**	**4.4119**	**0.1285**	**1.8063**

Table 6. **Method ablation on SSCBench-KITTI360 (val. set.)** We ablate different components of our method during inference (top) and training (bottom), demonstrating that each plays a significant role in achieving the best performance.

Notably, substituting semantic pruning with binary occupancy pruning [2, 4] leads to a substantial decline in performance, particularly in PSC metrics with -4.82/-1.86 in All PQ†/All PQ. This result is expected as semantic pruning not only balances supervision across classes, particularly smaller ones, but also provides additional information to the network. Removing our voxel-query semantic loss (Eq. 3) also leads to a remarkable decrease in performance, with a -2.24/-0.77 drop in All PQ†/All PQ, demonstrating its effectiveness without incurring additional computational costs. Lastly, the augmentations applied during inference (top rows) contribute to increased variation among subnetworks, thereby enhancing overall performance.

Semantic pruning. We further ablate the use of our semantic pruning in Tab. 7, by the replacing it with binary pruning as employed in [2, 4]. From the table, semantic pruning significantly enhances performance across most classes, notably for rare classes such as truck (+33.72/+2.67 PQ on Semantic KITTI/SSCBench-KITTI360), other-vehicle (+4.21/3.22 PQ on Semantic KITTI/SSCBench-KITTI360), and pole (+1.48/+2.22 PQ on Semantic KITTI/SSCBench-KITTI360). These im-

[1] https://github.com/cv-rits/LMSCNet
[2] https://github.com/yanx27/JS3C-Net
[3] https://github.com/SCPNet/Codes-for-SCPNet
[4] https://github.com/PRBonn/MaskPLS
[5] https://github.com/SCPNet/Codes-for-SCPNet/issues/8

	Method	car (3.92%)	bicycle (0.03%)	motorcycle (0.03%)	truck (0.16%)	other-veh. (0.20%)	person (0.07%)	bicyclist (0.07%)	motorcyclist. (0.05%)	road (15.30%)	parking (1.12%)	sidewalk (11.13%)	other-grnd (0.56%)	building (14.10%)	fence (3.90%)	vegetation (39.30%)	trunk (0.51%)	terrain (9.17%)	pole (0.29%)	traf.-sign (0.08%)	mean
PQ	w/o sem. pruning	22.13	9.82	17.97	11.17	7.11	2.91	0.00	0.00	75.00	11.38	24.20	0.00	3.53	**0.61**	8.21	**4.55**	31.05	8.21	2.61	15.04
	PaSCo (Ours)	**24.55**	**7.82**	**18.09**	**44.89**	**11.32**	**3.00**	0.00	0.00	**76.22**	**28.12**	**30.42**	**1.33**	**4.85**	0.27	**12.97**	4.22	**32.61**	**9.69**	**3.26**	**16.51**

(a) Semantic KITTI (val. set)

	Method	car (2.85%)	bicycle (0.02%)	motorcycle (0.01%)	truck (0.16%)	other-veh. (0.58%)	person (0.02%)	road (14.98%)	parking (2.31%)	sidewalk (6.43%)	other-grnd (2.05%)	building (15.67%)	fence (0.96%)	vegetation (41.99%)	terrain (7.10%)	pole (0.22%)	traf.-sign (0.06%)	other-structure (4.33%)	other-object (0.28%)	mean
PQ	w/o sem. pruning	12.46	0.00	2.20	5.00	1.34	**1.54**	68.78	1.63	18.73	0.00	0.71	0.06	0.17	0.00	0.37	3.29	0.00	0.68	6.50
	PaSCo (Ours)	**16.93**	0.00	**3.00**	**7.67**	**4.56**	0.91	**69.17**	**2.22**	**22.29**	**0.06**	**7.08**	**0.06**	**3.19**	0.00	**2.59**	**4.85**	0.00	**5.84**	**8.36**

(b) SSCBench-KITTI360 (val. set)

Table 7. **Ablation of sem. pruning on (a) Semantic KITTI (val. set) and (b) SSCBench-KITTI360 (val. set) for Panoptic Scene Completion class-wise performance.** Semantic pruning improves the performance of the majority of classes on both datasets.

provements can be attributed to a more balanced supervision among classes and the inclusion of additional semantic information.

