Underwater image segmentation based on computer vision and research on recognition algorithm

Ma Wenjuan¹ · Xu Feng¹

Received: 11 June 2021 / Accepted: 19 July 2021
© Saudi Society for Geosciences 2021

Abstract
Due to the continuous growth of the world’s population, the development and utilization of marine resources have received great attention. At present, marine fishing relies heavily on divers for underwater operations, which have the disadvantages of high risk, low efficiency, and high cost. Therefore, the development of underwater capture robot which can automatically detect, locate, and capture targets is of great significance for the development of marine economy. Underwater robot is an indispensable equipment for deep-sea operation and plays an irreplaceable role in the development of the ocean. When autonomous underwater vehicles perform underwater operations, they can use computer vision system to obtain clear underwater images and accurate target category information, which can help the manipulator select different grasping parts for different shapes and categories and improve work efficiency. The current underwater vision technology includes “acoustic vision” and “light vision.” Due to the influence of multichannel effect and blind area, the acoustic vision research in detecting and tracking underwater targets is not deep enough. Compared with the acoustic image processing system, the underwater optical vision system has the advantages of image and video capture and has higher real-time performance, which can aim at the target faster and more conveniently. Underwater vision optical system plays an important leading role in the detailed research of underwater vehicle sensing system, which can further improve the autonomous performance of underwater vehicle. In addition, considering the nature of underwater image imaging, we are developing an underwater image segmentation and recognition system based on image processing.

Keywords Computer vision · Underwater image · Image segmentation · Image recognition

Introduction
Image processing technology takes image processing as the main content and combines intelligent automatic control and other technologies to collect, process, and identify images, so that people can access them. At that time, similar functions were only used to process remote sensing images and medical images. Later, this technology has been widely used in many fields of public life (Selvam et al. 2015). As an important research topic in the field of image processing, target acquisition and tracking technology covers many fields such as image processing, pattern recognition, and artificial intelligence management, which has important research and practical significance. Target detection technology is the key to the manufacture of underwater fishing robot, and the results directly affect the subsequent planning and control of the machine (Nawab et al. 2016). The existing target recognition algorithms mainly focus on the target recognition task of land imaging, and the effect of land imaging is better. However, underwater image will be affected by image quality degradation, object clustering and occlusion, and the difficulty of obtaining large-scale data. Therefore, it is impossible to achieve high recognition accuracy only by existing methods (Ukah et al. 2019). Therefore, this paper discusses the problem of target detection in underwater scenes from different angles and takes the fast R-CNN target detection framework as the reference environment. Accurate detection of underwater target is the key to efficient operation of underwater vehicle (Papanikolaou et al. 2005). However, in the complex and changeable underwater environment, the images collected by
the equipment often encounter many problems, such as uneven light, poor contrast, blue-green, and blurred images. In addition, because underwater photography requires a lot of human and material resources, the collection of underwater data is very difficult, there is no high-quality underwater biological records, and the biological species are not rich enough, which makes the underwater robot can detect (Prasanna et al. 2012). In order to solve the problems of insufficient underwater biological imaging samples and poor image quality, this paper takes three kinds of underwater biological detection as examples and proposes the combination of image enhancement and Mask R-CNN structure to realize underwater biological detection (Ravindra and Mor 2019). In this case, target detection is achieved through a small data set (Panda et al. 2020). By comparing the recognition results of underwater target image before and after enhancement with those of the target recognition model based on Mask R-CNN and other models proposed in this paper, the effectiveness and superiority of this method can be verified. The research results presented in this paper can be used as an important reference point for the optical vision system of underwater vehicles and provide a wide range of applications for the study of marine biological resources, underwater fishing, and the construction of defense works (Prasad and Bose 2001).

Materials and methods

Data source

Due to the selective absorption of light by water and pollutants in water, compared with conventional images, underwater images usually have many disadvantages, such as poor visibility, poor contrast, poor clarity, loss of detail, and blue-green tone. In addition, it needs professional equipment and personnel to obtain underwater images, which makes it difficult to create a large number of underwater images. Most of the existing high-quality underwater biological records are fish records.

