Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Real-time indoor semantic map construction combined with the lightweight object detection network

To cite this article: Xumin Gao *et al* 2020 *J. Phys.: Conf. Ser.* **1651** 012142

View the article online for updates and enhancements.

Real-time indoor semantic map construction combined with the lightweight object detection network

Xumin Gao[1a*], Lin Jiang[2], Xingyu Guang[3] and Wenkang Nie[4]

[1]Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, Hubei, CHXX0138, China

[2]Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, Hubei, CHXX0138, China

[3]Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, Hubei, CHXX0138, China

[4]Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, Hubei, CHXX0138, China

[a]gaoxumin@mcfly.com.cn [*]Corresponding author's e-mail: 15071460998@163.com

Abstract. Aiming at the lack of semantic information of indoor objects in the grid map commonly used in the indoor mobile robot, this paper proposes a real-time indoor semantic map construction method combined with the lightweight object detection network: firstly, we proposes a lightweight object detection network which is called S-SSD (ShuffleNet-SSD) by improving the SSD (Single Shot MultiBox Detector) network, it can be used to extract the semantic information of indoor objects in real time; then the semantic information of indoor objects is transformed into the grid map which is created by mobile robot, so the indoor semantic map is constructed. The effectiveness and superiority of S-SSD in the indoor object detection of mobile robot are verified by the comparative experiment, and the effectiveness of the real-time indoor semantic map construction proposed in this paper is verified by the experiment of semantic map construction.

1. Introduction

Map creation is one of the key technologies of indoor mobile robot, it provides a basic guarantee for accurate localization and correct navigation [1-2]. At present, the main map types used by indoor mobile robot include grid map, feature map and topology map [3-5]. Grid map uses grids of the same size to describe the environment, although grid map is easy to create, it is inefficient for path planning [6]. Feature map uses geometric features such as lines, arcs to describe the environment, which can only adapt to very simple scenes [7]. Topology map uses nodes and line segments to describe the environment, but it is difficult to build a map in the large environment [8]. The common disadvantage of the above three types of map is lack of the semantic information of indoor objects, which makes mobile robot unable to complete some advanced tasks.

In recent years, with the development of deep learning, some classic networks such as R-CNN [9], Fast R-CNN [10], Faster R-CNN [11], YOLO [12], SSD [13] have been applied to object detection, it provides the possibility for mobile robot to detect indoor objects. Adhikari et al. [14] annotated the dataset of indoor objects by themselves, and used Fast R-CNN network to realize the detection of indoor objects. Tang et al. [15] applied SSD network to the task of indoor object detection. O'Keeffe et al. [16] pruned the YOLO model and transplanted it into the embedded system, which was applied to indoor object detection of mobile robot.

Therefore, this paper proposes a real-time indoor semantic map construction method combined with the lightweight object detection network: in order to consider both detection accuracy and speed, we propose a lightweight object detection network which is called S-SSD , it can be used to detect indoor objects in real time so that the semantic information of indoor objects can be extracted in real time for indoor robot. On this basis, the semantic information of indoor objects is transformed into the grid map, and the indoor semantic map is constructed.

2. Method

2.1. Network architecture of S-SSD

At present, the object detection networks which are commonly used include R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD. Considering the detection accuracy and speed, the SSD network is relatively best in detection performance, because SSD combines the grid regression idea of YOLO and the anchor mechanism of Faster R-CNN, and adds multi-scale feature detection, so it has the advantages of high detection accuracy and fast detection speed. SSD can meet the general tasks of object detection, but considering the requirement of real time, it can not meet the indoor object detection of mobile robot, because the indoor mobile robot is always in the moving state, it needs more faster detection speed than the general static object detection. To solve this problem, this paper proposes a lightweight object detection network which is called S-SSD by improving the SSD network, and it is applied to indoor object detection of mobile robot. The network architecture of S-SSD is shown in figure 1:

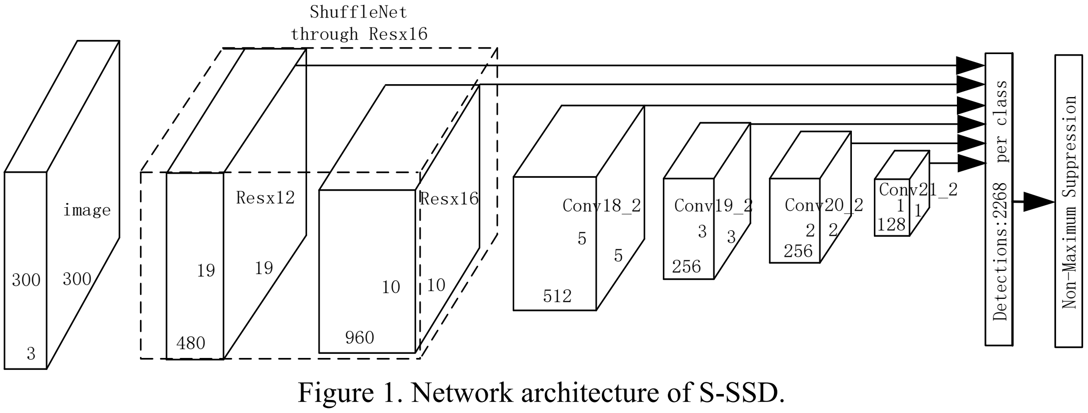


Figure 1. Network architecture of S-SSD.

What is different from the SSD network is that the S-SSD lightweight network uses the ShuffleNet [17] network with the removal of the last fully connected layer as the basic feature extraction network, while retaining the characteristics of multi-scale feature detection of SSD network. It uses Resx12, Resx16 in the front-end of network and multiple feature extraction layers in the back-end of network to realize multi-scale feature detection. Because ShuffleNet adopts the design ideas of depthwise separable convolution, pointwise group convolution and channel shuffle, the prediction time of the network is greatly reduced. Therefore, S-SSD combines the advantages of high detection accuracy of SSD and short prediction time of ShuffleNet, which can ensure the accuracy and speed of indoor object detection of mobile robot.

2.2. Indoor semantic map construction based on grid map

Firstly, we use Kinect v2 of mobile robot to get the RGB image of indoor scene, and use S-SSD lightweight network to detect indoor objects. At the same time, the depth image of the corresponding scene is obtained by the depth camera of Kinect v2. Through the registration parameters between the RGB image and the depth image obtained in the previous stage, we can calculate the spatial coordinates of indoor objects detected by S-SSD in the coordinate system of depth camera of Kinect v2. After that, we transform the spatial coordinates of indoor objects into the grid map created by mobile robot, so it completes the indoor semantic map construction. In this process, the main coordinate transformations which are involved are the following three aspects:

Suppose the coordinate system of depth camera of Kinect v2 is $O_C - X_C Y_C Z_C$, the coordinate system of mobile robot is $O_R - X_R Y_R Z_R$, the coordinate system of real world is $O_W - X_W Y_W Z_W$, the coordinate system of grid map is $O_G - X_G Y_G Z_G$

(1) The coordinate transformation between depth camera of Kinect v2 and mobile robot is as in equation (1):

$$\begin{bmatrix} X_R \\ Y_R \\ Z_R \\ 1 \end{bmatrix} = \begin{bmatrix} R_1 & T_1 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \qquad (1)$$

In equation (1), R_1 and T_1 are rotation matrix and translation matrix between the coordinate system of depth camera of Kinect v2 and the coordinate system of mobile robot, and the specific parameters can be obtained through calibration.

(2) The coordinate transformation between mobile robot and real world is as in equation (2):

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{bmatrix} R_2 & T_2 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_R \\ Y_R \\ Z_R \\ 1 \end{bmatrix} \qquad (2)$$

In equation (2), R_2 and T_2 are rotation matrix and translation matrix between the coordinate system of mobile robot and the coordinate system of real world, and the specific parameters can be obtained through calibration.

