Weijie Wei   Martin R. Oswald   Fatemeh Karimi Nejadasl   Theo Gevers
University of Amsterdam
{w.wei2, m.r.oswald, f.kariminejadasl, th.gevers}@uva.nl

Abstract

This supplementary material provides additional details about the implementation.

A. Implementation details

A-branch. We adopt HRNet [4] with object-contextual representations (OCR) [5], denoted as HRNet-OCR, as the backbone for A-branch. During training, the OCR loss is preserved while the original 2D segmentation head is removed. The intermediate features, also known as augmented representations as defined in the original paper, from HRNet-OCR are compressed to a total of 128 channels, thereby ensuring alignment with the output of the P-branch.

P-branch. We employ RandLA-Net [2] as the backbone for the P-branch and follow its official configuration for the SemanticKITTI dataset [1] with the following two modifications: Firstly, we double all feature channels in the RandLA-Net to accommodate the additional color features. Furthermore, we double the output channel for the last layer to ensure compatibility with the A-branch. Consequently, the encoder produces outputs with channel dimensions of 64, 128, 256, and 512, respectively. Secondly, we input the same point cloud to RandLA-Net twice and sum up the output features. Although the network does not change, the down-sampling within the network is random, leading to different features for the same point cloud in the end. This technique promotes the consistency of RandLA-Net.

GAF module. We adopt KPConv [3] as the point convolution in the GAF module and adhere to the configuration of the rigid KPConv. Accordingly, one single rigid KPConv encompasses a sphere with a radius of 0.5 meters, centered at the query point. Each kernel point exerts an influence on all support points within a sphere whose radius is 0.24 meters and centered on the kernel point.

References