Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070223
ISSN 2371-8412 Vol. 7 Num. 2

Research on Path Planning Algorithm Based on Fast Target Detection

Wei Zhang, Zhigao Cui, Nian Wang, Yunwei Lan

Xi'an Research Institute of High-Tech, Xi'an, 710025, China

Keywords: Mobile robot; Object detection; Two-dimensional grid map; Semantic navigation mapping

Abstract: As a key technology of robot navigation, path planning has garnered widespread attention and has been utilized in various applications such as mobile robots, unmanned aerial vehicles, and human-computer interaction. Recently, several studies advocate constructing semantic maps for path planning in the laboratory stage. However, these approaches require large storage space and high computing resource consumption, making it difficult to meet real-time requirements. We tackle this issue by building real-time semantic navigation map and propose a real-time path planning algorithm based on fast target detection. Specially, we first construct two-dimensional grid map using the Gamapping method and locate the target objection utilizing the object detection algorithm YOLOv3 retrained in an indoor experimental environment. Furthermore, by incorporating the category information and position information of the detected object into the two-dimensional grid map through a coordinate mapping mechanism, we combine the geometric metric information and visual detection information to build semantic navigation map for automatically planning a reasonable path. The experiments conducted on both qualitative and quantitative levels have demonstrated that our method achieves superior performance and practical application value.

1. Introduction

In recent years, Simultaneous Localization and Mapping (SLAM) technology has seen rapid pprogress, and has been successfully implemented in various fields such as mobile robots, unmanned aerial vehicles (UAVs) and driverless car. However, although traditional SLAM technology can effectively identify the geometric structure of the surrounding environment, it often ignores the understanding of semantic information. Due to the widespride application of deep learning, deep neural networks have made significant progress in object detection, semantic segmentation and other fields. As a result, some researchers are attempting to apply visual object detection algorithms and image semantic segmentation algorithms based on deep learning technology to SLAM research, thus ushering in a new era of semantic SLAM research.

For a semantic SLAM task, the construction of a semantic map is considered to be the most important and challenging step. Kuipers et al. [1] first proposed the concept of semantic map, emphasizin the importance of modeling external spatial knowledge, but their work was not yet practically implemented at that time. To address this issue, several recent studies [2-5] construct semantic maps using monocular visual SLAM, two-dimensional laser sensors, global conditional random fields, and semantic acquisition framework. Although these methods have made some progress, the construction of semantic maps is still in the laboratory stage and experiments are conducted in simulation environment. In contrast, there have been relatively few studies on constructing semantic maps in real-world environments. Furthermore, while visual SLAM technology can fully characterize the complex three-dimensional environment, the 3D map requires a significant amount of storage space and has poor robustness., Additionally, it has high hardware requirementsand cannot meet the real-time requirements. Therefore, under the existing technology and hardware conditions,

Therefore, we focus on the development of a method for constructing a real-time two-dimensional semantic map using two-dimensional laser sensors and visual sensors. To meet the practical requirements of autonomous navigation and human-computer interaction of mobile robots, we propose a two-dimensional semantic navigation map construction method based on fast object detection for real-time path planning. This method is designed. Firstly, based on the self-built mobile robot platform, the 2D raster map is drawn by Gamapping algorithm. Then, the retrained object detection algorithm YOLOv3 is eemployed to detect the target in real time. The category and position information of the target object is mapped to two-dimensional raster map by coordinate mapping. Finally, the interior 2D semantic map is constructed by combining geometric measurement information with visual detection information.

2. Proposed Method

The workflow of semantic navigation map construction based on fast object detection is shown in Figure 1. Firstly, based on two-dimensional laser sensor and odometer data, we use Gmapping algorithm to achieve incremental accurate raster mapping. Then, color images and depth images are collected in real time based on RGB-D cameras. The YOLOv3 algorithm is retrained for common indoor objects to realize real-time target detection of the images. Finally, the location information and category information of the detected object are transformed to the two-dimensional raster map to complete the construction of semantic map. Additionally, the proposed method can be divided into four parts: wheel odometer motion model, 2D laser sensor model, 2D raster map construction, object detection and coordinate mapping.

