Supplementary Material for
Panoptic-PolarNet: Proposal-free LiDAR Point Cloud Panoptic Segmentation

1. Discussion

Choice of the loss: We adopted a combination of cross-entropy loss and Lovasz softmax loss in the semantic head. Given the highly imbalanced class distribution in LiDAR point clouds, the cross-entropy loss will favor those classes that are the majority of points, like the road class and the building class. Conversely, Lovasz softmax loss optimizes directly on the mIoU Jaccard index, which treats all classes equally. Combining these two losses will force the network to optimize toward an overall accurate prediction while focusing more on hard classes. In the instance head, we chose the MSE loss instead of focal loss [1] for the heatmap regression. The reason is that we do not necessarily need a very accurate prediction of the center in the BEV due to the scarcity of instance overlaps. However, we need a monotonically decreasing heatmap from the center to the edge to have a proper keypoint selection in the NMS. Also, the experiments showed that focal loss decreases the PQ by 1.3%.

Data augmentation: We apply instance oversampling to compensate for two imbalances in the LiDAR point cloud: (1) The imbalance between “thing” and “stuff”. Points belong to “thing” classes usually consist of only a small portion of the point cloud. (2) The imbalance between different “thing” classes. For example, the most occurring class, car, has around 10^7 time more points than the least occurring class, motorcyclist in the SemanticKITTI dataset. Experiments show that even though this oversampling will decrease the segmentation accuracy in “stuff”, the overwhelmingly increases in “thing” can still provide a huge improvement to the PQ and mIoU. Our experiments also show that either simply putting instance points at any place in a point cloud or transform it through its center will decrease the PQ. We conclude that such simple augmentation ignores projection properties, introduces inconsistency into the LiDAR point cloud, and thus entangles BEV feature learning.

Proposal-free vs. proposal-based: Even though proposal-based panoptic segmentation methods dominate in the 2D domain, there are only a few existing approaches for LiDAR point clouds. We think there are two reasons. First, proposal-based methods rely heavily on the annotation of bounding boxes, whereas point cloud datasets do not necessarily provide such annotations. Second, most current proposal-based object detection methods, like what we assume in our instance head, are not designed to represent the scene along the Z-axis. Lacking proper representation makes it more challenging to achieve a competitive result while maintaining speed when modified into a panoptic segmentation network.

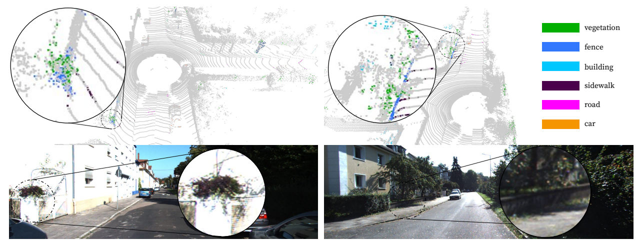


Figure 1: We highlight SAP-pruned points in color. Left: SAP prunes tangled vegetation and fence; Right: SAP prunes garden curbs that are annotated as the “ fence ”

End-to-end training: We only train the network to get an intermediate result and use a majority voting fusion to generate the final panoptic segmentation. Making the proposal-free panoptic segmentation network end-to-end trainable is still an open problem to explore in the future.

Self-adversarial Pruning Visualization: SAP is designed to remove ambiguous, noisy or/ and informative points. Since SemanticKITTI is a well-annotated dataset, we visually find SAP tends to remove challenging or ambiguous cases. Some examples are shown in Fig. 1.

2. Class-wise Results

We show the class-wise results of Panoptic-PolarNet on SemanticKITTI and nuScenes in Table 1 and Table 2. Our method has a similar panoptic segmentation performance in the corresponding classes among these two datasets. The low performance comes from the class that either has a small physical shape (like bicycle) or has a small number of instances in the dataset (like truck and construction vehicle). Despite being a more challenging dataset due to its significantly higher number of instances, nuScenes has fewer classes than SemanticKITTI, which makes it more distin-

Table 1: Class-wise results on **test** split of SemanticKITTI.

metrics	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mean
PQ	88.8%	33.0%	51.8%	35.2%	37.6%	57.3%	69.9%	52.4%	88.6%	42.6%	61.2%	1.6%	85.7%	46.0%	75.7%	54.5%	41.5%	48.5%	56.4%	54.1%
RQ	96.2%	46.7%	59.9%	38.7%	41.5%	65.7%	78.6%	57.7%	96.8%	56.2%	76.3%	2.8%	92.1%	62.3%	92.5%	73.6%	55.5%	66.0%	75.2%	65.0%
SQ	92.3%	70.7%	86.4%	90.9%	90.5%	87.2%	88.9%	90.9%	91.5%	75.9%	80.3%	55.6%	93.0%	73.8%	81.9%	74.1%	74.9%	73.4%	75.0%	81.4%
IoU	94.4%	38.7%	48.2%	46.2%	34.5%	51.1%	63.9%	24.9%	90.8%	61.3%	74.6%	16.5%	89.9%	61.1%	83.4%	66.7%	68.0%	56.8%	58.5%	59.5%

