Proceedings of 2020 IEEE
International Conference on Mechatronics and Automation
October 13 - 16, Beijing, China

A Novel Target Recognition System for the Amphibious Robot based on Edge Computing and Neural Network

Shuxiang Guo[1,2] *and* Handong Cheng[1]
[1]*Tianjin Key Laboratory for Control Theory&Applications In Complicated systems and Intelligent Robot Laboratory Tianjin University of Technology Binshui Xidao Extension 391,Tianjin,300384,China*
guo@eng.kagawa-u.ac.jp; 497210593@qq.com

Jian Guo[1*]
[2] *Department of Intelligent Mechanical Systems Engineering, Faculty of Engineering*
Kagawa University Takamatsu,Kagawa,Japan
*corresponding author : jianguo@tjut.edu.cn

Jigang Xu[3*]
[3]*Unit68709*
Qinghai Haidong,810700,China
*corresponding author : xujigang216@163.com

***Abstract* –In the past, the accuracy of target recognition was low and the transmission efficiency was low, in this paper, a target recognition system based on Edge calculation is proposed for the platform of Amphibious robot, which is mainly used for sea rescue and marine garbage search. Because of the flexible application of Amphibious robot to sea, land and air, the platform of Amphibious robot is equipped with camera to collect image information. We train neural networks on computers, We store the weight file on the edge calculation chip.The method of Edge calculation is used to run the Convolutional neural network to rescue the victims at sea, We have carried out many experiments to verify the accuracy of target recognition. Because the judgment of Edge nodes can greatly reduce the time of communication with terminals, thus improving the efficiency of rescue.**

Index Terms – Amphibious, Edge computing, Convolutional neural network.

I. INTRODUCTION

With the exploration of tourism and marine resources, more and more tourists are playing at the seaside. Survey ships and developers are targeting at the sea, but there is a risk of maritime distress. China's maritime safety awareness is relatively weak, awareness of the risk of maritime activities is still insufficient, the level of human skills is relatively low, marine facilities are insufficient and natural disasters are frequent, The existing maritime search and rescue is not timely enough, and the task of maritime rescue is a long way to go.

In recent years, great progress has been made in machine vision. Liu Jiaming and others from Beijing naval equipment department proposed a method of UAV recognition based on deep CNN (convolution neural network) [1]. They used SSD algorithm to detect UAV target. Then by training a learning network based on vgg16, we get a model to judge the target, but because the prediction accuracy of recognition is not high and the deep convolution neural network is very dependent on the data set.

In 2016, redomon et al, of Washington University put forward the Yolo algorithm [3], which uses the target from one end to the other for identification. The core of the algorithm is to predict through linear regression algorithm, so as to achieve the target's category label, coordinate and other information. However, this method has the disadvantage of insufficient accuracy. In 2017, they put forward and put forward the yolov2 algorithm [4], which is improved on the basis of Yolo. They select new basic network, add full convolution network and anchor mechanism, and carry out multi-scale training. The accuracy of this algorithm has been improved, but it is still unsatisfactory. In 2018, they improved Yolo algorithm again and put forward yolov3 algorithm [5], which uses classifier or locator to perform detection task again. They applied the model to multiple positions and scales of the image and selected the full convolution network darknet-53 Feature extraction is carried out, and the low-level position information and deep semantic information are highly integrated through the feature interaction layer, so those areas with high scores can be regarded as detection results, yolov3 greatly improves the recognition accuracy of single-level detector, but this algorithm is completed on the upper computer, if the personnel search and rescue needs to be carried out quickly to reduce the communication transmission time between the slave and the main section , we need to be able to complete independent judgment from the end, so we need to improve.

II. THE OVERVIEW OF AMPHIBIOUS ROBOT PLATFORM

The platform of amphibious robot is shown in Figure 1. Spherical amphibious robot is a kind of underwater robot. It can not only walk on land, but also complete the horizontal motion and rotation in water.Its working field covers land and sea. This platform has a hemispherical design [8], The lower body adopts the bionic quadruped's moving mode, and the water-proof steering gear and water sprayer move on land and sea respectively, The main controller controls the driving controller to control the steering of the waterproof steering gear, So it can turn and move on the beach, flat and muddy road.

