Underwater Target Detection Based on Deep Neural Network and Image Enhancement

Yunliang Zheng, Mengxue Yu*, Zi’ao Ma, Rong Liu, Yang Liu

1China Rescue and Salvage of Ministry of Transport of the People's Republic of China, Beijing, 100736, China
2Construction and Administration Bureau of South-to-North Water Diversion Middle Route Project Beijing, 100038, China
3The Medical College of Shihezi University, Shihezi University Shihezi 832000, China
4Department of Cardiothoracic surgery, Guangdong Second Provincial General Hospital, Guangzhou 510317, China

*e-mail: zyl_123202011@yeah.net
*Corresponding author: ymxbj@163.com
b*e-mail: MZAN@126.com
c*e-mail: liurong901107@163.com
d*e-mail: yang_liu123@163.com

Abstract—Underwater target detection tasks refer to the detection of targets contained in underwater images. Unlike traditional target detection tasks, for underwater targets, due to factors such as illumination, camera shake, complex background interference, and diversification of target types, the effect of target detection will be affected. In this paper, we propose a target detection algorithm based on image enhancement and deep network. The algorithm first enhances the image data to obtain a better contrast, and then uses a deep learning algorithm to separate the target and the background to improve the detection performance of the target. Experimental results show that the algorithm can achieve better detection performance.

1. INTRODUCTION

Underwater target detection tasks refer to the detection of targets contained in underwater images. Unlike traditional target detection tasks, for underwater targets, due to factors such as illumination, camera shake, complex background interference, and diversification of target types [1-3], the effect of target detection will be affected. In addition, the absorption and scattering of light by seawater will cause the lack of color characteristics of underwater images, making the task of target detection more difficult.

In recent years, underwater detection operation technology has become a research hotspot in the field of marine technology. To realize underwater detection and operation, underwater robots must be able to quickly and accurately identify and locate targets in the underwater environment that need to capture information. People rely on common marine fishing techniques such as longline fishing and
trawl fishing to measure fish density and distribution, or use correlation analysis to screen out key grid points that significantly affect fish resources to build models for prediction [4]. In addition, marine scientists also use underwater cameras to shoot video tapes, and then manually label fish to analyze fishing conditions. Precise fishery forecast data can solve the problem of the lack of fishery standard services based on the standard system in most fishery standard service systems, and can provide a basis for data decision-making for fishery standard revision guidelines. With the development of underwater photography technology and the popularization of related equipment, underwater photography is increasingly used in fisheries such as fish school monitoring and fish resource forecasting. The automatic processing of underwater video images greatly saves time and cost. However, this work faces many challenges, such as complex underwater background, distortion, low contrast, low resolution, and insufficient light propagation details. Fish target detection brings difficulties.

Due to the particularity of the underwater target detection task, the existing detection algorithms mostly rely on the gray information of the image. For example, [5] proposed an artificial target detection algorithm based on unconstrained underwater video, which mainly uses contour information. Complete target detection, but the detection speed is slower. In the literature [6], an underwater target detection algorithm based on region segmentation is proposed. This algorithm is not only robust, but also has a certain improvement in detection speed. But when the object is rotated or zoomed, the detection effect will be obviously affected. Literature [7] applies invariant moments to target detection, which solves the problems of translation and scaling. However, in the discrete case, the detection results are easily affected by the scaling factor. In response to the above problems, literature [8] proposes an underwater target detection algorithm based on invariant moments. This algorithm uses the minimum cross entropy to determine the threshold, which can ensure the integrity of the gray information, thereby achieving the fuzzy enhancement of the image; for the problem of uneven illumination in the underwater image, the algorithm uses Gray-gradient invariant moments to realize the segmentation of underwater images, which has better robustness and higher recall rate. Literature [9] proposed the use of visual saliency for underwater video detection and tracking, the difference between a single frame image and the mean value between its adjacent frames, and then the use of a saliency algorithm for target detection, which has achieved a relatively large breakthrough in speed, but the accuracy still cannot meet the expected requirements.

