2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA) | 978-1-6654-3265-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/AIEA53260.2021.00063

A loop closure detection method based on semantic segmentation and convolutional neural network

Hongyang Li[1st]
Changchun University of Science and Technology
Electronic Information Engineering
Changchun, China

Chengjun Tian[2nd] *
Changchun University of Science and Technology
Electronic Information Engineering
Changchun, China
* Corresponding author :1358663026@qq.com

Lequan Wang[3rd]
Changchun University of Science and Technology
Electronic Information Engineering
Changchun, China

Hongfu Lv[4th]
Changchun University of Science and Technology
Electronic Information Engineering
Changchun, China

***Abstract*-As artificial intelligence flourishes and many related technologies continue to develop, Visual Simultaneous Localization and Mapping, as the "vision" of robots, can utilize a large amount of environmental information. In addition, semantic segmentation can distinguish the background in the image from the moving objects. In recent years, convolutional neural networks have been successfully applied to image processing, and subsequently entered the field of V-SLAM research. Related researchers have tried to directly use convolutional neural networks to extract image features for loop closure detection, but the effect has not surpassed traditional methods. In order to make full use of image information, this paper proposes a loop closure detection and SLAM method based on semantic segmentation and convolutional neural network.**

Keywords-Simultaneous Localization and Mapping (SLAM); closed loop detection; Convolutional Neural Network (CNN); semantic segmentation

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) is a key functional module for robots to complete autonomous motion, that is, it is expected to realize the robot's autonomous localization and build environment maps simultaneously. However, at present, the two functions cannot be implemented synchronously, and the cooperation of the two functions is required to complete the SLAM task. The robot's SLAM function needs to rely on external sensors, mainly including lidar, camera, sonar and inertial odometer. Because the camera can obtain a wealth of information about the environment where the robot is located, the SLAM research in recent years has mainly focused on the Visual-SLAM (V-SLAM) direction. Loop closure detection is an important part of optimizing the SLAM function. Its function is to eliminate accumulated errors, correct the robot's positioning coordinates, and construct a globally consistent trajectory and map. Loop closure detection can provide the correlation between current data and all historical data, so it can provide constraints between the current frame and the historical frame [1].

Traditional V-SLAM generally uses artificial features to use the information obtained by the camera. The classic V-SLAM algorithm is Bag-of-Words (BoW). BoW contains an artificially constructed dictionary, and then according to the words contained in the image (feature descriptors) Combine the search dictionary and construct the corresponding feature vector to describe the whole image. Finally, the closed loop can be detected based on the vector difference between the image feature vectors [2]. The BoW algorithm can effectively run in various environments and has always been the mainstream algorithm in V-SLAM. Traditional local descriptors include HOG, SIFT, SURF, FREAK, ORB, BRISK, BRIEF and LIOP. Popular methods using feature point detectors and descriptors mainly include FAST, SIFT, SURF and ORB. Among them, the ORB feature points are detected by the FAST corner detector, and the orientation and rotation invariance feature descriptions are added, so that the real-time feature detection and description are more robust. The SIFT and SURF detectors include the direction of feature points and involve the histogram of the gradient calculation. In addition to the above-mentioned V-SLAM methods based on local features, there are also V-SLAM methods based on global features, such as the application of GIST and HOG [3]. However, the features mentioned in all the above methods are artificially designed, so they can only cope with limited scene changes. In addition, these artificial features can only contain some low-level information and cannot express complex spatial structure information, so they cannot handle drastic appearance changes.

In recent years, with the renaissance of deep learning, a large number of researchers have set their sights on the field of deep learning. Among them, convolutional neural networks (CNN) have been successfully applied to the field of image recognition and classification, and V-SLAM-related researchers will convolution Neural network is used in the loop closure detection module of V-SLAM, and the feature vector extracted by convolutional neural network is used to replace the traditional artificially designed feature vector [4]. This type of application has obtained better operating results. The general application method is to use the pre-trained convolutional

978-1-6654-3265-8/21/$31.00 ©2021 IEEE
DOI 10.1109/AIEA53260.2021.00063

neural network to extract the image feature vector directly to replace the traditional hand-designed feature vector. In addition, the dimensions of image feature vectors extracted by convolutional neural networks of different depths are different. Among them, although the low-dimensional feature vector of the image only contains the primary features and shallow information of the image, the low computing power required for computer processing will not burden the operation of the computer, which can improve the operating efficiency of the computer. In addition, the high-dimensional feature vector of the image contains more abstract and detailed information of the image, but the computer requires more computing power for processing, which is not conducive to the computer's understanding of the image and will reduce the operating efficiency. The high-dimensional feature vector and the low-dimensional feature vector of the image each have their own advantages and disadvantages, and they can be combined to obtain better results.