On land, the main controller controls the driver to complete the control of each steering gear and complete the movement of the whole action group . When underwater, the controller controls four water spray motors to complete the injection to move forward, backward and turn on the water surface and underwater , The spherical underwater robot can be used in many complex environments, and the design of multiple degrees of freedom is more sensitive. The bionic mechanism of the lower body makes the underwater robot move freely in the water [7].

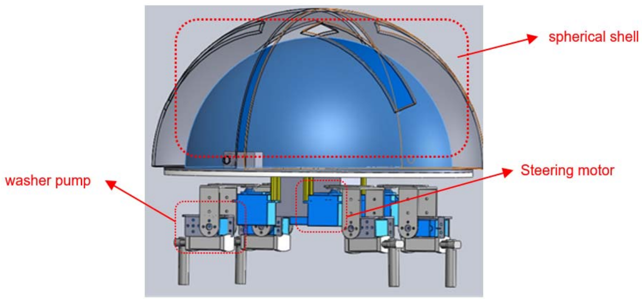

Fig. 1 The platform of Amphibious Robot.

The internal hardware of spherical robot is mainly composed of Arduino, stm32f4, raspberry pie and driving module. Because spherical robot often works in complex environment, we add BMI160 module to measure acceleration. In addition, Als-pt19 module wide range optical sensor is added, in which STM32F4 is used as the processing controller of these sensors[21]- [24].

III. Study of Neural Network and The Construction Of Edge Computing

A. Study of Neural Network

In order to improve the accuracy of detection, we use the method of deep learning for training, and we use convolution neural network for calculation. Convolution neural network is one of the most representative neural networks in the field of deep learning technology, and has made many breakthroughs in the field of image analysis and processing. The neuron in the neural network is the most basic functional unit of the neural network system, which simulates the biological characteristics. When the received signal value of neuron exceeds the set value, the neuron is in an active state, and then it will send a signal to the next neuron for signal transmission. The neural model is shown in Figure 2 [10]-[16].

Numerous neurons form a neural network. Unlike ordinary classifiers, neural network is a huge network. The output of the last layer we see is related to the neurons of each layer.

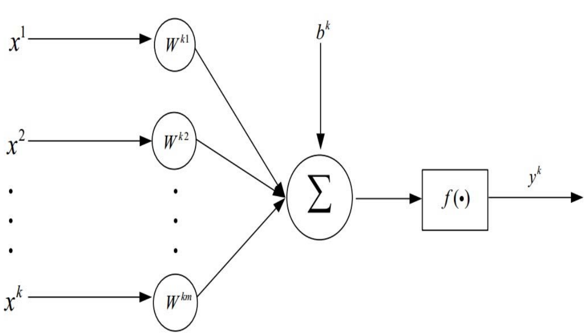

Fig. 2 Neuron model.

The model output formula is:

$$u^k = \sum_{j=1}^{k} w^{kj} x^j \quad (1)$$

$$y^k = f(u^k + b^k) \quad (2)$$

Where x^j is the input signal, w^{kj} is the weight, and b^k is the offset.

Because the nonlinear activation function is used in the neural network, the superposition of multiple linear functions is linear. In order to enhance the expression ability and learning ability of the network, the nonlinear activation function is introduced. When we use sigmoid and tanh as activation functions, we first need to normalize the input, otherwise all the activated signals will enter the flat area, the output of the hidden layer will all tend to be the same, and the original feature expression will not show. Therefore, the nonlinear activation function used in this paper is the Relu function. The Relu function is shown in Figure 3.

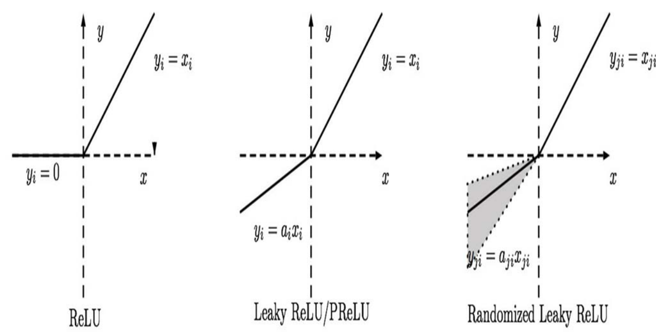

Fig. 3 Relu function.