C. Additional experiments

We provide further experiments for uncertainty estimation in Sec. C.1 and panoptic scene completion in Sec. C.2.

C.1. Uncertainty estimation

Robustness to Out Of Distribution (OOD). We extend our evaluation for Out Of Distribution, initially reported in Sec. 4.2 and Fig. 6. In Fig. 8, we again evaluate baselines MC Dropout [5], TTA [1], Deep Ensemble [8] along with PaSCo(M=1) and PaSCo, on the Robo3D dataset [7] which contains corrupted version of SemanticKITTI. All methods are trained on the clean version of SemanticKITTI.

Different from the main paper, we report performance on the 'clean' set (*i.e.*, the original SemanticKITTI) to better assess the effect of OOD. It's also important to note that better calibration may come at the cost of worse performance.

Hence, Fig. 8 presents not only uncertainty metrics (instance ece, SSC ece – *lower is better*) but also *two performance metrics* (All PQ† and mIoU – *higher is better*). The latter demonstrates that our better calibration (Fig. 8, top) comes along with better performance (Fig. 8, bottom) in almost all corruptions. This further demonstrates the superiority of PaSCo compared to both its one-subnet variation, PaSCo(M=1), and all other baselines.

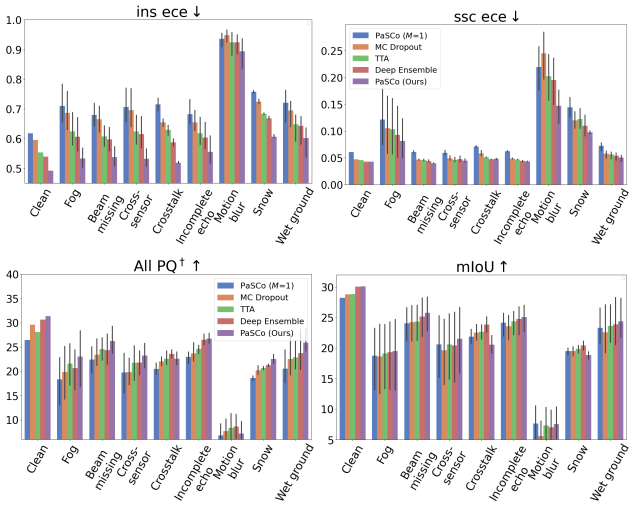


Figure 8. **Impact of Out Of Distribution data on PSC and uncertainty performance.** We evaluate PaSCo and baselines, trained on 'clean' Semantic KITTI on corrupted versions of the same set from the Robo3D [7] dataset. Top part reports, uncertainty ece metrics (*lower is better*), while bottom part reports performance metrics (*higher is better*). The types of corruptions are shown on the x-axis. Each bar represents the average metric for each type of corruption, with the error bars showing the minimum and maximum metric across intensities. PaSCo surpasses all baselines in terms of uncertainty measurement (*i.e.*, lower ece) and demonstrates comparable or better performance metrics (*higher* All PQ† and mIoU).

Qualitative results. Fig. 10a and Fig. 10b present additional qualitative results of uncertainty estimation on Se-

mantic KITTI and SSCBench-KITTI360 validation sets. We also illustrate the uncertainty of "stuff" class which we omitted for brevity in the main paper.

Overall, PaSCo(M=1) exhibits higher level of confidence across both voxel and instance uncertainties. Instances of small classes such as person, motorcycle, pole, and traffic light show higher uncertainty levels compared to larger objects like buildings, road and sidewalk. Furthermore, uncertainty tends to increase around the edges of instances, in areas with occluded views, and in regions with lower density of input points.