Table 2: Class-wise results on **validation** split of nuScenes.

metrics	barrier	bicycle	bus	car	construction vehicle	motorcycle	pedestrian	traffic cone	trailer	truck	driveable surface	other flat	sidewalk	terrain	manmade	vegetation	mean
PQ	41.5%	58.0%	70.4%	89.0%	36.4%	78.4%	85.1%	80.7%	49.7%	63.3%	95.7%	53.9%	67.7%	49.8%	84.3%	80.0%	67.7%
RQ	54.4%	68.6%	76.1%	95.9%	44.9%	86.0%	95.2%	91.8%	58.2%	69.4%	100.0%	66.3%	85.3%	65.1%	98.2%	94.7%	78.1%
SQ	76.3%	84.6%	92.5%	92.8%	81.2%	91.2%	89.4%	87.9%	85.3%	91.3%	95.7%	81.3%	79.4%	76.5%	86.0%	84.5%	86.0%
IoU	52.3%	28.1%	88.0%	90.3%	32.4%	71.7%	72.2%	52.8%	58.3%	76.6%	95.9%	68.8%	74.3%	73.4%	87.3%	85.5%	69.3%

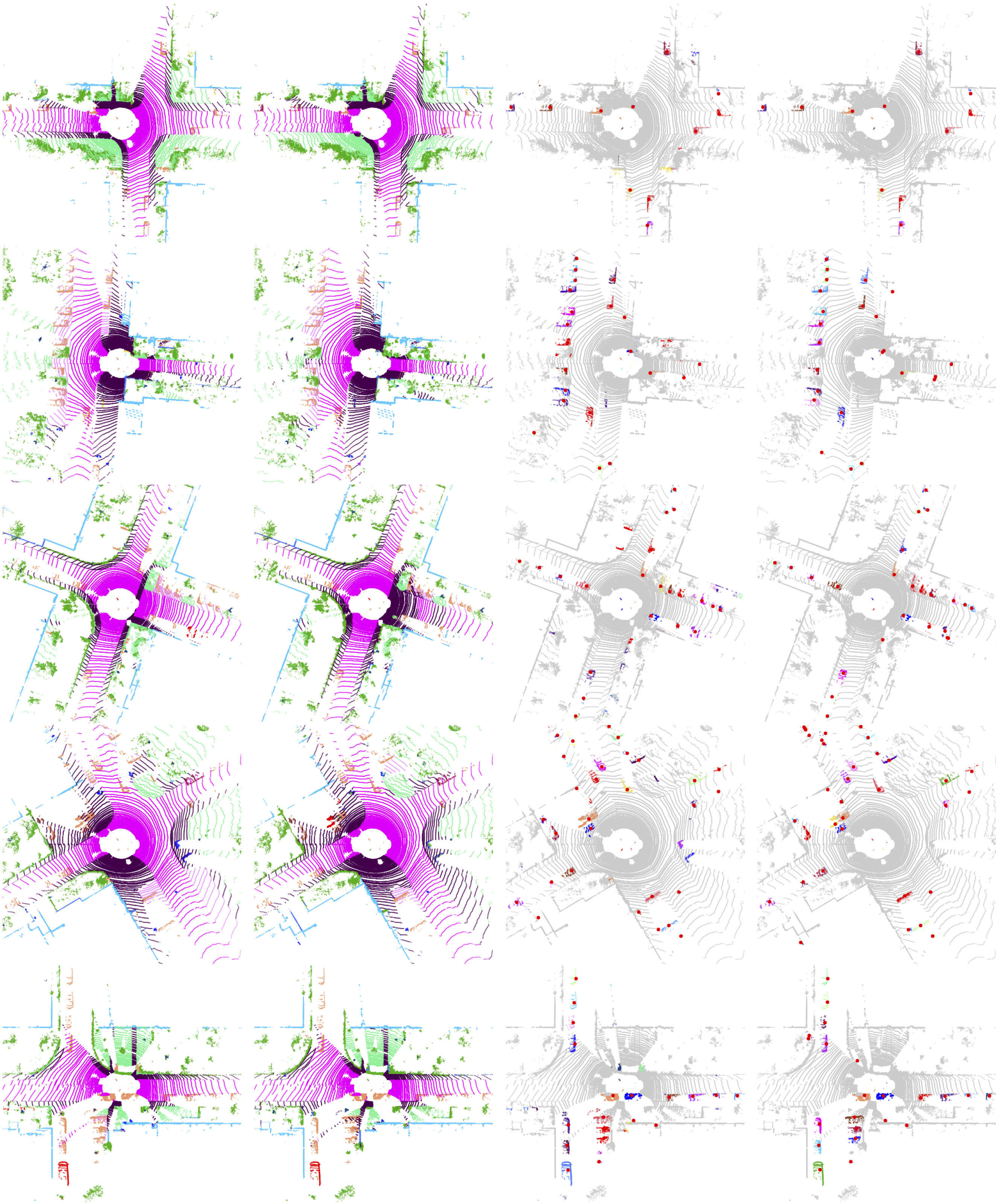
guishable and thus having higher PQ and mIoU.

