Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 9681455, 11 pages
https://doi.org/10.1155/2022/9681455

Research Article

The Robust Semantic SLAM System for Texture-Less Underground Parking Lot

Chongjun Liu and Jianjun Yao

College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Chongjun Liu; chongjunliu@hrbeu.edu.cn

Received 21 October 2021; Revised 9 April 2022; Accepted 19 April 2022; Published 6 July 2022

Academic Editor: Peng Hang

Automatic valet parking (AVP) is the autonomous driving function that may take the lead in mass production. AVP is usually needed in an underground parking lot, where the light is dim, the parking space is narrow, and the GPS signal is denied. The traditional visual-based simultaneous location and mapping (SLAM) algorithm suffers from localization loss because of inaccurate mapping results. A new robust semantic SLAM system is designed mainly for the dynamic low-texture underground parking lot to solve the problem mentioned. In this system, a 16-channel Lidar is used to help the visual system build an accurate semantic map. Four fisheye cameras mounted at the front, back, left, and right of the vehicle are also used to produce the bird's eye view picture of the vehicle by joint calibration. The vehicle can localize itself and navigate to the target parking lot with the semantic segmented picture and the preobtained semantic map. Based on the experiment result, the proposed AVP-SLAM solution is robust in the underground parking lot.

1. Introduction

The public traffic jam situation worsens with the increasing number of automobiles. Researchers in the automobile field are now devoting their effort to automatic driving systems to ease traffic pressure and present a safe way for driving. As one of the most promising and meaningful functions in automatic driving, the automatic valet parking (AVP) system has become the focus of scholars because it can provide drivers, particularly the new ones, an achievable and safe way to park vehicles under crowded parking conditions. This function can be achieved by providing vehicles with a high-definition (HD) map for vehicle path planning. Thus, the AVP function is achievable if the HD map and the global positioning system-inertial measurement unit (GPS-IMU) camera-based localization method can be used to locate a vehicle at a preknown place. However, a vehicle cannot possibly acquire environmental knowledge when this vehicle is located in an unknown place. A vehicle must locate itself and build an environment map while moving by itself to overcome the difficulty mentioned. Therefore, the SLAM problem was proposed in 1986 [1].

SLAM technology can be divided into two categories, namely, Lidar and Vision, depending on the sensors used. Lidar-based SLAM schemes are extensively analyzed by researchers [2]. Lidar can measure the angle and distance of obstacle points with higher accuracy, which is convenient for positioning and navigation. Lidar-based SLAM has high accuracy and no cumulative error when building maps. Excellent performance and dense point clouds can be obtained using the 3D Lidar. However, the 3D Lidar with 64 channels is expensive for commercialization [3]. The corridor of the underground parking lot is long and straight, with smooth walls on both sides, and it is easy to lose positioning only by relying on Lidar-based SLAM. Therefore, the vision-SLAM system receives attention from researchers worldwide because of its high perception ability and low cost.

In addition to vision-SLAM, other traditional feature methods and road-based feature methods are available. In traditional feature methods, sparse points, lines, and dense

planes in a real environment are taken as geometrical features, which can be used for vehicle localization [4]. Furthermore, corner features are widely used for visual mileage calculation [5–7]. The pose of the camera and feature positions can be estimated with these methods. Moreover, SIFT [8], SURF [9], BRIEF [10], and ORB [11] descriptors are widely used by researchers in describing the features to make these features unique. ORB-SLAM [12, 13] is a representative SLAM framework based on nonlinear optimization. The ORB features are used as tracking feature points while driving.

For the methods based on road features, lane lines, curbs, and traffic signs are widely used as landmarks, which can be used to localize the camera pose by comparing the landmarks with previously established maps. Compared with the traditional-feature-based method, road-feature-based methods use these landmarks, which are robust even the illumination conditions change. Yan [14] proposed a nonlinear optimization problem to localize a 6-DOF camera pose in terms of localization. In this method, the geometry of road markings and the odometry and epipolar geometry constraints of the vehicle were considered. The experiment results showed that submeter localization error is achieved on the road with sufficient road markings. Schreiber et al. [15] came up with a novel approach to establish precise and robust localization by using a stereo camera system and a highly accurate map with curbs and road markings. In this method, global navigation satellite systems are used only to obtain the initial location, and they are not used during a 50 km test. Ranganathan et al. [16] presented a new scheme for precise localization. In this scheme, the signs marked on the road were used to localize the automobile in a global coordinate. Furthermore, the mechanism combining road-mark-based map and sparse-feature-based map was adopted to obtain a high localization accuracy. In addition to location, many studies focus on mapping. Rehder et al. [17] proposed a novel approach to generate the local grid map by detecting the lane on the image taken by a camera. A globally consistent map can be constructed with the help of the local grid map. Jeong et al. [18] proposed a road-SLAM algorithm by considering the road markings obtained from images taken by a camera. In this algorithm, the random forest method was used to improve the matching accuracy by using a submap containing road information. Based on the experiment results, the accuracy of this mapping method can be improved to 1.098 m over 4.7 km of the path length. This result was validated by comparing the obtained data with the data from RTK-GPS.

In addition to pure vision odometer, vision-aided inertial navigation algorithm is becoming increasingly popular in the autonomous driving field. In this scheme, IMU is added into a vision-based scheme to improve localization precision. Mourikis et al. [19] proposed an extended Kalman filter-based algorithm for real-time vision-aided inertial navigation. The result showed that a very accurate pose estimation can be conducted with this sensor-fusing algorithm. Leutenegger et al. [6] came up with a keyframe-based visual-inertial odometry scheme. Although this scheme demands considerable computation resources, superior accuracy performance was obtained. The monocular visual-inertial system is the most commonly used VIO algorithm at present [6, 12]. In this scheme, a camera and a low-cost IMU are used to obtain high-accuracy localization.