C.2. Panoptic Scene Completion (PSC)

Quality of pseudo panoptic labels. DBSCAN is a classical strategy to approximates panoptic labels when unavailable [6, 13]. In the absence of full panoptic ground truth (GT), we validate the quality of our pseudo labels against the *single-scan* point-wise panoptic GT of SemKITTI for the val set, voxelizing both and evaluating where both are defined, in Tab. 8. This confirms that DBSCAN provides a good approximation of the true labels.

labels	All PQ†	All PQ	All SQ	All RQ	Thing PQ	Thing SQ	Thing RQ	Stuff PQ	Stuff SQ	Stuff RQ
HDBSCAN	80.07	80.05	90.41	83.07	63.24	89.03	70.29	**91.26**	**91.32**	**91.59**
DBSCAN (ours)	**88.17**	**88.15**	**92.70**	**90.12**	**83.50**	**94.76**	**87.91**	note that stuff classes are not clustered		

Table 8. Quality of Pseudo Labels on the SemKITTI (val set)

In Table I of MaskPLS [9], authors demonstrates that clustering with HDBSCAN (a density-aware DBSCAN) leads to reasonable panoptic segmentation on SemKITTI.

In Tab. 8, we find DBSCAN labels to be even more accurate than HDBSCAN since we operate on fairly homogenous density data (aggregation of multiple scans), while MaskPLS clusters single scans exhibiting high-varying density.

From Fig. 9, large objects and small objects are reasonably clustered by DBSCAN.

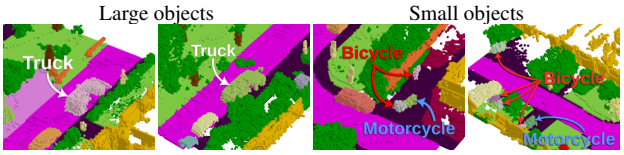


Figure 9. Examples of DBSCAN labels

Class-wise performance. Tab. 9 presents a class-wise comparison of our method, PaSCo, against baselines on (a) the Semantic KITTI (val. set) and (b) the SSCBench-KITTI360 (test set). For the important PQ metric, PaSCo outperforms all baselines in most classes on Semantic KITTI, with the exception of the fence category. Note that no method successfully predicted the bicyclist and motorcyclist classes. The SSCBench-KITTI360 dataset demonstrates its greater complexity with generally lower performance across all classes when compared to Semantic KITTI. Nonetheless, PaSCo still demonstrates strong performance, ranking first in 10 out of 18 classes and second in 3 out of 18. The effectiveness of PaSCo is further illustrated in its superior ability to detect masks and produce high-quality masks, as reflected by its first or second highest performance in most classes based on SQ and RQ metrics on both datasets.

Qualitative results. Fig. 11 presents further qualitative results on Semantic KITTI and SSCBench-KITTI360. PaSCo predicts a more complete scene geometry, as illustrated in rows 1, 3, 5 and 6. It also infers better instance quality, illustrated by the increased accuracy in instance shape in rows 2, 4, 5 and 6, and by the clearer separation observable in rows 2, 3, 4 and 5.

