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

Jie Chen, Member, IEEE, Jie Liu, Chang Liu, Member, IEEE, Jian Zhang, and Bing Han

Abstract—The Low frequency analysis and recording (LOFAR) spectrum is one of the key features of the underwater target, which can be used for underwater target recognition. However, the underwater environment noise is complicated and the signal-to-noise ratio of the underwater target is rather low, which introduces the breakpoints to the LOFAR spectrum and thus hinders the underwater target recognition. To overcome this issue and to further improve the recognition performance, we adopt a deep learning approach for underwater target recognition and propose a LOFAR spectrum enhancement (LSE)-based underwater target recognition scheme, which consists of preprocessing, offline training, and online testing. In preprocessing, a LOFAR spectrum enhancement based on multi-step decision algorithm is specifically designed to recover the breakpoints in LOFAR spectrum. In offline training, we then adopt the enhanced LOFAR spectrum as the input of convolutional neural network (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. 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.

Index Terms—Underwater target recognition, LOFAR spectrum, line spectrum enhancement, deep learning.

I. INTRODUCTION

The vast ocean contains rich mineral resources, marine living resources and chemical resources which can be exploited for economic benefits. Thus, the marine developments, e.g., the submarine prospecting, the oil platform monitoring, and the economic fish detection, are of great importance. Specifically, one of the key tasks in marine developments is the underwater target recognition [1], [2]. Deep learning (DL)-based underwater target recognition is a new way of realizing underwater target recognition in addition to the existing recognition methods to extract features and train classifiers manually. By 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.

Recently, DL techniques [3] have been exploited for wireless physical layer [4]–[7] and many effective and efficient DL-based schemes have been proposed for underwater target recognition. For example, [8] and [9] focused on underwater target recognition which didn’t have insufficient training samples. In the first step, the original audio was converted into LOFAR spectrum, and then generative adversarial networks (GAN) was used for sample expansion. In the second step, a 15% performance improvement could be obtained by using convolutional neural networks (CNNs) for feature learning and classification when the number of samples was more sufficient. [10] 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 [11] 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%. [12], [13] 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, [14]–[16] all used convolutional neural networks for feature extraction, but the application scenarios and the classifiers were different. [14] 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. [15] 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. [16] 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.

Note that the collected raw data is always seriously polluted by environmental noise, which introduces the breakpoints to the LOFAR spectrum. The so-called breakpoints directly affect the subsequent feature extraction and thus degrade the
Fig. 1. Deep learning underwater target recognition framework based on multi-step decision LOFAR line spectrum enhancement.

performance of the subsequent signal processing. Motivated by this, we first adopt a model-based approach to recover the breakpoints in LOFAR spectrum and then adopt a DL approach for task recognition. The main contributions of this paper are as follows.

(1) Different from 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, 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) In order 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, our proposed LOFAR-CNN method can achieve a recognition accuracy of 95.22% which outperforms the state-of-the-art methods.

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.

II. 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 Fig. 1. It is divided into four modules: sampling, feature preprocessing, offline training and online testing.

A. 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. 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. 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.

The RSSD algorithm regards resonance as the basis for signal decomposition, 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 by 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) [20] and Morphological Component Analysis (MCA) [21]. Its algorithm framework is shown in Fig. 2.

1) Morphological component analysis

Morphological component analysis is usually used to decompose signals with different morphological characteristics [22]. 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 in order 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 \(x_l\). \(\Phi_h, \Phi_l\) 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:

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

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 \(\Phi_{h,j}, \Phi_{l,j}\):

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

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

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) [19], 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 \(\alpha, \beta\) 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 \(\Phi_h, \Phi_l\).

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) 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 [23]. 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 Fig. 3 and Fig. 4.
The analysis filter bank of each layer is composed of high-pass filter $H_{\text{high}}(w)$, low-pass filter $H_{\text{low}}(w)$, and the corresponding scaling process, which are defined as follows:

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

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

\(\theta(w) = 0.5(1 + \cos(w))\sqrt{2 - \cos(w)}\) is the Daubechies filter with second-order disappearing moment \(20\). \(\alpha, \beta\) \((0 < \alpha < 1, 0 < \beta < 1)\) 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 $\alpha, \beta$ \([24]\) can be expressed as:

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

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 $\Phi_h, \Phi_l$ are constructed by selecting the oversampling rate $r$. The deepest level $J_{\text{max}}$ and the oversampling rate $r$ are defined as follows:

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

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

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

## III. 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 by 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 \([3]\). Because there are 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 get more difficult to identify. There are a large number of 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 \([25]\). 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 an underwater target by making full use of the advantages of deep neural network feature extraction.