In this paper, gradient descent method is used to optimize the neural network [17]. The purpose is to reduce the loss as much as possible. The core of this algorithm is to use the gradual descent method to solve the local optimal solution. The gradient descent formula is:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \quad (3)$$

The loss function is:

$$\mathrm{J}(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^i - y^i)^2) \quad (4)$$

Where m is the total number of samples. The purpose of optimization is to find the minimum value of $\mathrm{J}(\theta)$. The gradient direction refers to the partial derivative of $\mathrm{J}(\theta)$ to θ. The minimum loss is obtained by calculating the optimal parameter θ along the reverse direction of the gradient direction, This can reduce the problem of over fitting in the training process [18].

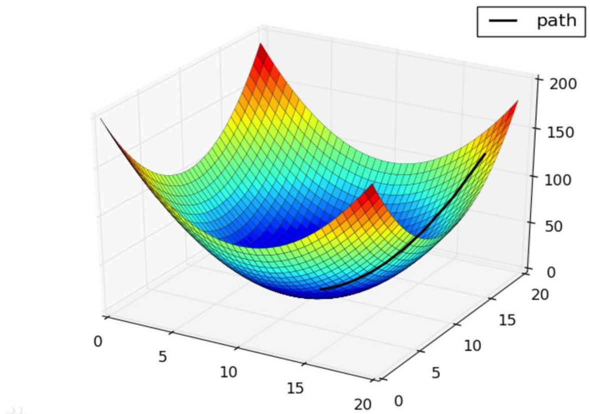

Fig. 4 Gradient Descent Graph.

The gradient descent diagram is shown in Figure 4 [19]. The black realization in the figure represents the process of gradient descent, which is iterated continuously to minimize the loss value. When we select different gradient points, the descent path is different, and even the local optimal solution may be different.

Convolution neural network includes convolution layer and pooling layer. Convolution layer is mainly used to convolute the input image signal. Convolution core in convolution layer is mainly used to perform linear convolution operation. The pooling layer is located between two adjacent convolutions. Its function is to reduce the parameters in the later full connection layer. We also use the method of weight sharing to reduce the parameters. The image convolution operation is shown in Figure 5.

The lower network in the graph is mapped to the upper network by convolution kernel, and the convolution operation is completed step by step.

Our neural network needs to run on the Edge Computing chip. First, we install Ubuntu system on the Edge Computing chip. Then we use Anaconda to build tensorflow and train in depth learning in tensorflow framework. We need to call opencv library to process the image. The information collected by the OV camera will be transmitted to the Edge calculation board for preprocessing. Since the neural network has been built and trained, the image signal will be used for target recognition and processing.

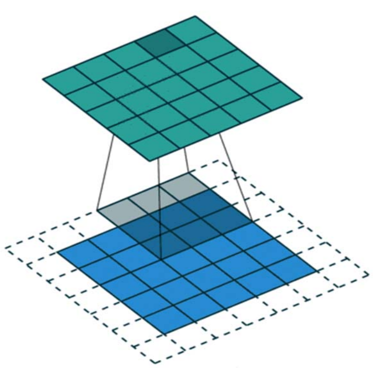

Fig. 5 The image convolution operation.

B. Study on Target Recognition Algorithm

The neural network algorithm is yolov3 algorithm, and the main frame of yolov3 algorithm is darknet53. The shortcut method is used to connect the two layers of networks that are not adjacent to each other, which solves the gradient disappearance caused by the increase of layers of darknet53. Darknet53 is a combination of full convolution and recursive structure.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 6 Darknet53 Network Structure.

The network structure of Darknet53 is shown in Figure 6 [5]. When receiving the image, it is reduced to 52, 26 and 13 layers through deep convolution. In the three layers of the darknet53 framework, there are full convolution feature extractors, the corresponding internal convolution kernel structure of the feature extractors, and multiple convolution kernels interleave to achieve the purpose. The input of the current feature layer has part of the output from the previous layer. Each feature layer has an output prediction result. Finally, the final prediction result is obtained by linear regression according to the confidence level [20].

