IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 15, 2021, accepted April 20, 2021, date of publication April 23, 2021, date of current version May 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075344

An Underwater Acoustic Target Recognition Method Based on Combined Feature With Automatic Coding and Reconstruction

XINWEI LUO, (Member, IEEE), YULIN FENG, (Student Member, IEEE), AND MINGHONG ZHANG

Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing 210096, China

Corresponding author: Xinwei Luo (luoxinwei@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11674057, Grant 91938203, Grant 11704069, and Grant 11874109; and in part by the Fundamental Research Funds for Central Universities under Grant 2242021K30019.

ABSTRACT Underwater acoustic target recognition is one of the main functions of the SONAR systems. In this paper, a target recognition method based on combined features with automatic coding and reconstruction is proposed to classify ship radiated noise signals. In the existing underwater acoustic target recognition systems, the target category features are mostly constructed based on the power spectrum according to a certain presupposed model, and some useful information in the data is discarded artificially. In the proposed recognition method, a feature extractor based on auto-encoding is designed. The feature extractor uses the restricted Boltzmann machine (RBM) to automatically encode the combined data of the power spectrum and demodulation spectrum of ship radiated noise without supervision and extracts the deep data structure layer by layer to obtain the signal feature vector. The extracted feature vector is sent to a Back Propagation (BP) neural network to realize target recognition. Due to the high cost of ship radiated noise acquisition, the sample size of ship radiated noise signals is often hard to meet the needs of neural network training. A method of data augmentation is designed by RBM auto-encoder to construct the expanded sample set, which improves the performance of the recognition system. The experimental results based on the actual ship's radiated noise show that the proposed method has better performance than the traditional methods.

INDEX TERMS Underwater acoustic, target recognition, ATR, restricted Boltzmann machine, auto-encoding; data augmentation.

I. INTRODUCTION

Sound is the only known form of energy that can travel long distance underwater. The classification and recognition of underwater acoustic targets are of great significance in the monitoring of ships at sea, the search for underwater targets, and maritime law enforcement, etc. Using the underwater acoustic signals received by hydrophones (hydrophone array), the underwater acoustic target recognition system analyzes the characteristics of underwater targets and distinguishes the types of targets by signal processing methods [1]–[3]. Underwater acoustic target recognition based on ship radiated noise is a research hotspot in the field of underwater acoustic signal processing. Scholars have carried out long-term and in-depth research on target recognition methods based on ship radiated noise, and the research mainly focuses on two directions: feature extraction and pattern recognition.

Feature extraction is a process of removing redundant information from original data to achieve dimensionality reduction. The mechanism of ship radiated noise is quite complicated. The engine, propeller, water pump, oil pump, and other sound sources excite the hull in water to radiate noise. Although the composition of ship radiated noise is complex and varied, it still contains a lot of useful information. The ship radiated noise signals sampled by sensors are complex and random, so some transformation methods are needed to describe these data. The power spectrum is an effective method to represent underwater acoustic signals due to its short-term stationary characteristics.

The associate editor coordinating the review of this manuscript and approving it for publication was Emrecan Demirors.

At present, most of the target recognition features of underwater acoustic signals are based on a power spectrum or spectrogram. Reference [4] incorporated spectral and wavelet domain information with different resolutions and introduced an underwater target classification framework based on these features. Reference [5] designed a deep learning recognition method based on time-domain data and spectrogram to classify civil ships, large ships, and ferries. According to [6], the Mel-frequency cepstrum coefficient (MFCC) spectrum is obtained by Mel frequency transformation of target noise, which is used as the feature of underwater acoustic targets. In these recognition methods, the target of feature extraction mainly focuses on the power spectrum or the spectrogram, but there are few types of research on feature extraction and recognition based on the demodulation spectrum.

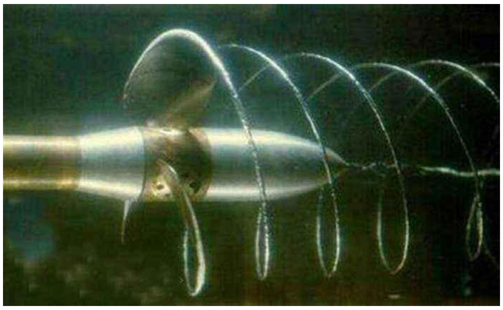


FIGURE 1. **Cavitation produced by rotating propellers in the water.**

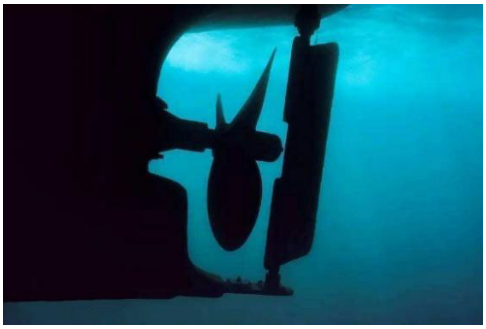


FIGURE 2. **Structure near the ship propeller.**

In fact, the radiated noise from ship propellers often has amplitude modulation-like characteristics that are a function of the shaft rotational rate and the number of blades. This phenomenon is caused by the periodic generation and collapse of cavitation in propeller blades under the condition of uneven pressure and wake inflow [7]. Figure 1 shows the cavitation produced by rotating propellers in the water and Figure 2 shows the structure near the ship propeller. The rhythm characteristic of ship radiated noise is an important feature of ship identification, and the rhythm feature is also the main basis for sonar operators to classify and recognize targets. This rhythmic feature is caused by propeller noise. There are abundant periodic modulation components in propeller radiated noise, which reflect the information of propeller type, propeller number, blade number, rotating speed, cavitation degree, and uneven flow field distribution. Demodulation analysis is an effective method to analyze the characteristics of propeller radiated noise. The rhythmic features contained in the demodulation spectrum are of great help to ship classification. Compared with the target recognition based on power spectrum only, the joint use of demodulation spectrum for target recognition can fully retain the information in the signal.