Another important technology in the field of image processing has been applied in autonomous driving and V-SLAM, namely semantic segmentation. Semantic Segmentation is to identify the content and its location in the image by searching all pixels belonging to the content in the image. Semantic segmentation generally involves three types of technical fields including image classification: recognizing image content; image detection and recognition: recognizing the location of each content of an image; image segmentation: understanding the meaning of an image. The usual semantic segmentation architecture can be defined as an encoder and a decoder [5].

Previous researchers combined convolutional neural networks into V-SLAM and tried to directly replace traditional artificial features with image features extracted by convolutional neural networks [6]. Although good results have been obtained, they have not exceeded the application effects of traditional artificial features. In addition to the problems faced by traditional V-SLAM, running a convolutional neural network requires a large amount of computing power, which will cause the system to run slowly. In order to solve these problems, this paper proposes a novel loop closure detection and SLAM method based on semantic segmentation and convolutional neural networks. Preprocessing of images through semantic segmentation can distinguish the images captured by the RGB-D camera during the robot movement into moving objects and fixed scenes, so as to perform loop closure detection more effectively.

II. PREPARATION WORK

This section will state the preparatory work before the experiment, mainly analyze the problems that will be encountered, and propose solutions for this paper.

There are two main types of problems in the loop detection process: one is false positives, also known as perceptual deviations, which are defined as different but similar scenes are judged as loops; the other is false negatives, also known as perceptual mutations, which are defined as the same scenes are Wrongly judged as non-loopback. A qualified loopback detection algorithm should try its best to overcome these two types of problems. Traditional loop detection algorithms generally use artificially designed features, which are susceptible to environmental factors and reduce the accuracy

Fig. 1 Physical map of the experimental platform robot

of loop detection. Convolutional neural networks have been widely used in image feature extraction in recent years. Studies have shown that the image features extracted by convolutional neural networks are more objective and less susceptible to environmental factors. However, general comparison algorithms will cause huge computational problems and reduce loop detection effectiveness.

In view of the above problems, this paper proposes a robot loop detection method. Through image preprocessing, key frame selection, establishment of high-dimensional feature vector database and low-dimensional feature vector database, etc., the accuracy and efficiency of loop detection are improved, and the SLAM process is eliminated. The cumulative error of the robot enhances the robustness of the robot in a complex environment.

III. MATERIALS AND METHODS

A. Experimental materials and experimental platform

The experimental hardware platform of this paper is a robot equipped with Jetson Nano processor, RGB-D depth camera and Mecanum kinematics structure. The software platform is the Ubuntu 18.04 system, the Robot Operating System (ROS) is installed in the Ubuntu system, and the ROS installation version is Melodic. The robot communicates with the PC in real time by means of local area network communication. The experimental robot is shown in Fig. 1.

B. Experimental method

This section will introduce the loop closure detection method proposed in this paper. The specific details are as follows.

1) Main process of loop closure inspection

The loop closure detection method proposed in this paper is as follows:

Step 1: Obtain the image of the robot movement process through the RGB-D camera carried by the robot, and then preprocess the image obtained by the camera;

Step 2: Obtain the low-dimensional feature vector of each frame of image after preprocessing, establish the low-

dimensional feature vector database Data(L), and use the key frame selection algorithm proposed in this paper to determine the specific number of key frames;

TABLE I. THE A AND Γ VALUES SET DURING THE EXPERIMENT AND THE CORRESPONDING LOOP CLOSURE DETECTION ACCURACY RESULTS

α	γ		
	0.85	0.9	0.95
0.70	0.7584	0.8277	0.8653
0.75	0.7962	0.8466	0.8879
0.80	0.8739	0.9218	0.9753
0.85	0.8366	0.8905	0.9187
0.90	0.8269	0.8708	0.8898

Step 3: Obtain the high-dimensional feature vector corresponding to the specific frame number of the key frame, and establish the key frame high-dimensional feature vector database Data(H);

Step 4: Obtain the high-dimensional and low-dimensional feature vectors of the current frame after preprocessing, and compare the high-dimensional feature vectors of the current frame with Data(H). When the comparison result is greater than α, determine the number of frames that may have a closed loop;

Step 5: According to the frame number range that may have a closed loop, use the low-dimensional feature vector of the current frame to compare with the corresponding frame of the frame number range that may have a closed loop in Data(L). When the comparison result is greater than the preset confidence parameter threshold γ, it is determined to be formed closed loop.