In this paper, the targeted detection of sea urchin, sea cucumber, and starfish is taken as an example. When using small data sets, image magnification and enhancement techniques are used to solve the above problems. The processing of underwater image data set is described in detail from four aspects: underwater biological data set structure, mini mask technology to reduce the use of video memory, underwater image enhancement, and test enhancement effect (Puthiyasekar et al. 2010).

The underwater biological image used in this paper is selected from the underwater robot target capture competition. The participants train their own models according to the data provided to detect underwater biological targets. In this paper, the selected images from UPRC 2018 are expanded and annotated to form a small underwater biological data set composed of 430 images (Rattan et al. 2005). In this paper, the images containing as many biological species as possible are selected to form the initial data set, and some representative images are shown in Figure 1.

Underwater image segmentation algorithm design

For any point I(i,j) and a certain distance d in the image, the increment of the point is determined, as shown in formula (1).

\[ s(d) = \sqrt{\text{sum}(E_d)/\text{num}(E_d)} \]  

The incremental element \( s(d) \) reflects the degree of difference between a point and a point with a distance equal to \( D \). If the \( D \) value is too large and the correlation is too low, the texture function may not be displayed correctly. In the article \( d = 1,2,3,4,5 \), each pixel can obtain 5 additional functions, which makes the Flip pixel point smaller and larger.

Formula (2) can be obtained from the above formula:

\[ \alpha_{\text{sum}}(x) = \lim_{s \to 0} \frac{\ln \mu_{\text{sum}}(B_s(x))}{\ln r} \]  

In this paper, it is defined as a cubic polynomial, such as formula (3):

\[ \ln \mu_{\text{sum}}(B_s(x)) = \alpha_{\text{sum}3}(x)(\ln r)^3 + \alpha_{\text{sum}2}(x)(\ln r)^2 \]
\[ + \alpha_{\text{sum}}(x)(\ln r) + \alpha_{\text{sum}0}(x) \]

Define the spectrum, as shown in formula (4):

\[ f(\alpha_{\text{sum}}) = FD(\alpha_{\text{sum}}) \]  

Therefore, the increment sign \( s(d) \) reflects the difference between the singular point and its surrounding environment, while the multifractal parameter reflects the local change and the overall change. In fact, for each pixel, this paper obtains 15 feature values from the image. In the process of segmentation, the objects will be classified according to some attributes. Therefore, we can cluster the objects. In this paper, we use the K-means clustering technology commonly used in image processing.

Target recognition experiment design

The average class distance is defined as formula (5) and formula (6).

\[ S_w = \frac{1}{2c} \sum_{i=1}^{c} \left( \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{N_{ij} - 1} \sum_{k=1}^{N_{ij}} \| x^{(i,j)} - x^{(i,k)} \| ^2 \right) \]  

\[ S_w = \frac{1}{2c} \sum_{i=1}^{c} \left( \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{N_{ij} - 1} \sum_{k=1}^{N_{ij}} \| x^{(i,j)} - x^{(i,k)} \| ^2 \right) \]

After output, the above formula can be simplified as formula (7):
In general, the average category vector can be used to represent categories. The mean vector of class I sample is $\mu(i)$. The mean vector of population sample is $\mu$. And the average distance between classes of class C is defined as formula (8):

$$S_b = \frac{1}{c} \sum_{i=1}^{c} \left[ \frac{1}{m} \sum_{l=1}^{m} \| \mu_{(i)} - \mu_l \|^2 \right]$$

The average distance between classes is very small, and the separability is very good. Therefore, the standard definition of distance is as follows: formula (9):

$$J_A = \frac{S_b}{S_w} = \frac{\frac{1}{c} \sum_{i=1}^{c} \sum_{l=1}^{m} \| \mu_{(i)} - \mu_l \|^2}{\frac{1}{c} \sum_{i=1}^{c} \sum_{l=1}^{m} \frac{1}{N_i-1} \sum_{j=1}^{N_i} \| x_{(i,k)} - \mu_{(i)} \|^2}$$

The higher the $J_A$ value is, the better the feature separability is. The size of template elements should not be as large as possible, and the most effective function should be selected to achieve effective pattern recognition.

**Results**

**Image enhancement effect analysis**

In view of the relatively poor appearance of DCP and CLAHE, DCP and CLAHE are deleted. Only the original image, MSRCR, and improved MSRCR are selected for further comparison, and the differences between them can be seen intuitively. As shown in Figure 2, the method proposed in this paper can obviously provide better visual experience.

