

Article

Underwater Target Recognition Based on Multi-Decision LOFAR Spectrum Enhancement: A Deep-Learning Approach

Jie Chen , Bing Han , Xufeng Ma and Jian Zhang

National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 610054, China; binghan@std.uestc.edu.cn (B.H.); xufeng_ma@126.com (X.M.); jianzhan_g@foxmail.com (J.Z.)

* Correspondence: jiechen@uestc.edu.cn; Tel.: +86-1592-810-9908



Citation: Jie, C.; Bing, H.; Xufeng, M.; Jian, Z. Underwater Target Recognition Based on Multi-Decision LOFAR Spectrum Enhancement: A Deep-Learning Approach. *Future Internet* **2021**, *13*, 265.
<https://doi.org/10.3390/fi13100265>

Academic Editors: Chang Liu, Shihao Yan, Qingqing Cheng, Minghui Min and Weijie Yuan

Received: 10 September 2021

Accepted: 5 October 2021

Published: 13 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Underwater target recognition is an important supporting technology for the development of marine resources, which is mainly limited by the purity of feature extraction and the universality of recognition schemes. The low-frequency analysis and recording (LOFAR) spectrum is one of the key features of the underwater target, which can be used for feature extraction. However, the complex underwater environment noise and the extremely low signal-to-noise ratio of the target signal lead to breakpoints in the LOFAR spectrum, which seriously hinders the underwater target recognition. To overcome this issue and to further improve the recognition performance, we adopted a deep-learning approach for underwater target recognition, and a novel LOFAR spectrum enhancement (LSE)-based underwater target-recognition scheme was proposed, which consists of preprocessing, offline training, and online testing. In preprocessing, we specifically design a LOFAR spectrum enhancement based on multi-step decision algorithm to recover the breakpoints in LOFAR spectrum. In offline training, the enhanced LOFAR spectrum is adopted as the input of convolutional neural network (CNN) and a LOFAR-based CNN (LOFAR-CNN) for online recognition is developed. Taking advantage of the powerful capability of CNN in feature extraction, the recognition accuracy can be further improved by the proposed LOFAR-CNN. Finally, extensive simulation results demonstrate that the LOFAR-CNN network can achieve a recognition accuracy of 95.22%, which outperforms the state-of-the-art methods.

Keywords: underwater acoustic communication; underwater target recognition; LOFAR spectrum; line spectrum enhancement; deep learning

1. Introduction

The ocean contains rich mineral resources, marine living resources and chemical resources. Its huge economic value has attracted the attention of almost all coastal countries in the world. Therefore, ocean development such as seabed exploration, oil platform monitoring, and economic fish detection is of great significance. The ability to accurately determine whether an underwater target is an ordinary aquatic fish or a dangerous obstacle with the help of underwater acoustic target-recognition technology is extremely critical to the safety of shipping vessels. Deep Learning (DL) is a discipline that specializes in how computers simulate or implement human learning behaviors to acquire new knowledge or skills, and reorganize the existing knowledge structure to continuously improve its own performance [1–3]. It is found that it also has a good performance in the field of communications. Underwater target recognition based on DL is a new method to realize underwater target recognition based on existing recognition methods. Using this method, it can automatically extract features from the original signal, compress feature vectors, fit the target map, reduce the impact of noise, avoid feature loss during manual extraction, improve generalization capabilities, and constantly improve the efficiency and accuracy of identification during the model process.

The application of deep learning in the field of underwater target recognition mainly involves three aspects. The first is the field of underwater recognition. Due to many reasons such as confidentiality and security, the collection and production of data sets are difficult. Therefore, researchers will use as many existing samples as possible, such as using Generative Adversarial Networks (GAN) to achieve sample expansion to meet the needs of deep learning with large data volumes. The second is the orthodox field of deep learning, such as computer vision and natural language recognition. Researchers start from optimizing and designing complex deep neural network structures, and only rely on neural networks to complete feature extraction. The third is the data preprocessing stage before inputting the neural network. In view of the serious pollution of the collected data by environmental noise, researchers perform denoising and spectral transformation on audio samples, or perform image denoising on sonar images. Its purpose is to make the sample features as obvious as possible through feature engineering, which is more conducive to the needs of deep neural network feature extraction.

In this paper, we are interested in underwater target-recognition methods based on deep learning. The Low-frequency analysis and recording (LOFAR) spectrum is widely used in the field of passive sonar ship target recognition due to its significant sound source information and relatively high signal-to-noise ratio. It transforms the signal from the time domain to the time-frequency domain. Sonars usually observe the line spectrum in the LOFAR spectrum to determine whether the target exists, and perform tracking and recognition. This type of method is mainly realized by extracting features and training classifiers. Unfortunately, because the data are always contaminated by environmental noise, breakpoints are introduced in the LOFAR spectrum, which reduces the performance of signal processing. To overcome this problem and further improve the performance of underwater target recognition, we use deep-learning methods for underwater target recognition, and propose an underwater target-recognition scheme based on LOFAR spectral enhancement (LSE). This solution can restore the breakpoints in the LOFAR spectrum and combine with Convolutional Neural Network (CNN) for online recognition, which reduces the impact of environmental noise and significantly improves the target recognition rate of existing algorithms.

1.1. Contributions

The main contributions of this paper are summarized as follows:

- (1) In contrast to the traditional algorithm, we use the decomposition algorithm based on resonance signal to preprocess the signal. Based on the multi-step decision algorithm with the line spectrum characteristic cost function [4], this paper proposes the specific calculation method of double threshold. In the purpose, this algorithm not only retains the continuous spectrum information in the original LOFAR spectrum, but also merges the extracted line spectrum with the original LOFAR spectrum. Finally, the breakpoint completion of the LOFAR spectrum is realized.
- (2) To further improve the recognition rate of underwater targets, we adopt the enhanced LOFAR spectrum as the input of CNN and develop a LOFAR-based CNN (LOFAR-CNN) for online recognition. Taking advantage of the powerful capability of CNN in feature extraction, the proposed LOFAR-CNN can further improve the recognition accuracy.
- (3) Simulation results demonstrate that when testing on the ShipsEar database [5], our proposed LOFAR-CNN method can achieve a recognition accuracy of 95.22% which outperforms the state-of-the-art methods.

1.2. Related Works

Recently, CNN has proven its powerful capability in many fields, such as computer vision, nature language processing, and wireless physical layer [6–8]. Convolutional neural networks are deep feedforward neural networks that include operations such as convolution calculations, pooled sampling, and nonlinear activation [4,9,10]. Compared

with the traditional feedforward neural networks such as MLP, three strategies in CNN make use of the spatial correlation of data which include weight sharing, local receptive field and down sampling. They reduce the risk of over fitting, the defect of gradient disappearance the complexity and parameter size of the network. However, they improve the generalization ability of the network. CNN was first proposed by LeCun [11] in 1990 and applied to the handwritten character detection system. In 2014, Szegedy [12] proposed GoogleLeNet which introduced the inception module. Receptive fields of different sizes enhanced the adaptability of the network to scale. The improved version [13,14] greatly reduces the parameter amount to enhance the nonlinearity of the network and speed up the calculation. The residual network was proposed by Kaiming. He [15] in 2015 adopted the idea of Shortcut Connection (SC) to solve the problem of network degradation. After full investigation and experimental verification, CNN is very suitable for underwater target recognition.

In addition, many effective and efficient DL-based schemes have been proposed for underwater target recognition. For example, Ref [16,17] focused on underwater target recognition which have sufficient training samples. In the first step, the original audio was converted into LOFAR spectrum, and then GAN was used for sample expansion. In the second step, a 15% performance improvement could be obtained using convolutional neural networks (CNNs) for feature learning and classification when the number of samples was more sufficient. Ref [18] combined competitive learning with deep belief network (DBN) and proposed a deep competitive network that used unlabeled samples to solve small number of samples in acoustic target recognition. This method could achieve a classification accuracy of 90.89%. To address the negative impact of redundant features on recognition accuracy and efficiency, the authors in [19] proposed a compressed deep competition network which combined network pruning with training quantization and other technologies and could achieve a classification accuracy of 89.1%. Ref [20,21] proposed a new time-frequency feature extraction method by jointly exploiting the resonance-based sparse signal decomposition (RSSD) algorithm, the phase space reconstruction (PSR), the time-frequency distribution (TFD), and the manifold learning. At the same time, a one-dimensional convolutional auto-encoder-decoder model was used to further extract and separate features from high-resonance components, which finally completed the recognition task and achieves a recognition accuracy of 93.28%. In addition, Ref [22–24] all used convolutional neural networks for feature extraction, but the application scenarios and the classifiers were different. [22] proposed an automatic target-recognition method of unmanned underwater vehicle (UUV), which adopted CNN to extract features from sonar images and used support vector machine (SVM) classifier to complete the classification. Ref [23] aimed to study different types of marine mammals. It also used the CNN+SVM structure to complete the feature extraction and classification recognition task. It compared the two classification and multi-class task scenarios. [24] adopted the civil ship data set and exploited the framework structure of CNN+ELM (extreme learning machine) as the underwater target classifier, which improved the recognition accuracy. We can see that with the in-depth research of scholars, the recognition rate of underwater targets based on deep learning has gradually increased.