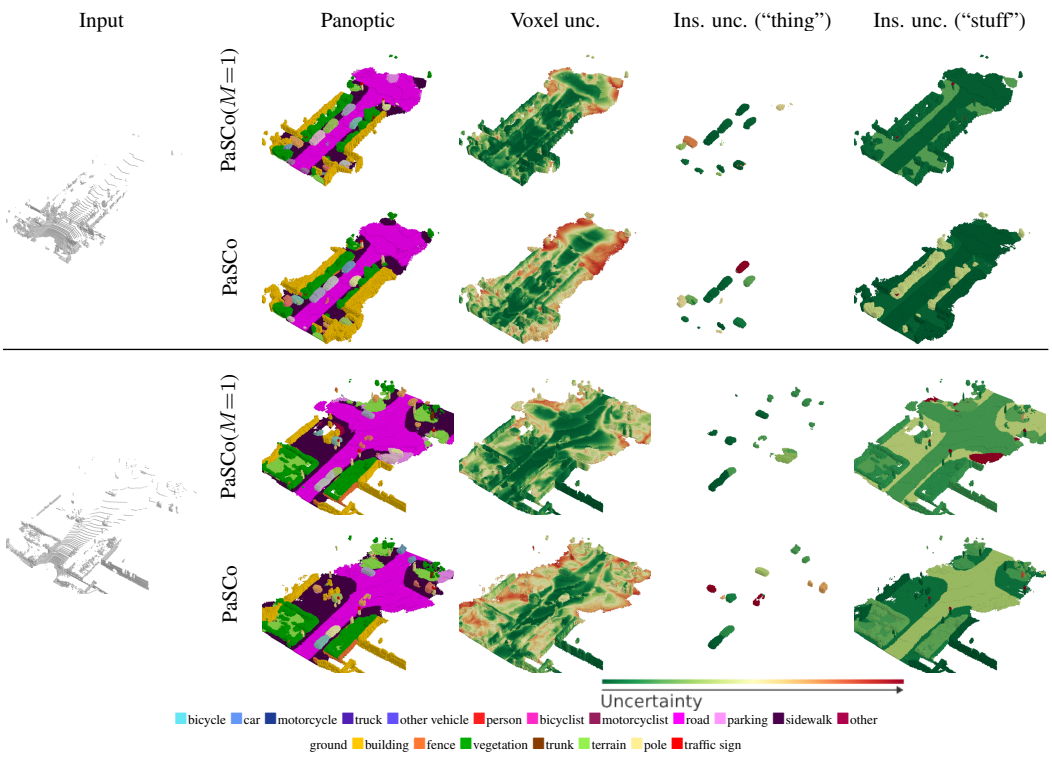
	Method	car (3.92%)	bicycle (0.03%)	motorcycle (0.03%)	truck (0.16%)	other-veh. (0.20%)	person (0.07%)	bicyclist (0.07%)	motorcyclist. (0.05%)	road (15.30%)	parking (1.12%)	sidewalk (11.13%)	other-grnd (0.56%)	building (14.10%)	fence (3.90%)	vegetation (39.30%)	trunk (0.51%)	terrain (9.17%)	pole (0.29%)	traf.-sign (0.08%)	mean
PQ	LMSCNet [10] + MaskPLS [9]	9.43	0.00	0.76	2.32	0.00	0.47	0.00	0.00	53.53	1.82	5.63	0.00	0.26	0.19	0.00	0.27	3.52	1.00	0.00	4.17
	JS3CNet [12] + MaskPLS [9]	9.57	1.07	4.19	17.54	0.91	0.12	0.00	0.00	58.45	5.32	15.89	0.00	1.02	1.33	0.00	0.76	13.63	0.28	0.00	6.85
	SCPNet [11] + MaskPLS [9]	18.44	4.84	6.72	4.42	2.79	1.81	0.00	0.00	63.89	7.92	19.92	0.00	3.11	3.28	0.13	2.29	21.55	1.99	0.17	8.59
	SCPNet* [11] + MaskPLS [9]	11.72	1.80	14.70	26.44	3.83	0.33	0.00	0.00	66.44	18.71	25.29	0.00	2.06	**4.12**	0.39	3.11	22.24	3.97	1.72	10.89
	PaSCo (Ours)	**24.55**	**7.82**	**18.09**	**44.89**	**11.32**	**3.00**	0.00	0.00	**76.22**	**28.12**	**30.42**	**1.33**	**4.85**	0.27	**12.97**	**4.22**	**32.61**	**9.69**	**3.26**	**16.51**
SQ	LMSCNet [10] + MaskPLS [9]	62.65	0.00	53.44	53.87	0.00	**69.00**	0.00	0.00	63.30	57.83	52.70	0.00	53.93	52.58	0.00	**59.76**	54.12	53.37	0.00	36.13