The formula calculates the information entropy, contrast, and clarity of the extended data set. According to the results of the original image, the data results are corrected to the appropriate size. The specific results are shown in Table 1.

From Table 1, it can be seen that the method proposed in this paper has the highest total score, and the improvement in clarity is larger than that of MSRCR, which is basically consistent with the sensory analysis above. In addition, compared with the original image, the quality of DCP even decreased, which is related to the original purpose of land scene defogging.

**Analysis of underwater image segmentation results**

Figure 3 is a deep-sea image with three types of textures: seawater, rock, and hydrothermal.

Figure 4 is the corresponding gray histogram.

It can be seen from the original image that the gray level distribution in the rock is very wide and overlaps with that of the hydrothermal fluid. It can be seen from the histogram that these three types of objects cannot be the same only through the gray threshold method. Therefore, it is necessary to use texture-based segmentation method.

As shown in Figure 5, the incremental features $\{s(1), s(2), s(3), s(4), s(5)\}$ are used as feature vectors, and the K-means clustering method divides pixels into two categories.
As can be seen from Figure 5, the rock is well divided, but there are many cracks between the rocks. Therefore, we will use mathematical operators for further processing, and the result is shown in Figure 6. The rock is more completely divided, and the white area in the picture represents a rock. After the rock is segmented, the remaining seawater and hydrothermal fluid use multifractal parameters as eigenvectors, as shown in Figure 7.

Experimental results verify the effectiveness of these two functions and lead to better segmentation results.

### Analysis of underwater image recognition results

The comparative experiment using Mask R-CNN structure is used to study the results and speed difference between using mini mask and not using mini mask. The specific results are shown in Table 2.

It can be seen that mini mask has little influence on the recognition accuracy, which may be due to the incorrect information on the tag. However, after using mini mask, the training time of the model will be greatly reduced. Mini mask is very important for a large number of GPU memory.

| Enhanced algorithm | Information entropy | Contrast | Clarity | Overall ratings |
|--------------------|---------------------|----------|---------|----------------|
| Original data set  | 1                   | 1        | 1       | 1              |
| DCP                | 1.031               | 1.064    | 0.765   | 0.945          |
| CLAHE              | 1.095               | 1.045    | 1.165   | 1.165          |
| MSRCR              | 1.135               | 1.165    | 1.648   | 1.304          |
| Improved MSRCR    | 1.154               | 1.065    | 1.762   | 1.346          |
This paper discusses four DCP algorithms, CLAHE, MSRCR, and improved MSRCR, to improve the mask recognition ability of R-CNN. The results are shown in Table 3. Finally, fivefold cross validation is used. The recall rate in the table is the ratio of the number of correctly identified targets to the actual targets in the test set, the accuracy rate is the ratio of the number of correctly identified targets to the number of identified targets, and map is the average AP value. The AP value of the image can be calculated from the fidelity curve, and the area of the triangle at the end of the curve can be subtracted from the area under the curve.

From Table 3, it can be seen that DCP for land scene area has the least improvement compared with Mask R-CNN, which is consistent with the abovementioned perceptual analysis and objective index results. Compared with MSRCR, the improved MSRCR presented in this paper shows slight improvement, which shows the nature of its extension. The performance improvement and objective analysis of the four enhancement algorithms are consistent with the objective analysis of the indicators. This part shows that the performance of deep network is proportional to the image quality.

In order to test the effectiveness of the proposed method, three kinds of depth detection models are selected: YOLOv3, SSD, R-CNN original, traditional detection model, and SIFT-based target detection model, which are used for comparative experiments. All experiments are based on the extended data set and are completed by the fivefold cross validation, and the results are shown in Table 4.

In Table 4, SIFT represents the SIFT-based model. It can be seen that the recall rate and SBP of this method are the best, which are 95.34% and 95.09%, respectively. Compared with the original Mask R-CNN, the data rate is 6.8% and map is improved by 12.36%. However, the speed of this method is
not high, which is not conducive to real-time identification of marine organisms. Although the model based on SIFT has the highest accuracy level, the recall rate is too slow and the overall performance is unsatisfactory.

