

## Research Article

# Design and Performance Evaluation of a Deep Neural Network for Spectrum Recognition of Underwater Targets

Dali Liu <sup>1,2</sup>, Xuchen Zhao,<sup>1</sup> Wenjing Cao,<sup>3</sup> Wei Wang,<sup>1</sup> and Yi Lu<sup>4</sup>

<sup>1</sup>School of Electronics and Information Engineering, Tiangong University, Tianjin, China

<sup>2</sup>State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

<sup>3</sup>School of Electrical Engineering and Automation, Tiangong University, Tianjin, China

<sup>4</sup>Tianjin Jinhang Computing Technology Research Institute, Tianjin, China

Correspondence should be addressed to Dali Liu; liudalidl@139.com

Received 21 March 2020; Revised 10 July 2020; Accepted 18 July 2020; Published 1 August 2020

Academic Editor: Anastasios D. Doulamis

Copyright © 2020 Dali Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the complexity of the underwater environment, underwater acoustic target recognition (UATR) has always been challenging. Although deep neural networks (DNN) have been used in UATR and some achievements have been made, the performance is not satisfactory when recognizing underwater targets with different Doppler shifts, signal-to-noise ratios (SNR), and interferences. In the paper, a one-dimensional convolutional neural network (1D-CNN) was proposed to recognize the line spectrums of Detection of Envelope Modulation on Noise (DEMON) spectrums of underwater target-radiated noise. Datasets of targets with different Doppler shifts, SNRs, and interferences were designed to evaluate the generalization performance of the proposed CNN. Experimental results show that compared with traditional multilayer perceptron (MLP) networks, the 1D-CNN model better performs in recognition of targets with different Doppler shifts and SNRs. The outstanding generalization ability of the proposed model shows that it is suitable for practical engineering applications.

## 1. Introduction

Underwater target recognition is one of the important functions of the sonar system, which plays a significant role in underwater situation perception. Due to the complex underwater environment, it is challenging for us to extract underwater target features. Therefore, underwater target recognition has always been a popular and challenging problem for sonar researchers [1].

Various sensors and signal processing methods are applied to extract features and identify underwater targets, such as Brillouin scattering [2], underwater optical imaging [3], and radiation noise feature extraction [4]. Based on these detection methods, some excellent projects used underwater interaction to raise awareness and access to European underwater cultural heritage [5]. For traditional underwater acoustic target recognition (UATR), the radiated noise of underwater targets and echoes reflected from these targets are the main signals applied to target recognition for passive

sonar [6, 7] and active sonar [8, 9], respectively. With the advantages of long detection distance and strong concealment, the passive recognition mode based on target-radiated noise has attracted more attention from researchers.

The passive recognition method mainly uses traditional detection of envelope modulation on noise (DEMON) and low-frequency analysis and recording (LOFAR) to extract target features. The UATR widely uses the DEMON method because it can get the invariable physical characteristics of the target. Wang et al. used the high-speed characteristic quantity (HSCQ) of the DEMON spectrum to classify high-speed underwater targets [10]. Wu et al. combined the GT cepstral coefficient with the characteristic vector of DEMON spectrum analysis to improve the accuracy and stability of target classification effectively [11]. Yao and Hui combined the analytical acoustic intensity flux (AAIF) with the DEMON spectrum to propose a multitarget detection technique [12]. Although the abovementioned results have achieved the feature extraction task, these processes often

require a large amount of human participation and are too dependent on subjective consciousness and experiences [13–15].

In the recent years, as the most popular deep learning model, the deep neural network (DNN) has attracted the interest of scholars in the field of UATR [16]. Yang et al. combined the auditory perception principle and convolutional neural network (CNN) to propose an auditory perception-inspired deep convolutional neural network (ADCNN), which used a CNN to extract features of different frequency components from signals and merged them at the fusion layer to achieve the classification of acoustic targets [17]. Choi et al. used the absolute values of matrix elements in the cross-spectrum density matrix (CSDM) to generate two additional matrices as input data and used them for training the CNN model [18]. Cao et al. combined the CNN architecture with a second-order pooling (SOP) and used a convolution layer to learn the local features of the data extracted by constant-Q transform (CQT), with achieving an end-to-end network to accomplish the classification of underwater targets [19]. Zhou et al. proposed a compound convolutional neural network based on the shared latent sparse (SLS) feature and deep belief network (DBN), using these two functions to learn fringe-based sonar images to improve the accuracy of classification [20]. Chen et al. proposed a method based on a convolutional neural network with residual units to recognize a time-frequency image of ship-radiated noise [21]. Wang et al. combined improved antinoise power-normalized cepstral coefficients (ia-PNCC) with a CNN and applied multitaper and normalized Gammatone filter banks to improve the antinoise capacity [22]. At the same time, the multilayer neural network and robust adaptive controller also optimized the target tracking control of the underdriven autonomous underwater vehicles (AUV), improving the anti-interference ability of underwater target recognition in the face of complex marine environments [23, 24]. Besides, multidimensional fusion networks [25] and the support vector machine (SVM) [26] are also used for UATR tasks.

Although the neural network has made some achievements in the field of radiated noise recognition of underwater targets, in actual engineering applications, there are still many problems that affect the recognition accuracy. Due to the relative movement between the underwater target and the sonar system, the sonar receives an underwater radiated noise containing the Doppler frequency shift. Also, due to the influence of the environment noise and the target distance, the signal-to-noise ratio (SNR) of the received radiated noise is not constant. Besides, because of environmental or human interferences, redundant spectrum lines may appear in the DEMON spectrum of the received radiated noise, or some spectrum lines may weaken or even disappear. When training a DNN, the training dataset cannot cover all the abovementioned situations. Though a trained DNN can recognize the targets in the dataset very well, it is challenging to recognize targets that are different from the training samples. As we all know, the acquisition of radiated noise data of underwater targets is costly. Therefore, this paper aims at improving the generalization ability of a DNN

to identify the underwater targets with different Doppler shifts, SNRs, and interferences in the case of a limited dataset.

First, the model of underwater target-radiated noise was built, and then, the subband processing was performed to obtain the DEMON spectrums. The DEMON spectrums of different kinds of targets constituted the datasets to train and validate the DNN. The DEMON spectrums of radiated noise were composed of line spectrums and continuous spectrums, and line spectrums were used for target classification and recognition in the paper. Validation datasets with different Doppler shifts, SNRs, and interferences were designed to evaluate the generalization performance of the DNN. By adjusting the structure and parameters of DNN, a one-dimensional CNN (1D-CNN) was proposed for the recognition of the radiated noise of underwater targets. To further understand the performance of the designed network, a multilayer perceptron (MLP) network was designed to compare with it. By evaluating the performance of the designed networks, the effects of the Doppler shift, SNR, and interferences on the network performance were obtained. The proposed 1D-CNN can recognize the radiated noise of targets in different situations and has outstanding generalization ability.