	JS3CNet [12] + MaskPLS [9]	59.88	53.79	55.17	57.73	55.70	62.50	0.00	0.00	65.98	55.70	54.53	0.00	52.62	53.41	0.00	55.01	56.38	**57.68**	0.00	41.90
	SCPNet [11] + MaskPLS [9]	66.69	57.78	65.30	55.30	**65.15**	61.01	0.00	0.00	68.56	58.72	55.81	0.00	54.94	54.45	51.04	55.58	59.86	52.97	**57.14**	49.49
	SCPNet* [11] + MaskPLS [9]	63.98	54.46	60.54	54.03	58.55	52.31	0.00	0.00	70.61	59.25	56.69	0.00	53.61	**55.70**	52.84	56.05	58.76	53.55	56.62	48.29
	PaSCo (Ours)	**70.10**	**57.84**	**67.00**	**67.33**	62.15	60.14	0.00	0.00	**77.52**	**62.62**	**59.95**	**54.71**	**55.87**	51.29	**52.85**	57.50	**63.88**	54.78	55.17	**54.25**
RQ	LMSCNet [10] + MaskPLS [9]	15.05	0.00	1.42	4.30	0.00	0.67	0.00	0.00	84.56	3.15	10.69	0.00	0.48	0.37	0.00	0.45	6.50	1.87	0.00	6.82
	JS3CNet [12] + MaskPLS [9]	15.98	2.00	7.59	30.38	1.63	0.20	0.00	0.00	88.59	9.55	29.14	0.00	1.93	2.49	0.00	1.39	24.17	0.48	0.00	11.34
	SCPNet [11] + MaskPLS [9]	27.65	8.38	10.29	8.00	4.28	2.96	0.00	0.00	93.18	13.50	35.69	0.00	5.66	6.03	0.25	4.12	36.00	3.76	0.30	13.69
	SCPNet* [11] + MaskPLS [9]	18.32	3.31	24.29	48.94	6.54	0.63	0.00	0.00	94.09	31.58	44.61	0.00	3.84	**7.39**	0.73	5.56	37.84	7.42	3.04	17.80
	PaSCo (Ours)	**35.03**	**13.51**	**27.00**	**66.67**	**18.21**	**4.98**	0.00	0.00	**98.32**	**44.91**	**50.73**	**2.44**	**8.69**	0.52	**24.54**	**7.33**	**51.05**	**17.70**	**5.91**	**25.13**