### A. 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:

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

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 specific calculation steps are as follows:

1. Framing and windowing. Divide the sampling sequence of the signal into $K$ frames and each frame contains $N$ sampling points. 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. In order 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:

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

(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.

**B. Analysis and Construction of Line Spectrum Cost Function**

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

\[
O(\eta) = \frac{\lambda F(\eta) + \mu T(\eta)}{A(\eta)},
\]

(18)

where \(\eta\) 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, \(\lambda\) and \(\mu\) are weighting coefficients. The definitions of \(A(\eta)\), \(F(\eta)\), and \(T(\eta)\) are as follows:

\[
A(\eta) = \sum_{i=1}^{N} a(P_i),
\]

(19)

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

(20)

\[
T(\eta) = \sum_{i=1}^{N} g(P_i).
\]

(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:

\[
d(P_{i-1}, P_i) = f(P_{i-1}) - f(P_i),
\]

(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:

\[
g(P_i) = \begin{cases} 
1 & a(P_i) < \varepsilon \\
0 & \text{others}
\end{cases}
\]

(23)

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

\[
P(w) = |STFT\{a(t)\}|^2,
\]

(24)

\[
\varepsilon = A_{\text{average}} = \sqrt{\frac{\sum_{M,N} P(w)}{M \times N}},
\]

(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 get the average amplitude of the LOFAR spectrum of marine environment interference noise, that is the threshold \(\varepsilon\) for determining whether the point \(p_i\) is a breakpoint.

It can be analyzed from the cost function: when the point on the path passes or is close to the line spectrum, the sum of the point amplitude on the path increases, while the frequency gradient, the number of breakpoints and the cost function \(O(\eta)\) decrease. The path which the target cost function is the smallest is considered to have a line spectrum, so the problem of line spectrum detection is transformed into the problem of finding the optimal path \(\eta\) and minimizing the cost function about the path \(\eta\).

**C. Sliding Window Line Spectrum Extraction Algorithm based on Multi-step Decision**

In this section, for the problem of minimizing the cost function mentioned in the previous section, a sliding window line spectrum extraction algorithm based on multi-step decision is used to search for the optimal path. As shown in Fig. 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.

In order to cover a line spectrum in a search window, the size of the window is related to the following two points:

(1) The line spectrum width of the ships’ radiated noise is related to its center frequency. The Doppler frequency shift caused by the ship’s motion will also broaden the line spectrum to a certain extent, so the size of the window needs to ensure that the line spectrum is completely contained in the window;
(2) The size of the STFT frame, which comes from the process of calculating the LOFAR spectrum, determines the frequency resolution in the LOFAR. The size of the window can be calculated by combined with (1).

The specific steps of the sliding window line spectrum extraction algorithm based on multi-step decision are as follows:

(1) Define the search window size \( L \);
(2) Define the ternary vector of each point at time \( t_i \);
(3) From time \( t_2 \) to time \( t_N \), find the optimal path from 2 to \( N \) line by line in the search window.
(4) Get the optimal path with length \( N \) in the search window at time \( t_N \):

\[
O(\eta^*) = \min_{k \leq i \leq k + L - 1} O(\eta^*_{P_k^i}).
\]  

(5) A counter is set for each time-frequency point in the LOFAR spectrum, and the counter value is initialized to 0. If the objective function value \( O(\eta^*) \) corresponding to the optimal path \( \eta^* \) in the search window is greater than the threshold \( \gamma \), the counter values corresponding to \( N \) points on the optimal path will be respectively increased by 1. The specific threshold value is:

\[
\gamma = \min_{1 \leq r \leq M} O(\eta_{\text{noise}}^*).
\]  

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

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 \((0, P_i^j, 0, 0)\).
(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, \( V(P_i) = \{ P_{i-1}^k, \ldots, P_{i-1}^{k+L-1} \} \), the optimal path \( \eta^*_{P_i} \) to the length \( i \) of the point \( P_i \) is obtained from the optimal path \( \eta^*_{P_{i-1}} \) of \( P_{i-1} \) in \( V(P_i) \), that is \( \eta^*_{P_i} = \eta^*_{P_{i-1}} \cup \{ P_i \} \), where \( k \leq j \leq k + L - 1 \) satisfies:

\[
O(\eta^*_{P_i}) = \min_{P_{i-1}^j \in V(P_i)} O(\eta^*_{P_{i-1}} \cup \{ P_i \}),
\]

\[
j = \arg\min_{P_{i-1}^j \in V(P_i)} O(\eta^*_{P_{i-1}} \cup \{ P_i \}),
\]

where \( \{ P_i \} \) represents the set of points \( P_i \).

