

Underwater target recognition methods based on the framework of deep learning: A survey

Bowen Teng¹ and Hongjian Zhao

Abstract

The accuracy of underwater target recognition by autonomous underwater vehicle (AUV) is a powerful guarantee for underwater detection, rescue, and security. Recently, deep learning has made significant improvements in digital image processing for target recognition and classification, which makes the underwater target recognition study becoming a hot research field. This article systematically describes the application of deep learning in underwater image analysis in the past few years and briefly expounds the basic principles of various underwater target recognition methods. Meanwhile, the applicable conditions, pros and cons of various methods are pointed out. The technical problems of AUV underwater dangerous target recognition methods are analyzed, and corresponding solutions are given. At the same time, we prospect the future development trend of AUV underwater target recognition.

Keywords

Deep learning, AUV, dangerous target recognition, few-shot target recognition, environmental interference

Date received: 13 September 2020; accepted: 4 November 2020

Topic Area: Mobile Robots and Multi-Robot Systems

Topic Editor: Antonio Fernandez-Caballero

Associate Editor: Zhenxue Chen

Introduction

With the development of technology and the continuous growth of military power, countries around the world are shifting their military priorities to the ocean. Limited by the natural conditions of the ocean and the physical limits of human beings, it is obviously impossible to exploit marine resources solely by humans. Autonomous underwater vehicle (AUV) is capable of performing underwater tasks independently. Therefore, AUV equipped with visual image acquisition equipment is often used for real-time detection in the underwater environment, which has also made autonomous underwater robots widely used in military fields, for example, mine detection, intelligence collection, and off-shore defense.

In 1997, based on the neural network classifier k -nearest neighbor attractor and the optimal discriminatory filter classifier, Naval Surface Warfare Center (Dahlgren,

Virginia, USA) extracted and classified the features of each detected minefield.¹ This method reduced false alarms and laid the foundation for deep learning in underwater dangerous target recognition. In 2003, for improving the efficiency of clearing landmines and unexploded ordnance, Carnegie Mellon University proposed a new method to deal with sensor uncertainty, which uses geometrical and topological features instead of sensor uncertainty models.² Therefore, it speeds up the demining process. To reduce

Industrial robotics engineering division, Beijing Research Institute of Automation for Machinery Industry Co., Ltd, Beijing, PR China

Corresponding author:

Hongjian Zhao, Industrial robotics engineering division, Beijing Research Institute of Automation for Machinery Industry Co., Ltd, Beijing 100120, PR China.

Email: zhaohj95@163.com



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

the impact of the complex and changeable underwater environment, Cao et al.³ proposed a method named stacked autoencoder (SAE)-softmax that joint SAE and softmax classifying underwater targets. This method yielded the highest recognition rate up to 94.12% out of radial basis function support vector machines (SVM) and probabilistic neural network (PNN) methods. Based on transfer-reinforcement learning, Cai et al.⁴ proposed a multi-AUV target recognition approach, which reduces the impact of complex environment while improving the efficiency of target recognition. The average recognition accuracy is 82.82% out of other six methods in the case of turbid water quality, object occlusion, insufficient light, complex background, and overlapping targets.

Recently, deep learning has gained great attention in the field of target recognition. The recognition accuracy is improved through a training model on a large number of samples. However, the underwater target recognition technology is different from the land or air target recognition method. It is affected by the scattering effect of the water medium on the target information, resulting in the blurred or distorted target information. At the same time, affected by the complex underwater environment, such as time-varying ocean currents, uneven illumination, turbid water quality, and so on, it is difficult to collect target image information. Moreover, the appearance and shape of hostile dangerous targets are diverse, resulting in insufficient samples for training model and reducing recognition accuracy.

At present, for the lack of dataset, transfer learning⁵⁻⁷ can be used to train model on a dataset with a large number of land or air targets and then transfer the model to the underwater target field. Generative adversarial network (GAN)^{8,9} is a new method, which can autonomously generate underwater target images to increase the number of samples. Image preprocessing, image restoration, and reinforcement learning also can be used to reduce the impact of the underwater environment interference.

Most researchers conducted in-depth research on image processing to improve the accuracy of target recognition. However, the information extracted from a single image is limited. Aiming at the insufficient acquisition of target information, Cai et al.¹⁰⁻¹² introduced multiview light field reconstruction into the target recognition field. The target information can be collected through multiple views,¹³ that is, multi-AUV is used to recognize underwater dangerous targets. Luo et al.¹⁴ introduced the GAN network to the field of multi-AUV target recognition, which not only increased the accuracy of target recognition but also reduced the impact of underwater complex environment.

The above methods and studies mainly introduce deep learning algorithm into the field of underwater target recognition. But various difficulties are encountered in the process of recognizing dangerous targets in a real underwater environment, such as environmental problems, interference problems, information collection problems, and sample information missing. In this article, we will mainly

discuss about different types of underwater dangerous target recognition technology, summarize the existing methods, sum up the problems and technical difficulties of various technologies in the process of underwater dangerous target recognition, and look forward to the future development direction.

Underwater dangerous target recognition technology

Mine recognition technology

Mines are widely used in modern naval battles and play an important role. Mines can not only strike submarines and block maritime traffic routes but can also cause serious psychological burdens on enemy personnel. The application of advanced technology in mine weapons makes modern mines more concealed and intelligent. It is very difficult to accurately find and eliminate them in the vast sea area.

The automatic mine recognition is the current development trend.¹⁵ In 1997, Shin et al.¹⁶ proposed a method of integrated wide-band compression and mine detection in shallow water areas. This method combines the target recognition algorithm and image compression to achieve excellent detection performance while minimizing the computational complexity of the algorithm. Gleckler and Fetzer¹⁷ used an integration method of an underwater laser rangefinder and a digital camera to detect and measure the mine information. It can locate dangerous targets and recognize them. Miao et al.¹⁸ introduced an approach of mine target recognition based on basic vision. This idea came from the essential shape characteristics of the mine target. According to the physical meaning of the geometric moment, it combines the regional feature and the boundary feature to construct three shape descriptors suitable for the mine target. It uses the threshold judgment method to realize the mine target recognition. This method has higher accuracy (more than 94%) and better stability than the method based on moment invariants. It is more suitable for the recognition of underwater targets with specific shapes and the circumstance when the targets are partially occluded.

With the development of deep learning in the target recognition field, the recognition accuracy is getting higher and higher, which has become one of the main methods of target recognition. Some researchers¹⁹ used unsupervised processing technology to detect mine-like targets on the collected image. The AUV equipped with sonar detection equipment is used to detect the changes in image texture and image intensity in the area so as to determine the mine target buried under the sea, as shown in Figure 1. Although this method can detect mine targets through unsupervised training, the error rate is relatively high.