(3) The coordinate transformation between the real world and the grid map is as in equation (3):

$$\begin{cases} x_g = ceil(\frac{x}{r}) \\ y_g = ceil(\frac{y}{r}) \end{cases} \qquad (3)$$

In equation (3), (x, y) represents the coordinate value of a point in the coordinate system of real world, (x_g, y_g) represents the coordinate value of the point in the coordinate system of grid map, r is the length value of the grid cell in the grid map corresponding to the length value in the real world.

3. Experiment

3.1. The platform of indoor mobile robot

The experimental platform of indoor mobile robot used in this paper, is as shown in figure 2, it mainly includes control system, laser radar, Kinect v2. The GPU of control system is GXT 1050ti. The laser radar is used to create grid map, and Kinect v2 is used to detect indoor objects and obtain the spatial coordinates of corresponding indoor objects.

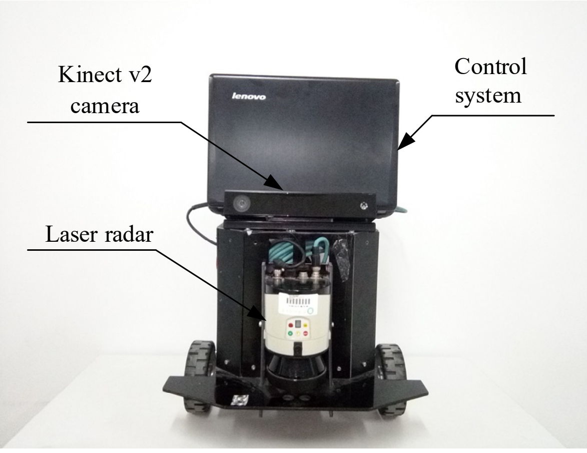


Figure 2. The experimental platform of indoor mobile robot.

3.2. Comparative experiment of indoor object detection

We screened out 12000 images containing indoor objects from Indoor_CVPR09 [18], NYU2 [19], and SUN [20], and labeled the nine common indoor objects (bed, refrigerator, washer, sofa, cabinet, chair, table, closestool, door) to be as the dataset for indoor object detection in this paper.