The mask accuracy of the proposed method is 43.27. However, due to the use of mini mask and underwater features, the edge of the contour becomes blurred, so the annotation is not accurate. Therefore, the accuracy of mask in this paper is only for reference. The final results of target detection and instance segmentation are shown in Figure 8.

**Discussion**

**Overview of computer vision image processing based on fractal**

The commonly used fractal dimension algorithms include Keller box dimension algorithm, Peli covering algorithm based on morphological operator, and Salki differential box counting algorithm. In practical application, different methods for determining dimensions are used for different research objects. When calculating the maximum and minimum gray values of each pixel in each grid, Li and Wang (2000) proposed the double pyramid maximum and minimum values, which improved the speed of fractal dimension estimation.

| Mini mask | Recall rate/% | Accuracy/% | mAP/% | Training time/hour |
|-----------|---------------|------------|-------|-------------------|
| No        | 95.46         | 97.31      | 96.54 | 24.8              |
| Have      | 95.43         | 98.02      | 95.90 | 19.2              |

Li solves the problem of human target in multitemporal background of remote sensing image and proposes to use the improved rectangle size adjustment algorithm to calculate the rectangle size of image after high pass filtering, so that the background and target can be more fully separated (Biney and Christopher 1991). At present, the application of multifractal theory in image processing and analysis is mainly in finding the multifractal spectrum of image, and the theory is widely used in image segmentation and edge detection. Arnéodo et al. (2000) used 2D continuous wavelet transform to calculate the maximum modulus of wavelet, calculated the multifractal spectrum based on it, and analyzed the image. Li and YU (2003) proposed a capacitive multifractal image analysis method. Capacitance is the generalization of measurement (Chadha 1999). By calculating multiple fractal spectra under different multiples, the texture properties are described from different aspects, and a new image texture classification algorithm based on adaptive fuzzy clustering is proposed (Demirel et al. 2008). In this paper, an edge detection method based on multifractal spectrum is proposed. Firstly, the coarse image retention index of image pixels is calculated, and the kernel estimation method is used to estimate the multifractal singular spectrum corresponding to the coarse image spectrum index, and then, the edge points of the image are obtained according to the multifractal spectrum value of the image. Shi et al. (2006) used multifractal model to describe the complex characteristics of ocean disturbance, and on this basis, they proposed a modeling and simulation method of ocean disturbance based on multifractal wavelet model (Drury et al. 1991). The ocean is the second space for human survival. The rational utilization, development, and monitoring of marine resources will play an important role in the sustainable development of human society. The emergence of underwater robot violates the original limitation of artificial diving and provides a powerful tool for human beings to explore and develop deep-sea resources. Most of the underwater intelligent vision systems limited by underwater imaging conditions mainly rely on “acoustic vision” and “vision” technology. Especially, the research of underwater image processing technology based on visible light is very important. Conventional image processing technology highly depends on noise interference and imaging conditions and is not...

![Fig. 7 Final segmentation result](image-url)

Table 2   Results of using and not using mini mask

| Enhanced algorithm | Recall rate/% | Accuracy/% | mAP/% | Training time/hour |
|-------------------|---------------|------------|-------|-------------------|
| Original data set | 83.45         | 94.26      | 84.36 |                   |
| DCP               | 87.30         | 96.95      | 87.54 |                   |
| CLAHE             | 89.29         | 97.96      | 91.26 |                   |
| MSRCR             | 94.25         | 97.64      | 94.48 |                   |
| Improved MSRCR    | 95.43         | 98.02      | 95.90 |                   |

Table 3   Influence of enhancement algorithms on the detection accuracy of Mask R-CNN

| Enhanced algorithm | Recall rate/% | Accuracy/% | mAP/hour |
|-------------------|---------------|------------|----------|
| Original data set | 83.45         | 94.26      | 84.36    |
| DCP               | 87.30         | 96.95      | 87.54    |
| CLAHE             | 89.29         | 97.96      | 91.26    |
| MSRCR             | 94.25         | 97.64      | 94.48    |
| Improved MSRCR    | 95.43         | 98.02      | 95.90    |
suitable for underwater image processing (Edet and Offiong 2002). There are few underwater image processing algorithms based on fractal theory, so it is necessary to realize this aspect.

