Robotica (2022), **40**, pp. 4011–4030
doi:10.1017/S026357472200073X

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Research on rapid location method of mobile robot based on semantic grid map in large scene similar environment

Hengyang Kuang[1], Yansheng Li[1,*], Yi Zhang[1], Yong Wan[1] and Gengyu Ge[2]

[1]School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China and [2]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
*Corresponding author. E-mail: liyansheng@cqupt.edu.cn

Received: 30 December 2021; **Revised:** 7 April 2022; **Accepted:** 8 May 2022; **First published online:** 8 June 2022

Keywords: mobile robot, grid map, instance segmentation, semantic map, large scene similar environment, global localization

Abstract

Aiming at the problem that adaptive Monte Carlo localization (AMCL) algorithm is difficult to localize in large scenes and similar environments. This paper uses a semantic information-assisted approach to improve the AMCL algorithm. This method realizes the robust localization of the robot in the large scenes and similar environments. Firstly, the 2D grid map created by simultaneous localization and mapping using lidar can obtain highly accurate indoor environmental contour information. Secondly, the semantic object capture is achieved by using a depth camera combined with an instance segmentation algorithm. Then, the semantic grid map is created by mapping the semantic point cloud through the back-projection process of the pinhole camera. Finally, semantic grid maps are used as a priori information to assist in localization, which will be used to improve the initial particle swarm distribution in the global localization of the AMCL algorithm and thus will solve the robot localization problem in this environment. The experimental evidence shows that the semantic grid map solves the environmental information degradation problem caused by 2D lidar as well as improves the robot's perception of the environment. In addition, this paper improves the localization robustness of the AMCL algorithm in large scenes and similar environments, resulting in an average localization success rate of about 90% or even higher, and further reduces the number of iterations. The global localization problem of robots in large scenes and similar environments is effectively solved.

1. Introduction

With the rapid development and wide application of mobile robots, higher requirements are placed on mobile robots, such as diversification of functions, specialization of scenes, and ontology intelligence. Among them, the robot needs to have a high scene generalization ability, which enables the robot to ensure that the robot can work normally in different environments. Therefore, the mobile robot's capture of environmental information and the mobile robot's pose estimation relative to the environment are very important. Conventional navigation and localization of mobile robots include GPS, Wi-Fi, RFID, IMU, and other methods [1–4]. However, for indoor environments, these methods have some limitations. In order to solve the problem of mobile robot indoor environment feature capture and localization with a priori map, many researchers have done a lot of work on simultaneous localization and mapping (SLAM) technology.

For environmental map creation, it is a difficult problem to create a map using noisy and uncertain observations while needing to know the relative poses of the mobile robot itself in the created map. When the mobile robot cannot get the environment map and its own poses at the same time, researchers have proposed the SLAM scheme [5]. The environmental maps constructed by this scheme can be divided into two categories according to the type of sensors. One type is based on laser sensor mapping algorithms, such as Cartographer [6], Hector-SLAM [7], Gmapping [8], and Karto-SLAM [9]. The other type is based on vision sensor, such as ORB-SLAM [10,11], RGB-D-SLAM [12,13], VINS-Mono [14], and DSO [15]. Among them, the Cartographer algorithm is the mainstream solution for obtaining 2D grid maps for robots because it can provide high accuracy and high robustness of grid maps with loop-back detection. The problem of environmental information capture is solved by instant localization and map construction. Chuang Qian et al. [16] proposed to construct a higher quality indoor map using IMU combined with lidar. Based on the already available grid map, the localization problems can be divided into three categories including location tracking, global localization, and kidnapped robot problem according to the difficulty of localization. Position tracking occurs locally in the grid map, and the predicted position often lies near the real position, which is often approximated by Gaussian distribution. For such relatively simple localization problems, Monte Carlo localization (MCL) [17] algorithms have been able to solve. Unfortunately, traditional MCL does not perform very well in position tracking and global localization [18]. When the initial position of the robot is unknown in the grid map, such localization problems evolve into global localization or kidnapped robot problem. In this case, the initial pose may exist at any location in the grid map. In this case, the Gaussian distribution cannot be solved for this type of localization problem; however, the uniform distribution is frequently used to estimate it. Thrun et al. proposed the adaptive MCL (AMCL) algorithm, and Kullback–Leibler divergence (KLD) sampling makes this localization algorithm more effective, and this algorithm can solve part of the kidnapped robot problem. Of course, there are some other localization methods that can solve the robot localization problems [19–21]. However, in some environments the laser point cloud features are degraded, such as chairs, seats, and other items in the room. The environmental information captured by these objects in the 2D lidar is just displayed as a few small points, and the characteristics of such objects are greatly lost. In another case, in multiple similar environments, the existing localization method based on particle filtering will fail. In the face of these two types of environments, although grid maps can be created, AMCL localization algorithms often fail in global localization. Therefore, how to reduce the influence of object features, such as chairs and tables, to create more realistic maps has become a hot topic for researchers. McCormac et al. [22] combined SLAM and Convolutional Neural Networks (CNN) to obtain semantic maps. Salas-Moreno et al. [23] proposed a real-time localization and mapping paradigm, which uses 3D object recognition to skip low-level geometric processing and directly generate incrementally constructed maps at the object-oriented level. Shichao Yang and Scherer [24] propose a general approach for monocular 3D target detection and SLAM mapping without an a priori target model and demonstrate for the first time that semantic target detection and geometric SLAM can mutually benefit from each other in a unified framework. Nicholson et al. [25] combined SLAM techniques with object detection to achieve estimation of optimal camera positions and 3D landmark representation of objects in the environment. Because of the introduction of object-based landmarks, such as pairwise quadratic surfaces, this work has positive implications for the acquisition of semantic information for mobile robots. Kundu [26] uses a method of fusion of semantic segmentation and visual SLAM to establish a three-dimensional semantic map. The above work paved the way for the study of semantic SLAM. Most of the above methods are based on visual SLAM as the main framework. Although visual SLAM performs well in localization, the maps created are difficult to compete with laser SLAM in terms of accuracy and robustness, and there is a cumulative error in the construction of maps created by visual SLAM. It is not conducive to robot navigation and localization. Mobile robots need high-precision maps to meet the needs of localization and navigation in practical applications in indoor environments. Therefore, in this paper, laser SLAM is chosen to build maps.

This paper proposes to use instance segmentation algorithm to create semantic grid map based on 2D lidar and RGB-D camera. Firstly, the 2D grid map is created by using lidar to achieve SLAM to obtain high-precision indoor environment contour information. Secondly, semantic object capture is obtained using a depth camera combined with an instance segmentation algorithm. Semantic grid maps are also created by the back-projection process of the pinhole camera for semantic point cloud mapping, and such maps can solve the problem of environmental information degradation and similarity. This paper also improves the localization algorithm to achieve robust localization and fast convergence of the robot under the degraded features and similar environment of the grid map. Experimental evidence shows that the proposed method in this paper is effective.

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

In summary, the main contributions of this paper are as follows:

1. Proposed a semantic grids map creation system based on 2D lidar and RGB-D camera combined with instance segmentation algorithm, and storage method of semantic grids map.
2. Based on the AMCL algorithm, a new initial particle distribution method is proposed. Experiments show that this method has outstanding performance in global localization under large scenes and similar environments.
3. Compared with the existing AMCL algorithm, this method reduces the maximum particle number burden, shortens the number of iterations, and improves the robustness of the AMCL algorithm for localization under large scenes and multiple similar environments.

The rest of this article is organized as follows. Section 2 analyzes the disadvantages of current grid maps and robot localization based on grid maps. Section 3 introduces the overall scheme of semantic map construction and the specific process of semantic map realization. After that, we analyzed the localization algorithm and improved the design. Then, Section 4 conducts experimental verification and analysis. Finally, Section 5 summarizes this article.

2. Actual problem analysis

2.1. Problem analysis of grid map

In indoor environments, mobile robots usually use 2D lidar to create scale maps. In the process of building a map, the position of the robot to the map changes with the movement of the robot at all times, as well as the observations captured by multiple sensors may be noisy. There may be errors in the values obtained by multiple detections of the same obstacle. So the observation value at one time cannot be used as the basis for identifying the obstacle, the prediction of the observation value is needed to reduce the error and obtain a relatively accurate scale map. Therefore, researchers introduced a probability grid map to describe the scale relationship of the indoor environment, and the grid probability model can be expressed as follows:

$$\log\left(odd(s|z)\right) = \log\frac{P(z|s=1)}{P(z|s=0)} + \log\left(odd(s)\right) \tag{1}$$

where $\log\frac{P(z|s=1)}{P(z|s=0)}$ is the measurement model; $\log\left(odd(s)\right)$ is the grid state value before the update. The grid state update is only related to the measurement value of the first term. Because the grid map adopts binary representation, it is convenient for the navigation, localization, and path planning of mobile robots. Therefore, grid maps are widely used in the field of indoor mobile robots. Currently, the mainstream algorithmic frameworks for building laser SLAM-based grid maps can be divided into two categories: graph-based optimization and filter based. Among them, the representative Cartographer algorithm based on graph optimization creates maps with high accuracy and robustness compared to the Gmapping algorithm using particle filtering. More importantly, Cartographer algorithm can create stable grid maps even in larger environments. The Cartographer is currently the dominant algorithm for research and application. Therefore, this paper uses the Cartographer algorithm to create grid maps. As shown in Fig. 1, a grid map of 8 m wide by 10 m long is created using the Cartographer algorithm.

