Research on Guideline Navigation System Based on Deep Learning

Shouqun Ming[1]
Inner Mongolia University of Technology
Institute of aviation
Hohhot Inner Mongolia,China
511663431@qq.com

Lili Wang[2]
Inner Mongolia University of Technology
Engineering Training Teaching Department
Hohhot Inner Mongolia,China
793767319@qq.com

Zhimin Hong[3]
Inner Mongolia University of Technology
College of Science
Hohhot Inner Mongolia, China
1219078977@qq.com

Lele Kang[4]
Inner Mongolia University of Technology
College of Science
Hohhot Inner Mongolia,China
751940750@qq.com

Bowen Zhuang[5]
Inner Mongolia University of Technology
Institute of Chemical Technology
Hohhot Inner Mongolia, China
2298464072@qq.com

2021 4th International Conference on Intelligent Robotics and Control Engineering (IRCE) | 978-1-6654-1348-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IRCE53649.2021.9570850

***Abstract*—In order to solve the problem of the environmental information that the vision impairment cannot perceive itself due to its own vision defects, a navigation system based on Faster RCNN depth learning is proposed. The navigation system utilizes the improved RESNET50 for the backbone network training data, and the visual sensor recognizes the environmental target, the identified blind, obstacles, etc. will be admitted to the thresholded thresholded threshold, the determination position, follow the blind trajectory and avoid the obstacle. Complete the path planning of the four-wheeled mobile robot to the outdoor environment. Through the ROS robot operating system, build an experimental platform, verify the improved algorithm, effectively prove to meet the navigation needs of the vision impairment.**

Keywords—Deep Learning, Navigation, Image Processing, Outdoor robot, Vision barrier

I. INTRODUCTION

The number of global vision impairrs is growing, and the navigation problem out of vision impairrs needs to be resolved. Through the literature [1], it is shown that vision impairrs need information in the outdoor: (1) Environmental factors, pedestrians, vehicles, rugged roads, etc .; (2) traffic information: blind road, traffic light information, etc. In order to replace the above information, the literature [2] proposes a STM32-based intelligent bribined stick system that uses traditional ultrasonic sensors to avoid impairment to achieve guide, this system is convenient, but ultrasonic sensors are unable to individual obstacles Identification, and only 2m detection range is not conducive to the actual application. In order to expand the detection range, weavier detection information, literature [3] proposed a wearable guide robot system based on bicomputer camera, this system utilizes a double visual positioning, the accuracy is higher than the ultrasonic sensor, but the algorithm structure is complex, It is necessary to suppress noise and voids during the calculation.

In order to solve the above problems, the paper proposes a FasterRCNN neural network model navigation system based on improved RESNET50 for backbone network. The system advantage is that there is no need to extract the specified feature, which is the training model through the COCO data set. Compared to the literature [2] ultrasonic sensor guide system, the system increases the detection range of 5m in front of the detection, and the amount of information is extremely rich, and the detection is continuous (multiple systems can be set separately), To achieve multi-task detection effect (simultaneously identify blind roads, obstacles, traffic signs, etc.), relative to the literature [3] double visual positioning system, the system algorithm is simple, proves that the system has effectively reduced algorithm complexity, improvement The guidelines and robustness of the guide robot navigation performance, which provides an effective guarantee for vision impairrs.

Fund Projects: National Natural Science Foundation of China (81860605); Inner Mongolia University of Technology Innovation Experimental Project for College Students (2020303005).

II. NAVIGATION SYSTEM RESEARCH PRINCIPLE

The robot carried by the research system of the thesis mainly includes two parts: hardware equipment and navigation system algorithm. The main movement method is four-wheeled movement [4-6]. The four-wheeled movement method is relatively simple and improves the response speed[7-8]. The software algorithm mainly uses the improved ResNet50 as the backbone network Faster RCNN neural network model.