When we need to carry out forward propagation, the size transformation of tensor is realized by changing the step size of convolution kernel, such as strip = (2,2), which is equivalent to reducing the Edge length of image to one-half of the original, so the area of image is reduced to one quarter of the original

We use the regression method to predict the border, and the network coordinate formula is:

$$b_x = \sigma(t_x) + c_x \tag{5}$$

$$b_y = \sigma(t_y) + c_y \tag{6}$$

$$b_w = p_w e^{t_w} \tag{7}$$

$$b_h = p_h e^{t_h} \tag{8}$$

Where c_x , c_y is the coordinate offset of the network, p_w , p_h is the side length of the preset anchor box, the final frame coordinate value is b_x, b_y, b_w, b_h and the network learning goal is t_x, t_y, t_w, t_h.

Yolov3 model is based on the basic network of darknet53 to extract features. Compared with the network model of yolov2, yolov3 model is faster, because Darknet53 can achieve higher floating-point measurement operation. Through the method of multi-scale detection, the boundary box and category label of the target are predicted by regression on the recognition samples.

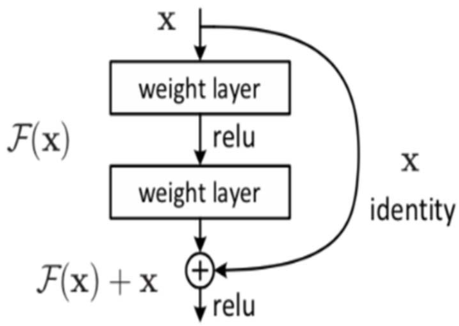

Fig. 7 Residual Block Diagram.

The residual block diagram is shown in Figure 7. In the figure, X represents the input, F(x) indicates the output of the residual block before the activation function of the second layer. Because of the addition of residual network and the addition of shortcut connections in residual network, it is easier to be optimized.

When we have completed the construction of neural network and trained the data, the external sensor detects the image signal and transmits it to the Edge calculation board. At this time, the Edge calculation board will make a judgment. First, it will divide a collected picture into network lattice, and then each cell will detect the target whose center point is exactly in the lattice. Next, the cell will to estimate the bounding box and the confidence score of the bounding box. The confidence of the predicted bounding box is mainly the probability of the bounding box containing the target and the accuracy of the bounding box. In addition, classification will be carried out, and each cell will judge the probability value of the category of this lattice.

The first generation Yolo algorithm will produce the disadvantage that a cell predicts how many bounding boxes it only predicts a group of category probability values. Then, the later Yolo algorithm improves and binds the predicted value of category probability with the bounding box to solve this problem. In addition, yolov3 added the residual network for the biggest improvement. Another point is to use the feature pyramid networks for object detection to achieve multi-scale detection, which can achieve faster detection. If the category label of the person is detected, the signal will be returned to the main control board.

C. The Construction of Edge Computing

We have added a cloud platform under the Amphibious robot to support the OV camera to collect image information [9]. We have added an Edge calculation board inside the Amphibious robot, which is mainly used for image recognition. The information collected by the OV camera will be transmitted to the Edge calculation board, which uses raspberry Pie performs neural network operation to complete Edge calculation.

As Google released the armv7 version of tensorflow in 2017, it is now possible to install the tensorflow framework on raspberry pi. We chose the fourth generation raspberry pie. The internal parameters are 4G of running memory and 64g of storage. It supports dual frequency Wi Fi, Bluetooth 5.0, two micro HDMI 2.0 interfaces (4K 60fps), Gigabit network interface, Mipi DSI interface, Mipi CSI camera interface, stereo headset interface, two USB 3.0, two USB 2.0, and the extension interface is still 40 pin GPIO.

We installed the tensorflow framework on raspberry pi, and then installed the openmv library. And then we installed numpy on raspberry pi. Finally, install version 1.14 of tensorflow on raspberry pi. We import the trained weight file to raspberry pi.

IV. Experiment and Result Analysis

Our Edge board uses the fourth generation raspberry pi, then we install the raspberry pie system. Then set up the network of raspberry pi. The computer is connected with raspberry pie through the network cable, and the computer is connected with Wi-Fi to share the network between the wireless network of notebook and raspberry pi. Then we install opencv and tensorlfow on the raspberry pi. Limited performance due to Edge computing, We adopted Mini Yolo V3 model. It turns out that small models can bring faster speeds. At the same time, we carried out a large-scale model experiment on the upper computer.

After deep learning training, we have carried out multiple experiments for comparison and analysis. We have carried out the comparison between yolov3 and other algorithms in the prediction value of VOC data set, and the comparison results are shown in Table I.