Compared with traditional machine learning methods, deep learning methods have great advantages in feature extraction. In recent years, the development of deep convolutional neural network algorithms has significantly improved the accuracy of target detection, and can self-learn the characteristics of the data itself without any prior knowledge, and automatically extract it layer by layer [10-12]. When the number is deep enough, the data can be deeply mined to extract high-quality features with resolving power. In recent years, through layer-by-layer learning, deep learning, especially convolutional neural networks (convolutional neural networks, CNN), has made breakthroughs in the field of target detection.

Although the underwater target detection algorithm based on deep learning has certain advantages in accuracy and speed, it also has some limitations that cannot be ignored in the face of complex underwater images. Due to the diversity of underwater target shape and scale, based on anchor, it is difficult for the deep learning algorithm of the point frame to obtain a high recall rate. In addition, the underwater target morphology is quite different, and the learning difficulty of the characteristics of different types of samples is quite different, which will also affect the target detection effect and increase the instability of the model. To this end, this article first enhances the image data to get a better contrast, and then uses the deep learning algorithm to separate the target and the background to improve the detection performance of the target.

2. DEEP CONVOLUTIONAL NETWORK
Deep CNN is a type of neural network with a special structure. "Depth" indicates that it has a large number of layers, usually more than 15 layers. The first few layers of the entire network are convolution layers and pooling layers. Alternate structure, and then several layers are fully connected
layers. The basic principle is: first learn different features by the convolutional layer, and then converge the spatial shape into the high-dimensional feature space by the pooling layer [13-15]. Multi-layer alternating convolution and pooling can learn hierarchical feature representations. Finally, a classifier is learned by the fully connected layer in the high-dimensional feature space.

As shown in Figure 1, all nodes of the convolutional layer and the pooling layer in the deep CNN are arranged into a series of 2-dimensional arrays, called "feature maps". In the convolutional layer, the input of each hidden layer node only contains nodes in one local neighborhood in the first layer [16]. The nodes in the first layer in the local neighborhood are multiplied by a weight matrix, and then through a non-linear activation function "regularized linear unit" ReLU, the result of the operation is used as the output value of the node of the convolutional layer. Each hidden layer node can be regarded as a feature detector, because when a certain feature it represents appears in its input, the node has a larger response value [17]. All nodes on the same feature map are restricted to share the same connection weight, so each feature map detects the same feature at different positions of the image. Due to local connection and weight sharing, the number of independent parameters that need to be learned from data in deep CNN is greatly reduced [18-20]. In the next pooling layer, each pooling layer feature map corresponds to 1 convolutional layer feature map. Each node of the pooling layer takes a node in a local neighborhood in the adjacent previous convolutional layer as input, and then down-sampling. The usual down-sampling method is to only keep the maximum value of all nodes in a local neighborhood, and ignore the rest of the node values [21]. A deep CNN contains many combinations of convolutional layers and pooling layers. After obtaining various features of the image from the global to the details through the alternating convolutional layer and the pooling layer, the deep CNN realizes the final classification result output through the fully connected layer.

![Deep CNN architecture](image_url)

Figure 1. Deep CNN architecture

The great success of deep CNN in the industry is due to: 1) the improvement of the algorithm; 2) the acquisition of massive data; 3) the popularization of high-performance computing resources such as graphics processing units, and the improvement of the algorithm is the leap forward of the deep CNN. The key factor of development.
3. DEEP CNN-BASED UNDERWATER TARGET DETECTION

3.1. STFT and Image Enhancement
The time-frequency characteristic spectrograms used in this paper are all obtained by short time Fourier transform (STFT). The short-time Fourier transform is currently the most commonly used mathematical method to reflect the characteristics of signal time-frequency distribution. Usually, the process of converting the time domain signal to the frequency domain is completed by Fourier transform. The inner product form of the Fourier transform is:

\[ S(f) = \langle s(t), e^{j2\pi ft} \rangle \] (1)

It can be seen from the above formula that the Fourier transform is a global transformation of the signal in the entire time domain. It is not suitable for analyzing non-stationary random signals related to both the time and frequency domains, nor is it suitable for analyzing the local performance of the signal in a certain period of time [22]. Therefore, it is necessary to use the two-dimensional joint representation of the time domain and the frequency domain to obtain the time-frequency characteristics of the signal. The idea of the short-time Fourier transform is to treat the non-stationary process as the superposition of several short-term stationary processes, and realize the estimation of the non-stationary process by analyzing the short-term stationary processes one by one. Its specific method is: using a certain window function, sliding continuously from the beginning of the time axis to the end of the time axis, each time the signal in the window is intercepted and the Fourier transform is performed to generate the frequency spectrum of the time period in the window, and finally the signals from each segment are obtained. Time-varying spectral array formed.