## 2. Signal Processing and Samples Generation

*2.1. DEMON Processing of Underwater Radiated Noise.* The passive sonar mainly uses ship-radiated noise to detect and locate underwater targets. Because the SNR of the target-radiated noise is extremely low and the effective signal is completely masked in the noise, the received signal needs to be processed to improve the SNR firstly.

The power spectrums of the ship's radiated noise are mixed spectrums composed of line spectrums and continuous spectrums. This paper focuses on line spectrums, which can reflect some periodic stable vibration sources on the target. The line spectrums consist of a series of discrete sinusoidal components with frequencies  $f_1, f_2, \dots, f_M$ , and the signal with the single-frequency carrier can be expressed as

$$L(t) = \left[ 1 + \sum_{i=1}^M p_i \sin(2\pi f_i t + \phi_i) \right] \cdot \cos \omega t, \quad (1)$$

where  $f_i$ ,  $p_i$ , and  $\phi_i$  are the frequency, modulation factor, and random initial phase of the  $i$ -th line spectrum, respectively,  $M$  is the number of the sinusoidal components, and  $\omega$  is the carrier frequency.

The signal model of the received ship-radiated noise can be written as

$$S(t) = L(t) + N(t), \quad (2)$$

where  $N(t)$  is the Gaussian white noise and  $L(t)$  is the underwater environment noise.

DEMON spectrum analysis is one of the important methods to recognize ship-radiated noise, and it can show us the inherent physical characteristics of the target, including the axial frequency and leaf frequency of the propeller [27].

The different frequencies and amplitudes of line spectrums in DEMON spectrums are the main features to identify different underwater targets.

To simplify the analysis,  $L(t)$  is supposed to have only one component, which can be written as

$$L'(t) = A(1 + m \sin \Omega t) \cdot \cos \omega t, \quad (3)$$

where  $A$  is the amplitude of the signal,  $m$  is the modulation factor,  $0 < m < 1$ , and  $\Omega$  is the frequency of the modulation signal.

To obtain the modulation signal  $\sin \Omega t$ , signal  $L'(t)$  is processed by absolute value demodulation; then,

$$\begin{aligned} |L'(t)| &= A(1 + m \sin \Omega t) \cdot |\cos \omega t|, \\ &= A(1 + m \sin \Omega t) \cdot \frac{4}{\pi} \cdot \left( \frac{1}{2} + \frac{1}{3} \cos 2\omega t - \frac{1}{3 \cdot 5} \cos 4\omega t + \frac{1}{5 \cdot 7} \cos 6\omega t + \dots \right), \\ &= \frac{2A}{\pi} + \frac{2A}{\pi} \cdot m \cdot \Omega t + \frac{4A}{3\pi} \cos 2\omega t \\ &\quad + \frac{2mA}{3\pi} \sin(\Omega + 2\omega)t + \frac{2mA}{3\pi} \sin(\Omega - 2\omega)t + \dots \end{aligned} \quad (4)$$

The signal  $|L'(t)|$  is mainly composed of a constant component,  $\sin \Omega t$ , and other high-frequency components related to the carrier frequency  $\omega$ . Therefore, the modulation signal can be obtained after  $|L'(t)|$  is filtered.

The discrete modulation signal,  $x(n)$ , is analyzed by Fourier transform, which can be written as

$$X(f) = \sum_{i=1}^N x(n) \cdot e^{-j2\pi f n N}. \quad (5)$$

Then, we obtain the DEMON spectrum of the radiated noise of underwater targets. As the carrier frequency  $\omega$  is unknown, the received radiated noise should be processed by a series of band pass filters before DEMON processing to avoid losing part of feature spectrums [28, 29].

In this paper, the radiated noise signal with the length of about 4.7 seconds and the sampling rate of 44.1 kHz was filtered by six band pass filters, and DEMON processing was applied to each obtained subband signals. The sampling rate of low pass filtered signals was reduced to 882 Hz before the Fourier transform, in order to reduce the calculation burden. Finally, the FFT length was set to 4096, and half of the FFT results constituted the DEMON spectrums with the length of 2048. Figure 1 shows the proposed DEMON processing.

The  $i$ -th channel DEMON spectrum is  $X_i(k)$ , where  $0 \leq i \leq 5$  and  $k$  is the  $k$ -th value of discrete Fourier transform  $0 \leq k \leq 2047$ . In order to further reduce the calculation of the DNN and retain the feature information as much as possible, for each frequency value  $k$ , the maximum value of the six channels was selected to compose a new DEMON spectrum, which can be described as

$$X'(k) = \max_i X_i(k). \quad (6)$$

The  $X_i(k)$  can form a dataset with 6-channel samples that contain all the features of the underwater target, and  $X'(k)$  can form a single-channel dataset by data compression, which can reduce the time required for training and validating the DNN.

## 2.2. Datasets Generation

**2.2.1. Design of Training and Validation Datasets.** To evaluate the effects of Doppler shift, SNR, and interferences on the performance of the DNN, the samples in the validation dataset need to contain all these parameters. It was not convenient to change these parameters in real underwater target-radiated noise, and it was not difficult to control the parameters in the simulated data described in equation (2) and generate different samples.

In this paper, seven kinds of targets were labeled 0 to 6, where targets 0–3 were real radiated noise, and targets 4–6 were simulated radiated noise. The DEMON spectrum of real target-radiated noise constituted dataset A, which was divided into dataset A1 and dataset A2. The two datasets were used for network training and performance evaluation, respectively. The simulated DEMON spectrum constituted dataset B, which was also divided into two parts: datasets B1 and B2, whose functions were the same as those of datasets A1 and A2. The validation datasets for evaluating the effects of Doppler shift, SNR, and interferences on the network performance were labeled as dataset C, D, and E, respectively. Figure 2 shows the generation and function of all datasets.

### 2.2.2. Generation of Validation Datasets

#### (1) Generation of the dataset with different Doppler shifts (dataset C)

Due to the relative movement of the underwater target and the sonar platform, there will be a Doppler frequency shift in the received radiated noise of the underwater target. The sonar receives a radiated noise written as

$$L_1(t) = L(\eta t), \quad (7)$$

where  $L(t)$  is the target-radiated noise,  $\eta$  is the scaling factor of the Doppler effect, and  $\eta = c/(c + v) \approx 1 - (v/c) = 1 - f_d$ ;  $v$  is the relative radial velocity,  $c$  is sound speed in the water, and  $f_d = v/c$  is the relative Doppler frequency shift.

By changing the value of  $f_d$ , a series of target-radiated noises with different Doppler frequency shifts were generated, and samples of the validation dataset C were generated after DEMON spectrum processing. In this paper, the range of  $f_d$  was set from  $-0.02$  to  $0.02$  in steps of  $0.001$  so that 41 kinds of validation samples with different Doppler frequency shifts were generated. In the training dataset B, the parameter  $f_d$  was set to zero to represent the data without frequency shift.