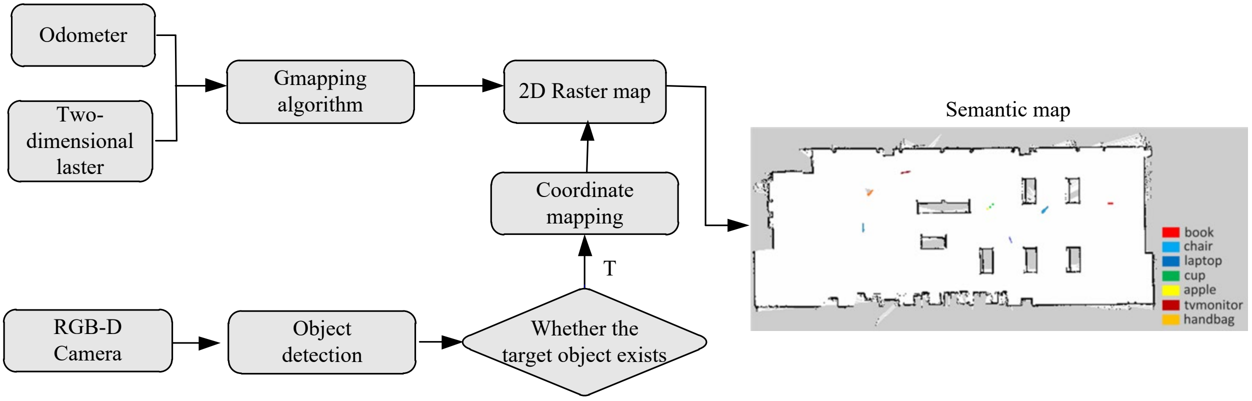

Figure 1: Overall workflow of the proposed method

2.1 Wheel odometer motion model

The wheel odometer relies on the pulse variation output by photoelectric encoder in a certaintime to calculate the moving distance and rotation angle of the wheel, so as to estimate the relative variation of the mobile robot's pose. Suppose the resolution of photoelectric encoder is P, the deceleration ratio of reducer is η, the diameter of robot wheel is D, and the axle length of vehicle is W, then the displacement increment of the wheel within a unit time can be calculated as

$$\begin{cases} \delta = \dfrac{\pi D}{\eta P} \\ Y_{odom} = \delta N, \end{cases} \tag{1}$$

where δ is the distance that wheel rotates within a unit time, and N is the number of pulses output by the wheel within a unit time. Assume that the increment of the left and right encoders within a sampling interval is $\Delta \mathrm{n}$, $\Delta \mathrm{m}$ respectively, the displacement variation of left and right wheels can be obtained as

$$\begin{cases} \Delta Y_{odom\_left} = \delta \Delta n \\ \Delta Y_{odom\_right} = \delta \Delta m. \end{cases} \tag{2}$$

Within a sampling interval, the displacement increment of mobile robot can be expressed as the average displacement increment of the left and right wheels, and its angle increment can be further calculated as

$$\begin{cases} \Delta D = (\Delta Y_{odom\_left} + \Delta Y_{odom\_right}) / 2 \\ \Delta \theta = (\Delta Y_{odom\_right} - \Delta Y_{odom\_left}) / W, \end{cases} \tag{3}$$

where ΔD and $\Delta \theta$ respectively represent the displacement increment and angle increment of robot within a sampling interval. Let the pose of robot at the t-th time be $[x_t, y_t, \theta_t]^{\mathrm{T}}$, the pose of robot at $(t+1)$-th time can be calculated as

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} x_t + \Delta D \cdot \cos(\theta_t + \Delta \theta) \\ y_t + \Delta D \cdot \sin(\theta_t + \Delta \theta) \\ \theta_t + \Delta \theta \end{bmatrix} \tag{4}$$

2.2 Two-dimensional laser sensor model

Two-dimensional laser sensors work by transmitting a beam of light and receiving a beam reflected back by an obstacle, and use the time of beam travels to calculate the distance of the obstacle. In this paper, the used dimensional laser sensor has two main functions: Firstly, the real-time observation information with the environment map is matched and combined with the odometer data to complete the robot pose estimation; Secondly, after the precise positioning of the robot, the map is incrementally constructed from the two-dimensional laser sensor data at the current moment.