As shown in Fig. 1, the grid map consists of three different shades of gray, with the three shades of gray indicating three environmental states. The light gray indicates the free state, and the robot can move within the range of this color; however, the dark gray indicates the unknown state, which means that these areas are not determined in the process of building the map, and therefore the robot cannot enter; the black indicates the occupation state, and the black area indicates that it is an obstacle, and the robot is also inaccessible.

It can be seen that the grid map only describes the corresponding environmental information by these three states, as shown in Fig. 1. The amount of map information is single, containing only the outline

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

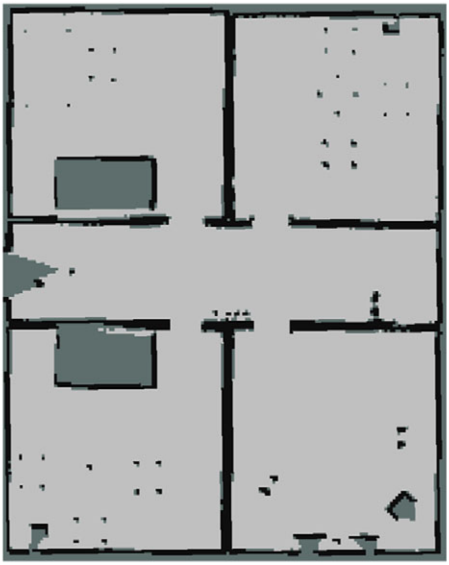


Figure 1. *Grid map of four rooms and corridors created by 2D lidar.*

information of the indoor environment, and the feature difference between points and lines is small, and the environmental information is seriously lost. This will lead to a reduction in the perception of mobile robots in the indoor environment. Furthermore, due to the limitation of the single-line laser characteristic of 2D lidar, when exploring objects like tables and chairs, only the leg obstacles of the objects can be scanned, and then the mapping to the grid map consists of only a few black points. The serious abstraction of objects further enhances the degradation of environmental features and the problem of environmental similarity. Since grid maps cannot meet the needs of mobile robots in semantic navigation, similar environment localization, and human–robot interaction, this paper proposes to build a semantic grid map based on the traditional grid map with an instance segmentation algorithm to address the typical problems of grid map appeal. In this way, the capabilities of mobile robots in semantic navigation, similar environment localization, and human–robot interaction are improved.

2.2. Analysis of localization algorithm problems based on grid map in large scenes and similar environments

Currently, the AMCL algorithm, developed from the MCL algorithm, is the most popular localization algorithm and is widely used in indoor mobile robots. The motion of an indoor mobile SLAM robot cannot be localized with the assistance of GPS like an outdoor robot can. Therefore, its localization needs to be built on top of a grid map like Fig. 1. Using the grid map as a priori information, the AMCL algorithm is used to perform the estimation of the positional pose to determine the robot's position in the indoor environment. Although the AMCL algorithm performs well in solving the global localization and kidnapped robot problem for some scenarios, in certain cases this algorithm is relatively difficult to localize and even fails frequently. For example, classrooms, hotels, hospital wards, and other places are prone to similar room layout. In addition, due to the limitation of the single-line laser of the 2D lidar, the lidar is mainly composed of points when scanning the environment, which further aggravates the feature degradation in such an environment. Therefore, a grid map created in such an environment using 2D lidar is prone to have multiple local areas similar to each other, which resembles the environment shown in Fig. 1. In a similar environment, it will result in a better match between the lidar scan results and the grid map when the AMCL algorithm is performed to update the particle weights. In turn, the illusion of higher particle weights appears in these similar environments after several iterations, making the AMCL

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

algorithm resample many fake particles still during the resampling process. During the iterative process from week to week, these fake particles are constantly involved in the estimation of the bit pose of the mobile robot, generating a high number of iterations and long convergence times. More serious is the problem that leads to frequent failure of AMCL algorithm localization and algorithm failure.

In addition, large scenes are also a kind of problem that needs special consideration. The relationship between scene size and robot is relative. For example, in the case of a handling robot in a factory depot, the size of the factory depot may be a factor in determining the scene size. When the size of the scene is beyond the perception range of the robot's sensors and affects the normal activities of the robot, then the environment can be defined as a large scene. However, for mobile service robots, the number of rooms may be the determining factor for scene size. As mentioned earlier for hospitals, hotels, and other places, the size and layout of rooms have certain similarities. When the number of rooms in these places reaches a certain order of magnitude, the number of similar local areas also increases, and when the robot needs global localization in this environment, it requires a large number of initial particles for initial pose assumptions, which increases the computational burden of localization and localization efficiency. Therefore, solving the balance problem between the fast convergence of global localization and the number of particles in large scenes is also the object of study in this paper. The AMCL algorithm has a serious impact on the localization effect by the maximum initial number of particles in large scenes. The core idea of the AMCL algorithm is based on particle filtering. In particle filtering, every particle at every moment represents a possible hypothesis of a real robot at that moment. On the other hand, based on the assumption that x_t is included in particle set χ_t, the probability is similar to the posterior $bel(x_t)$ of Bayesian filtering:

$$x_t^i \sim p(x_t|z_{1:t}, u_{1:t}) \tag{2}$$

From Eq. (2), it can be seen that the more dense the number of samples in the state space region, the greater the probability that the real state of the robot falls into that region. Therefore, for the algorithm to have a good localization effect in large scenes, the number of particles per unit area of the grid map needs to be as high as possible. Although the higher the particle density, the better the localization of the robot. However, the computational burden on the processor also increases. More importantly, when the total number of particles reaches a certain value, the algorithm will cause serious particle storm problems in the convergence process, resulting in a longer convergence process and large time consumption, as shown in Fig. 2. Therefore, a reasonable selection of the initial number of particles has more scope for improving the localization performance of the AMCL algorithm in large scenes.

According to Fig. 2, it can be seen that the maximum initial number of particles is 30,000. When the number of iterations is 20, most of the particles have been discarded, leaving more particles remaining in the region with high local similarity. Most of the remaining particles are due to the similar environment, which leads to a high degree of laser matching, so that the particles at this areas continue to maintain a high weight during the weight update process, but cannot be discarded during the resampling process. Therefore, for global localization in large scenes and similar environments, the reasonable distribution of initial particles and the selection of the maximum initial number of particles become crucial.

For the two specific problems of the appealed AMCL algorithm, this paper uses semantic information to assist and redesign the initial particle distribution of the algorithm to solve the localization problems of similar environments and large scenes. The adaptability of AMCL algorithm to different scenes is improved.

3. Methods

3.1. General scheme of semantic map construction

As shown in Fig. 3, the system is divided into two parts, one part is laser SLAM, which uses 2D lidar and odometer information with Cartographer algorithm to create a grid map. The other part is to generate semantic point clouds. First, the RGB images captured by the depth camera are used for the instance segmentation algorithm to capture the instance objects. Meanwhile, the RGB image is aligned with