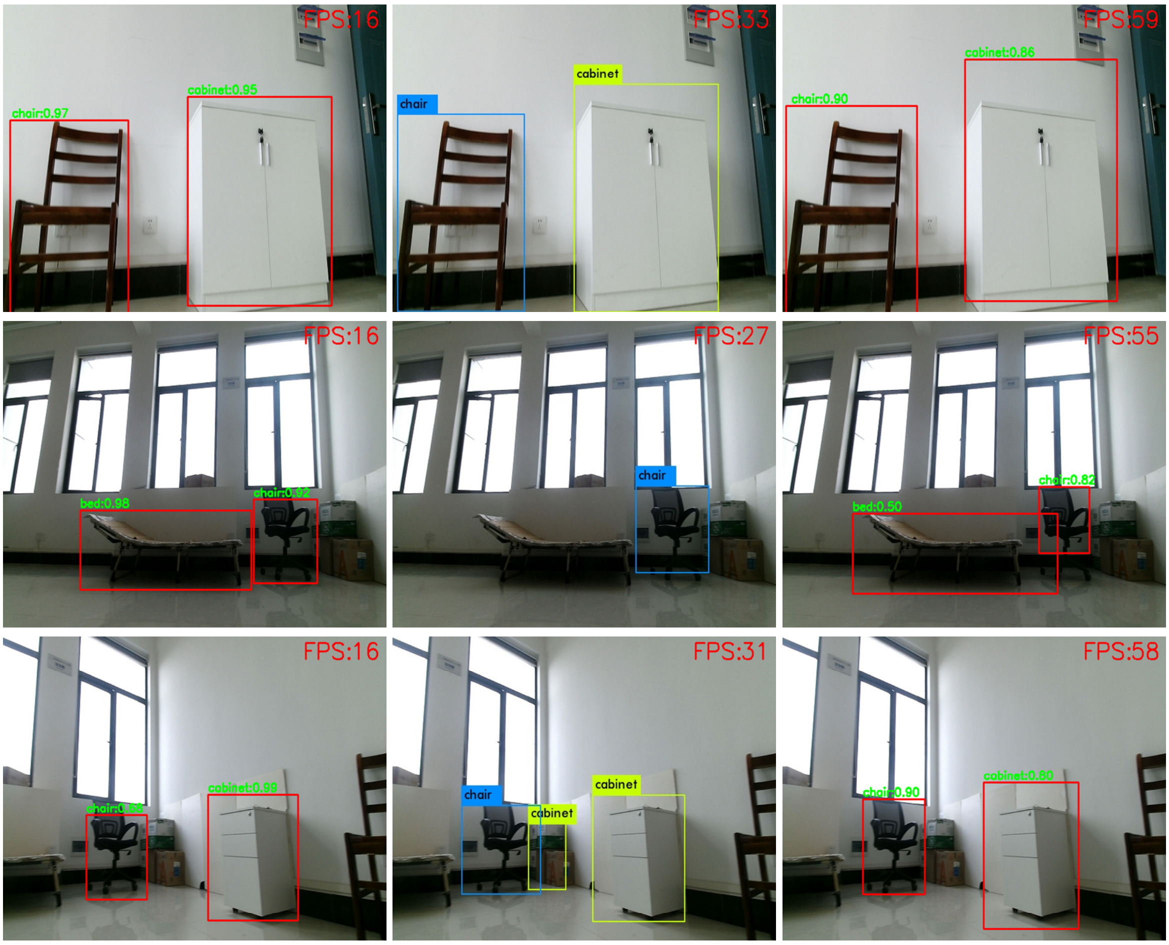
In order to verify the superiority of the S-SSD lightweight network proposed in this paper, we use SSD network, Tiny-yolo lightweight network which is commonly used to detection task of indoor mobile robot and the S-SSD lightweight network proposed in this paper to train the dataset under the same conditions. Then, we compare mAP (mean Average Precision) and mFPS (mean Frames Per Second) of different detection networks on the test set. The comparison results are shown in table 1:

Table 1. The comparison results of different detection networks on test set

Network	mAP(%)	mFPS
SSD	78.8	16
Tiny-yolo	67.6	31
S-SSD	76.7	58

It can be seen from table 1 that although SSD has the highest detection accuracy, the detection speed of mFPS is only 16, which can not meet the real-time requirement of object detection of indoor mobile robot. Tiny-yolo is very low in detection accuracy which is only 67.6%, it can not meet the accuracy requirement of object detection of indoor mobile robot. Compared with SSD, the detection accuracy of S-SSD is only slightly reduced by 2.1%, and the detection speed of mFPS can achieve 58. Thus considering the detection accuracy and speed, S-SSD lightweight network proposed in this paper is better than SSD network and Tiny-yolo lightweight network.

In order to more intuitively verify the effectiveness and superiority of the S-SSD lightweight network in the real-time object detection of indoor mobile robot, we use different detection networks to carry out the comparative experiments of real-time indoor object detection in real indoor scene, and we intercept detection results of three different time to compare, the comparison results are shown in figure 3:



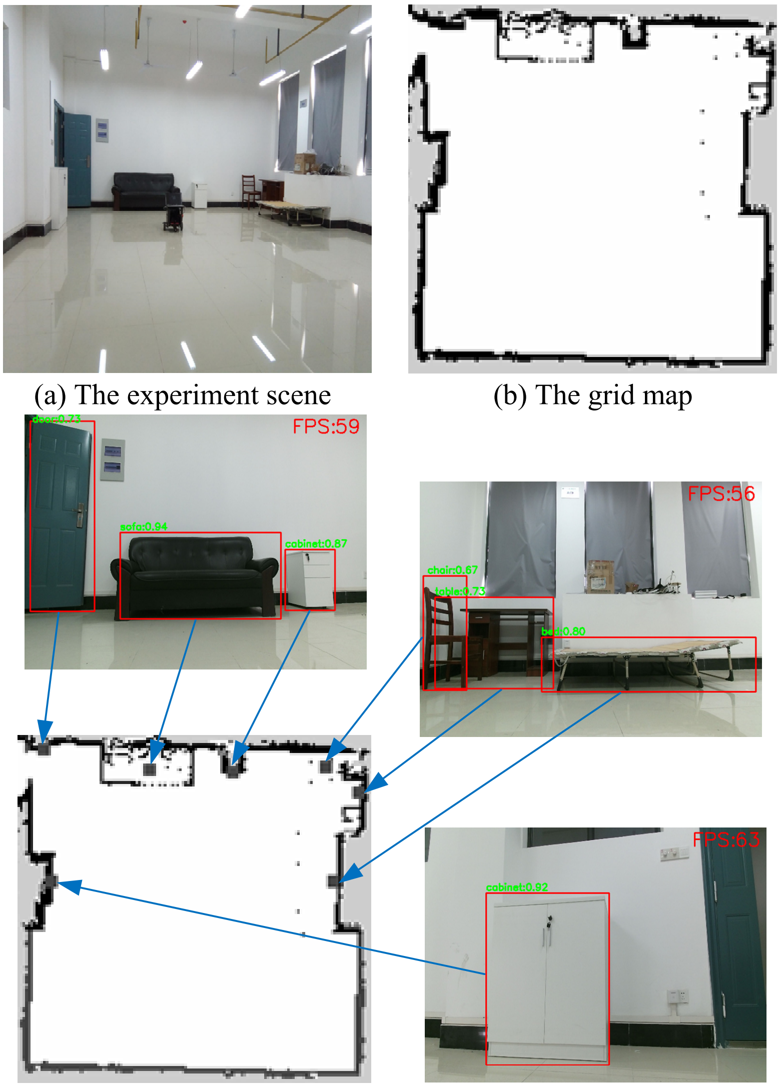
(a) SSD (b) Tiny-yolo (c) S-SSD

Figure 3. The comparison results of real-time indoor object detection

It can be seen from figure 3, in terms of detection accuracy, SSD network and S-SSD lightweight network can detect all indoor objects in scene, and Tiny-yolo has problems of missing detection and false detection. In terms of detection speed, SSD can not meet the real-time requirement of indoor object detection of mobile robot, while Tiny-yolo and S-SSD can meet it. So considering detection accuracy and speed, S-SSD can meet the requirement of detection accuracy and speed of indoor object detection of mobile robot.

3.3. The experiment of real-time indoor semantic map construction

In order to verify the effectiveness and superiority of the real-time semantic map construction proposed in this paper, we use the indoor mobile robot platform to carry out the experiment of real-time semantic map construction. The experimental results are shown in figure 4. The figure 4(a) is the indoor experiment scene; The figure 4(b) is the grid map created by mobile robot. The figure 4(c) is the result of real-time semantic map construction based on grid map, we use the gray rectangular grid with the size of 3x3 to describe the average location of objects which are dedected by mobile robot in the semantic map.



(a) The experiment scene (b) The grid map

(c) The semantic map based on grid map

Figure 4. The result of real-time semantic map construction by mobile robot

It can be seen from figure 4(b) that the grid map lacks the description of indoor objects. Compared with the original grid map, the semantic map in figure 4(c) adds the semantic information of indoor objects, which makes the indoor mobile robot have a ability to complete more complex tasks while having the basic location and navigation functions.

4. Conclusion

In this paper, a real-time indoor semantic map construction method combined with the lightweight object detection network is proposed. Firstly, the real-time detection of indoor objects is realized through the lightweight object detection network which is called S-SSD, and then the spatial coordinates of objects are transformed into the grid map created by the mobile robot, so as to complete the construction of indoor semantic map. Finally, the effectiveness and superiority of the real-time indoor semantic map construction are verified by experiment.