Williams and Fakiris²⁰ constructed a set of classifiers, controlling the relative importance of each target in the learning phase of a given classifier through the modulation

factor. They inferred the quantity of classifiers and all the other relevant model parameters from the training dataset automatically. This method improves the utilization rate of underwater target information and significantly improves the accuracy of target recognition, as shown in Figure 2. To extract multilayer features from sonar image, Guo and Chen²¹ proposed the naïve Bayes Poisson gamma belief network (PGBN) model based on PGBN and Bayes'

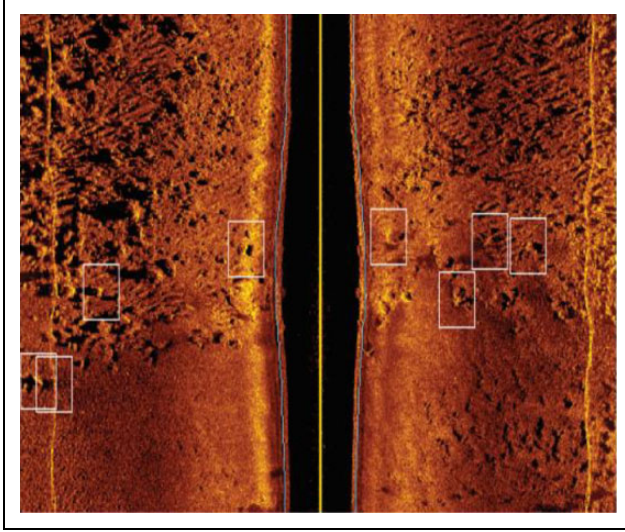


Figure 1. Mine target recognition based on sonar image. The figure shows a vehicle turn and two mine-like features.

theorem, which improve the training efficiency of the model. Moreover, the recognition accuracy can reach 93.85%, which is better than PGBN and other models, such as three-layer restricted Boltzmann machine, similarity deep belief network (DBN), DBN, SVM, and kernel SVM.

In the process of underwater target detection, how to reduce background interference is also extremely challenging. Based on the unsupervised network, Xie et al.²² proposed a feature extraction approach to extract the mine intrinsic attributes. They constructed a spectrum regularization unsupervised network (SRUN) to distinguish target information from background information. Target detection is not only based on image features but also on the basis of the difference between the known target spectrum and the collected information pixel spectrum. Figure 3 shows the schematic of the proposed approach, which comprises the following steps. First, the SRUN was proposed to extract compact features in hyperspectral images. Then, the effective nodes are selected and further weighted adaptively. Finally, the background information is suppressed to gain the detection map. Experimental results on several datasets indicate that the proposed SRUN-based target detection algorithm is more suitable for targets at the sub-pixel level and those with structural information.

To increase the accuracy of mine target recognition, Giovanneschi et al.²³ proposed drop-off minibatch decentralized online dictionary learning. It takes an advantage of the fact that a large number of the training data may be

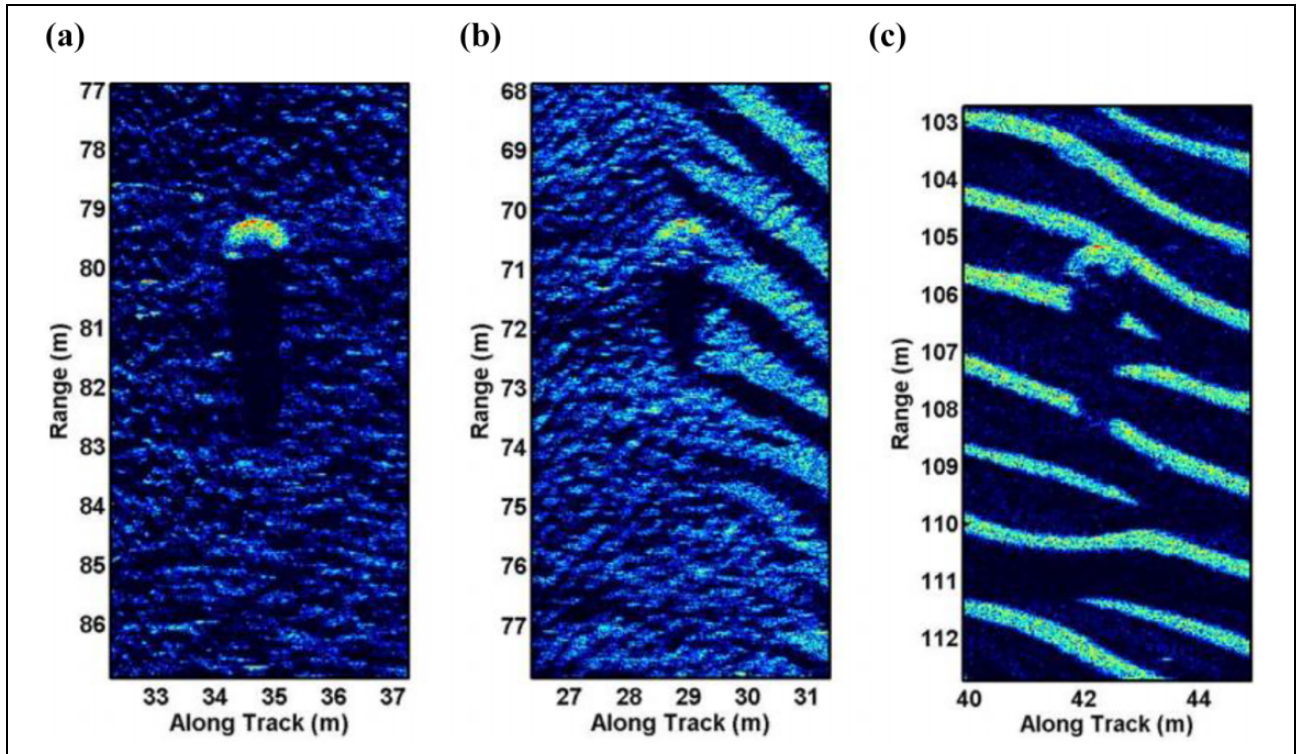


Figure 2. Example synthetic aperture sonar image chip of a truncated-cone-shaped target (a) on the seabed, (b) on the board of the flat seabed and ripples, and (c) on seabed characterized by sand ripples.

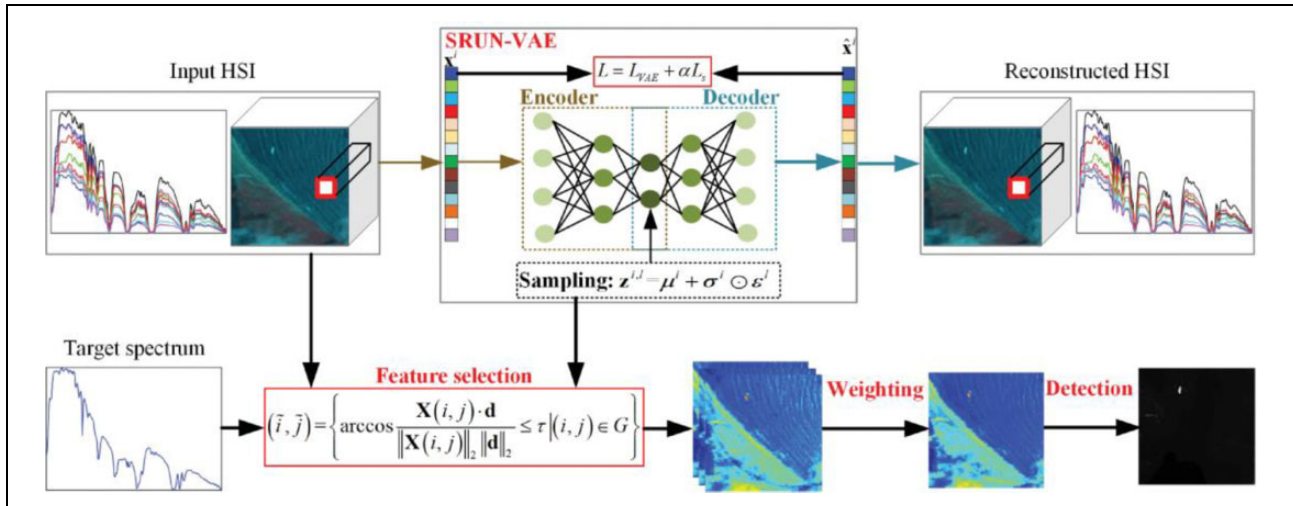


Figure 3. Schematic of the SRUN-based target detection method. SRUN: spectrum regularization unsupervised network.

correlated. With this method, they trained the model on a small batch iterative manner and deleted samples that are no longer relevant. This method is faster and retains similar classification performance as the classical online dictionary learning and online dictionary learning correlation-based variant methods.

