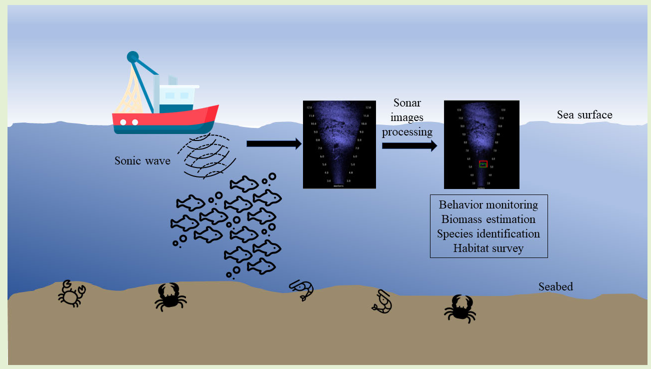
Deep Learning Algorithms for Sonar Imagery Analysis and Its Application in Aquaculture: A Review

Yingqian Chai, Huihui Yu, Ling Xu, Daoliang Li, and Yingyi Chen

***Abstract*—With the rapid development of underwater sensor techniques, using acoustic techniques to obtain and analyze underwater images is an effective approach due to the superior nature of sonar, especially in turbidity and dim underwater environment. In recent years, sonar image processing based on deep learning (DL) is being developed and applied to automatic target recognization tasks. This article outlines research of DL-based algorithms for sonar imagery, including denoising, feature extraction, classification, detection, and segmentation. This review shows that sonar image classification is the most studied and transfer learning has put into spotlight. Besides, datasets and applications are listed that can be used in aquaculture. Currently known applications include behavior analysis, biomass estimation, species identification, and habitat survey. Furthermore, smart aquaculture, as one of the important application fields of sonar technique, the challenges and future research directions in aquaculture are discussed. The purpose of this work aims to help researchers find effective DL algorithms for sonar image processing, especially the potential application in aquaculture.**



***Index Terms*— Aquaculture, deep learning (DL), image processing, sonar image.**

I. Introduction

WITH the rapid development of underwater sensor techniques, acoustic sensors are mostly considered powerful devices to sense underwater objects and targets compared with the extreme fading of electromagnetic waves and visual signals [1], [2], especially in certain environment, such as turbidity and low light in water [3]. Sound navigation and ranging (SONAR) [4] is a technique and device that uses acoustic waves to measure distances and detect underwater objects. Using sonar to collect image data can effectively avoid the influence of water visibility during the imaging processing [5]. Acoustic signals are mainly divided into active signals and passive signals. The echo principle of active sonar and passive sonar is shown in Fig. 1. In this article, we focus on three active sonars, which are side-scan sonar (SSS), forward-looking sonar (FLS), and synthetic aperture sonar (SAS). The classifications and applications of sonars are shown in Table I

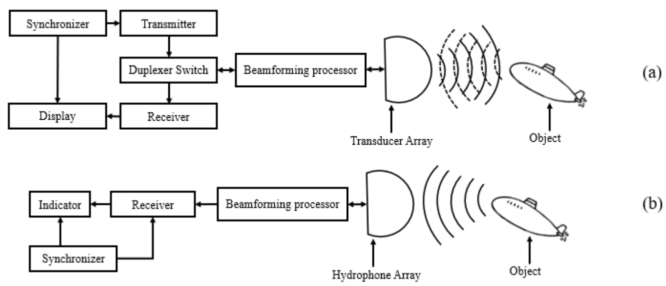


Fig. 1. Echo principle. (a) Active sonar. (b) Passive sonar.

Manuscript received 8 February 2023; revised 8 October 2023; accepted 10 October 2023. Date of publication 27 October 2023; date of current version 30 November 2023. This work was supported in part by the National Natural Science Foundation of China "Analysis and feature recognition on feeding behaviour of fish school in facility farming based on machine vision" under Grant 62076244, in part by the Beijing Digital Agriculture Innovation Consortium Project under Grant BAIC10-2023, and in part by the National Natural Science Foundation of China "Intelligent identification method of underwater fish morphological characteristics based on binocular vision" under Grant 62206021. The associate editor coordinating the review of this article and approving it for publication was Prof. Dongsoo Har. *(Corresponding author: Yingyi Chen.)*

Yingqian Chai, Ling Xu, Daoliang Li, and Yingyi Chen are with the National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China, also with the Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China, also with the Beijing Engineering and Technology Research Centre for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China, and also with the College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China (e-mail: chaiyq93@163.com; lingxu@cau.edu.cn; dliangl@cau.edu.cn; chenyingyi@cau.edu.cn).

Huihui Yu is with the School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China (e-mail: yuhh1990@126.com).

Digital Object Identifier 10.1109/JSEN.2023.3324438

TABLE I
CLASSIFICATIONS AND APPLICATIONS OF SONARS

Sonars	Applications
FLS	Obstacle avoidance and local path planning, the depth measure and the topography of the sea floor
SSS	Quick and cursory scan of the ocean floor
SAS	Imaging of geomorphology and small targets in bottom and burial, submarine pipeline detection

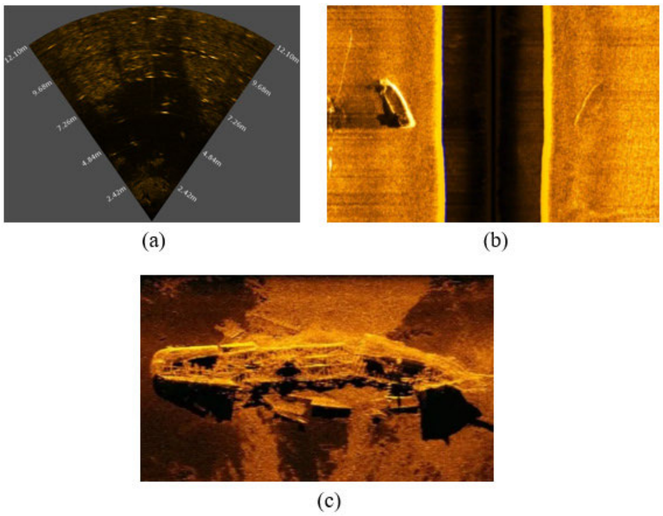


Fig. 2. Comparison of three sonar images. (a) FLS. (b) SSS. (c) SAS.

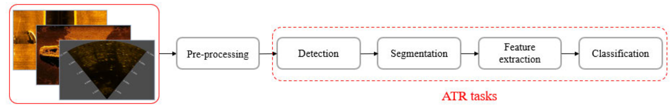


Fig. 3. Flowchart of preprocessing and ATR tasks for sonar images.

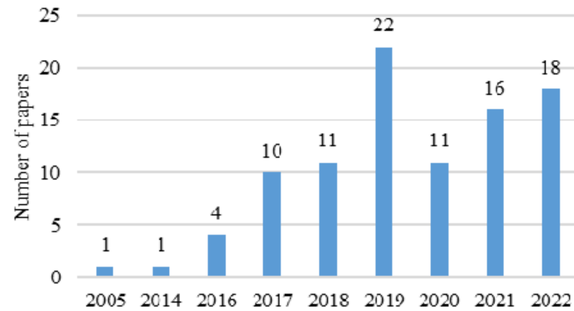


Fig. 4. Distribution of papers for different years.

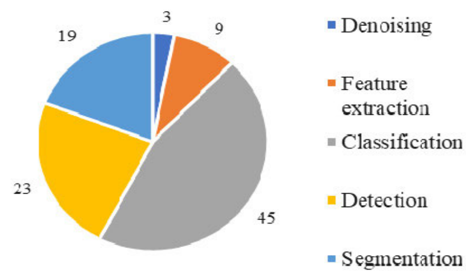


Fig. 5. Number of papers of sonar image processing based on DL.

and Fig. 2 shows three sonar images. The sonar images consist of three different regions, including highlight area, shadow area, and background. These regions provide the basis to process sonar images.

Deep learning (DL) widely has been used in image processing. With the development of DL, many methods for optical image processing are used for sonar image automatic target recognition (ATR) tasks, including feature extraction, classification, detection, and segmentation. However, unlike optical imaging, sonar imaging does not have plenty of information on detailed content, color, shape, and texture [6]. Meanwhile, there is a lot of noise. These factors lead to the late application of DL in sonar images than optical images.

We learn about some review articles related to sonar image processing [7], [8], [9]. As far as we know, however, no survey gives a complete view on sonar image processing, including preprocessing and ATR tasks of SSS, FLS, and SAS, and applications of sonar image processing based on DL for aquaculture. We mainly focus on the above content.

This article is organized as follows. Section II presents various models and methods of sonar image processing based on DL, such as sonar image denoising, feature extraction, classification, detection, and segmentation. Section III introduces the applications of sonar image processing based on DL in aquaculture. Section IV describes the common key challenges and future research directions. Discussion and conclusion are in Section V.

II. SONAR IMAGE PROCESSING BASED ON DL

DL is an effective tool to analyze and process sonar images for ATR tasks. Fig. 3 shows preprocessing and the typical processing chain of ATR tasks. Fig. 4 plots the number of papers published from different years. The number of papers and proportion of different sonar image processing tasks is shown in Fig. 5. In this section, DL models for sonar image denoising and ATR tasks are described in detail.

A. Sonar Image Denoising

With strong noises and low resolution, the quality of sonar images is far inferior to that of optical images. Also, the difference between objects and background is small, which makes it difficult to read useful information and adversely affects ATR tasks. Therefore, sonar image preprocessing is needed before executing ATR tasks. In this section, we mainly focus on sonar image denoising preprocessing.

Sonar image denoising focuses on traditional methods. The classical denoising algorithms mainly include both spatial domain filtering and transform domain filtering [10], [11], according to different filtering space. The former [12], [13] includes median filtering and Wiener filtering. The latter [14] has wavelet transform, Fourier transform, principal component analysis, and so on. Noise is divided into additive noise and multiplicative noise [15], [16]. The most common noise in sonar images is speckle noise, which is multiplicative noise. A lot of works used traditional denoising methods to denoise the speckle noise [17], [18], [19], [20]. In general, spatial filtering has a good visual smoothing effect, but the loss of image edge details is more. The transform domain

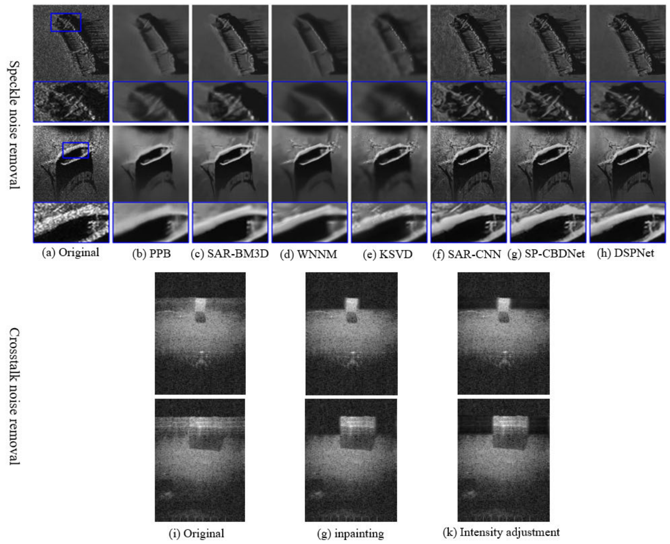


Fig. 6. Removal effect of speckle noise and crosstalk noise. (a) Original. (b) PPB. (c) SAR-BM3D. (d) WNNM. (e) KSVD. (f) SAR-CNN. (g) SP-CBDNet. (h) DSPNet. (i) Original. (j) Inpainting. (k) Intensity adjustment.

filtering has a good edge preserving effect, but the removal of high-frequency contexts of signal leads to the loss of details.