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

\[
O(\eta^*) = \min_{k \leq i \leq k + L - 1} O(\eta^*_{P_{n-1}}).
\]

(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 \( \eta^* \) in the search window is greater than the threshold \( \gamma \), we would consider that there is a line spectrum on the optimal path, and the counter values corresponding to \( N \) 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 \( O(\eta_{\text{noise}}) \) of the optimal path \( \eta_{\text{noise}}^* \) in the \( r \)th observation window is obtained, where \( 1 \leq r \leq M - L + 1 \) then the threshold is:

\[
\gamma = \min_{1 \leq r \leq M} O(\eta_{\text{noise}}^*).
\]

(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.

IV. UNDERWATER TARGET RECOGNITION FRAMEWORK DESIGN

A. Design of underwater target recognition framework based on convolutional neural network

Recently, CNN has proven its powerful capability in many fields, such as computer vision, nature language processing, and wireless physical layer [26–31]. Convolutional neural networks are deep feedforward neural networks that include operations such as convolution calculations, pooled sampling, and nonlinear activation [3]. [17]. Compared with the traditional feedforward neural networks like 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 [32] in 1990 and applied to the handwritten character detection system. In 2014, Szegedy [33] proposed GoogleLeNet which introduced the inception module. Receptive fields of different sizes enhanced the adaptability of the network to scale. The improved version [34], [35] 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 [36] in 2015 adopted the idea of Shortcut Connection (SC) to solve the problem of network degradation.

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 [I]. For this CNN network structure, it refers some ideas of the Inception module which uses different
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 III. 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.

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 have 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 [10]–[13], [37]–[39].

B. Experimental software and hardware platform

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

C. 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.

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

V. NUMERICAL RESULTS

A. 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 [18], which was recorded by David et al.
TABLE IV
EXPERIMENTAL HARDWARE PLATFORM AND SOFTWARE SUPPORT

| lab environment | Configuration          |
|-----------------|------------------------|
| operating system| Ubuntu 16.04           |
| Graphics card   | GTX 1080ti             |
| Programming language and version | 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) |

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

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

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

Energy Distribution of High resonant Component signal

![Energy Distribution of High resonant Component signal](image)

Energy Distribution of Low resonant Component signal

![Energy Distribution of Low resonant Component signal](image)

for extracting low resonance components are \( Q_l = 1, r_l = 3 \), \( J_l = 3 \).

From the energy percentage of each frequency band in Fig. 6, the energy distribution of the low resonance component is mostly concentrated in the higher frequency band (greater than 1000Hz), while the energy distribution in the low frequency band is very small. Comparing with Fig. 7 we find that higher energy distribution of the original signal comes from the low resonance component. In Fig. 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:

\[
C_{A,B} = \frac{\int_{f_1}^{f_2} N_A(f) \star N_B(f) df}{\sqrt{\int_{f_1}^{f_2} N_A^2(f) df \int_{f_1}^{f_2} N_B^2(f) df}},
\]

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 the Table V 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.

For the line spectrum enhancement algorithm based on multi-step decision, the experimental results are shown in Fig. 9 and Fig. 10, which are the LOFAR spectrum of the original signal and the LOFAR spectrum after line spectrum enhancement. In Fig. 9 there is a obvious line spectrum in the part marked by white circles, but the line spectrum is broken in the part marked by black circles. In Fig. 10 the line spectrum
TABLE V
SPECTRAL CORRELATION COEFFICIENTS BETWEEN THE TWO TYPES OF ORIGINAL SIGNALS AND THEIR HIGH RESONANCE COMPONENTS

| signal       | $S_{\text{Original}_A(t)}$, $S_{\text{Original}_B(t)}$ | $S_{\text{high}_A(t)}$, $S_{\text{high}_B(t)}$ |
|--------------|--------------------------------------------------------|------------------------------------------------|
| $C_{A,B}$    | 0.7161                                                 | 0.7074                                         |

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.