Most of the above methods are based on the physical meaning of geometric moments to recognize targets. But the shape characteristics of the mines are more prominent. Researchers can train a better model by combining the description area and boundary characteristics to establish a descriptor suitable for the shape of the mine targets or using the deep learning algorithm. In the future, researchers should focus on improving the accuracy of various shapes of mine target recognition as well as excellent anti-interference ability and timeliness.

Underwater manmade target recognition technology

At present, in addition to lethal mines, there are also many manmade devices with detection, inspection, and strike capabilities. How to accurately recognize underwater manmade equipment is one of the current key research directions. Olmos et al.²⁴ proposed an approach for detecting the manmade targets in unconstrained underwater videos. This algorithm can only detect targets with known contours. But when the image quality is poor, it directly reduces the recognition accuracy of the target.

In recent years, scholars used deep learning technology for underwater target recognition, which can improve target recognition accuracy and recognize more types of targets. For the purpose of reducing the impact of different environments on target recognition, Parma University used multiple datasets to study the potential of vision-based target detection algorithms in underwater scenes.²⁵ Through the training of multiple datasets, the algorithm can accurately recognize targets in different underwater environments

and provide new ideas for subsequent research on multidata information fusion. Yu et al.²⁶ built a model composed of five convolutional layers and three fully connected layers based on convolutional neural network (CNN) deep learning theory. In the training procedure, both labeled in-air images and unlabeled underwater images are used to train the model. In the last two layers, the maximum mean distance feature metric is added to regularize. This method shows good robustness when recognizes underwater targets, with a recognition accuracy of up to 55.07%. The specific process is shown in Figure 4.

In the underwater target recognition procedure, accurate extraction of target feature information is the main factor that affects the recognition accuracy. Hussain and Zaidi²⁷ deblurred the image by reducing the noise in the image and predicting the Euclidean shape. Ma et al.²⁸ extracted the targets of interest in underwater images by applying color-based algorithms. Then, they used the improved two-dimensional (2D) Otsu algorithm to remove the background color noise. Furthermore, a robust algorithm based on shape signature was used to recognize the shape type of a regular object. The experimental results indicate an ideal outcome with an average recognition rate of shape type (approximately 90%). For the purpose of improving the real-time performance of the underwater target recognition algorithm, Qing et al.²⁹ proposed a new method based on wavelet transform and improved Hough transform. According to the experimental results, the proposed algorithm can accurately detect the straight lines that existed on manmade objects in complex underwater background. It has excellent real-time performance, that is, only 17.22 ms per image of the best result.

In the process of recognizing dangerous targets, it not only needs the target be accurately recognized but also needs to calculate the target's status information, such as position, movement direction, and travel speed. Chen and Xu³⁰ established a DBN model and a stacked denoising

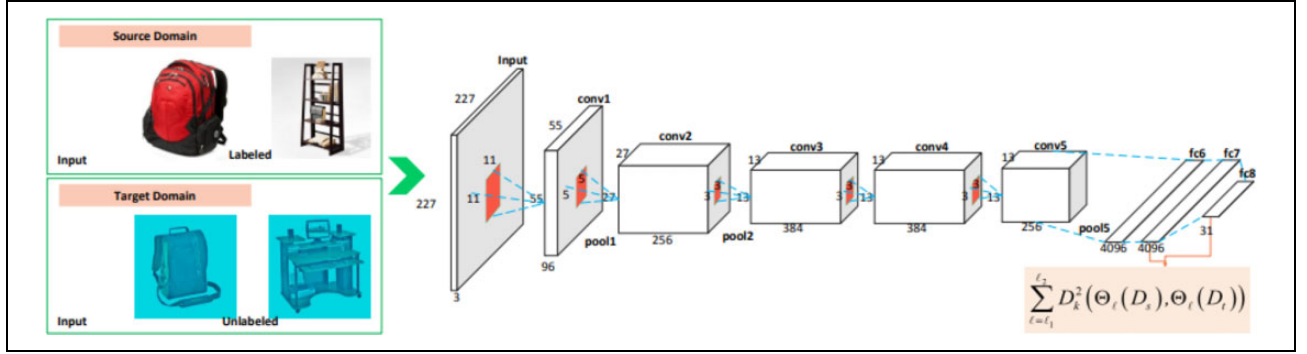


Figure 4. The training process of CNN-based target recognition. Conv means the convolutional layer and fc means the fully connected layer. CNN: convolutional neural network.

Table 1. The experiments results of different algorithms.

Experiment	SVM	GRNN	PNN	DBN	SDAE
Underwater acoustic data of three kinds of target	96.2%	94.2%	92.5%	96.8%	98.2%
The same target at different navigation states	90.4%	90.2%	87.6%	92.2%	92.1%
Experiment data of different states of one target	88.6%	86.2%	84.8%	90.5%	91.8%

SVM: support vector machine; GRNN: general regression neural network; PNN: probabilistic neural network; SDAE: stacked denoising autoencoder.

autoencoder model. They compared the underwater acoustic simulated data of different types of targets and different states of one target and experimental data of different states of one target recognized by DBN and other models (SVM, GRNN, PNN, and SDAE). Table 1 presents the detail of experiments results.

The introduction of deep learning technology promotes the development of underwater target detection research. On this basis, numerous researchers have proposed more powerful models. All of the above algorithms are used for specific application scenarios, which have certain universalities but also have limitations. We summarize the advantages and disadvantages of the above algorithms here, as given in Table 2.

Few-shot target recognition

Due to the diverse shapes of underwater artificial devices in the real environment, it is hard to collect target images and train a satisfactory model. These factors lead to low target recognition accuracy in the real environment. At present, transfer learning theory can effectively transfer the source domain training model to the target domain. Because of the convenience of collecting samples on land and in the air, the trained model can be transferred to underwater targets by training on existing targets. Based on this theory,

Table 2. Comparison of underwater dangerous target recognition methods.

Methods	Advantages	Disadvantages
Target recognition based on shape feature	Simple algorithm and fast recognition speed	Affected by known information, the anti-interference ability is weak
Unsupervised recognition technology	High recognition accuracy	The recognition accuracy is limited by the quantity and quality of the training dataset
Deep learning theory based on CNN	High recognition accuracy, reducing the need for training samples	The training process is complicated and the preparation process takes a long time

CNN: convolutional neural network.

Xiamen University integrated deep learning and transfer learning to recognize underwater manmade targets.²⁶ This method is superior to traditional methods in underwater manmade target recognition tasks. It is suitable for long-term research and development.

