Research on Underwater Image Semantic Segmentation Method Based on SegNet

Jialun Wang, Fusheng Zha(✉), and Xiuwen Bi

Institute of Robotics, School of Mechanical and Electrical Engineering, Harbin University of Technology, Harbin, China
zhafusheng@hit.edu.cn

Abstract. At present, there is little research on underwater image semantic segmentation, and the effect of training using the traditional semantic segmentation network model on land is not ideal. This paper analyzes the current development of underwater image semantic segmentation, establishes a multi-objective underwater image semantic segmentation dataset, and proposes a network model for the impact of underwater complex conditions on semantic segmentation. Experiments show that this network has better segmentation effect than SegNet network in processing underwater images.

Keywords: Underwater image · Semantic segmentation · Image enhancement · Edge extraction

1 Introduction

The ocean is an important resource of a country, known as the "sixth continent". Seawater covers 71% of the earth's surface, but the rate of human exploration is only 5%, and the utilization and development of the ocean is less than 1% [1]. Today, when land exploration is relatively perfect, it has become a new goal for many countries to focus on the ocean and develop and protect marine resources in the 21st century. Responding to the new global situation aims to safeguard and develop China's maritime rights and interests and break through the island chain blockade. China has also accelerated its strategic deployment for the ocean. Through the acquisition, transmission, processing and application of seabed information, it has a vital impact on the more perfect protection of seabed resources and the rational development and utilization of seabed resources [2].

Through semantic segmentation of the underwater images, the unmanned underwater vehicle can identify the area or target of interest to its own side (including coral, underwater pipeline, submarine, etc.), which can not only realize the relevant research of underwater resource exploitation and environmental monitoring, but also provide necessary conditions for subsequent map construction; At the same time, real-time underwater image semantic segmentation also provides the possibility for applications such as navigation.

As a key technology of image processing, semantic segmentation has been more and more favored by relevant researchers in recent years. Its huge application space in many

L. Yan and J. Na (Eds.): ICFPMCE 2022, AHE 10, pp. 131–140, 2023.
https://doi.org/10.2991/978-94-6463-022-0_12

fields such as driverless [3, 4], augmented reality [5], medical image analysis [6] and three-dimensional reconstruction [7, 8] has also attracted more and more scholars to study it. Yuan et al. proposed a new Otsu threshold segmentation method, which improves the speed and accuracy of semantic segmentation. However, due to the image obtained by sonar, it does not have the richness of color and line of camera image. Padmavathi et al. proposed an algorithm combining fuzzy algorithm, threshold algorithm and c-means clustering algorithm. It is observed that there is no gradient phenomenon in the segmented image, and a better image is obtained than the conventional fuzzy c-means threshold method [9]. Liu et al. established the corresponding weight according to the relationship between image edge and center, and proposed an image segmentation method [10]. The segmentation effect of the main target of the image is good, but the segmentation effect of the background is poor. Moreover, due to the introduction of additional information, the time of image processing is slightly longer than that of other algorithms. Couprie et al. filter and extract features from the depth map and RGB channels, and use the classifier to classify the super-pixel segmented image. However, when the object is small or there is noise, there will be segmentation errors [11]. Islam et al. [12] introduced several underwater datasets and proposed suim-net structure. Compared with fcn8 and other network structures, although the segmentation accuracy decreases slightly, it has been significantly improved in speed, which can meet the requirements of real-time to a certain extent. Sun Yajing uses multi-scale convolution kernel fusion to segment underwater images using sonar, and improves some performance [13]. Ma Zhiwei et al. Proposed the US net model to improve the segmentation accuracy of underwater images by using the generated pseudo tags and edge extraction. This method improves the extraction of sea urchin and starfish biological images, and there is no in-depth discussion on the complex underwater environment [14].

At present, most underwater segmentation directly adopts traditional algorithms and ignores the impact of water on the quality of semantic segmentation. However, underwater images have many problems, such as color degradation, reduced contrast and blurred boundary. Simply adopting the algorithms commonly used on land often leads to problems such as unable to clearly determine the image boundary and misjudgment of semantic information. As shown in Fig. 1, affected by color degradation, the original seafloor rocks are identified as covered water plants. Therefore, this paper improves the effect of semantic segmentation by image enhancement and edge extraction.

The second chapter introduces the establishment of underwater image semantic segmentation data set, the third chapter introduces the applied network model and training process, and the fourth chapter draws a conclusion through experiments.

(a) (b) (c)

Fig. 1. Semantic error segmentation image. (a) is underwater image, (b) is annotated image, (c) is actual semantic segmentation result.