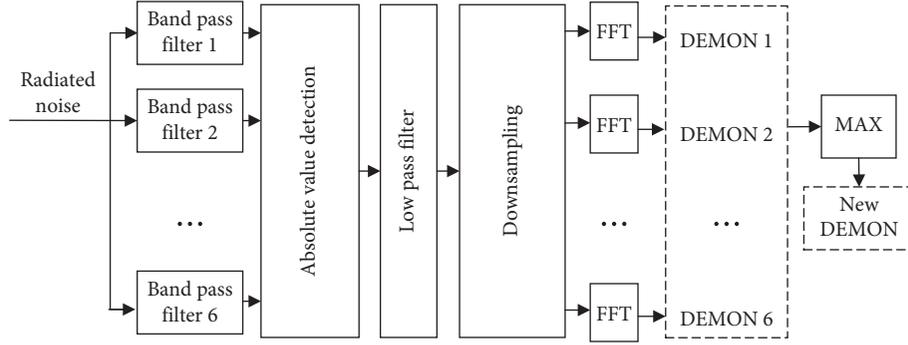


FIGURE 1: The proposed DEMON Processing.

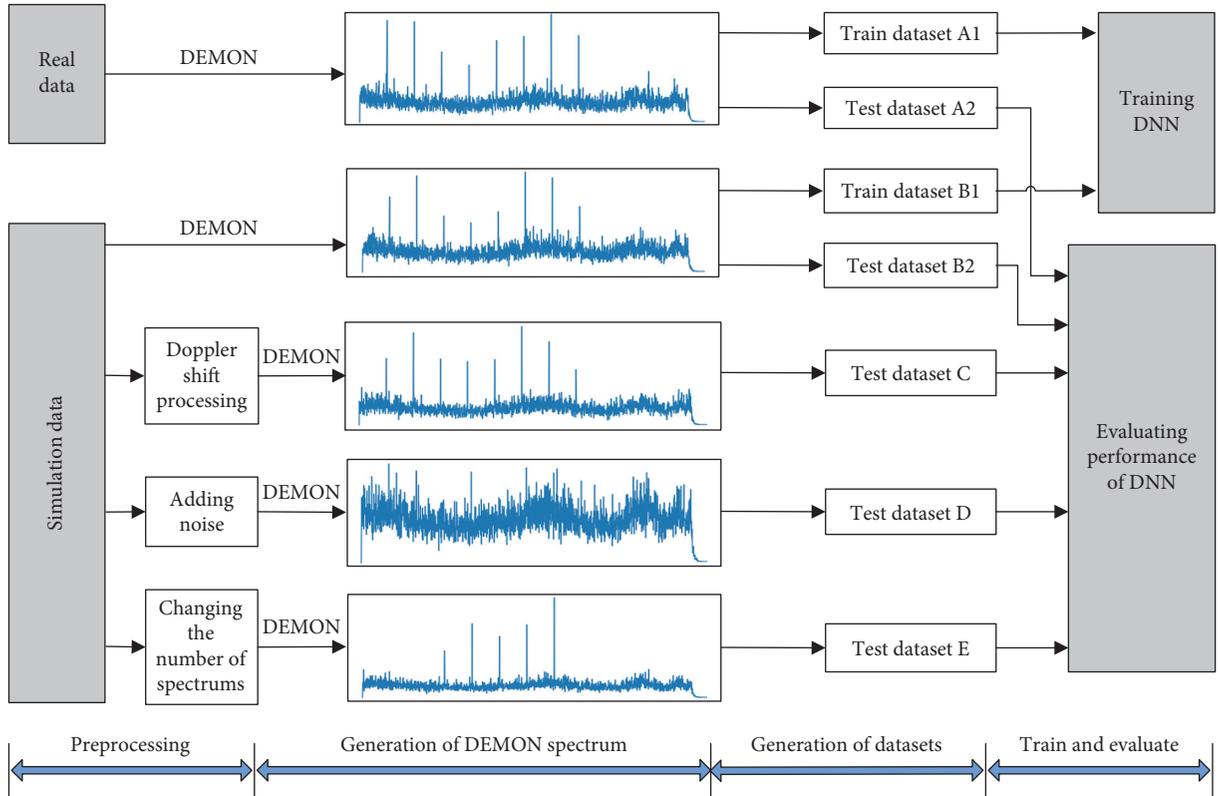


FIGURE 2: Design of training and validation datasets.

(2) Generation of the dataset with different SNRs (dataset *D*)

Due to the influence of the underwater environment noise and the distance between the sonar and the target, the sonar received very different SNRs of the target-radiated noise that can be expressed as

$$L_2(t) = L(t) + \delta N(t), \quad (8)$$

where  $N(t)$  is Gaussian white noise, and  $\delta$  is the coefficient that can control the noise power. By changing the value of  $\delta$ , a series of target-radiated noises with different SNRs were generated, and samples of the validation dataset *D* were generated after DEMON spectrum processing. In this paper,

the range of  $\delta$  was set from 1.0 to 4.0, and the interval was 0.2 so that 16 kinds of validation samples with different SNRs were generated. Taking the power of the maximum line spectrum as the signal power, the SNR of the radiated noise was about 95 (19.8 dB) when  $\delta = 1.0$ . In dataset *B*,  $\delta$  was set to 1.0.

(3) Generation of the dataset with different interferences (dataset *E*)

Due to environmental or human interferences, redundant line spectrums may appear in the DEMON spectrums of the radiated noise received by the sonar, or some line spectrums may weaken or even disappear. During the simulation, according to equation (1), by changing the number  $M$  of sinusoidal signals in  $L(t)$ , a series of target-radiated

noises with various interferences can be generated. Furthermore, samples of the validation dataset  $E$  were generated after DEMON spectrum processing. In this paper, the number of line spectrums was set from  $M - 3$  to  $M + 4$ , so that eight kinds of validation samples with different interferences were generated. When the number of line spectrums is less than  $M$ , the line spectrums are lost. Otherwise, it means that there are interferences in the received radiated noise. In dataset  $B$ , the number of line spectrum was the standard value  $M$ , and there were no interferences and spectrum loss.

### 2.2.3. Dimensions of the Training and Validation Datasets.

We did actual surveys on four different ships to get four kinds of real underwater target-radiated noise in dataset  $A$  and 3000 samples for each kind. The DEMON spectrum of each sample was 2048 in size. Dataset  $A$  was split into training dataset  $A1$  with 9600 samples and validation dataset  $A2$  with the remaining 2400 samples, which means that all the samples were split into a 80% training dataset and 20% validation dataset. Dataset  $B$  contained three kinds of simulation targets whose frequency values of the line spectra were generated randomly within a reasonable range, and its split way was the same as that of dataset  $A$ . Each kind of target in datasets  $A$  and  $B$  had the same number of samples. Datasets  $C$ ,  $D$ , and  $E$  also contained four kinds of real targets and three kinds of simulated targets, respectively, and the sample number of each kind was 500. For example, for a specific target in dataset  $C$ , the number of samples with relative Doppler shift 0.001 was 500. In order to evaluate the performance differences between the 6-channel DEMON spectrums and the single-channel DEMON spectrums, this paper designed a 6-channel dataset  $X1$  and a single-channel dataset  $X2$ , respectively, and both datasets were composed of dataset  $A1$ ,  $A2$ ,  $B1$ ,  $B2$ ,  $C$ ,  $D$ , and  $E$ . Tables 1 and 2 list the detailed information of datasets.