1.3. Organization

The rest of this article is organized as follows. The second section introduces the model of the system. The third section introduces the deep-learning underwater target signal recognition framework based on multi-step decision LOFAR line spectrum enhancement. The fourth section is the experimental verification and simulation results of our proposed algorithm framework. The fifth section is the summary of the article.

Some notations in this paper are shown in the following. $\|\cdot\|_2$ and $\|\cdot\|_1$ respectively represent the L2 norm and L1 norm. $STFT\{\cdot\}$ is short-time Fourier transform. Term $E(\cdot)$ is the statistical expectation. $argmin$ represents the variable value when the objective function is minimized.

2. System Model

In this paper, we consider a deep-learning underwater target-recognition framework based on multi-step decision LOFAR line spectrum enhancement which is shown in Figure 1. It is divided into four modules: sampling, feature preprocessing, offline training and online testing.

2.1. Signal Decomposition Algorithm Based on Resonance

In traditional signal processing, Fourier transform is usually used to analyze in the frequency domain or time-frequency domain, but these methods are only valid for periodic stationary signals [25]. However, due to the generation mechanism of ship radiated noise and the complex channel conditions in the marine environment, the ship radiated noise collected by hydrophones is usually the mixture of oscillating signals and transient non-oscillating signals [20]. The harmonic component (or oscillation component) of the ship's radiated noise plays an important role in the identification of underwater targets. Therefore, a signal decomposition algorithm based on resonance that effectively responds to nonlinear signals is used to preprocess the signal. Based on the oscillation characteristics rather than the frequency or scale, the method can obtain a signal composed of multiple simultaneous and continuous oscillations (high-resonance component). To some extent, it weakens the transient non-oscillation signal of uncertain duration (low-resonance component) and gaussian white noise (residual component) which is conducive to feature extraction.

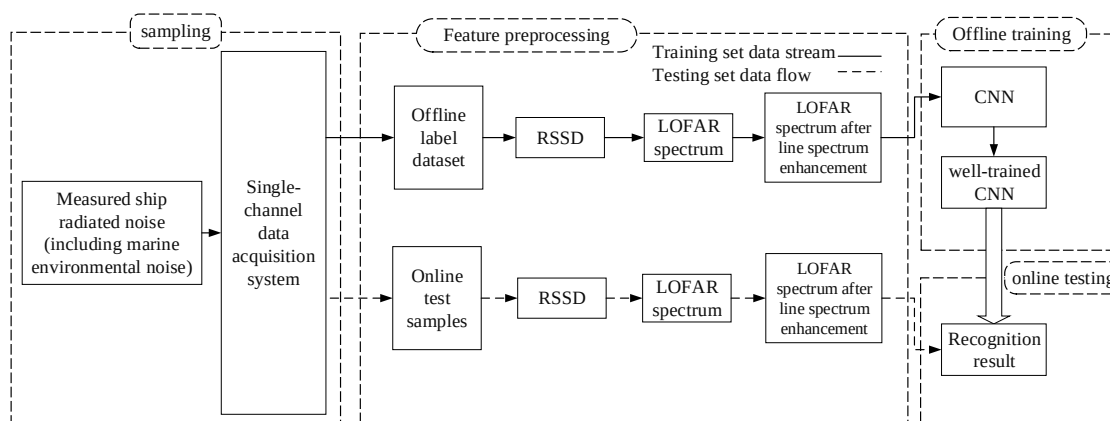


Figure 1. Deep-learning underwater target-recognition framework based on multi-step decision LOFAR line spectrum enhancement.

The RSSD algorithm regards resonance as the basis for signal decomposition [26], and the Q factor quantifies the degree of signal resonance. Specifically, high-resonance signals exhibit a higher degree of frequency aggregation in the time domain, more simultaneous oscillating waveforms with a larger Q factor. Low-resonance signals appear non-oscillating and indefinite transient signal with a smaller Q factor. Therefore, the basic theory of the RSSD algorithm is that using two different wavelet basis functions (corresponding to Q factors of different sizes), we can find a sparse representation of a complex signal and reconstruct the signal.

The algorithm mentioned in this section is divided into adjustable Q -Factor Wavelet Transform (TQWT) [27] and Morphological Component Analysis (MCA) [28]. Its algorithm framework is shown in Figure 2.

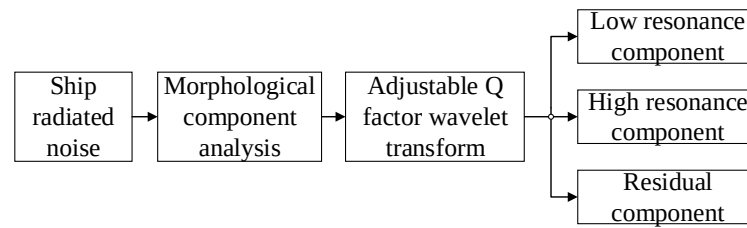


Figure 2. Signal decomposition algorithm based on resonance.

2.1.1. Morphological Component Analysis

Morphological component analysis is usually used to decompose signals with different morphological characteristics [29]. The ship radiated noise with oscillating and non-oscillating component has different morphological characteristics. So, the MCA algorithm can be used to separate and extract the ship radiated noise to construct the optimal sparse representation for its high-resonance and low-resonance component.

Considering the discrete ship radiated noise sequence, the signal can be sparsely expressed as:

$$x = \Phi_h w_h + \Phi_l w_l + n, \quad (1)$$

where w_h, w_l are the wavelet coefficients corresponding to the high resonant component x_h and the low resonant component x_l . Φ_h, Φ_l are wavelet basis functions corresponding to x_h, x_l . n represents the residual components of the signal which removes first two.

The purpose of MCA is to obtain an optimal representation w_h, w_l of the high-resonance component and low-resonance component of the signal. This problem can be solved by minimizing the following objective function:

$$J(w_l, w_h) = \|x - \Phi_h w_h - \Phi_l w_l\|_2^2 + \sum_{j=1}^{J_h+1} \lambda_{h,j} \|w_h^j\|_1 + \sum_{j=1}^{J_l+1} \lambda_{l,j} \|w_l^j\|_1. \quad (2)$$

Here, J_h and J_l represent the number of decomposition layers of x_h and x_l . w_h^j and w_l^j are the wavelet coefficients of the high-resonance component and the low-resonance component of the j th layer, respectively. $\lambda_{h,j}, \lambda_{l,j}$ are the normalized coefficients of $w_{h,j}, w_{l,j}$ and their values are related to energy of $\Phi_{h,j}, \Phi_{l,j}$:

$$\lambda_{l,j} = k_{l,j} \|\Phi_{l,j}\|_2, j = 1, 2, \dots, J_l + 1, \quad (3)$$

$$\lambda_{h,j} = k_{h,j} \|\Phi_{h,j}\|_2, j = 1, 2, \dots, J_h + 1, \quad (4)$$

where $k_{l,j}, k_{h,j}, (k_{l,j} + k_{h,j} = 1)$ are the proportionality coefficient of the energy distribution of the high-resonance component and the low-resonance component. $k_{l,j} = k_{h,j} = 0.5$ are selected to balance the energy distribution of the two components.