Semantic segmentation is a new image clustering task at the pixel level, and it is widely applied in perception in the automatic driving domain and medical image diagnosis [20–23]. In recent years, deep convolution neural network has been widely used in semantic segmentation tasks [24], and the majority of the networks are based on various convolution network structures. Among them, U-Net [20, 25] is widely accepted and improved as the basic network that can be trained with pictures taken by a camera and could classify pixels of the pictures into parking lines and signs. The basic framework of U-Net is shown in Figure 1. These classified results are critical data for building maps or localizing automobiles. The residual network [26] can achieve good results to adapt to highly complex segmentation scenarios through a very deep layer depth and a large number of parameters. However, lightweight networks, such as ERF-Net [27], consider the real-time performance and accuracy with the method of distillation [28] to be deployed to edge computing devices, such as onboard computers.

Automotive valet parking is a complex function that can be equipped on vehicles and help drivers, particularly new drivers, park their cars in a carport. However, the light conditions in underground parking lots are usually very dim, and smooth walls, floors, and columns can be found inside. All these conditions complicate the parking task. Moreover, traditional vision equipment is influenced and becomes unfit in this scenario. In order to solve the above problems, we first build a vehicle platform with a 16-channel Lidar and four surrounding cameras. The robot operating system (ROS) is adopted to call both the camera and Lidar for collecting data. Also, a method consisting of image semantic segmentation, lidar supplemental mapping, semantic mapping, and localization is proposed. The rest of this paper is organized as follows. The detailed system architecture is introduced in Chapter 2. Then, the methodology proposed in this study is described in chapter 3. Finally, the experiment results are presented to show the robustness of the proposed AVP-SLAM solution in chapter 4, and our conclusion is drawn in chapter 5.

2. System Architecture

Four surround-view cameras and a 16-channel Lidar are applied in the proposed mapping and localization system, as shown in Figure 2. The framework for this system consists of two parts. One is offline mapping, and the other is for localization. For the offline mapping system, the 16-channel Lidar is used to provide the odometry and build the point cloud map. The semantic information is added to this map by keyframe matching. We select Lidar keyframes at 0.2 seconds intervals. We use ROS to call both the camera and Lidar for collecting data, so we have their own time stamps in the header of the message. Based on the time stamp, we select the semantic map corresponding to the Lidar to overlay the semantic map according to the position and posture of the

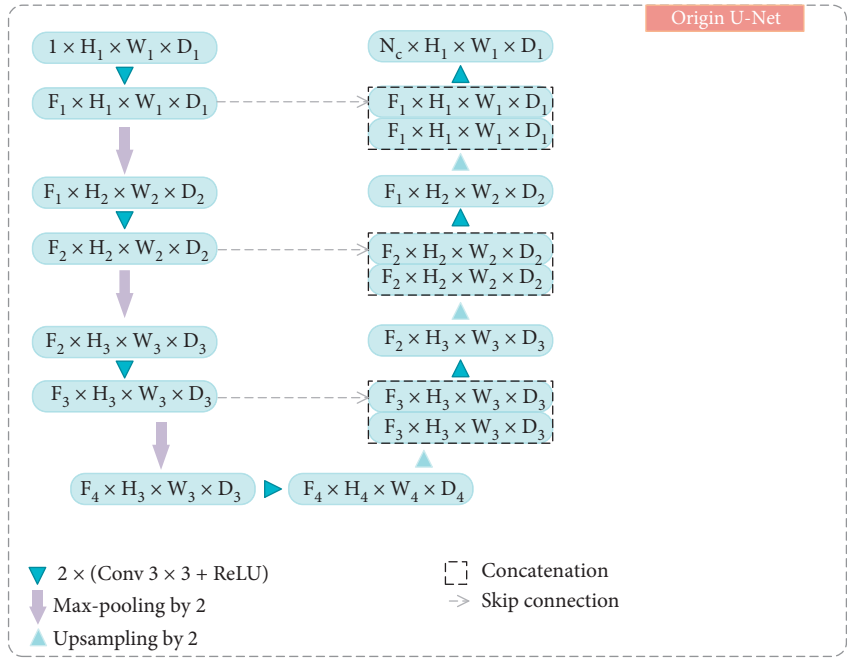


Figure 1: Basic framework of U-Net.

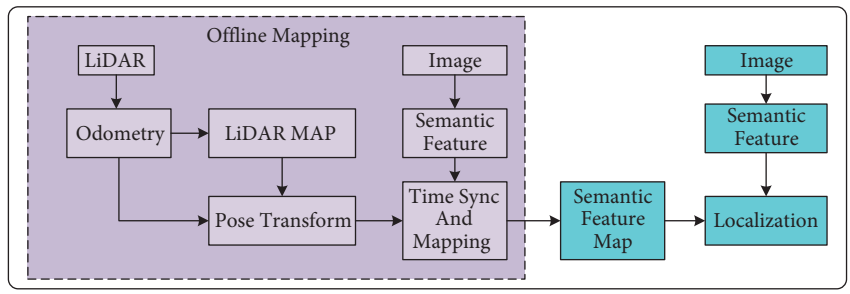


Figure 2: System summary.

Lidar frames. We drive a data collection vehicle across the road in an underground parking lot, select 3707 point cloud keyframes, and then select the corresponding image in the simultaneously recorded image data. The ORB features are extracted from the global map to build the visual semantic dataset and obtain the initial pose for localization. The global map is divided into several zones, and the dataset is established based on the number of character types. Furthermore, the dictionary can be built by zones. The initial pose can be determined with ORB features of the semantic image. Then, the localization can be done with the pose data in the last frame and the obtained real-time semantic data.