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

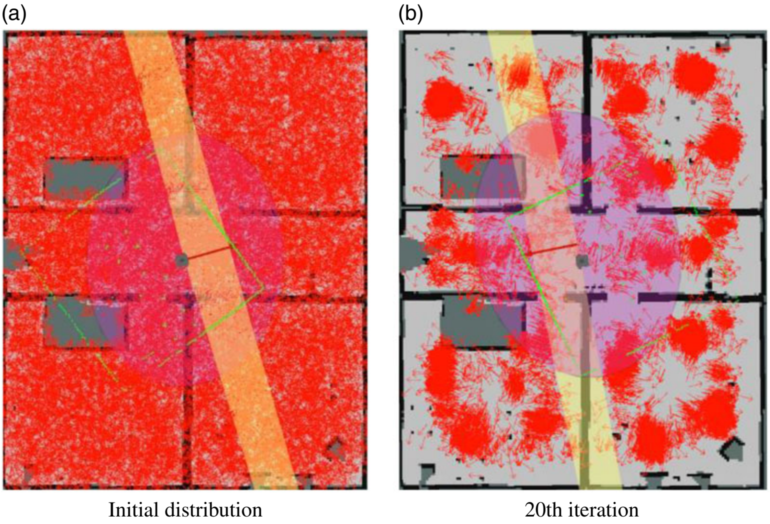


Figure 2. *Convergence process for a maximum initial particle number of 30,000.*

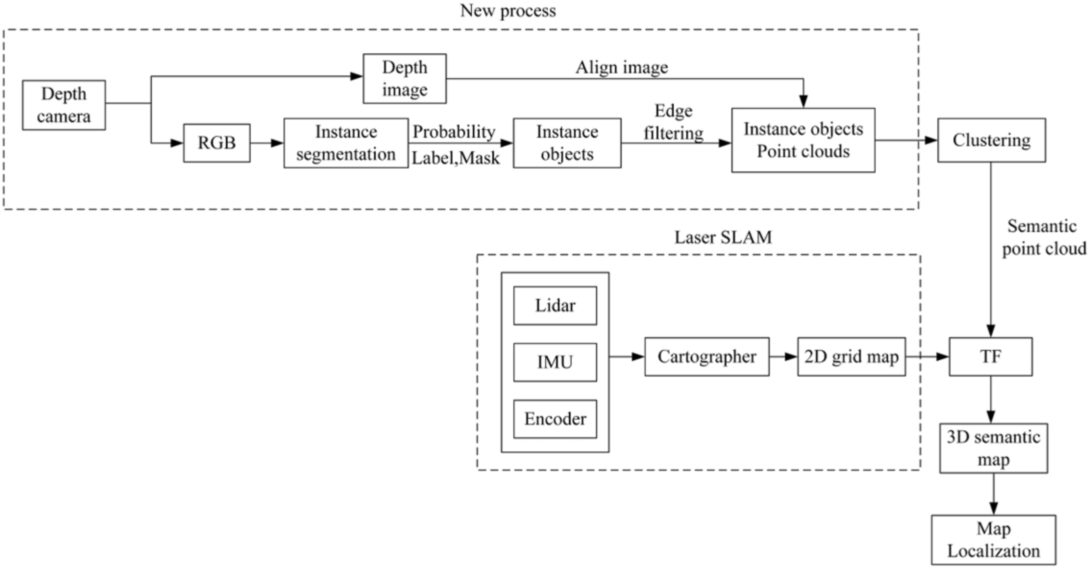


Figure 3. *The system framework for implementing semantic grid map by instance segmentation.*

the depth image and the pixel points corresponding to the instance object are transformed to the camera coordinate system. Second, the semantic point cloud is transformed by coordinates and these 3D semantic point clouds are fused with 2D grid maps to obtain semantic grid maps. Finally, the semantic grid map is localized to realize the secondary loading of semantic information to meet the more advanced localization and navigation tasks of mobile robots.

3.2. Semantic information obtaining method

In order to build a semantic grid map, this paper introduces deep learning related methods to obtain semantic objects for the purpose of adding semantic information to the grid map. In indoor environments,

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

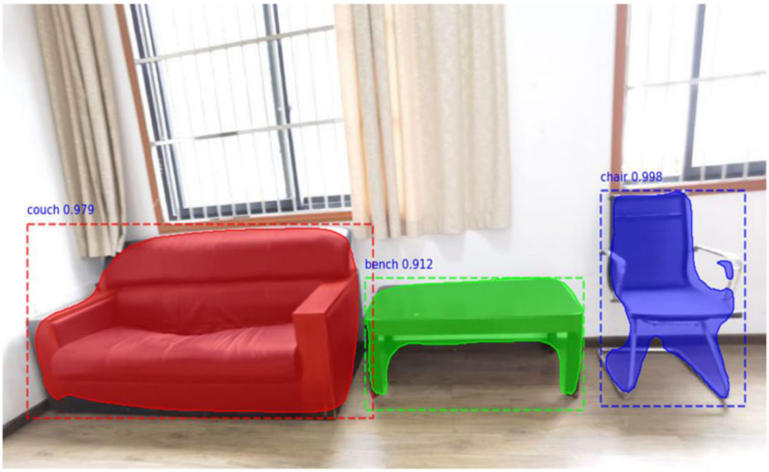


Figure 4. *Instance segmentation effect.*

objects can be classified into three categories: static, semi-static, and dynamic. Among them, dynamic objects, such as people, pets, and other objects, may move at any time and whose relative position is not fixed. The information of such objects captured by lidar and camera sensors for map building will affect the robustness of the map and indirectly affect the localization and navigation tasks of mobile robots. Therefore, when capturing semantic objects, it is necessary to eliminate dynamic objects in the environment to ensure high robustness of semantic information. The other type is static objects, such as beds, tables, doors, and windows. The posture of the objects is relatively fixed and not easy to change. More importantly, lidar captures such objects with significant loss of object features, which can easily lead to degradation of environmental features and thus exacerbate the problem of environmental similarities, making the robot less capable in localization. Therefore, the characteristics, attributes, types, and other information of such objects can be added to the grid map to compensate for the disadvantages of the current grid map. The task of semantic segmentation is to assign the pixels of the image to the corresponding category labels according to the existing classification [27–29]. Comparing semantic segmentation and instance segmentation, semantic segmentation is difficult to distinguish the classification problem of cross pixel points of the same category, while the individual pixel points segmented by the instance segmentation method are relatively independent, which can solve the classification problem of different individuals of the same kind of objects. In order to obtain more accurate segmentation results and facilitate the point cloud processing of a single object, this article chooses to combine with instance segmentation. Comparing with segmentation algorithms such as Faster R-CNN [30], YOLACT [31], and Mask R-CNN [32], among which Mask R-CNN can detect the target of the image and also discriminate each object, the network has good scalability and high recognition accuracy. Therefore, this paper chooses Mask R-CNN as the instance segmentation algorithm. Figure 4 shows the segmentation result of Mask R-CNN instance.

Mask R-CNN can recognize thousands of categories. In the indoor environment, it mainly contains common categories such as bed, sofa, and table. In Fig. 4, the scene built in an indoor environment, Mask R-CNN recognizes three categories of couch, bench, and chair at the same time and the corresponding recognition accuracy. Meanwhile, the mask image of each individual is given, as shown in Fig. 5.

3.3. Fusion of semantic information and grid map

In order to map the semantic objects to the grid map, the following work was done to achieve the fusion with the grid map. First, it is required to determine the 3D relationship of the semantic objects in the

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

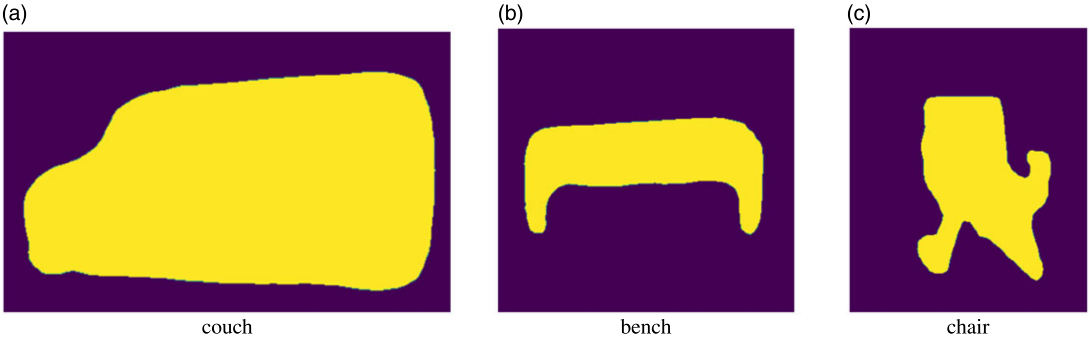


Figure 5. *The mask image of each instance.*

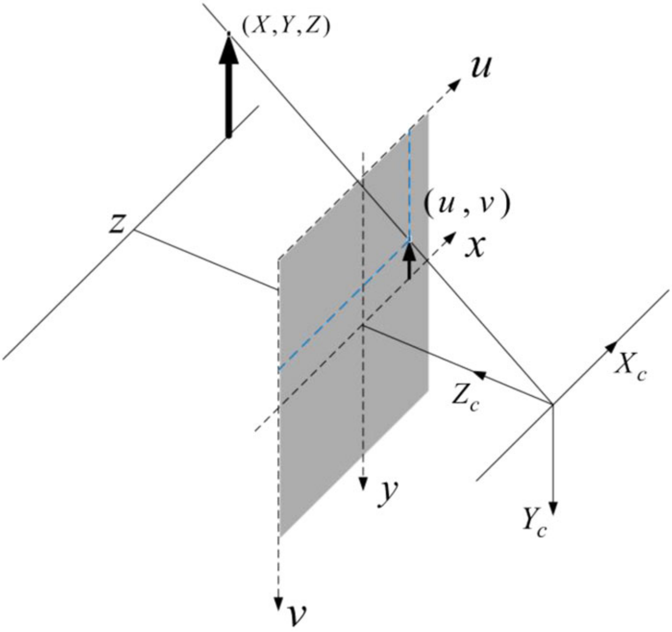


Figure 6. *Pinhole camera projection process.*

camera coordinate system. According to Fig. 6, the camera coordinate system is $(X_C, Y_C, Z_C)^{\mathrm{T}}$. The pixel coordinate system is $(u, v)^{\mathrm{T}}$, which is represented in dark gray in the figure.