Compared with traditional methods, image denoising based on DL can achieve better results by learning the statistical characteristics of images. For multiplicative speckle noise, a convolutional neural network (CNN)-based residual learning method [21], [22] was proposed to estimate the Gaussian noise and contributed a deep blind DeSPeckling network (DSPNet). The method introduced the feature pyramid network (FPN) and the atrous spatial pyramid pool (ASPP) and consisted of two subnets. The two subnets were noise estimation network in the logarithmic domain (Log-NENet) to estimate noise level map and denoising network in the logarithmic domain (Log-DNNet) to reduce random noise. The multiscale mixed loss function improved the robustness and generalization ability of network. The DSPNet is capable of achieving a balance between noise removal and significant image detail preservation. For the distortion of the target shape caused by crosstalk noise in FLS images, CNN [23] was used to detect and remove crosstalk noise. YOLOv3, which was reduced the input layers and the filter size of the last layer into one-fourth, detected crosstalk noise. Two methods were used to remove noise. The first method is inpainting by filling the detection region with adjacent shadow pixel value. The second method is intensity adjustment, which is that each pixel multiplies a weight value determined by analyzing the detected region pixel intensity. This method preserved important information for a single image. Fig. 6 shows the removal effect of spackle noise and crosstalk noise.

This overview shows that sonar image denoising usually uses traditional filtering methods. DL-based approaches have just started. There are many obstacles to denoise sonar image, mainly including the absence of datasets and complexity of noise. In other fields, especially optical images, some supervised and unsupervised DL methods were used. These methods can be referred to sonar images. We hope that denoising preprocessing removes the strong noise and preserve the weak original edge characteristic information as much as possible. This is still a challenge.

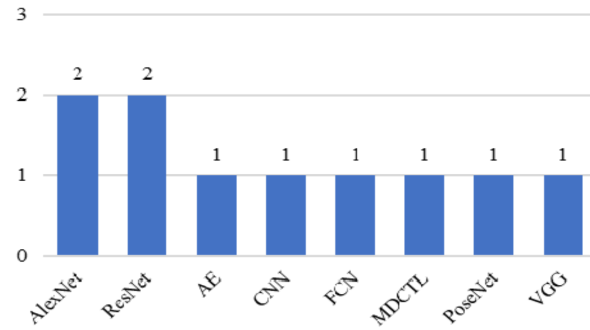


Fig. 7. Number of papers for different feature extraction methods.

TABLE II
SUMMARY OF FEATURE EXTRACTION BASED ON DL FOR SONAR IMAGES

Paper	Method	Object
[24]	AE	MLO
[26]	AlexNet	Objects of interest
[25]	VGG-16, VGG-19, VGG-f, AlexNet	Mine
[27]	ResNet-50	Archaeological sites
[29]	PoseNet	Objects of harbor, coast
[30]	FCN	Seabed topography
[28]	ResNet-50	Shipwreck, plain wreck
[31]	MDCTL	wreck, airplane, seabed background
[32]	CNN	Shipwreck, corpse, aircraft, sand wave

B. Feature Extraction

It is mainly to extract the discriminating features of objects in sonar images. Before using DL methods, feature extraction was mainly based on handcrafted features, which were designed for specific objects. This section focuses on feature extraction based on DL. Fig. 7 summarizes the number of papers for different feature extraction methods. Papers are listed in Table II.

Before using DL for feature extraction of sonar images, artificial neural network (ANN) was used for feature representation learning [24]. Transfer learning is an important method for sonar image feature extraction. The first paper using DL was published in 2017 [25], [26]. Fine-tuning pre-training CNNs, including VGG-16, VGG-19, VGG-f, AlexNet, ResNet-50, and PoseNet [25], [26], [27], [28], [29], were used to extract features, which were combined with support vector machine (SVM) to achieve better performance. Among them, VGG-19 and AlexNet had good performance. ResNet had better generalization ability. Meanwhile, PoseNet used triplets. Besides, for the problems of low accuracy, discontinuous edge contour, and loss of detail features, an improved fully convolutional network (FCN) [30], channel attention module (CAM) [31], and a deep separable residual module (DSRM) [32] were used. The improved FCN added a batch normalization layer to improve the skip structure and used a mini-batch gradient descent method to accurately extract

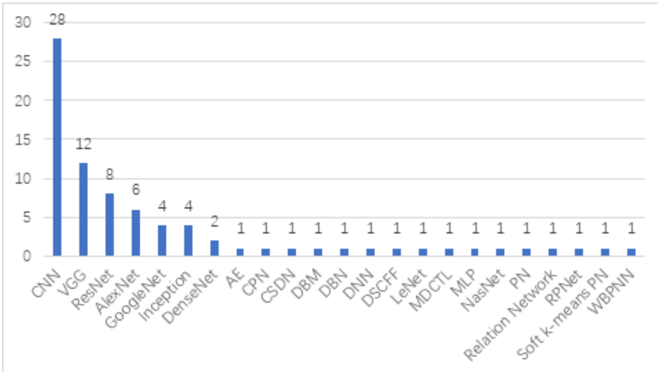


Fig. 8. Number of papers for different classification methods.

and fuse detailed information. The mean intersection over union (IoU) is 83.05%. The CAM learned more features by combining the maximum pooling and the global average pooling results. Each channel outputted different features and CAM paid more attention to outputting features. The DSRM extracts multiscale features and a multichannel feature fusion method (MCFFM) to enhance the transfer capability of feature. The DSRM decreased model training parameters to reduce computational complexity. The MCFFM used channel merging technology to stitch the feature information of different stages.

As seen from the overview, compared with traditional feature extraction methods, DL-based methods can capture more discriminating features and provide better performance. However, few articles study feature extraction, which is mostly included in subsequent ATR tasks.

C. Classification

Sonar image classification distinguishes different classes of objects according to their respective features reflected in images. In the case of limited samples and zero sample, scholars have conducted different studies. This section reviews different classification models with different sample sizes and sonar types, including SSS, FLS, and SAS. Fig. 8 shows the number of papers for different classification methods.

1) Public Datasets: Some studies used publicly available datasets to complete classification task. The classification methods are shown in Table III. The table shows the classification tasks carried out in the paper using different methods and datasets.

Three datasets of CKI, TDI-2017, and TDI-2018 were used to classify images, including a submerged dummy, with noise effects such as polarizing noise. The CKI dataset was captured in clean water testbed. TDI-2017 and TDI-2018 were acquired from real cloudy water in 2017 and 2018, respectively. GoogleNet [33] and AlexNet [34] were used for the binary classification task of body and background. For the multiclass classification task, the dataset *SeabedObjects-KLSG* contains wreck, drowning victim, airplane, mine, and seafloor images. Huo et al. [35] and Cheng et al. [31] used the dataset. Transferring and fine-tuning pretrained CNNs [35], [36], [37], e.g., VGG and ResNet, were used to achieve more satisfactory results. Huo et al. [35] demonstrated that the model improved to an accuracy of 97.76% trained by using

TABLE III
OBJECT CLASSIFICATIONS OF USING PUBLIC DATASETS

Paper	Method	Object	Dataset
[33]	GoogleNet	Submerged body vs. background	CKI, TDI-2017, TDI-2018
[34]	AlexNet, GoogleNet	Submerged body vs. background	CKI, TDI-2017, TDI-2018
[35]	CNN, VGG-16, VGG-19	wreck, drowning victim, airplane, mine, seafloor	the *SeabedObjects-KLSG*
[31]	MDCTL	shipwreck, airplane, seabed background	the SeabedObjects-KLSG-II
[36]	VGG16, ResNet, ResNet+	Ten types of marine debris	FLS Marine Debris Dataset
[37]	AlexNet, LeNet, DenseNet-52, DenseNet-100, DenseNet-151, ResNet-3-2, ResNet-3-3, ResNet-4-2, VGG-13, VGG-16	Reef, mud, sand wave	CIFAR-10

a dataset that included 70% of the *SeabedObjects-KLSG* and 30% of semisynthetic data of airplanes and drowning victims produced by optical images. Cheng et al. [31] proposed a multidomain collaborative transfer learning (MDCTL) method based on multiscale repeated attention mechanism (MSRAM). The MDCTL method combined low-level characteristic similarity between SSS and synthetic aperture radar (SAR) images and high-level representation similarity between SSS and optical images to efficiently improve the ability of feature extraction. The MSRAM combined space and the CAM extract rich contour information to achieve a better classification accuracy of 99.21% and generalization capability of model. For the classification of other datasets, Aleem et al. [36] used VGG16 and ResNet-50 to complete ten types of marine debris classification on FLS marine debris dataset. Feature maps from regional proposal network (RPN) of faster-region-CNN (RCNN) were used for the classification in the final step. Using ResNet-50 over VGG16 can learn more features. Qin et al. [37] used different CNNs to achieve sediment classification on dataset CIFAR-10. The best performance is a 3.459% error rate by fine-tuning ResNet.

2) Limited Training Samples: Datasets of sonar images are very scarce. In this section, we reviewed the classification methods of sonar images with limited data or simulation data. Table IV shows the classification methods carried out in the papers for different classification tasks and sonars.

