Week 1 Report

Doğaç Güzelkokar, Niklas Bubeck

June 2022

1 Week 1 Goals

The main goal of the first week was to determine the project direction in terms of data domains and architectural or methodical improvements. For this purpose, we reviewed a list of papers on the tasks of panoptic segmentation and 4D panoptic segmentation. Additionally, we looked into the RADAR sensor from CARLA.

2 Literature Review

We mainly reviewed the papers on 4D panoptic segmentation next to papers on the panoptic segmentation. In essence, 4D panoptic segmentation works focus on successfully associating instances across consecutive scans, with less emphasis on architectural improvements on the panoptic segmentation network, where panoptic segmentation and instance association are separately done.

In the "4D Panoptic LiDAR Segmentation" paper, the authors get the semantic labels and instance labels in parallel. In the instance segmentation part, the main steps can be summarized like this: first, the object centers are found, then, the embeddings and variations for each point are calculated. Following this, these calculated values are taken from only the "thing"(cars, persons, etc.) points from the previous n scans and all points in the current scan, consecutively, each point in the current scan is assigned to an instance, based on a normal distribution where the distribution centers are object centers and the point variations are received from the model output, if the probability of a point belonging to an instance is higher than some threshold value, then that point is assigned to that instance.

Another approach, which overtook the above-mentioned work on the SemanticKITTI benchmark, was "Contrastive Instance Association for 4D Panoptic Segmentation using Sequences of 3D LiDAR Scans". In this work, semantic segmentation is also done in parallel. What they are doing differently is that they first get instance segmentations for each scan, then, they find the same instance in different scans, following up on this, a contrastive learning process is set up, using the same instances in different scans as positive samples and all other instances as negative samples. Through this, they can produce point embeddings where the points belonging to the same instance have similar embeddings, yet points belonging to different instances have different embeddings. Finally, having these embeddings, they find the closest matches between instances in different scans based on the embedding distances.

3 Discussion

Based on the existing works, we came up with a few points of focus. Working on the SemanticKITTI dataset would be more convenient as we think there could be better approaches still based on point clouds.

Panoptic segmentation seems not to be the main focus for 4D panoptic segmentation. Maybe here, we can be content with a pretrained panoptic segmentation network, using the SOTA network as our backbone. Similar to the contrastive learning approach, we could look into different ways to associate instances across scans. Also, even though the contrastive learning approach involves motion information implicitly, we wonder that if the motion information is given explicitly, would it improve the results? Additionally, would adding non-contrastive embeddings(to have better performance in local regions), and having two embedding spaces where the overall distance metric between instances is a linear combination of the distances in these two embedding spaces improve the results?