**Research status of underwater image enhancement**

Compared with the traditional land image, underwater image has many problems, including uneven brightness, blue-green tone, increased noise, dark detail loss, and poor contrast, which makes it difficult to identify underwater targets. Zheng et al. (2015) used dark channel to improve the image quality and underwater image segmentation in the case of poor visibility. Qiu et al.’s (2019) MSR is used to enhance underwater image and improve the detection accuracy of depth detector for sea cucumber. Common underwater image enhancement technologies usually include histogram equalization technology, Retinex-based technology, dark channel priority-based technology, and deep learning-based technology (Ediagbonya et al. 2015).

Histogram correction can fine-tune the image contrast and even the brightness distribution of the whole image. When the distribution of image pixel values is similar, this method works well. However, if the image contains a large number of areas that are too bright or too dark, the contrast of these areas will not be fully improved. Hummel (1977) proposed an adaptive histogram to solve this problem and first applied it to the cockpit display. However, when the processes regions with constant contrast, it often increases the histogram of these regions, which also produces noise (Godt et al. 2006). Therefore, it is necessary to limit the contrast enhancement of these regions to reduce the noise. This technique is called limiting the flattening of contrast histogram. On the basis of adjusting the global histogram, Duan et al. (2010) proposed using adaptive histogram adjustment to reset the contrast of local region in the image. Both global histogram and local histogram use the same hue mapping operator in their adjustment methods. Huang et al. (2018) are used by the method to optimize the image quality of the underwater scene of RGB and CIE lab.

Retinex theory is proposed by land, and the key is to render the image as the product of lighting components and reflection components, so as to reconstruct the real scene seen by human.

| Model          | Recall rate/% | Accuracy/% | mAP/%  | Frames/sec |
|----------------|---------------|------------|--------|------------|
| SIFT           | 3.79          | 100        | 2.78   | 0.65       |
| SSD            | 69.01         | 79.14      | 70.26  | 3.65       |
| YOLOv3         | 76.65         | 89.26      | 79.10  | 10.76      |
| Mask R-CNN     | 83.59         | 94.65      | 84.36  | 0.76       |
| Method of this article | 95.43 | 98.20 | 95.90 | 0.96       |

Fig. 8  Object detection and instance segmentation results
eyes. After more than 40 years of development, Retinex with rich functions balances effect and speed and is widely used. Rahman et al. (2011) added color restoration based on MSR to reduce color shift caused by image enhancement. Lei et al. (2018) proposed an improved MSR method, which can effectively eliminate the distortion of underwater image and improve the appearance of underwater image by improving the gray level of RGB channel. This paper introduces the principle of MSCR in detail.

He et al. (2011) proposed an algorithm, which is used to obtain information about the depth of field based on the dark channel in the image and then defog the foggy image according to it, which is called dark channel prior. The algorithm has a good effect in removing fog in the ground scene, but it is difficult to please people in the underwater scene (Huang and Jin 2008). Dai et al. (2018) have done color compensation based on the bright channel of underwater image to enhance the image quality. Ma and Wang et al. (2019) improved the DCP algorithm to make it suitable for underwater scenes. Their method uses quadtree search to estimate atmospheric illumination values, while multiscale merging is used to calculate depth maps. In Tang et al.’s (2018) method, the depth map is calculated not only from the dark channel, but also from the difference between the bright channel and the dark channel.

Fabbri et al. (2018) used generative countermeasure network to enhance underwater image. Xu and Sun et al. (2016) thought that convolutional neural network based coding and decoding network was used to enhance underwater image. Compared with traditional methods, these deep learning-based image processing technologies are time-consuming, slow due to the lack of GPU or other deep learning computer chips, and put forward higher requirements for computing resources. On the one hand, a large number of low quality underwater images are needed, and they correspond to the basic principles of high quality one by one. Since this paper uses deep learning technology to identify marine organisms, covering the deep network to enhance the image will significantly reduce the network performance in real time. In addition, this article does not provide a high-quality version of the data set used. Therefore, the research of image enhancement in this paper mainly focuses on non-deep learning technology.

**Research status of underwater target recognition**

Usually, target recognition can involve many research fields, including target recognition, classification, and subdivision. Traditionally, finding and identifying objects in an image or video are usually called target. The segmentation of the target pixel plane based on the captured target is called instance segmentation. The segmentation of all pixels in the image and the purpose of using them for a specific target or background are called semantic segmentation.