The hardware equipment is divided into two parts. One part is a body equipped with vision sensors, a mini computer (CPU is E3, equipped with GTX1050TI 4G), batteries and other components, and the other part is a control lever for guidance from robots to visually impaired people. Equipped with a switch button, the visually impaired can turn off the robot navigation system algorithm as needed, and combine the visual algorithm with the voice module algorithm. Since visual sensors are mainly used to identify blind roads and obstacles, visual algorithms are the most important and difficult.

The system research steps are shown in the flowchart of the navigation system in Figure 1. First, input the collected images into the Faster RCNN neural network model for training to obtain training data. Then write the training data into the onboard computer, and the four-wheeled mobile robot will navigate.

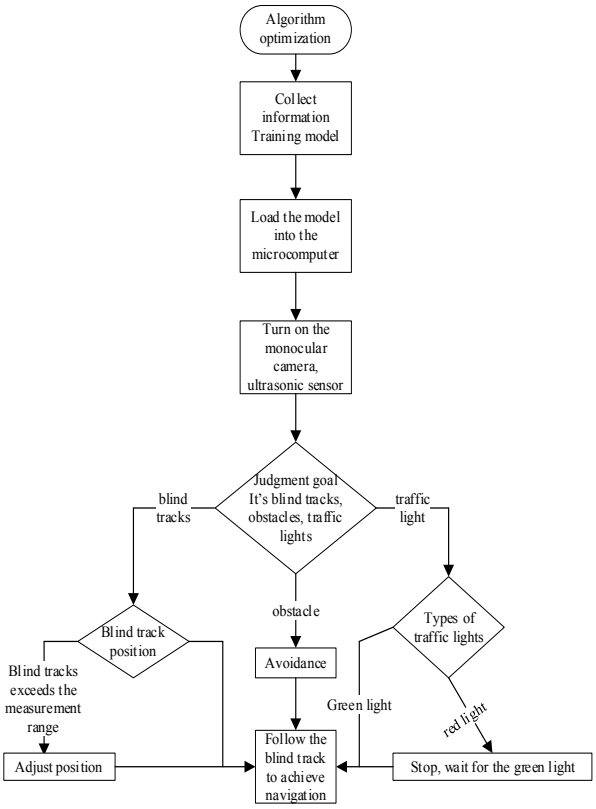

Fig. 1. Navigation system flow chart

III. FASTER RCNN PRINCIPLE

Faster RCNN is improved by RCNN and fast RCNN[9-11]. In target detection, Faster RCNN abandons tradition The method of Selective Search[12-15], directly apply RPN to calculate Candidate frame, so that the calculation rate of target recognition has been greatly increased Improve [16]. The network used in this article is based on the original Faster RCNN network improves the backbone network, which includes three levels:Improved RESNET50, RPN, Fast RCNN. Network flow The process diagram is shown in Figure 2.

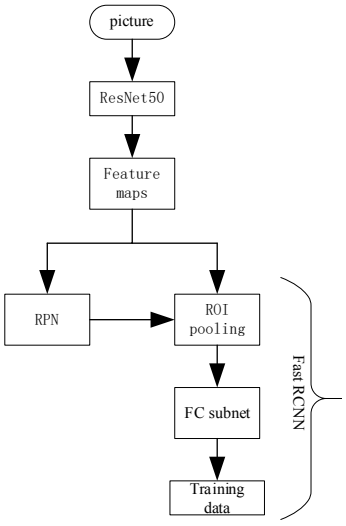

Fig. 2. Network flow chart

A. Improved RESNET50 Network

With the rapid development of deep learning, the sparse connection structure classification model (Google Net), a large-scale image recognition depth convolution network (VGG), the deep residual network (RESNET)[17-20], etc. Insensus learning classic image detection model, due to network degradation and The existence of problems such as gradient disappearance, resulting in difficult to use increased networking methods to increase training accuracy [21-24]. Therefore, RESNET50 [25] uses a cross-layer connection method to solve the above problems. It is passed directly to a deeper network to build and effectively training data using a cross-layer connection.