In theory, the original power spectrum and demodulation spectrum data can be directly used as the input of the recognition system for pattern recognition. However, the amount of original spectrum data is still large, which brings great pressure to the recognition system. To reduce the complexity of classification, spectral data need to be further compressed to reduce the data dimension. The commonly used data compression methods include Karhuner-Loeve Transform (KLT), principal component analysis, and autoregressive model, etc [8], [9]. Because of the good recognition ability of the human auditory system for ship radiated noise, the recognition method based on hearing models has also been deeply studied. To simulate the recognition process of the human auditory system, the Mel band-pass filter bank is used to simulate the decomposition of the cochlear acoustic signal, Discrete Cosine Transform (DCT) is used to simulate the energy conversion caused by hair cell vibration, and pattern recognition method is used to analyze the acoustic spectrum characteristics. Signal recognition based on human auditory acoustics has been successfully applied in the field of speech recognition, which includes preprocessing, framing, Mel-frequency cepstrum coefficients (MFCC)/ Gammatone frequency cepstral coefficients (GFCC) feature extraction, hidden Markov model (HMM) pattern matching, and so on. Compared with the autoregressive model, acoustic target recognition using auditory features has better performance [10]. The above feature extraction is a process of filtering the original information based on a subjective set model to reduce the amount of information [11]. When the data features are inconsistent with the subjective assumptions, the performance of the obtained features will be greatly reduced.

The adaptive feature extraction based on deep neural network (DNN) has better generalization performance than traditional feature extraction and has fewer requirements on SNR and distribution of samples. Therefore, adaptive feature extraction is quite suitable for underwater acoustic target recognition [12]. Adaptive feature extraction algorithms based on DNN include convolution neural network [13], generative countermeasure network [14], and deep Boltzmann machine (DBM) [15]. DBM has the following characteristics suitable for adaptive feature extraction. The DBM is a stack of restricted Boltzmann machines (RBM), so it can complete feature extraction self-supervised. DBM is a multi-layer structure, the number of cells in each layer can be adjusted freely, so it can adapt to different types of input. DBM completes the data auto-encoding based on the data distribution characteristics of the whole training set. This multi-layer structure can extract the high-level probability features of the data and has strong generalization ability. Research shows that the performance of adaptive feature extraction is better than that of feature extraction.

Using DBM, the optimal reconstruction of input data based on probability distribution is obtained by self-supervised optimization of parameters. RBM auto-encoder model is widely used in speech, image, and other signal processing fields. In underwater acoustic target recognition, RBM auto-encoder can effectively reduce the dimension of the original spectrum data and extract the high-level data distribution characteristics of the original data [16], which is helpful for subsequent pattern recognition.

The pattern recognition algorithm divides samples into several categories according to their characteristics. Traditional pattern recognition algorithms include linear discriminant analysis (LDA), support vector machine (SVM) [17], Gaussian mixture model (GMM) [18], etc. In recent years, the neural network method has also been widely used in pattern recognition of underwater acoustic signals [16], [19].

Traditional unsupervised clustering methods such as K-means and Gaussian mixture model (GMM) have many limitations in practical application. First of all, the number of clusters must be given when clustering, which is difficult to achieve in practical applications. Secondly, it is assumed that sample characteristics obey specific distribution, and GMM requires samples to obey Gaussian mixed distribution. However, actual samples are often difficult to fit Gaussian distribution, which will lead to distribution mismatch and reduce clustering effect [16]. In the case of Marine environmental noise, the signal-to-noise ratio of underwater acoustic signals is relatively low, so the clustering model needs to be modified according to the actual environment to achieve a better classification effect. DNN can improve the shortcomings of traditional pattern recognition methods. DNN is a model with a complex structure, which can fit arbitrary distribution samples. DNN has more adjustable parameters, which can effectively use large-scale samples to improve the generalization ability and recognition accuracy of the recognition system. Referring to the network structure of DNN, Reference [20] extracts the features of underwater acoustic signals based on RBM auto-encoder and uses BP classifier to obtain better recognition results than traditional recognition methods. However, the method in [20] only takes a short-time power spectrum as the feature input and ignores the unique rhythmical characteristics of ship radiated noise. Besides, when the number of samples is small, the recognition effect of this method will decrease obviously.

DNN often uses a large amount of data to carry out network training, but massive training samples can not always be obtained in actual measurement. Especially for demodulation spectrum features, it is a process of slow change over time, which requires the sample data to reach a certain length of time (second level) so that the features can be extracted effectively. However, the training sample resources of ship-radiated noise are not abundant, which leads to the problem of insufficient training samples. In order to solve this problem, the method adopted in this paper is data augmentation, which is one of the most commonly used techniques in deep learning. Data augmentation is mainly used to expand the training data set, make the data set as diverse as possible, and make the training model have stronger generalization ability. Therefore, in practical application, data augmentation becomes an important part of the preprocessing of model training.

In this paper, an auto-encoder based on the Boltzmann machine is constructed to extract the adaptive features of the power spectrum and demodulation spectrum data of underwater acoustic signals. After the greedy pre-training layer by layer, the Markov chain Monte Carlo method is adopted for overall optimization. Based on the BP neural network, a classification system of underwater acoustic targets is designed. To solve the problem of an insufficient sample size of underwater acoustic targets during network training, a data augmentation method is designed.

In Section 2, the power spectrum and demodulation spectrum calculation methods of ship radiated noise are analyzed, and the method of acquiring signal characteristics based on RBM is extracted. In Section 3, a classifier combining RBM auto-encoder and BP is proposed, and a method of data augmentation is designed to improve the performance of the classifier. In Section 4, the performance of the proposed underwater acoustic target recognition method is analyzed with the ShipsEar database. Section 5 summarizes the article.

II. FEATURE EXTRACTION BASED ON UNDERWATER ACOUSTIC SIGNAL SPECTRUM

Ship radiated noise signal is a random signal. During the navigation, the propeller, rotary or reciprocating machinery, various pumps, and other sound sources stimulate the hull to radiate sound into the water. According to different excitation sources, ship radiated noise can be regarded as a combination of mechanical noise, propeller noise, and hydrodynamic noise [21].

The ship radiated noise has the characteristics of approximately stationary in a short time. Through time-frequency transformation, a complex ship radiated noise signal can be converted into a more regular power spectrum, which can more effectively describe the nature of the noise. The power spectrum of the signal contains the characteristics of ship vibration and navigation state. The power spectrum has been widely used in feature construction of underwater acoustic target recognition, including Low-Frequency Analysis Recording (LOFAR), MFCC, GFCC, and so on. However, the envelope fluctuation information of ship radiated noise caused by propeller modulation is not included in the power spectrum. It is necessary to consider the modulation spectrum in feature construction.