In order to realize the real-time positioning of robot, it is usually necessary to convert the laser coordinates to the world coordinates, that is, to establish a two-dimensional laser sensor observation model. The distance between the measuring point and the laser transmitting point z_t, and the angle ε_t between the measuring point and the horizontal coordinate of the two-dimensional laser sensor can be obtained by using the two-dimensional laser sensor. As shown in Figure 2, if the posture of robot at the moment is $[x_t, y_t, \theta_t]^{\mathrm{T}}$, then we obtain their world coordinate as

$$\begin{bmatrix} x_{z_t} \\ y_{z_t} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} + z_t \begin{bmatrix} \cos(\theta_t + \varepsilon_t) \\ \sin(\theta_t + \varepsilon_t) \end{bmatrix} + \begin{bmatrix} \xi_x \\ \xi_y \end{bmatrix} \tag{5}$$

where $[x_{z_t}, y_{z_t}]^{\mathrm{T}}$ represents the coordinates of the laser measuring point in the world coordinate, $[\xi_x, \xi_y]^{\mathrm{T}}$ represents the measurement noise, and generally follows the Gaussian distribution of zero mean.

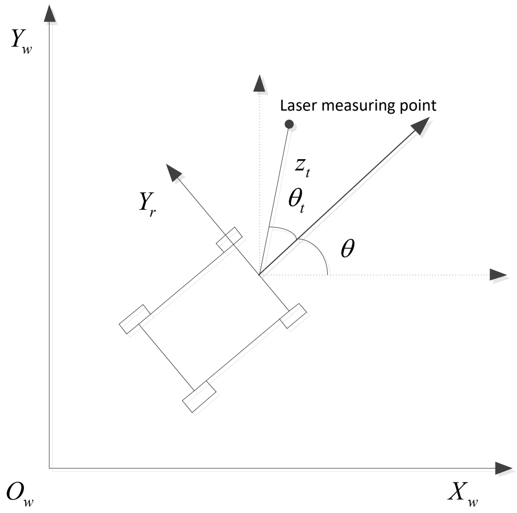

Figure 2: Schematic diagram of coordinate transformation that measures points of two-dimensional laser sensor.

2.3 Two-dimensional raster map construction

In this paper, the two-dimensional laser map is constructed using Gmapping algorithm. The core of Gmapping algorithm is particle filter algorithm, which can be divided into two parts: robot pose estimation by particle filter and global map update by Kalman filter. Gmapping algorithm takes odometer information as input and data of two-dimensional laser sensor as observation values. Its combined posterior probability density distribution is shown as

$$p\left(x_{1:t}, m \middle| z_{1:t}, u_{1:t-1}\right) = p\left(m \middle| x_{1:t}, z_{1:t}\right) \bullet p\left(x_{1:t} \middle| z_{1:t}, u_{1:t-1}\right) \tag{6}$$

where $x_{1:t} = \left[x_1, x_2, \cdots, x_t\right]$ represents the robot pose sequence within the moment t, $m = \left[x_1, y_1, x_2, y_2, \cdots x_t, y_t\right]$ represents the location of environmental features, $z_{1:t} = \left[z_1, z_2, \cdots, z_t\right]$ represents the observation sequence of two-dimensional laser sensor within the moment t, and $u_{1:t-1} = \left[u_1, u_2, \cdots, u_{t-1}\right]$ represents the odometer measurement value within the moment $t-1$.

The method based on particle filtering randomly distributes several particles in the map, and each particle represents a possible motion trajectory [6]. In the process of robot driving, the particle set is iteratively converging, in which the particles with high weight are retained, while the particles with low weight are abandoned, so as to ensure that the optimal particle set consistent with the robot trajectory is retained eventually. Based on particle filtering algorithm, Gmapping algorithm simultaneously takes the odometer motion model and the real-time observation information of two-dimensional laser sensor as the proposed distribution [7], so that the sampling is distributed in the likelihood function region of the observation model to the maximum extent, which increases the probability of obtaining the optimal particle and realizes the construction of a more accurate map. The sampling proposal distribution is shown as

$$p\left(x_t \left| m_{t-1}^{(i)}, x_{t-1}^{(i)}, z_t, u_{t-1}\right.\right) = \frac{p\left(z_t \left| m_{t-1}^{(i)}, x_t\right.\right) p\left(x_t \left| x_{t-1}^{(i)}, u_{t-1}\right.\right)}{p\left(z_t \left| m_{t-1}^{(i)}, x_{t-1}^{(i)}, u_{t-1}\right.\right)} \tag{7}$$

The corresponding probability $p\left(x_{1:t}, m \middle| z_{1:t}, u_{1:t-1}\right)$ can be obtained by combining the position sequence $x_{1:t}$ and the observation information $z_{1:t}$ of two-dimensional laser sensor, so that the local map observed in real time can be continuously fused into global map, and the global map can be updated until the construction of global map is realized.