3. Semantic Mapping and Localization Methodology

3.1. Image Processing. Four surround-view cameras are used in this project. The position and visual angle of each camera should be adjusted to have a good surrounding picture of the vehicle, as shown in Figure 3. The four purple points on the vehicle are cameras with the fish lens that looks downward. The dashed line is the field of view for each camera. Figure 3 shows that four overlapping areas exist between every two adjacent cameras. Thus, the camera should be calibrated offline, and the weight of each camera should be appropriately set to integrate four separate pictures into one picture. The well-calibrated result can be seen in Figure 4(a). The synthesized picture taken by the camera during driving in the underground parking lot is shown in Figure 4(b). The results shown in Figure 4 indicate that the cameras are well-calibrated to provide enough visual information that can be used to localize the vehicle.

3.2. Image Semantic Segmentation. After theoretical exploration and practical verification, this study adopts U-Net with attention mechanism [29] to perform semantic segmentation tasks, which can make the network sensitive to the characteristics of specific locations. Data enhancement method is also used in increasing the training samples to overcome the disadvantages caused by the size limitation of the dataset. Attention coefficient $\alpha \in [0, 1]$ preserves the activation for specific tasks by identifying remarkable image regions and simplifying feature responses. The output of the

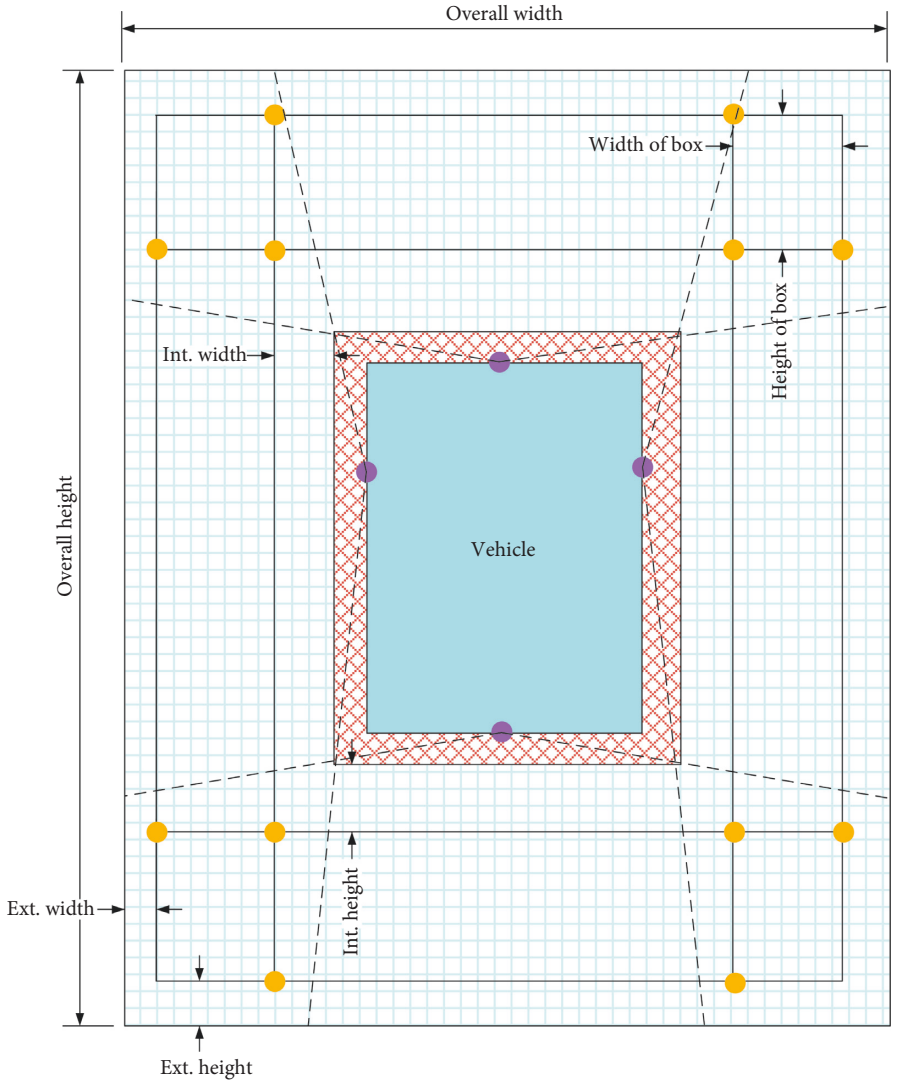


Figure 3: Configuration of four surround-view cameras used in the AVP-SLAM system.

(a) (b)

Figure 4: Synthesized picture. (a) Well-calibrated synthesized pictures. (b) Synthesized picture during driving.

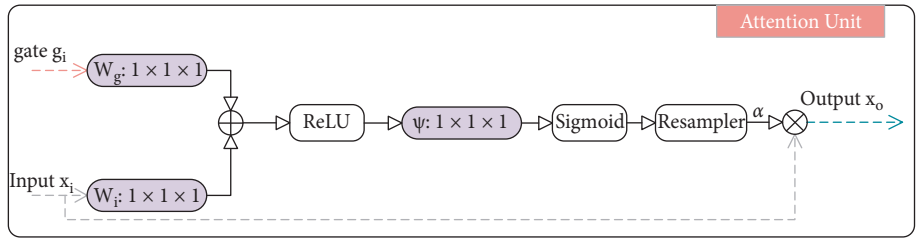


Figure 5: Schematic diagram of attention unit.