In the construction of target category features, the common features are the line spectrum of the power spectrum, the shape of the power spectrum, line spectrum of demodulation spectrum, MFCC, GFCC, etc. However, these features are extracted based on the pre-set model and need prior information in practical application. When the environment changes, the classification performance of these features will be significantly reduced.

This section introduces the method of spectrum analysis of ship radiated noise signal, and proposes an improved deep Boltzmann machine for automatic coding feature extraction of the spectrum of ship radiated noise signal.

A. SPECTRAL ANALYSIS OF UNDERWATER ACOUSTIC SIGNAL

There are two kinds of typical spectrum analysis for ship radiated noise signals. One is power spectrum analysis, which is used to analyze the energy distribution of noise signals in the frequency domain. The other is demodulation spectrum analysis, which is used to analyze the periodic characteristics of the envelope in the noise signal.

To reduce the random fluctuation of the power spectrum estimation and reduce the energy leakage of each frequency component, windowing and averaging are carried out based on the periodogram. At this time, $x(n)$ is divided into overlapping K-sections, each signal length is M, and the power spectrum $S(k)$ of the signal can be expressed as

$$S(k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-j2\pi\frac{n}{N}k}\right|^2 \tag{1}$$

where k is the frequency index.

To reduce the random fluctuation of the power spectrum estimation and reduce the energy leakage of each frequency component, windowing and averaging are carried out based on the periodogram. At this time, $x(n)$ is divided into overlapping k-sections, each signal length is m, and the power spectrum $S(k)$ of the signal can be expressed as

$$S(k) = \frac{1}{KMU}\sum_{i=0}^{K-1}\left|\sum_{n=iL}^{iL+M-1} x(n)w(n-iL)e^{-j2\pi\frac{(n-iL)}{M}k}\right|^2 \tag{2}$$

where K is the number of data sections, M is the length of each data section, L is the step length of adjacent data sections, $w(n)$ is a window function, and U is the coefficient participating in amplitude normalization. Here $U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n)$.

Detection of envelope modulation on noise (DEMON) is a method for demodulation and analysis of ship radiated noise signals. In the DEMON method, the envelope component of the signal is obtained firstly, and then the envelope is analyzed by spectrum to extract the line spectrum features of shaft frequency, blade frequency, and harmonic frequencies, which are important information for target detection and classification. The block diagram of the DEMON process is shown in Figure 3.

The main process of DEMON processing is divided into three steps. Firstly, the frequency band of cavitation noise is estimated, and then the signal is filtered by using this frequency band. Then the envelope of the signal is extracted. Finally, the envelope signal is Fourier transformed to obtain the DEMON spectrum. When square law detection is used, the calculation formula of the demodulation spectrum $D(k)$ is as follows.

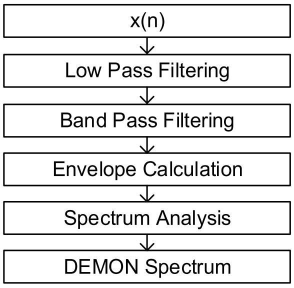


FIGURE 3. Block diagram of DEMON process.

$$D(k) = \left|\sum_{n=0}^{N-1}[x(n)^* h_L(n)^* h(n)]^2 w(n)e^{-j2\pi\frac{n}{N}k}\right|^2 \tag{3}$$

where * is the linear convolution operator, $h_L(n)$ is a low-pass filter function, $h(n)$ is a band-pass filter function, and $w(n)$ is a window function.

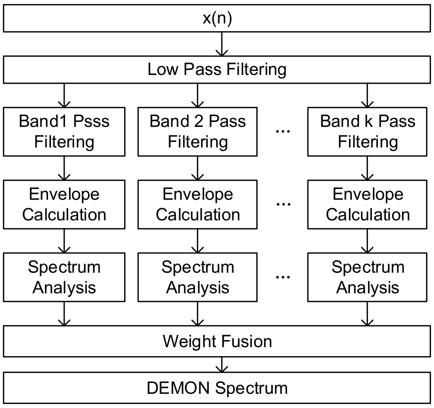


FIGURE 4. Block diagram of multi-band DEMON process.

According to [22], [23], the spectrum of cavitation noise of a propeller is uneven, and the theoretical maximum frequency is inversely proportional to the diameter of the cavitation bubble in water. The energy distribution and modulation degree of the wideband modulated signal of the actual ship noise are uneven. Non-uniform modulation will lead to the degradation of the quality of the demodulation spectrum obtained by the direct DEMON process, so single-band DEMON is difficult to adapt to the feature extraction of non-uniform modulation signal in the actual target signal. Therefore, the actual DEMON processing uses multi sub-band demodulation to obtain the modulation spectrum of each sub-band, and then comprehensively process to obtain the modulation characteristics. The block diagram of a multi sub-band DEMON is shown in Figure 4.

At this time, the calculation formula of demodulation spectrum $D(k)$ is as follows.

$$D(k) = \sum_{i=1}^{K} \alpha_i \left| \sum_{n=0}^{N-1} [x(n)^* h_L(n)^* h_i(n)]^2 w(n) e^{-j2\pi \frac{n}{N} k} \right|^2 \tag{4}$$

where * is the convolution operator, $h_L(n)$ is a low-pass filter function, $h_i(n)$ is i-th band-pass filter of K pre-set frequency band filters. $w(n)$ is a window function. α_i is the weighting coefficient of the demodulation spectrum in each frequency band. In engineering applications, we usually set α_i as follows [24].

$$\alpha_i = \frac{\sigma_i^2}{\sum\limits_{j=1}^{K} \sigma_j^2} \tag{5}$$

where σ_j^2 is the variance of the demodulation spectrum of j-th frequency band.

Reference [25] points out that ship radiated noise signal has the property of cyclostationarity, so the method of cyclic modulation spectrum (CMS) analysis is suitable for the modulation analysis of ship radiated noise. CMS is an alternative method of multiband summation, which can detect modulation frequency and carrier frequency simultaneously. When using CMS for DEMON analysis, it can achieve higher resolution in modulation frequency and carrier frequency, and significantly reduce the calculation time. The block diagram of DEMON based on CMS is shown in Figure 5.