Through decomposition of the Augmented Lagrangian Shrinkage Algorithm (SALSA) [26], the optimal wavelet coefficients can be obtained by solving the optimization problem of the formula. Therefore, the optimal expressions for the high-resonance component and the low-resonance component obtained by the MCA algorithm are:

$$x_h^* = \Phi_h w_h^*, \quad (5)$$

$$x_l^* = \Phi_l w_l^*. \quad (6)$$

In summary, the purpose of the RSSD algorithm is to construct the optimal sparse representation of the high and low-resonance components of the ship radiated noise. The specific steps can be expressed as follows:

- (1) Select the appropriate filter scaling factor α, β according to the waveform characteristics of the signal. Then calculate the parameters Q_h, r_h, J_h corresponding to the high-resonance component, and the parameters Q_l, r_l, J_l corresponding to the low-resonance component. At last, construct the corresponding wavelet basis function Φ_h, Φ_l .
- (2) Reasonably set the weighting coefficient $\lambda_{h,j}, \lambda_{l,j}$ of the L1 norm of the wavelet coefficients of each layer. Obtain the optimal wavelet coefficient w_h^*, w_l^* by minimizing the objective function through the SALSA algorithm.
- (3) Reconstruct the optimal sparse representation x_h^*, x_l^* of high-resonance components and low-resonance components.

2.1.2. Adjustable Q-Factor Wavelet Transform

TQWT is a discrete wavelet transform that can flexibly adjust the constant Q factor according to the resonance of the processed signal, which has an overcomplete basis and can be perfectly reconstructed [30]. This section uses the TQWT toolbox to complete simulation experiments and signal processing. The implementation framework consists of two filter banks which are analysis filter bank and integrated filter bank. They are shown in Figures 3 and 4. A filter bank refers to a group of filters. They have a common input, or a common output. The analysis filter bank has a common input to obtain multiple subband signals. In contrast, the integrated filter has a common output, combining multiple subband signals into a single signal to restore the original analyzed signal.

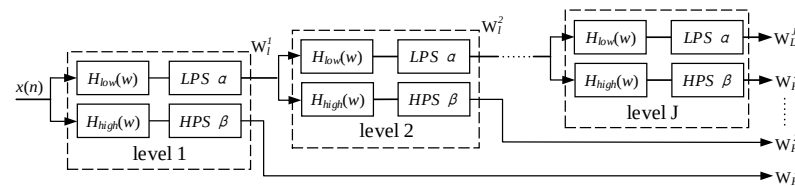


Figure 3. Analysis filter bank.

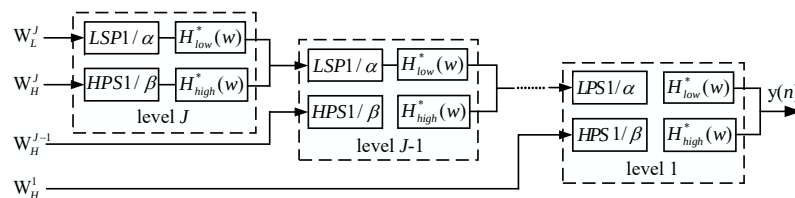


Figure 4. Integrated filter bank.

The analysis filter bank of each layer is composed of high-pass filter $H_{high}(w)$, low-pass filter $H_{low}(w)$, and the corresponding scaling process, which are defined as follows:

$$H_{high}(w) = \begin{cases} 0 & |w| \leq (1-\beta)\pi \\ \theta(\frac{\alpha\pi-w}{\alpha+\beta-1}) & (1-\beta)\pi \leq w \leq \alpha\pi \\ 1 & \alpha\pi \leq |w| \leq \pi \end{cases} \quad (7)$$

$$H_{low}(w) = \begin{cases} 1 & |w| \leq (1-\beta)\pi \\ \theta(\frac{w+(\beta-1)\pi}{\alpha+\beta-1}) & (1-\beta)\pi \leq w \leq \alpha\pi \\ 0 & \alpha\pi \leq |w| \leq \pi \end{cases} \quad (8)$$

$\theta(w) = 0.5(1 + \cos(w))\sqrt{2 - \cos(w)}$ is the Daubechies filter with second-order disappearing moment [27]. α, β ($0 < \alpha < 1, 0 < \beta < 1$) are the scaling factors after the signal passes through the low-pass and high-pass filters, respectively. The scaling process of low-pass and high-pass are defined as:

$$Y(w) = X(\alpha w), |w| \leq \pi, \quad (9)$$

$$Y(w) = \begin{cases} X(\beta w + (1 - \beta)\pi) & 0 \leq w \leq \pi \\ X(\beta w - (1 - \beta)\pi) & -\pi < w < 0 \end{cases} \quad (10)$$

The Q factor quantifies the degree of signal resonance, and its definition is f_c/BW , where f_c represents the center frequency of the signal and BW represents the bandwidth.

If the sampling frequency of the original input signal is f_s , then the center frequency f_c , the filter bank level j and α , β [31] can be expressed as:

$$f_c = \alpha^j \frac{2 - \beta}{4\alpha} f_s. \quad (11)$$

Similarly, bandwidth BW can be expressed as:

$$BW = 0.5\beta\alpha^{j-1}\pi. \quad (12)$$

Therefore, the Q factor is derived as:

$$Q = \frac{2 - \beta}{\beta}. \quad (13)$$

After the original signal passes through the filter bank, the output of the low-pass channel is iteratively inputted to the deeper level filter bank until the preset level J . At the same time, the wavelet basis functions Φ_h , Φ_l are constructed by selecting the oversampling rate r . The deepest level $J_{\max}$ and the oversampling rate r are defined as follows:

$$r = \frac{\beta}{\alpha + 1}, \quad (14)$$

$$J_{\max} = \left\lceil \frac{\log(\beta N/8)}{\log(1/\alpha)} \right\rceil. \quad (15)$$

In summary, in the TQWT algorithm, Q , r , J can be calculated by selecting α , β , and α , β selection is only determined by the inherent oscillation characteristics of the signal. Therefore, it can flexibly select α , β according to the specific requirements of Q , r , J . For the input signal of ship radiated noise, we need to set Q_h , r_h , J_h to extract its high-resonance information and set Q_l , r_l , J_l to extract its low-resonance information.

3. LOFAR Spectral Line Enhancement Based on Multi-Step Decision

The line spectrum has been widely used in the field of passive sonar ship target recognition because of its significant sound source information and relatively high signal-to-noise ratio. The Low-Frequency Analysis Representation (LOFAR) spectrum transforms the signal received by the passive sonar from time domain to time-frequency domain using the short-time Fourier transform (STFT), which can reflect the signal in the two dimensions of time domain and frequency domain. Scientists observe the line spectrum in the LOFAR spectrum to determine the presence or absence of the target, and perform tracking and recognition [9]. Because there is more demand of the stealth technology of the ship and the radiated noise of the ship's target is greatly reduced, the signal-to-noise ratio of the ship radiated noise received by the hydrophone array is also decreasing. The line spectrum components become more difficult to identify. There are many research results on automatic detection and extraction of line spectrum under low signal-to-noise ratio.

In this paper, we study from the multi-step decision algorithm based on the line spectrum feature cost function proposed by Di Martino [32]. Then we propose a specific calculation method of double threshold, and retain the continuous spectrum information in the original LOFAR spectrum. At last, we combine the original LOFAR spectrum with the extracted line spectrum, and complete the recognition and detection of underwater target by making full use of the advantages of deep neural network feature extraction.

3.1. Structure LOFAR Spectrum

The LOFAR spectrum is calculated by short-time Fourier transform (STFT). Unlike the traditional Fourier transform, which requires signal stability, STFT is suitable for non-stationary signals. It takes advantage of the short-term stationary characteristics of the signal. After windowing and framing the signal, the Fourier transform is performed to obtain the signal at time-frequency. Then it is more accurately characterize the distribution of signal frequency components and time nodes. The calculation formula is as follows:

$$STFT\{s(t)\} = \int_{-\infty}^{\infty} s(t)w(t - \tau)e^{-j\omega\tau}d\tau, \quad (16)$$

where $STFT\{\cdot\}$ is short-time Fourier transform, $s(t)$ is the signal to be transformed and $w(t)$ is the window function (truncating function). The process of calculating the LOFAR spectrum can be compared with the "LOFAR Spectrum" in the Feature Processing stage in Figure 1. The specific calculation steps are as follows:

- (1) Framing and windowing. The sound signal is unstable globally, but can be regarded as stable locally. In the subsequent speech processing, a stable signal needs to be input. Therefore, it is necessary to frame the entire speech signal, i.e., to divide it into multiple segments. We divide the sampling sequence of the signal into K frames and each frame contains N sampling points. The larger the N and K , the larger the amount of data, and the closer the final result is to the true value. Due to the correlation between the frames, there are usually some points overlap between the two frames. Framing is equivalent to truncating the signal, which will cause distortion of its spectrum and leakage of its spectral energy. To reduce spectral energy leakage, different truncation functions which are called window function can be used to truncate the signal. The practical application window functions include Hamming window, rectangular window and Hanning window, etc.
- (2) Normalization and decentralization. The signal of each frame needs to be normalized and decentralized, which can be calculated by the following formula:

$$s''(t) = \frac{s'(t) - E[s'(t)]}{\max(|s'(t)|)}. \quad (17)$$

Here, $s'(t)$ is the normalization of $s(t)$, which makes the power of the signal uniform in time. $s''(t)$ is the decentralization of $s(t)$, which makes the mean of the samples zero.