SSS is often used to capture targets on seabed, including mine-like object (MLO), nonmine-like object (NMLO), other significant object (OSO), shipwreck, and plane seabed sediment. Nevertheless, available sonar images are limited. Under the condition of limited training samples, the training of DL models is very difficult. For the classification of SSS image, transfer learning is still a valuable method. CNN [38], [39], [40], [41], [42], [43], [44], VGG [45], [46], ResNet [47],

TABLE IV
SUMMARY OF CLASSIFICATION METHODS IN LIMITED TRAINING SAMPLES

Index	Paper	Method	Task	Sonar
SSS	[38]	CNN	MLO vs. background	——
	[49]	InceptionNet	MLO, NMLO, FAO	The Marine Sonic Technology (MST) SSS, Interferometric SAS
	[43]	CNN	MLO	The Marine Sonic Technology (MST) SSS
	[45]	VGG-11，ResNet-18	Ship & plane wreck vs. others	——
	[39]	AW-CNN, CNN, DBN	Shipwreck, plain wreck, stone, tire, shoal, sand ripple	——
	[44]	CNN, AlexNet, VGG-16, VGG-19, ResNet-50, Inception-v3, GoogleNet	MLO, OSO, background	The Marine Sonic Technology (MST) SSS
	[46]	CNN, VGG	MLO, NMLO, FAO	The Marine Sonic Technology (MST) SSS, Interferometric SAS
	[48]	CNN	Seabed sediment (reef, mud, sand wave)	DF1000/560D digital dual-frequency SSS
	[40]	CNN	Boat, plane, sand, stone	——
	[47]	PN, Relation Network, Soft k-means PN, CPN, CNN, ResNet-18, ResNet-50	Anchor, cube, plane, boat, pyramid	——
	[41]	CNN	Archaeological sites vs. background	Edgetech 2205 Dual Frequency 600/1600 kHz
	[51]	DSCFF, VGG16, ResNet50	Sphere, cylinder, truncated cone	——
	[42]	AlexNet, Inception, VGG-16	Wreck, rock, mule, unknown, nothing	——
	[50]	DNN	Airplane, shipwreck, others	——
FLS	[52]	CNN	9 types of debris objects and background	ARIS Explorer 3000
	[53]	CNN	Jellyfish, sediment, artefacts, fish, seaweed, background	ARIS Explorer 3000
	[54]	WBPNN	Seabed sediment	——
	[55]	RPNet	Sand and muddy sand, mixed sediment, coarse sediment	Simrad/Kongsberg EM2040, Reson Seabat 7125
SAS	[71]	MLP	Mines	——
	[58]	AE, CNN	Seabed sediment (Flat sand, sandy ridges, rocky sand, rocky, sand ripple)	The Small Synthetic Aperture Minehunter II (SSAMII)
	[72]	GoogleNet	Seabed sediment (Fine, sand, coarse, mixed sediment)	——
	[25]	VGG-16, VGG-19, VGG-f, AlexNet	Mines	——
	[65]	CNN	Mines	——
	[66]	AlexNet, VGG-16, DenseNet-161, Inception-ResNet-v2, NasNet-large, CNN	Mines	the HISAS 1030 SAS
	[73]	CSDN	Target vs. noise	——
	[59]	CNN	Rock vs. cylinder	——
	[60]	CNN	Target vs. clutter	——
	[56]	CNN	Mines	CMRE MUSCLE SAS
	[75]	DBM	Target vs. clutter	CMRE MUSCLE SAS
	[57]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[61]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[67]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[68]	CNN, VGG	Target vs. clutter	CMRE MUSCLE SAS
	[63]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[69]	CNN，VGG-16	Target vs. clutter	CMRE MUSCLE SAS
	[62]	CNN	UXO vs. Non-UXO	CMRE MUSCLE SAS
	[64]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[70]	CNN	Target vs. clutter	CMRE MUSCLE SAS
	[74]	CNN, DensNet-121, ResNet-18	Target vs. clutter	CMRE MUSCLE SAS

AlexNet [48], and InceptionNet [49] were used to classify objects in SSS images. The combination of VGG-11 and ResNet-18 [45] effectively avoided overfitting. Other strategies contribute to reach high accuracy, such as fine-tuning last full connected layer with SVM classifier and all parameters of networks, combining CNN with a hierarchical Gaussian process (HGP) classifier and machine learning (ML), and using adaptive weights CNN (AW-CNN) using the weights from deep belief network (DBN). Ochal et al. [47] compared several few-shot learning (FSL) methods, which were prototypical network (PN), relation network, soft k-means PN, and consistent PN (CPN). On most datasets, FSL methods achieved better performance than baseline models. In addition, using generative adversarial network (GAN) to extend the dataset [40], [44] effectively improves the performance of classification. In extreme situations, no appropriate SSS image of specific targets is available, called the "zero-shot learning problem" [50]. A zero-shot classification method was developed using deep neural network (DNN) to training pseudo SSS images. The images were synthesized through a fixed style-transfer method using common optical images and any available SSS images. For a great deal of parameters and high time complexity problem, a depthwise separable convolution feature fusion (DSCFF) [51] was proposed. This method made full use of effective information to reach 90.1% accuracy and had better performance and lighter model compared with VGG16 and ResNet50. Comparing the computation speed and accuracy of different convolution layer depth CNNs, convolutional layers beyond five layers improved accuracy slowly but at greater cost [43].

FLS can be used to search underwater targets and seafloor topography. CNNs were used to classify debris objects [52] and jellyfish [53]. For the classification of seabed sediment, wavelet-BP neural network (WBPNN) [54] and random patches network (RPNet) [55] were used. The WBPNN was with self-adaptive learning rate and momentum factor and introduced network parameter initialization. The RPNet avoided the influence of noise and small sample. At the same time, for the oversmoothness and misclassifications, RPNet used a decision fusion algorithm and improved the accuracy of classification compared to a single classifier.

SAS can address acoustic signal and achieve the imaging of geomorphology and targets in bottom and burial of marine. Similarly, CNNs [25], [56], [57], [58], [59], [60] are the most used method. A CNN [56], [57] with four convolutional layers had the best performance. Fine-tuning CNN [58] was used to classify seafloor types with a small set of labeled data and the accuracy achieved 88.2%. Synthetic images [59] were used to train CNN, which performed as same as a CNN trained on real data. A binary classification task [60] was completed and its result showed that training with low confident false alarms and augmented data performed better. Target-concept transfer learning, sensor transfer learning, and intramodality transfer learning were trained from mines to unexploded ordnance (UXO) [61], [62]. Among them, intramodality transfer learning achieved superior performance. Moreover, averaging the classification prediction results, fusing feature extraction data [63], modifying the final dense layer and output layer [64], changing the number of fully connected layers [65], and fine-tuning the whole network [66] provided richer information and achieved better performance. VGG and NasNet-large pretrained on ImageNet achieved the best performance after fine-tuning. Besides, phase information, frequency spectrum, amplitude content, and 2-D power spectral density (PSD) [67], [68], [69] of SAS images are beneficial to obtaining better performance of classification that CNN was combined with the information. Using the phase information [67] alone could obtain better classification performance than using the amplitude content of the imagery. The performance of the PSD [68] in addition to the amplitude was the best. Meantime, a VGG network pretrained on ImageNet achieved the best performance by fine-training. The parameters of CNN [69], which combined with the above information, were less. In addition, some auxiliary content, including image quality, local background environment, object shape, orientation, and length, can also be available. Besides, tiny CNNs [70], multilayer perceptron (MLP) [71], GoogleNet [72], and compound CNN (CSDN) [73] were also used for classification. To reduce the high false alarm rate, structural prior-driven regularized DL (SPDRDL) [74] was proposed, which incorporated structural similarity prior and structural scene context priors. Besides, multiview SAS image classification was researched combining image-fusion technique and a DL-based on Boltzmann machines (DBM) [63], [75].

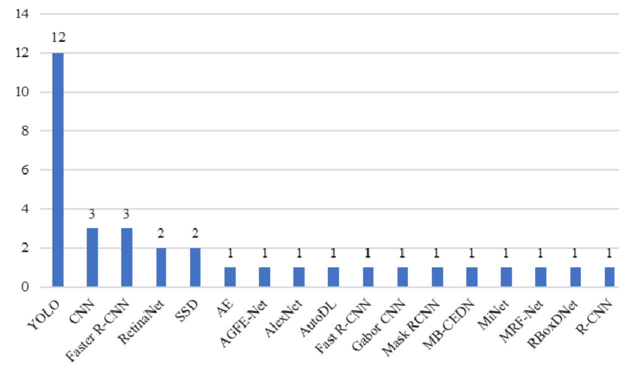


Fig. 9. Number of papers for different detection methods.

Classification is the most studied task of sonar images at present. As can be seen from the overview, CNN and VGG network are popular methods for classifying sonar images. Transfer learning is a feasible method pretraining on optical images and fine-tuning on sonar images.

D. Detection

Object detection of sonar images aims to recognize possible regions of interest (ROIs), predict the location and size of each object with the boundary frame, and give the confidence degree.

At present, object detection methods are mainly divided into two categories: two-stages and one-stage algorithms. DL detection methods based on attention mechanism are also considered in Section II-D3. The number of papers for different detection methods is shown in Fig. 9. Table V lists the detection methods for different detection objects.

1) Two-Stage Detection: Two-stage detection algorithm divides the detection problem into two stages. It generates regional proposals and then classifies the candidate regions. This method has high precision, but the speed is low. Some papers compared both one- and two-stage methods, which are discussed in one-stage detection. Faster R-CNN [76] was used to detect objects using three active learning methods, including

TABLE V
SUMMARY OF DETECTION METHODS

Index	Paper	Method	Object	Sonar
Two-stages	[76]	Faster R-CNN, YOLOv1, SSD	Shipwreck, plane, wreckm corpse	Colletcted from internet
	[77]	Mask RCNN	Metal ball, metal triangle, metal circle, bottle, tire, mine model, iron chain	Tritech Gemini 720i
One-stage	[92]	AE, CNN	MLO	——
	[25]	VGG-16, VGG-19, VGG-f, AlexNet	Mines	——
	[79]	YOLOv2	seafloor anomaly, rock	——
	[87]	YOLOv2，RBoxDNet	Subsea isolation valve, shipwreck	Tritech Gemini 720i
	[84]	YOLOv3-DPFIN	Line target	——
	[78]	YOLOv1, Faster R-CNN	shipwreck	Colletcted from internet
	[89]	RetinaNet	Rock	EG&G DF-1000 SSS
	[64]	CNN	MLO	CMRE MUSCLE SAS
	[88]	Gabor CNN, R-CNN, Fast R-CNN, Faster R-CNN, Tiny YOLOv3, YOLOv3, SSD300	MLO	Marine Sonic Technology (MST) SSS
	[81]	YOLOv3	Shipwreck	——
	[90]	RetinaNet	Rock	EG&G DF-1000 SSS
	[86]	TR-YOLOv5s	Shipwreck, submarine container	Edgetech 4200-MP, Shark-S150D, Benthos SIS-1624
	[80]	YOLOv2, YOLOv3, CNN	Targets	——
	[91]	MiNet	MLO	Klein 3500, EdgeTech 2205
	[85]	YOLOv4	person, kenaifish, geese, shark, alligator, propeller, pipe, ammunitionbox, umbrella, shipwreck	ARIS Explorer 3000，ARIS Explorer 1800
	[82]	YOLOv3	AUV	——
	[83]	YOLOv3	Obstacle	——
Attention mechanism	[93]	MRF-Net	Fishing net	Gemini 720i FLS
	[94]	AGFE-Net	cube, sphere, cylinder, human body, tires, circle cage, square cage, metal bucket, airplane, ships	Tritech Gemini 1200i, SSS of Canadian Imagenex Company
	[95]	AutoDL	Shipwreck, human, aircraft, background	——
	[96]	MB-CEDN	Anchor, shipping container, square crate and debris, undersea pipe	Multi-element SAS mounted on Hydroid REMUS 600 UV

uncertainty selection, uncertainty and diversity selection, and location information selection. Uncertainty and uncertainty and diversity selection saved 50% of labeling cost and had the same performance compared to random selection. Besides, single shot multibox detector (SSD) and YOLOv1 also were used as detector and SSD and faster R-CNN outperform YOLOv1. A 32-layer feature extraction network was constructed using residual blocks to replace ResNet50/101 in mask RCNN [77]. Besides, the Adagrad optimizer was used to improve the network performance. This method reduced the training parameters and the mean average precision (mAP) is 96.97%.

