

# SUPPLEMENTARY MATERIALS FOR INSTANCE-AWARE 3D SEMANTIC SEGMENTATION POWERED BY SHAPE GENERATORS AND CLASSIFIERS

**Anonymous authors**

Paper under double-blind review

## A MORE RESULTS

We show more qualitative results for Waymo Open Dataset Sun et al. (2020) in Figure 2, for SemanticKITTI Behley et al. (2019) in Figure 3, and for ScanNet Dai et al. (2017) in Figure 4.

## B INSTANCE CLUSTERING RESULTS

We show the comparison between the ground-truth and the clustered instance labels in the Figure 1. The ground-truth labels only has 7 categories while we have about 13 categories. We use red boxes to highlight the noisy clustering in each result. The noisy patterns are not in majority and our network is robust to those patterns, e.g. the max-pool of per-point features from two pedestrians still gives us pedestrian classification.

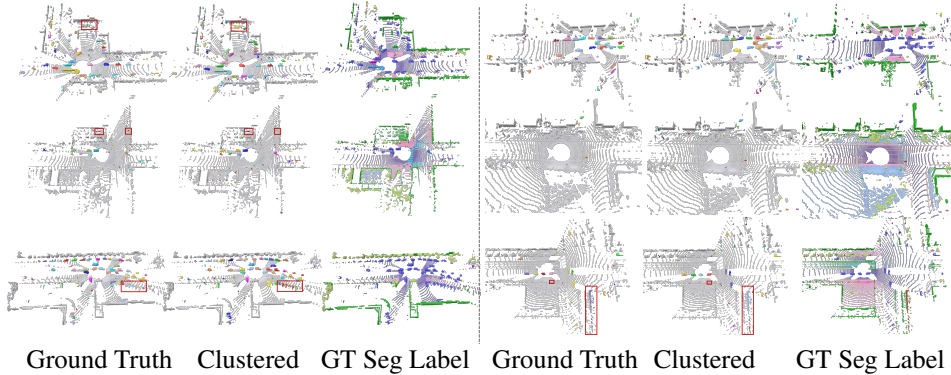


Figure 1: Instance clustering results on the validation set of the Waymo Open Dataset. Note that for clustered results, 14 semantic categories are chosen while there are only 7 categories in the ground truth instance labels. Red boxes highlights the noisy clustering results.

## C MORE ABLATION STUDIES

In the main paper we show the ablation studies on different instance categories, and different instance heads. Here we show more ablation studies on the loss weights for the instance classification and reconstruction in Table 1.

## D FAILURE CASES

As described in the Section 6 in the main paper, one limitation of our method is that the consistency on the mis-classified instances might cause lower IoU than the inconsistent predictions. Figure 5 shows two examples of such situations. On the hard examples, our method might make inaccurate instance classifications, thus all points on that instance are mis-segmented. Since the IoU is computed over all points, the IoU of a misclassified instance (the second row in Figure 5) is lower than the partially corrected predictions (the first row in Figure 5). However, we still think keeping the

$\lambda_1$	0.05	0.1	0.2	0.5
mIoU	68.53	68.93	68.77	68.34

(a) Ablation study on the loss weight of the instance classification head.

$\lambda_1$	0.005	0.01	0.02	0.05
mIoU	68.67	68.93	68.56	68.13

(b) Ablation study on the loss weight of the shape reconstruction head.

Table 1: Ablation study on the loss weights of the instance classification head and shape reconstruction head. We show the mIoU on the validation set of Waymo Open Dataset.

instance consistency is important because of the natural clustering properties of instances. We will develop more metrics to measure this rationality other than IoU.

## E MODEL ARCHITECTURE

We include some training details in the Section 4.5 in the main paper. Here we introduce the detailed network architectures. The whole pipeline is in Figure 6. The source code of our model can also be found in the appended code folder.

## F SOURCE CODE

We append the source code for our paper in the sourcecode folder. Please refer to the README.md for detailed instructions.

## G LIDARMULTINETYE ET AL. (2023) BASELINE

The source code for LidarMultiNetYe et al. (2023) is not yet released. We’ve contacted the author for the code and they haven’t replied to us for 3 months. We implemented the code by ourselves following the descriptions in the paper and supplementary materials. To keep the fairness of comparison, we only include the GCP module, described as the main novelty contribution in the paper. The results in the main paper didn’t contain the object detection and BEV segmentation learning because they requires more supervision. The second stage refinement is also not included because it relies on the results of object detection and BEV segmentation.

We include the source code for our implementation in this source code folder. Please refer to the README.md for detailed instructions.

## REFERENCES

- Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Juergen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *ICCV*, 2019.
- Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, pp. 5828–5839, 2017.
- Pei Sun, Henrik Kretschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, and Benjamin Caine. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020.
- Dongqiangzi Ye, Zixiang Zhou, Weijia Chen, Yufei Xie, Yu Wang, Panqu Wang, and Hassan Foroosh. Lidarmultinet: Towards a unified multi-task network for lidar perception, 2023.