Based on a cycle-consistent adversarial network and a conditional generation adversarial network, Li et al.³¹ proposed a trainable end-to-end system of an underwater multistyle generation adversarial network to solve the problem of fewer underwater image dataset. The system can generate diverse underwater images from aerial images using hybrid countermeasures and unpaired methods. Chen et al.³² proposed a new two-level feature alignment method. With it, a typical deep domain adaptation network can deal with the domain shift problem between two modalities in data generating process. For evaluating the quality of the generated images, Liu et al.³³ used similarity values, structural similarity index, and multiscale structural similarity index to calculate the color and structure similarity level. Rao et al.³⁴ introduced a multimodal model, which can complete the recognition task based on experience in

Table 3. Comparison of different methods for few-shot underwater target recognition.

Methods	Advantages	Disadvantages
Transfer learning	The land and air targets have a lot of samples. They can be smoothly transferred to the target domain.	After the transfer to the target domain, the model is fixed and is greatly affected by the existing target domain.
Supervised learning	The recognition accuracy is high and overfitting is not easy to occur.	Model parameters are difficult to obtain. The model is less flexible and sensitive to abnormal samples.
Generative adversarial network	There is no need to design a model that follows any kind of factorization. The generator network and the discriminator network will automatically adjust the network.	No need to premodel. The model is too unstable to control.

the case of fewer training sample images. The method proposed by Cho et al.³⁵ can generate simulated images through simple underwater images, which makes target recognition more accurate. To compute the similarity between the template image and sonar image, they define a correlation array of $S_j(i)$ and $T_k(i)$ as

$$R_{j,k}(i) = \sum_{\lambda=1}^{r_t} S_j(i + \lambda) T_k(\lambda) \quad (1)$$

where $R_{j,k}(i)$ is the correlation array for $1 \leq i \leq r_s - r_t + 1$.

The problem of few-shot image recognition can be solved not only by generating new samples but also by transfer learning. Jin and Liang³⁶ proposed a framework for underwater few-shot image recognition based on transfer learning. Firstly, an improved median filter was used to suppress the noise of fish images. A classical operation is used to describe the denoising results quantitatively. The peak signal-to-noise ratio (PSNR) for RGB images is computed using the standard formula

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{(3|X|)^{-1} \sum_{c=R,G,B} \sum_{x \in X} (y_c(x) - \hat{y}_c(x))^2} \right) \quad (2)$$

where x is a 2D spatial coordinate that belongs to the image domain $X \subset \mathbb{Z}^2$, the subscript $c \in \{R, G, B\}$ denotes the color channel, y is the original image, and \hat{y} denotes the denoised image. The larger the PSNR, the better the denoising.

Table 4. Comparison of five methods on different datasets.^a

Models	CSDN	VGG-DBN	SSD-DBN	RFCN-DBN	SCDAE-CNN
Dataset A	93.34%	89.83%	77.90%	83.61%	67.86%
Dataset B	92.27%	89.16%	78.27%	82.43%	67.59%

CSDN: compound convolutional neural network; VGG: visual geometry group; DBN: deep belief network; SSD: single shot multibox detector; RFCN: region-based fully convolutional network; SCDAE: stacked convolutional denoising auto-encoder; CNN: convolutional neural network.

^aDataset A is collected in the Philippine Sea, includes the air gun samples and bomb samples with depths of 50 and 220 m. Dataset B is collected in the South China Sea, only contains bomb samples with the depths of 7, 50, and 300 m.

Then, the neural network is pretrained by the ImageNet that is the largest image recognition database in the world. Finally, they used the preprocessed target image to fine-tune the pretrained neural network. Thus, the recognition accuracy on the test dataset reaches 85.08%, which has made a significant improvement.

Traditional point-based feature methods often perform poorly because of biofouling, corrosion, and other effects that lead to dramatic changes in target visual appearance. Li et al.³⁷ used supervised learning to relearn the target and combined the particle filtering framework to automatically recognize the target. The solutions for few-shot target recognition are given in Table 3.

Target recognition under environmental interference conditions

Due to the harsh underwater environment, the quality of the collected target images is poor. The change of target state and the object shelter also has a huge impact on the target recognition procedure. Zhou et al.³⁸ introduced a compound convolutional neural network based on shared latent sparse feature and DBN. Experimental result shows that this approach is more stable for different dataset and has the highest accuracy of up to 93.34%. Experimental result is presented in Table 4.

To effectively recognize targets of different depths and reduce radiation noise, Yang et al.³⁹ combined deep long short-term memory network (DLSTM) and deep autoencoder neural network together. They used pretrained DLSTM model and softmax classifier to classify ship radiation noise. Based on the long short-term memory network, Zhang and Xing⁴⁰ proposed a novel method, which integrates multiple feature data and softmax classifier to effectively remove underwater noise interference. In multiple experiments, the best results reach the accuracy of 97%. The feature fusion schematic is shown in Figure 5.

In the underwater target recognition process, the light intensity changes greatly as the depth increases. When the illumination of the target surface is uneven, shadows will be generated, which will cause a part of appearance