D. Experimental verification of underwater target recognition based on convolutional neural network (CNN)

a) CNN network offline training process

According to the frame structure of underwater target recognition in Fig. 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 is 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 getting a size of 1024 * 64 LOFAR spectrum sample. The sampling frequency of audio is 52734 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 VI. 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.

4) The sample obtained in step (3) is treated with LOFAR spectrum enhancement as in the previous section. 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. It will be used as the input layer of the convolutional neural network shown in the figure.

b) Identification of measured ship radiated noise

The testing data set adopts the same feature preprocessing

TABLE VI
VARIOUS SAMPLE TRAINING SETS AND TESTING SETS

| ID | Training set | Testing set |
|----|--------------|-------------|
| 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 | 11449 603 |
| Total | 3838 | 2176 |

TABLE VII
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%  |
as the training set and inputs the trained model to complete the test.

Fig. 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%, respectively. 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 VII.

Fig. 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.

VI. CONCLUSION

In this paper, we have studied underwater target recognition using the LOFAR spectrum. Firstly, 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. Specially, 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.

REFERENCES

[1] 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, pp. 1104–1116, Mar. 2019.
[2] M. Jouhari, K. Ibrahimi, H. Tentmbe, and J. Ben-Othman, “Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things,” IEEE Access, vol. 7, pp. 96879–96899, Jul. 2019.
[3] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning. MIT Press Cambridge, 2016.
[4] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, “Deep learning for wireless physical layer: Opportunities and challenges,” China Commun., vol. 14, no. 11, pp. 92–111, Nov. 2017.
[5] C. Liu, J. Wang, X. Liu, and Y.-C. Liang, “Deep CM-CNN for spectrum sensing in cognitive radio,” IEEE J. Sel. Areas Commun., vol. 37, no. 10, pp. 2306–2321, Oct. 2019.
[6] C. Liu, Z. Wei, D. W. K. Ng, J. Yuan, and Y.-C. Liang, “Deep transfer learning for signal detection in ambient backscatter communications,” IEEE Trans. Wireless Commun., vol. 20, no. 3, pp. 1624–1638, Mar. 2021.
[7] X. Liu, C. Liu, Y. Li, B. Vucetic, and D. W. K. Ng, “Deep residual learning-assisted channel estimation in ambient backscatter communications,” IEEE Wireless Commun. Lett., vol. 10, no. 2, pp. 339–343, Feb. 2021.
[8] G. Jin, F. Liu, H. Wu, and Q. Song, “Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal,” J. Exp. Theor. Artif. Intell., vol. 32, no. 2, pp. 205–218, Aug. 2019.
[9] F. Liu, Q. Song, and G. Jin, “Expansion of restricted sample for underwater acoustic signal based on generative adversarial networks,” in Proc. Int. Conf. Graph. Image Proc. (ICGIP), vol. 11069, pp. 1106948–1106957, Chengdu, China, 2019.
[10] H. Yang, S. Shen, X. Yao, M. Sheng, and C. Wang, “Competitive deep-belief networks for underwater acoustic target recognition,” Sensors, vol. 18, no. 4, pp. 952–965, Mar. 2018.
[11] S. Shen, H. Yang, and M. Sheng, “Compression of a deep competitive network based on mutual information for underwater acoustic targets recognition,” Entropy, vol. 20, no. 4, pp. 243–256, Apr. 2018.
[12] J. Yan, H. Sun, H. Chen, N. U. R.Junejo, and E. Cheng, “Resonance-based
time-frequency manifold for feature extraction of ship-radiated noise,” Sensors,
vol. 18, no. 4, pp. 936–957, Mar. 2018.
[13] X. Ke, F. Yuan, and E. Cheng, “Underwater acoustic target recognition
