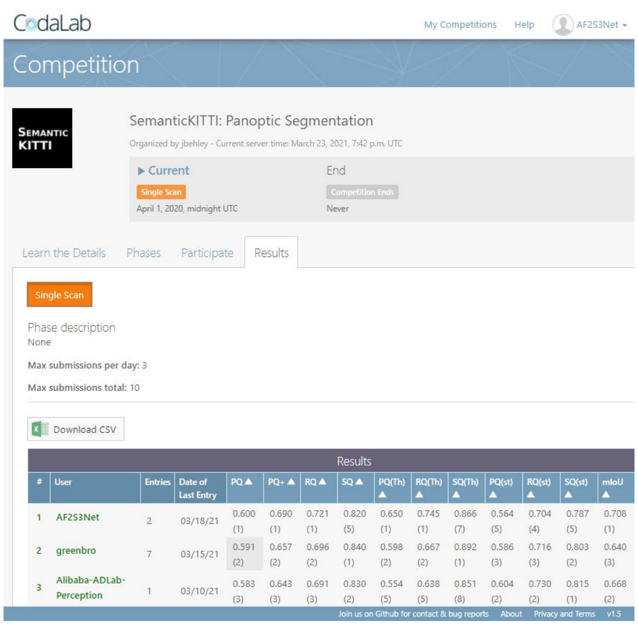
Supplementary Material for
GP-S3Net: Graph-based Panoptic Sparse Semantic Segmentation Network

To better evaluate the proposed method, provide the SemanticKITTI benchmark results and more detailed test results related to SemanticKITTI, we include the supplementary materials with the following information.

In Fig. 1, we present a screen-shot of the public leaderboard of SemanticKITTI, taken at 2021-03-17. Our method achieves the first rank with mean PQ of 60%. We also provide detailed class-wise comparison of our method with state-of-the-art approaches based on different metrics of PQ, RQ, and SQ. It can be observed that the the proposed method is superior compared to other benchmarks on PQ and RQ with a large margin of 2.6% and 3.4%, respectively. The values of SQ are within the range of other approaches.



Results

#	User	Entries	Date of Last Entry	PQ	PQ+	RQ	SQ	PQ(Th)	RQ(Th)	SQ(Th)	PQ(st)	RQ(st)	SQ(st)	mIoU
1	AF2S3Net	2	03/18/21	0.600 (1)	0.690 (1)	0.721 (1)	0.820 (5)	0.650 (1)	0.745 (1)	0.866 (7)	0.564 (5)	0.704 (4)	0.787 (5)	0.708 (1)
2	greenbro	7	03/15/21	0.591 (2)	0.657 (2)	0.696 (2)	0.840 (1)	0.598 (2)	0.667 (2)	0.892 (1)	0.586 (3)	0.716 (3)	0.803 (2)	0.640 (3)
3	Alibaba-ADLab-Perception	1	03/10/21	0.583 (3)	0.643 (3)	0.691 (3)	0.830 (2)	0.554 (5)	0.638 (5)	0.851 (8)	0.604 (2)	0.730 (2)	0.815 (1)	0.668 (2)

Figure 1: Screenshot of the SemanticKITTI public leaderboard. Our method achieves $\mathbf{1^{st}}$ with mean PQ of 60%.

Moreover, some examples from the validation set of SemanticKITTI are provided to help the reader in analyzing the experimental results. Fig. 2 is one of the challenging frames with large quantity of instances in SemanticKITTI validation sequence. This graph-based approach is highly effective to handle these complex scenes as observed. Note that the colors on each instance differ when comparing to the ground truth since the predicted instance ID is not identical with the labeled ones.

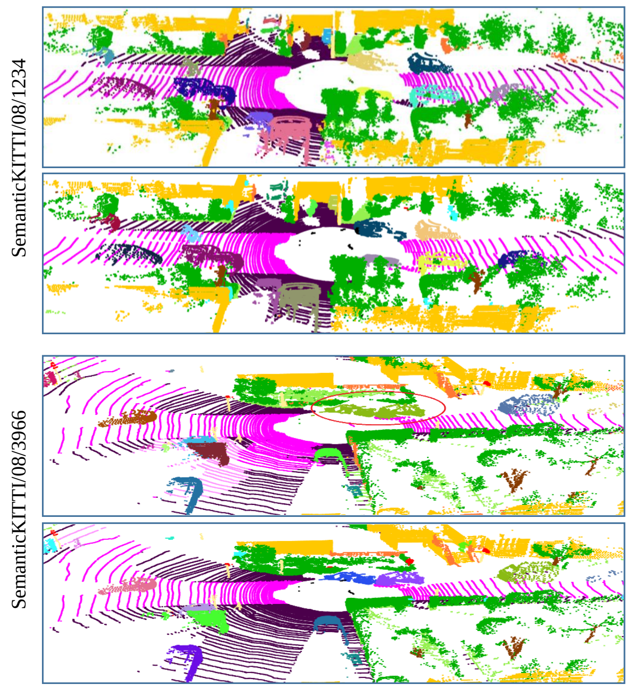


Figure 2: Two examples of LiDAR panoptic segmentation. For each frame, prediction (top) and ground truth (bottom) are shown from SemanticKITTI validation set.