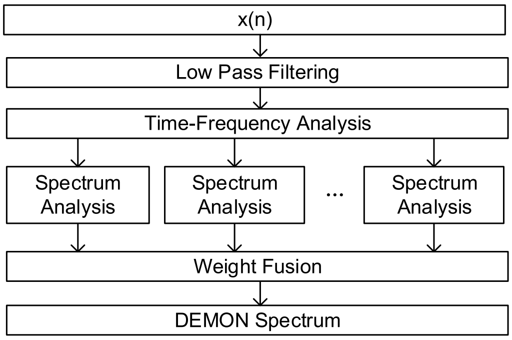


FIGURE 5. Block diagram of multi-band DEMON process based on T-F Analysis.

In DEMON based on CMS, the calculation formula of demodulation spectrum $D(k)$ is as formula (6) and (7).

$$X(p, q) = \frac{1}{M} \left| \sum_{n=pL}^{pL+M-1} [x(n)^* h_L(n)] w_1(n - pL) e^{-j2\pi \frac{n-pL}{M} q} \right|^2 \tag{6}$$

$$D(k) = \sum_{q=0}^{Q-1} \alpha_q \left| \sum_{p=0}^{\mathrm{P}-1} X(p, q) w_2(n) e^{-j2\pi \frac{p}{\mathrm{P}} k} \right|^2 \tag{7}$$

where $X(p, q)$ is the spectrogram of $x(n)$. p is the time index and $p = 0, 1, \ldots, \mathrm{P} - 1$. P is the number of data sections. q is the frequency index and $q = 0, 1, \ldots, \mathrm{Q} - 1$. Q is the number of frequency points to be analyzed. M is the length of each data section, L is the step length of adjacent data sections, $w_1(n)$ and $w_2(n)$ are window functions. α_q is the weighting factor of frequency band q. Similar to multi-band DEMON process, α_q can be set according to the variances of DEMON spectrum of each frequency band.

Due to the characteristics of high operation efficiency and high resolution of carrier and modulation frequency, the demodulated spectrum of the target signal is obtained by using the method of DEMON based on CMS in this paper. Figure 6 and Figure 7 show the photos of two ships (a passenger and a ro-ro) and their radiated noise signal waveform, power spectrum, and DEMON spectrum respectively.

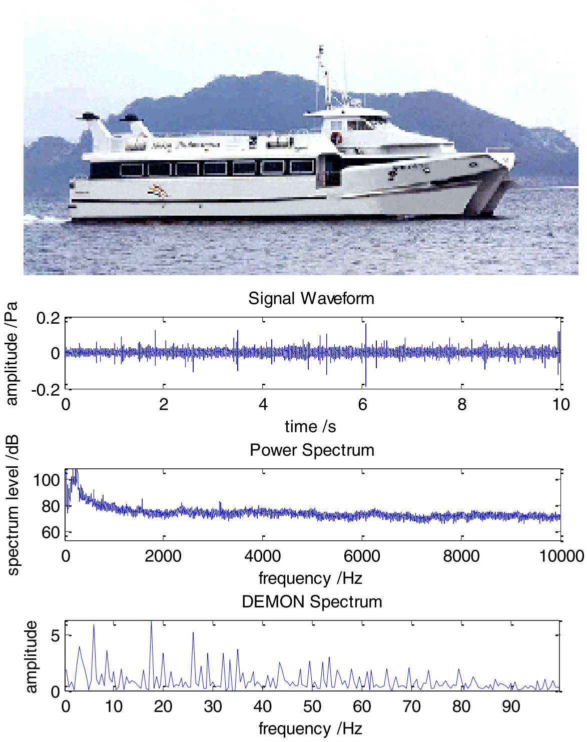


FIGURE 6. Photo of a passenger and the signal waveform, power spectrum and the DEMON spectrum of the passenger's radiated noise.

B. FEATURE EXTRACTION BASED ON RBM AUTO-ENCODER

The hand-craft feature extraction method selects the key components of the signal as features based on the Presupposed model, thus reducing the dimension of the original signal. The traditional feature extraction method obtains the power spectrum and DEMON spectrum by time-frequency conversion of the original signal, and obtains the key features of the signal manually, including line spectrum strength, power spectrum shape, modulation frequency, and modulation line structure. Figure 8 depicts this process. The traditional feature extraction method based on the analysis of ship noise effectively

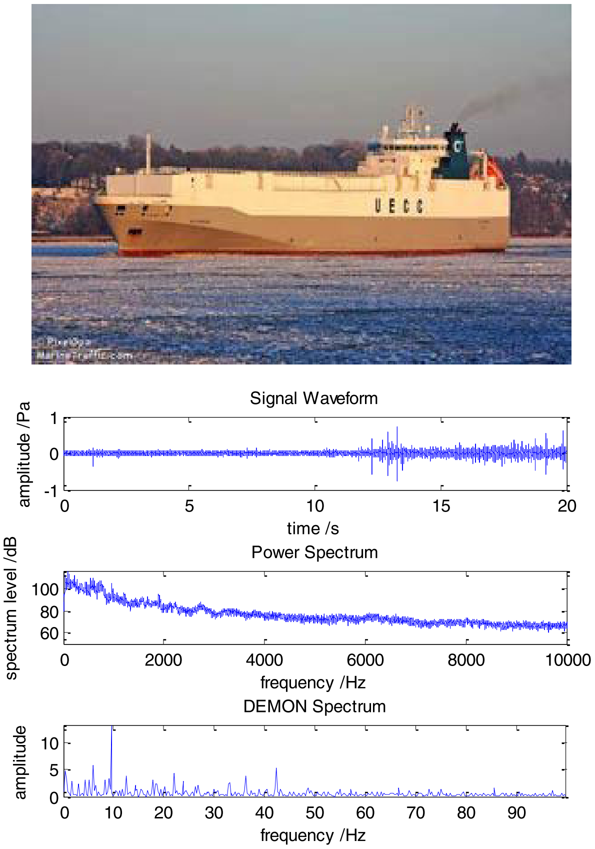


FIGURE 7. Photo of a ro-ro and the signal waveform, power spectrum and the DEMON spectrum of the ro-ro's radiated noise.