(a) Semantic KITTI (val. set)

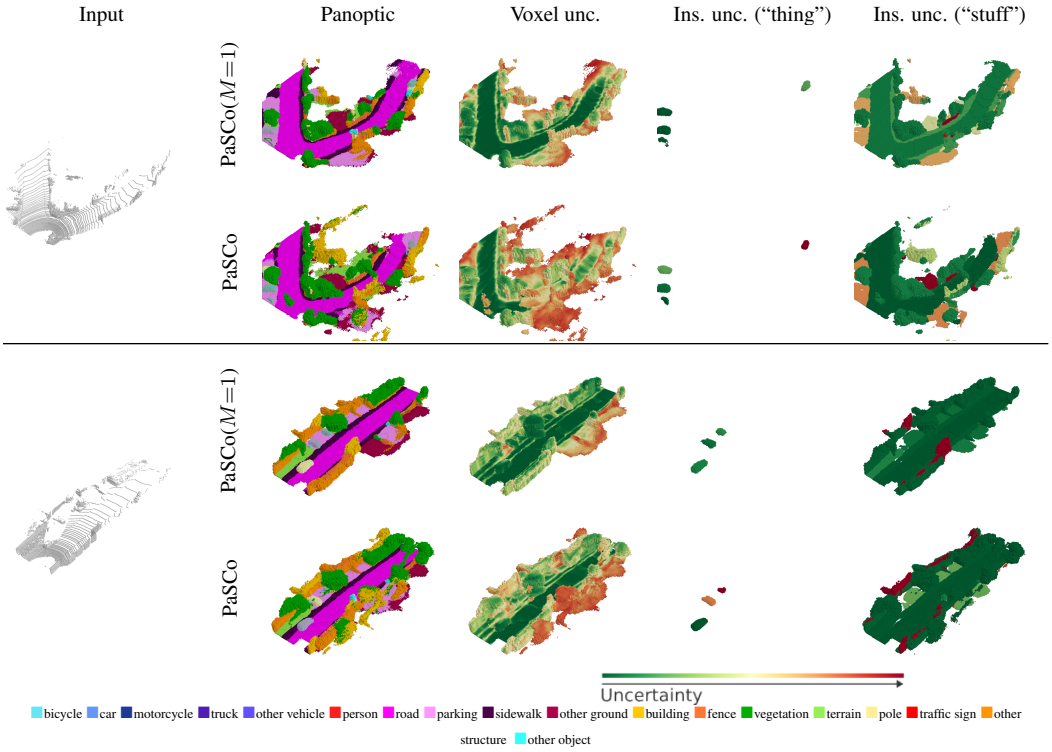
	Method	car (2.85%)	bicycle (0.02%)	motorcycle (0.01%)	truck (0.16%)	other-veh. (0.58%)	person (0.02%)	road (14.98%)	parking (2.31%)	sidewalk (6.43%)	other-grnd (2.05%)	building (15.67%)	fence (0.96%)	vegetation (41.99%)	terrain (7.10%)	pole (0.22%)	traf.-sign (0.06%)	other-structure (4.33%)	other-object (0.28%)	mean
PQ	LMSCNet [10] + MaskPLS [9]	4.64	0.00	0.00	0.00	0.00	0.67	49.87	0.31	5.75	0.00	0.00	0.00	0.00	1.59	0.13	4.24	0.00	0.00	4.14
	JS3CNet [12] + MaskPLS [9]	13.77	0.00	0.81	3.58	0.48	1.50	63.13	1.63	23.99	0.12	0.14	0.19	0.00	4.36	1.51	6.55	0.09	0.44	6.79
	SCPNet [11] + MaskPLS [9]	**17.77**	0.00	**2.56**	1.45	1.69	**1.91**	46.77	0.54	22.06	0.04	0.14	0.37	0.00	5.28	2.02	7.48	**0.14**	0.29	6.14
	SCPNet* [11] + MaskPLS [9]	15.87	0.00	1.53	3.80	0.97	1.72	57.55	**3.34**	30.65	**0.16**	0.04	**0.68**	0.00	**5.66**	2.46	9.62	0.08	0.26	7.47
	PaSCo (Ours)	14.85	**0.11**	1.73	**9.01**	**1.97**	1.60	**72.05**	1.47	**35.82**	0.00	**24.29**	0.53	**8.96**	5.07	**3.15**	**14.47**	0.11	**1.38**	**10.92**
SQ	LMSCNet [10] + MaskPLS [9]	54.47	0.00	0.00	0.00	0.00	**67.99**	66.89	58.36	53.46	0.00	0.00	0.00	0.00	54.17	56.27	65.93	0.00	0.00	26.52
	JS3CNet [12] + MaskPLS [9]	58.34	0.00	59.77	52.01	52.67	67.64	69.77	56.09	57.48	53.95	52.22	53.28	0.00	54.62	54.94	64.73	55.20	58.21	51.16
	SCPNet [11] + MaskPLS [9]	**60.89**	0.00	**61.42**	52.88	**57.20**	58.39	63.06	55.22	56.97	**55.56**	52.45	55.16	0.00	54.96	54.54	65.31	**61.72**	55.60	51.18
	SCPNet* [11] + MaskPLS [9]	59.32	0.00	54.73	**55.96**	53.73	65.01	69.17	**59.13**	58.09	52.13	50.13	55.66	0.00	55.33	55.10	63.70	52.62	52.18	50.67
	PaSCo (Ours)	58.22	**53.03**	57.14	55.89	55.53	65.39	**76.15**	58.04	**58.51**	0.00	**56.18**	**61.65**	**55.28**	**57.99**	**56.83**	**67.85**	54.74	**61.42**	**56.10**
RQ	LMSCNet [10] + MaskPLS [9]	8.51	0.00	0.00	0.00	0.00	0.99	74.56	0.53	10.75	0.00	0.00	0.00	0.00	2.94	0.24	6.43	0.00	0.00	6.45
	JS3CNet [12] + MaskPLS [9]	23.61	0.00	1.36	6.88	0.91	2.22	90.49	2.91	41.72	0.22	0.28	0.36	0.00	7.98	2.75	10.11	0.15	0.76	10.71
	SCPNet [11] + MaskPLS [9]	**29.18**	0.00	**4.17**	2.75	2.95	**3.27**	74.16	0.98	38.72	0.08	0.28	0.67	0.00	9.60	3.71	11.45	**0.23**	0.53	10.15
	SCPNet* [11] + MaskPLS [9]	26.76	0.00	2.79	6.79	1.81	2.65	83.20	**5.65**	52.77	**0.30**	0.09	**1.23**	0.00	**10.24**	4.47	15.10	0.16	0.50	11.92
	PaSCo (Ours)	25.51	**0.21**	3.03	**16.13**	**3.55**	2.45	**94.62**	2.53	**61.22**	0.00	**43.24**	0.86	**16.21**	8.75	**5.54**	**21.33**	0.19	**2.24**	**17.09**