We realize most of the errors occur when the *things* classes are close to each other on the side of the ego vehicle, as shown in Fig. 2. In this case, the model segments the two parked vehicles as one instance. This is due to the limitation of the graph construction step where we downsample the point cloud to clusters via HDBSCAN [3]. This process could potentially cluster different instances together when they are very close to each other. Thus, future works include designing an efficient and effective clustering algorithm to build the graph from semantic information.

Method	Car	Truck	Bicycle	Motorcycle	Other-vehicle	Person	Bicyclist	Motorcyclist	Road	Sidewalk	Parking	Other-ground	Building	Vegetation	Trunk	Terrain	Fence	Pole	Traffic-sign	mPQ
R.Net [23] + P.P [21]	66.9	6.7	3.1	16.2	8.8	14.6	31.8	13.5	90.6	63.2	41.3	6.7	79.2	71.2	34.6	37.4	38.2	32.8	47.4	37.1
KPC. [30] + P.P [21]	72.5	17.2	9.2	30.8	19.6	29.9	59.4	22.8	84.6	60.1	34.1	8.8	80.7	77.6	53.9	42.2	49.0	46.2	46.8	44.5
LPSAD [22]	76.5	7.1	6.1	23.9	14.8	29.4	29.7	17.2	90.4	60.1	34.6	5.8	76.0	69.5	30.3	36.8	37.3	31.3	45.8	38.0
PanopticTrackNet [17]	70.8	14.4	17.8	20.9	27.4	34.2	35.4	7.9	**91.2**	66.1	50.3	10.5	81.8	75.9	42.0	44.3	42.9	33.4	51.1	43.1
Panoster [10]	84.0	18.5	36.4	44.7	30.1	61.1	69.2	51.1	90.2	62.5	34.5	6.1	82.0	77.7	55.7	41.2	48.0	48.9	59.8	52.7
DS-Net [15]	**91.2**	28.8	45.4	47.2	34.6	63.6	71.1	58.5	89.1	61.2	32.3	4.0	83.2	79.6	**58.3**	43.4	50.0	55.2	65.3	55.9
EfficientLPS [29]	85.7	30.3	37.2	47.7	**43.2**	70.1	66.0	44.7	91.1	**71.1**	**55.3**	**16.3**	**87.9**	**80.6**	52.4	**47.1**	**53.0**	48.8	61.6	57.4
GP-S3Net [Ours]	84.1	**36.7**	**41.8**	**77.0**	42.1	**73.0**	**81.1**	**84.2**	91.1	64.5	38.9	11.9	80.4	68.1	53.2	37.2	46.9	**55.7**	**72.4**	**60.0**

Table 1: Class-wise PQ scores on SemanticKITTI test dataset.

Method	Car	Truck	Bicycle	Motorcycle	Other-vehicle	Person	Bicyclist	Motorcyclist	Road	Sidewalk	Parking	Other-ground	Building	Vegetation	Trunk	Terrain	Fence	Pole	Traffic-sign	mRQ
Panoster [10]	92.9	22.3	50.6	54.9	36.2	72.3	77.5	61.4	99.0	77.5	46.9	8.7	88.9	92.8	74.5	55.3	63.5	66.3	77.0	64.1
DS-Net [15]	**97.5**	32.4	**62.2**	56.3	38.9	74.3	78.4	62.7	96.8	76.7	42.6	6.4	89.3	95.7	**77.5**	58.3	65.5	74.0	81.9	66.7
EfficientLPS [29]	94.3	33.0	54.1	54.5	46.8	79.2	74.7	47.6	97.4	**86.1**	**69.5**	**22.9**	**94.7**	**96.2**	73.8	**62.1**	**69.4**	69.0	79.7	68.7
GP-S3Net [Ours]	92.4	**41.8**	58.2	**89.3**	**48.3**	**83.7**	**90.5**	**92.0**	**99.6**	82.5	51.1	16.3	87.8	86.8	72.7	50.3	61.4	**75.8**	**89.7**	**72.1**

Table 2: Class-wise RQ scores on SemanticKITTI test dataset.

Method	Car	Truck	Bicycle	Motorcycle	Other-vehicle	Person	Bicyclist	Motorcyclist	Road	Sidewalk	Parking	Other-ground	Building	Vegetation	Trunk	Terrain	Fence	Pole	Traffic-sign	mSQ
Panoster [10]	90.4	82.6	**71.9**	81.4	83.1	84.5	89.3	83.1	91.1	80.7	73.5	69.8	92.2	83.8	74.7	74.5	75.7	73.7	77.7	80.7
DS-Net [15]	**93.6**	88.9	71.0	83.8	89.0	85.6	**90.7**	93.3	92.0	79.8	75.8	61.4	**93.2**	83.2	**75.2**	74.4	76.3	**74.5**	79.7	82.3
EfficientLPS [29]	90.9	**92.0**	68.8	**87.5**	**92.2**	**88.5**	88.3	**94.0**	**93.5**	**82.6**	**79.5**	71.0	92.9	**83.9**	70.9	**75.7**	**76.5**	70.8	77.3	**83.0**
GP-S3Net [Ours]	91.1	87.9	71.8	86.2	87.3	87.2	89.6	91.6	91.4	78.2	76.0	**72.7**	91.7	78.4	73.2	73.9	76.4	73.5	**80.6**	82.0

Table 3: Class-wise SQ scores on SemanticKITTI test dataset.