2) One-Stage Detection: One-stage algorithm is an end-to-end object detection method, which uses a network to output class and location of objects. This method has high speed but low precision, especially for the small target regression. There are many studies concentrated on target detection algorithms based on you only look once (YOLO) series. YOLOv1 [78], YOLOv2 [79], YOLOv3 [80], [81], [82], [83], YOLOv3-dual-path feature fusion neural network (DPFIN) [84], lightweight YOLOv4 [85], and integration of transformer module and YOLOv5s (TR-YOLOv5s) [86] were used to detect objects. For rotated object detection, an end-to-end network (RBoxNet) [87] was proposed and a pipeline composed of YOLOv2 and RBoxNet to detect targets and determine target position and rotation. The RBoxNet reached the optimum tradeoff between accuracy and speed, while YOLOv2+RBoxNet was the fastest at 16.19 FPS. YOLOv3 outperformed YOLOv2 and CNN at 98.2% of a true positive rate. Aiming at sonar datasets with small effective sample size and low SNR, YOLOV3-DPFIN extracted effective features through dual-path network (DPN) module and fusion transition module, and improved multiscale prediction by dense connection. The experiment showed that mAP was 84.4% with 56 fps on two simulated sonar datasets. FPN, K-means, and binary classification cross-entropy function were used to improve YOLOv3. The lightweight YOLOv4 improved feature extraction network CSPDARKnet-53 of YOLOv4 to reduce the model parameters and network depth, replaced the PANet feature enhancement module with the adaptive spatial feature fusion module (ASFF), and increased the number of fused feature layers. The performance of the model in detection accuracy and speed improved compared to YOLOv4 and YOLOv4-tiny. TR-YOLOv5s introduced attention mechanism to complete real-time target detection in view of the characteristics of sparse targets and poor features in SSS images. The model was verified that the mAP was 85.6%, macro-F_2 was 87.8%, and the real-time recognition speed was about 0.068 s per image. In addition, Gabor-based detector [88] was proposed to detect MLOs and compared to other detectors. The method extracted both weak and strong features, which effectively handled multiple scales MLOs. A parameterized Gabor layer was introduced to extract features. The cascaded layers embedded the Gabor filtering modules. These improvements enhanced detection speed and precision. Besides, RetinaNet [89], [90], MiNet [91], CNN [64], [92], and AlexNet [25] were also used to detect objects. RetinaNet detection framework [89] based on ResNet needed to contain at least 3 × 3 pixels in the image to detect rock and using a single-stage residual network [90] could improve the resolution of the images. MiNet [91] was light to mount on autonomous underwater vehicle (AUV) at 9.9 MB and 0.0122 giga floating-point operations per second (GFLOPS) and used an incremental training procedure to train the model. However, using a VGG-like CNN [92] to detect MLO caused high false alarm rate.

3) Attention Mechanism: Except for one- and two-stage detection methods, attention mechanism-based models were also used for object detection. Multiple receptive field network (MRF-Net) [93], adaptive global feature enhancement network (AGFE-NET) [94], automatic DL (AutoDL) [95], and multibranch convolutional encoder–decoder network (MB-CEDN) [96] were used to detect underwater objects. Aiming at the problem of redundancy and poor generalization ability of transfer model, AutoDL was introduced. The number of parameters was only 15.2% of that of ResNet50, but the method achieved high detection precision and speeds.

Reviewing all work of detection methods, YOLO series are widely used to detect objects in sonar images. Basic CNNs after pretraining and fine-tuning are also popular target detection methods. Meanwhile, fast R-CNN, faster R-CNN, YOLO, and SSD were compared. In general, fine-tuning a pretrained DL model is a good strategy. Different methods are required for different detection tasks. Different tasks and the lack of dataset result in no uniform standard for measuring detection methods.

E. Segmentation

Image segmentation is the technology and process of dividing pixels of images into specific classes with unique properties and proposing the object of interest. The ROI for sonar images is the highlight and shadow areas caused by an object, which gets its size and shape. The result of sonar image segmentation is a masked image that specific regions are assigned to specific labels. Image segmentation techniques based on DL are commonly divided into two categories: semantic segmentation and instance segmentation.

An overview is in the following section. Similarly, the number of papers for different segmentation methods is presented in Fig. 10 and Table VI lists semantic and instance segmentation methods.

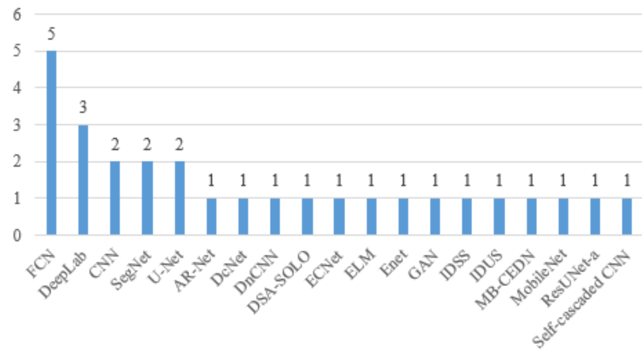


Fig. 10. Number of papers for different segmentation methods.

1) Semantic Segmentation: It only assigns a class label to each pixel. FCN [30], [97], [98], DeepLab [99], DeepLabV3+ [100], CNN [6], [32], [101], SegNet [102], and U-Net [103] were used for semantic segmentation of sonar images. FCN could join a ladder network architecture to fine-tune. The Markov random field could optimize the smoothing results of FCN and could be combined with DeepLab to identify seabed information. DeepLabV3+ added a symmetrical information synthesis module (SISM) to weaken the strong echoes of water column. The model was trained using a coarse-to-fine segmentation strategy. The result achieved 1.1 pixels of mean error. A CNN framework (AR-Net) [101] integrated the advantages of SegNet network structure symmetry and U-Net. The effectiveness was verified by comparing six different segmentation models, including FCN, Deeplab, U-Net, SegNet, ResUNet-a, and AR-Net.

Real-time segmentation tasks and encoder–decoder networks are hotly discussed. Efficient network (ENet) [102], efficient convolutional network (ECNet) [104], dilated CNN (DcNet) [105], and multibranch convolutional encoder–decoder network (MB-CEDN) [96] are encoder–decoder networks and achieved real-time segmentation. In general, an encoder learns rich hierarchical features and a decoder restores full input-size resolution feature maps. In ENet, a single stream DNN with multiple side outputs optimized edge segmentation and weighted loss was used to alleviate the influence of the imbalance classification. DcNet solved various noises, such as the intensity inhomogeneity problem, reduced the spatial dimension of input images, and recovered the details of the target. Its core network is a block connection named DCblock, which used dilated convolution and depthwise separable convolution between the encoder and the decoder to attain more context of images. MB-CEDN was a semisupervised learning model and was proposed for multitarget segmentation. In addition, self-cascaded CNN (SC-CNN) [106], online sequential extreme learning machine (OS-ELM) [107], iterative deep unsupervised segmentation (IDUS), and iterative deep semisupervised segmentation (IDSS) [108] also accomplished real-time segmentation. SC-CNN was available for speckle noise and intensity inhomogeneity and simultaneously used local and global features. OS-ELM combined with time-varying gain (TVG) correction, speed correction, and ensemble empirical mode decomposition (EEMD) denoising. Real-time of IDUS and IDSS outperformed supervised

TABLE VI
SUMMARY OF DETECTION METHODS

Index	Paper	Method	Task	Sonar
Semantic segmentation	[97]	FCN	Target	——
	[98]	FCN	Highlight, shadow, background	Marine Sonic SSS
	[99]	DeepLab	Sand waves	——
	[102]	ENet, SegNet	Sand waves	Shark-S450D SSS
	[103]	U-Net, FCN	Seagrass, pothole	——
	[104]	ECNet	Lines	Shark-S450D
	[6]	CNN	Fish, non-fish	Blueprint Oculus M750d
	[30]	FCN	Seabed topography	——
	[30]	GAN	Cylinder, ball, brick	DIDSON
	[105]	DcNet	Sand wave, seabed reef	Dual-frequency SSS
	[107]	ELM, MobileNetV1	Sand waves	Shark-S450D SSS
	[100]	DeepLabV3+	Water column	Klein SSS, EdgeTech SSS, Benthos SIS-1624, DeepVision DE340, etc.
	[106]	Self-cascaded CNN	Highlight, shadow, background	Marine Sonic SSS and EdgeTech
	[109]	DnCNN	Drowning victim, sunken ship, aircraft	Lcocean and other sonars
	[96]	MB-CEDN	Debris, ordnance, munitions downed aircraft, shipwreck	Multi-element SAS mounted on Hydroid REMUS 600 UV
	[32]	CNN	Shipwreck, corpse, aircraft, sand wave	Tritech 1200i SSS
	[108]	IDUS, IDSS	Seven seabed types	HFSAS of UUV
	[101]	FCN, Deeplab, U-Net, SegNet, ResUNet-a, AR-Net	Artificial reef	NORBIT-iWBMS, Teledyne Reson SeaBat T50P
Instance segmentation	[110]	DSA-SOLO	Ship, plane, body	——

methods and applied to platform of low size, weight, and power. Similarly, MRF is available for refining the result. Besides, GAN [30] and deep CNN for image denoising (DnCNN) [109] were also used for segmentation. DnCNN integrated receptive field block and search attention mechanism to complete segmentation of a single object image. It retains more edge information and details.

2) Instance Segmentation: It has not only the class label but also an ID. A double split attention segmentation objects by locations (DSA-SOLO) [110] was proposed for SSS image segmentation, which introduced a DSA attention module to fuse spatial attention and channel attention. The model embedded DSA into the FPN network of SOLOv2 and was trained by a public dataset called sonar common target detection (SCTD) [95].