TABLE I
Forecast Value Comparison

Detector	Training Data Set	mAP
SSD 300	VOC2007	74.6
YOLOv1	VOC2007	63.3
YOLOv2	VOC2007	69.1
YOLOv3	VOC2007	79.6
Mini-YOLOv3	VOC2007	76.3

We trained the data set of VOC2007.The experimental results of yolov1 and SSD (Single Shot MultiBox Detector) are from literature, and the experimental results of yolov2 are from literature. In order to make a comparison of yolov3, this paper adopts the comparative experiment of the same data set. It can be seen from the chart that SSD is better than yolov1 and yolov2, but yolov3 is the best performance result. The Mini-YoloV3 detection speed is the fastest.

Although yolov2's darknet-19 still has a great advantage in speed, compared with yolov2's Yolov3 is not so fast, but on the basis of real-time, it pursues the accuracy of recognition. Because we use the way of Edge computing to rescue the ocean, which reduces the time of Edge node sending to the terminal, we must pursue the accuracy of Edge node judgment.

In order to reduce the communication time, we use the concept of edge computing to identify the target because of the emergency time of the sea rescue.We use a totally different method to solve the problem of target detection. It gives the collected image to the computing board for one-time forward processing of neural network. However, although SSD is also used for one-time forward processing of neural network, we can see in the chart that yolov3 achieves higher precision and faster operation speed than SSD. Therefore, Mini-Yolov3 algorithm based on neural network is adopted on the Edge node.

Next, ten groups of experiments were carried out respectively. Five groups of pictures came from the Internet, and five groups took pictures in their own reality.

This platform can be carried on amphibious robots to search for garbage at sea. Figure 10 shows the Edge Computing chip mounted on the Amphibious Robot. We can detect many kinds of objects. Figure 8 shows the edge computing board mounted on a spherical robot.

From the ten experiments, we can see that the recognition algorithm using convolutional neural network is much faster and more accurate than the traditional computer vision using sliding window to find objects of different areas and sizes. All of our ten experiments can accurately identify the position of people from the image.

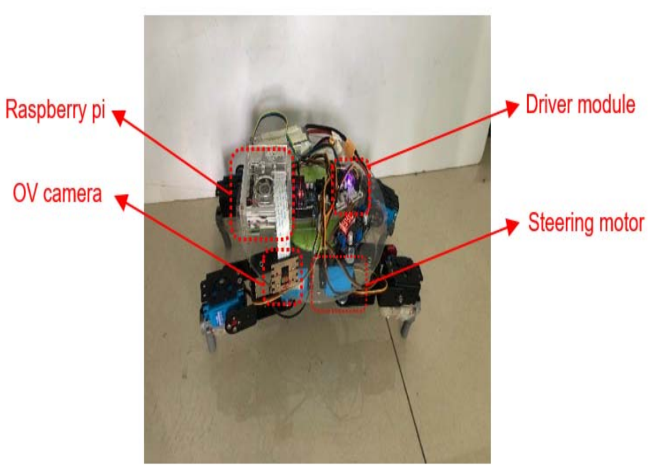

Fig. 8 Edge Computing chip carrying diagram.

Because we mainly carry out rescue operations in the ocean, we collect experiments on people blocking different parts, so as to verify the feasibility of our experiments.

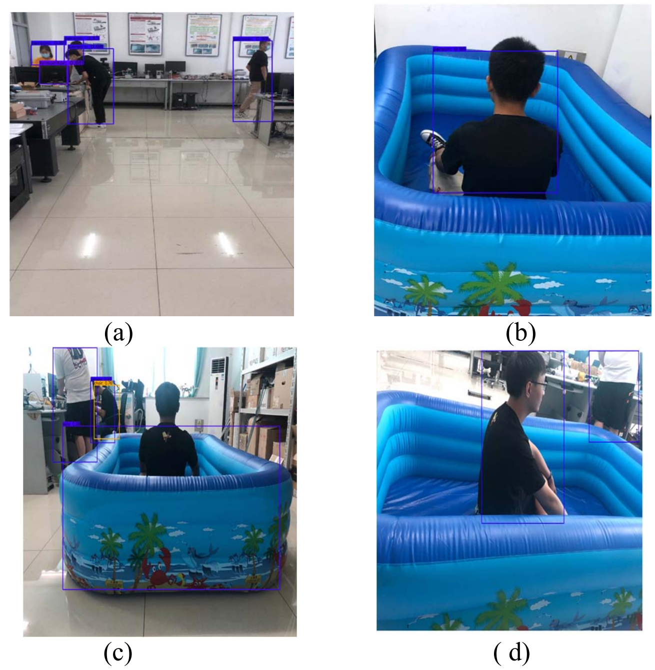

Fig. 9 Target detection results.