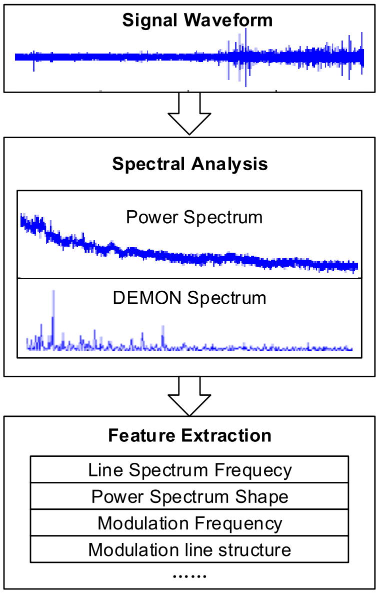


FIGURE 8. Traditional feature extraction process of underwater acoustic signal.

picks up the key information of underwater acoustic signal but ignores the role of sonar's hearing in underwater acoustic target recognition. GFCC uses several band-pass filters to simulate the frequency band effect of the human ear and simulates the process of the human ear transforming vibration into a neural signal through DCT transformation, and finally obtains the energy of the original signal in different frequency bands as characteristics. GFCC method uses an auditory simulation method to supplement some information abandoned by traditional feature extraction methods and achieves a higher recognition rate. Hand-craft features inevitably discard some useful information artificially. Adaptive feature extraction based on DNN can customize the feature extraction model according to the data distribution of the training set and has better generalization ability than hand-craft features. In order to obtain a better recognition effect for underwater acoustic targets, adaptive feature extraction based on RBM auto-encoder is introduced into the underwater acoustic target recognition system.

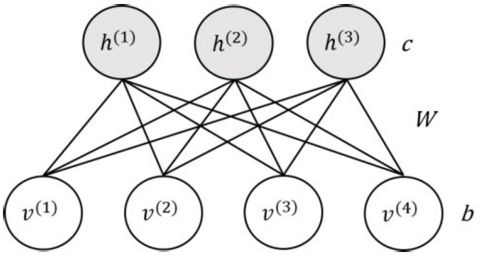


FIGURE 9. RBM structure.

Boltzmann machine (BM) is an algorithm that transforms the probability of data into energy and optimizes the model with the idea of annealing. The restricted Boltzmann machine (RBM) can be used as a self-monitoring encoder to reduce the dimension of the original spectrum. The structure of the restricted Boltzmann machine is shown in Figure 9. RBM is a two-layer model. The lower layer is visible layer units v, and the upper layer is hidden units h. The interlayer neurons are fully connected, the connection weight is W, and the neurons in the layer are not connected. b and c are the offset of the visible layer and the hidden layer respectively. The state of neurons in RBM is binary, and the state probability of the whole model is controlled by its energy. The energy and probability models of RBM are as follows.

$$E(v, h) = -b'v - c'h - h'Wv \tag{8}$$

$$P(x) = \sum_h p(x, h) = \sum_h \frac{e^{-E(x,h)}}{Z} \tag{9}$$

where b', c', h' are the transpose of b, c, h. v is the visible layer unit of the system. x is the visible layer unit corresponding to any hidden layer unit h. $E(v, h)$ is the joint energy function of the hidden layer and the visible layer. $E(x, h)$ is the joint energy function of any hidden layer and the corresponding visible layer. $P(x)$ is the likelihood function of visible layer x, and $p(x, h)$ is the joint probability distribution of x and h. $Z = \sum_x e^{-F(x)}$ is the partition function and $F(x) = -log \sum_h e^{-E(x,h)}$ is the free energy.

The probability distribution of the hidden layer of RBM can be determined by the network parameters and the states of visible layer units. The network parameters are randomly initialized and the input is used as the initial state of the visible units. The hidden layer state is calculated by the forward process, and the visible layer state is reconstructed by the reverse process. So the RBM parameter solution is actually the solution of Markov chain Monte Carlo (MCMC).

$$\begin{aligned} p(h_i = 1|v) &= sigm(c_i + W_i v) \\ p(v_j = 1|h) &= sigm(b_j + W_j' h) \end{aligned} \tag{10}$$

where $sigm(x)$ is sigmoid function. h_i is the i-th unit in hidden layer. v_j is the j-th unit in visible layer. W' represents the transpose of the connection weight matrix W. c_i is the offset of hidden layer unit h_i, b_j is the offset of visible layer unit v_j.

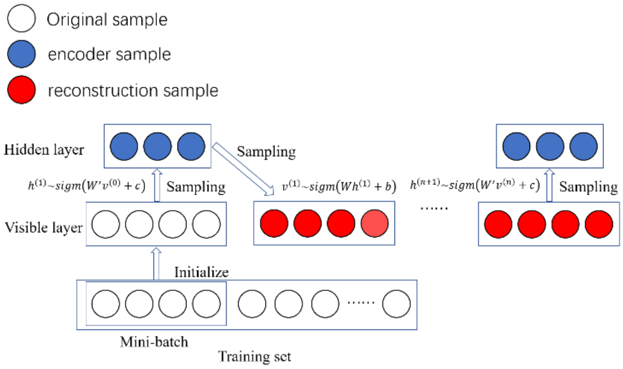


FIGURE 10. Contrast divergence algorithm.

The common RBM algorithm is the contrast divergence algorithm (CD), and its algorithm flow is shown in Figure 10. By encoding and reconstructing the subset of the original data set, the updating gradient of model parameters is obtained. The formula is as follows.

$$\begin{aligned} -\frac{\partial \log p(v)}{\partial W_{ij}} &= E_v[p(h_i|v) \cdot v_j] - v_j^{(i)} \cdot sigm(W_i \cdot v^{(i)} + c_i) \\ -\frac{\partial \log p(v)}{\partial c_i} &= E_v[p(h_i|v)] - sigm(W_i \cdot v^{(i)}) \\ -\frac{\partial \log p(v)}{\partial b_i} &= E_v[p(v_i|h)] - v_j^{(i)} \end{aligned} \tag{11}$$

where W_{ij} is the i-row j-column element in matrix W. W_i is the i-th row in the matrix W. $E_v[]$ is expectation of variables in brackets. $v^{(i)}$ is the i-th sample in data set. $v_j^{(i)}$ is the j-th unit of i-th sample. c_i is a constant.

In RBM training, Equation 8, 9 describes how the network fits the training sample probabilistic features through the relationship between parameters and neuron states. Equation 10 provides a general method for calculating probability in RBM. Equation 11 is used to update network parameters to reduce the difference between the reconstructed sample and the original sample, to improve the description of the overall probability characteristics of the sample set by the network.