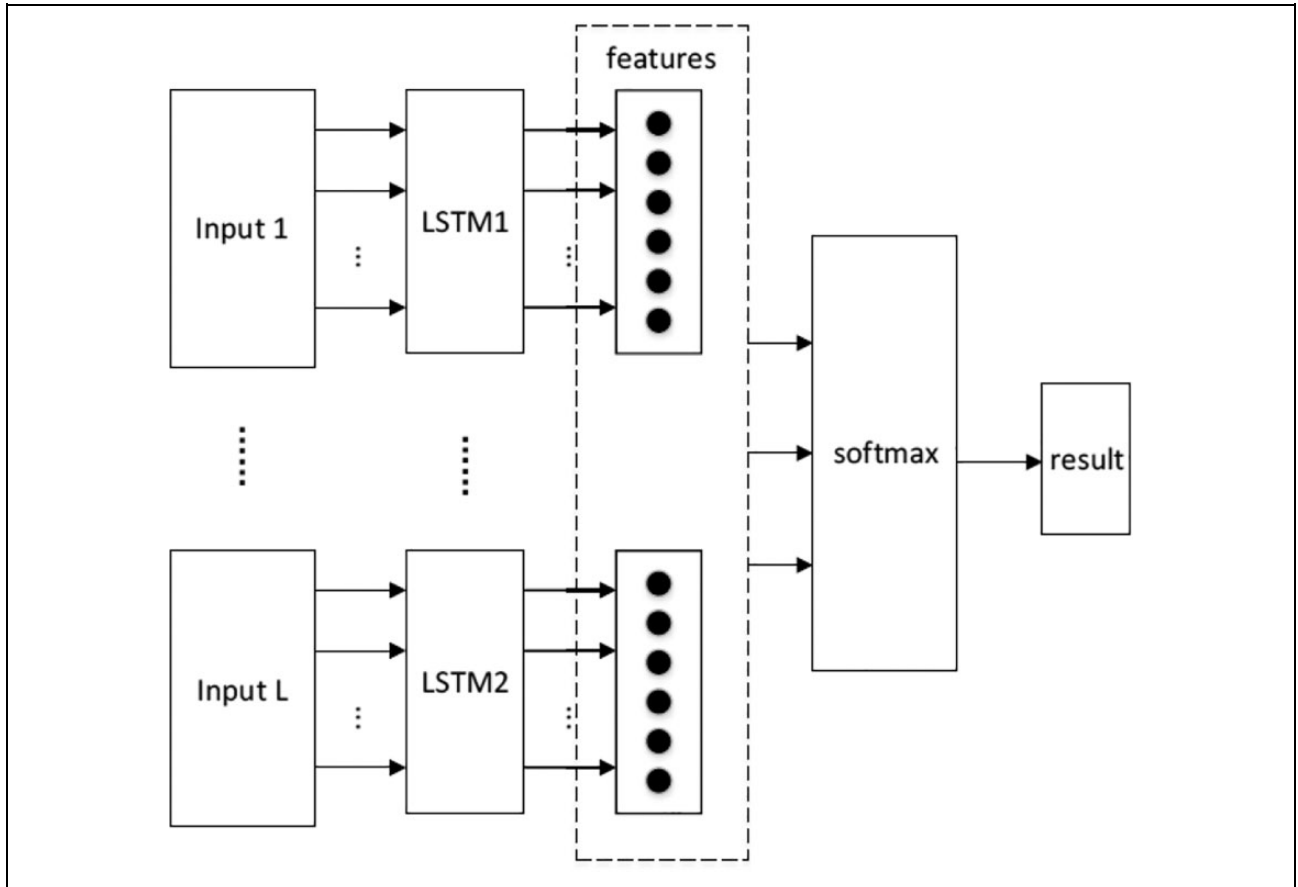


Figure 5. Mutilclass feature fusion recognition.

information loss. Zhang and Negahdaripour⁴¹ combined shadow information to reconstruct the shape of three-dimensional targets to minimize the impact of shadows. Song et al.⁴² used the AUV equipped with visual image acquisition equipment to compensate the target with different light intensity so that the algorithm could extract the image features and color features of the target image. In this way, it can reduce the influence of uneven illumination on target information.

Aiming at the shortcomings of traditional backpropagation (BP) neural network, such as slow convergence and tending to local minima, Tang et al.⁴³ proposed a novel approach of BP neural network design based on immune genetic algorithm. This algorithm overcomes the problems of genetic algorithm in search efficiency, individual diversity, and premature. It effectively improves the convergence performance.

Because of the turbidity, absorption, and scattering of the water, the images collected underwater become blurred, which greatly affects the accuracy of target recognition. To reduce these impacts, Li et al.⁴⁴ proposed an effective defogging model to restore the visibility, color, and natural appearance of underwater images. Ding et al.⁴⁵ proposed a new underwater image enhancement strategy combining adaptive color correction and model-

based defogging. By contrast with original underwater images, enhanced images help to reveal more feature points. This strategy effectively improves the quality of underwater images and makes the algorithm more accurate to recognize underwater target.

Due to the problems of low contrast, blue-green projection and low visibility, the captured underwater environment images appear green and blue.⁴⁶ This leads to distortion of the captured images. Ahn et al.⁴⁷ proposed a data enhancement method based on the principle of retina to promote the visibility of captured images. Chuang et al.⁴⁸ used feature learning technology and error-proof classifiers to preprocess the collected images to improve the image clarity. Zhang et al.⁴⁹ applied visual inspection to underwater image feature extraction. Before underwater image preprocessing, dark channel is applied to eliminate haze and enhance the contrast of underwater images. Robustness and real time of the algorithm have been greatly improved. Yu et al.⁵⁰ proposed a novel framework named underwater GAN for image restoration. It uses a convolutional patchGAN classifier to learn structure loss. Based on the underwater image generator model, a more realistic image is generated through simulation. The influence of abnormal image contrast in the target recognition process is reduced.

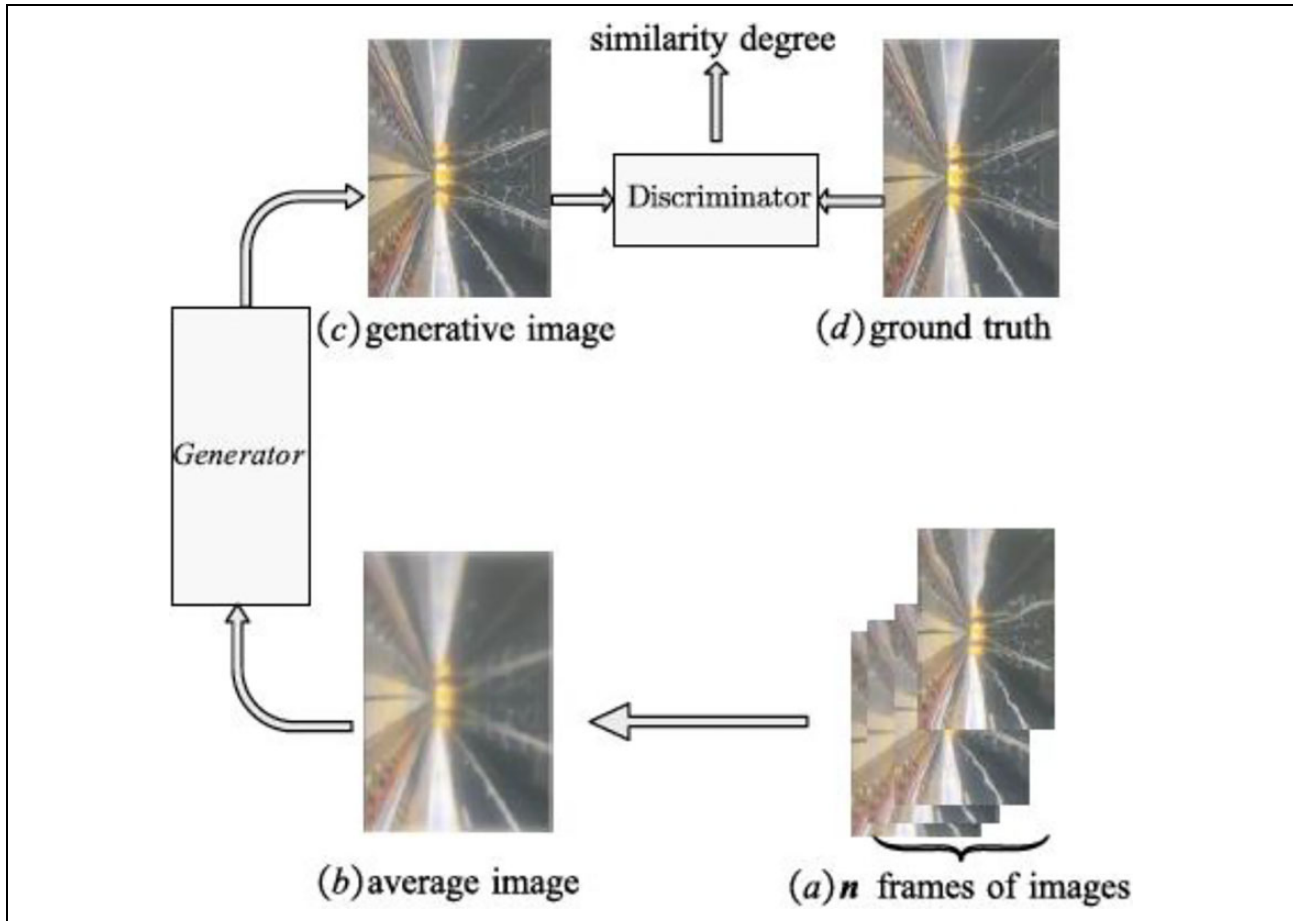


Figure 6. Network structure of He's method.