## 3. Design and Training of DNNs

**3.1. Design of DNNs.** There are many types of deep neural networks, and different networks are suitable for different types of tasks. We tried a few standard networks and finally chose CNN, a type of feedforward DNNs [30, 31] with a deep structure that includes convolutional calculations. It is one of the representative algorithms of deep learning, and CNNs have the capability of representational learning which can shift invariant classification of the input information according to its hierarchical structure. Therefore, it is also called "Shift-Invariant Artificial Neural Networks" (SIANN).

This paper proposed a one-dimensional CNN (1D-CNN) to recognize the DEMON spectrums of underwater target-radiated noise. It was expected that the network could have good generalization ability for the DEMON spectrums with different Doppler frequency shifts, SNRs, and interferences. Because there was no empirical information about the network structure, different parameters such as the layer

number and kernel size were combined, and the CNNs with different combinations were trained, respectively. In this process, the number of layers was set from 2 to 5, the kernel size was set from  $1 \times 2$  to  $1 \times 10$ , and the kernel number was set from 2 to 10. When the number of CNN layers exceeded three, its performance increased very slightly. Figure 3 shows the optimal model of CNNs we have tried for the DEMON spectrums recognition. It consists of two convolutional layers, two pooling layers, a fully connected layer, and a hidden layer. The input of each neuron was connected to the output of the previous layer, which was used to extract local features [32, 33]. Four convolution kernels, each with the size of 5, were distributed in the convolution layer  $C1$ , whose input was a DEMON spectrum with 2048 elements. The output of layer  $C1$  was four-channel data, and each channel also had 2048 elements. The pooling layer  $S1$  was used for pooling data from the  $C1$  layer, and the data were compressed into 512 so that this layer could use a smaller dimension to represent the feature data of the previous layer without losing valid information [34]. The  $C2$  layer contained five convolution kernels, each with the size of 3; furthermore, its output was five-channel data, and each had 512 elements. When the data were pooled again by the  $S2$  layer, the data length was reduced to 128. Then, the data were flattened and sent to the fully connected layer. Finally, after processed by the proposed deeply interconnected structure, the input samples were classified or identified.

The paper proposed two CNNs that were trained by a single-channel dataset and a 6-channel dataset, respectively. The only difference between the two CNNs is the size of the input layers. The input layer size of the CNN trained by a single-channel dataset is  $1 \times 2048$  and the 6-channel dataset is  $6 \times 2048$ . Figure 3 shows the structure of 1D-CNN trained by a single-channel dataset, and Table 3 lists the detailed information of the network parameters.

When evaluating network performance, in order to understand it more comprehensively, an MLP network [35] was designed to be compared with CNNs. Similarly, when selecting the MLP network parameters, the number of layers was set from 2 to 5 and the number of neurons was set from 32 to 2048 and was traversed various collocations for model selection. When the complexity of network reached a certain level, the increase in the number of layers was not helpful for performance improvement. Considering that a simple structure reduces the number of calculations for network propagation, the structural parameters of the MLP listed in Table 4 were finally selected.

**3.2. Training of DNNs.** Before training the networks, all network parameters were initialized by the Glorot initialization method that extracts samples from the truncated normal distribution with mean value  $\mu = 0$  and standard deviation  $\sigma = \sqrt{2/(\text{fan\_in} + \text{fan\_out})}$ , where  $\text{fan\_in}$  and  $\text{fan\_out}$  are the numbers of input and output units of the weight tensors, respectively. For all convolutional layers, the rectified linear unit (RELU) was used as the activation function. During the training process, the dropout layer was added after the fully connected layer, and the dropout rate

TABLE 1: The dimensions of the 6-channel dataset.

	Dataset	No. of events	No. of samples	No. of pixels	No. of channels	Data dimension
Training	X1-A1	1	9600	2048	6	$9600 \times 6 \times 2048$
	X1-B1	1	7200	2048	6	$7200 \times 6 \times 2048$
Validation	X1-A2	1	2400	2048	6	$2400 \times 6 \times 2048$
	X1-B2	1	1800	2048	6	$1800 \times 6 \times 2048$
	X1-C	41	3500	2048	6	$41 \times 3500 \times 6 \times 2048$
	X1-D	16	3500	2048	6	$16 \times 3500 \times 6 \times 2048$
	X1-E	8	3500	2048	6	$8 \times 3500 \times 6 \times 2048$

TABLE 2: The dimensions of the single-channel dataset.

	Dataset	No. of events	No. of samples	No. of pixels	No. of channels	Data dimension
Training	X2-A1	1	9600	2048	1	$9600 \times 2048$
	X2-B1	1	7200	2048	1	$7200 \times 2048$
Validation	X2-A2	1	2400	2048	1	$2400 \times 2048$
	X2-B2	1	1800	2048	1	$1800 \times 2048$
	X2-C	41	3500	2048	1	$41 \times 3500 \times 2048$
	X2-D	16	3500	2048	1	$16 \times 3500 \times 2048$
	X2-E	8	3500	2048	1	$8 \times 3500 \times 2048$

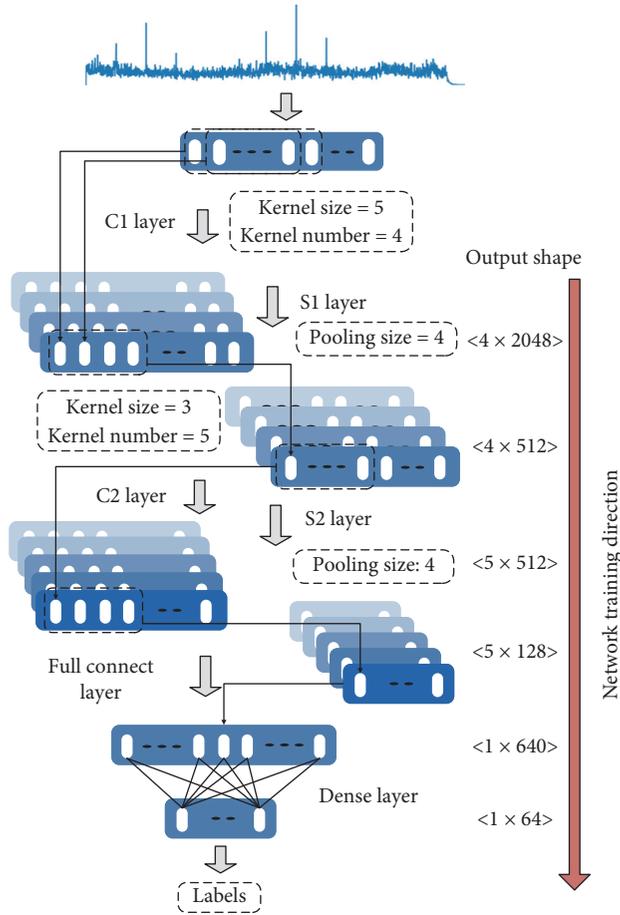


FIGURE 3: The architecture of the proposed 1D-CNN.

was set to 0.3 to avoid overfitting. One hundred and twenty-eight samples were extracted randomly from the dataset as a batch and input into the networks for training. The softmax layer converted the vector of 7 elements output by the last

TABLE 3: The structure and parameters of the proposed CNN.