2.4 Object detection and coordinate mapping

YOLOv3 is one of the best algorithms in the field of target detection at present, which can achieve good results in terms of detection speed and accuracy [8]. Therefore, this paper adopts the YOLOv3 algorithm trained for specific indoor objects to carry out target detection. In addition, the algorithm has good robustness for detecting objects or small objects that are very close to the camera in the image, which is very advantageous for mobile robots traveling at a relatively low speed.

After the completion of the two-dimensional raster map and object detection, the position information and category information of objects detected by the object detection algorithm need to be further mapped to the two-dimensional raster map, that is, the position information of objects in the camera coordinate system is mapped to the two-dimensional raster map coordinate system. In general, the robot coordinate system and the camera coordinate system have the following transformation relationship.

$$\begin{cases} \left[x_r, y_r, z_r\right]^{\mathrm{T}} = R_{cr}\left[x_c, y_c, z_c\right]^{\mathrm{T}} + T_{cr} \\ R_{cr} = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \\ T_{cr} = \left[\Delta x, \Delta y, \Delta z\right]^{\mathrm{T}}, \end{cases} \tag{8}$$

Where $\left[x_r, y_r, z_r\right], \left[x_c, y_c, z_c\right]$ are the coordinates of three-dimensional space points in the robot coordinate system and the camera coordinate system respectively. R_{cr}, T_{cr} are the rotation matrix and translation vector between the robot coordinate system and the camera coordinate system respectively. α are the tilt Angle of the camera. $\left[\Delta x, \Delta y, \Delta z\right]$ are the projected distance between the origin of the two coordinate systems on the three coordinate axes. They all depend on the installation positions of the camera and the laser sensor.

Similarly, the robot coordinate system has the following transformation relation with the world coordinate system.

$$\begin{cases} \left[x_w, y_w, z_w\right]^{\mathrm{T}} = R_{wr}\left[x_r, y_r, z_r\right]^{\mathrm{T}} + T_{wr} \\ R_{wr} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ T_{wr} = \left[\Delta x', \Delta y', 0\right]^{\mathrm{T}}, \end{cases} \tag{9}$$

where $\left[x_w, y_w, z_w\right], \left[x_r, y_r, z_r\right]$ are the coordinates of three-dimensional space points in the world coordinate system and the robot coordinate system respectively; R_{wr}, T_{wr} are the rotation matrix and translation vector between the world coordinate system and the robot coordinate system respectively; β is the angles between the two coordinate systems; $\left[\Delta x', \Delta y', 0\right]^{\mathrm{T}}$ are the corresponding coordinate axis distance between the origin of the two coordinate systems.

According to the resolution of the constructed raster map, the coordinates under the world coordinate system can be converted into discrete raster coordinates as

$$\begin{bmatrix} x_g \\ y_g \end{bmatrix} = \mathrm{int}\left(\begin{bmatrix} x_w \\ y_w \end{bmatrix} / \mathit{resolution}\right), \tag{10}$$

Where $\left[x_g, y_g\right]$ and $\left[x_w, y_w\right]$ respectively represent raster coordinates and world coordinates, and *resolution* represents the resolution of the constructed raster map.

In order to highlight the objects contained in the map, the generated semantic navigation map containing object annotation information is further optimized. For a two-dimensional raster map, the obstacles where the marked points are detected are marked uniformly to form a closed pattern, and different kinds of objects are represented by different colors, thus completing the construction of a semantic navigation map.

3. Experimental results and analysis

In this paper, the computer used for deep neural network training and testing of target detection is cpu i9 processor, main frequency 3.7GHz, memory 64GB, NVIDIA GTX 1080Ti graphics card, and Ubuntu 14.04 system.

