Weakly Supervised Segmentation on Outdoor 4D point clouds with Temporal Matching and Spatial Graph Propagation

A. Supplementary Material

A.1. Quality of Pseudo Label

To further understand our temporal and spatial propagation mechanisms, we compare the quality of pseudo labels in each stage in Table A. Note that all the results in Table 1 are models trained with the corresponding updated pseudo labels in Table A. The labels used in Baseline-A are the initial annotations augmented with the super-voxel segmentation, which contains 0.0057% points in total, as stated in Section 3. After approaching the Baseline-A, the naive pseudo label generation in Baseline-B directly updates the points with high confidence scores as the pseudo labels for the model training.

In SPP stage, our temporal matching(both greedy matching(Model-A) and optimal transport(Model-B)) generates 1000 times more pseudo labels, and around 5% improvement of the pseudo label quality. Considering low qualities and similar quantities of pseudo labels in Model-A and Model-B, the performances of Model-A and Model-B are similar.

After training a new model in SPP stage, we further update the pseudo labels with our temporal matching and spatial propagation in DSP stage. Combining temporal matching and spatial propagation(Model-D and Model-E) generates about 20 times more than onnly temporal matching. However, there are ∼20% drawbacks of the pseudo label quality with spatial propagation. Referring to Table 1, the quantity of pseudo labels in the DSP stage benefits more than just the quality of pseudo labels.

	Baseline-A	Baseline-B	Model-A	Model-B	Model-C	Model-D	Model-E
U-Pre	0.001	20.0	1.0	1.0	20.0	20.0	20.0
mIoU	73.1	46.7	77.7	78.1	51.4	56.2	56.6

Table A. **The performance of the pseudo label in the training set of SmeanticKITTI**.U-Pre is the percentage of updated points. Initial is the initial annotation with supervoxel segmentation.

A.2. More Qualitative Results of Generated Pseudo Labels on SemanticKITTI

In this section, we show more qualitative results on SemanticKITTI in Figures A, respectively. Through these results, we demonstrate that our proposed method generate a high quality pseudo labels.

A.3. Limitation

Although our experiments show the effectiveness of our proposed framework, there are still two major limitations in our framework. Firstly, as shown in Figure 7, building long-range connections between annotated regions and distance frames is still a challenge in our setting. The frames far from the first frame do not share many annotated regions with the first frame, and the valuable objects, like cars and people, in the frames far from the first frame are usually not in the annotated regions. Consequently, the quantity and quality of pseudo labels decrease significantly with the frame distance increasing, limiting our proposed method's performance. Secondly, there are many hyper-parameters in both temporal matching and spatial graph propagation. The choices of hyper-parameters in our two-stages framework costs extra efforts to approach the best performance of our work on different datasets. Moreover, the two-stage design increases the uncertainty of the training process. The uncertainty of the training also costs extra efforts. As our work aims to achieve a effective model with minimum effort, simplifying the two-stage framework is a valuable topic.

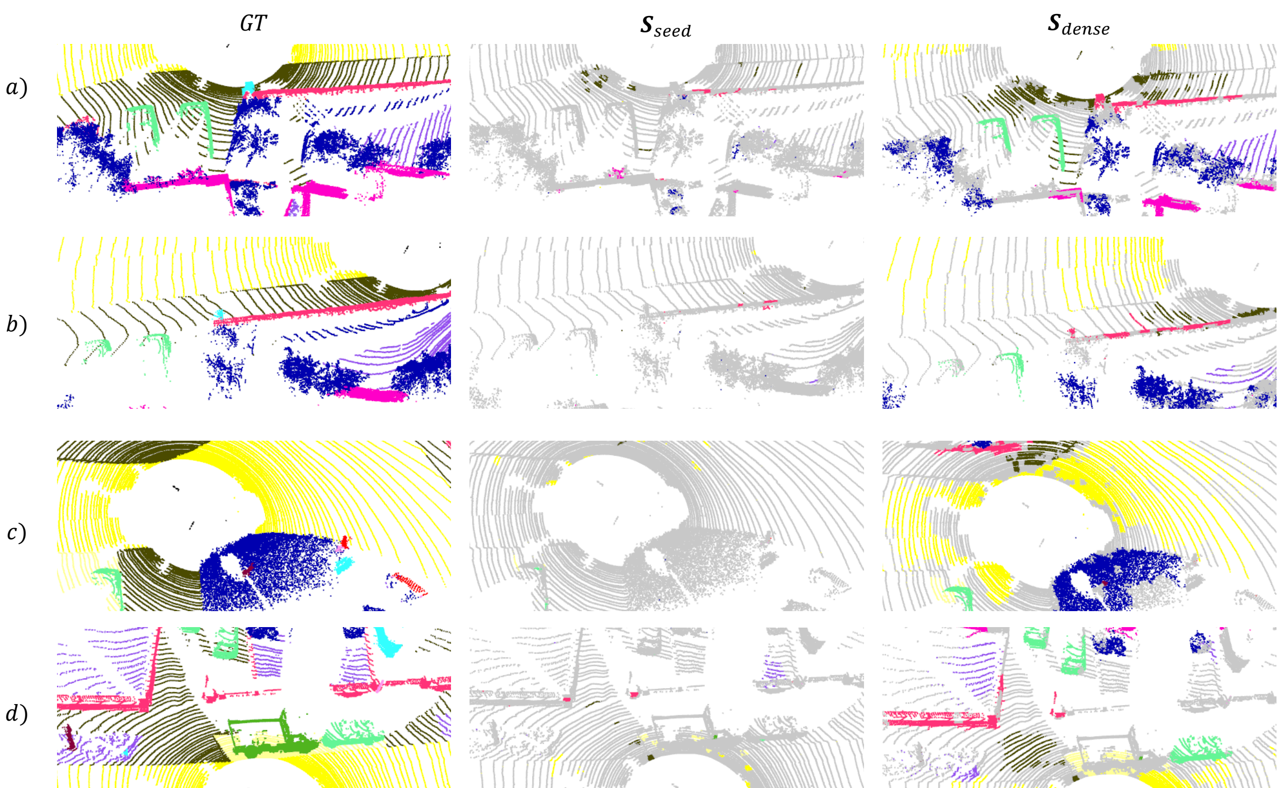

Figure A. **More Qualitative results of generated pseudo labels in stage 1 and stage 2**.