The pixel coordinates containing semantic information need to be converted to the camera coordinate system. According to the principle of pinhole imaging, it can be expressed as follows:

$$z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathrm{K}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where K is the camera internal reference matrix; u, v is the pixel coordinates of the mask image. Then, the depth information d corresponding to the semantic pixels is obtained by aligning the RGB image and the depth image. Combining Eq. (3), the pixel $(u, v, d)^{\mathrm{T}}$ converted to the three-dimensional point

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

cloud P_c, $(X_c, Y_c, Z_c)^T$ under the camera coordinates can be expressed as follows:

$$\begin{cases} X_c = (u - c_x)Z_c/f_x \\ Y_c = (v - c_y)Z_c/f_y \\ Z_c = d/1000 \end{cases} \tag{4}$$

At this time, the pixel points containing semantics have been transformed to the 3D space under the camera coordinate system. Since the semantic point cloud contains some noisy points, some of them are filtered by clustering. Finally, the relative relationship between the camera coordinate system and the map coordinate system is obtained, and the semantic point cloud $\mathrm{P_c}$ converted to the map coordinate system $\mathrm{P_m}$ can be expressed as:

$$\mathrm{P_m} = \mathrm{R} \times \mathrm{P_c} + \mathrm{T} \tag{5}$$

where R is the rotation matrix and T is the translation vector. The semantic point cloud in the camera coordinate system is converted to the map coordinate system by Eq. (5). So far, the entire semantic point cloud mapping process is completed.

3.4. Localize semantic grid map

The semantic point cloud features of objects in indoor environments extracted by instance segmentation and point cloud transformation are referred to here as semantic landmarks. After secondary map loading, these semantic landmark features can be used for semantic-assisted localization, semantic navigation, human–robot interaction, etc. of the robot. Usually, grid maps are localized in map.png and map.yaml formats, which do not contain the preservation of semantic landmarks. Therefore, it is an important task to localize these semantic landmarks as well as grid maps. In this paper, we propose the idea of container $\{C_i\}$ to store semantic landmarks. In the containers, each container contains one or more semantic landmark objects of the same class $C_i\{SemanticLabels_{1,2,3...}\}$. The relevant parameters of the corresponding semantic landmark can be obtained through the landmark object, such as the semantic category, the point cloud clustering center, and the pose of the point cloud constituting the semantic landmark in the map coordinate system. Finally, the container is localized with the grid map in a configuration file format. When the robot loads the map twice, semantic landmarks and high-precision grid maps are loaded simultaneously to improve the robot's ability to perceive and generalize the environment. More importantly, it overcomes some environmental information degradation problems caused by laser point cloud capture, which makes the robot promising for higher-order navigation and localization tasks and adaptation to more demanding environments.

3.5. Improved localization algorithm

The core of the AMCL algorithm is to estimate the posterior poses of the robot by combining probabilistic motion and perceptual models with particle filtering. The AMCL algorithm is a combination of Augmented_MCL and KLD_Sampling_MCL. It can solve the global localization of the robot and the kidnapped robot problem in some situations. The main flow of the Augmented_MCL algorithm is shown in Algorithm 1.

In Algorithm 1, the set $\chi_t = \{x_t^{[1]}, x_t^{[2]}, x_t^{[3]}, \ldots, x_t^{[M]}\}$ of M particles is used to express the confidence level $bel(x_t)$. The initial confidence is obtained from the M particles randomly generated by the prior distribution. The 4th line of the algorithm is the prediction stage. The pose at the new time is estimated according to the motion model. x_{t-1} is the state estimator at time $t-1$ and u_t is the control variable at time $t-1$ to t. The state transition distribution is obtained from the motion model, and the hypothetical particle state $x_t^{[i]}$ is randomly sampled at time t from this distribution.

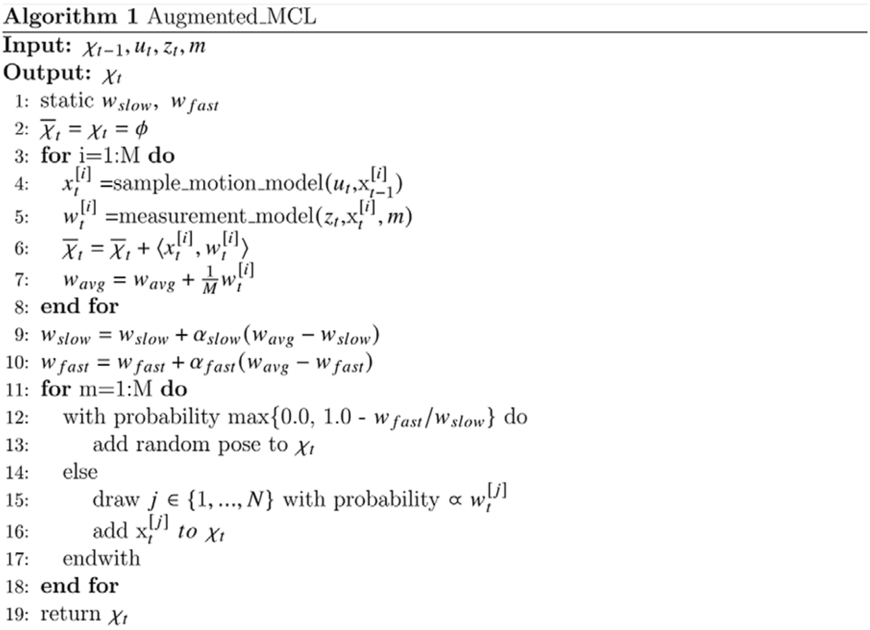
$$x_t^{[i]} \sim p(x_t|x_{t-1}, u_t) \tag{6}$$

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

In line 5 of the Algorithm 1, the particle importance weights $w_t^{[i]}$ are updated according to the observed model, and the observed values $w_t^{[i]}$ can be expressed as follows:

$$w_t^{[i]} = p(z_t|x_t^{[i]}) \tag{7}$$

The new particle set is obtained in line 6 of the Algorithm 1 and lines 11–18 are the resampling process. AMCL is combined with KLD to achieve dynamic adjustment of the number of particles according to the localization convergence. The resampling process differs between AMCL and Algorithm 1 and is not described in detail here.



```
Algorithm 1 Augmented_MCL
Input: $\chi_{t-1}, u_t, z_t, m$
Output: $\chi_t$
 1: static $w_{slow}$, $w_{fast}$
 2: $\overline{\chi}_t = \chi_t = \phi$
 3: for i=1:M do
 4:     $x_t^{[i]}$ =sample_motion_model($u_t$,$x_{t-1}^{[i]}$)
 5:     $w_t^{[i]}$ =measurement_model($z_t$,$x_t^{[i]}$, $m$)
 6:     $\overline{\chi}_t = \overline{\chi}_t + \langle x_t^{[i]}, w_t^{[i]} \rangle$
 7:     $w_{avg} = w_{avg} + \frac{1}{M} w_t^{[i]}$
 8: end for
 9: $w_{slow} = w_{slow} + \alpha_{slow}(w_{avg} - w_{slow})$
10: $w_{fast} = w_{fast} + \alpha_{fast}(w_{avg} - w_{fast})$
11: for m=1:M do
12:     with probability max{0.0, 1.0 - $w_{fast}/w_{slow}$} do
13:         add random pose to $\chi_t$
14:     else
15:         draw $j \in \{1, \ldots, N\}$ with probability $\propto w_t^{[j]}$
16:         add $x_t^{[j]}$ $to$ $\chi_t$
17:     endwith
18: end for
19: return $\chi_t$
```

For the problems of localization in similar environments and large scenes in this algorithm of AMCL, this paper proposes the use of semantic landmark-assisted localization to solve them. According to Monte Carlo sampling, it is known that the a priori particle distribution affects the accuracy of the a posteriori positional estimation. As shown in Fig. 2, all the particles of the AMCL algorithm adopt only one distribution in the initial state, that is, the form of uniform distribution. However, in this paper, these initial particles will be divided into two parts. One part of the particles obeys the original global uniform distribution, which is the same as the present AMCL algorithm, in which the particle samples are collected uniformly in the global free area of the grid map. This part of particles ensures the initial particle diversity of AMCL algorithm and makes the localization have certain robustness; the other part of initial particles adopts semantic landmarks to assist the distribution, and the semantic information of the grid map can improve the robot's ability to perceive the environment and improve the robot's ability to recognize similar environments. The robot can be assisted by these semantic information in the process of localization to narrow down the estimated range of the robot in the global grid map, enhance the initial particle density in the key possible regions, and reduce the localization difficulty, so as to solve the localization problem of AMCL in large scenes and similar environments.