In order to avoid interference caused by detection of unnecessary objects, 20 types of common objects in indoor scenes were selected based on COCO data set and PASCAL VOC data set [9], and a total of 6000 images were collected for training. Before training, pre-selected training sets are uniformly made into VOC format xml files and converted into YOLO format txt files. In the training, the original image was randomly clipping, mirroring, translation, stretching, rotation and other data augmentation processing.

Figure 3 shows the effect diagram of target detection in the actual experimental environment by using the YOLOv3 network. In order to quantitatively highlight the advantages of this algorithm in target detection accuracy and detection speed, it was quantitatively compared with YOLOv2 and Faster RCNN algorithms, and the experimental results were shown in Table 1.

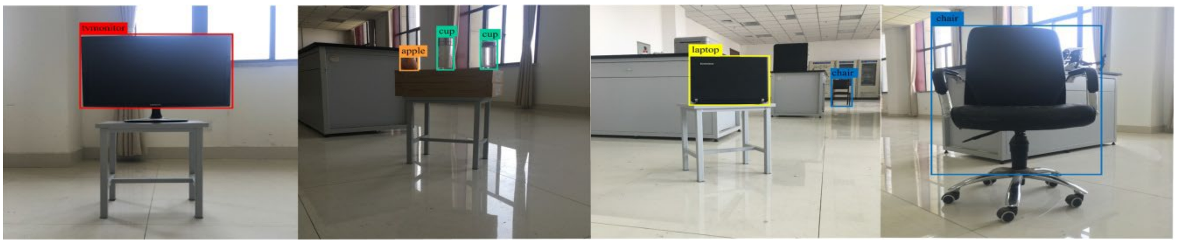

Figure 3: Schematic diagram of YOLOv3 algorithm target detection results

Table 1: Comparison of indoor target detection accuracy

Category	YOLOv3	YOLOv2	Faster RCNN
chair	73.3	52.6	54.4
bed	74.5	53.8	59.2
door	77.4	70.2	73.1
sofa	69.2	61.6	63.6
cup	86.8	71.3	85.5
handbag	79.3	72.2	76.5
person	92.9	85.6	89.4
cat	93.3	86.5	88.4
dog	94.2	88.2	85.9
apple	69.8	62.3	57.3
banana	65.6	52.3	52.2
book	72.5	60.3	62.1
laptop	83.4	74.4	80.1
tvmonitor	77.5	62.3	60.7
bicycle	92.3	84.7	80.5
refrigerator	73.1	63.0	63.2
suitcase	85.7	79.8	82.2
backpack	73.9	66.3	66.3
sports ball	69.6	54.2	60.5
vase	70.2	60.5	69.8
AVG	78.73	68.11	70.55

As can be seen from Table 1, the detection accuracy of YOLOv3 is higher, which is 15.60% and 11.59% higher than that of YOLOv2 and Faster RCNN respectively. In addition, in terms of detection speed, YOLOv3 is also very fast, which can fully meet the requirements of real-time semantic map construction in this paper. This paper conducted experiments on a desktop computer equipped with GTX 1080Ti, and the detection speed is shown in Table 2. mAP (mean Average Precision) indicates the average detection accuracy, and FPS(Frames Per Second) indicates the number of frames detected per second.

Table 2: Comparison of indoor target detection speed

Detecting algorithm	mAP/%	FPS
YOLOv3	78.73	43
YOLOv2	68.11	45
Faster RCNN	70.55	7

In this paper, an autonomous mobile robot is used to construct a semantic navigation map. As shown in Figure 4, the robot integrates laser sensor, wheel odometer, depth camera, monocular camera, IMU measuring element, 16-channel ultrasonic wave and other devices. Among them, the laser sensor is the RPLIDAR A2 LIDAR, which can scan and measure in a 360-degree all-round way, so as to obtain the map information of the plane point cloud in the space where the robot is located.