As seen from the above review, research focuses on semantic segmentation. FCN and encoder–decoder are typical semantic segmentation architectures. MRF is available for postprocessing the result of segmentation. Meanwhile, due to the lack of public datasets, each segmentation algorithm is used for specific segmentation tasks and targets. It is difficult to say which is the best method of segmentation. Instance segmentation also needs to be explored.

F. DL and Sonar Image Processing

DL techniques have been used for image processing tasks, including feature extraction, classification, detection, and segmentation. Also, DL models have been applied to various types of images, such as optical images, medical images, and microscopic images. Meanwhile, as mentioned above, DL has also been applied to the field of sonar image processing. DL models can capture more feature information and improve the accuracy of ATR tasks. Many DL methods were used to process sonar images to complete classification, detection, and segmentation of sonar images. This helps us easier to obtain the information of sonar images. However, most DL methods typically require large amounts of labeled and clean data for training. This is a huge challenge for sonar image processing. Sonar images are not easy to collect and have severe noise. These problems make the models overfitting and precision degradation. Meanwhile, no uniform standard measures the performance of DL models. Thus, sonar image processing utilizing DL techniques is still a valuable and meaningful task.

III. APPLICATIONS IN AQUACULTURE

Sonar sensors are powerful tools for intelligence aquaculture in muddy waters or sea environments. With the development of DL, sonar images [4] can be used for behavior monitoring and analysis, biomass estimation, identification and classification of marine biology, and habitat survey. Sonars can accurately sense and detect wide-range underwater scenes. Here, we review some typical applications based on DL in aquaculture.

A. Datasets Available for Aquaculture

Here are three public datasets available for aquaculture, including DIDSONARawFishDatasets, sonar image counting

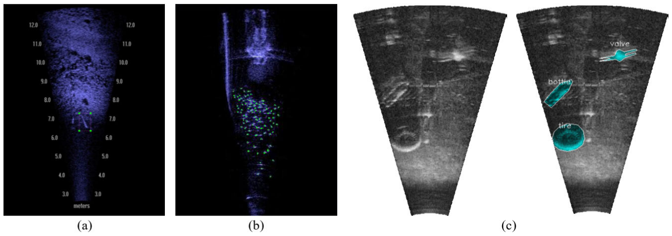


Fig. 11. Sample images of datasets. (a) DIDIDSONARawFishDatasets [119]. (b) Sonar image counting dataset [112]. (c) Marine debris dataset [113].

dataset, and marine debris dataset. Their sample images are shown in Fig. 11 and the details are listed in Table VII.

DIDIDSONARawFishDatasets was collected from the Ocqueoc River, a tributary of Lake Huron in northern Michigan, USA [111]. Sonar image counting dataset with low resolution (about 360 × 360) has severe noises and unstable visual characteristics. A region of interested is only around a third of the whole images. It includes 537 images and manually labeled dotted annotations. Among them, 417 images are used for training and validation, while 120 images are for testing [112]. Marine debris dataset was divided randomly into three sets of training, validation, and testing; 1000 images were selected for training, 251 for validation, and 617 for testing [113].

B. Behavior Monitoring and Analysis

Behavior analysis of aquatic organisms plays an essential role in aquaculture. Long-term monitoring is dispensable for realizing the growth and health status of the aquatic organisms in aquaculture. Sonar is a harmless device and allows continuous monitoring and observation using images.

Some highly accurate imaging sonars [5], [114] have been used to observe fish to improve the efficiency of surveillance of fish. For monitoring of sardines at night, sonar images collected by ARIS Explorer 1800 were translated into optical images using cGAN [115]. For detection and identification of migrating American Eel Anguilla rostrata [10], CNN was used to distinguish between eels and noneel objects in sonar images of ARIS Explorer 1800. French et al. [53] and Gorpincenko et al. [116] used ARIS Explorer 3000 to collect jellyfish sonar images and detected jellyfish blooms using DNN.

To sum up, studies include nighttime fish monitoring, migration behavior monitoring, and reproductive behavior monitoring that apply sonar images and DL to behavior analysis and monitoring. Feeding behavior, abnormal behavior, and so on have not been studied.

C. Biomass Estimation

Effective biomass estimation is beneficial for aquaculture management. Biomass estimation is a widespread application field of imaging sonar, such as fish counting. However, fishes in sonar images exhibit unreliable appearance characters and vary significantly in shape and size. These render great challenges for counting. Traditional biomass estimation used manual counting and semiautomated counting [117], which are slow and time-consuming. A reliable solution for estimation is to use DL.

High-resolution sonar images contribute to accurate assessment of aquatic resource. Recent counts regression network [112], CNNs [118], and YOLO [119] were used to count fish. Besides, the automated counting of tuna schools was completed by a morphometric classification model (MCM) using medium-range sonar images.

D. Identification and Classification of Marine Biology

Biodiversity is vital to marine ecosystems. The identification and classification of biology is necessary to monitor biodiversity. It is a big challenge for species classification to use sonar images. Morphological features (e.g., size and shape) and motion features (e.g., swimming speed, direction, and trajectory) are key information to identify and classify biology [117]. Traditional species identification methods include feature- and shadow-based methods, but these features are difficult to design and extract. DL-based methods are an efficient solution. Kandimalla et al. [119] used YOLO and mask-RCNN to classify eight distinct species of fish on DIDSONARawFishDatasets. French et al. [53] and Gorpincenko et al. [116] used CNN to class six kinds of jellyfish, fish, seaweed, artifacts, sediment, and background. Zang et al. [10] classed eels and noneel objects using CNN. Zang et al. [10], French et al. [53], and Gorpincenko et al. [116] used ARIS series sonars to collet sonar images. However, the existing DL-based methods for identification and classification of biology are relatively few. These methods focused on classifying specific species.

E. Habitat Survey

It is vital for biodiversity and environment. Habitat mapping is an important task, which help monitoring, managing, and preserving ecosystems. Sonars help acquiring underwater images, but resolution and clarity are low. For marine habitat mapping, Feldens et al. [89] and Feldens [90] used a single-stage residual network to improve the resolution of the rock sonar images and the detection precision for small rocks. Caretti et al. [120] used U-net to classify reef pixels for sonar images. Besides, a Gemini 720i multibeam FLS is used for detecting fishing nets in real time in the sea using MRF-Net [93].

To sum up, DL-based sonar image processing has relatively few applications in aquaculture. Most applications are still in the manual and software semiautomated phase. Sonar image processing based on DL still has great development potential in complex water environment such as large seine net and deep-sea aquaculture.

IV. Challenges and Future Research Directions

As mentioned earlier, DL shows its advantages and is an effective way for sonar image processing. However, the intrinsic characteristics of sonar images and the absence of datasets cause a great challenge for DL model training and application in aquaculture. Hence, it is essential to analyze and consider challenges and future research directions from four aspects: sonar image, dataset, method, and application.

TABLE VII
SUMMARY OF THE PUBLIC DATASETS FOR AQUACULTURE

Dataset	sonar	Number of species	object	Number/size of images	Application	Reference
DIDSONARawFishDatasets	DIDSON	8	Eight kinds of fish	524 clips	Fish classification and fishery assessment	[111]
Sonar image counting dataset	Online collection	1	Fish	537 images	Fish counting	[112]
Marine debris dataset	ARIS Explorer 3000	12	marine debris and distractor objects (tires, hooks, valves, etc.)	1868 images	Detection and segmentation of marine debris	[113]

A. Intrinsic Characteristic of Sonar Images

Sonar imaging is highly susceptible to noise. Acoustic images have less effective information and have no obvious texture features and spectral features, such as local variations, color, and energy, only geometrical features such as length and area. There is a lot of noise in sonar images, such as environmental noise, reverberation, and self-noise. Environmental noise is the background and ambient noise, which is ubiquitous in underwater. Due to multipath effects, Doppler effect, sidelobe interference, and the acoustic wave scattering and reflection formed by different waters and underwater geological types, they create reverberation. In addition, self-noise is caused by sonar equipment itself. All these factors lead to strong speckle noise, low contrast, weak feature information, and poor resolution of sonar images [12]. This makes it hard to distinguish targets, clutter, and background signatures.

Objects in sonar images have deformation, edge blur, and other problems [121], and the resolution of sonar images is low. These data characteristics cause serious interference to identify and detect targets and make it difficult to process and extract feature information. A typical example is that when some objects such as fishes are far away from sonar, they will become tiny and circular. It is quite confused to tell the difference between objects and noise [112].

B. Absence of Sonar Image Dataset

The dataset construction is changing and time-consuming. For sonar images, there are very few open-access datasets due to acquisition difficulty, high cost, and confidentiality. Thus, the lack of dataset, high cost of data annotation, and the insufficiency of data diversity are challenging problems for DL. Hence, building public baseline datasets is a future research direction. At present, different measures for collecting and augmenting sonar data for DL-based algorithms are practiced widely.

The common ways to collect underwater sonar images are mainly divided into four categories: public datasets, collected from the Internet, experimental acquisition, and synthesis and simulation of realistic datasets. They are listed and described in the following section.

1) Public Datasets in Aquaculture: It can be seen from Section III-A that the number of public datasets is scarce for sonar images in aquaculture. There are only three public datasets. Moreover, the cost of human annotation is expensive. The low data volume or bias caused by incomplete annotations is also a question that needs to be solved. Therefore, it is vital to establish large public datasets.

2) Collected From Internet: Some research collected sonar images from the Internet, but there are some flaws. First, images are unstructured and there are a large number of useless images for target. Second, images are obtained in different environments using different sonars. Finally, the number and size of sonar image are unbalanced and different.

3) Experiment Acquisition: In Section II, many papers used different sonars to collect sonar images. This method can obtain the corresponding data according to experimental needs and personalized scheme. Though it is the best way, there are still some problems. Underwater environments are complex and sonar sensors are expensive. Many experimenters do not have the appropriate experimental conditions. Besides, collecting and processing data need to spend a lot of time.

4) Synthesis and Simulation Realistic Datasets: For some practical applications where data are difficult and insufficient to obtain, synthesizing sonar images is desirable [122]. To solve this question, some techniques are deployed, including sonar data simulation [123], [124], GAN [125], [126], data augmentation [127], and transfer learning [128]. However, the gap between synthesized and real sonar image data fundamentally limits the performance of the analysis and process of sonar images.

The performance of DL models grows logarithmically as the pretraining dataset increases [129]. However, the absence of public sonar image datasets and expensive data annotation cost are the most challenging problems. Building and sharing sonar image datasets is an important research direction. Moreover, the data structures with various labels can be better learned using plentiful variations. Therefore, obtaining rich and abundant useful samples is critical for training and building a high-performance DL model. Research on augmenting and generating sonar images is worth considering.

C. Method

DL is a powerful tool and has made breakthroughs in sonar image processing. It needs large-scale data. The small training set can lead to low network efficiency and overfitting. Meanwhile, the strong noise and clarity of sonar images also pose a challenge to ATR tasks. Therefore, we consider the following ways to improve the DL performance.