The process of semantic landmark-assisted sampling starts with determining the relative relationship description between the robot and the semantic landmarks. The semantic landmarks in the semantic grid map are determined and localized during the construction of the semantic layer. The absolute poses of the semantic landmarks in the grid map are known after secondary loading, and there is a one-to-one correspondence between the semantic landmarks and the poses of different objects in the real-world coordinates. At any moment, the semantic information observed by the robot through the camera

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

contains a distance information, an orientation information, and a label information after segmentation by instances, which can be represented by a multidimensional vector as:

$$L(z_t) = \{C_t^1, C_t^2, \ldots\} = \left\{ \begin{pmatrix} r_t^1 \\ \theta_t^1 \\ label_1 \end{pmatrix}, \begin{pmatrix} r_t^2 \\ \theta_t^2 \\ label_2 \end{pmatrix}, \quad \ldots \right\} \tag{8}$$

where z_t denotes the observed quantity at time t, r denotes the distance between the robot and the semantic landmark, and θ denotes the orientation of the robot and the semantic landmark. *label* denotes the semantic label, where it is assumed that there is one and only one semantic landmark of each class, but the actual situation may contain k identical semantic label signatures. The grid map adopts a right-angle coordinate system, and if the absolute pose of the robot is (x, y, α), the absolute pose of the label $(k = 1)$ is (x_l, y_l), and the relative relationship between the robot and the semantic labels can be expressed as:

$$\begin{cases} r = \sqrt{(x_l - x)^2 + (y_l - y)^2} \\ \theta = \text{atan2}\,(y_l - y, x_l - x) \end{cases} \tag{9}$$

In the process of global localization, the pose of the robot is unknown, but the relative pose of the corresponding semantic point cloud information can be obtained through the key frame captured at the current moment, and the number of semantic point cloud information is n. The distance r relationship between the robot and the semantic landmark can be expressed as:

$$r = \frac{1}{n} \sum_{i=1}^{n} label_{z_i} \tag{10}$$

The semantic label of the key frame at the current moment can be used to obtain the absolute position (x_l, y_l) of the semantic landmark of the same name in the grid map. At this point, it can be determined that the robot is in a circular region with position (x_l, y_l) as the center and r as the radius. Therefore, the particle state with the landmark-assisted posterior can be expressed as:

$$p(x_t | C_t^i, label_j, m) \tag{11}$$

After determining the relative positions of the robot and the semantic landmarks, the particle distribution strategy in the ring region is then determined. The closer to the ring center radius region, the more dense the particle distribution is, and away from the center radius region the more sparse the particle distribution is. In this paper, a one-dimensional Gaussian model is used to implement the particle distribution in the annular landmark region. δ indicates the dispersion of the sampled particles with the radius center of the sampled annular region, and δ can be expressed as:

$$\delta = \eta \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \tag{12}$$

where $x \in (-3\sigma, 3\sigma)$, $\mu = 0$. After determining the relative position and the distribution problem of sampling the ring region, the specific flow of particle sampling in the initial state of AMCL can be expressed as shown in Algorithm 2.

Algorithm 2 provides a detailed description of the initial state distribution process of the AMCL algorithm. Line 1 of the algorithm divides the total number of input particles M into two parts proportionally. γ_t denotes the set of particles obeying the global uniform distribution and λ_t denotes the set of particles obeying the landmark sampling distribution. Thus, the initial state of particles can be expressed as:

$$\chi_t = \lambda_t\{x_t^i | i \le N\} \cup \gamma_t\{x_t^j | j \le M - N\} \tag{13}$$

Line 6 determines the relative distance between the landmark and the robot, r is the locus sampling radius, and δ is the dispersion coefficient. Lines 8–11 show the global poses (x_i, y_i, θ) for each particle after landmark sampling, where α obeys a uniform distribution of [0,1].

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

Algorithm 2 Particles Initial State Landmark Sampling

```
Input: M,$f^i$, $r$, $m$
Output: $\overline{\chi}_t$
Determine the number of semantic sampling particles N=$\varepsilon M$.
$\gamma_t = \phi$
find _semantic_landmarks($f_i$)
while $s \leq N$ do
    for i=1:k do
        Particles conform to Gaussian model distribution
        $r_i = \frac{1}{n}\sum_{j=1}^{n} l_{z_j} \pm \delta$
        $Sure\ label_i\ coordinates\ in\ the\ map\ coordinate\ system,\ (x_{l_i}, y_{l_i})$
        $x_i = rand([-r_i, r_i])$
        $y_i = \pm\sqrt{r_i^2 - (x_i - x_{l_i})^2} + y_{l_i}$
        $\alpha$ follow [0,1] uniform distribution
        $\theta = \alpha * 2\pi - \pi$
        Put the coordinates $(x_i, y_i, \theta)$ $into$ $\lambda_t^{[i]}$
        Judge the range of s and update the status
    end for
end while
$\overline{\chi}_t = \chi_t = \lambda_t \cup \gamma_t$
return $\overline{\chi}_t$
```

On the one hand, the distribution of the set λ_t increases the number of sampled initial particles in the focal similar region, which increases the density of particles per unit area in the focal region. The probability that the particle poses of the initial state represent the real poses of the robot is greater, improving the ability of the AMCL algorithm to cope with global localization in similar regions and reducing the probability of AMCL localization failure. On the other hand, compared with the original one that only single obeys the global uniform distribution, semantic landmark sampling reduces the particle burden of AMCL algorithm that requires high density in large scenes and similar environments, indirectly reducing the number of particles and computational burden. Therefore, the initial particle distribution based on landmark sampling will enable the AMCL algorithm to make significant performance improvements in terms of algorithm localization success rate and convergence iteration efficiency while reducing the maximum number of initial particles.

4. Experimental verification and analysis

Considering the physical space limitation of the laboratory, it is difficult to build a real similar environment, so this paper uses the Gazebo plug-in in ROS as a platform to build the environment. In Gazebo, it is convenient to build the experimental scenarios according to the experimental requirements. In order to ensure the consistency and persuasiveness of the experiment, a 12 × 14m experimental environment built using Gazebo is shown in Fig. 7, which is similar to Fig. 2. In which, the size and dimensions of the room are the same. Second, the robot model is chosen as TURTLEBOT3 (waffle). The CPU of the laptop is Intel(R) Core (TM)i7-9750@2.6GHz 5.59GHz, the GPU is NVIDIA GeForce RTX 2080, and the RAM is 16GB. Various programming languages and coding rules are used in the experimental part of this paper. For the SLAM part, the coding conforms to the ROS programming specification and is mainly implemented in C++ and Python. Example segmentation and point cloud processing are implemented using python. The specific software framework of the experiment is shown in Fig. 8.

According to the creation process of the semantic map, the grid map is first created using the Cartographer algorithm, and a new process is started to capture the 3D semantic point cloud information. Then, the semantic point cloud information is transformed and fused to obtain the semantic grid map. Finally, the semantic landmarks are localized and saved. As shown in Fig. 9, the semantic grid map is loaded again.

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

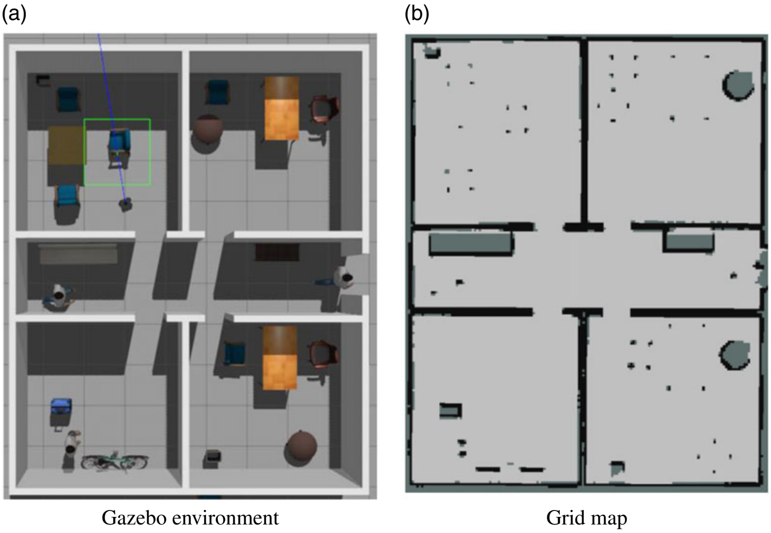


Figure 7. *Gazebo environment and environment grid map.*

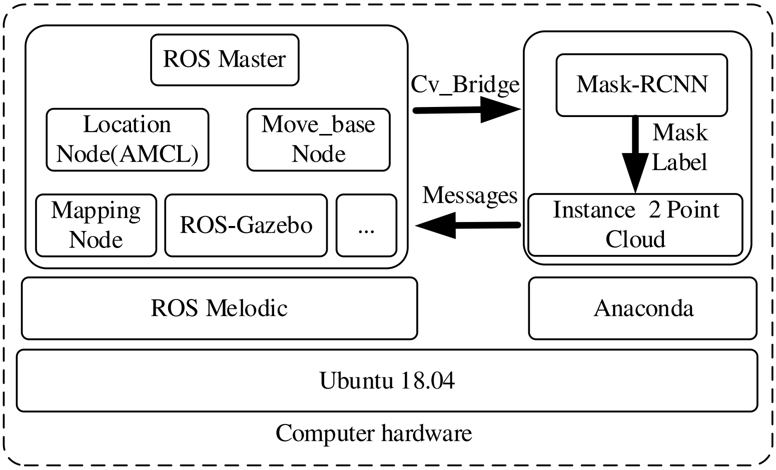
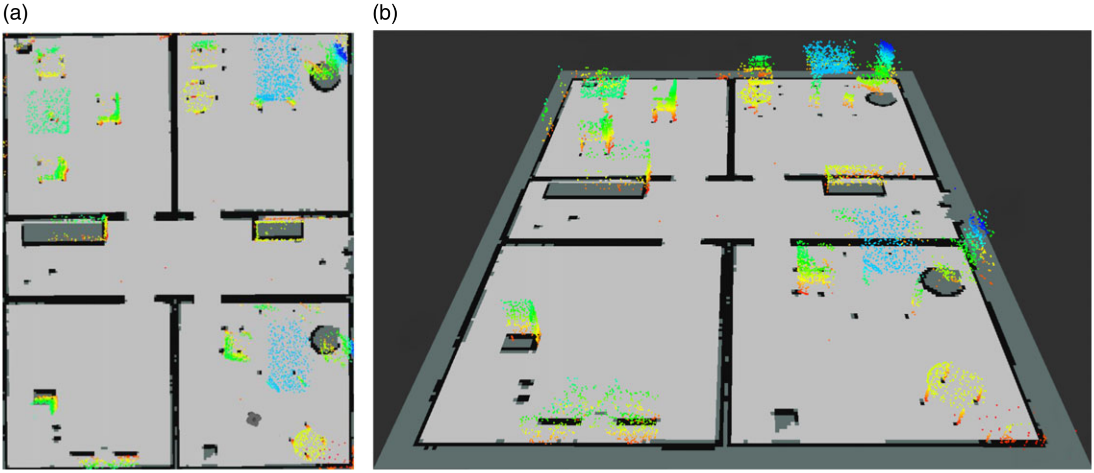