Single-channel dataset CNN			
Layer type	Kernel number	Kernel size	Output shape
Input	—	—	$1 \times 2048$
Conv_1D	4	$1 \times 5$	$4 \times 2048$
MaxPooling1D	1	$1 \times 4$	$4 \times 512$
Conv_1D	5	$1 \times 3$	$5 \times 512$
MaxPooling1D	1	$1 \times 4$	$5 \times 128$
Flatten	—	—	$1 \times 640$
Dense	—	64	$1 \times 64$
Output	—	7	$1 \times 7$

TABLE 4: The structure and parameters of the proposed MLP.

Single-channel mode for MLP		
Layer type	Neuron size	Output shape
Input_layer	—	$1 \times 2048$
Dense_1	1024	$1 \times 1024$
Dense_2	256	$1 \times 256$
Dropout_1	—	$1 \times 256$
Dense_3	128	$1 \times 128$
Dropout_2	—	$1 \times 128$
Output_layer	7	$1 \times 7$

layer of the network into the probability distribution, which can determine which kind of target the input sample was. Back-propagation was performed by calculating the error between the predicted value and the real label, thereby updating the weights of the network. Finally, the training epoch was set to 15, and the parameters of the entire network were optimized by the Adam Optimization Algorithm [36]. The invariant structure information in a large number of training samples was extracted layer by layer in the network, and a method for underwater target recognition was realized.

## 4. Experiments and Discussion

*4.1. Experimental Setup.* The proposed CNN model was trained on the training datasets *A1* and *B1*, and the performance of the model was evaluated on datasets *C*, *D*, and *E*. In order to understand the CNN performance more thoroughly, a carefully designed MLP network was compared with the CNN. This study focused on observing the changing trend of accuracy with different parameters, so it mainly evaluated the performance in the following three aspects:

- (1) When the received radiated noise contains a Doppler shift, is the proposed CNN trained by samples with no Doppler shift able to recognize the DEMON spectrums? Dataset *C* was used to evaluate the CNN and MLP networks for the recognition of targets with the Doppler shift.
- (2) When the SNR of the received radiated noise declines, can the proposed CNN trained by samples with high SNRs recognize the DEMON spectrums? Dataset *D* was used to evaluate the CNN and MLP networks for recognizing low-SNR targets.
- (3) When the received radiated noise contains interferences, can the proposed CNN trained by samples with no interferences recognize the DEMON spectrums? Dataset *E* was used to evaluate the CNN and MLP networks for recognizing targets with interferences.

Besides, the performances of two CNNs trained by 6-channel samples and signal-channel samples, respectively, were evaluated. Dataset *X1* and *X2* were used to evaluate the two CNNs about Doppler shift, SNR, and interferences.

*4.2. Performance Evaluation of Doppler Shifts.* Dataset *C* was used to evaluate the CNN trained by dataset *A1* and *B1* to understand the performance of the proposed CNN in recognizing targets with Doppler shift. Figure 4 compares the evaluation results of CNN (blue line) with the performance of a well-designed MLP network (red line). The horizontal axis is the relative Doppler frequency shift, and the vertical axis is the accuracy of the networks for target recognition.

The figure highlights the high recognition accuracy of the MLP network when the relative Doppler shift is small ( $|f_d| \leq 0.003$ ). However, as the Doppler shift increases, the accuracy declines rapidly. The results show that the MLP network is sensitive to the Doppler effect, and it is difficult for it to recognize the DEMON spectrums of underwater moving targets if trained by stationary samples. For the proposed CNN, the recognition accuracy remains almost constant when the relative Doppler shift is in a broad range ( $|f_d| \leq 0.015$ ). When the relative Doppler shift increases to 0.02, the recognition accuracy slightly decreases and still maintains a value of more than 93%. The results show that the CNN has higher Doppler tolerance in DEMON spectrums recognition, and the proposed CNN trained by samples of stationary targets can still recognize fast-moving targets.

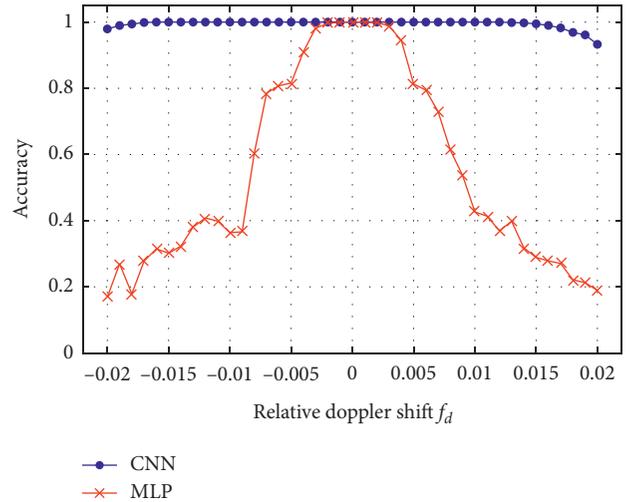


FIGURE 4: The performance of the proposed CNN and MLP networks for Doppler shifts.

To further evaluate the recognition performance of CNN for dataset *C*, we analyzed the results of samples with the maximum relative Doppler frequency shift ( $f_d = -0.02$ ), whose accuracy in Figure 4 was the worst. The classification results are shown in Figure 5, where the vertical axis is the actual label and the horizontal axis is the predicted label. There are 7 kinds of targets, and the sample number of each kind is 500. The color depth reflects the accuracy of recognition.

The precision and recall can be expressed as

$$\text{precision}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}, \quad (9)$$

$$\text{recall}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k},$$

where  $k$  is the sample label of the  $k$ -th kind, True Positive (TP) is correctly identified, False Positive (FP) is incorrectly identified, and False Negative (FN) is incorrectly rejected.

Therefore, each kind of  $F1$ -score can be written as

$$F1_k = \frac{2 \cdot \text{precision}_k \cdot \text{recall}_k}{\text{precision}_k + \text{recall}_k}. \quad (10)$$

By calculating the average value of the  $F1$ -score of each kind,  $F1_{\text{average}}$  can be written as

$$F1_{\text{average}} = \frac{1}{N} \sum_{k=0}^{N-1} F1_k, \quad (11)$$

where  $N$  is 7.