Acknowledgments

This research was supported by National Natural Science Foundation of China (NO. 51505347) and Natural Science Foundation of Hubei Province of China (NO. 2018CFB626).

References

[1] Junchao, L., Zhijun, C., Xiaochao, F. (2019) Research on accurate matching of target positioning for indoor mobile robot. Modern Electronics Technique. 6(31): 172-183

[2] Biswas, J., Veloso, M. (2011) Depth camera based indoor mobile robot localization and navigation. Proceedings IEEE International Conference on Robotics & Automation. 32:1697-1702.

[3] Horvath, E., Pozna, C.R. (2017) Probabilistic occupancy grid map building for Neobotix MP500 robot. Workshop on Positioning. 32: 121-133.

[4] Wang, H.M, Hou, Z.G., Ma, J. (2007) Sonar Feature Map Building for a Mobile Robot. IEEE International Conference on Robotics & Automation. 15: 156-164.

[5] Schwertfeger, S., Birk, A. (2016) Map evaluation using matched topology graphs. Autonomous Robots. 40(5): 761-787.

[6] Xiao, Y., Zhang, C., Luo, J. (2020). Integrating the radiation source position into a grid map of the environment using a mobile robot. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 45:976-987.

[7] Wen, Y., Lian, P., Hui, T. (2011). Mobile robot simultaneous localization and mapping based on feature map. Computer Measurement & Control. 111: 69-76

[8] Gunathillake, A, Huang, H.L., Savkin, A.V. (2019) Sensor-network-based navigation of a mobile robot for extremum seeking using a topology map. IEEE transactions on industrial informatics. 12(11):110-123.

[9] Girshick, R., Donahue J., Darrell, T.,et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE Conference on Computer Vision & Pattern Recognition. 19:136-145.

[10] Girshick, R. (2016) Fast r-cnn. 2015 IEEE International Conference on Computer Vision (ICCV). 31:104-112.

[11] Ren, S., He, K., Girshick, R. (2017) Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence. 37(4): 137-149.

[12] Redmon, J., Divvala, S., Girshick, R. (2015) You only look once: unified, real-time object detection. 76: 192-205

[13] Liu, W., Anguelov, D., Erhan, D. (2016) SSD: Single Shot MultiBox Detector. European Conference on Computer Vision. 33: 95-109.

[14] Adhikari, B., Peltomaki, J., Puura, J. (2018) Faster Bounding Box Annotation for Object Detection in Indoor Scenes. 2018 7th European Workshop on Visual Information Processing (EUVIP). 45: 33-45

[15] Tang, C., Ling, Y.S, Zheng, K.D. (2018) Object detection method of multi view SSD based on deep learning. Infrared and laser engineering. 047 (001): 290-298

[16] O'Keeffe, S., Villing, R. (2018) Evaluating pruned object detection networks for real-time robot vision. In: IEEE International Conference on Autonomous Robot Systems & Competitions. Torres Vedras. pp. 91-96.

[17] Zhang, X., Zhou, X., Lin, M. (2017) Shufflenet: an extremely efficient convolutional neural network for mobile devices. arXiv preprint arXiv: 1707.01083v2.

[18] Quattoni, A., Torralba, A. (2009) Recognizing indoor scenes. In: IEEE Conference on Computer Vision & Pattern Recognition. Miami. pp. 751-765.

[19] Silberman, N., Hoiem, D., Kohli, P. (2012) Indoor Segmentation and Support Inference from RGBD Images. In: European Conference on Computer Vision. Heidelberg. pp. 171-185.

[20] Xiao, J., Hays, J., Ehinger, K. A. (2010) SUN database: Large-scale scene recognition from abbey to zoo. In: Computer Vision & Pattern Recognition. Heidelberg. San Francisco. pp. 97-109.