Figure 8. *Experimental software framework diagram.*

The semantic grid map built by capturing indoor objects such as sofa, table, and chair is implemented on the basis of the grid map. Compared with a single grid map, as shown in Fig. 7(b), the semantic grid map established by the semantic point cloud mapping system proposed in this paper has a richer understanding of the environment, such as the category, contour, and pose of some objects. It improves the robot's ability to understand the environment. The implementation of the semantic grid map enables the mobile robot to perform advanced tasks such as semantic-assisted localization, semantic navigation, and human–robot interaction.

In order to verify the effectiveness of the semantic information-assisted localization method proposed in this paper, the following comparison experiments are made with this paper's method and the AMCL algorithm on the basis of Fig. 7(a). The existing AMCL algorithm goes to the maximum initial number of particles of 2000 and 20,000, respectively, and the improved algorithm takes the maximum initial number of particles of 6000. First, Fig. 10(a) shows the initial state when the maximum number of particles of the AMCL algorithm is 2000, and all particles are evenly distributed in the free area of the

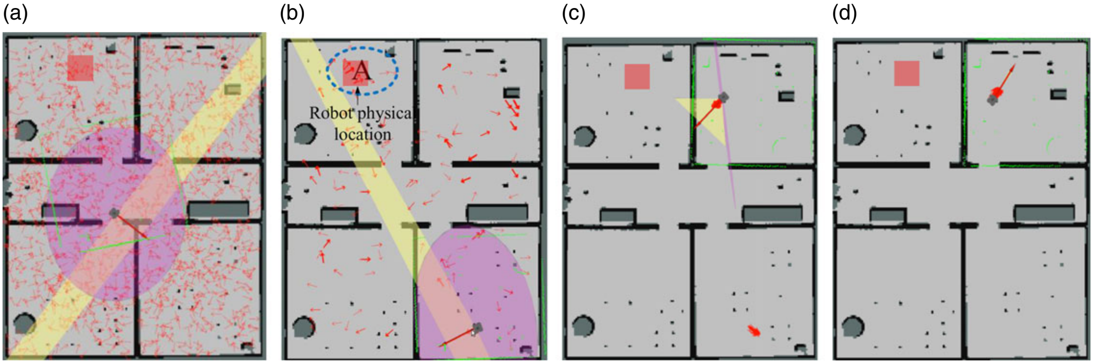
https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press



Top view

Front view

Figure 9. *Semantic grid map.*



Initial distribution

10th iteration

100th iteration

163th iteration

Figure 10. *Convergence process for a maximum initial particle number of 2000.*

grid map. The real pose of the robot is located in the light red rectangular box in the figure, and the algorithm starts iterating when the mobile robot starts moving. Among them, the process in Fig. 10(b) and (c), there are several local regions with a high number of particles; however, the particles carrying the real poses of the mobile robot are discarded in the process of continuous iteration, such as the particles in region A in Fig. 10(b). Figure 10(d) shows the effect of 163 iterations. At this time, Max_weight converges to 1, but it can be seen from the figure that the posterior pose of the AMCL algorithm is wrong. Because the environment is similar, the matching degree of the laser is high, but the particles that can accurately reflect the pose of the robot are discarded. Naturally, when the maximum number of initial particles is small, the localization robustness is not high, and the AMCL algorithm is prone to failure. Nevertheless, compared to Figure 11(a), when the maximum number of particles in the initial state is 20,000, the initial particle density in the free area of the grid map is high at this time, and the particles near the real pose increase. In the process of continuous iteration, these particles have higher weights and are not easily discarded. On the contrary, those particles in similar regions have high weights with lidar matching, but they will slowly decrease in the iterative process compared with the particles in the real poses. As a result, the probability of successful localization of the robot also increases relatively, and the robustness of localization increases. As shown in Fig. 11(b), the ellipse region is the predicted

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

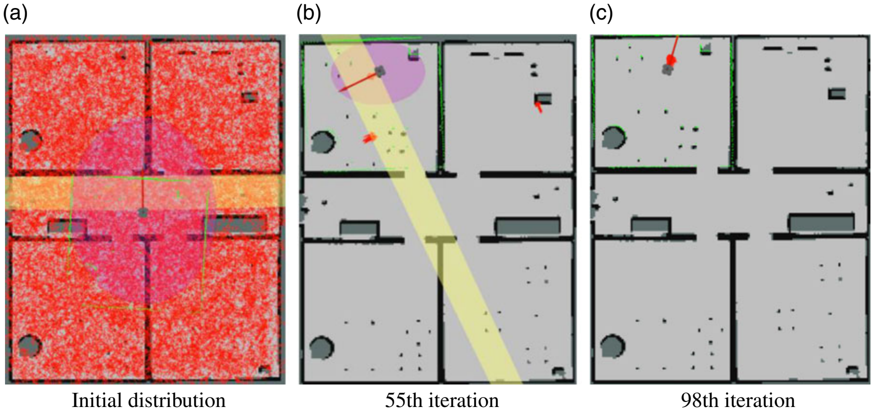


Figure 11. *Convergence process for a maximum initial particle number of 20,000.*

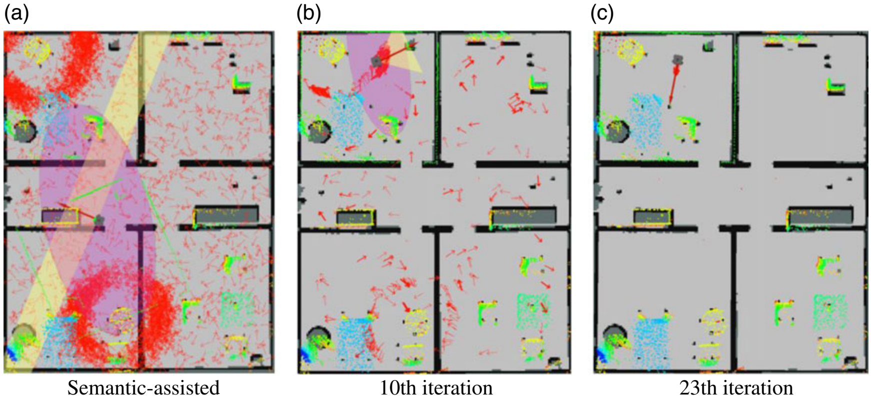


Figure 12. *Improving the convergence process for the maximum initial particle number of 6000 for the initial distribution.*

position of AMCL algorithm, and after 55 iterations, the robot has been localized in the real position, and the Max_weight is 0.556 at this time. Finally, the Max_weight converges to 1 at 98 iterations, and the localization is successful, as shown in Fig. 11(c).