The experimental test values of α and γ are shown in Table 1 for the accuracy of loop closure detection after the test.

2) Image preprocessing process

The specific method of image preprocessing mentioned in the above algorithm is as follows:

Step 1: After the continuous video information is processed by the computer, the image specifications are unified;

Step 2: Use the semantic segmentation algorithm to process the image after the unified specification;

Step 3: Obtain the background part and the dynamic object part in the scene contained in the image;

Step 4: Keep the background part contained in the image;

Step 5: Finally, input the processed image to the subsequent steps.

3) Key frame selection algorithm flow

The specific steps of the key frame selection method proposed in this paper are as follows:

Step 1: The video image information after image preprocessing is input into the key frame selection network;

Step 2: The initial frame must be selected as the key frame, the second frame is set as the reference frame, and the subsequent frames of the reference frame are the frames to be measured;

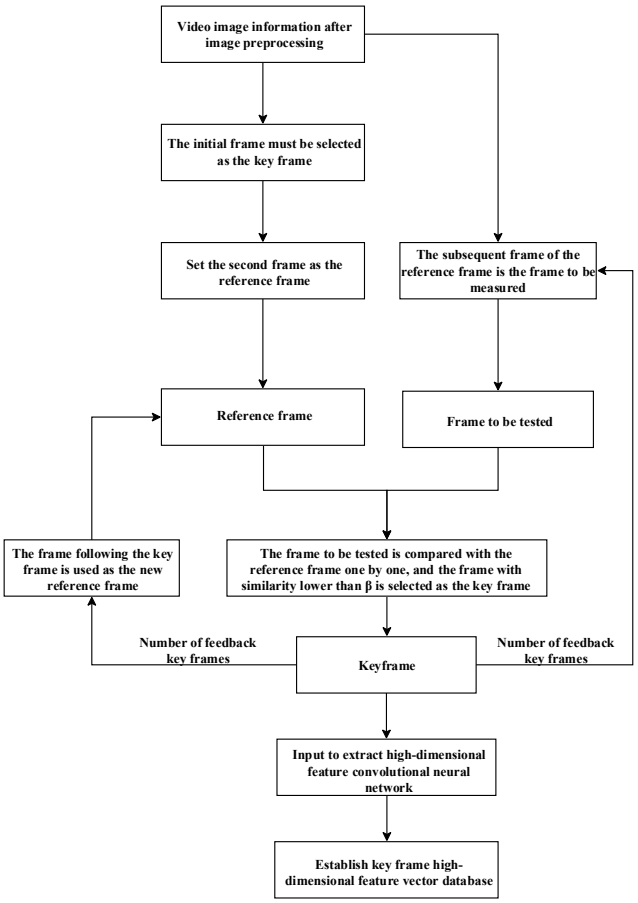

Fig. 2 Key frame selection algorithm flowchart

Step 3: Compare the frame to be tested and the reference frame one by one, select the frame with a similarity lower than β as the key frame, feedback the number of key frame frames, and use the subsequent frame of the key frame as the new reference frame, and the subsequent frame of the reference frame is the frame to be tested, repeat the aforementioned comparison method to continue to select key frames;

Step 4: Input the selected key frames into the convolutional neural network for extracting high-dimensional features;

Step 5: Establish the key frame high-dimensional feature vector database Data(H).

In the experiment process of this paper, the β value is preset to 0.6. The specific flow chart of the key frame selection algorithm is shown in Fig. 2.

C. Actually test the SLAM effect

This section mainly elaborates the actual SLAM effect, in order to show the effectiveness of the method in this paper, a practical test is made.

Loop closure detection is an important module in the V-SLAM process, and its effect can be directly reflected in whether the cumulative error can be eliminated and whether a