- (3) Perform Fourier transform on each frame signal and arrange the transformed spectrum in the time domain to obtain the LOFAR spectrum.

3.2. Analysis and Construction of Line Spectrum Cost Function

The definition of the line spectrum feature cost function is as follows:

$$O(\eta) = \frac{\lambda F(\eta) + \mu T(\eta)}{A(\eta)}, \quad (18)$$

where η represents a summation path along the time axis in the observation window of the LOFAR graph, and the length of the path is N . $A(\eta)$ characterize the amplitude characteristics of the line spectrum, $F(\eta)$ is the frequency continuity of the line spectrum, and $T(\eta)$ is the trajectory continuity of the line spectrum, λ and μ are weighting coefficients. The definitions of $A(\eta)$, $F(\eta)$, and $T(\eta)$ are as follows:

$$A(\eta) = \sum_{i=1}^N a(P_i), \quad (19)$$

$$F(\eta) = \sum_{i=3}^N |d(P_{i-2}, P_{i-1}) - d(P_{i-1}, P_i)|, \quad (20)$$

$$T(\eta) = \sum_{i=1}^N g(P_i). \quad (21)$$

Each pixel on the summing path is $P_i (1 \leq i \leq N)$, which means a point on the i line of the time axis. $a(P_i)$ characterizes the amplitude of the point P_i . $d(P_{i-1}, P_i)$ characterizes the frequency gradient at two points in the path, which is defined as follows:

$$d(P_{i-1}, P_i) = f(P_{i-1}) - f(P_i), \quad (22)$$

where $f(P_i)$ represents the frequency of the point P_i . $g(P_i)$ characterizes the breakpoint identification, which is defined as follows:

$$g(P_i) = \begin{cases} 1 & a(P_i) < \varepsilon \\ 0 & \text{others} \end{cases}. \quad (23)$$

If the amplitude of the point P_i is less than ε , it is regarded as a breakpoint and recorded as 1, otherwise it is recorded as 0. Regarding the calculation of the threshold ε , the original algorithm is mostly set by empirical values, and a new calculation method is proposed as follows:

$$P(w) = |STFT\{n(t)\}|^2, \quad (24)$$

$$\varepsilon = A_{average} = \sqrt{\frac{\sum P(w)}{M * N}}, \quad (25)$$

where $n(t)$ represents the marine environmental noise. The sampling sequence of the interference noise in the marine environment is subjected to STFT transformation which can obtain the LOFAR spectrum. At the same time, the instantaneous power $p(w)$ of each time-frequency point is calculated. M, N represent the points of frequency domain and time domain of LOFAR spectrum. The power of all time and frequency points is summed and averaged to obtain the average power. Take a square to obtain the average amplitude of the LOFAR spectrum of marine environment interference noise, which is the threshold ε for determining whether the point p_i is a breakpoint.

It can be analyzed from the cost function: the problem of line spectrum detection is transformed into the problem of finding the optimal path η and minimizing the cost function about the path η .

3.3. Sliding Window Line Spectrum Extraction Algorithm Based on Multi-Step Decision

In this section, a sliding window line spectrum extraction algorithm based on multi-step decision is used to search for the optimal path. As shown in Figure 5, in this algorithm, a window which can slide along the frequency axis and cover the whole time axis is set in the LOFAR spectrum. We search the optimal path in this window. The reason for setting the window is that there may be multiple line spectrum co-existing in the LOFAR spectrum. By properly setting the size of the window, the search range of the path can be limited to a certain region of the LOFAR spectrum. Then the line spectrum in each window can be extracted, which can avoid that only the strongest spectral line is extracted in the whole LOFAR spectrum.

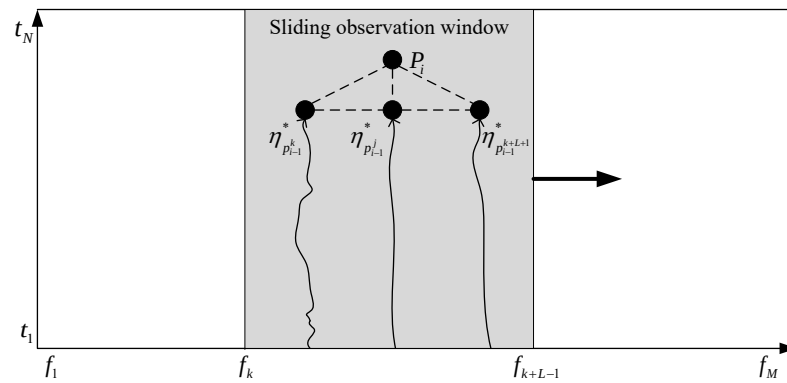


Figure 5. Frequency-domain sliding window multi-step decision dynamic tracking line spectrum.

To cover a line spectrum in a search window, the size of the window is related to the line spectrum broadening and the frequency resolution in the LOFAR spectrum. The specific calculation steps are as follows:

- (1) The length of the frequency axis in the LOFAR spectrums M . The start point is f_1 , and the end point is f_N . The length of the time axis is N . The start point is t_1 , and the end points t_N . The search window size is defined as L .
- (2) Each point in the figure is defined as P_i^j , representing the time-frequency pixel on the j th column on the frequency axis and the i th row on the time axis, where $1 \leq j \leq M$, $1 \leq i \leq N$. $\eta_{P_i^j}^*$ represents the optimal path from t_1 to t_N in the observation window, $A(\eta_{P_i^j}^*), F(\eta_{P_i^j}^*), T(\eta_{P_i^j}^*)$ defines as a set of ternary vectors for points P_i^j , and the triplet of each point at t_1 is initialized to $(a(P_i^j), 0, 0)$.
- (3) From t_2 to t_N , find the optimal path with length from 2 to N in the search window line by line. In the figure, P_i is set to any point in t_1 , the start position of the observation window is f_k , and the corresponding end position is f_{k+L-1} . At t_{i-1} , the neighboring L points of P_i form a set as follows, $V(P_i) = \{P_{i-1}^k, \dots, P_{i-1}^{k+L-1}\}$, the optimal path $\eta_{P_i}^*$ to the length i of the point P_i is obtained from the optimal path $\eta_{P_{i-1}^j}^*$ of $P_{i-1}^j \in V(P_i)$, i.e., $\eta_{P_i}^* = \eta_{P_{i-1}^j}^* \cup \{P_i\}$, where $k \leq j \leq k+L-1$ satisfies:

$$O(\eta_{P_i}^*) = \min_{P_{i-1}^j \in V(P_i)} O(\eta_{P_{i-1}^j}^* \cup \{P_i\}), \quad (26)$$

$$j = \operatorname{argmin}_{P_{i-1}^j \in V(P_i)} O(\eta_{P_{i-1}^j}^* \cup \{P_i\}), \quad (27)$$

where $\{P_i\}$ represents the set of points P_i .

- (4) At t_N , the optimal path of the k points P_N^j in the search window is $\eta_{P_N^j}^*$, where $k \leq j \leq k+L-1$, then the optimal path of length N in the search window is:

$$O(\eta^*) = \min_{k \leq j \leq k+L-1} O(\eta_{P_N^j}^*). \quad (28)$$

- (5) Set a counter for each time-frequency point in the LOFAR spectrum, and the counter value is initialized to 0. If the value of the objective function $O(\eta^*)$ corresponding to the optimal path η^* in the search window is greater than the threshold γ , we would consider that there is a line spectrum on the optimal path, and the counter values corresponding to the N points on the optimal path are increased by 1 respectively. The specific steps of threshold calculation are as follows:
First, the input of the algorithm is changed from the LOFAR spectrum of ship radiation noise to the LOFAR spectrum of marine environmental noise. The corresponding cost function $O(\eta_{noise}^r)$ of the optimal path η_{noise}^r in the r th observation window is obtained, where $1 \leq r \leq M - L + 1$ then the threshold is:

$$\gamma = \min_{1 \leq r \leq M} O(\eta_{noise}^r). \quad (29)$$

- (6) Slide the search window with a step size of 1. Repeat the above steps until the observation window slides to the end. The output count value graph is the traced line spectrum.