based on supervised feature-separation algorithm,” Sensors, vol. 18,
no. 12, pp. 4318–4342, Dec. 2018.
[14] P. Zhu, J. Isacs, B. Fu, and S. Ferrari, “Deep learning feature extraction for
target recognition and classification in underwater sonar images,” in
Proc. IEEE Annu. Conf. Decis. Control. (CDC), pp. 2724–2731,
Melbourne, Australia, Dec. 2017.
[15] C. McQuay, F. Sattar, and P. F. Driessen, “Deep learning for hydrophone
big data,” in Proc. IEEE Pac. Rim Conf. Commun., Comput. Signal
Process. (PACRIM), pp. 1–6, Victoria, Canada, Aug. 2017.
[16] G. Hu, K. Wang, Y. Peng, M. Qiu, J. Shi, and L. Liu, “Deep learning
methods for underwater target feature extraction and recognition,”
Comput. Intell. Neurosci., vol. 2018, pp. 1–10, May 2018.
[17] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction
and functional architecture in the visual cortex,” J. Physiol., vol. 160,
no. 1, pp. 106–154, Jan. 1962.
[18] D. Santos-Dominguez, S. Torres-Guijarro, A. Cardenal-López, and
A. Pena-Gimenez, “underwater vessel noise database,” Appl. Acoust.,
vol. 113, pp. 64–69, Dec. 2016.
[19] W. Huang, H. Sun, Y. Liu, and W. Wang, “Feature extraction for rolling
element bearing faults using resonance sparse signal decomposition,”
Exp. Tech., vol. 41, no. 3, pp. 251–265, Feb. 2017.
[20] I. W. Selesnick, “Wavelet transform with tunable q,” IEEE Trans. Signal
Process., vol. 59, no. 8, pp. 3560–3575, Apr. 2011.
[21] J.-L. Starck, M. Elad, and D. L. Donoho, “Image decomposition via the
combination of sparse representations and a variational approach,” IEEE
Trans. Image Process., vol. 14, no. 10, pp. 1570–1582, Oct. 2005.
[22] K. F. Al-Raheem, A. Roy, K. Ramachandran, D. K. Harrison, and
S. Grainger, “Rolling element bearing faults diagnosis based on autocorrelation
of optimized: wavelet de-noising technique,” Int. J. Adv. Manuf.
Technol., vol. 40, no. 3–4, pp. 393–402, Jan. 2009.
[23] M. J. Shensa, “The discrete wavelet transform: wedding the a trous
and mallat algorithms,” IEEE Trans. Signal Process., vol. 40, no. 10,
pp. 2464–2482, Oct. 1992.
[24] H. Wang, J. Chen, and G. Dong, “Feature extraction of rolling bearing’s
early weak fault based on cemd and tunable q-factor wavelet transform,”
Mech. Syst. Signal Process., vol. 48, no. 1–2, pp. 103–119, Oct. 2014.
[25] J.-C. Di Martino, J. P. Haton, and A. Laporte, “Lofargram line tracking
by multistage decision process,” in Proc. IEEE Int. Conf. Acoust Speech
Signal Process. (ICASSP), vol. 1, pp. 317–320, Minneapolis, USA,
Apr. 1993.
[26] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, “Deep residual learning
for ship-radiated noise based on multimodal deep learning,” IEEE Trans.
Veh. Technol., vol. 69, no. 5, pp. 5307–5319, May 2020.
[27] J. Xie, C. Liu, Y.-C. Liang, and J. Fang, “Activity pattern aware spectrum
sensing: A CNN-based deep learning approach,” IEEE Commun. Lett.,
vol. 23, no. 6, pp. 1025–1028, Jun. 2019.
[28] J. Xie, J. Fang, C. Liu, and L. Yang, “Unsupervised deep spectrum
sensing: A variational auto-encoder based approach,” IEEE Trans. Veh.
Technol., vol. 69, no. 5, pp. 5307–5319, May 2020.
[29] J. Xie, J. Fang, C. Liu, and X. Li, “Deep learning-based spectrum sensing
in cognitive radio: A CNN-LSTM approach,” IEEE Commun. Lett.,
vol. 24, no. 10, pp. 2196–2200, Oct. 2020.
[30] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard,
W. E. Hubbard, and L. D. Jackel, “Handwritten digit recognition with a
back-propagation network;” in Proc. Adv. neural inf. proces. syst.,
pp. 396–404, Denver, USA, Nov. 1989.
[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,
V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions;” in
Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR),
pp. 1–9, Boston, USA, Jun. 2015.
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking
the inception architecture for computer vision,” in Proc. IEEE Comput.
Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2818–2826, Las
Vegas, USA, Jun. 2016.