Sonar image denoising based on DL is worth considering. The strong speckle and crosstalk noise make it difficult to process ATR tasks. Hence, how to denoise and reserve high- and low-frequency information for sonar images is an essential research direction. For supervised learning, clean and noisy–clean image pairs are needed. It is difficult and expensive to collect clean sonar images. Hence, self-supervised methods and unsupervised learning are promising approaches.

Overfitting is a serious problem for sonar image processing due to the lack of datasets. A successful DL model is impeded by the lack of large annotated sonar images. Transfer learning is an effective solution for insufficient and unbalanced data. In addition, data enhancement, dropout, regularization, and residual connection are also considered.

The state-of-the-art models on optical images apply to sonar images. Some methods, e.g., video swin transformer [130] and swin transformer [131], have not yet been applied to sonar images. As seen from Section II, there are fewer articles based on attention mechanism. Besides, comparison with traditional methods is necessary to show better performance of DL models.

D. Application

As seen from Section III, research of sonar image processing based on DL in aquaculture is relatively few. Some applications are still in the manual and semiautomated stages. Many of these studies are in laboratory settings, and we need more real-world data to complete. For behavior analysis, including movement behavior and feeding behavior, these behaviors also need to be studied. In aquaculture, there are more complex scenarios for fishery, such as multitarget, high density, small target, and motion blur. These factors make noise more serious and target identification more difficult [132]. It is feasible to use sonar image geometrical features such as length and area to estimate length of organisms such as fish. Biomass estimation and classification of species are also worth exploring in aquaculture. The length features of sonar images can be applied.

V. Discussion and Conclusion

This article reviews sonar image processing based on DL and its application in aquaculture. First, this article overviews the methods of sonar image denoising and ATR tasks based on DL. Then, applications in aquaculture are introduced. Finally, challenges and future research directions are discussed for sonar image processing and application in aquaculture. The literature review shows that the lack of public dataset and the severe noise of sonar images are the most challenging problems. Methods based on CNN and transfer learning are important ways for sonar image processing.

Speckle noise and crosstalk noise of sonar images are the main interference factors. The noise removal of sonar images is mostly done by traditional methods, and only two articles are based on DL. Classification tasks are the most studied. CNN is the most popular classification method, followed by VGG and ResNet. The use of DL in sonar images is a few years behind other images. For detection and segmentation tasks, the application of the most advanced methods still needs to work hard. How to better apply DL and the most advanced methods is worth discussing. The lack of public datasets in aquaculture is also worthy of attention. GAN and transfer learning are feasible solutions. GAN can use optical images to generate more sonar images. Fine-tuning pretraining network is a reliable method for low volume of dataset by transfer learning. The research of sonar image processing based on DL in aquaculture is relatively few. Migration behavior and reproduction behavior have been studied to some extent, but movement behavior and feeding behavior still need to be considered. Some complex scenarios for fishery, such as multitarget, high density, and small target, can cause more serious noise. This is also a problem that needs to be solved. In the end, length estimation, biomass estimation, and classification of species are also worth exploring in aquaculture.

References

[1] H. K. Alaie and H. Farsi, "Passive sonar target detection using statistical classifier and adaptive threshold," *Appl. Sci.*, vol. 8, no. 1, p. 61, Jan. 2018.

[2] A. Abu and R. Diamant, "Enhanced fuzzy-based local information algorithm for sonar image segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 445–460, 2020.

[3] Y. Huang, W. Li, and F. Yuan, "Speckle noise reduction in sonar image based on adaptive redundant dictionary," *J. Mar. Sci. Eng.*, vol. 8, no. 10, p. 761, Sep. 2020.

[4] U. Anitha, S. Malarkkan, G. D. A. Jebaselvi, and R. Narmadha, "Sonar image segmentation and quality assessment using prominent image processing techniques," *Appl. Acoust.*, vol. 148, pp. 300–307, May 2019.

[5] F. Martignac, A. Daroux, J. Bagliniere, D. Ombredane, and J. Guillard, "The use of acoustic cameras in shallow waters: New hydroacoustic tools for monitoring migratory fish population. A review of DIDSON technology," *Fish Fisheries*, vol. 16, no. 3, pp. 486–510, Sep. 2015.

[6] J. H. Christensen, L. V. Mogensen, and O. Ravn, "Deep learning based segmentation of fish in noisy forward looking MBES images," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 14546–14551, 2020.

[7] B. Teng and H. Zhao, "Underwater target recognition methods based on the framework of deep learning: A survey," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 6, pp. 1–12, 2020.

[8] L. C. F. Domingos, P. E. Santos, P. S. M. Skelton, R. S. A. Brinkworth, and K. Sammut, "A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance," *Sensors*, vol. 22, no. 6, p. 2181, Mar. 2022.

[9] Y. Steiniger, D. Kraus, and T. Meisen, "Survey on deep learning based computer vision for sonar imagery," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105157.

[10] X. Zang, T. Yin, Z. Hou, R. P. Mueller, Z. D. Deng, and P. T. Jacobson, "Deep learning for automated detection and identification of migrating American eel *Anguilla rostrata* from imaging sonar data," *Remote Sens.*, vol. 13, no. 14, p. 2671, Jul. 2021.

[11] J. Xiao, W. Zou, S. Zhang, J. Lei, W. Wang, and Y. Wang, "Video denoising algorithm based on improved dual-domain filtering and 3D block matching," *IET Image Process.*, vol. 12, no. 12, pp. 2250–2257, Dec. 2018.

[12] X. Wang, Q. Li, J. Yin, X. Han, and W. Hao, "An adaptive denoising and detection approach for underwater sonar image," *Remote Sens.*, vol. 11, no. 4, p. 396, Feb. 2019.

[13] Z. Tang, G. Ma, J. Lu, Z. Wang, B. Fu, and Y. Wang, "Sonar image mosaic based on a new feature matching method," *IET Image Process.*, vol. 14, no. 10, pp. 2149–2155, Aug. 2020.

[14] W. Chen, Z. Liu, H. Zhang, M. Chen, and Y. Zhang, "A submarine pipeline segmentation method for noisy forward-looking sonar images using global information and coarse segmentation," *Appl. Ocean Res.*, vol. 112, Jul. 2021, Art. no. 102691.

[15] J. Liu and S. Osher, "Block matching local SVD operator based sparsity and TV regularization for image denoising," *J. Sci. Comput.*, vol. 78, no. 1, pp. 607–624, Jan. 2019.

[16] J. Chanussot, F. Maussang, and A. Hetet, "Scalar image processing filters for speckle reduction on synthetic aperture sonar images," in *Proc. OCEANS MTS/IEEE*, vol. 4, Oct. 2002, pp. 2294–2301.

[17] W. Tian, Z. Chen, J. Shen, F. Huang, and L. Xu, "Underwater sonar image denoising through nonconvex total variation regularization and generalized Kullback–Leibler fidelity," *J. Ambient Intell. Hum. Comput.*, vol. 13, no. 11, pp. 5237–5251, Nov. 2022.

[18] Y. Jin, B. Ku, J. Ahn, S. Kim, and H. Ko, "Nonhomogeneous noise removal from side-scan sonar images using structural sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1215–1219, Aug. 2019.

[19] H. Liu, F. Yang, S. Zheng, Q. Li, D. Li, and H. Zhu, "A method of sidelobe effect suppression for multibeam water column images based on an adaptive soft threshold," *Appl. Acoust.*, vol. 148, pp. 467–475, May 2019.

[20] L. Zheng and K. Tian, "Center affine filter based adaptive image despeckling method with preserving details," *Appl. Acoust.*, vol. 155, pp. 16–23, Dec. 2019.

[21] Y. Lu, M. Yang, and R. W. Liu, "DSPNet: Deep learning-enabled blind reduction of speckle noise," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 3475–3482.

[22] Y. Lu, R. W. Liu, F. Chen, and L. Xie, "Learning a deep convolutional network for speckle noise reduction in underwater sonar images," in *Proc. 11th Int. Conf. Mach. Learn. Comput. (ICMLC)*, Aug. 2019, pp. 445–450.

[23] M. Sung, H. Cho, T. Kim, H. Joe, and S.-C. Yu, "Crosstalk removal in forward scan sonar image using deep learning for object detection," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9929–9944, Nov. 2019.

[24] J. C. Isaacs, "Representational learning for sonar ATR," *Proc. SPIE*, vol. 9072, Jun. 2014, Art. no. 907203.

[25] J. McKay, I. Gerg, V. Monga, and R. G. Raj, "What's mine is yours: Pretrained CNNs for limited training sonar ATR," in *Proc. OCEANS, Anchorage*, Sep. 2017, pp. 1–7.

[26] P. Zhu, J. Isaacs, B. Fu, and S. Ferrari, "Deep learning feature extraction for target recognition and classification in underwater sonar images," in *Proc. IEEE 56th Annu. Conf. Decis. Control (CDC)*, Dec.2017, pp. 2724–2731.

[27] J. Rutledge et al., "Intelligent shipwreck search using autonomous underwater vehicles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 6175–6182.

[28] G. Divyabarathi, S. Shailesh, and M. V. Judy, "Object classification in underwater SONAR images using transfer learning based ensemble model," in *Proc. Int. Conf. Adv. Comput. Commun. (ICACC)*, Oct. 2021, pp. 1–4.

[29] P. O. C. S. Ribeiro et al., "Underwater place recognition in unknown environments with triplet based acoustic image retrieval," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 524–529.

[30] H. Wang, N. Gao, Y. Xiao, and Y. Tang, "Image feature extraction based on improved FCN for UUV side-scan sonar," *Mar. Geophys. Res.*, vol. 41, no. 4, pp. 1–17, Dec. 2020.

[31] Z. Cheng, G. Huo, and H. Li, "A multi-domain collaborative transfer learning method with multi-scale repeated attention mechanism for underwater side-scan sonar image classification," *Remote Sens.*, vol. 14, no. 2, p. 355, Jan. 2022.

[32] Z. Wang, J. Guo, W. Huang, and S. Zhang, "Side-scan sonar image segmentation based on multi-channel fusion convolution neural networks," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5911–5928, Mar. 2022.

[33] H.-T. Nguyen, E.-H. Lee, and S. Lee, "Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body," *Sensors*, vol. 20, no. 1, p. 94, Dec. 2019.

[34] S. Lee, B. Park, and A. Kim, "A deep learning based submerged body classification using underwater imaging sonar," in *Proc. 16th Int. Conf. Ubiquitous Robots (UR)*, Jun. 2019, pp. 106–112.

[35] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020.

[36] A. Aleem, S. Tehsin, S. Kausar, and A. Jameel, "Target classification of marine debris using deep learning," *Intell. Autom. Soft Comput.*, vol. 32, no. 1, pp. 73–85, 2022.