4. Underwater Target-Recognition Framework Based on CNN

CNN has proven its powerful capability in many fields, such as pattern recognition, computer vision, nature language processing, and wireless physical layer [1,7,8,10]. It is composed of input layer, convolutional layer, activation function, pooling layer, and fully connected layer. The input layer can handle multi-dimensional data; similar to other neural networks, because of the use of gradient descent learning, the input features of CNN need to be standardized. The convolutional layer is used for feature extraction. The activation layer strengthens the expressive ability of the neural network. The pooling layer compresses the input features and extracts the main features [33–35]. The fully connected layer connects all the features and sends the output value to the classifier.

From the LOFAR spectrum of the measured underwater acoustic signal which is extracted through multi-step judgment, we design a convolutional neural network structure according to its characteristics. The specific network parameters can be seen in Table 1. For this CNN network structure, it refers some ideas of the inception module which uses different sizes of convolution kernels and weighs the characteristics of the global and local information distribution. This network structure selects different convolution kernels and pooling kernels for preliminary feature extraction. The output of each sub-layer is cascaded and passes through several convolutional layers and pooling layers. Finally, the flatten layer extends the feature map into vectors and the network completes the classification by the dense layer. Convolution and pooling performed in parallel in the network obtain features of different information scales. The network has strong feature extraction capabilities for the positional relationship of line spectrum on different frequency points in the LOFAR spectrum.

The network parameters of CNN have been marked in Table 1. $(p * q) * r$ means the size of the convolution kernel is $(p * q)$, r means the number of channels. $stride = m * n$ means the step size is $m * n$. Conv and MaxPool are convolution layer and max pooling layer, respectively. CNN training and optimization hyperparameters are shown in Table 2.

Table 1. The CNN network model parameters of measured dataset.

Input Layer $(1024) \times 64 \times 1$				
Conv+ ReLU $(7 \times 7) \times 16$ stride = 2×1	Conv+ ReLU $(7 \times 5) \times 16$ stride = 2×1	Conv+ ReLU $(5 \times 5) \times 16$ stride = 2×1	Conv+ ReLU $(3 \times 3) \times 16$ stride = 2×1	MaxPool (3×3) Conv+ReLU $(1 \times 1) \times 16$ stride= 1×1
Fileter concatenation				
ReLU+MaxPool (3×3)				
Conv+ReLU $(5 \times 5) \times 16$ stride = 2×1				
MaxPool (3×3)				
Conv+ReLU $(5 \times 5) \times 16$ stride = 2×1				
MaxPool (3×3)				
Conv+ReLU $(3 \times 3) \times 32$ stride = 2×2				
MaxPool (3×3)				
Flattern				
Dense (4)				

Table 2. CNN training, optimization hyperparameters.

Item	Value
Optimizer	adam
Learning rate	0.01
Number of samples	200
Training round	30
Loss function	Cross entropy loss function

Adam was originally proposed by Diederik Kingma of OpenAI and Jimmy Ba of the University of Toronto [36]. It is a first-order optimization algorithm that can replace the traditional stochastic gradient descent process. It can iteratively update neural network weights based on training data.

5. Numerical Results

5.1. Source of Experimental Data

The experimental data used in this article is divided into two parts: The first part of the underwater acoustic database is named ShipsEar [5], which was recorded by David et al. in the port of Vigo and it is vicinity on the Atlantic coast of northwestern Spain. The second part is based on the four types of signals simulated by the ship radiated noise. By mixing with the audio No. 81–92 in the database which are treated as the pure marine environment background noise, the simulated actual ship radiated noise under different signal-to-noise ratios is obtained.

Vigo Port is one of the largest ports in the world with a considerable cargo and passengers. Taking advantage of the high traffic intensity of the port and the diversity of ships, it can record the radiated noise of many different types of ships on the dock, including fishing boats, ocean liners, Roll-on/Roll-off ships, tugboats, yachts, small sailboats, etc. The ShipsEar database contains 11 ship types (marine environmental noise) and a total of 90 audio recordings in "wav" format, with audio lengths varying from 10 s to 11 min.

By extracting and summarizing audios in the database, it is divided into four categories according to the size of the ship types collected which is shown in Table 3. In addition, the date and weather conditions of the collected audios, the coordinates and driving status of the ship's specific position, the number, depth and power gain of hydrophones,

atmospheric and marine environmental data are also listed in detail. The information can be used as a reference in the study.

Table 3. Four types of ship targets.

Item	Value
W	Fishing boat, trawler, mussel harvester, tugboat, dredge
X	Motorboat, pilot boat, sailboat
Y	Passenger ferry
Z	Ocean liner, ro-ro ship

Because of military security considerations in the field of underwater target recognition, military databases are mostly kept secret. However, due to the inconvenience of collection and the high cost of civil databases, there are few public civilian databases for researchers to use. After the emergence of the ShipsEar database, it has been used in the application research of ship radiated noise separation, denoising, classification, etc. It is also common to use this database to complete research in the field of deep learning [18–21,37–39].

5.2. Experimental Software and Hardware Platform

The hardware platform and software support required to complete the deep-learning experiment are shown in Table 4.

Table 4. Experimental hardware platform and software support.

Lab Environment	Configuration
Operating system	Ubuntu 16.04
Graphics card	GTX 1080ti
Programming language and software	Pycharm 2019.1 + Python 3.6
	Matlab R2016b
Deep-learning library and software toolbox	Keras 2.3 (tensorflow backed)
	Librosa Audio processing library (python)
	TQWT Toolbox (Matlab language)

During the whole experiment, we mainly use two pieces of software, Pycharm 2019.1 and MATLAB R2016B. Referring to Figure 1, we used MATLAB to complete the entire process of feature extraction. The algorithm in Section 3 is based on MATLAB for the calculation of LOFAR spectrum samples. For the algorithm in Section 4, we used Pycharm and Python to complete the task of CNN-based underwater target recognition. In addition, the depth library and software toolbox used are Keras (offline training and online testing), Librosa (all phasis) and TWQT (sampling and feature preprocessing).

5.3. Multi-Step Decision LOFAR Line Spectrum Enhancement Algorithm Validity Test

In this section, the audio data of ShipsEar (a database of measured ship radiated noise) is used to verify the effectiveness of the algorithm.

In the process of testing the effectiveness of the multi-step decision LOFAR line spectrum enhancement algorithm, morphological component analysis is required. In Section 2.1.1, we detailed the use of the RSSD algorithm to construct the optimal sparse representation of the high and low-resonance components in the ship radiated noise. In the RSSD algorithm, it is necessary to select an appropriate filter scaling factor according to

the waveform characteristics of the signal, to calculate the parameters corresponding to the two types of resonance components, and construct the corresponding wavelet basis functions. In this regard, we have completed the calculation of the parameters with the help of MATLAB. For the signal decomposition algorithm based on resonance, the parameters setting for extracting high-resonance components are $Q_h = 4$, $r_h = 3$, $J_h = 32$, and the parameters setting for extracting low-resonance components are $Q_l = 1$, $r_l = 3$, $J_l = 3$. The energy distribution of the low-resonance component signal and the high-resonance component signal is finally calculated as follows:

From the energy percentage of each frequency band in Figure 6, the energy distribution of the low-resonance component is mostly concentrated in the higher frequency band (greater than 1000 Hz), while the energy distribution in the low-frequency band is very small. Compared to Figure 7, we find that higher energy distribution of the original signal comes from the low-resonance component. In Figure 8, most of the energy of the high-resonance component is concentrated in the low-frequency narrow band, and the narrow band energy distribution characteristic is usually regarded as a line spectrum. In previous studies, the low-frequency line spectrum is the main manifestation of mechanical noise and propeller cavitation noise in the LOFAR spectrum. It is also an important basis for the identification of ship radiated noise. Therefore, the separated high-resonance component retains the main features of underwater target recognition well.