There are 49 layers of convolution and single layer fully connected in RESNET50[26], of which CONV2~CONV5 contain 3, 4, 6, and 3 residual blocks respectively, and each residual block contains 3 layers of convolution Product[27], but the 3*3max pooling in CONV1 and CONV2 is combined into a 1-layer convolution, and we get: a convolutional layer[28]. RESNET50 contains two different types of mapping[29], one type is type mapping, as shown in the right poly line in the structure of a residual unit in Figure 3, and the other type is residual mapping, its mapping expression:

$$H(x) = F(x) + x \tag{1}$$

$$F(x) = (\omega_3 \delta(\omega_2 \delta(\omega_1 x))) \tag{2}$$

$$H(x) = x \tag{3}$$

Among them: refers to the output of the residual network, refers to the output after the convolution operation, is the convolution operation, and represents the activation function. For the above formula, as long as the value is always equal to 0, an identity mapping function can be formed as equation (3). For this paper, only when the residual block is not less than two layers can it be effective[30].

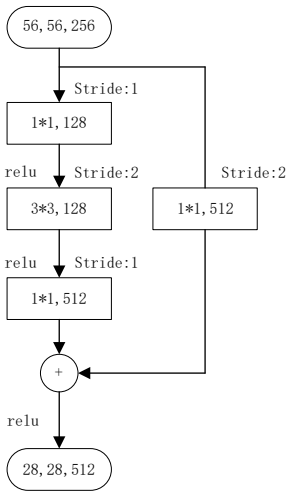

Fig. 3. A residual unit structure

Compared to the research, traffic lights, blind, obstacles, the characteristics of the traffic light, blind, and in the case of ensuring their accuracy, it needs to be treated in real time. Therefore, the 7 * 7 convolution layer of the RESNET50 first layer network is changed to 3 3 * 3 super positional layer. Single convolution layer Feature Map is calculated as follows:

$$\frac{N-i}{1} + 1 = Z \tag{4}$$

where: N is the size of the image size, i is the size of the FEATUREMAP outputted, Z is the size of the Feature Map output. A 7 * 7 and three 3 * 3 convolution layers, the final resulting of the Feature Map size is equal. The improved advantage is to use a number of nonlinear activation functions to ensure accurate accuracy of its identification, and effectively reduce the amount of various parameters, and set the first volume of the product input and output. The size is x, after improved The number of network parameters is 27 * x * x; reduces (49 * x2-27 * x2) = 22 * x2 parameters. Therefore, it is convenient to deploy to the mobile terminal to meet real-time processing requirements, easy to train, effectively improve the performance of the model [30-31].

B. Region Proposal Networks(RPN)

The RPN network flow chart is shown in Figure 4. First, RESNET50 is used to extract features of the collected image information such as blind roads to obtain feature maps. Configure the appropriate sampling multiple, and select the feature layer with the configured sampling multiple as the detection layer. A series of anchor frames of different sizes and different proportions are preset according to the selected detection layer. Then, a limited number of candidate regions are obtained by second classification and regression of the preset anchor frame. The combined function of the two-class loss function and the regression loss function used in the previous period is called the image loss function[32]. Its definition is L_{union} (5):

$$L(\{t_i\}+\{p_i\}) = \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_i^*) + \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \tag{5}$$

Among them: i represents the anchor point index in this batch of processing; p_i represents the predicted probability of taking an i as the object; when the true label $p_i^* = 1$, it means that the i anchor point is a positive sample, $p_i^* = 0$ It means that the i anchor point is a negative sample; t_i means the vector of 4 parameterized coordinates of the preset bounding box; t_i^* refers to the deviation between the border of the candidate area and the real target border. L_{cls} represents the two-category loss, that is, the log loss of the two classes; L_{reg} represents the regression loss; N_{cls} and N_{reg} represent the normalization parameters of the classification loss and the regression loss, respectively; λ represents the balance weight. The loss function of formula (6):

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \tag{6}$$

where R is the SmoothL1 robust function in[33].