The conventional underwater target detection algorithm without deep learning has good applicability for some underwater scenes with a single target and small changes. Compared with the segmented objects, it shows that the active contour model has excellent performance, adaptability, and accuracy, but when the target and background color are the same, the performance is poor (Karim 2011). From the biological point of view, the hierarchical background model shows a good ability to detect underwater objects. The recognition rate is about 80% by using the discrimination algorithm to remove the background and the Gaussian coverage model to detect the moving target in the video. The algorithm is simple and fast and can be recognized online. Chen et al. (2014) based on a variety of functions, combined with machine learning and weighted information about each function to segment images, proved higher reliability. Zheng et al. (2015) combined dark channel priority, wavelet transform kernel, and hierarchical multiscale decomposition algorithm for image segmentation, which can capture more details and perform well in the case of poor visibility. Wang et al. (2015) proposed a fast underwater image segmentation method based on improved Markov random field model, which combines Markov random field model with rigid clustering method and proves that the segmentation result has the best speed.

Although the above non-deep learning technology has achieved some practical success in some cases, compared with conventional optical imaging, underwater imaging usually has many problems, such as uneven illumination, poor contrast, image clarity, detail and bluish green tone, and poor versatility of conventional models. It is difficult to achieve satisfactory detection rate for multiple underwater targets with multiple categories. In Pascal VOC data set, the R-CNN target detection algorithm proposed by Girshick et al. (2014) is 30% higher than the previous algorithms in map. After that, there are many deep learning algorithms for target recognition. In the work of Ren et al. (2015), R-CNN has been gradually optimized from fast R-CNN and fast R-CNN to Mask R-CNN and evolved into a classical two-step structure for target recognition and instance segmentation. As the representative of one-stage target recognition framework, SSD and YOLO, as the representative of single-stage target detection framework, are better in speed (Khalid et al. 2017).

With the increasing popularity of object detection algorithm based on deep learning, there are some related researches in the field of underwater surveillance. Konovalov et al. (2019) established a classifier based on Xception model, collected multiple training and test data sets, and finally achieved 99.94% AUC. Salman et al. (2016) trained and tested the two data sets, respectively, which are similar to Konovalov et al.’s (2019) data set merging method. In their deep convolution neural network model, only the output of convolution layer is used to combine features, while the output of lower sampling layer is skipped. This method helps to
save resources, but it is not suitable for completely preserving information. Qin et al. (2016) proposed a deep structure to identify live fish in water by combining convolution neural network, PCA, block histogram, spatial pyramid, and linear SVM. In the deep learning architecture, the masking information of foreground fish is extracted, and the original image is replaced to recognize the fish. Ahmad et al. (2019) replace the three channels of RGB image with the output of Gaussian mixture model, the output of optical flow algorithm, and the gray image of the original image in an innovative way. The model can recognize multiple targets well, but it is only suitable for two kinds of classification. Xu and Matzner (2018) used the YOLO framework to identify fish in water (Magesh et al. 2017). The data sets were from three different hydropower plants, but the map value was only 0.5392.

As we all know, the model based on deep learning is usually not suitable for small sample size. Different from traditional land photography, underwater photography requires professional equipment and personnel, including underwater photographers and rescue workers. In addition, the complex and changeable environment of underwater scene usually leads to poor imaging effect. As a result, there are few data sets that can be used to detect underwater targets, resulting in the lack of appropriate underwater deep learning technology. Even if there are multiple data sets, most of them are only used to define fish targets, Qin, etc. The modified SSD was used to identify sea cucumber. Mahmood et al. (2016) combined the elements extracted by hand with those extracted from VGG network to identify coral reefs in the ocean, achieving the highest classification accuracy in MLC data set.

In order to solve the shortage of related samples, Cecotti and Jha (2018) have proved that Gan imaging can improve the robustness of depth model. Jahic et al. (2019) used Gan to create data sets to optimize handwritten digit recognition using depth model. In this paper, Gan is also used to generate images to expand the data sets. O’Byrne et al. (2018) creatively use artificial images instead of real images to train models. They use Segne model to implement semantic segmentation, train on artificial data sets, and detect the biological deposition on underwater buildings in the real world (Muhammad et al. 2016). However, if you can, real photos are the best choice. Jian et al. (2019) provided the mued data set, which contains hundreds of underwater targets, but it is only suitable for detecting observable objects, and it is difficult to summarize into tasks such as image classification, target detection, instance segmentation, and semantic segmentation.