In addition, Spectral Correlation Coefficient (SCC) [40] can also be used to measure the effectiveness of the RSSD algorithm. The physical significance of the spectral correlation coefficient is measuring the similarity of the power spectrum of the two signals, which is defined as follows:

$$C_{A,B} = \frac{\int_{f_1}^{f_2} N_A(f) * N_B(f) df}{\sqrt{\int_{f_1}^{f_2} N_A^2(f) * \int_{f_1}^{f_2} N_B^2(f) df}}, \quad (30)$$

where $N_A(f)$ and $N_B(f)$ represent the power spectrum of the two types of signals A and B, respectively. f_1 and f_2 represent the range of the power spectrum. This means that the radiated noise of the two types of ships with a higher degree of difference has a smaller spectral correlation coefficient. It can be seen from Table 5 that the spectral correlation coefficients in the high-resonance components of signals A and B are smaller than their original spectral correlation coefficients. It means we can enhance the degree of difference between the two signals by extracting the high-resonance components of the signal.

Energy Distribution of Low Resonant Component Signal

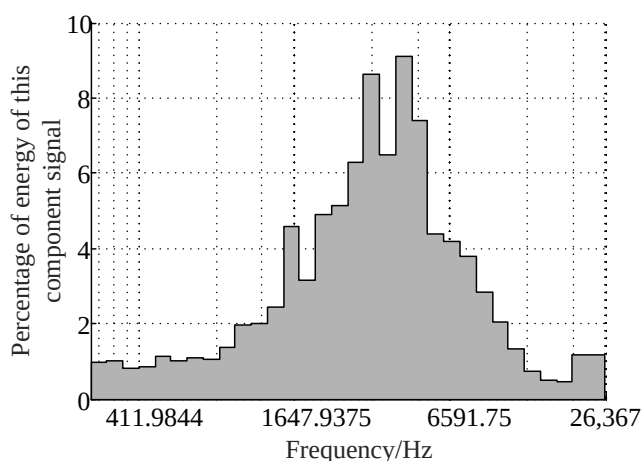


Figure 6. Percentage of total energy of each frequency band of low-resonance component signal.

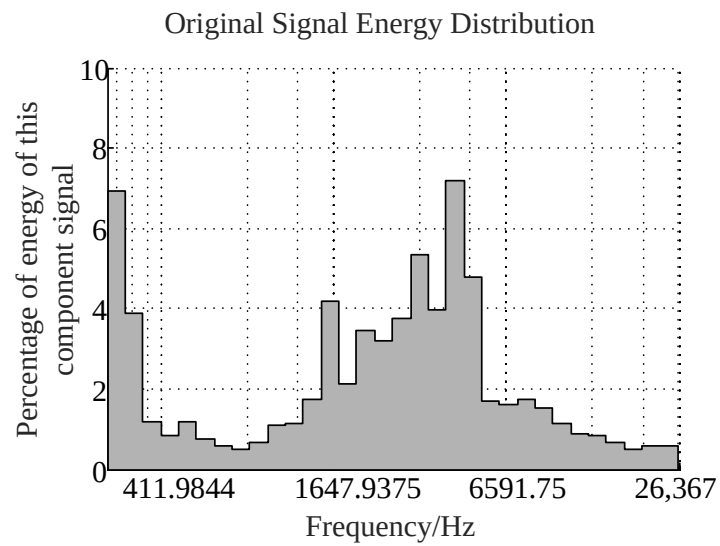


Figure 7. The percentage of total energy of each frequency band of the original signal.

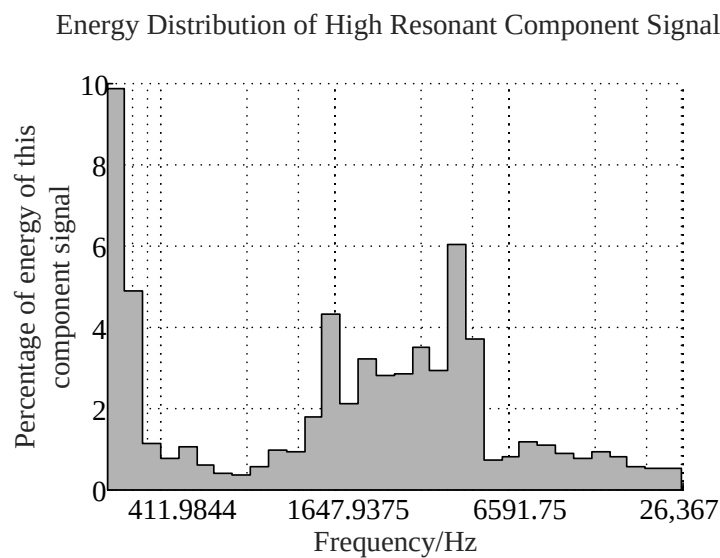


Figure 8. Percentage of total energy of each frequency band of high-resonance component signal.

Table 5. Spectral correlation coefficients between the two types of original signals and their high-resonance components.

Signal	$S_{Original_A}(t), S_{Original_B}(t)$	$S_{high_A}(t), S_{high_B}(t)$
$C_{A,B}$	0.7161	0.7074

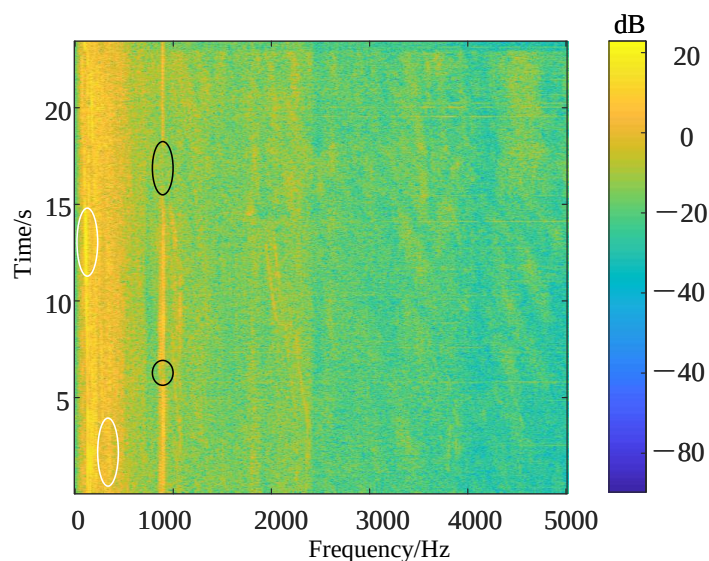


Figure 9. LOFAR spectrum of the original signal.

For the line spectrum enhancement algorithm based on multi-step decision, the experimental results are shown in Figure 9 and 10, which are the LOFAR spectrum of the original signal and the LOFAR spectrum after line spectrum enhancement. In Figure 9, there is an obvious line spectrum in the part marked by white circles, but the line spectrum is broken in the part marked by black circles. In Figure 10, the line spectrum indicated by the white circles are extended to completeness, and the vacant part of the line spectrum indicated by the black circles is also completed. Therefore, even if the line spectrum in the LOFAR spectrum has “breakpoints”, “broken lines” or only a short line due to noise interference, the line spectrum enhancement algorithm can still extend and complete the line spectrum.

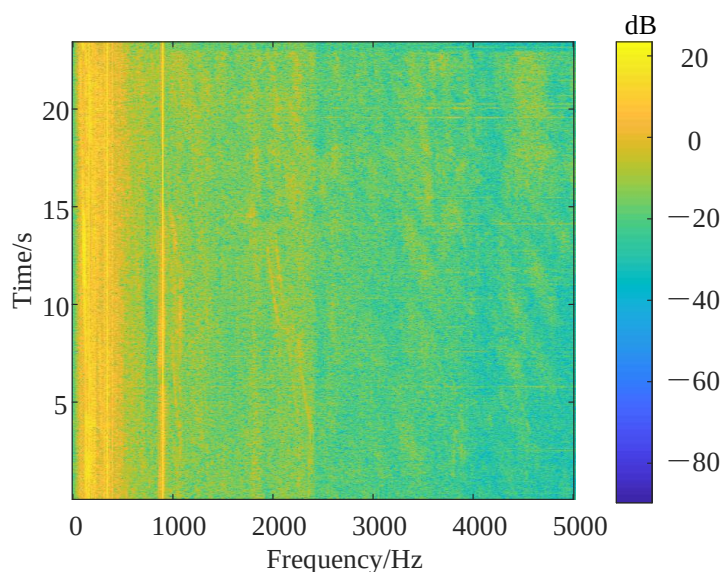


Figure 10. LOFAR spectrum after line spectrum enhancement.