3. Qualitative Results

We show the visualization examples of Panoptic-PolarNet on SemanticKITTI and nuScenes in Figure 2 and Figure 3 respectively. Our method can make accurate instance predictions regardless of the distance and point density variation. We can also visually verify that nuScenes has significantly more instances than SemanticKITTI. And most of those instances belong to some challenging classes that have a small number of points. There are also duplicated instance predictions within a short distance. This could be fixed by introducing class-wise prior knowledge in the grouping stage in the future.

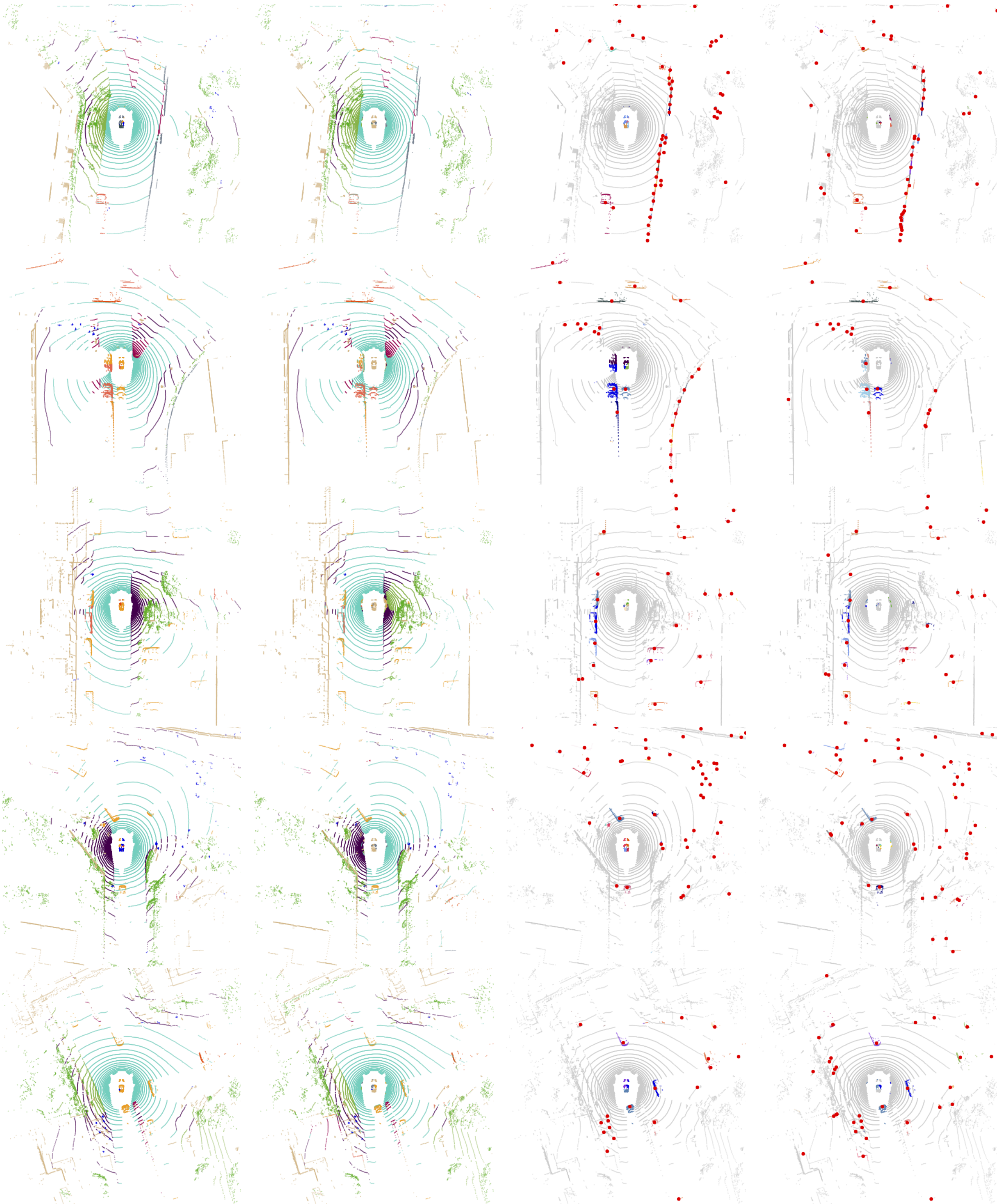
References

[1] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.



(a) Semantic Ground Truth (b) Semantic Prediction (c) Instance Ground Truth (d) Instance Prediction

Figure 2: Visualization of Panoptic-PolarNet on the SemanticKITTI dataset. The red dots in the instance prediction represent the center for each instance.



(a) Semantic Ground Truth (b) Semantic Prediction (c) Instance Ground Truth (d) Instance Prediction

Figure 3: Visualization of Panoptic-PolarNet on the nuScenes dataset. The red dots in the instance prediction represent the center for each instance.