2 Dataset Establishment

The establishment of the data set consists of two parts. One part is provided by the Minnesota interactive robotics and Vision Laboratory. This part of the data has 1635 groups of underwater pictures and their corresponding cutting masks. It includes 8 parts: human divers (HD), aquatic plants and seaweed (PF), wrecks and ruins (WR), underwater equipment and robots (RO), coral reefs and invertebrates (RI), fish and vertebrates (FV), seabed and rock (SR) and background water body (BW). The color and corresponding code are shown in the Table 1.

The sample dataset is shown as Fig. 2.

From the above information, it can be judged that the dataset generally meets the needs of this subject. However, due to the lack of torpedo and mine, underwater pipeline and pipeline, which need to be classified, it is necessary to add underwater torpedo and mine, underwater pipeline and pipeline, which need to be trained on the basis of the existing dataset. By cutting and sorting the obtained pictures, and collecting the underwater images of the model through the underwater experimental platform built by the laboratory, a series of pictures about the two objectives mentioned above can be obtained, as shown in the Fig. 3.

Mark the collected pictures, and set the colors and codes as shown in the Table 2.

Table 1. Color and corresponding code of original dataset

target classification	colour	code
Background waterbody	(0,0,0)	BW
Human divers	(0,0,255)	HD
Plants/sea-grass	(0,255,0)	PF
Wrecks/ruins	(0,255,255)	WR
Robots/instruments	(255,0,0)	RO
Reefs and invertebrates	(255,0,255)	RI
Fish and vertebrates	(255,255,0)	FV
Sand/sea-floor/ rocks	(255,255,255)	SR

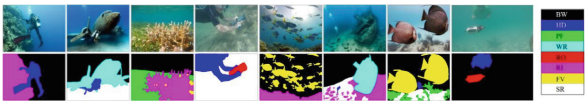

Fig. 2. Sample display of original dataset.

Fig. 3. Partially collected pictures.

Table 2. Color and code of target

target classification	colour	code
Torpedoes and mines	(160,32,240)	TM
canal and pipeline	(255,64,64)	CP

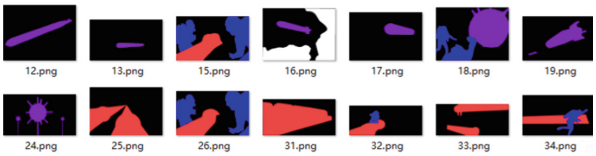

Fig. 4. Semantic annotation.

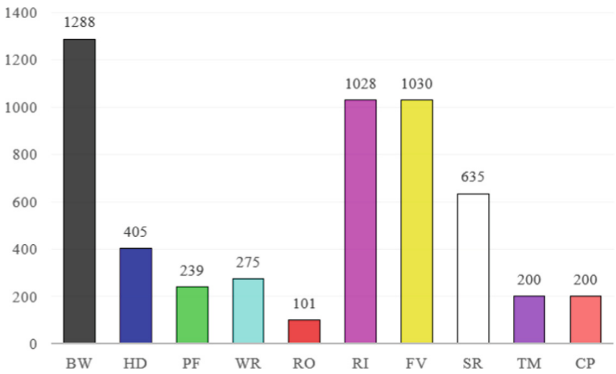

Fig. 5. Number of target categories in each picture.

Labelme is used for the above image annotation. Labelme is an open image annotation software with graphical interface. It is compiled using python, and the graphical interface uses QT (pyqt). Labelme can label images in the form of line segments, rectangles, polygons and circles for tasks such as target detection, semantic segmentation and image classification. By rewriting the bottom layer of labelme, set the colors of each category

to the colors required by this task, that is, the colors in Table 1 and Table 2. Mark some pictures shown in the figure above, and the results are shown in the Fig. 4.

After data set expansion and balance, the quantity comparison of each target is shown in the Fig. 5.

3 UW-SEGNET

The overall framework of underwater image semantic segmentation network based on segnet is shown in the Fig. 6.

We can improve the quality of image by improving contrast and color restoration, and then improve the accuracy of semantic segmentation. Because the density of water is much higher than that of air, there are various small floating particles in the water, including sediment, microorganisms and so on. When light is reflected from the target to the camera, due to the influence of these particles and the absorption and scattering effect of water, the underwater image will have low contrast. Therefore, improving the image contrast is beneficial to improve the effect of underwater image semantic segmentation. In most shallow water images, the red color distribution is generally 50–150, while the blue-green color distribution is generally 70–210. Therefore, the sensitivity of underwater images to RGB three channels is inconsistent. The common way is to stretch the color range directly to the whole space of 0–255, and using the same stretching method will lead to red over compensation. Therefore, each color space is stretched by η. The stretching formula is as follows:

$$O_{min} = I_{min} - \frac{(I_{max} - I_{min}) \times \eta}{2} \tag{1}$$

$$O_{max} = I_{max} + \frac{(I_{max} - I_{min}) \times \eta}{2} \tag{2}$$

$$P_o = (P_i - I_{min})\left(\frac{O_{max} - O_{min}}{I_{max} - I_{min}}\right) + O_{min} \tag{3}$$