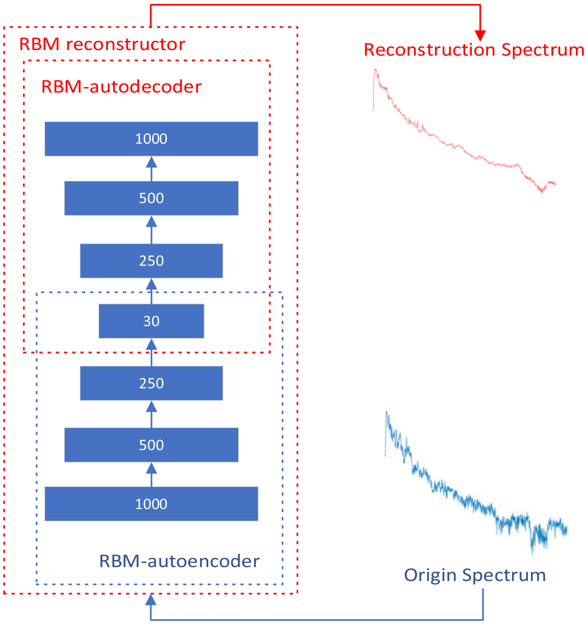


FIGURE 11. RBM reconstructor structure.

The single-layer RBM acts as an auto-encoder and takes the original data set as the self-monitoring training set to complete the probabilistic feature compression of the original spectrum input. The output of a single-layer RBM auto-encoder is the probability characteristic of the compression dimension. The full connection structure allows any input to affect any output, which is beneficial to the information utilization of the fusion spectrum. RBM can encode the input spectrum nonlinearly, which allows us to construct a stacked RBM auto-encoder to auto-encoder the original spectrum layer by layer. Its structure is shown in Figure 10. The internal data model of the multi-layer RBM structure is more complex. In multi-layer auto-encoding, the neural network learns more complex abstract probability features of the original spectrum, thus obtaining lower reconstruction error than single-layer RBM auto-encoder or principal component analysis (PCA). Figure 11 shows the structure of an RBM reconstructor. Numbers in the figure represent the number of neurons in each layer. In Figure 11, the auto-encoder part compresses the original spectrum layer by layer and obtains the best parameters of the model in the training. The parameters of the auto-decoder are symmetrical with the auto-encoder. The RBM reverse process is carried out on the RBM auto-encoding results layer by layer, and the reconstruction results of the original spectrum are obtained. The reconstructed spectrum is the result of the original spectrum under the effect of the overall probability characteristics of the training set. This process is actually a denoising process for a single sample, making a single sample more "gregarious".

The detailed implementation process of the RBM auto-encoder is shown in Algorithm 1.

Algorithm 1 RBM Auto-Encoder

```
Take input from original spectrum and normalization
Construct training set and initialize network parameters
The training set data is divided into multiple minipatches
for 1:layer
  for 1:epoch[i]
    for batch = 1:batch_num
      Implement the forward process based on Equation (10), and obtain the hidden unit state
      Implement the reconstruction process based on Equation (10), and obtain the visable unit state
      Update network parameters based on Equation (11)
    end
  end
end
Implement the forward process from input to the last hidden layer, and obtain the feature output
Implement the reconstruction process from feature output to visbale layer, and obtain the input reconstruction
```

When RBM is used to extract actual spectral features, the input structure of RBM should be consistent with the spectral structure, which requires the input dimension of RBM to be adjusted according to the fusion spectrum dimension. Besides, the energy of the underwater acoustic signal is concentrated in the low-frequency band, and the high-frequency components are often meaningless noise and clutter for classification. Therefore, according to the observation of the data, the signals with frequencies below 8 kHz are used for underwater acoustic target recognition. At the same time, compared with other acoustic signals (such as speech signals), the underwater acoustic signal has longer stability, and propeller noise has obvious periodicity. Therefore, each sample needs a longer duration, and the number of visible layer units of RBM varies with the spectral length at 2 seconds.

On the other hand, the hyperparameters of the model need to be adjusted manually to achieve good training and reconstruction performance. In order to extract the features of each hidden layer, the number of visible units is reduced. The learning rate needs to be considered comprehensively according to the convergence speed and reconstruction error of data training. A large learning rate will accelerate the convergence speed, but it will lead to the vibration of the reconstruction error. A small learning rate may fall into the local minimum value, but it has the opportunity to achieve a better learning effect. The capacity of the minibatch should be set flexibly according to the size of the data set. Reasonable minibatch can accelerate training and effectively use a larger training set.

By learning the structure of different tag data, DNN can identify its characteristics. DNN needs a large number of labeled samples to get a good training effect. In General, the smaller the sample of the training data set, the worse the performance of DNN. In underwater acoustic target recognition, it is generally considered that samples are difficult to obtain and there are few marked samples. In the construction of an underwater acoustic target recognition system, the number of labeled samples is often difficult to meet the requirements of DNN. In the classification process, a small number of samples will lead to overfitting.

Common methods to augment audio data include noise injection, time shift, dynamic range gain, and equalization. However, these methods have some limitations, namely the lack of diversity of samples, which cannot improve the classification effect in practical application [27]. This paper presents a method of data augmentation based on RBM auto-encoder. RBM auto-encoder can achieve nonlinear data compression and extract high-level probability distribution of data. The original data can be decoded by the multi-layer RBM structure symmetrical to the encoder. This decoder can reconstruct the original data through the high-level probability distribution characteristics of the original data, and obtain the results that conform to the overall characteristics of the sample. These reconstruction results are not the accurate restoration of original data, because the encoding and decoding process of RBM includes random sampling from data probability to corresponding neuron state. The reconstruction result of the decoder is the result of the probability characteristics of the sample population acting on the single sample restoration, which can weaken the uniqueness of a single sample and obtain new samples with statistical similarity.

RBM auto-encoder is used to reconstruct the samples, and the reconstructed data is used to increase the number of training data, which is conducive to preventing overfitting and facilitating the classification of small samples. The experiment shows that the target recognition accuracy can be improved by using the sample expansion training set reconstructed by the RBM auto-encoder.

The reconstructor was tested with four sets of data. Each dataset contains 600 samples. The samples in data set A are sinusoidal function with noise, the samples in data set B are higher frequency sinusoidal function with noise, the samples in data set C are cosine function with noise, and the samples in data set D are power spectrum data of the real signal. According to the method of algorithm 1, the 4-layer RBM reconstructor is trained. Figure 12 shows the input data and reconstruction data of the reconstructor.