(b) SSCBench-KITTI360 (test set)

Table 9. **Class-wise performance on (a) Semantic KITTI (val. set) and (b) SSCBench-KITTI360 (test set) for Panoptic Scene Completion.** We report the performance of our method and baselines for each class across the two datasets. Our approach exceeds the performance of baseline methods in most classes, particularly in the crucial PQ metric. The SSCBench-KITTI360 dataset exhibits a higher level of complexity compared to Semantic KITTI, as evidenced by its overall lower performance metrics. PaSCo also shows superior performance in SQ and RQ by being either first or second in most classes.



(a) **Semantic KITTI.**



(b) **Qualitative uncertainty comparison on SSCBench-KITTI360 (val. set.**

Figure 10. **Additional qualitative uncertainty comparison on Semantic KITTI and SSCBench-KITTI360.** PaSCo offers more insightful estimates of uncertainty, particularly in smaller instances with incomplete geometry, along the boundaries of segments, in areas with sparse input points, and in extrapolated regions.

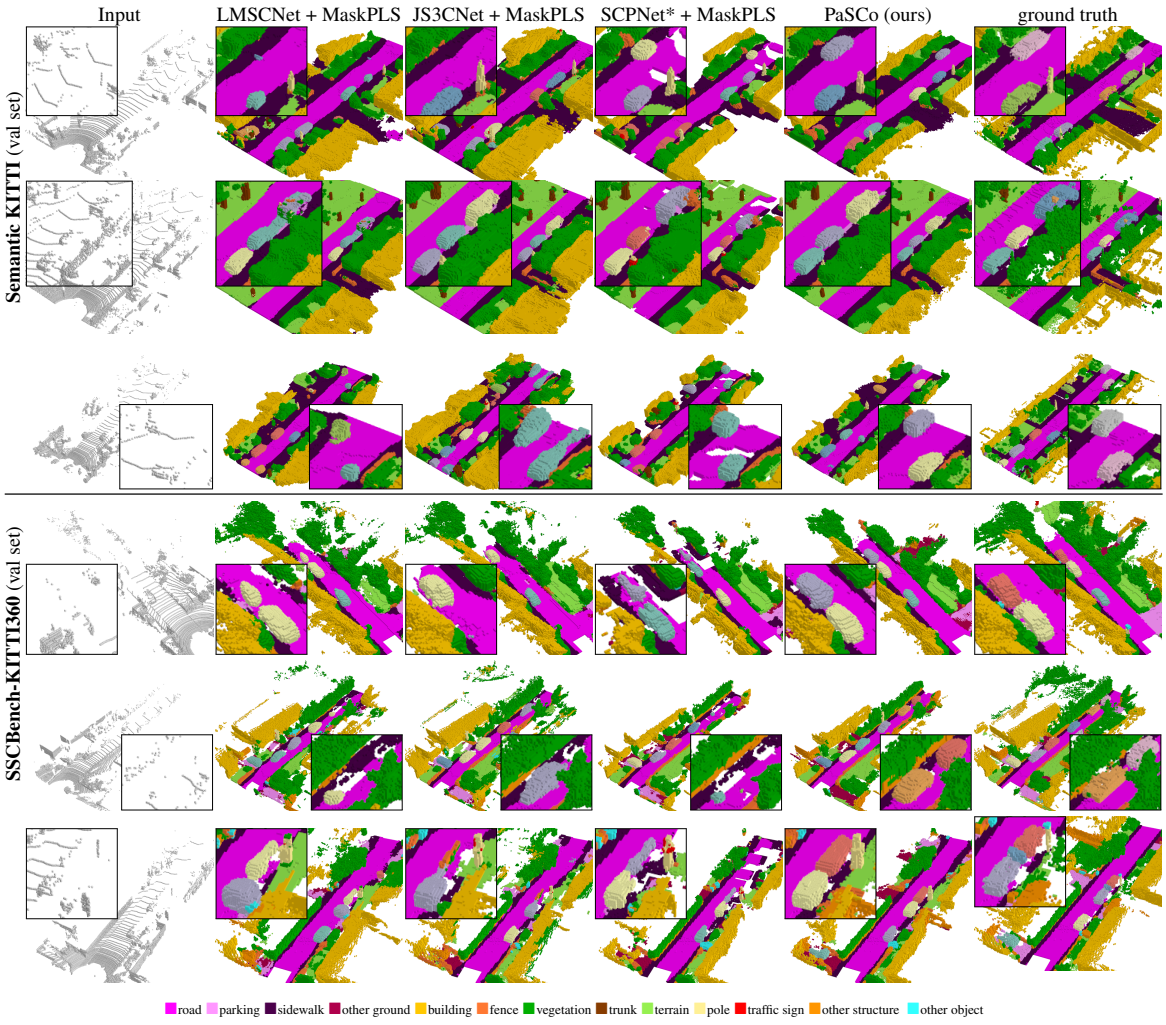


Figure 11. **Additional qualitative results on Panoptic Scene Completion**. PaSCo demonstrates improved instance quality, evident from superior geometry (rows 1, 3 to 6), and better separation (rows 2 to 4). Additionally, it predicts more accurate scene structure, with less missing geometry (rows 1, 3, 5, and 6).

References

[1] Murat Seckin Ayhan and Philipp Berens. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In *MIDL*, 2018. 2

[2] Ran Cheng, Christopher Agia, Yuan Ren, Xinhai Li, and Bingbing Liu. S3cnet: A sparse semantic scene completion network for lidar point clouds. In *CoRL*, 2020. 1