Therefore, the traditional non-deep learning method has good cognitive ability in simple situations and single purpose. However, their generalization is weak, so it is difficult to deal with complex and changing underwater conditions, especially in the task of target recognition, multiple targets, and multiple classification. Deep learning can only solve this problem and provide high generalization performance. However, due to the small number and variety of seabed data sets, it is difficult to widely deploy deep learning underwater. Therefore, this paper attempts to use a small number of samples to provide satisfactory detection of multiple underwater targets and multiple categories of targets.

Conclusion

At present, the most commonly used methods to perceive underwater environment information are underwater optical vision and underwater acoustic vision. People always think underwater acoustic vision is an important means for robots to perceive underwater environment. It has done a good research on the navigation and collection of target position information of underwater robot. But there is blind area in underwater sonar detection. When the robot is close to the target, acoustic vision cannot provide stable target information. Because the subject of this paper is underwater target and the underwater environment is different from land environment, a new solution for underwater imaging performance is proposed. Underwater target detection in this paper includes image segmentation and feature extraction. Due to the influence of light, underwater noise, and other factors on the absorption and scattering of water, underwater images are fuzzy and unstable. This paper discusses an underwater image segmentation algorithm based on two-dimensional fuzzy Otsu. It extends the maximum variance method between categories from one dimension to two-dimensional and adds fuzzy theory. It combines the double advantages of the maximum variance between class method and fuzzy theory, which makes the algorithm more stable and wider. When using with underwater robot, it is the main guarantee for safe and effective operation to detect and identify the surrounding targets correctly. For this reason, underwater robots are usually equipped with a light and sound vision system to obtain comprehensive information. The system based on optical vision is more intuitive than acoustic image processing system, which is very important for the research of underwater robot sensor system in detail, and can further improve the autonomous performance of underwater robot. Therefore, underwater target detection based on optical vision has always been a hot area underwater research. This paper studies the underwater target detection based on deep learning and analyzes the main reasons to limit the accuracy of underwater target detection. It also solves the problem of underwater image degradation and false negative sample caused by the difficulty of accurately marking underwater target detection records and the small sample problem caused by the difficulty of obtaining underwater image in a large range. In small sample scenarios, the method of target assignment based on foreground segmentation is proposed for detecting underwater targets with weak control and strong control. The effectiveness of the method is
tested by experiments, and the development and implementation of underwater target detection system are completed. In order to solve the problem of underwater image degradation, this paper proposes and implements an underwater image enhancement method based on generation countermeasure network. The underwater style transmission data set is created by combining MSRCR and dehaz net. The network model design is completed to improve the underwater image and loss function. The method can simultaneously achieve color correction and detail restoration of underwater images.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Open access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made. The images or other third party material in this article is included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahmad S, Ahmad S S, Faisal S et al (2019) Automatic fish detection in underwater videos by a deep neuralnetwork-based hybrid motionlearning system[J]. ICESJ Mar Sci 2019

Arnéodo A, Decoster N, Roux SG (2000) A wavelet-based method formultifractal image analysis. I. Application and test applications on isotropic and anisotropic random rough surfaces. Europ Phy J B 2000(15):567–600

Biney CA, Christopher AB (1991) Trace metal concentrations in fish and sediments from the WIWI: a small urban river in Kumasi, Ghana. Trop Ecol 32(2):197–206

Chadha DK (1999) A proposed new diagram for geochemical classification of natural waters and interpretation of chemical data. Hydrogeol J 7(5):431–439

Chadha DK (1999) A proposed new diagram for geochemical classification of natural waters and interpretation of chemical data. Hydrogeol J 7(5):431–439

Chen Z, Wang H, Xu L et al (2014) Visual-adaptation-mechanism based underwater object extraction[J]. Opt Laser Technol 56:119–130

Dai C, Lin M, Wang Z et al (2018) Underwater image enhancement based on bright channel color compensation and fusion [J]. Acta Optica Sinica 38(11):86–95