TABLE 1. MSE of the reconstrction samples and orignal samples.

MSE	origin A	origin B	origin C	origin D
reconstruction A	0.2245	0.5607	0.5350	0.3672
reconstruction B	0.5377	0.2356	0.5472	0.4679
reconstruction C	0.5393	0.5649	0.2335	0.5196
reconstruction D	0.1759	0.2406	0.2144	0.0115

The mean square error (MSE) of the reconstruction samples and orignal samples are shown in Table 1.

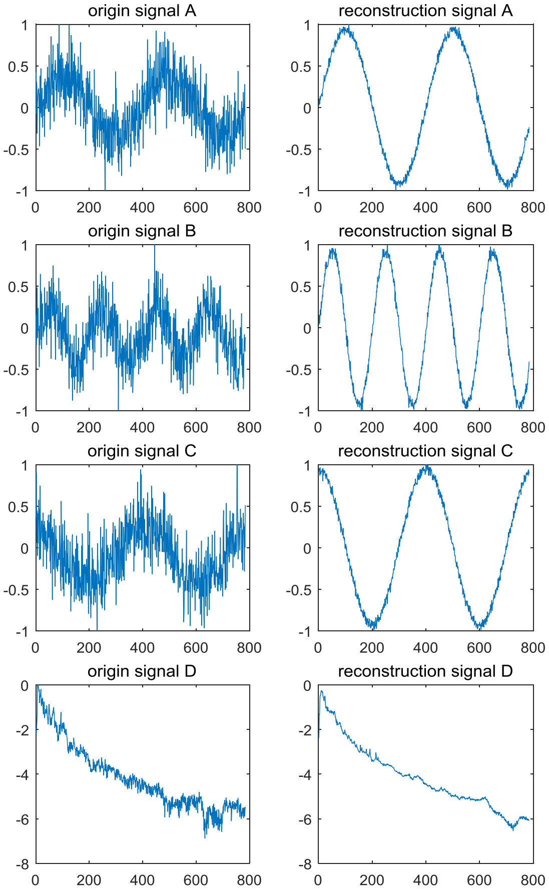


FIGURE 12. Origin data (left) and reconstructed data (right).

As can be seen from Figure 12 and Table 1, the reconstructed sample has good similarity with the original sample, and retains certain randomness.

III. TARGET RECOGNITION SYSTEM

This paper constructs an underwater acoustic target recognition system based on the RBM auto-encoder and BP neural network. Its structure is shown in Figure 13. In target recognition, the input of the model is the spectrum of the underwater acoustic signal, and the high-level features of data are extracted by layer by layer auto-coding of stacked RBM, and the target recognition is based on these features by BP neural network. BP neural network is the most common DNN, which is often used in the classification and recognition of complex signals. BP neural network is a multi-layer network composed of a large number of neurons and their connections. The output expectation of the network can be consistent with the actual situation by the gradient descent method.

The training process of the BP neural network is to use labeled samples to calculate the parameter update gradient by comparing the difference between the model output and the expected output, to calculate the connection weight matrix W and neuron threshold θ of each layer. Because of the deep layers of the neural network, the calculation of the gradient generally depends on the chain derivation of the network layer by layer. The training process can be divided into forwarding propagation and backpropagation. During forward propagation, the state of the back layer neurons is determined by the parameters of the front layer neurons and the network. In this process, the states of all hidden layer neurons are calculated, and the states of input neurons are given by samples.

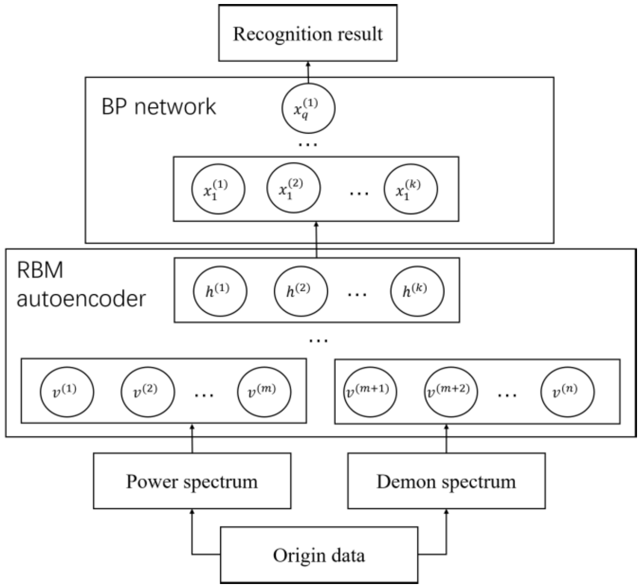


FIGURE 13. Deep clustering system structure.

$$x_j^{(p+1)} = f(\sum_{i=0}^{n-1} W_{ij}^{(p)} x_i^{(p)} - \theta_j^{(p)}), \quad j \in [0, n-1] \quad (12)$$

where $x_i^{(p)}$ is the ith neuron in the pth layer. $W_{ij}^{(p)}$ is the connection weight of the corresponding neuron in the current layer, θ_j is the threshold of the corresponding neuron, and $f(x)$ is the activation function.

In the process of back-propagation, the training parameters are gradient solved according to the neuron values and training errors in the model. The training errors can be measured by the loss function $E(W, \theta)$

$$E(W, \theta) = \sum_{i \in D} (t_i - y_i)^2 \quad (13)$$

where D is the training set, t_i is the label of the data, and y_i is the forward propagation output.

Using a chain derivative to calculate gradient can reduce the computational complexity and speed up the training of the BP neural network.

$$\begin{aligned} \frac{\partial E(W,\theta)}{\partial W} &= \frac{\partial E(W,\theta)}{\partial z^l} \frac{\partial z^l}{\partial W} \\ \frac{\partial E(W,\theta)}{\partial \theta} &= \frac{\partial E(W,\theta)}{\partial z^l} \frac{\partial z^l}{\partial \theta} \end{aligned} \quad (14)$$

where the calculation of $z^l = W^{(l)} \cdot \overrightarrow{x^{(l)}} - \theta^{(l)}$, $\frac{\partial E(W,\theta)}{\partial z^l}$ depends on activation function form.