For samples with the maximum relative Doppler frequency shift ( $f_d = -0.02$ ), the calculation result of  $F1_{\text{average}}$  was 0.8256.

*4.3. Performance Evaluation of SNRs.* Due to underwater environment noise and the distance between the target and sonar system, the SNR of the received radiated noise is not constant. In order to understand the performance of the

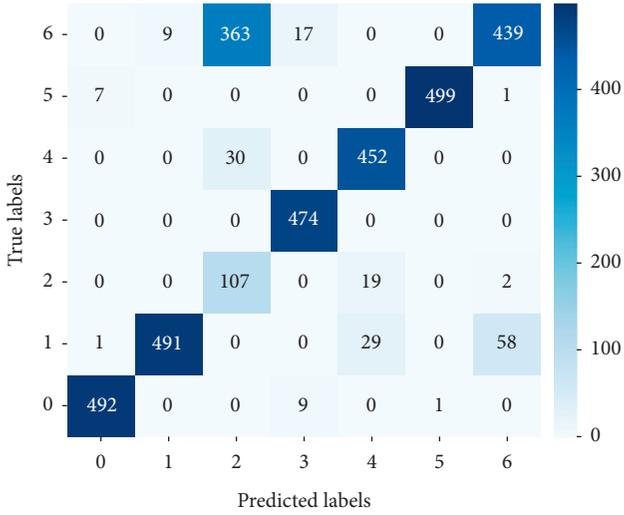


FIGURE 5: The classification results of samples with  $f_d = -0.02$ .

proposed CNN for recognizing targets with various SNRs, dataset  $D$  was used to evaluate the CNN and MLP network as shown in Figure 6. The horizontal axis is the value of  $\delta$ , which can control the noise power, and the vertical axis is the accuracy of the networks.

According to the figure, both CNN and MLP networks can recognize targets well at high SNR values ( $\delta \leq 1.8$ ). However, as the value of  $\delta$  increases, the accuracies of CNN and MLP networks decline rapidly. The proposed CNN performs better than the MLP network, and the accuracy of the CNN is still higher than 80% when  $\delta \leq 2.8$ . The SNR is a crucial factor that can affect underwater target recognition for the DNN. If the SNR is in a reasonable range, the proposed CNN can recognize underwater targets with various SNRs.

We used the same method to obtain the classification results for the samples with the accuracy of 80% ( $\delta = 2.8$ ), as shown in Figure 7. Also, the calculation result of the  $F1_{\text{average}}$  was 0.6511.

**4.4. Performance Evaluation of Interferences.** Due to environmental or human interferences, redundant spectrums may appear in the DEMON spectrums of the received radiated noise, or some spectrums may weaken or even disappear. In order to understand the performance of the proposed CNN in recognizing targets with interferences, dataset  $E$  was used to evaluate the CNN and MLP network. Figure 8 shows the results of the comparison; the horizontal axis is the number of interferences, and the vertical axis is the accuracy of the CNN and MLP networks. When the interferences are less than zero, spectrums with the same number are lost.

According to the figure, both the MLP network and the proposed CNN can recognize underwater targets with interferences well. A small disturbance of the number of feature spectrums little influences the performance of the DNN.

The recognition results were good enough, and further analysis was unnecessary.

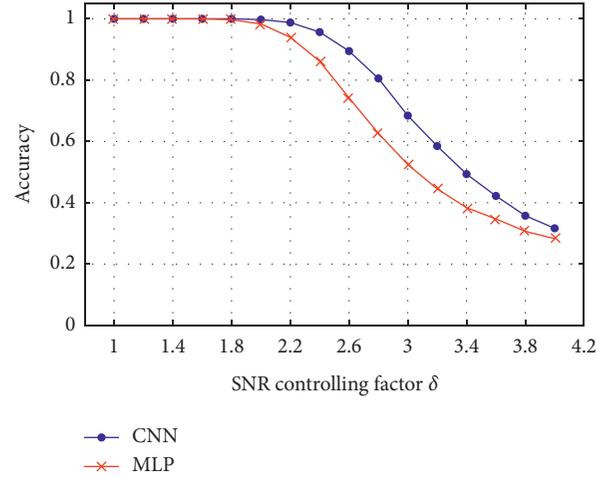


FIGURE 6: The performance of the proposed CNN and MLP networks for SNRs.

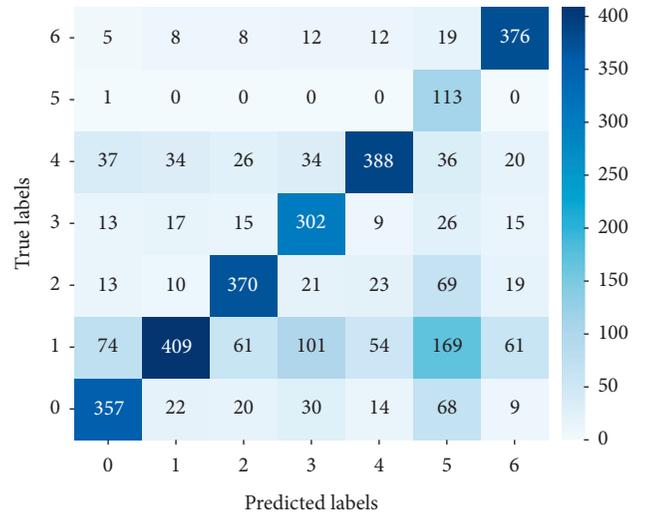


FIGURE 7: The classification results of samples with  $\delta = 2.8$ .

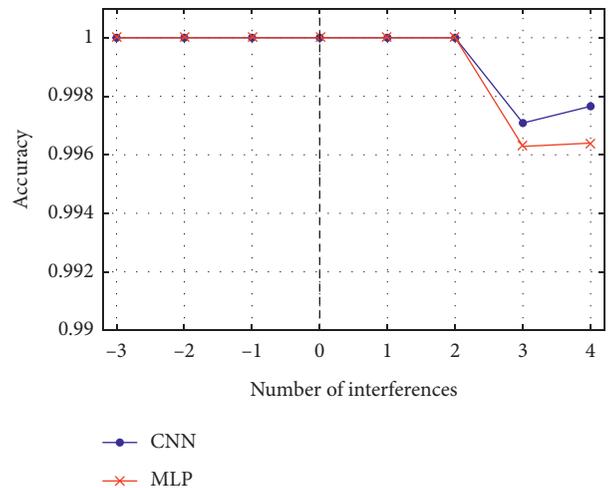


FIGURE 8: The performance of the proposed CNN and MLP networks for interferences.