Figure 9 shows the experimental results. In our marine rescue plan, the scene of people at sea is simulated, and different parts of people are blocked respectively, which can be accurately identified. Then we will immediately carry out rescue.

The task we need the Edge node to complete is to have the ability of independent judgment, rather than relying on the terminal to judge and then conduct command processing, which will waste a lot of practice, thus missing the best rescue time.

In the 100 groups of experiments, we counted the prediction rate of each image and selected the average value to make a, In order to observe the data intuitively, take an average value every ten groups. Figure 10 shows the confidence data graph.

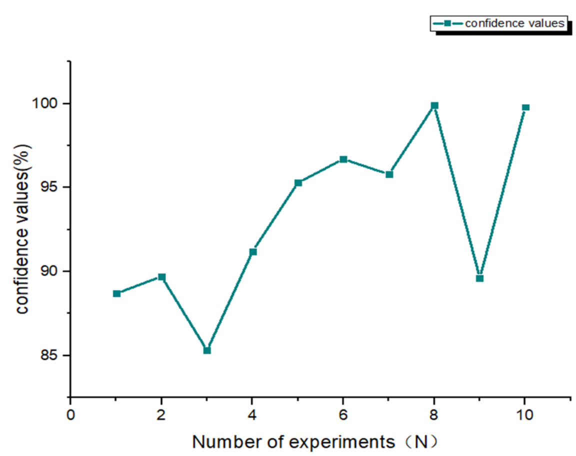

Fig. 10 Confidence values data graph.

From the picture, we can see that the highest prediction rate of the ten experiments is 0.999, and the lowest prediction rate is 0.857 of the third group. Therefore, when we recognize the target, the edge computing node will feedback information. we will send signals to the main control board for the next series of marine rescue operations. We use the concept of edge computing to avoid the communication time of image transmission, and carry edge computing nodes directly on the spherical amphibious robot.

V. CONCLUSIONS AND FUTURE WORK

In this paper, We proposed a target recognition system based on edge computing, The purpose is to improve the speed of ocean rescue.we used the tensorflow framework to build the convolution neural network, and chosed the miniYOLOV3 detection algorithm. We called the openCV Library in tensorflow. After initialization, we had adjusted the parameters, and each border box followed a confidence value. In the first step, all lower than the confidence threshold would be excluded. Next, the construction was completed in the Edge node. This paper mainly dealt with the fast and real-time rescue of the Edge node, which ensured the rescue of personnel as fast as possible. We could processed and analyzed the data in real time or faster, so that the data processing was closer to the source, rather than the external data center or cloud, which could shorten the delay time, and we greatly improved the efficiency of rescue. In the follow-up work, we need to carry out the denoising and defogging treatment in the complex environment, such as the fog weather or the environment with low visibility, and we need to test the neural network training model to judge and rescue the personnel.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (61703305), Key Research Program of the Natural Science Foundation of Tianjin (18JCZDJC38500) and Innovative Cooperation Project of Tianjin Scientific and Technological (18PTZWHZ00090).

REFERENCE

[1] Jian Guo, Shuxiang Guo, et al, "Design and CharacteristicEvaluation of a Novel Amphibious Spherical Robot", *Microsystem Technologies,* Vol.23, No.6, pp.1-14,2017.

[2] Girshick R, Donahue J, Darrell T, et al, "Rich Feature Hierarchies forAccurate Object Detection and Semantic Segmentation "，*Computer Vision and Pattern Recognition (CVPR)* , vol. 139, pp.580-587, 2014.

[3] Redmon J, Divvala S, Girshick R, et al, "You Only Look Once: Unified, Real-Time Object Detection" ,*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , 2016,pp.779-788.

[4] Redmon J, Farhadi A, et al, "YOLO9000: Better, Faster, Stronger " , *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017,pp.6517-6525.