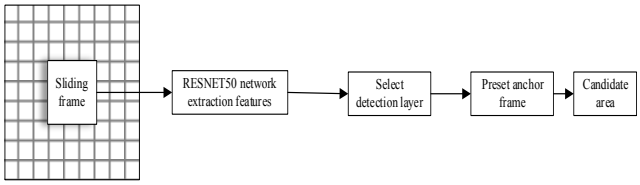

Fig. 4. RPN network flow chart

C. Fast RCNN Principle

Fast RCNN mainly includes three steps: (1) Pass through the matching of candidate regions in the ROI Pooling layer[33] (positive sample: best match or $IOU \geq \theta_+$; negative sample: $IOU < \theta_-$; ignore sample: non-positive and non-negative). Go through the region again According to the sampling multiple, the candidate area is mapped to the feature, and then the candidate area is equally divided into sub-areas of the specified size (if the position of the sub-region is not an integer, it is rounded according to the rule of rounding), and finally in each specified Take the maximum value among the sub-regions of the size to obtain the feature of the specified size. (2) Input the features of the specified size into the FC sub-network, and output the required features. (3) Use the Soft max function to get the classification label and the SmoothL1 function to get the return true value.

Soft max loss function[34-37]:

$$f_{y_i} = W_{y_i}^T x_i \tag{7}$$

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \tag{8}$$

where: N is the number of training data, f_j represents the j element of the vector of class score f ($j \in 1, K$], K is the number of classes), y_i represents the label, and x_i is the i input feature. f_{y_i} is expressed as the product of the i column of W and the i input feature.

IV. Analysis of Training Results

The platform system used for training data is ubuntu16.04, and the computer used for training data is configured with GPU: NVIDIA Quadro RTX 5000; floating-point operations per second: 10.89 TFLOPS; graphics card memory: 16 GB. The training data is the collected 282 images, as shown in Fig. 5: Examples of obstacles, traffic lights (stop/go), and blind road samples. The initial learning rate parameter is set to 0.0003.

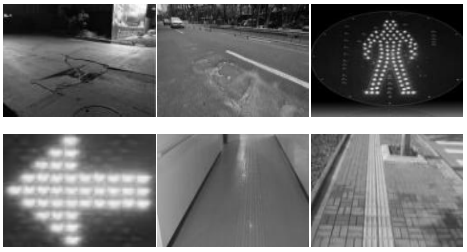

Fig. 5. Examples of obstacles, traffic lights (stop/go), and blind road samples

When the image set is used for training in the network, the initial learning rate set in the '.config' file is 0.0003, because the experiment found that the learning rate decays to 0.000003 when the step is greater than 900,000, and the learning rate is 0.00003 when the step is less than or equal to 900,000. From the gradient update graph in Figure 6, it can be found that when the learning rate is set to 0.0003 and batch_size=1 (one sample is trained each time), about 7 batches are trained per second (when training 10000 batches, the test is suspended, so the figure above There is a recess in the middle). Therefore, under the high learning rate at this time, there is better convergence and Faster training speed. In this experiment, a total of 20,000 batches were trained, about 71 rounds. You can get the network loss diagram as shown in Figure 7:

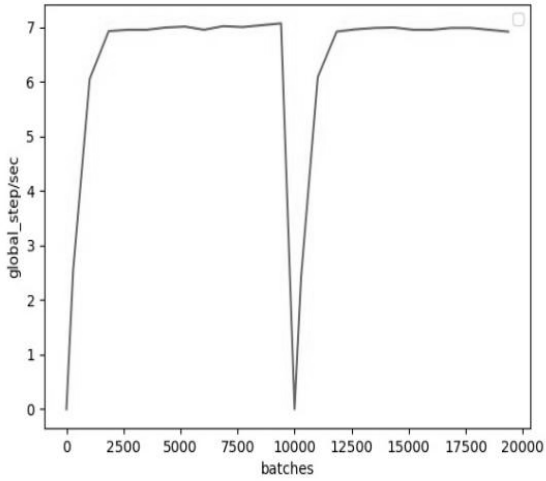

Fig. 6. Gradient update graph

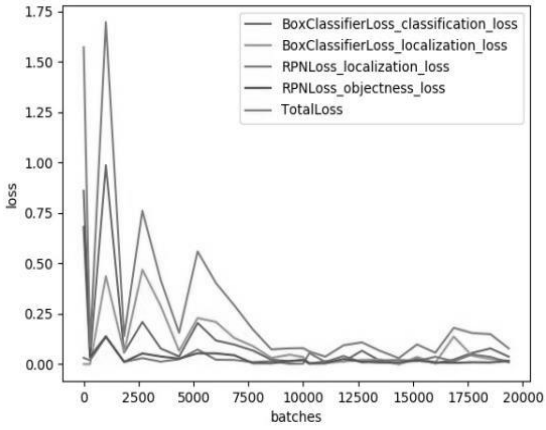

Fig. 7. Network loss diagram

In Figure 6, gobal_step/sec is a performance indicator that shows how many batches of data can be processed per second during model training (referred to as gradient update).

Figure 7 Data meaning:

Loss of regional candidate network:

- localization_loss: localization loss of RPN.
- object ness_loss: loss of the classifier

Final classification loss:

- Box Classifier Loss/classification_loss: classify the detected objects into various categories of loss
- Box Classifier Loss/localization_loss: localization loss
- Total Loss: total loss

It can be seen from Figure 6 that when the loss of the bounding box regressor, the loss of the classifier, the loss of the detected objects classified into various categories, the classification_loss and localization_loss in the final loss, and the total loss of 9000 batches, the loss value is almost Reached a value of 0, and from 9,000 batches to 20,000 batches, their loss value fluctuated between 0 and 0.25, and tended to 0. So this shows that the network model fits the data well, there is no over-fitting phenomenon, and effective learning is carried out, so this network model is suitable for identifying and navigating for the visually impaired.

In the search for multiple tasks, we will consider two questions: one is how much of the information in the searched area is the information you need, and the other is how much of the information you need is searched out. Therefore, use precision and recall two performance indicators to express. As shown in Figure 8 Precision-recall curve:

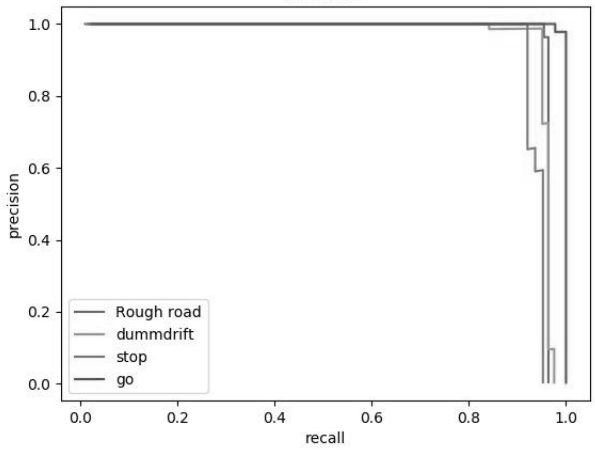

Fig. 8. Precision-Recall curve

Figure 8 above shows the precision-recall curves corresponding to the blind road, obstacles (dummdrift), traffic lights (stop), and traffic lights (go). The four types of PR curves are shown in Figure 8. When the recall value is 0 At ~0.85, the precision value of the four types of curves is almost 1, indicating that the sample types are all positive samples at this time. Only when the recall value is 0.85~1.0, fluctuations begin to appear, and then begin to decline, and the positive samples decrease. The area enclosed by the lower left of the four types of PR curves is the AP (Average-Precision) value. It can be seen from Figure 8 that the geometric area of the AP value is relatively large. According to the data output, the four types of AP values in Table 1 can be obtained. Substituting the trained model into the test program, several types of sample recognition test images are obtained, as shown in Fig. 9: Obstacles, traffic lights (stop/go), and blind road verification images.