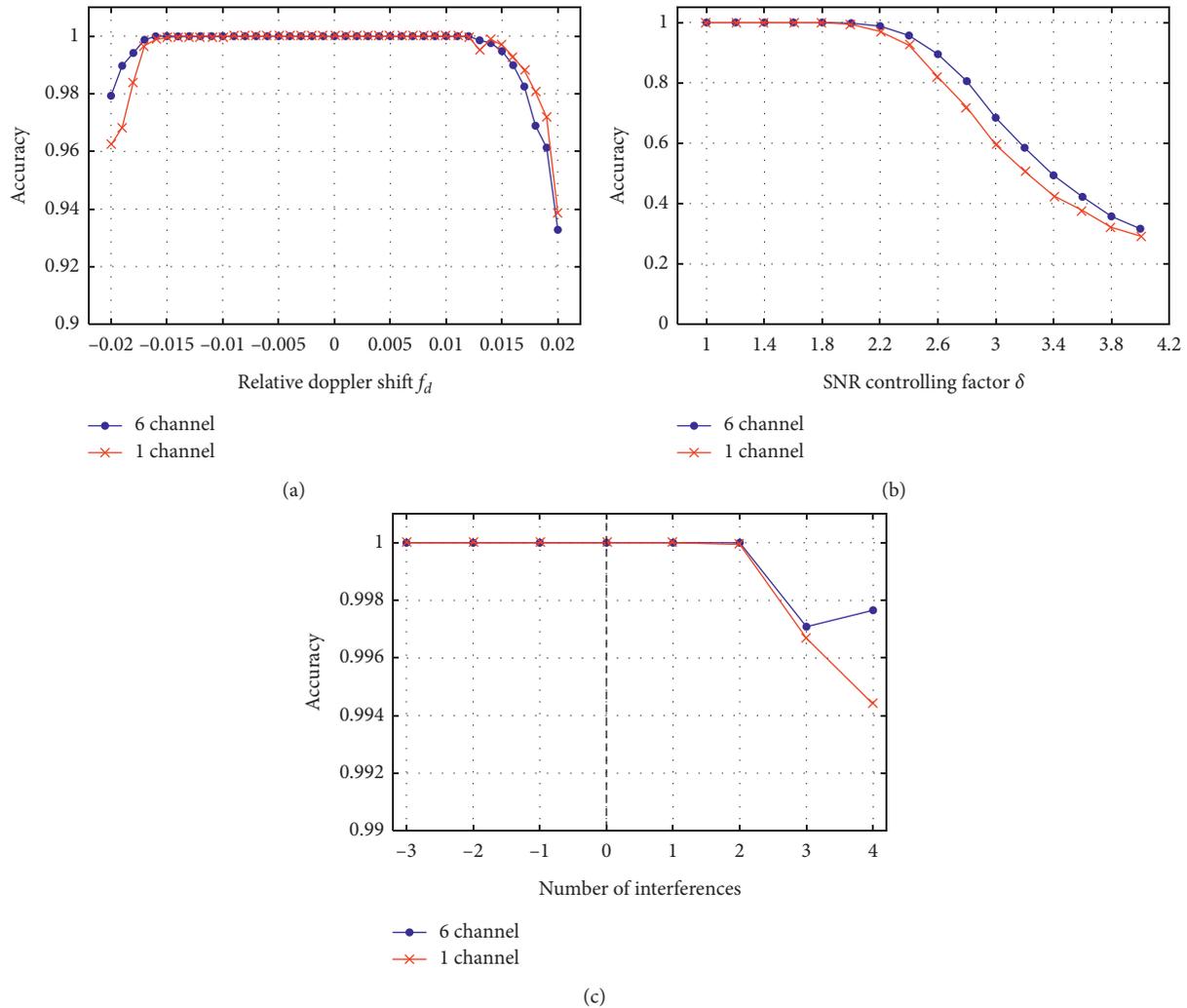


FIGURE 9: The performance comparisons of the 6-channel CNN and single-channel CNN. (a) Doppler shifts; (b) SNRs; and (c) interferences.

TABLE 5: The time consumption of the single-channel CNN and 6-channel CNN.

	Type	Dataset	Time (s)
Training (20 epochs)	6-channel	X1-A&X1-B	210.19
	Single-channel	X2-A&X2-B	44.06
Validation	6-channel	X1-C	158.58
		X1-D	52.65
	Single-channel	X1-E	21.06
		X2-C	31.03
		X2-D	10.25
		X2-E	4.10

**4.5. Performance Evaluation of Channel Numbers.** The received radiated noise was divided into six bands before absolute value demodulation order to obtain the DEMON spectrums with six channels. Then, the 6-channel samples were compressed into single-channel DEMON spectrums by selecting the maximum values to reduce the calculation burden. Dataset X1 of 6-channel samples and dataset X2 of

single-channel samples were used to train and evaluate two CNNs; Figure 9 compares the performances.

In the figure, the CNN trained by 6-channel samples has slightly better accuracy than that trained by single-channel samples in the aspects of Doppler shift, SNR, and interferences. However, at the cost of a little accuracy loss, the total size of samples shrinks, which can reduce the training and validating time and storage of the CNN significantly. Table 5 shows the training and validation time required for single-channel and multichannel datasets based on dual “NVIDIA GeForce GTX1080Ti” graphics cards. During the training process, the time required for the 6-channel CNN was 210.19 seconds and that of the single-channel CNN was 44.06 seconds. During the validating process, taking X1-C and X2-C for evaluating Doppler performance as an example, the time required for the 6-channel CNN was 158.58 seconds and that of the single-channel CNN was 31.03 seconds. The results show that the single-channel CNN can greatly reduce the calculation time for training and validation. Therefore, a trade-off should be made between the recognition accuracy and the available resources when

recognizing DEMON spectrums of underwater radiated noise.

## 5. Conclusions

A one-dimensional CNN for recognizing the DEMON spectrums of underwater target-radiated noise was proposed, and the performances of recognition of targets with different Doppler shifts, SNRs, and interferences were evaluated. Compared with MLP networks, the proposed 1D-CNN has higher Doppler tolerance and can recognize underwater targets with lower SNRs. At the cost of a little accuracy loss, compared with the CNN trained by 6-channel samples, the CNN trained by single-channel samples can reduce the training and validating time and needs less storage. The simulation results show that the proposed 1D-CNN trained by single-channel samples has excellent performances in the DEMON spectrums recognition of underwater targets and is suitable for practical engineering applications.

## Data Availability

The datasets used to support the findings of this study have not been made available because the datasets are too large. The datasets can be obtained from the author through e-mail.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 61601322), State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences (Grant no. SKLA202002), and the Independent Innovation Project of the Third Research Academy, CASIC (Grant no. YuanJi [2019]740).