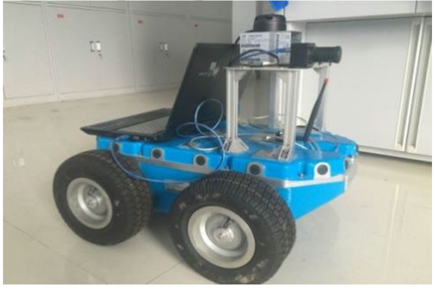

Figure 4: Autonomic mobile robot

In order to verify the effect of the 2D semantic navigation map construction method proposed in this paper, the built mobile robot is used to conduct experiments in the real environment. Figure 5 shows the two-dimensional semantic navigation map constructed based on the overall scene of the laboratory. As we can see, the black area represents the obstacles in the environment, the white area indicates that there are no obstacles in this area, while the gray area represents the area that the mobile robot has not explored, and the colored marked figure in the figure is the detected target object.The proposed method has a clear and concise expression of the environment. The experimental results prove that the proposed method can well complete the task of semantic map construction, so as to effectively support the robot to complete advanced tasks such as autonomous navigation and human-computer interaction.

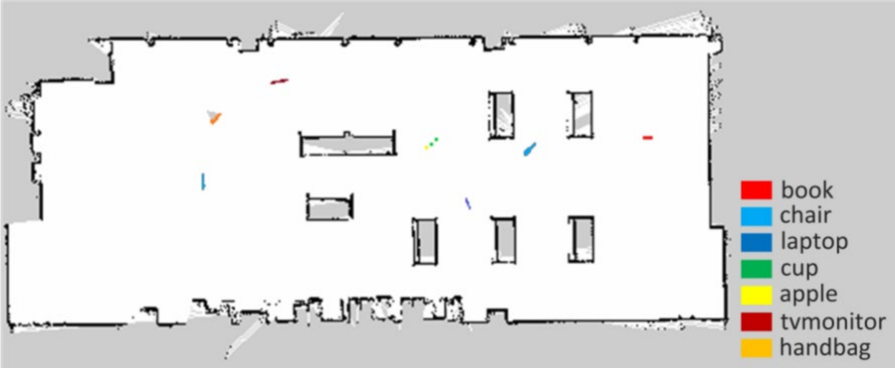

Figure 5: 2D semantic navigation map of laboratory scene

4. Conclusion

In this paper, we propose a real-time semantic navigation map construction method based on fast object detection, and verify the path planning effect by conducting indoor experiments in real scenes. Firstly, based on the laser sensor and odometer data, the Gmapping algorithm is used to construct a two-dimensional raster map, and the expression of the real scene is better realized by parameter optimization. Then, the YOLOv3 algorithm is used to train the indoor object image dataset, and the coordinate mapping mechanism is established to map the category information and location information of the detected target object into a two-dimensional raster map. The experimental results show that the proposed method achieves good results both in detection speed and detection accuracy when recognizing common indoor objects. At the same time, the object category and location information recognized by the object detection algorithm are fused into the grid map to complete the construction of the semantic map, which can effectively generate a reasonable path and assist the robot to complete more advanced tasks such as autonomous navigation.

References

[1] B. Kuipers. Modeling spatial knowledge. Cognitive Science, 1978, 2(2): 129-153.

[2] J. Civera, D. Gálvez-López, L. Riazuelo, et al. Towards semantic SLAM using a monocular camera. International Conference on Intelligent Robots and Systems, 2011: 1277-1284.

[3] Z. Liu, G. V. Wichert. Extracting semantic indoor maps from occupancy grids. Robotics & Autonomous Systems, 2014, 62(5): 663-674.

[4] W. T. Jiang, X. J. Gong, J. L. Liu. Dense semantic map construction of large-scale scene based on incremental computing. Journal of Zhejiang University, 2016, 50(2):385-391.

[5] J. S. Yu, H. Wu, G. H. Tian Cloud-based semantic library design and robot semantic map construction. Robot, 2016, 38(4):410-419.

[6] X. Y. Zhan. Research on Simultaneous localization and indoor map Construction Algorithm based on Lidar. Harbin Insitute of Techmology, 2017

[7] H. Chou, X. L. Ping W. Y. Gao, et al. Research on multi-sensor information fusion algorithm for constructing mobile robot map. Mechanical manufacturing, 2017, 55(8):1-4.

[8] J. Redmon, A. Farhadi. YOLOv3: An Incremental Improvement. 2018, arXiv, 4.

[9] M. Everingham, L. V. Gool, C. K. Williams, et al. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 2010, 88(2): 303-338.