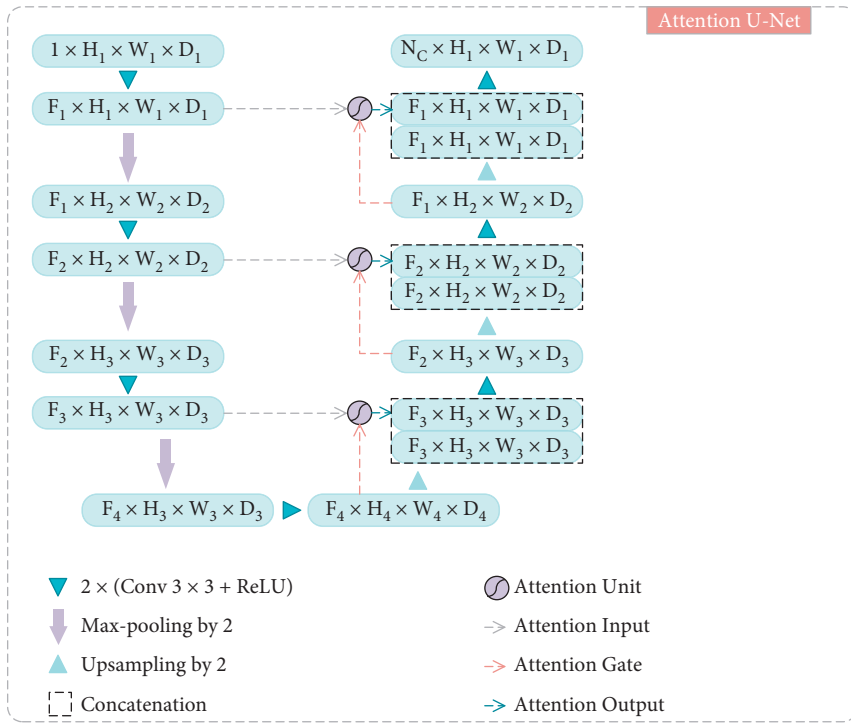


Figure 6: Structure of ATT-U-Net.

unit is the multiplication of the characteristic graph input and the attention coefficient. Each input pixel matrix is $x_i \in \mathbb{R}^{F_L}$, which has a corresponding single-scale feature, and F_L presents the number of feature maps at layer L. Feature-map x_L is obtained at the output of layer L by sequentially applying a linear transformation. For multiple segmentation classes, multidimensional attention coefficients can be used to learn a classified subset of objects in each dimension. A gate vector $g_i \in \mathbb{R}^{F_g}$ can be used to determine the high-attention area by acting on each pixel I. Gate coefficient can be obtained using the additional attention, which can be expressed as follows [27, 28]:

$$\begin{aligned} q_{\text{att}}^{l} &= \psi^T\left(\sigma_1\left(W_x^T x_i^L + W_g^T g_i + b_g\right)\right) + b_\psi, \\ \alpha_i^L &= \sigma_2\left(q_{\text{att}}^L\left(x_i^L, g_i; \Theta_{\text{att}}\right)\right). \end{aligned} \tag{1}$$

$\sigma_2(x_i, c) = (1/1 + e^{-x_{i,c}})$ represents sigmoid activation function. The attention unit is defined by the parameter set Θ_{att}, including linear transformation $W_x \in \mathbb{R}^{F_L \times F_{int}}$, $W_g \in \mathbb{R}^{F_g \times F_{int}}$, $\psi \in \mathbb{R}\hat{}(F_{\text{int}} \times 1)$, and offset $b_\psi \in \mathbb{R}$, $b_g \in \mathbb{R}^{F_{\text{int}}}$. The structure of the attention unit is shown in Figure 5. The structure of the attention unit added U-Net (ATT U-Net) is shown in Figure 6. The convolution parameter updating the rules of L-1 layer is as follows. The function $f(x^L; \Phi^L) = x^{L+1}$ applied in convolution layer L is characterised by trainable kernel parameters Φ^L.

$$\begin{aligned} \frac{\partial\left(x_i^L\right)}{\partial\left(\Phi^{L-1}\right)} &= \frac{\partial\left(\alpha_i^L f\left(x_i^{L-1}; \Phi^{L-1}\right)\right)}{\partial\left(\Phi^{L-1}\right)} \\ &= \alpha_i^L \frac{\partial\left(f\left(x_i^{L-1}; \Phi^{L-1}\right)\right)}{\partial\left(\Phi^{L-1}\right)} + \frac{\partial\left(\alpha_i^L\right)}{\partial\left(\Phi^{L-1}\right)} x_i^L. \end{aligned} \tag{2}$$

The first gradient term α_i^L on the right is scaled. In the multidimensional attention unit, α_i^L corresponds to a vector containing each grid scale. In each subattention unit, supplementary information is extracted and fused to define the output of the residual connection. In order to reduce the number of training parameters and the computational complexity of attention units, the linear transformation $(1 \times 1 \times 1)$ without any spatial support is implemented, and

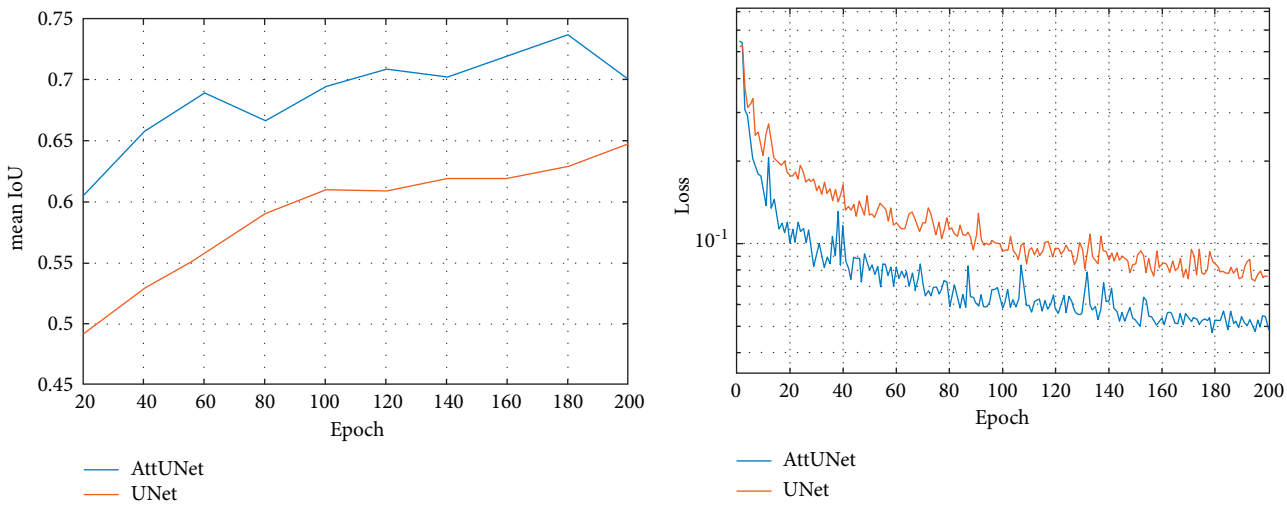


Figure 7: Mean IoU and loss curve during training with 200 epoch.