Since the underwater environment is accompanied by time-varying ocean currents, the signal obtained by the imaging sensor has a corresponding relationship with time. When the variable ocean currents cause fluctuations in the image refractive index in the imaging path, the task of target recognition is more difficult.⁵¹ Florida Atlantic University⁵² used compressed line sensing to reconstruct images after ocean current interference so that the imaging system can recover target information under various turbulence intensities. The network in the literature⁵³ refers to the network structure of Kupyn et al.⁵⁴ It restores the underwater distorted image sequence through GAN. Moreover, the training process is directed by the Wasserstein distance. Smaller the distance means higher similarity between real and fake image. This method can effectively restore the distorted images and make the images restoration degree higher. It reduces the impact of time-varying ocean currents on target information collection. The network architecture is shown in Figure 6.

In the future, multi-AUV underwater target recognition will inevitably develop in the direction of real-time, high accuracy, and autonomy. The accuracy of target recognition in complicated underwater environments needs to be further improved. Table 5 summarizes the classification of

target recognition methods to reduce the impact of the complex underwater environment.

Different algorithm under the same dataset

With the development of deep learning, target recognition technology has also made considerable progress. Deep learning has a strong learning ability. It can learn useful information in images from a large amount of training dataset and effectively detect objects in images. Since R-CNN⁵⁵ was proposed by Girshick, the field of target recognition has gained great attention and become an emerging research hotspot. Since then, many new models have been proposed, such as fast R-CNN,⁵⁶ faster R-CNN,⁵⁷ FPN,⁵⁸ YOLO,⁵⁹ SSD,⁶⁰ and so on. These algorithms have their own advantages and disadvantages. Some researchers apply them to the same dataset to verify the capabilities of different algorithms.

Wang et al.⁶¹ proposed a new underwater target detection dataset, called UDD, which contains three categories (sea cucumber, sea urchin, and scallop) with a total of 2227 images. YOLOv3, RetinaNet, and other networks were selected for comparison. The comparison results are given

Table 5. Summarize of target recognition method in complex underwater environment.

Problems	Research method	Principles and characteristics
Radiation noise	1) Filter processing	Through the noise reduction preprocessing of the collected information, it is suitable for the situation with low noise.
	2) Radiated noise modeling	The radiated noise was modeled by the neural network and then denoise based on the model information.
Uneven light	1) 3D object reconstruction	Through the known target information combined with shadow information to reconstruct the shape of the three-dimensional object, the accuracy of target recognition is increased.
	2) Illumination compensation	Filling up the target with lighting equipment is greatly affected by the distance of the light source under water.
	3) Algorithm fix	The algorithm is used to unify the brightness of the pixels of the collected image to reduce the impact of uneven illumination on the recognition algorithm.
Turbid water	1) Dehazing algorithm	According to the principle of minimum information loss and optical characteristics of underwater imaging, the influence of water turbidity on image quality is eliminated. The visibility of the image is increased.
	2) Image enhancement	Dehaze the image through adaptive color correction based on the atmospheric scattering model.
Low contrast	1) Contrast enhancement algorithm	Through the contrast enhancement algorithm, image artifacts are reduced and the target details are clearer.
Time varying ocean current	1) Image restoration	Using deep learning technology to use underwater distortion image sequence for image restoration.
	2) Imaging reconstruction	Through the reconstruction of the image after turbulence, the influence of time-varying ocean current on the target image is reduced.

in Table 6. To make a fair comparison, all models were trained from scratch with the same hyperparameters and data augmentation methods were used with the same parameter settings.

Table 6. Comparisons for different algorithm on UDD dataset.

Method	Backbone	Params	FPS	mAP
SSD ⁶⁰	MobileNetV2	3.05M	11	22.7%
YOLOv3 ⁶²	DarkNet-53	61.9M	32	46.8%
RetinaNet ⁶³	ResNet-18	19.81M	14	24.6%
RetinaNet ⁶³	ResNet-50	36.15M	10	34.2%
FCOS ⁶⁴	ResNet-50	31.84M	27	44.9%
Foveabox ⁶⁵	ResNet-50	36.02M	28	30.0%
FreeAnchor ⁶⁶	ResNet-50	36.15M	25	32.7
RPDet ⁶⁷	ResNet-50	36.6M	22	45.1%
GA-RetinaNet ⁶⁸	ResNet-50	37.15M	12	36.1%
CenterNet ⁶⁹	DLA-34 ⁷⁰	18.12M	33	36.6%

Table 7. Comparisons for different algorithms on URPC 2018 dataset.

Method	Sea urchin	Sea cucumber	Scallop	Starfish	mAP
Faster R-CNN ⁵⁷	58.4%	78.2%	27.1%	68.1%	58.0%
FPN ⁵⁸	61.7%	85.5%	33.9%	72.9%	63.5%
R-FCN ⁷²	66.4%	87.5%	40.3%	75.7%	66.5%
Deformable faster R-CNN ⁷¹	66.7%	86.4%	41.5%	76.3%	67.5%
Deformable R-FCN ⁷¹	89.9%	90.1%	73.5%	89.2%	85.7%

As presented in Table 6, CenterNet was the best performer at 33FPS, followed by YOLOv3 and Foveabox at 32FPS and 28FPS, respectively. In terms of accuracy, YOLOv3 performed best at 46.8%, followed by RPDet at 45.1% and FCOS at 44.9%. Overall, YOLOv3 performed best, ranking among the best in terms of accuracy and detection speed.

Underwater Robot Picking Contest in 2018 provided an underwater object target detection dataset, including sea urchins, sea cucumbers, scallops, and starfish. To test the detection effect of different algorithms, Zhang et al.⁷¹ tested the regular faster R-CNN, FPN, and R-FCN. They also tested faster R-CNN and R-FCN with deformable convolution. All the models used the same hyperparameters and were tested on the same computer. The experimental results are presented in Table 7.

In the traditional target detection algorithm, R-FCN has the best performance, with mAP reaching 66.5%, which is much better than faster RCNN and FPN. Moreover, compared with the original network, the performance of the network is improved after the new deformable convolution is adopted. The best performer was deformable R-FCN at 85.7%.

In recent years, the field of underwater target recognition has developed rapidly, and new algorithms are proposed every year. These algorithms have fast detection speed and high accuracy, but they are generally targeted

at specific application scenarios. The poor universality of algorithms is always a big problem in this field. In the future, the research direction should focus on developing the algorithm with strong universality.

Summary

With the continuous development of underwater weapons and equipment, researchers pay more attention to underwater safety issues. The underwater dangerous target recognition has become one of the focuses of research. Due to the vast sea area and the complex underwater environment, it is difficult to collect dangerous target images. To solve these problems, many scholars have conducted research on few-shot target recognition, such as transferring the trained models of aerial or land targets to the field of underwater dangerous targets through transfer learning, increasing the number of samples through reinforcement learning, and using GAN to perform dangerous target image generation. The target recognition accuracy of deep learning can be improved by increasing the training samples. Some scholars used methods, such as target reconstruction, image defogging, and image restoration, to reduce the actual underwater interference environment (such as uneven illumination, turbid water quality, and time-varying ocean currents, etc.) on the target image to improve the accuracy of target recognition.