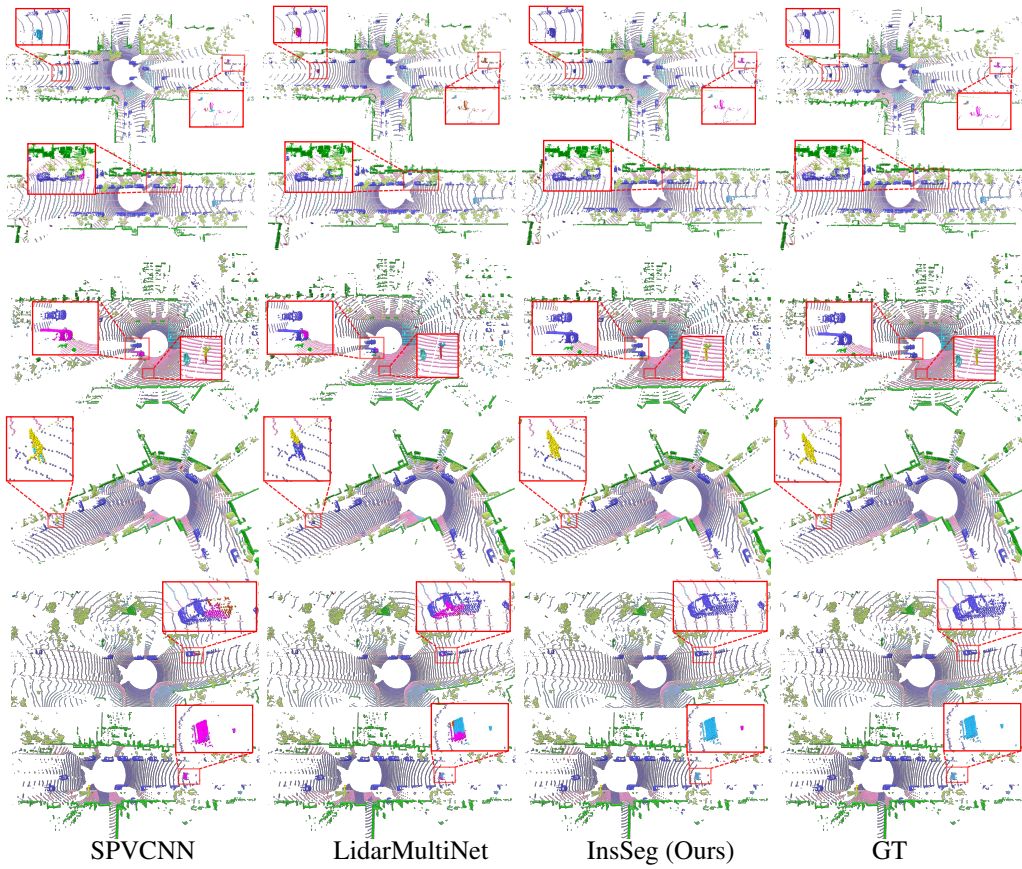


Figure 2: More qualitative results on Waymo Open DatasetSun et al. (2020).

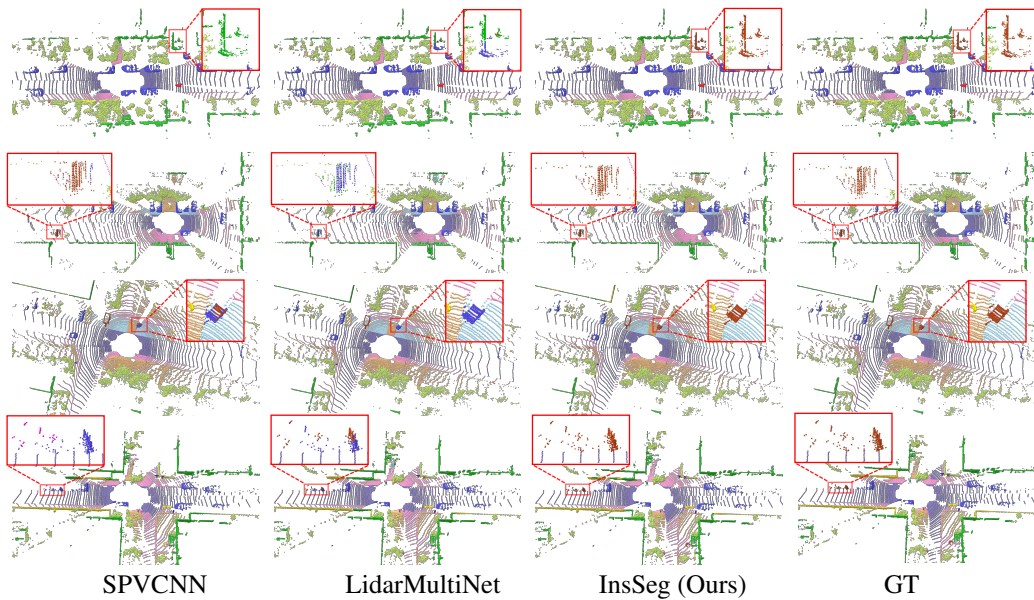


Figure 3: More qualitative results on SemanticKITTIBehley et al. (2019).