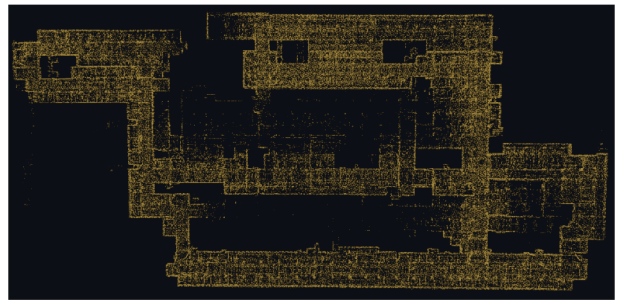
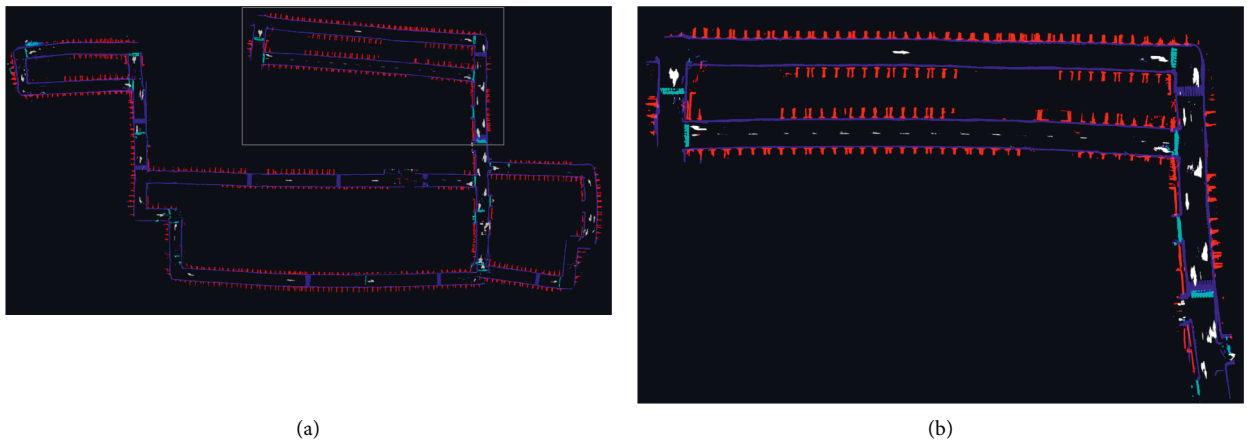


Figure 8: Point cloud map by Lidar.



(a) (b)

Figure 9: Optimized semantic map. (a) Global view. (b) Zoom-in view.

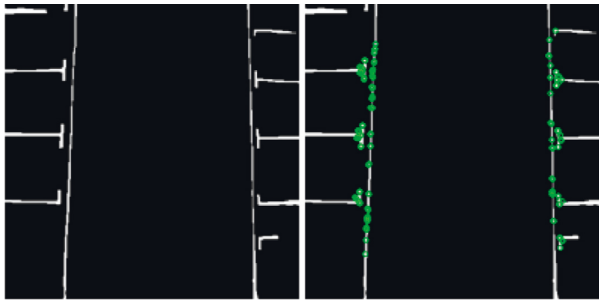
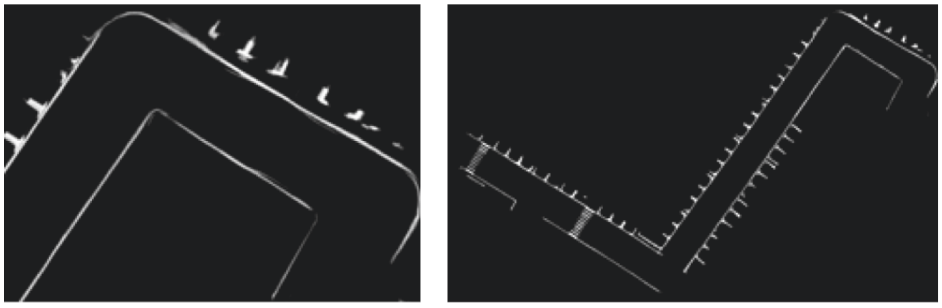


Figure 10: ORB feature detection result.



(a) (b)

Figure 11: (a) Area map with the highest ORB matching scores. (b) Local map made by ICP method.

the input characteristic map is downsampled to the same resolution as the gated signal. The relevant linear transformation couples the feature diagrams and rearranges them to the low dimensional space to implement the gating operation. Low-dimensional feature maps, such as the first residual connection, do not perform a gating function, because they cannot represent input data in high-dimensional space. We use depth supervision to force the medium feature map to be semantically recognizable at each image scale, which ensures that the attention unit has the ability to affect the response of a wide range of image foreground content at different scales.

A performance comparison between ATT U-Net and the original U-Net is shown in Figure 7. As shown in Figure 7, the decrease of loss and the IoU performance when ATT-U-Net is used are faster and better than those when traditional U-Net is used. Cross-entropy loss was used here.