I_{min} is minimum input pixel value, 50 or 70, I_{max} is maximum input pixel value, 150 or 210, O_{min} is minimum output pixel value, O_{max} is maximum output pixel value, P_o is

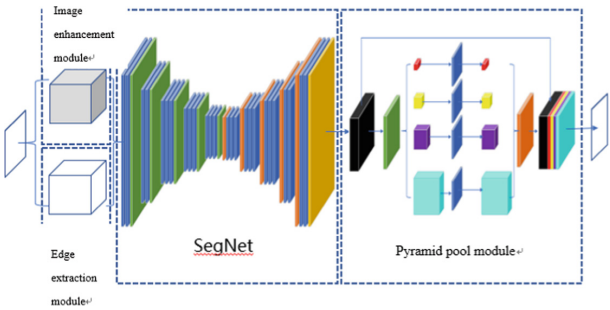

Fig. 6. Underwater semantic segmentation network model.

Fig. 7. Image enhancement effect.

input pixel value, P_i is output pixel value, η is tensile coefficient, considering the pixel distribution of each color channel, η is 50%.

As the light of different colors will attenuate in varying degrees when propagating underwater, the red light with the longest wavelength will attenuate the fastest, and will completely fade and disappear at about 5 m, followed by green light, which will fade to disappear at about 20 m, while the blue light with the shortest wavelength can be transmitted to a longer distance. Color distortion will lead to the main color of blue-green in the underwater image, making some originally blue-green targets blurred. In serious cases, it will also lead to misjudgment in semantic segmentation, such as identifying white reefs or sea sand on the seabed as aquatic plants attached to reefs. Therefore, it is necessary to use color correction to balance the three RGB colors to the correct color before attenuation.

After the contrast of the underwater image is improved, the color of the image is corrected. Firstly, the RGB three channel color is converted into the CIE LAB color model. In CIE LAB, the L component is equivalent to the brightness term, which is linearly sliding stretched by the Rayleigh distribution. The upper and lower bounds are set to 100 and 0 respectively, the setting range of A and B is −128 ~ 127, a and B represent red and yellow respectively, that is, (127, 0) represents red, (0, −128) represents blue, The stretching formula is given by the following formula:

$$P\chi = I\chi \times (\varphi^{1-\left|\frac{I\chi}{128}\right|}),\ \chi \in \{a, b\} \tag{4}$$

where Iχ and Pχ represent input and output pixels respectively, and the optimal solution of φ given by others can be taken as 1.3.

The result of image enhancement is shown in Fig. 7.

Due to the difference between underwater image semantic segmentation and land semantic segmentation, due to the influence of water, in addition to the low contrast and color degradation, there will also be problems of fuzzy details and boundaries and low image definition. The human eye cannot observe the target range to be detected, and these problems will lead to the same problems in semantic segmentation, resulting in unsatisfactory semantic segmentation results. In the practical application of edge detection, Sobel operator is often used. Sobel operator contains two 3 × 3, respectively calculate the brightness change of an image on a vector and the brightness change on the normal vector. When the threshold increases, the feature extraction of underwater image

Fig. 8. Effect drawing of edge extraction.

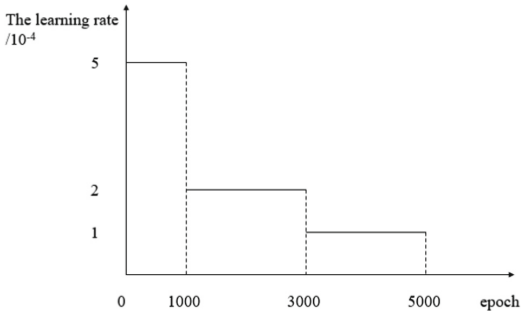

Fig. 9. Learning rate setting.

is clearer, so this paper takes the threshold $G_{max} = 255$. The result of edge extraction is shown in Fig. 8.

After inputting the image, the image is enhanced and edge extracted respectively, and then the two are fused. The original RGB three channel image is transformed into a four channel image with edge information, and then it enters the main SegNet for encoding and decoding. Finally, the pyramid pooling module is used to improve the ability of the image to perceive the context, and finally a segmented image with semantic information is output. The loss function adopts the binary cross entropy loss function, and the function formula is as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y^{(i)} \log y_p^{(i)} + (1 - y^{(i)}) \log(1 - y_p^{(i)}) \tag{5}$$

N is number of samples, $y_p^{(i)}$ is estimate, $y^{(i)}$ is true value.

The learning rate setting is shown in the Fig. 9.