The local enhancement of the spectrum image refers to the enhancement of image visual effects and enhancement of image quality by improving image contrast, brightness, and clarity. The contrast limited adaptive histogram equalization (CLAHE) algorithm is an integration and improvement of the histogram equalization (HE) algorithm and the adaptive histogram equalization (AHE) algorithm, and the underwater image local enhancement effect is better. Therefore, in this paper, the CLAHE algorithm is used to locally enhance the image to pave the way for the subsequent target detection.

For a discrete signal sequence \( x(n) \), the data intercepted by the window function at time \( n \) is:

\[ x_w(m) = x(m)w(n-m) \] (2)

In the formula, \( w(n) \) is a window function, and its energy is generally concentrated near \( m \) or its center point \( m=0 \). The Fourier transform of this sequence is:

\[ X_{STFT}(n, f) = \sum_m x_w(m)e^{-j2\pi nf} = \sum_m x(m)w(n-m)e^{-j2\pi nf} \] (3)

3.2. Target Detection
The target detection algorithm based on image enhancement and deep network is to first enhance the image data to obtain better contrast, and then use the deep learning algorithm to separate the target and the background to improve the detection performance of the target. The detection principle framework is shown in Fig. 2 as follows.

![Figure 2. The target detection framework](image-url)
4. RESULTS
In order to verify the superiority of the proposed algorithm, the following experiments are designed in this paper. First, the images of underwater flexible targets are collected for annotation, and an experimental data set is made. This paper selects 6 underwater biological models as the detection targets, and then verifies the accuracy and real-time performance of underwater target detection by comparing the training and test results under different network models. In the experiment, a total of 1133 pictures of 6 kinds of underwater flexible targets were collected for labeling, and then divided into proportions, including 227 test data and 906 training data. In the experiment, first perform the pre-training of the image classification model for the basic network of each target detection. The training and test data sets are Imagenet data set, the learning rate is set to 0.05, the training method is multistep, the training batch size is 14, the maximum number of training times max_iter is 600000, and a test is performed every 10000 trainings. Experimental results show that the proposed algorithm can achieve 90% detection probability. Table 1 shows the detection performance of each algorithm under different signal-to-noise ratios (SNR).

| SNR (dB) | CNN    | Energy |
|---------|--------|--------|
| -12     | 0.0151 | 0      |
| -11     | 0.0171 | 0      |
| -10     | 0.0154 | 0      |
| -9      | 0.0208 | 0      |
| -8      | 0.025  | 0      |
| -7      | 0.038  | 0      |
| -6      | 0.051  | 0      |
| -5      | 0.0769 | 0      |
| -4      | 0.109  | 0      |
| -3      | 0.167  | 0      |
| -2      | 0.234  | 0      |
| -1      | 0.313  | 0      |
| 0       | 0.415  | 0.0003 |
| 1       | 0.571  | 0.0009 |
| 2       | 0.609  | 0.0021 |
| 3       | 0.782  | 0.0457 |
| 4       | 0.876  | 0.1031 |
| 5       | 0.839  | 0.1526 |
| 6       | 0.974  | 0.2261 |
| 7       | 0.934  | 0.2301 |
| 8       | 0.971  | 0.491  |
| 9       | 0.977  | 0.603  |
| 10      | 0.99   | 0.781  |

Table 1 shows the detection performance results of different algorithms. It can be seen from the results that compared with the classic energy detector, the performance of CNN has been greatly improved. The above detection results fully illustrate the performance advantages of the proposed algorithm.
5. CONCLUSION
In this paper, we have proposed a target detection algorithm based on image enhancement and deep network. Image enhancement is to increase the contrast of the image and make the target and background information more obvious. With the help of the marked target and background data, the deep learning network is used for training, and the detection network is obtained. Simulation experiments have proved the performance advantages of this method.