5.4. Experimental Verification of Underwater Target Recognition Based on Convolutional Neural Network (CNN)

5.4.1. CNN Network Offline Training Process

According to the frame structure of underwater target recognition in Figure 1, the specific settings and steps can be divided into:

- (1) Read the high-resonance component signals in sequence, then windowing and framing the signal. We choose Hanning window (Hanning), and the window size is 2048 (i.e., FFT points are 2048). The overlap between frames is 75%.
- (2) The signal of each frame is normalized and decentralized. The power of the signal is uniform in time and the average value of the sample is 0. It means the data are limited to a certain range, which can eliminate singular sample data. At the same time, it can also avoid the saturation of neurons and accelerate the convergence rate of the network.
- (3) First, perform Fourier transform on each frame signal. Second, take the logarithmic amplitude spectrum of the transformed spectrum and arrange it in the time domain. Last, take 64 points on the time axis as a sample, which means obtaining a size of $1024 * 64$ LOFAR spectrum sample. The sampling frequency of audio is 52,734 Hz, and the duration of each sample is about 0.62 s. The numbers of training and testing sets of various samples are shown in Table 6. The ID in the table is the label of the audio in the ShipsEar database. The corresponding type of ship for each segment can be obtained according to the ID. The type of ship corresponding to audio is used as a label for supervised learning of deep neural networks.

Table 6. various sample training sets and testing sets

	ID	Training Set	Testing Set
		Number of Samples	Number of Samples
W	46, 48, 66, 73, 74, 75, 80, 93, 94, 95, 96	836	531
X	21, 26, 29, 30, 50, 52, 57, 70, 72, 77, 79	837	516
Y	6, 7, 8, 10, 11, 13, 14	1016	526
Z	18, 19, 20	1149	603
Total		3838	2176

- (4) The sample obtained in step (3) is treated with LOFAR spectrum enhancement. The specific sample processing process and calculation process are in Section 3.3. Then the LOFAR spectrum with enhanced line spectrum characteristics is obtained. The LOFAR spectrum is a two-dimensional matrix, which can be regarded as a single-channel image. After that as shown in Figure 1, the data enhanced by the multi-step decision LOFAR spectrum is input into the CNN network for subsequent identification.

5.4.2. Identification of Measured Ship Radiated Noise

The testing data set adopts the same feature preprocessing as the training set and inputs the trained model to complete the test.

Figure 11 shows the standardized confusion matrix. The recognition accuracy of the radiated noise of the four types of ships is different. Among them, the recognition effect of the Y-signal is the best and the recognition accuracy rate reaches 100.00%. The recognition

accuracy rates of the W-type and Z-type are slightly worse, and they are 95.10% and 97.68%. Additionally, the recognition effect of the X-type is the worst, which is only 87.60%. In summary, the total recognition accuracy rate is 95.22%. The recognition accuracy of four kinds of measured ship radiated noise is shown in Table 7.

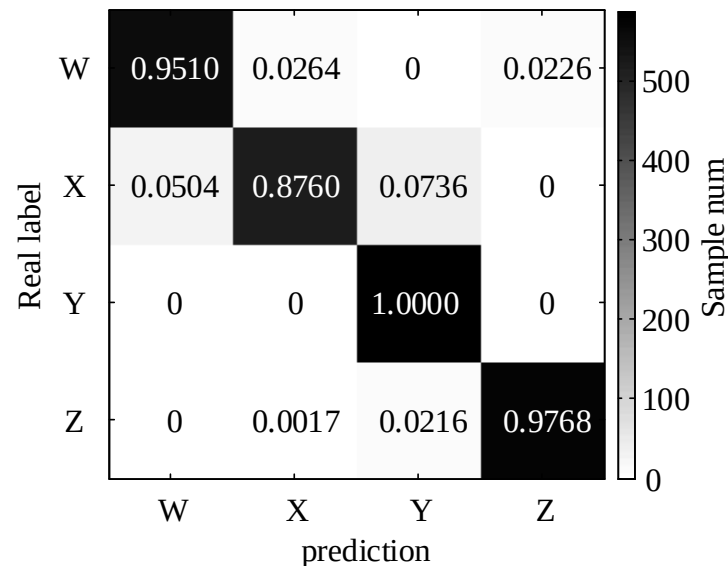


Figure 11. Confusion matrix of four types of measured ship radiated noise under CNN.

Table 7. The recognition accuracy rate of the four types of measured ship radiated noise.

Recognition Rate	Class W	Class X	Class Y	Class Z	Average
CNN	95.10%	87.60%	100.0%	97.78%	95.22%

Figure 12 shows the ROC curve and the corresponding AUC value of the four types of signals. The horizontal axis uses a logarithmic scale to enlarge the ROC curve in the upper left corner. The ROC curves of the signals of W, Y, and Z are relatively close to the (0, 1) point, and their classification effects are relatively good. However, the ROC curve of the signals of type X is closest to the 45-degree line, so the classification effect is worst. Judging from the AUC value, the AUC of the Z-type signal is the highest, which reaches 0.9981. The AUC values of the W-type and Y-type signals have respectively reached 0.9952 and 0.9925. The AUC value of the X-type signal is only 0.9702. Therefore, the classification effect of the X-type signal is also inferior to the other three types of signals.

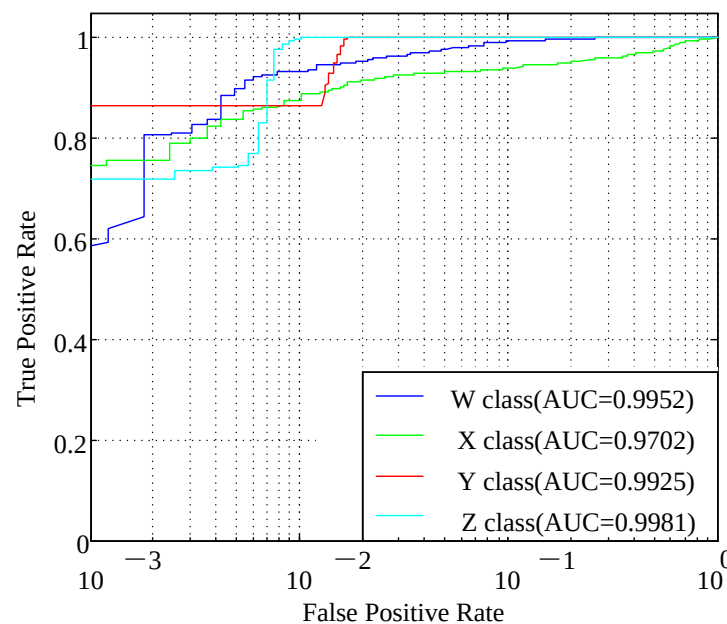


Figure 12. The ROC curve and AUC value of four types of measured ship radiated noise under CNN.

6. Conclusions

In this paper, we have studied underwater target recognition using the LOFAR spectrum. First, a deep-learning underwater target-recognition framework based on multi-step decision LOFAR line spectrum enhancement is developed, in which we use CNN for offline training and online testing. Under the developed underwater target-recognition framework, we then use the LOFAR spectrum as the input of CNN. In particular, on calculating the LOFAR spectrum of the high-resonance component, we use the algorithm based on resonance and design the LOFAR spectrum line enhancement algorithm which is based on multi-step decision. To the best of our knowledge, the difference between the radiated noise of different types of ships is enhanced, and the broken line spectrum can be detected and enhanced. Finally, we conduct extensive experiments in terms of the detection performance, scalability, and complexity. The results have shown that the LOFAR-CNN method can achieve the highest recognition rate of 95.22% with the measured ship radiation noise which can further improve the recognition accuracy compared with other traditional method.

7. Future Works

This paper uses deep-learning methods to provide a framework to realize underwater target recognition. This algorithm shows excellent underwater target-recognition ability, and has great application value in many aspects such as seabed exploration, oil platform monitoring and economic fish detection. It can not only predict dangerous objects in advance, improve the safety of ship navigation, but also create greater economic benefits. However, there are still some shortcomings that need to be resolved.