[37] X. Qin, X. Luo, Z. Wu, and J. Shang, "Optimizing the sediment classification of small side-scan sonar images based on deep learning," *IEEE Access*, vol. 9, pp. 29416–29428, 2021.

[38] I. Dzieciuch, D. Gebhardt, C. Barngrover, and K. Parikh, "Non-linear convolutional neural network for automatic detection of mine-like objects in sonar imagery," in *Proc. 4th Int. Conf. Appl. Nonlinear Dyn. (ICAND)*, in Lecture Notes in Networks and Systems, vol. 6, 2017, pp. 309–314.

[39] X. Wang, J. Jiao, J. Yin, W. Zhao, X. Han, and B. Sun, "Underwater sonar image classification using adaptive weights convolutional neural network," *Appl. Acoust.*, vol. 146, pp. 145–154, Mar. 2019.

[40] Y. Xu, X. Wang, K. Wang, J. Shi, and W. Sun, "Underwater sonar image classification using generative adversarial network and convolutional neural network," *IET Image Process.*, vol. 14, no. 12, pp. 2819–2825, Oct. 2020.

[41] N. Nayak, M. Nara, T. Gambin, Z. Wood, and C. M. Clark, "Machine learning techniques for AUV side-scan sonar data feature extraction as applied to intelligent search for underwater archaeological sites," in *Field and Service Robotics* (Springer Proceedings in Advanced Robotics), vol. 16. Singapore: Springer, 2021, pp. 219–233.

[42] D. Polap, N. Wawrzyniak, and M. Wlodarczyk-Sielicka, "Side-scan sonar analysis using ROI analysis and deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4206108.

[43] D. Gebhardt, K. Parikh, I. Dzieciuch, M. Walton, and N. A. V. Hoang, "Hunting for naval mines with deep neural networks," in *Proc. OCEANS, Anchorage*, Sep. 2017, pp. 1–5.

[44] S. L. Phung et al., "Mine-like object sensing in sonar imagery with a compact deep learning architecture for scarce data," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2019, pp. 1–7.

[45] X. Ye, C. Li, S. Zhang, P. Yang, and X. Li, "Research on side-scan sonar image target classification method based on transfer learning," in *Proc. OCEANS MTS/IEEE Charleston*, Oct. 2018, pp. 1–6.

[46] A. Bouzerdoum et al., "Improved deep-learning-based classification of mine-like contacts in sonar images from autonomous underwater vehicles," in *Proc. Underwater Acoust. Conf. Exhib. Ser. Conf.*, 2019, pp. 179–186.

[47] M. Ochal, J. Vazquez, Y. Petillot, and S. Wang, "A comparison of few-shot learning methods for underwater optical and sonar image classification," in *Proc. Global Oceans, Singap., U.S. Gulf Coast*, Oct. 2020, pp. 1–10.

[48] X. Luo, X. Qin, Z. Wu, F. Yang, M. Wang, and J. Shang, "Sediment classification of small-size seabed acoustic images using convolutional neural networks," *IEEE Access*, vol. 7, pp. 98331–98339, 2019.

[49] P. Chapple, T. Dell, and D. Bongiorno, "Enhanced detection and classification of mine-like objects using situational awareness and deep," in *Proc. Underwater Acoust. Conf. Exhib.*, 2017, pp. 529–536.

[50] C. Li, X. Ye, D. Cao, J. Hou, and H. Yang, "Zero shot objects classification method of side scan sonar image based on synthesis of pseudo samples," *Appl. Acoust.*, vol. 173, Feb. 2021, Art. no. 107691.

[51] W. Gong, J. Tian, and J. Liu, "Underwater object classification method based on depthwise separable convolution feature fusion in sonar images," *Appl. Sci.*, vol. 12, no. 7, p. 3268, Mar. 2022.

[52] M. Valdenegro-Toro, "Object recognition in forward-looking sonar images with convolutional neural networks," in *Proc. OCEANS MTS/IEEE Monterey*, Sep. 2016, pp. 1–6.

[53] G. French et al., "JellyMonitor: Automated detection of jellyfish in sonar images using neural networks," in *Proc. 14th IEEE Int. Conf. Signal Process. (ICSP)*, Aug. 2018, pp. 406–412.

[54] L. He, J. Zhao, Z. Qiu, and J. Feng, "High-accuracy seabed sediment classification using multi-beam acoustic backscatter data," in *Proc. OCEANS, Chennai*, Feb. 2022, pp. 1–23.

[55] J. Wan et al., "MBES seabed sediment classification based on a decision fusion method using deep learning model," *Remote Sens.*, vol. 14, no. 15, p. 3708, Aug. 2022.

[56] D. P. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2497–2502.

[57] D. P. Williams, "Demystifying deep convolutional neural networks for sonar image classification," *Proc. 4th Underwater Acoust. Conf.*, Skiathos, Greece, Sep. 2017, pp. 513–520.

[58] J. L. Chen and J. E. Summers, "Deep neural networks for learning classification features and generative models from synthetic aperture sonar big data," in *Proc. Meetings Acoust.*, 2016, vol. 29, no. 1, Art. no. 032001.

[59] A. I. Karjalainen, R. Mitchell, and J. Vazquez, "Training and validation of automatic target recognition systems using generative adversarial networks," in *Proc. Sensor Signal Process. Defence Conf. (SSPD)*, May 2019, pp. 1–5.

[60] A. P. Galusha, J. Dale, J. Keller, and A. Zare, "Deep convolutional neural network target classification for underwater synthetic aperture sonar imagery," *Proc. SPIE*, vol. 11012, May 2019, Art. no. 1101205.

[61] D. P. Williams, "Convolutional neural network transfer learning for underwater object classification," *Proc. Inst. Acoust.*, vol. 40, no. Pt2, pp. 123–131, 2018.

[62] D. P. Williams, "Transfer learning with SAS-image convolutional neural networks for improved underwater target classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 78–81.

[63] D. C. D'Ales, D. P. Williams, and S. Dugelay, "Target classification using multi-view synthetic aperture sonar Imagery," *Proc. 5th Underwater Acoust. Conf. Exhib.*, 2019, pp. 227–233.

[64] T. Berthomier, D. P. Williams, B. d'Alès, and S. Dugelay, "Exploiting auxiliary information for improved underwater target classification with convolutional neural networks," in *Proc. Global Oceans, Singap., U.S. Gulf Coast*, Oct. 2020, pp. 1–10.

[65] K. Zhu, J. Tian, and H. Huang, "Underwater object images classification based on convolutional neural network," in *Proc. IEEE 3rd Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2018, pp. 301–305.

[66] N. Warakagoda and Ø. Midtgaard, "Transfer-learning with deep neural networks for mine recognition in sonar images," *Proc. Inst. Acoust.*, vol. 40, pp. 115–122, Oct. 2018.

[67] D. P. Williams, "Exploiting phase information in synthetic aperture sonar images for target classification," in *Proc. OCEANS, MTS/IEEE Kobe Techno-Oceans (OTO)*, May 2018, pp. 1–6.

[68] I. Gerg and D. Williams, "Additional representations for improving synthetic aperture sonar classification using convolutional neural networks," *Proc. Inst. Acoust.*, vol. 40, pp. 11–22, Aug. 2018.

[69] D. Williams, R. Hamon, and I. Gerg, "On the benefit of multiple representations with convolutional neural networks for improved target classification using sonar data," in *Proc. 5th Underwater Acoust. Conf. Exhib.*, Jul. 2019, p. 2912.

[70] D. P. Williams, "On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery," *IEEE J. Ocean. Eng.*, vol. 46, no. 1, pp. 236–260, Jan. 2021.

[71] I. Quidu, N. Burlet, J.-P. Malkasse, and F. Florin, "Automatic classification for MCM systems," in *Proc. Eur. Oceans*, vol. 2, 2005, pp. 844–847.

[72] T. Berthold, A. Leichter, B. Rosenhahn, V. Berkhahn, and J. Valerius, "Seabed sediment classification of side-scan sonar data using convolutional neural networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2017, pp. 1–8.

[73] X. Zhou, K. Yang, and R. Duan, "Deep learning based on striation images for underwater and surface target classification," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1378–1382, Sep. 2019.

[74] I. D. Gerg and V. Monga, "Structural prior driven regularized deep learning for sonar image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4200416.

[75] D. P. Williams and S. Dugelay, "Multi-view SAS image classification using deep learning," in *Proc. OCEANS MTS/IEEE Monterey*, Sep. 2016, pp. 1–9.

[76] L. Jiang, T. Cai, Q. Ma, F. Xu, and S. Wang, "Active object detection in sonar images," *IEEE Access*, vol. 8, pp. 102540–102553, 2020.

[77] Z. Fan, W. Xia, X. Liu, and H. Li, "Detection and segmentation of underwater objects from forward-looking sonar based on a modified mask RCNN," *Signal, Image Video Process.*, vol. 15, no. 6, pp. 1135–1143, Sep. 2021.

[78] L. Xu, X. Wang, and X. Wang, "Shipwrecks detection based on deep generation network and transfer learning with small amount of sonar images," in *Proc. IEEE 8th Data Driven Control Learn. Syst. Conf. (DDCLS)*, May 2019, pp. 638–643.

[79] D. Einsidler, M. Dhanak, and P.-P. Beaujean, "A deep learning approach to target recognition in side-scan sonar imagery," in *Proc. OCEANS MTS/IEEE Charleston*, Oct. 2018, pp. 1–4.

[80] Y. Steiniger, J. Groen, J. Stoppe, D. Kraus, and T. Meisen, "A study on modern deep learning detection algorithms for automatic target recognition in sidescan sonar images," *Proc. Meetings Acoust.*, vol. 44, no. 1, 2021, Art. no. 070010.

[81] T. Yulin, S. Jin, G. Bian, and Y. Zhang, "Shipwreck target recognition in side-scan sonar images by improved YOLOv3 model based on transfer learning," *IEEE Access*, vol. 8, pp. 173450–173460, 2020.

[82] X. Cao, L. Ren, and C. Sun, "Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1968–1981, Mar. 2023.

[83] X. Cao, L. Ren, and C. Sun, "Research on obstacle detection and avoidance of autonomous underwater vehicle based on forward-looking sonar," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 16, 2022.

[84] W. Kong et al., "YOLOv3-DPFIN: A dual-path feature fusion neural network for robust real-time sonar target detection," *IEEE Sensors J.*, vol. 20, no. 7, pp. 3745–3756, Apr. 2020.

[85] X. Fan, L. Lu, P. Shi, and X. Zhang, "A novel sonar target detection and classification algorithm," *Multimedia Tools Appl.*, vol. 81, no. 7, pp. 10091–10106, Mar. 2022.