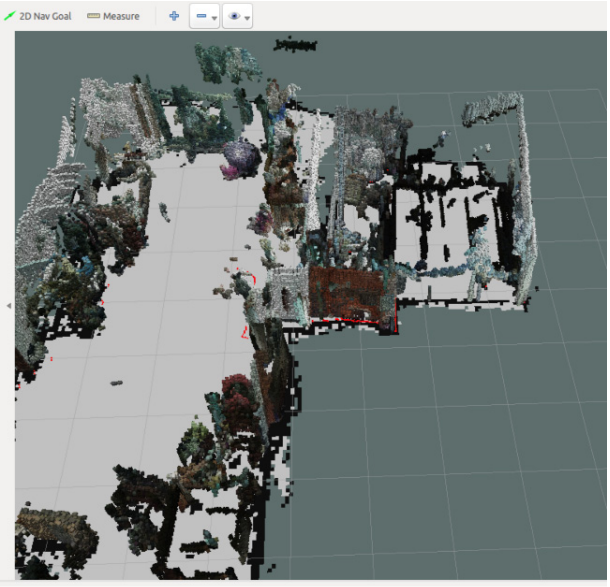

Fig. 3 Visual SLAM mapping renderings

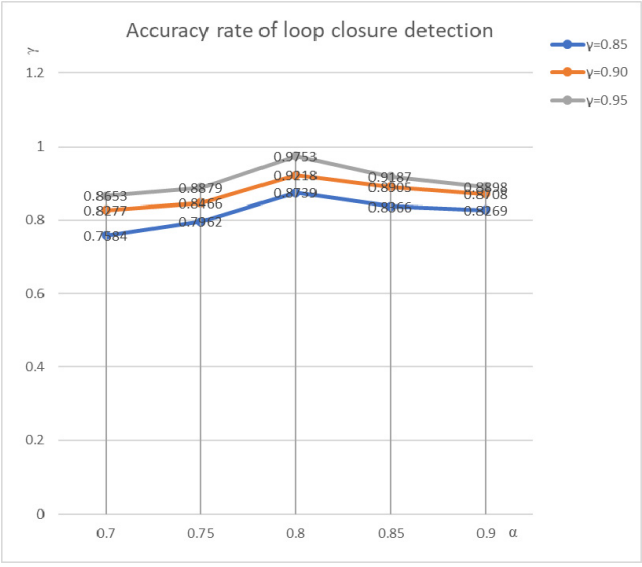

Fig. 4 Experimental result graph

globally consistent map can be established. In order to test the effectiveness of the method in this paper, real experiments have been carried out and the following effects have been obtained. As shown in Fig. 3.

IV. Results and Discussion

Draw a data diagram based on the experimental data, as shown in Fig. 4. From the data, it is not difficult to find that the accuracy of closed loop detection is relatively highest when α is 0.8 and γ is 0.95, which also shows that the algorithm proposed in this paper has good robustness and reliability. Compared with the traditional method that directly uses convolutional neural network to extract image features for loop closure detection, the method proposed in this paper has improved, at the same time, it improves the accuracy and efficiency of loop closure detection, and makes better use of the environmental information contained in the image. In addition, the combined use of semantic segmentation and convolutional neural networks also significantly enhances the accuracy of loop closure detection.

In addition, the β in the experiment process of this paper is preset to 0.6, and no comparison experiments with other values have been made. The purpose of fixing the value of β is to prevent the step of selecting key frames from affecting the main experimental objectives. Experimental results show that when β value is 0.6, α is 0.8 and γ is 0.95, the accuracy of loop closure detection is relatively highest, and this algorithm achieves the expected purpose.

V. Conclusions

This paper confirms the effectiveness of this method through experiments. Compared with the traditional method of directly applying convolutional neural network, it improves the accuracy and efficiency of loop closure detection. However, in the course of the experiment, we also found the shortcomings of this method's insufficient coping ability during fast movements, and it needs to be improved. We will continue to overcome this shortcoming in the future.

From the discussion in this article, we can see that the application prospects of convolutional neural networks are still very impressive. Convolutional neural networks have huge application potential in V-SLAM. In the foreseeable future, convolutional neural networks are likely to completely replace and surpass the traditional V-SLAM algorithm. However, due to hardware limitations and many deficiencies in the current algorithms, further improvement and research by related researchers are needed.

References

[1] Gao, X. (2017) Unsupervised Learning to Detect Loops Using Deep Neural Networks for Visual SLAM System. Autonomous Robots,1: 1-18.

[2] Bai, D. (2018) CNN Feature Boosted SeqSLAM for Real-Time Loop Closure Chinese Journal of Electronics, 1-18.

[3] Liu, Q. (2019) Loop closure detection using CNN words, Intelligent Service Robotics 1: 303-318.

[4] Wang, Y. (2020) Robust Loop Closure Detection Integrating Visual–Spatial–Semantic Information via Topological Graphs and CNN Features. Remote Sensing 23.

[5] Wang, Z. (2019) Manifold Regularization Graph Structure Auto-Encoder to Detect Loop Closure for Visual SLAM. IEEE Access 1: 59524-59538.

[6] Bampis, L. (2018) Fast Loop-Closure Detection Using Visual-Word-Vectors from Image Sequences:. The International Journal of Robotics Research 1: 62-82.