International Journal of Automotive Technology, Vol. 24, No. 3, pp. 787–797 (2023)
DOI 10.1007/s12239–023–0065–y

pISSN 1229–9138/ eISSN 1976–3832

IMPROVED 3D SEMANTIC SEGMENTATION MODEL BASED ON RGB IMAGE AND LIDAR POINT CLOUD FUSION FOR AUTOMANTIC DRIVING

Jiahao Du, Xiaoci Huang*, Mengyang Xing and Tao Zhang

School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

(Received 21 March 2022; Revised 2 November 2022; Accepted 9 January 2023)

ABSTRACT–LiDAR point cloud semantic segmentation algorithm is crucial to the environmental understanding of unmanned vehicles. At this stage, in autonomous vehicles, effectively integrating the complementary information of LiDAR and camera has become the focus of research. In this work, a network framework (called PI-Seg) for LiDAR point clouds semantic segmentation by fusing appearance features of RGB images is proposed. In this paper, the perspective projection module is introduced to align and synchronize point clouds with images to reduce appearance information loss. And an efficient and concise dual-flow feature extraction network is designed, a fusion module based on a continuous convolution structure is used for feature fusion, which effectively reduces the amount of parameters and runtime performance, and more suitable for autonomous driving scenarios. Finally, the fused features are added to the LiDAR point cloud features as the final output features, and the point cloud category label prediction is realized through the MLP network. The experimental results demonstrate that PI-Seg has a 5.3% higher mIoU score than SalsaNext, which is also a projection-based method, and still has a 1.4% performance improvement compared with the latest Cylinder3D algorithm, and in quantitative analyses the mAP value also has the best performance, showing that PI-Seg is better than other existing methods.

KEY WORDS : Point cloud semantic segmentation, Multi-sensor fusion, Perspective projection, Dual-flow network

1. INTRODUCTION

An autonomous vehicle is that can achieve fully autonomous driving in a dynamic environment. To ensure high-precision perception of the surroundings, self-driving vehicle possess different sensors and efficient algorithms. Semantic scene understanding is the fundamental task of autonomous driving, which provides accurate environmental information for path planning and behavioral decision-making to ensure the safety of autonomous driving. And the primary method of scene understanding is to assign class labels to the input data through semantic segmentation to help unmanned vehicles recognize and understand the surrounding environment (Cao *et al*., 2020; Wang *et al*., 2021).

With the help of the deep learning and many of open-source datasets, 2D semantic segmentation methods based on camera sensors have made significant progress. Since RGB images have rich appearance information, the 2D semantic segmentation method can provide more detailed segmentation results However, the RGB image does not have spatial information, and the camera is vulnerable to changes in lighting, which leads to poor semantic scene understanding in autonomous driving. Therefore, researchers want to make up for the shortcomings of camera sensors through 3D semantic segmentation.

So far, the research on 3D semantic segmentation has mainly focused on three directions. (1) Based on 3D voxel (Wu *et al*., 2015; Maturana and Scherer, 2015; Han *et al*., 2016). Before directly processing the point cloud, the researchers utilize a 3D voxel grid to fully preserve the 3D shape information. However, due to the low voxel ratio on the surface of the 3D shape, sparse voxelization may result. And the resolution of the 3D voxel cannot be too high, otherwise the network structure will be too complicated. (2) Based on original point cloud. Qi *et al*. (2017) designed PointNet, which is a pioneering work that directly processes the original 3D point set. Based on this, subsequent research has achieved a good segmentation effect. However, as shown in Figure 1 left, due to the sparseness of the collected outdoor scene point clouds and the lack of semantic appearance information, this method has a high error in fine-grained segmentation tasks. Therefore, for the shortcomings of the two methods of using only the camera and only using the LiDAR, a straightforward solution is to fuse perceptual data from two sensors for 3D semantic segmentation. (3) Based on fusion images (Qi *et al*., 2016; Kalogerakis *et al*., 2017; Cortinhal *et al*., 2020; Kochanov *et al*., 2020). These

**Corresponding author*. e-mail: 06060005@sues.edu.cn