[86] Y. Yu, J. Zhao, Q. Gong, C. Huang, G. Zheng, and J. Ma, "Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5," *Remote Sens.*, vol. 13, no. 18, p. 3555, Sep. 2021.

[87] G. Neves, M. Ruiz, J. Fontinele, and L. Oliveira, "Rotated object detection with forward-looking sonar in underwater applications," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112870.

[88] H. Thanh Le, S. L. Phung, P. B. Chapple, A. Bouzerdoum, C. H. Ritz, and L. C. Tran, "Deep Gabor neural network for automatic detection of mine-like objects in sonar imagery," *IEEE Access*, vol. 8, pp. 94126–94139, 2020.

[89] P. Feldens, A. Darr, A. Feldens, and F. Tauber, "Detection of boulders in side scan sonar mosaics by a neural network," *Geosciences*, vol. 9, no. 4, p. 159, Apr. 2019.

[90] P. Feldens, "Super resolution by deep learning improves Boulder detection in side scan sonar backscatter mosaics," *Remote Sens.*, vol. 12, no. 14, p. 2284, Jul. 2020.

[91] J. M. Topple and J. A. Fawcett, "MiNet: Efficient deep learning automatic target recognition for small autonomous vehicles," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1014–1018, Jun. 2021.

[92] K. Denos, M. Ravaut, A. Fagette, and H.-S. Lim, "Deep learning applied to underwater mine warfare," in *Proc. OCEANS, Aberdeen*, Jun. 2017, pp. 1–7.

[93] R. Qin et al., "Multiple receptive field network (MRF-Net) for autonomous underwater vehicle fishing net detection using forward-looking sonar images," *Sensors*, vol. 21, no. 6, p. 1933, Mar. 2021.

[94] Z. Wang, S. Zhang, W. Huang, J. Guo, and L. Zeng, "Sonar image target detection based on adaptive global feature enhancement network," *IEEE Sensors J.*, vol. 22, no. 2, pp. 1509–1530, Jan. 2022.

[95] P. Zhang, J. Tang, H. Zhong, M. Ning, D. Liu, and K. Wu, "Self-trained target detection of radar and sonar images using automatic deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4701914.

[96] I. J. Sledge, M. S. Emigh, J. L. King, D. L. Woods, J. T. Cobb, and J. C. Príncipe, "Target detection and segmentation in circular-scan synthetic aperture sonar images using semisupervised convolutional encoder–decoders," *IEEE J. Ocean. Eng.*, vol. 47, no. 4, pp. 1099–1128, Oct. 2022.

[97] J. Chen and J. E. Summers, "Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (SAS) images," in *Proc. 173rd Meeting Acoust. Soc. Amer. 8th Forum Acusticum*, vol. 30, 2017, Art. no. 055018.

[98] Y. Song, Y. Zhu, G. Li, C. Feng, B. He, and T. Yan, "Side scan sonar segmentation using deep convolutional neural network," in *Proc. OCEANS, Anchorage*, Sep. 2017, pp. 1–4.

[99] F. Yu et al., "Segmentation of side scan sonar images on AUV," in *Proc. IEEE Underwater Technol. (UT)*, Apr. 2019, pp. 1–4.

[100] G. Zheng, H. Zhang, Y. Li, and J. Zhao, "A universal automatic bottom tracking method of side scan sonar data based on semantic segmentation," *Remote Sens.*, vol. 13, no. 10, p. 1945, May 2021.

[101] Z. Dong, Y. Liu, L. Yang, Y. Feng, J. Ding, and F. Jiang, "Artificial reef detection method for multibeam sonar imagery based on convolutional neural networks," *Remote Sens.*, vol. 14, no. 18, p. 4610, Sep. 2022.

[102] K. Li et al., "Real-time segmentation of side scan sonar imagery for AUVs," in *Proc. IEEE Underwater Technol. (UT)*, Apr. 2019, pp. 1–5.

[103] M. Rahnemoonfar and D. Dobbs, "Semantic segmentation of underwater sonar imagery with deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 9455–9458.

[104] M. Wu et al., "ECNet: Efficient convolutional networks for side scan sonar image segmentation," *Sensors*, vol. 19, no. 9, p. 2009, Apr. 2019.

[105] X. Zhao, R. Qin, Q. Zhang, F. Yu, Q. Wang, and B. He, "DcNet: Dilated convolutional neural networks for side-scan sonar image semantic segmentation," *J. Ocean Univ. China*, vol. 20, no. 5, pp. 1089–1096, Oct. 2021.

[106] Y. Song, B. He, and P. Liu, "Real-time object detection for AUVs using self-cascaded convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 46, no. 1, pp. 56–67, Jan. 2021.

[107] R. Nian et al., "Towards characterizing and developing formation and migration cues in seafloor sand waves on topology, morphology, evolution from high-resolution mapping via side-scan sonar in autonomous underwater vehicles," *Sensors*, vol. 21, no. 9, p. 3283, May 2021.

[108] Y.-C. Sun, I. D. Gerg, and V. Monga, "Iterative, deep synthetic aperture sonar image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4206615.

[109] H. Xu, L. Zhang, M. J. Er, and Q. Yang, "Underwater sonar image segmentation based on deep learning of receptive field block and search attention mechanism," in *Proc. 4th Int. Conf. Intell. Auto. Syst. (ICoIAS)*, May 2021, pp. 44–48.

[110] H. Huang, Z. Zuo, B. Sun, P. Wu, and J. Zhang, "DSA-SOLO: Double split attention SOLO for side-scan sonar target segmentation," *Appl. Sci.*, vol. 12, no. 18, p. 9365, Sep. 2022.

[111] E. McCann, L. Li, K. Pangle, N. Johnson, and J. Eickholt, "An underwater observation dataset for fish classification and fishery assessment," *Sci. Data*, vol. 5, no. 1, Oct. 2018, Art. no. 180190.

[112] L. Liu, H. Lu, Z. Cao, and Y. Xiao, "Counting fish in sonar images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3189–3193.

[113] D. Singh and M. Valdenegro-Toro, "The marine debris dataset for forward-looking sonar semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 3734–3742.

[114] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (DIDSON)," in *Proc. Int. Symp. Underwater Technol.*, Apr. 2002, pp. 187–192.

[115] K. Terayama, K. Shin, K. Mizuno, and K. Tsuda, "Integration of sonar and optical camera images using deep neural network for fish monitoring," *Aquacultural Eng.*, vol. 86, Aug. 2019, Art. no. 102000.

[116] A. Gorpincenko, G. French, P. Knight, M. Challiss, and M. Mackiewicz, "Improving automated sonar video analysis to notify about jellyfish blooms," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4981–4988, Feb. 2021.

[117] Y. Wei, Y. Duan, and D. An, "Monitoring fish using imaging sonar: Capacity, challenges and future perspective," *Fish Fisheries*, vol. 23, no. 6, pp. 1347–1370, Nov. 2022.

[118] R. M. Connolly et al., "Out of the shadows: Automatic fish detection from acoustic cameras," *Aquatic Ecol.*, pp. 1–12, May 2022.

[119] V. Kandimalla, M. Richard, F. Smith, J. Quirion, L. Torgo, and C. Whidden, "Automated detection, classification and counting of fish in fish passages with deep learning," *Frontiers Mar. Sci.*, vol. 8, pp. 1–15, Jan. 2022.

[120] O. N. Caretti, D. R. Bohnenstiehl, and D. B. Eggleston, "Spatiotemporal variability in sedimentation drives habitat loss on restored subtidal oyster reefs," *Estuaries Coasts*, vol. 44, no. 8, pp. 2100–2117, Mar. 2021.

[121] V. Padmaja, V. Rajendran, and P. Vijayalakshmi, "Study on metal mine detection from underwater sonar images using data mining and machine learning techniques," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 5, pp. 5083–5092, May 2021.

[122] W. Wang, F. Wen, Z. Yan, and P. Liu, "Optimal transport for unsupervised denoising learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2104–2118, Feb. 2023.

[123] R. Cerqueira, T. Trocoli, G. Neves, S. Joyeux, J. Albiez, and L. Oliveira, "A novel GPU-based sonar simulator for real-time applications," *Comput. Graph.*, vol. 68, pp. 66–76, Nov. 2017.

[124] C. Huang, J. Zhao, Y. Yu, and H. Zhang, "Comprehensive sample augmentation by fully considering SSS imaging mechanism and environment for shipwreck detection under zero real samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5906814.

[125] Y. Jiang, B. Ku, W. Kim, and H. Ko, "Side-scan sonar image synthesis based on generative adversarial network for images in multiple frequencies," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1505–1509, Sep. 2021.

[126] M. Sung et al., "Realistic sonar image simulation using deep learning for underwater object detection," *Int. J. Control, Autom. Syst.*, vol. 18, no. 3, pp. 523–534, Mar. 2020.

[127] X. Zhang, Z. Wang, D. Liu, Q. Lin, and Q. Ling, "Deep adversarial data augmentation for extremely low data regimes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 15–28, Jan. 2021.

[128] S. Lee, B. Park, and A. Kim, "Deep learning based object detection via style-transferred underwater sonar images," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 152–155, 2019.

[129] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 843–852.

[130] Z. Liu et al., "Video Swin transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 3192–3201.

[131] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[132] T. Zhou, J. Si, L. Wang, C. Xu, and X. Yu, "Automatic detection of underwater small targets using forward-looking sonar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4207912.

Yingqian Chai received the B.S. degree in computer science and technology from China Agricultural University, Beijing, China, in 2019, where she is currently pursuing the M.S. degree with the College of Information and Electrical Engineering.

Her research interests include computer vision, deep learning, and sonar image processing.

Huihui Yu received the Ph.D. degree in agricultural information technology research from China Agricultural University, Beijing, China, in 2008.

She is a Lecturer of Computer Science and Technology with the School of Information Science and Technology, Beijing Forestry University, Beijing. She has worked in the research domains of agricultural information acquisition and information processing.

Ling Xu received the M.S. degree from the China Academy of Agricultural Mechanization Science, Beijing, China, in 2021. She is currently pursuing the Ph.D. degree with the College of Information and Electrical Engineering, China Agricultural University, Beijing.

Her research interests include computer vision, computer graphics, and machine learning.

Daoliang Li received the Ph.D. degree in agricultural engineering from China Agricultural University, Beijing, China, in 1999.

He is currently a Professor with the Department of Electronic Engineering, China Agricultural University, and the Director of the National Innovation Center for Digital Fishery, Beijing. His research interests include the agricultural Internet of Things, agricultural artificial intelligence, and agricultural information processing.

Yingyi Chen received the Ph.D. degree in agricultural engineering from China Agricultural University, Beijing, China, in 2008.

He is currently a Professor with the Department of Computer Engineering, China Agricultural University. His research interests include the agricultural Internet of Things, agricultural artificial intelligence, and agricultural information processing.