Demirel S, Tuzen M, Saracoglu S, Soyak M (2008) Evaluation of various digestion procedures for trace element contents of some food materials. J Hazard Mater 152(3):1020–1026

Drury CF, Voroney RP, Beauchamp EG (1991) Availability of NH4+-N to microorganisms and the soil internal N cycle. Soil Biol Biochem 23(2):165–169

Duan J, Bressan M, Dance C et al (2010) Tone-mapping high dynamicrange images by novel histogram adjustment[J]. Pattern Recognition 43(5):1847–1862

Edet AE, Offiong OE (2002) Evaluation of water quality pollution indices for heavy metal contamination monitoring. A study case from Akpabuyo-Odukpani area, Lower Cross River Basin (southeastern Nigeria). Geojournal 57:295–304

Ediagbonya TF, Nmema E, Nwachukwu PC, Teniola OD (2015) Identification and quantification of heavy metals, colloids and anions in water bodies using enrichment factors. J Environ Anal Chem 2:146

Fabri C, Islam MJ, Sattar J (2018) Enhancing underwater imagery using generative adversarial networks[C] 2018 IEEEInternational Conference on Robotics and Automation (ICRA). Brisbane, QLD, Australia: IEEE 2018:7159–7165

Girshick R, Donahue J, Darrell T et al (2014) Rich feature hierarchies for accurate object detection and semanticsegmentation[C]. IEEE Conference on Computer Vision & Pattern Recognition.Columbus, OH, USA: IEEE, 2014:580–587

Godt J, Scheidig F, Grosse-Siestrup C, Esche V, Brandenburg P, Reich A, Gorneberg DA (2006) The toxicity of cadmium and resulting hazards for human health. J Occup Med Toxicol 1(1):1–6

He K, Sun J, Tang X et al (2011) Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence 33(12):2341–2353

Huang D, Wang F, Song W, et al (2018) Underwater image enhancement with adaptive histogram stretching under different color models [J]. Chinese J Image Graph 023(005):640–651

Huang SW, Jin JY (2008) Status of heavy metals in agricultural soils as affected by different patterns of land use. Environ Monit Assess 139(1):317–327

Hummel R (1977) Image enhancement by histogram transformation[J]. Computer Graphics and Image Processing 6(2):184–195

HajicB, Guelfi N, Ries B (2019) Software engineering for dataset augmentation using generative adversarial networks[C] 10th IEEE International Conference on Software Engineering and Service Science. Beijing, China: IEEE 2019:59–66

Jian M, Qi Q, Yu H et al (2019) The extended marine underwater environment database and baseline evaluations[J]. Appl Soft Comput 80:425–437

Karim Z (2011) Risk assessment of dissolved trace metals in drinking water of Karachi, Pakistan. Bull Environ Contam Toxicol 86(6):676–678

Khalid S, Shahid M, Dumat C, Niazi NK, Bibi I, Gul Bakhit HFS, Abbas G, Murtaza B, Javed HMR (2017) Influence of groundwater and wastewater irrigation on lead accumulation in soil and vegetables: implications for health risk assessment and phytoremediation. Int J Phytoremed 19(11):1037–1046

Konovolav DA, Saleh A, Bradley M et al (2019) Underwater fish detection with weak multi-domain supervision[C]. Int Joint Conference on Neural Networks. Budapest, Hungary: IEEE, 2019:1–8

Lei F, Zhu L, Wang X (2018) Research on underwater color image with improved multi-scale Retinex color gray scale[J]. Small microcomputersystem 039(001):185–188

Li H, Yu B (2003) A new image segmentation algorithm based on new multifractal features. Opt Prec Eng 12(6):627–631

Li Q, Wang M (2000) Fast recognition method of fruit defects based on multifractal features. Chinese J Ina Graph 5(2):144–148

Ma M, Huang L (2019). Improved underwater image restoration algorithm based on dark channel a priori [J]. Modern Computer (Professional Edition), 637(01):82–85

Magesh NS, Chandrasekar N, Elango L (2017) Trace element concentrations in the groundwater of the Tamiraparani river basin, South India: insights from human health risk and multivariate statistical techniques. Chemosphere 185:468–479