3.3. Lidar Supplemental Mapping. In this study, a 16-channel Lidar is used to build the segmentation map supplementarily. The SC-LeGO-LOAM framework [30–32] is used to assist in building the map. ROS/C++ is selected as the code framework. An image-based segmentation method [33] divides the distance map made by Lidar into multiple groups of clusters, and classes with less than 30 points are discarded as environmental noise to improve the efficiency of processing and the accuracy of feature extraction. The mark (ground point or segmentation point), coordinates in the distance graph, and the distance to the sensor for each point can be obtained by segmentation. These characters of the ground and segmentation points are used for character extraction. In the loop detection of Lidar-based SLAM, the scan context descriptor encodes the radar point cloud and scores the similarity of loop detection. The established Lidar point cloud map after a series of optimization is shown in Figure 8, providing pose information for the construction of the semantic map.

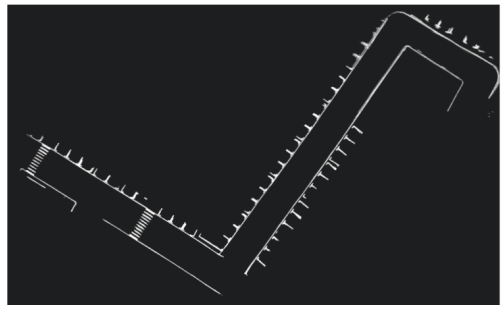


Figure 12: Global view of the matched area map within the local map.

3.4. Semantic Mapping. After obtaining the point cloud map, we can obtain the pose information of every frame with high accuracy. The semantic map and Lidar map with the same timestamp are combined based on the pose information of every Lidar frame. After being matched, the pose transformation between keyframes is used to accumulate the semantic map and obtain pose transformation frames. The established semantic maps obtain the overlapped parts because of the pose data error for point cloud and matching error. The iterative closest point (ICP) algorithm is used in

Table 1: The IoU result.

Lane line	Parking line	Speed bump	Traffic signs	Average IoU
0.78	0.68	0.84	0.78	0.77

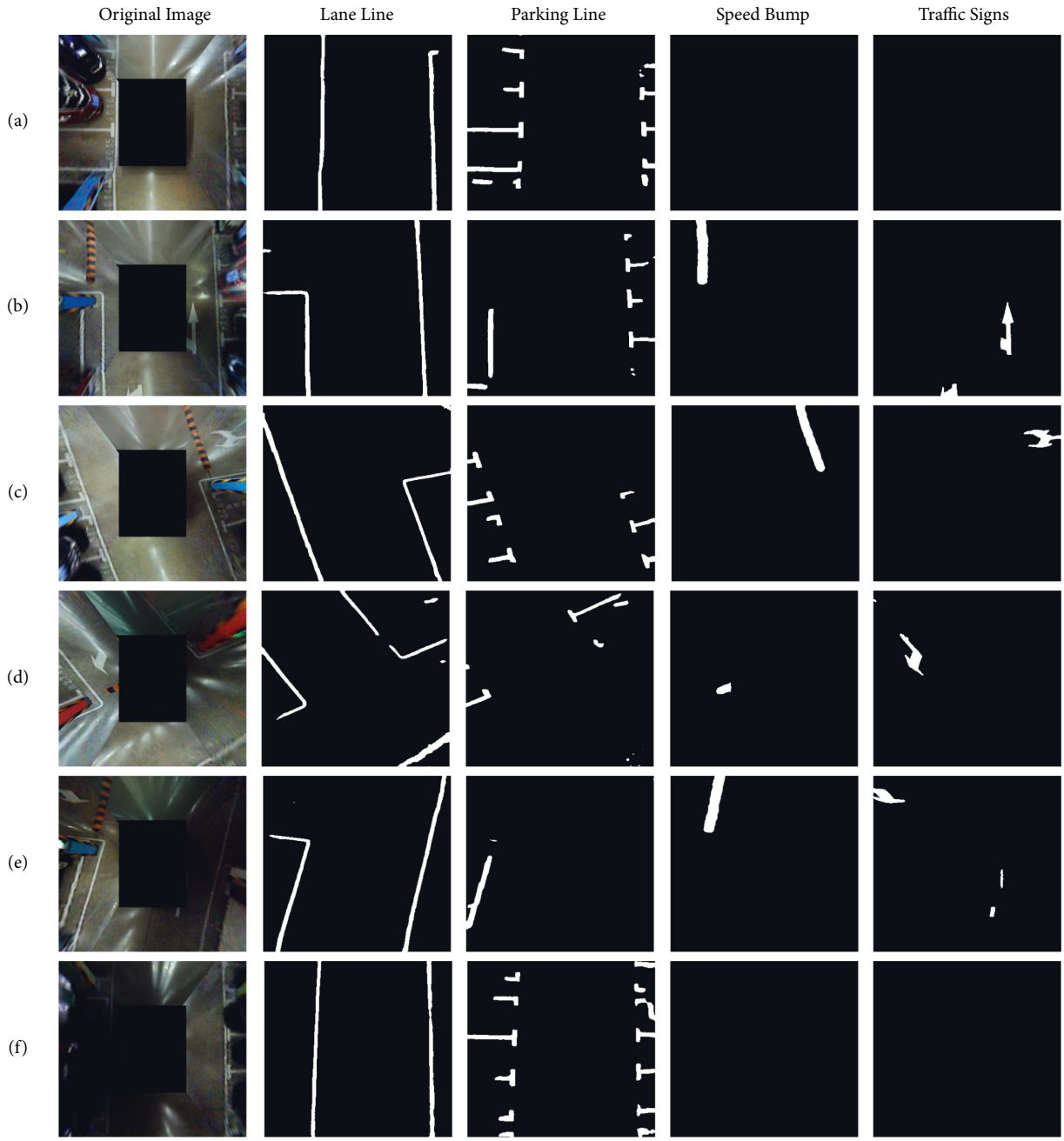


Figure 13: Semantic segmentation result.

calculating the best fusion transformation for the overlapped parts between two maps to solve the problem mentioned. The optimized global map is shown in Figure 9.