With the development of the cluster system, multiple AUVs are used for collaborative work to collect target information from different angles and reduce the limitation of collecting information from a single perspective. The comprehensive utilization of various marine information can offset or reduce the impact caused by the special underwater environment of the ocean. It will be an important research direction to further improve underwater target recognition.

For the development of diversified shapes of underwater dangerous targets, as well as the shapes of unknown enemy dangerous targets, the accuracy of target recognition cannot be guaranteed only by training dataset. Current metalearning can make algorithms to have learning capabilities. Target recognition methods based on metalearning may enable the higher recognition accuracy of underwater dangerous targets.

In summary, the underwater target recognition method will develop in the direction of intelligence, autonomy, high precision removal rate, strong robustness, and real-time performance. It will play a more powerful role in the military and civilian fields.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Key Research and Development Project [2019YFB1311000].

ORCID iD

Bowen Teng  <https://orcid.org/0000-0001-7235-4930>

References

1. Dobeck GJ, Hyland JC, Smedley L, et al. Automated detection/classification of sea mines in sonar imagery. *Proc SPIE* 1997; 3079: 90–110.
2. Acar EU, Choset H, Zhang Y, et al. Path planning for robotic demining: robust sensor-based coverage of unstructured environments and probabilistic methods. *Int J Rob Res* 2003; 22: 441–466.
3. Cao X, Zhang X, Yu Y, et al. Deep learning-based recognition of underwater target. In: *IEEE international conference on digital signal processing (DSP)*, Beijing, China, 16 October 2016, pp. 89–93.
4. Cai L, Sun QK, Xu T, et al. Multi-AUV collaborative target recognition based on transfer-reinforcement learning. *IEEE Access* 2020; 8: 39273–39284.
5. Einsidler D, Dhanak M, and Beaujean P. A deep learning approach to target recognition in side-scan sonar imagery. In: *OCEANS 2018 MTS/IEEE*, Charleston, SC, USA, 22–25 October 2018, pp. 1–4.
6. Jin L, Liang H, and Yang C. Accurate underwater ATR in forward-looking sonar imagery using deep convolutional neural networks. *IEEE Access* 2019; 7: 125522–125531.
7. Berg H and Hjelmervik KT. Classification of anti-submarine warfare sonar targets using a deep neural network. In: *OCEANS 2018 MTS/IEEE Charleston*, Charleston, SC, USA, 22–25 October 2018, pp. 1–5.
8. Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017, pp. 95–104.
9. Deng W, Zheng L, Ye Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: *2018 IEEE/CVF conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, 18–23 June 2018, pp. 994–1003.
10. Cai L, Luo P, Zhou G, et al. Maneuvering target recognition method based on multi-perspective light field reconstruction. *Int J Distrib Sens Netw* 2019; 15: 1–12.
11. Cai L, Luo P, Zhou G, et al. Multiperspective light field reconstruction method via transfer reinforcement learning. *Comput Intell Neurosci* 2020; 3: 1–14.
12. Cai L, Luo P, and Zhou G. Multistage analysis of abnormal human behavior in complex scenes. *J Sens* 2019; 2019: 1–10.
13. Cai L and Sun QK. Multiautonomous underwater vehicle consistent collaborative hunting method based on generative adversarial network. *Int J Adv Rob Syst* 2020; 17: 1–10.

14. Luo P, Cai L, Zhou G, et al. Multiagent light field reconstruction and maneuvering target recognition via GAN. *Math Prob Eng* 2019; 2019: 1–10.
15. Stack J. Automation for underwater mine recognition: current trends & future strategy. *Proc SPIE* 2011; 8017: 1–21.
16. Shin FB, Kil DH, Dobeck GJ, et al. An integrated approach to bandwidth reduction and mine detection in shallow water with reduced-dimension image compression and automatic target recognition algorithms. *Proc SPIE* 1997; 3079: 203–212.
17. Gleckler AD and Fetzer GJ. Multipurpose underwater imaging and ranging camera for low-visibility mine countermeasure (MCM) missions. *Proc SPIE* 1999; 3711: 141–150.
18. Miao XK, Feng Z, Xu YJ, et al. Mine object recognition method research of AUV based on vision. *Ocean Eng* 2012; 30: 158–164.
19. Chapple PB. Unsupervised detection of mine-like objects in seabed imagery from autonomous underwater vehicles. In: *OCEANS 2009*, Biloxi, MS, USA, 26–29 October 2009, pp. 1–6.
20. Williams DP and Fakiris E. Exploiting environmental information for improved underwater target classification in sonar imagery. *IEEE Trans Geosci Remote Sens* 2014; 52: 6284–6297.
21. Guo D and Chen B. SAR image target recognition via deep Bayesian generative network. In: *2017 international workshop on remote sensing with intelligent processing (RSIP)*, Shanghai, China, 19–21 May 2017, pp. 1–4.
22. Xie W, Yang J, Lei J, et al. SRUN: spectral regularized unsupervised networks for hyperspectral target detection. *IEEE Trans Geosci Remote Sens* 2019; 99: 1–12.
23. Giovanneschi F, Mishra KV, Gonzalez-Huici MA, et al. Dictionary learning for adaptive GPR landmine classification. *IEEE Trans Geosci Remote Sens* 2019; 99: 1–20.
24. Olmos A, Trucco E, and Lane D. Automatic man-made object detection with intensity cameras. In: *OCEANS IEEE*, Biloxi, MS, USA, 29–31 October 2002, pp. 1555–1561.
25. Rizzini DL, Kallasi F, Oleari F, et al. Investigation of vision-based underwater object detection with multiple datasets. *Int J Adv Rob Syst* 2015; 12: 1–13.
26. Yu X, Xing X, Zheng H, et al. Man-made object recognition from underwater optical images using deep learning and transfer learning. In: *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Calgary, Canada, 15–20 April 2018, pp. 1852–1856.
27. Hussain SS and Zaidi SSH. Underwater man-made object prediction using line detection technique. In: *Proceedings of the 2014 6th international conference on electronics, computers and artificial intelligence (ECAI)*, Bucharest, Romania, 23–25 October 2014, pp. 1–6.
28. Hou GJ, Luan X, Song DL, et al. Underwater man-made object recognition on the basis of color and shape features. *J Coast Res* 2016; 32: 1135–1141.
29. Qing ZL, Xiao LL, and Nana GU. Fast line detection algorithm for underwater man-made objects. *Comput Eng Appl* 2014; 163: 1533–1548.
30. Chen Y and Xu X. The research of underwater target recognition method based on deep learning. In: *2017 IEEE international conference on signal processing, communications and computing (ICSPCC)*, Xiamen, China, 22–25 October 2017, pp. 1–5.
31. Li N, Zheng Z, Zhang S, et al. The synthesis of unpaired underwater images using a multistyle generative adversarial network. *IEEE Access* 2018; 6: 54241–54257.
32. Chen C, Xie W, Huang Y, et al. Weakly-supervised man-made object recognition in underwater optimal image through deep domain adaptation. In: *International conference on neural information processing*, Siem Reap, Cambodia, 13–16 December 2018, pp. 311–322. Cham: Springer.
33. Liu M, Yuan F, Zhu Y, et al. Generating underwater images by GANs and similarity measurement. In: *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, Kobe, Japan, 28–31 May 2018, pp. 1–5.
34. Rao D, Deuge MD, Nourani-Vatani N, et al. Multimodal learning for autonomous underwater vehicles from visual and bathymetric data. In: *2014 IEEE international conference on robotics and automation (ICRA)*, Hong Kong, China, 31 May–7 June 2014, pp. 3819–3825.
35. Cho H, Gu J, and Yu SC. Robust sonar-based underwater object recognition against angle-of-view variation. *IEEE Sens J* 2016; 16: 1013–1025.
36. Jin L and Liang H. Deep learning for underwater image recognition in small sample size situations. In: *OCEANS 2017-Aberdeen*, Aberdeen, UK, 19–22 June 2017, pp. 1–4.
37. Li J, Eustice RM, and Johnson-Roberson M. Underwater robot visual place recognition in the presence of dramatic appearance change. In: *OCEANS 2015-MTS/IEEE Washington*, Washington, DC, USA, 19–22 October 2015, pp. 1–6.
38. Zhou X, Yang K, and Duan R. Deep learning based on striation images for underwater and surface target classification. *IEEE Signal Process Lett* 2019; 26: 1378–1382.
39. Yang H, Xu G, Yi S, et al. A new cooperative deep learning method for underwater acoustic target recognition. In: *OCEANS 2019-Marseille*, Marseille, France, 17–20 June 2019, pp. 1–4.
40. Zhang S and Xing S. Intelligent recognition of underwater acoustic target noise by multi-feature fusion. In: *2018 11th international symposium on computational intelligence and design (ISCID)*, Hangzhou, China, 8–9 December 2018, pp. 212–215.
41. Zhang SM and Negahdaripour S. 3-D shape recovery of planar and curved surfaces from shading cues in underwater images. *IEEE J Oceanic Eng* 2002; 27: 100–116.
42. Song D, Sun W, Ji Z, et al. Color model selection for underwater object recognition. In: *2014 international conference on information science, electronics and electrical engineering*, Sapporo, Japan, 26–28 April 2014, pp. 1339–1342.
43. Tang XD, Pang YJ, Li Y, et al. Object auto-recognition for underwater targets. In: *2009 Chinese control and decision conference*, Guilin, China, 17–19 June 2009, pp. 4612–4616.
44. Li C, Guo J, Cong R, et al. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans Image Process* 2016; 25: 5664–5677.