[3] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, 2019. 1

[4] Angela Dai, Christian Diller, and Matthias Nießner. Sg-nn: Sparse generative neural networks for self-supervised scene completion of rgb-d scans. In *CVPR*, 2020. 1

[5] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 2016. 2

[6] Fangzhou Hong, Lingdong Kong, Hui Zhou, Xinge Zhu, Hongsheng Li, and Ziwei Liu. Unified 3d and 4d panoptic segmentation via dynamic shifting networks. *TPAMI*, 2024. 3

[7] Lingdong Kong, Youquan Liu, Xin Li, Runnan Chen, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Robo3d: Towards robust and reliable 3d perception against corruptions. In *ICCV*, 2023. 2

[8] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*, 2017. 2

[9] R. Marcuzzi, L. Nunes, L. Wiesmann, J. Behley, and C. Stachniss. Mask-Based Panoptic LiDAR Segmentation for Autonomous Driving. *RA-L*, 2023. 3, 4

[10] Luis Roldão, Raoul de Charette, and Anne Verroust-Blondet. Lmscnet: Lightweight multiscale 3d semantic completion. In *3DV*, 2020. 4

[11] Zhaoyang Xia, Youquan Liu, Xin Li, Xinge Zhu, Yuexin Ma, Yikang Li, Yuenan Hou, and Yu Qiao. Scpnet: Semantic scene completion on point cloud. In *CVPR*, 2023. 1, 4

[12] Xu Yan, Jiantao Gao, Jie Li, Ruimao Zhang, Zhen Li, Rui Huang, and Shuguang Cui. Sparse single sweep lidar point cloud segmentation via learning contextual shape priors from scene completion. In *AAAI*, 2021. 4

[13] Shuquan Ye, Dongdong Chen, Songfang Han, and Jing Liao. Learning with noisy labels for robust point cloud segmentation. In *ICCV*, 2021. 3

[14] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *CVPR*, 2021. 1