In the same grid map, localization experiments are conducted using the improved localization method in this paper. In the initial state, the maximum number of particles is taken as 6000. 6000 particles are then divided into two parts according to the algorithm design requirements, of which 2000 are used for the global uniform distribution of the grid map and the rest of the particles are sampled according to the assistance of semantic landmarks. First, the RGB images and depth information captured by the camera are combined with the instance segmentation algorithm to obtain the semantic labels and distance information of the robot's view at this time. As shown in Fig. 12(a), in the top view of the map, the semantic label at this point is "table," and the "table" is matched with the loaded global semantic landmark. The initial particle sampling is performed in the result of the matching. As can be seen in Fig. 12(a), since there are two "table" instances of the global semantic landmark, two identical uniform distributions are sampled at the corresponding locations based on the semantic landmark and distance information, respectively. So far, the initial particle distribution assisted by semantic landmarks has been determined. Compared with the initial particle distribution of the AMCL algorithm, the particles of this method are assisted by semantic landmarks to increase the particle density in the area where the robot may exist. More importantly, it reduces the burden of the maximum number of particles. On the basis of the initial

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

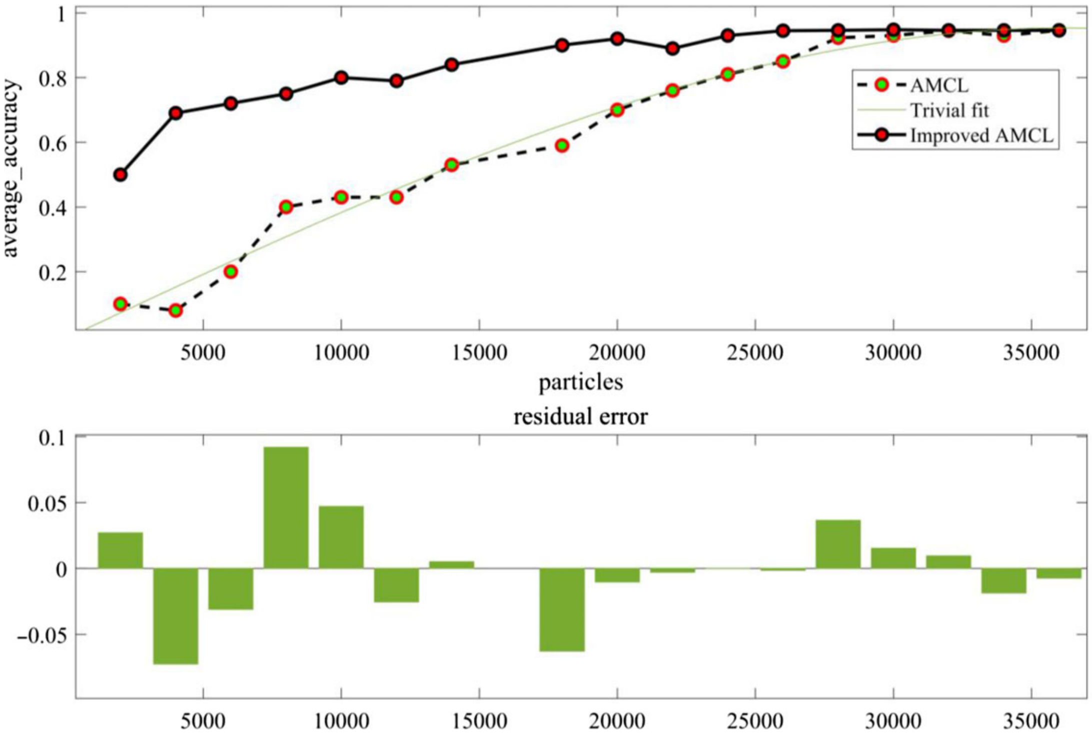


Figure 13. *Graph of maximum initial particle number versus average success rate of AMCL localization and residuals.*

state, after 10 iterations of the improved AMCL algorithm, most of the particles have been gradually discarded in the iterative process. As shown in Fig. 12(b), the convergence area of the algorithm, that is, the elliptical area in the figure, has been localized near the real position of the robot, and the Max_weight at this time is 0.862. Finally, when the number of iterations is 23, at this time Max_weight is 1, a global localization convergence process is completed, and the a posteriori positional results match the real poses of the robot. The above experiments demonstrate the superiority of the algorithm proposed in this paper on the basis of semantic grid maps. In large scenes and similar environments, the semantic grid map improves the robot's ability to perceive the environment and increases the ability to distinguish the environment. From the convergence process, it can be seen that the improved AMCL algorithm, which uses less number of particles and increases the particle density of possible regions, improves the localization ability of AMCL algorithm.

To further compare the localization effect of the improved algorithm with the existing AMCL algorithm, the experiments were repeated extensively in the setting of Fig. 7. In this set of experiments, the area size is fixed and the initial number of particles is variable. First, the initial number of particles is divided into 17 groups from 2000 to 34,000, and then, each group does 60 global localization experiments separately to find the average success rate of localization in the corresponding group respectively. Finally, a nonlinear fit was made based on the results of the appealing large number of localization experiments, as shown in Fig. 13. As can be seen from the figure, when the initial number of particles is small, the probability of successful localization is relatively low and the residuals with the fitted curve are relatively scattered, so the robustness of the AMCL algorithm at this time is low for localization in such similar environments and the algorithm is prone to failure. However, as the number of initial particles increases, the density of particles per unit size increases, and the localization success rate also increases, and the average success rate of localization and the residuals error of the fitted curve decrease. At this time, the AMCL algorithm is more robust in localization. Comparing the initial number of particles

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

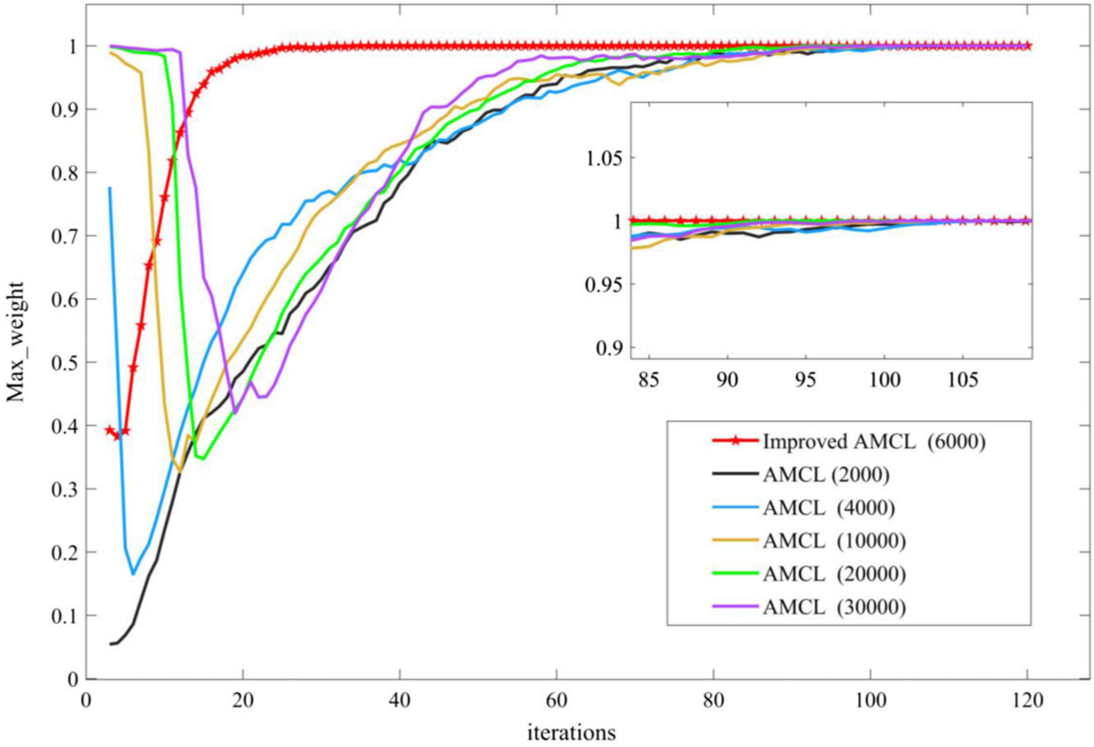


Figure 14. *Comparison of particle weight convergence between semantic landmark distribution and original distribution.*

with 2000 and 36,000, it can be found that the increase in the number of particles during the experiment leads to a sharp increase in computation and a particle storm that causes a long time to maintain a large number of particles without convergence. Therefore, the number of particles cannot be increased blindly. Comparing the localization success rate of the improved algorithm, we can find that the average success rate of localization in the same environment is still around 80% even when the maximum initial number of particles is below 10,000, which is much higher than the existing AMCL algorithm. With the increase of the maximum initial particle number, the localization rate also increases. Therefore, the improved algorithm in this paper has high localization success rate and better robustness in large indoor scenes and similar environments.

In addition, the improved algorithm still performs outstandingly in terms of the number of iterations and convergence time. As shown in Fig. 14, the curves without the pentagram labeled are the convergence process of the AMCL algorithm with the maximum number of particles between 2000 and 30,000, respectively; the curve in red and labeled with the pentagram is the convergence process of the algorithm in this paper taking the maximum number of particles as 6000. It is found that the number of iterations in the convergence process of existing AMCL algorithm is higher in such environment. Although the Max_weight reaches 0.9 or more in about 60 iterations, the number of iterations is high because several local particles do not converge completely, and the highest one even reaches 105 iterations. Nevertheless, in the process of convergence of the algorithm proposed in this paper, Max_weight is close to 1 after about 20 iterations, and basically completely converges when the number of iterations is 35 times. As a result, the algorithm proposed in this paper significantly reduces the number of iterations for global localization in such environments.

Finally, the improved AMCL algorithm in this paper is compared with the traditional AMCL algorithm for localization convergence time, as shown in Table I. The amount of robot state control is the

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

Table I. *Comparison of positioning convergence time.*

Number of initial state particles	AMCL localization time (s)	Improved AMCL localization time (s)
2000	12	7
6000	18	9
10,000	20	15
20,000	44	17
30,000	78	28

same, and the same amount of state transfer is guaranteed. The improved algorithm in this paper is more advantageous in the overall convergence time. Especially after the number of particles is greater than 10,000, the improved AMCL algorithm is much faster than the traditional AMCL algorithm in terms of convergence time.