3.5. Semantic Localization. The ORB features are extracted from the global map and used to make a visual semantic word bag to determine the initial position and posture. The global map is divided into regions. The word bags and dictionaries are established based on the number of the feature points and regions, respectively. Then, the ORB features are extracted from the initial semantic image input, and a word bag is built. This word bag is used to score the similarity in the dictionary, and the area where the vehicle is located is determined by the score. The ORB feature detection is shown in Figure 10. After the vehicle localization

area is obtained using ORB feature matching, the ICP method is used to overlay several continuous semantic maps into a new one. The local map localization process is shown in Figures 11 and 12. The purple marks are the matched pixel points, whereas the green points and the white points in Figure 12 are the pixel points for the local map and the area map, respectively. After the location of the local map is obtained, the semantic map of the current frame is used to match with the local map and obtain the vehicle posture and location. The new local map from the global map can be chosen to be used for the matching and localization in the next step with the help of the current vehicle posture.

4. Results and Discussion

Several experiments were performed to validate the proposed AVP-SLAM system. All the presented data were taken from the vehicle platform. Four cameras mounted at the front, rear, left, and right sides of the vehicle with fish lens were used in this SLAM system. Furthermore, a 16-channel Lidar was used to help build the map. A Neousys computer with 32G RAM size and 11 G video memory was used for good system efficiency. The front image was taken by the front camera at 30 Hz with a resolution of 1920×1080 pixels. The images taken by the rear, left, and right cameras were recorded at 20 Hz with a resolution of 640×480 pixels. After image stitching optimization and synchronization, an image with a resolution of 1090×860 pixels was output by the system at 18 Hz.

4.1. Real Semantic Segmentation Experiment Result. The experiments were performed with several harsh external conditions to test the robustness performance of the proposed AVP-SLAM algorithm. There is no open-source dataset of the underground parking lot with semantic segmentation for bird's eye view. We have made an AVP dataset, and it is a single underground parking lot dataset for bird's eye view. Please refer to the link of supplementary materials for the data set. The performance of network ATT U-Net in the 520th epoch verification set is the highest, and the loss value is 0.38. The loss value decreases in the subsequent training process, but its performance in the verification set decreases, and the network is overfitted. Finally, the ATT U-Net parameters of the 520th epoch are used for the subsequent segmentation process. The IoU result is shown in Table 1. The final semantic segmentation result with ATT U-Net is shown in Figure 13. Although the parking line is blocked by the car, the parking line can still be clearly seen in Figure 13(a). Figure 13(b) has various marks in the same figure, but they are accurately identified. Although the reflected light overlaps with the white traffic signs, the traffic signs are segmented correctly, as shown in Figure 13(c). This case is the same as the speed bumps and traffic signs, which are shielded by other parked automobiles, as shown in Figures 13(d) and 13(e). Based on the result shown in Figure 13(f), the semantic segmentation result was not affected by the relatively dim light in the parking lot. The result of the U-Net segmentation mechanism on the underground parking lot is very good. In our system, the semantic is visualized and output, and the type of map is also a pixel map. The output of the system is 10fps-12fps, which can meet the real-time positioning requirements of low-speed vehicles. In summary, every specific feature was segmented precisely under different environments.

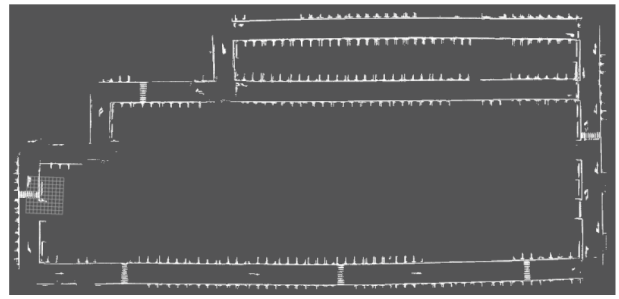


Figure 14: Global optimized semantic map.

4.2. Mapping and Localization. The experiment was performed in a dim underground parking lot. We used an additional Lidar in this system because the semantic map precision was easily affected by the initial values because of the relatively large error during matching between frames and the semantic map. Thus, we adopted Lidar, which is used for building maps. The SC-LeGO-LOAM framework was used to build the map, and the ROS/C++ was used as the code framework. Furthermore, the loop detection with scene context algorithm was used to optimize the mapping precision. The semantic map could be built with the posture data in the established Lidar map by matching the corresponding semantic image. Finally, the global optimized semantic map is shown in Figure 14.

Localization precision is more important than mapping precision because the automobile can localize and drive itself to the correct destination position even with an imprecise semantic map. In our experiment, the localization experiment was performed with the previously established and optimized semantic map. The initial position of the automobile is regarded as known data. These data are usually saved in NVM when the vehicle was parked the last time.

The final localization result is shown in Figure 15. The red line is used to show the motion trail for the experimental vehicle. During the experiment, the vehicle could constantly localize itself from start to end. The detailed localization result can be found in the video material.

4.3. Real Application: Autonomous Valet Parking. The proposed AVP-SLAM system was used under real autonomous valet parking cases in the underground parking lot. The preestablished semantic map was used by the vehicle to localize itself in this parking lot and guide itself to the prechosen parking lot automatically, as shown in Figure 16. Additional detailed experiment results can also be found in the shared video materials. In conclusion, a good SLAM result can be provided with the proposed AVP-SLAM system.

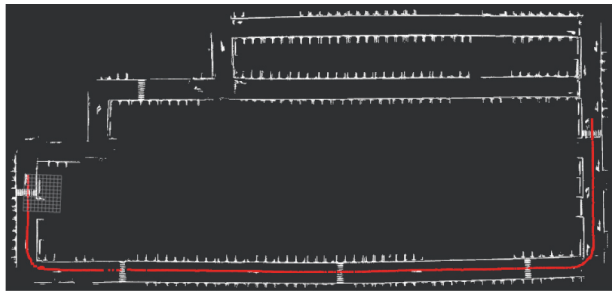


Figure 15: Experimental localization result in the parking lot.