## References

- [1] H. Wen, Y. Yang, H. Zhou et al., "Classification and analysis on the ocean underwater acoustic detection signals," *Journal of Applied Oceanography*, vol. 34, no. 2, pp. 272–278, 2015.
- [2] R. Henker, T. Schneider, A. Wiatrek et al., "Optimisation of optical signal delay in slow-light systems based on stimulated Brillouin scattering," pp. 180–185, IET Conference Publications, Galway, Ireland, 2008.
- [3] H. Zhang, L. Wan, and X. Tang, "Underwater target recognition using artificial fish-swarm algorithm," in *Proceedings of the 2009 Chinese Conference on Pattern Recognition. (CCPR 2009)*, Nanjing, China, November 2009.
- [4] Q. Li, "A new method of passive ranging for underwater target: distance information extraction based on wave guide invariant," *Acta Acustica*, vol. 40, no. 2, pp. 138–143, 2015.
- [5] D. Skarlatos, P. Agrafiotis, T. Balogh et al., "Project imar-culture: advanced VR, iMmersive serious games and augmented reality as tools to raise awareness and access to European underwater cultural heritage," *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, vol. 10058, pp. 805–813, 2016.
- [6] S. Henclik, "Underwater acoustic target tracking with fixed passive sonar system," *Acta Acustica United with Acustica*, vol. 88, no. 5, pp. 674–677, 2002.
- [7] Y. Zhang, J. Sun, and Y. Zhang, "Research on acoustic signal detection simulation for passive sonar," in *Proceedings of the 2010 International Conference on Computational and Information Sciences (ICIS 2010)*, pp. 948–951, Chengdu, China, 2010.
- [8] T. Kim and K. Bae, "HMM-based underwater target classification with synthesized active sonar signals," in *Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011)*, vol. E94-A, pp. 1805–1808, Barcelona, Spain, October 2011.
- [9] R. Harrison, C. Yang, C. Lin et al., "Classification of underwater targets with active sonar," in *Proceedings of the 1st IEEE Regional Conference on Aerospace Control Systems (AEROCOS 1993)*, pp. 534–538, Westlake Village, CA, USA, 1993.
- [10] S. Wang, Y. Wang, Y. Wang et al., "Feature extraction and classification of spectrum of radiated noise of underwater high speed vehicle," *Journal of Electronics and Information Technology*, vol. 39, no. 11, pp. 2684–2689, 2017.
- [11] Y. Wu, Y. Yang, L. Yang et al., "Underwater target recognition based on constant-beamwidth waveform fidelity and interference-suppression," *Journal of Northwestern Polytechnical University*, vol. 33, no. 5, pp. 843–848, 2015.
- [12] A. Yao and J. Hui, "Single vector sensor DEMON spectrum detection technique," *Technical Acoustics*, vol. 25, no. 1, pp. 66–70, 2006.
- [13] X. Yang and B. Wang, "Review of techniques for underwater moving target spectrum feature extraction and enhancement," *Underwater & Physical Acoustics*, vol. 26, pp. 69–72, 2007.
- [14] Z. Song, Y. Ding, X. Zhao et al., "Target recognition method of LOFAR spectrum based on the underwater," *Journal of Naval Aeronautical and Astronautical University*, vol. 26, no. 3, pp. 283–286, 2011.
- [15] J. Liu, H. Yang, L. Zhong et al., "Underwater target recognition based on line spectrum and support vector machine," in *Proceedings of the 2014 International Conference on Mechatronics, Control and Electronic Engineering (ICMRA 2014)*, vol. 536, Zhuhai, China, March 2014.
- [16] N. Wang, M. He, H. Wang et al., "A fast reduced-dimension convolution model for underwater target recognition," *Journal of Harbin Engineering University*, vol. 40, no. 7, pp. 1327–1333, 2019.
- [17] H. Yang, J. Li, S. Shen, and G. Xu, "A deep convolutional neural network inspired by auditory perception for underwater acoustic target recognition," *Sensors*, vol. 19, no. 5, p. 1104, 2019.
- [18] J. Choi, Y. Choo, and K. Lee, "Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning," *Sensors*, vol. 19, no. 16, p. 3492, 2019.
- [19] X. Cao, R. Togneri, X. Zhang, and Y. Yu, "Convolutional neural network with second-order pooling for underwater target classification," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 3058–3066, 2019.
- [20] X. Zhou, K. Yang, and R. Duan, "Deep learning based on striation images for underwater and surface target classification," *IEEE Signal Processing Letters*, vol. 26, no. 9, pp. 1378–1382, 2019.
- [21] Y. Chen, S. Du, H. Quan et al., "Underwater target recognition method based on convolution residual network," in *Proceedings of the 2nd Franco-Chinese Acoustic Conference (FCAC 2018)*, vol. 283, Le Mans, France, June 2018.

- [22] N. Wang, M. He, J. Sun et al., "IA-PNCC: noise processing method for underwater target recognition convolutional neural network," *Computers, Materials & Continua*, vol. 58, no. 1, pp. 169–181, 2019.
- [23] O. Elhaki and K. Shojaei, "Neural network-based target tracking control of underactuated autonomous underwater vehicles with a prescribed performance," *Ocean Engineering*, vol. 167, pp. 239–256, 2018.
- [24] K. Shojaei, "Three-dimensional neural network tracking control of a moving target by underactuated autonomous underwater vehicles," *Neural Computing and Applications*, vol. 31, no. 2, pp. 509–521, 2019.
- [25] X. Wang, A. Liu, Y. Zhang, and F. Xue, "Underwater acoustic target recognition: a combination of multi-dimensional fusion features and modified deep neural network," *Remote Sensing*, vol. 11, no. 16, p. 1888, 2019.
- [26] X. Qiao, J. Bao, H. Zhang, F. Wan, and D. Li, "Underwater sea cucumber identification based on principal component analysis and support vector machine," *Measurement*, vol. 133, pp. 444–455, 2019.
- [27] S. Wang, J. Qiu, and S. Wang, "Enhancement of ship radiated noise DEMON spectrum SNR based on correlation properties theory of principles of system dynamics," *Ship Science and Technology*, vol. 35, no. 8, pp. 24–27, 2013.
- [28] J. Qiu and J. Zhang, "Research on ship-radiated noise beat tune," in *Proceedings of the 7th International Symposium on Computational Intelligence and Design (ISCID 2014)*, vol. 1, pp. 573–576, Hangzhou, China, March 2014.
- [29] Y. Cheng, X. Gao, and H. Liu, "A method for ship propeller blade-number recognition based on template matching," *Technical Acoustics*, vol. 29, no. 2, pp. 228–231, 2010.
- [30] L. Yan, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [31] L. Yan, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.
- [32] Y. Bengio and L. Yann, "Convolutional networks for images, speech, and time-series," *Handbook of Brain Theory and Neural Networks*, pp. 255–258, MIT Press, Cambridge, MA, USA, 1995.
- [33] N. Cohen, O. Sharir, and A. Shashua, "On the expressive power of deep learning: a tensor analysis," in *Proceedings of the 29th Conference on Learning Theory (COLT 2016)*, vol. 49, no. 1, pp. 698–728, New York, NY, USA, June 2016.
- [34] N. Srivastava, G. Hinton, A. Krizhevsky et al., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [35] T. Ebert, O. Bänfer, and O. Nelles, "Multilayer perceptron network with modified sigmoid activation functions," *Artificial Intelligence and Computational Intelligence*, vol. 6319, pp. 414–421, 2010.
- [36] D. P. Kingma and J. Ba, "ADAM: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, May 2015.