Therefore, after comparison experiments, it is found that the semantic grid map mapping system proposed in this paper overcomes the problem of environmental feature degradation. Second, the localization algorithm assisted by using semantic landmark information is more adaptable to large scenes and similar environments compared to the AMCL algorithm. Although in the case of reducing the maximum initial particle number, there is a significant reduction in the number of iterations of the algorithm and the achievement of robust localization effects.

5. Conclusion

The related work in this paper is about solving the problem that the AMCL algorithm is difficult to locate in a priori large scenes and similar environments. Firstly, a mapping system using instance segmentation algorithm to construct semantic grid graph is proposed by using the combination of lidar and depth camera. And the innovative method of storing semantic landmarks in containers is proposed, which can facilitate the later loading and use of semantic information. Next, the grid map incorporating semantic landmarks improves the robot's perception of the environment and solves the problem of degradation of environmental features due to the limitation of 2D laser point cloud. Secondly, the method of using semantic landmark information to assist in solving global localization is also proposed, which improves the initial particle distribution for global localization of the AMCL algorithm. Experimentally, it is proved that in the same environment, Algorithm 2 enables the AMCL algorithm to achieve a localization success rate of about 90% or even higher, while the maximum number of initial particles required and the number of iterations in the convergence process are much less than the original AMCL algorithm, reducing the influence of the localization success rate of AMCL algorithm depending on the maximum number of particles. It realizes robust and fast localization of robots in large scenes and similar environments.

Data availability. Not applicable.

Author contributions. All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Hengyang Kuang, Yansheng Li, and Yi Zhang. The first draft of the manuscript was written by Hengyang Kuang, Yong Wan, Gengyu Ge and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding. This research is supported by National Natural Science Foundation of China (61803058) and China Postdoctoral Science Foundation (2021MD703939).

Conflicts of interest. The authors have no competing interests to declare that are relevant to the content of this article.

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

References

[1] W. Cui, Q.D. Liu, L.H. Zhang, H.X. Wang, X. Lu and J. Li, "A robust mobile robot indoor positioning system based on Wi-Fi," *Int. J. Adv. Robot. Syst.* **17** (2020).

[2] G. Lee, B.C. Moon, S. Lee and D. Han, "Fusion of the SLAM with Wi-Fi-based positioning methods for mobile robot-based learning data collection, localization, and tracking in indoor spaces," *Sensors* **20** (2020).

[3] B. Tao, H. Wu, Z. Gong, Z. Yin and H. Ding, "An RFID-based mobile robot localization method combining phase difference and readability," *IEEE Trans. Automat. Sci. Eng.* **18**, 1406–1416 (2021).

[4] H. G. Min, X. Wu, C. Y. Cheng and X. Zhao, "Kinematic and dynamic vehicle model-assisted global positioning method for autonomous vehicles with low-cost GPS/Camera/In-vehicle sensors," *Sensors* **19** (2019).

[5] C. Debeunne and D. Vivet, "A review of visual-LiDAR fusion based simultaneous localization and mapping," *Sensors* **20** (2020).

[6] W. Hess, D. Kohler, H. Rapp and D. Andor, "Real-Time Loop Closure in 2D LIDAR SLAM," **In:** *Proceedings of IEEE International Conference on Robotics and Automation* (2016) pp. 1271–1278.

[7] S. Kohlbrecher, O. Von Stryk, J. Meyer and U. Klingauf, "A Flexible and Scalable SLAM System with Full 3D Motion Estimation," **In:** *Proceedings of IEEE International Symposium on Safety, Security, and Rescue Robotics* (2011) pp. 155–160.

[8] G. Grisetti, C. Stachniss and W. Burgard, "Improved techniques for grid mapping with Rao-Blackwellized particle filters," *IEEE Trans. Robot.* **23**, 34–46 (2007).

[9] K. Konolige, G. Grisetti, R. Kummerle, W. Burgard, B. Limketkai and R. Vincent, "Efficient Sparse Pose Adjustment for 2D Mapping," **In:** *Proceedings of IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings* (2010) pp. 22–29.

[10] R. Mur-Artal, J. M. M. Montiel and J. D. Tardos, *ORB-SLAM: A Versatile and Accurate Monocular SLAM System* (2015).

[11] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Trans. Robot.* **33**, 1255–1262 (2017).

[12] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers and W. Burgard, "An evaluation of the RGB-D SLAM system," **In:** *Proceedings of IEEE International Conference on Robotics and Automation* (2012) pp. 1691–1696.

[13] K. Tateno, F. Tombari and N. Navab, "When 2.5D is Not Enough: Simultaneous Reconstruction, Segmentation and Recognition on Dense SLAM," **In:** *Proceedings of IEEE International Conference on Robotics and Automation* (2016) pp. 2295–2302.

[14] T. Qin, P. Li and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.* **34**, 1004–1020 (2018).

[15] R. Wang, M. Schworer and D. Cremers, "*Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras*," **In:** *Proceedings of IEEE International Conference on Computer Vision* (2017) pp. 3923–3931.

[16] C. Qian, H.J. Zhang, J. Tang, B.J. Li and H. Liu, "An orthogonal weighted occupancy likelihood map with IMU-aided laser scan matching for 2D indoor mapping," *Sensors* **19** (2019).

[17] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*. MIT (2006).

[18] C.-Y. Li, I. H. Li, Y.-H. Chien, W.-Y. Wang and C.-C. Hsu, "Improved Monte Carlo Localization with Robust Orientation Estimation based on Cloud Computing," **In:** *Proceedings of IEEE Congress on Evolutionary Computation* (2016) pp. 4522–4527.

[19] S. Zhao, J. Gu, Y. Ou, W. Zhang, J. Pu and H. Peng, "IRobot Self-Localization using EKF," **In:** *Proceedings of IEEE International Conference on Information and Automation* (2016) pp. 801–806.

[20] X. Xu, F. Pang, Y. Ran, Y. Bai, L. Zhang, Z. Tan, C. Wei and M. Luo, "An indoor mobile robot positioning algorithm based on adaptive federated Kalman Filter," *IEEE Sens. J.* **21**, 23098–23107 (2021).

[21] H. Yu, J. Wang, B. Wang, H. Han and G. Chang, "Generalized total Kalman filter algorithm of nonlinear dynamic errors-in-variables model with application on indoor mobile robot positioning," *Acta Geodaetica et Geophysica* **53**, 107–123 (2018).

[22] J. Mccormac, A. Handa, A. Davison and S. Leutenegger, "Semantic Fusion: Dense 3D Semantic Mapping with Convolutional Neural Networks," **In:** *Proceedings of 2017 IEEE International Conference on Robotics and Automation* (2017) pp. 4628–4635.

[23] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. J. Kelly, A. J. Davison, and IEEE, "SLAM Plus Plus: Simultaneous Localisation and Mapping at the Level of Objects," **In:** *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2013) pp. 1352–1359.

[24] S. Yang and S. Scherer, "CubeSLAM: Monocular 3-D Object SLAM," *IEEE Trans. Robot.* **35**, 925–938 (2019).

[25] L. Nicholson, M. Milford and N. Sunderhauf, "QuadricSLAM: Dual quadrics from object detections as landmarks in object-oriented SLAM," *IEEE Robot. Automat. Lett.* **4**, 1–8 (2019).

[26] A. Kundu, Y. Li, F. Dellaert, F. Li and J. M. Rehg, "Joint Semantic Segmentation and 3D Reconstruction from Monocular Video," **In:** *Proceedings of 13th European Conference on Computer Vision (ECCV)* (2014) pp. 703–718.

[27] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A.L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Patt. Anal. Mach. Intell.* **40**, 834–848 (2018).

[28] M. Siam, M. Gamal, M. Abdel-Razek, S. Yogamani, M. Jagersand and IEEE, "RTSEG: Real-Time Semantic Segmentation Comparative Study," **In:** *Proceedings of IEEE International Conference on Image Processing (ICIP)* (2018) pp. 1603–1607.

[29] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou and G. Cottrell, "Understanding Convolution for Semantic Segmentation," **In:** *Proceedings of IEEE Winter Conference on Applications of Computer Vision* (2018) pp. 1451–1460.

https://doi.org/10.1017/S026357472200073X Published online by Cambridge University Press

[30] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Patt. Anal. Mach. Intell.* **39**, 1137–1149 (2017).

[31] D. Bolya, C. Zhou, F. Xiao and Y. J. Lee, “YOLACT: Real-Time Instance Segmentation,” **In:** *Proceedings of 17th IEEE/CVF International Conference on Computer Vision* (2019) pp. 9156–9165.

[32] K. He, G. Gkioxari, P. Dollar and R. Girshick, “Mask R-CNN,” *IEEE Trans. Patt. Anal. Mach. Intell.* **42**, 386–397 (2020).

Cite this article: H. Kuang, Y. Li, Y. Zhang, Y. Wan and G. Ge (2022). “Research on rapid location method of mobile robot based on semantic grid map in large scene similar environment”, Robotica **40**, 4011–4030. https://doi.org/10.1017/S026357472200073X