45. Ding X, Wang Y, Zhang J, et al. Underwater image dehaze using scene depth estimation with adaptive color correction. In: *OCEANS 2017-Aberdeen*, Aberdeen, UK, 19–22 June 2017, pp. 1–5.
46. Ghani ASA, Nasir AFA, and Tarmizi WFW. Integration of enhanced background filtering and wavelet fusion for high visibility and detection rate of deep sea underwater image of underwater vehicle. In: *IEEE 2017 5th international conference on information and communication technology (ICoICT)*, Malacca, Malaysia, 17–19 May 2017, pp. 1–6.
47. Ahn J, Yasukawa S, Sonoda T, et al. Image enhancement and compression of deep-sea floor image for acoustic transmission. In: *IEEE Oceans-Shanghai*, Shanghai, China, 10–13 April 2016, pp. 10–13.
48. Chuang M, Hwang J, and Williams K. A feature learning and object recognition framework for underwater fish images. *IEEE Trans Image Process* 2016; 25: 1862–1872.
49. Zhang L, He B, Song Y, et al. Underwater image feature extraction and matching based on visual saliency detection. In: *OCEANS 2016-Shanghai*, Shanghai, China, 10–13 April 2016, pp. 1–4.
50. Yu X, Qu Y, and Hong M. Underwater-GAN: underwater image restoration via conditional generative adversarial network. In: *International conference on pattern recognition*, Beijing, China, 20–24 August 2018. Cham: Springer.
51. Hou W, Goode W, and Kanaev A. Underwater image quality degradation by scattering. In: *2012 Oceans-Yeosu* Yeosu, South Korea, 21–24 May 2012, pp. 1–5.
52. Ouyang B, Hou W, Gong C, et al. Experimental study of a compressive line sensing imaging system in a turbulent environment. *Appl Opt* 2016; 55: 8523.
53. He C and Zhang Z. Restoration of underwater distorted image sequence based on generative adversarial network. In: *2019 IEEE 8th joint international information technology and artificial intelligence conference (ITAIC)* Chongqing, China, 24–26 May 2019, pp. 866–870.
54. Kupyn O, Budzan V, Mykhailych M, et al. DeblurGAN: blind motion deblurring using conditional adversarial networks. In: *2018 IEEE/CVF conference on computer vision and pattern recognition* Salt Lake City, UT, USA, 18–23 June 2018, pp. 8183–8192.
55. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *2014 IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, 23–28 June 2014, pp. 580–587.
56. Girshick R. Fast R-CNN. In: *2015 IEEE international conference on computer vision (ICCV)*, Santiago, Chile, 11–18 December 2015, pp. 1440–1448.
57. Ren SQ, He KM, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017; 39: 1137–1149.
58. Lin TY, Dollar P, Girshick R, et al. Feature pyramid networks for object detection. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)* Honolulu, HI, USA, 21–26 July 2017, pp. 936–944.
59. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27–30 June 2016, pp. 779–788.
60. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi-box detector. In: *2016 European conference on computer vision (ECCV)* Amsterdam, the Netherlands, 8–16 October 2016, pp. 21–37.
61. Wang Z, Liu C, Wang S, et al. UDD: an underwater open-sea farm object detection dataset for underwater robot picking. 2020, arXiv eprints, arXiv: 2003.01446.
62. Redmon J and Farhadi A. YOLOv3: an incremental improvement. 2018, arXiv eprints, arXiv: 1804.02767.
63. Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2017; 99: 2999–3007.
64. Tian Z, Shen C, Chen H, et al. FCOS: fully convolutional one-stage object detection. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 27 October–2 November 2019, pp. 9626–9635.
65. Kong T, Sun F, Liu H, et al. Foveabox: beyond anchor-based object detector. Online Referencing, <http://arxiv.org/abs/1904.03797> (2019).
66. Zhang X, Wan F, Liu C, et al. FreeAnchor: learning to match anchors for visual object detection. 2019, arXiv eprints, arXiv:1909.02466.
67. Yang Z, Liu S, Hu H, et al. RepPoints: point set representation for object detection. In: *2019 IEEE/CVF international conference on computer vision (ICCV)*, Seoul, Korea, 27–28 October 2019, pp. 9656–9665.
68. Wang J, Chen K, Yang S, et al. Region proposal by guided anchoring. In: *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Long Beach, CA, USA, 15–20 June 2019, pp. 2960–2969.
69. Zhou X, Wang D, and Krähenbühl P. Objects as points. 2019, arXiv eprints, arXiv:1904.07850.
70. Yu F, Wang D, and Darrell T. Deep layer aggregation. In: *2018 IEEE/CVF conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, 18–23 June 2018, pp. 2403–2412.
71. Zhang D, Li L, Zhu Z, et al. Object detection algorithm based on deformable convolutional networks for underwater images. In: *2019 2nd China symposium on cognitive computing and hybrid intelligence (CCHI)*, Xi'an, China, 21–22 September 2019, pp. 274–279.
72. Dai JF, Li Y, He KM, et al. R-FCN: object detection via region-based fully convolutional networks. 2016, arXiv eprints, arXiv:1605.06409v2.