Figure 16: Test vehicle.

5. Conclusions

In this study, a camera-Lidar combined with the SLAM solution was proposed. In this scheme, a 16-channel Lidar was used in assisting the visual system, that is, four surrounding-view cameras with fish lens, to build the map. Moreover, the semantic features, lane lines, parking lines, speed bumps, traffic signs, and other visual features could be detected using ATT U-Net even under harsh situations. Thus, a complete semantic map was built based on the detected features. With the preobtained map, the vehicle could localize itself during driving.

Furthermore, a real AVP experiment was performed to validate the proposed SLAM solution. The result showed that the vehicle can park itself in a correct parking lot autonomously in a dim underground parking lot. Thus far, the proposed SLAM solution is only effective with the AVP scenario. We will continue to develop this solution in the application field in a much more difficult environment.

Data Availability

The data are available through https://pan.baidu.com/s/1ioeXqYIlocYpsQb0KB-_Ng. The extraction code is lxq8.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

C. J. Liu completed the principle derivation, scheme design, scheme implementation, experimental verification, and article writing. J. J. Yao completed the system design and experimental scheme design.

Supplementary Materials

The data, video materials, and code are available through https://pan.baidu.com/s/1ioeXqYIlocYpsQb0KB-_Ng. The extraction code is lxq8. (*Supplementary Materials*)

References

[1] R. Smith, "On the representation of spatial uncertainty," *Int.j.robotics Res*, 1986.

[2] Y. Zhou, "A survey of VSLAM," *CAAI Transactions on Intelligent Systems*, 2018.

[3] J. Yue, *Lidar Data Enrichment Using Deep Learning Based on High-Resolution Image: An Approach to Achieve High-Performance Lidar-Based SLAM Using Low-Cost Lidar*, 2020.

[4] G. Chen, H. Cao, J. Conradt, H. Tang, F. Rohrbein, and A. Knoll, "Event-based neuromorphic vision for autonomous driving: a paradigm shift for bio-inspired visual sensing and perception," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 34–49, 2020.

[5] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[6] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2014.

[7] T. Qin, P. Li, and S. Shen, "Vins-mono: a robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," in *Proceedings of the 9th European conference on Computer Vision*Springer-erlag, Graz Austria, May 2006.

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: binary robust independent elementary features," computer vision - ECCV 2010," in *Proceedings of the 11th European Conference on Computer Vision*, Heraklion, Crete, Greece, September 2010.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, Barcelona, Spain, Novemer 2011.

[12] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[13] R. Mur-Ar tal and J. D. T Ar Dos, "ORB-SLAM2: an open-source SLAM system for monocular," *Stereo and RGB-D Cameras*, 2016.

[14] L. Yan, "Monocular localization in urban environments using road markings," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, Redondo beach, CA, USA, June 2017.

[15] M. Schreiber, C. Knoppel, and U. Franke, "LaneLoc: lane marking based localization using highly accurate maps," in *Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV)*, Gold Coast City, Australia, June 2013.

[16] A. Ranganathan, D. Ilstrup, and W. Tao, "Light-weight localization for vehicles using road markings," *Intelligent Robots and Systems (IROS)*, 2013.

[17] E. Rehder and A. Albrecht, "Submap-based SLAM for road markings," in *Proceedings of the Intelligent Vehicles Symposium*, Seoul, South Korea, 2015.

[18] J. Jeong, Y. Cho, and A. Kim, *Road-SLAM: Road Marking Based SLAM with Lane-Level Accuracy*, pp. 1736–1473, IEEE, New Jersy, NJ, USA, 2017.

[19] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," robotics and automation," in *Proceedings of the 2007 IEEE International Conference on IEEE*, Washington, DC, USA, August 2007.

[20] W. Weng and X. Zhu, "INet: convolutional networks for biomedical image segmentation," *IEEE Access*, 991, p.2021.

[21] M. Z. Alom, *Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation*, 2018.

[22] Y. Ye, *Universal Semantic Segmentation for Fisheye Urban Driving Images*, 2020.

[23] H. Zhao, "Pyramid scene parsing network," *IEEE Computer Society*, 2016.

[24] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional Networks for Biomedical Image Segmentation*, Springer International Publishing, New York, NY, USA, 2015.

[25] S. Pasban, "Infant brain segmentation based on a combination of VGG-16 and U-Net deep neural networks," *IET Image Processing*, vol. 14, 2021.

[26] C. Szegedy, "Inception-v4," *Inception-ResNet and the Impact of Residual Connections on Learning*, 2016.

[27] E. Romera, "ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, pp. 1–10, 2017.

[28] Y. Hou and Z. C. T.-W. C. C. Ma, "Inter-region affinity distillation for road marking segmentation," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.

[29] O. Oktay, *Attention U-Net: Learning where to Look for the Pancreas*, 2018.

[30] Z. Ji and S. Singh, "LOAM: lidar odometry and mapping in real-time," in *Proceedings of the Robotics: Science and Systems Conference*, Berkeley, CA, USA, July 2014.

[31] T. Shan and B. Englot, "LeGO-LOAM: lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (IROS) IEEE, Madrid, Spain, October 2018.

[32] G. Kim and A. Kim, "Scan context: egocentric spatial descriptor for place recognition within 3D point cloud map," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (IROS) IEEE, Madrid, Spain, October 2018.

[33] I. Bogoslavskyi and C. Stachniss, "Fast range image-based segmentation of sparse 3D laser scans for online operation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems IEEE*, pp. 163–169, Daejeon, Korea, October 2016.