[5] Redmon J, Farhadi A, et al, " YOLOv3: An Incremental Improvement " ，IEEE Computer Vision and Pattern Recognition, 2018,pp.520-529.

[6] Simonyan K, Zisserman A, et al, "Very Deep Convolutional Networks for Large-Scale Image Recognition " , *international conference on learning representations*, 2015,pp.568-576.

[7] Watanabe, K, Iwatani, Y,Nonaka K,Hashimoto K, et al, "A visual-servo-based assistant system for unmanned helicopter control" , *RSJ International Conference on Intelligent Robots and Systems*, 2008,pp. 822–827.

[8] Huiming Xing, Shuxiang Guo et al , "Hybrid Locomotion Evaluation for a Novel Amphibious Spherical Robot", *Applied Sciences*, Vol.8, No.2, pp.1-8,2016.

[9] Guo S, He Y, et al. Modeling and experimental evaluation of an improved amphibious robot with compact structure. *Robotics and Computer-Integrated Manufacturing*, pp,51:37-52,2018.

[10] Shuxiang Guo, Yanlin He et al, "Modal and fatigue analysis of critical components of an amphibious spherical robot " , Microsystem Technologies, Vol.23, No.6, pp.1-15, 2016.

[11] Xihuan Hou, Shuxiang Guo, et al, "Hydrodynamic Analysis-Based Modeling and Experimental Verification of a NewWater-Jet Thruster for an Amphibious Spherical Robot", *Sensors*, Vol.19, No.259, pp.1-15,2016.

[12] Bell S, Lawrence Zitnick C, Bala K, et al, "Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks" , *IEEE Computer Vision and Pattern Recognition*, 2016,pp.2874-2883.

[13] Huiming Xing, Shuxiang Guo, et al, "Hybrid Locomotion Evaluation for a Novel Amphibious Spherical Robot", *Applied Sciences*, Vol.8, No.2, pp.1-15, 2018.

[14] Shuxiang Guo, Shaowu Pan et al, "A system on chip-based real-time tracking system for amphibious spherical robots", *International Journal of Advanced Robotic Systems,* Vol.14, No.4, pp.1-9,2018.

[15] Yao X, Cheng Q, Zhang G Q, et al, "A Novel Independent RNN Approach to Classification of Seizures against Non-seizures", *arXiv* ,pp.1903,2015.

[16] Wildes R P, Asmuth J C, Green G L, et al, "A machine-vision system for iris recognition ", *Machine Vision & Applications* , 1996,pp.1-8.

[17] Fu C Y, Liu W, Ranga A, et al, "DSSD : Deconvolutional Single Shot Detector", *IEEE Conference on Computer Vision and PatternRecognition (CVPR)*. 2017,pp.1-8,.

[18] Li, S.; van der Horst, E.; Duernay, P.; De Wagter, C.; de Croon, G.C.H.E, et al, "Visual model-predictive localization for computationally efficient autonomous racing of a 72-gram drone", *arXiv* , 2019,pp.10110.

[19] Namiki, A.; Matsushita, S.; Ozeki, T.; Nonam`i, K , et al, "Hierarchical processing architecture for an air-hockey robot system", *Robotics and Automation, Karlsruhe*, 2013,pp. 1187–1192.

[20] Shuxiang Guo, Yanlin He, et al "Modal and fatigue analysis of critical components of an amphibious spherical robot", *Microsystem Technologies*, Vol.23,No.6, pp.1-15,2016.

[21] Maoxun Li, Shuxiang Guo, et al, "Designand performance evaluation of amphibious spherical robot" , *Roboticsand Autonomous Systems*, Vol.64, No.6, pp.21-34, 2018.

[22] Liang Zheng, Shuxiang Guo, et al, "The communication and stability evaluation of amphibious spherical robots", *Microsystem Technologies*, Vol.24, No.6, pp.1-12,2016.

[23] Jian Guo, Shuxiang Guo et al, Li L. "Design and characteristic evaluation of a novel amphibious spherical robot", *Microsystem Technologies*, Vol.23, No.6, pp.1999-2012, 2017.

[24] Yanlin He , Shuxiang Guo et al, "Preliminary Mechanical Analysis of an Improved Amphibious Spherical Father Robot " , *Microsystem Technologies*, Vol.22, No.8, pp.2051-2066,2016.