TABLE I. AP VALUES OF FOUR TYPES

Type name	Average recognition accuracy (AP)
Rough road	0.963306
Dummdrift	0.959844
stop	0.999517
go	0.940470

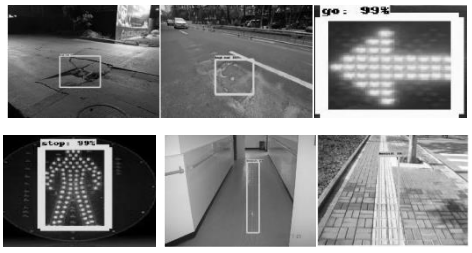

Fig. 9. Obstacles, traffic lights (stop/go), and blind road verification diagram

Therefore, the analysis from Figure 8 and Table 1 and Figure 9 shows that the model performance is very good, the system training is accurate, and the robot's navigation needs are met.

According to the above experimental results, in the training process of 20,000 batches, each loss continuously decreases as the training progresses, and eventually approaches 0. When the recall value is lower than 0.85, the model precision value is close to 1, and the model is accurate The rates are all above 94%. The training results show that the system has fast convergence speed, high recognition accuracy, high average recognition accuracy, and it is not easy to produce errors. In addition, the over-fitting process is well restrained, and the image recognition effect is better. This model is suitable for the research on the navigation direction of the blind guide robot proposed in the paper.

V. Analysis of Mobile Navigation Results

The map simulated by the navigation is a circular map with a size of 3m*3m. Run the written voice navigation algorithm, turn on the camera, identify blind roads, obstacles, etc., perform binarization, and calculate the center position offset of the identified blind roads. The specified offset range is 0~0.1m. If the blind track offset error is within the specified range within the visible center of the camera, the robot will follow the blind track and go straight. Broadcast information about the blind track. If there is an obstacle, the robot will deflect and move to avoid the obstacle. At the same time, the voice broadcast prompts the visually impaired to respond, so as to achieve the navigation effect. The navigation effect is shown in Figure 10, the actual navigation without obstacles is shown in Figure 11, and the actual navigation with obstacles is shown in Figure 12 as follows:

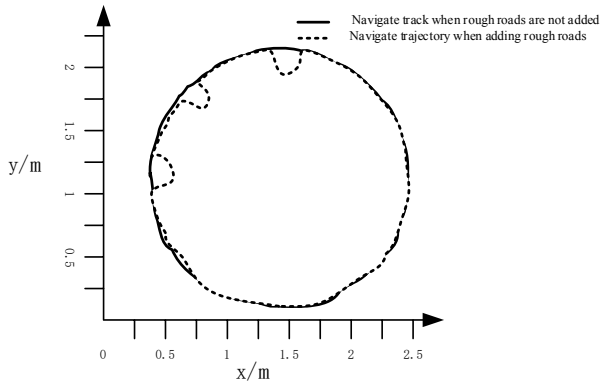

Fig. 10. Navigation effect diagram

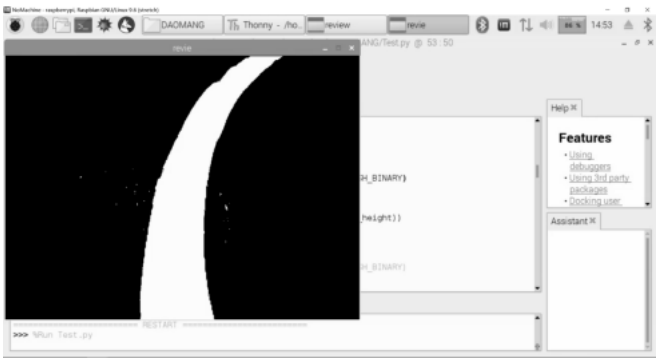

Fig. 11. Part of the actual navigation picture when there are no obstacles

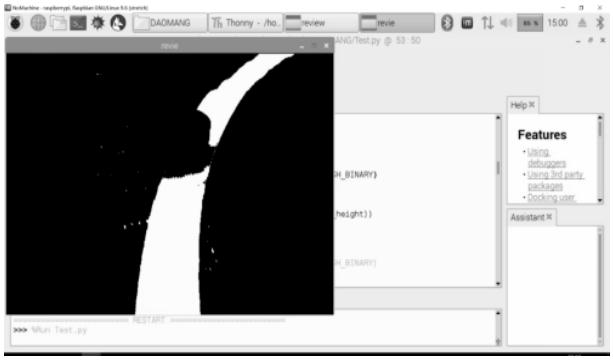

Fig. 12. Part of the actual navigation picture with obstacles

It can be seen from Figure 10 that the robot can navigate steadily for a week in a simulated map with only blind tracks. When an obstacle (a circular bump simulated here) is added, it can be clearly seen from Figure 12 that it is at the position of the obstacle. It can be seen from 10 that when there is an obstacle, the robot makes an effective trajectory offset. The obstacle recognition accuracy in actual operation differs from the recognition accuracy when tested on the training machine by 6.92%, which can meet the real-time navigation function of the blind guide robot. Make it safer for the visually impaired to travel.

VI. Conclusion

The thesis starts from solving the problem of traveling for the visually impaired, and uses the recognition advantages of deep learning to propose a navigation system based on Faster RCNN deep learning. The thesis builds this system on four-wheeled mobile robot research. The results show that the training loss value of the system is close to 0, the average recognition accuracy rate in actual navigation reaches 89.66%, the reaction interval of the four-wheeled robot is within 0.1 seconds, and the blind track center offset value is 0~0.1m Fluctuating within the range, the navigation route of the four-wheeled robot conforms to the estimated route, and can return to the estimated route again after avoiding obstacles. The research shows that the system of this paper has high recognition accuracy, fast calculation speed, stable operation, small error, and meets the navigation requirements, which verifies the effectiveness of the research method. Moreover, the robot equipped with the system has low cost and simple maintenance, which greatly improves the practical application value. However, the four-wheeled movement used in this article cannot cross a high level, so the next step will be an in-depth study of this issue.

References

[1] Wu Zhaohan, Rong Xuewen, Fan Yong. Overview of the research status of blind guide robots[J]. Computer Engineering and Applications, 2020, 56(14):1-13.

[2] Song Yu'e, Liu Yehui, Zhang Xiaoyan, Wang Chengguo, Lin Haonan. The design of smart guide stick based on STM32[J]. Electronic Devices, 2020, 43(05): 1180-1184.

[3] Zhu Aibin, He Dayong, Luo Wencheng, Chen Wei. Research on wearable blind guide robot based on binocular vision method[J]. Machine Design and Research, 2016, 32(05): 31-34.

[4] Wang Guansheng,Zheng Jianghua,Wahhaf Khalick,Zhang Yang,Yao Juhui.Research and application of navigation/path guidance aids for the visually impaired[J].Computer Applications and Software,2012,29(12):147-151.

[5] Muduli U R, Beig A R, Al Jaafari K, et al. Interrupt-Free Operation of Dual-Motor Four-Wheel Drive Electric Vehicle Under Inverter

Failure[J]. IEEE Transactions on Transportation Electrification, 2020, 7(1): 329-338.

[6] Soumeur M A, Gasbaoui B, Abdelkhalek O, et al. Comparative study of energy management strategies for hybrid proton exchange membrane fuel cell four wheel drive electric vehicle[J]. Journal of Power Sources, 2020, 462: 228167.

[7] Wu H, Si Z, Li Z. Trajectory tracking control for four-wheel independent drive intelligent vehicle based on model predictive control[J]. IEEE Access, 2020, 8: 73071-73081.