4 Evaluate

This paper evaluates image enhancement and semantic segmentation, that is, the effect of image enhancement is evaluated locally, and then the effect of semantic segmentation is evaluated as a whole.

Mean square error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and enhanced underwater image quality evaluation index (UCIQE) are used to analyze the effect of image enhancement. Use the above evaluation methods to evaluate the three image enhancement methods of AHS, UCM and ICM, and sort out the data in the following Table 3.

Table 3. Image enhancement evaluation of AHS, UCM and ICM

type	PSNR	SSIM	UCIQE
AHS	19.4349	0.8602	5.2649
UCM	19.6845	0.8038	6.0683
ICM	15.1225	0.9485	4.5444

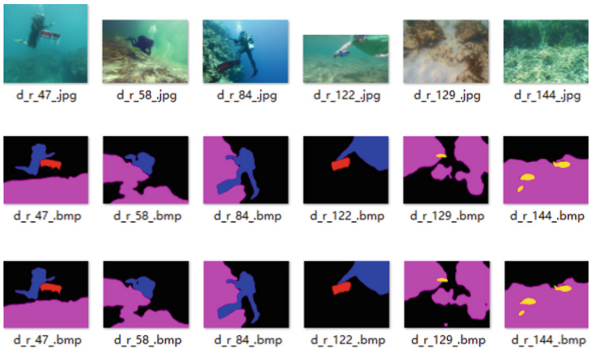

Fig. 10. Semantic segmentation result diagram: the first row is underwater image, the second row is SegNet segmentation result, and the third row is UW-SegNet segmentation result.

Table 4. Comparison of network training results

	Recall	Precision	MIoU	F1
UW-SegNet	97.631	88.400	90.339	92.414
SegNet	95.890	86.300	85.683	90.240

It can be seen from the data in the Table 3 that the contrast and brightness of AHS have been improved, and a good balance has been achieved in each index. To sum up, this paper adopts the AHS method to enhance the preprocessing of underwater images.

The results of semantic segmentation and segnet network training are shown in the Fig. 10.

Semantic segmentation generally has five evaluation indexes, namely pixel accuracy (PA), precision, recall, mean intersection over Union (MIoU) and F-measure. The above values can be obtained from the confusion matrix. It can be seen from Table 4 that the method used in this paper has better semantic segmentation effect than the method of directly using SegNet training.

5 Conclusion and Prospect

According to the problems of color degradation, low contrast and edge blur of underwater image, based on the color distribution and other information of underwater image, this

paper proposes two semantic segmentation network front-end modules: image enhancement and edge extraction, which achieve better segmentation effect than SegNet in semantic segmentation of underwater image. However, since the above tests are carried out on a few kinds of classification, the result of multi-objective semantic segmentation is still not ideal, and effectively improving the effect of multi-objective semantic segmentation will be the next work.

References

1. Cong, R., et al.: Research progress of underwater image enhancement and restoration driven by deep learning. Signal Process. **36**(09), 1377–1389 (2020)
2. Guo, J., et al.: Research progress of underwater image enhancement and restoration methods. Chin. J. Image Graph. **22**(03), 273–287 (2017)
3. Wang, J.: Research on key technologies of unmanned vehicle environment perception system. University of Science and Technology of China (2016)
4. Zhang, Q.: Environment modeling of driverless vehicle road scene. Xi'an University of Technology (2018)
5. An, Z., et al.: design and research of augmented reality head up display system combined with image semantic segmentation. J. Opt. **38**(07), 85–91 (2018)
6. Liu, Y., Chen, S.: overview of medical image segmentation methods. Electron. Technol. **30**(08), 169–172 (2017)
7. Wang, G.: Binocular stereo 3D reconstruction based on image segmentation and image cutting. Dalian University of Technology (2013)
8. Shan, J., et al.: real-time 3D semantic map construction in indoor scene. J. Instrum. **40**(05), 240–248 (2019)
9. Yuan, X., Martínez, J.F., Eckert, M., et al.: An improved Otsu threshold segmentation method for underwater simultaneous localization and mapping-based navigation. Sensors **16**(7), 1148 (2016)
10. Padmavathi, G., Muthukumar, M., Thakur, S.K.: Non linear Image segmentation using fuzzy c means clustering method with thresholding for underwater images. Int. J. Comput. Sci. Issues (IJCSI), **7**(3), 35 (2010)
11. Liu, Y.: An efficient image segmentation approach based on graph theory. J. Appl. Sci. Eng. **21**(1), 117–124 (2018)
12. Smeulders, A.W., et al.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
13. Islam, M.J., et al.: Semantic segmentation of underwater imagery: dataset and benchmark. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2020)
14. Sun, Y.: Research on sonar image target recognition technology based on deep learning. Southeast University (2019). https://doi.org/10.27014/d.cnki.gdnau.2019.003356