- (1) Most studies on underwater target recognition do not disclose data sources for reasons such as confidentiality. There is also a lack of unified and standardized data sets in the industry. The actual measured ship radiated noise data set used in this article is already one of the few publicly available underwater acoustic data sets. However, the data set itself is seriously disturbed by marine environmental noise. The number of samples of various types of ships is unevenly distributed, and the total number of samples is also insufficient. Therefore, how to combine underwater target recognition with deep learning under limited conditions is a big problem.

- (2) Because the data set is seriously disturbed, this paper adopts a series of feature-enhancing preprocessing methods to improve the recognition rate, and has achieved excellent results. In fact, further reducing the impact of ocean noise and evaluating the impact of various neural networks on the recognition effect can be considered in the future work.

Author Contributions: Conceptualization, J.C.; methodology, J.C., B.H., X.M. and J.Z.; software, B.H. and X.M.; validation, B.H. and X.M.; formal analysis, J.C. and B.H.; investigation, J.C. and B.H.; resources, J.C. and B.H.; data curation, J.C., B.H., X.M. and J.Z.; writing—original draft preparation, J.C. and B.H.; writing—review and editing, J.C. and B.H.; visualization, J.C. and B.H.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China National Key R&D Program under grant 2020YFB1807700.

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

- Xie, J.; Fang, J.; Liu, C.; Li, X. Deep learning-based spectrum sensing in cognitive radio: A CNN-LSTM approach. *IEEE Commun. Lett.* **2020**, *24*, 2196–2200.
- Liu, C.; Yuan, W.; Li, S.; Liu, X.; Ng, D.W.K.; Li, Y. Learning-based Predictive Beamforming for Integrated Sensing and Communication in Vehicular Networks. *arXiv* **2021**, arXiv:2108.11540.
- Xie, J.; Fang, J.; Liu, C.; Yang, L. Unsupervised deep spectrum sensing: A variational auto-encoder based approach. *IEEE Trans. Veh. Technol.* **2020**, *69*, 5307–5319.
- Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Phy.* **1962**, *160*, 106–154.
- Santos-Domínguez, D.; Torres-Guijarro, S.; Cardenal-López, A.; Pena-Gimenez, A. ShipsEar: An underwater vessel noise database. *Appl. Acoust.* **2016**, *113*, 64–69.
- Ciaburro, G.; Iannace, G. Improving Smart Cities Safety Using Sound Events Detection Based on Deep Neural Network Algorithms. *Informatics* **2020**, *7*, 23, doi:10.3390/informatics7030023.
- Liu, C.; Wei, Z.; Ng, D.W.K.; Yuan, J.; Liang, Y.C. Deep transfer learning for signal detection in ambient backscatter communications. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 1624–1638.
- Ciaburro, G. Sound Event Detection in Underground Parking Garage Using Convolutional Neural Network. *Big Data and Cognitive Computing* **2020**, *4*, 20, doi:10.3390/bdcc4030020.
- Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016.
- Liu, C.; Wang, J.; Liu, X.; Liang, Y.C. Deep CM-CNN for spectrum sensing in cognitive radio. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2306–2321.
- LeCun, Y.; Boser, B.E.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.E.; Jackel, L.D. Handwritten digit recognition with a back-propagation network. In Proceedings of the Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, Denver, CO, USA, November 1989; pp. 396–404. Available online: <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf> (accessed on 4 October 2021).
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-first AAAI conference on artificial intelligence (AAAI), San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, 17–19 June 2016; pp. 770–778.
- Jin, G.; Liu, F.; Wu, H.; Song, Q. Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. *J. Exp. Theor. Artif. Intell.* **2019**, *32*, 205–218.
- Liu, F.; Song, Q.; Jin, G. Expansion of restricted sample for underwater acoustic signal based on generative adversarial networks. May 2019. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP), Chengdu, China, 12–14 December 2018; Volume 11069, pp. 1106948–1106957.

18. Yang, H.; Shen, S.; Yao, X.; Sheng, M.; Wang, C. Competitive deep-belief networks for underwater acoustic target recognition. *Sensors* **2018**, *18*, 952–965.
19. Shen, S.; Yang, H.; Sheng, M. Compression of a deep competitive network based on mutual information for underwater acoustic targets recognition. *Entropy* **2018**, *20*, 243–256.
20. Yan, J.; Sun, H.; Chen, H.; Junejo, N.U.R.; Cheng, E. Resonance-based time-frequency manifold for feature extraction of ship-radiated noise. *Sensors* **2018**, *18*, 936–957.
21. Ke, X.; Yuan, F.; Cheng, E. Underwater Acoustic Target Recognition Based on Supervised Feature-Separation Algorithm. *Sensors* **2018**, *18*, 4318–4342.
22. Zhu, P.; Isaacs, J.; Fu, B.; Ferrari, S. Deep learning feature extraction for target recognition and classification in underwater sonar images. In Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, Australia, 12–15 December 2017; pp. 2724–2731.
23. McQuay, C.; Sattar, F.; Driessen, P.F. Deep learning for hydrophone big data. In Proceedings of the 2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), Victoria, QC, Canada, 21–23 August 2017; pp. 1–6.
24. Hu, G.; Wang, K.; Peng, Y.; Qiu, M.; Shi, J.; Liu, L. Deep learning methods for underwater target feature extraction and recognition. *Comput. Intell. Neurosci.* **2018**, *2018*, 1–10.
25. Kubáčková, L.; Burda, M. Mathematical model of the spectral decomposition of periodic and non-periodic geophysical stationary random signals. *Stud. Geophys. Geod.* **1977**, *21*, 1–10.
26. Huang, W.; Sun, H.; Liu, Y.; Wang, W. Feature extraction for rolling element bearing faults using resonance sparse signal decomposition. *Exp. Tech.* **2017**, *41*, 251–265.
27. Selesnick, I.W. Wavelet transform with tunable Q-factor. *IEEE Trans. Signal Process.* **2011**, *59*, 3560–3575.
28. Starck, J.L.; Elad, M.; Donoho, D.L. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. Image Process.* **2005**, *14*, 1570–1582.
29. Al-Raheem, K.F.; Roy, A.; Ramachandran, K.; Harrison, D.K.; Grainger, S. Rolling element bearing faults diagnosis based on autocorrelation of optimized: wavelet de-noising technique. *Int. J. Adv. Manuf. Technol.* **2009**, *40*, 393–402.
30. Shensa, M.J. The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Trans. Signal Process.* **1992**, *40*, 2464–2482.
31. Wang, H.; Chen, J.; Dong, G. Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform. *Mech. Syst. Signal Proc.* **2014**, *48*, 103–119.
32. Di Martino, J.C.; Haton, J.P.; Laporte, A. Lofargram line tracking by multistage decision process. In Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Minneapolis, MN, USA, 27–30 April 1993; Volume 1, pp. 317–320.
33. Liu, C.; Liu, X.; Ng, D.W.K.; Yuan, J. Deep Residual Learning for Channel Estimation in Intelligent Reflecting Surface-Assisted Multi-User Communications. *IEEE Trans. Wirel. Commun.* **2021**. doi:10.1109/TWC.2021.3100148.
34. Liu, X.; Liu, C.; Li, Y.; Vucetic, B.; Ng, D.W.K. Deep residual learning-assisted channel estimation in ambient backscatter communications. *IEEE Wirel. Commun. Lett.* **2020**, *10*, 339–343.
35. Liu, C.; Yuan, W.; Wei, Z.; Liu, X.; Ng, D.W.K. Location-aware predictive beamforming for UAV communications: A deep learning approach. *IEEE Wirel. Commun. Lett.* **2020**, *10*, 668–672.
36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Chen, Z.; Li, Y.; Cao, R.; Ali, W.; Yu, J.; Liang, H. A New Feature Extraction Method for Ship-Radiated Noise Based on Improved CEEMDAN, Normalized Mutual Information and Multiscale Improved Permutation Entropy. *Entropy* **2019**, *21*, 624–640.
38. Yuan, F.; Ke, X.; Cheng, E. Joint Representation and Recognition for Ship-Radiated Noise Based on Multimodal Deep Learning. *J. Mar. Sci. Technol. Eng.* **2019**, *7*, 380–397.
39. Ke, X.; Yuan, F.; Cheng, E. Integrated optimization of underwater acoustic ship-radiated noise recognition based on two-dimensional feature fusion. *Appl. Acoust.* **2020**, *159*, 107057–107070.
40. Hou, W. Spectrum autocorrelation. *Acta Acust* **1988**, *2*, 46–49.