REFERENCES
[1] Kattawar G W, MJ Raković. Virtues of Mueller matrix imaging for underwater target detection[J]. Applied Optics, 1999.
[2] Tucker J D, Azimi-Sadjadi M R. Coherence-Based Underwater Target Detection From Multiple Disparate Sonar Platforms[J]. IEEE Journal of Oceanic Engineering, 2011, 36(1):37-51.
[3] Dubreuil M, Delrot P, Leonard I, et al. Exploring underwater target detection by imaging polarimetry and correlation techniques[J]. Applied optics, 2013, 52(5):997-1005.
[4] Maire F D, Prasser D, Dunbabin M, et al. A Vision Based Target Detection System for Docking of an Autonomous Underwater Vehicle. 2009.
[5] Hasan, Mohammed, A, et al. A modified block FTF adaptive algorithm with applications to underwater target detection.[J]. IEEE Transactions on Signal Processing, 1996.
[6] Gunes A. Joint underwater target detection and tracking with the Bernoulli filter using an acoustic vector sensor[J]. Digital Signal Processing, 2016, 48(C):246-258.
[7] Jezequel V, Audo F, Pellen F, et al. Experimentally based simulations on modulated lidar for shallow underwater target detection and localization[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2010, 7825(2).
[8] Mukherjee, K, Gupta, et al. Symbolic Analysis of Sonar Data for Underwater Target Detection[J]. IEEE Journal of Oceanic Engineering: A Journal Devoted to the Application of Electrical and Electronics Engineering to the Oceanic Environment, 2011.
[9] Li X, Jia H, Mei Y. Underwater target detection based on fourth-order cumulants beamforming[J]. Journal of the Acoustical Society of America, 2017, 142(4):2524-2524.
[10] Nour A, Fabrice P, Le B G, et al. Extra-cavity radiofrequency modulator for a lidar radar designed for underwater target detection[J]. Applied Optics, 2017, 56(26):7367.
[11] Wei Z, J Yang, Min S. A Method of Underwater Acoustic Signal Classification Based on Deep Neural Network[C]// 2018 5th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2019.
[12] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization[J]. International Journal of Computer Vision, 2020, 128(2):336-359.
[13] Szegedy C, Toshev A, D Erhan. Deep Neural Networks for object detection[C]// Advances in neural information processing systems. 2013.
[14] Hodo E, Bellekens X, Hamilton A, et al. Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey[J]. 2017.
[15] Jung S I, Hong K S. Deep network aided by guiding network for pedestrian detection.[J]. Pattern Recognition Letters, 2017, 90:43-49.
[16] Song Y, H Zhang, Liu L, et al. Rail Surface Defect Detection Method Based on YOLOv3 Deep Learning Networks. IEEE, 2019.
[17] Kurdthongmee W. Accelerate the Detection Frame Rate of YOLO Object Detection Algorithm[J]. Springer, Cham, 2019.
[18] Yip M C, Gonzalez M M, Valenta C R, et al. Deep learning-based real-time detection of neurons in brain slices for in vitro physiology[J]. Scientific Reports, 2021, 11(1).
[19] Etxeberria-Garcia M, Ezaguirre F, Plazaola J, et al. Embedded object detection applying Deep Neural Networks in railway domain[C]// 2020 23rd Euromicro Conference on Digital System Design (DSD). 2020.
[20] Shafiee M J, Chywł B, Li F, et al. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video[J]. Journal of Computational Vision and Imaging Systems, 2017, 3(1).

[21] Carolis B D, Ladogana F, Macchiarulo N. YOLO TrashNet: Garbage Detection in Video Streams[C]// 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE, 2020.

[22] Sirmaek B, Unsalan C. Urban-Area and Building Detection Using SIFT Keypoints and Graph Theory[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(4):1156-1167.