[8] Chen W, Liang X, Wang Q, et al. Extension coordinated control of four wheel independent drive electric vehicles by AFS and DYC[J]. Control Engineering Practice, 2020, 101: 104504.

[9] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. arXiv preprint arXiv:1506.01497, 2015.

[10] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.

[11] He Z, Zhang L. Multi-adversarial faster-rcnn for unrestricted object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6668-6677.

[12] Wan S, Goudos S. Faster R-CNN for multi-class fruit detection using a robotic vision system[J]. Computer Networks, 2020, 168: 107036.

[13] Benjdira B, Khursheed T, Koubaa A, et al. Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3[C]//2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS). IEEE, 2019: 1-6.

[14] Zhong Z, Sun L, Huo Q. An anchor-free region proposal network for Faster R-CNN-based text detection approaches[J]. International Journal on Document Analysis and Recognition (IJDAR), 2019, 22(3): 315-327.

[15] Hu B, Wang J. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network[J]. IEEE Access, 2020, 8: 108335-108345.

[16] Luo Qing,Zhou Wei,Ma Zijun,Xu Haixia.Dermoscopy image classification method based on FL-ResNet50[J].Progress in Laser and Optoelectronics,2020,57(18):232-240.

[17] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.

[18] Farooq M, Hafeez A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs[J]. arXiv preprint arXiv:2003.14395, 2020.

[19] Jiang Y, Chen L, Zhang H, et al. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module[J]. PloS one, 2019, 14(3): e0214587.

[20] Yamazaki M, Kasagi A, Tabuchi A, et al. Yet another accelerated sgd: Resnet-50 training on imagenet in 74.7 seconds[J]. arXiv preprint arXiv:1903.12650, 2019.

[21] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[22] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[25] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[26] Chen Guojun,Chen Wei,Yu Hanqi.Research on target tracking method of monocular vision underwater robot based on deep learning[J].Machine Tool & Hydraulics,2019,47(23):79-82.

[27] Tao Zhu,Liu Zhengxi,Xiong Yunyu,Li Zheng.Pedestrian head detection based on deep neural network[J].Computer Engineering and Science,2018,40(08):1475-1481.

[28] Wang Heng,Li Xia,Liu Xiaofang,Xu Wenlong.Research on breast cancer pathological image classification based on ResNet50 network[J].Journal of China Jiliang University,2019,30(01):72-77.

[29] Wang Wencheng, Jiang Hui, Qiao Qian, Zhu Hanhao, Zheng Hong. Research on Classification and Recognition of Ten Fish Images Based on ResNet50 Network[J].Rural Economy and Technology,2019,30(19):60-62.

[30] Long Jinyi,Zhou Hua.Real-time detection of outdoor parking spaces based on AlexNet neural network[J].Chinese Science and Technology Papers,2021,16(03):295-300.

[31] Zhang Wuyang,Zhang Wei,Song Fang,Long Lin.Monocular visual obstacle avoidance method for quadrotor UAV based on deep learning[J].Computer Applications,2019,39(04):1001-1005.

[32] Chen Y, Lin Z, Zhao X, et al. Deep learning-based classification of hyperspectral data[J]. IEEE Journal of Selected topics in applied earth observations and remote sensing, 2014, 7(6): 2094-2107.

[33] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision.2015:1440-1448.

[34] Ma Xiaoming,Zhang Zhili,Gao Huimin.Simulation of Intelligent Target Positioning of Sorting Robot Based on Deep Learning[J].Computer Simulation,2020,37(06):314-317.

[35] Mei J, Xiao C, Szepesvari C, et al. On the global convergence rates of softmax policy gradient methods[C]//International Conference on Machine Learning. PMLR, 2020: 6820-6829.

[36] Li Y, Wang T, Kang B, et al. Overcoming classifier imbalance for long-tail object detection with balanced group softmax[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10991-11000.

[37] Liu Y, He L, Liu J. Large margin softmax loss for speaker verification[J]. arXiv preprint arXiv:1904.03479, 2019.