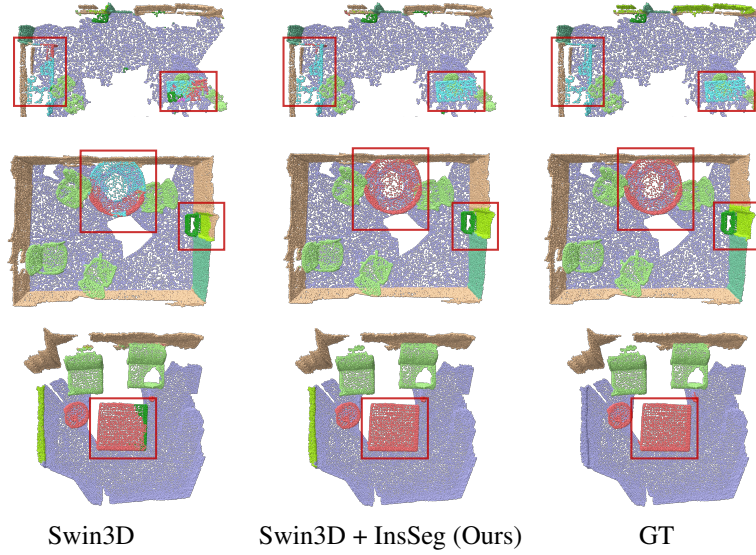


Figure 4: More qualitative results on ScanNetDai et al. (2017).

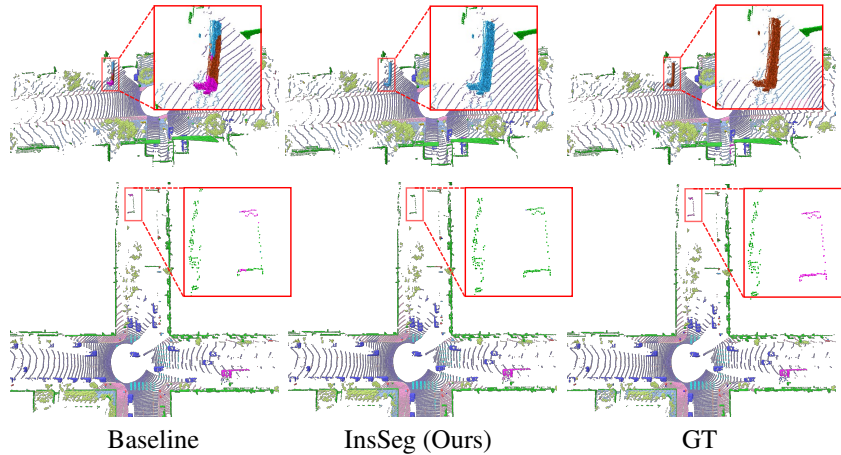


Figure 5: Failure cases. We show two examples where the our method made wrong classification and the whole instance is mis-segmented. This would cause lower IoU than the inconsistency results predicted by baseline.

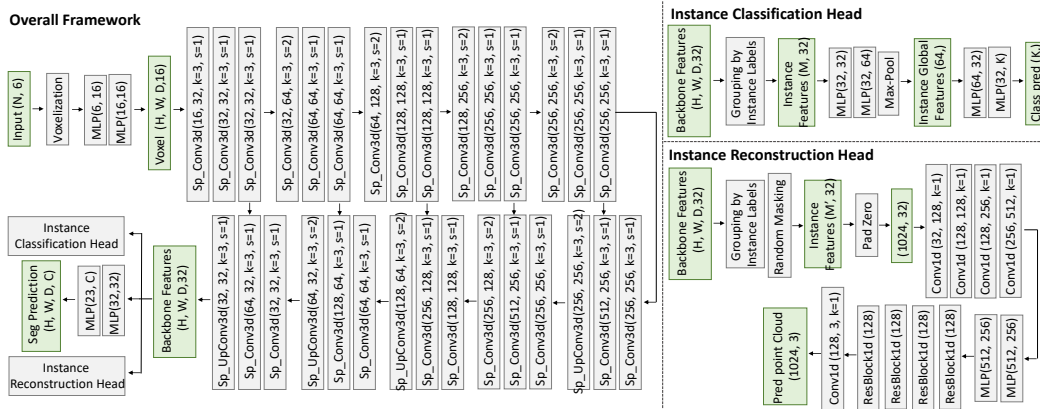


Figure 6: Detailed model architecture. The left part shows the overall framework, including the backbone and per-voxel segmentation head. The top right -art shows the architecture of the instance classification head. the botton right part shows the architecture of the shape reconstruction head. You can also find the source code for this in the attached code folder.