BP neural network can adapt to different input scales by adjusting the number of neurons, and can also classify more complex problems by increasing the number of hidden layers. Compared with GMM, K-means, and other conventional underwater acoustic target classifiers, BP neural network is more flexible in structure and needs less manual intervention. At the same time, BP neural network has a complex structure, which is competent for the classification of a large number of data samples. BP neural network is an effective method to solve the complex classification problem, which can be used as an underwater acoustic target classifier. However, the low signal-to-noise ratio of the underwater acoustic signal and the few labeled samples affect the good classification results of the BP neural network.

When the number or distribution of input samples is not ideal, the model may have fitting problems, including overfitting and underfitting. As a high-capacity deep network, BP neural network can solve complex tasks, but when the number of samples is too small, the details and noise of samples are learned by the BP neural network, which will lead to overfitting of the model. Overfitting mainly shows that the performance gap between the training set and the test set is large. In order to avoid overfitting, the common method is to increase the diversity of samples. The RBM auto-encoder can reconstruct the original samples in a statistical sense, and the reconstructed samples conform to the probability characteristics of the sample set data. Using an RBM auto-encoder to reconstruct the training set can avoid overfitting.

BP neural network with more than 5 layers can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details [28], [29]. Therefore, we choose the 5-layer BP network structure. We set the dimension of the first hidden layer neuron of BP neural network to 500, which is close to the dimension of conventional underwater acoustic power spectrum data, and the dimension of the later hidden layer decreases on the basis of the former. We set the number of neurons in each layer as 50-500-200-50-1.

Because of the problem of overfitting of the BP neural network in a small training set, the system uses the reconstruction sample of the RBM auto-encoder to expand the training set, as shown in Figure 14. The sample reconstructed by the RBM auto-encoder is a new sample with the overall probability characteristics of the data set and achieves the effect of denoising the original data. Firstly, the original spectrum is randomly divided into a training set and test set according to 6:1 random sampling. Then the RBM auto-encoder is run on the training set to get the appropriate RBM parameters and the reconstruction samples of the training set. Then, the BP neural network is trained by using the neural network training set. Finally, the underwater acoustic target recognition results are obtained by the RBM auto-encoder and BP neural network.

In a word, the RBM auto-encoder BP neural network system has the following advantages. RBM auto-encoder reduces the dimension of the original spectrum and extracts high-level probability features layer by layer, which increases the separability of data, which is conducive to BP neural network classification. RBM auto-encoder can reconstruct and denoise the training samples, and reduce the overfitting probability of the BP neural network by adding additional samples. Both BP neural network and RBM auto-encoder are self-learning algorithms and do not need to design too many parameters, which is conducive to the application of an underwater acoustic recognition system in complex situations.

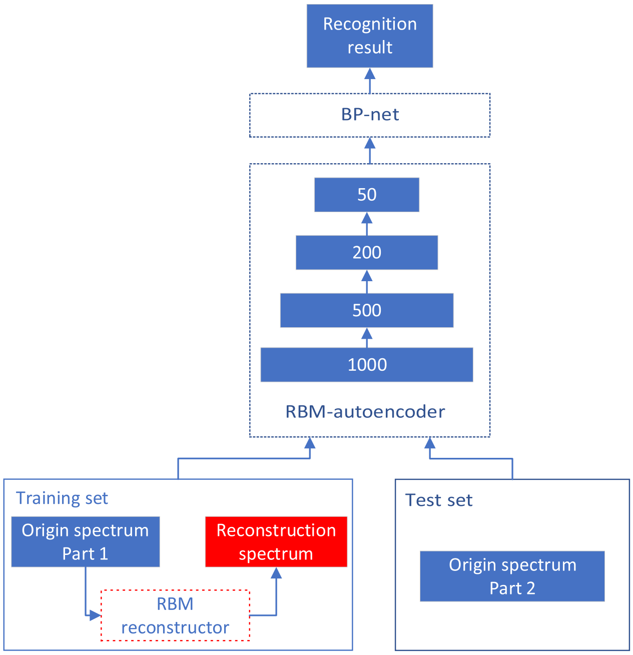


FIGURE 14. The structure of the proposed target recognition system.

IV. EXPERIMENT

The experimental process is divided into four stages. Firstly, the original underwater acoustic time-domain data are preprocessed in different ways, and then aligned and spliced in the time axis, to obtain the original input samples. In the second step, the original input samples corresponding to the training set are sent to the RBM auto-encoder, and the reconstructed samples are obtained by auto-encoding and decoding. In the third step, the original input samples are randomly divided into a training set and test set according to 6:1, and the reconstructed samples are added to the training set. These samples are sent to the RBM auto-encoder to complete feature extraction. Finally, the dataset was used to train and test the underwater acoustic target recognition system to evaluate the performance of different systems.

The number of layers of the RBM auto-encoder was set to 4, and the number of units in each layer was 985, 500, 200, and 50. The learning rate during training was set to 0.001. The value of the weight matrix was initialized randomly at $[-0.001, 0.001]$, the offset value was initialized to 0, and the number of iterations for each layer was set to 100. The Boltzmann machine took the normalized spectrum of the segmented signal as input and completed the feature extraction through the four-layer RBM auto-encoder. BP neural network obtained the output features of the RBM auto-encoder to complete target recognition.

To verify the performance improvement of the proposed method in improving the recognition rate and reducing overfitting, we arranged a control experiment to use the training set without reconstruction samples to train the network. Besides, to verify the performance of the feature fusion algorithm in increasing the separability of the data, we arrange a comparative experiment to use a separate algorithm to preprocess the time-domain underwater acoustic data.

A. ACOUSTIC SIGNAL PREPROCESS

The data samples processed in this study are single-channel audio signals. The signal was divided into 2 seconds frames with a 50% overlap between frames, and a Hanning window was added to the signal to suppress high-frequency interference and energy leakage.

By calculating the power spectrum and DEMON spectrum of the windowed signal, the time-domain signal is converted to the frequency domain. The power spectrum provides the frequency domain energy distribution of the signal, and the DEMON spectrum provides the analysis of the ship's modulation noise. The subsequent target recognition system uses this information to classify and recognize ships.

The normalized power spectrum $\hat{S}(k)$ and the normalized DEMON spectrum $\hat{D}(k)$ are taken as the system input. The normalized spectrum is calculated as follows.

$$\begin{aligned} \hat{S}(k) &= \frac{S(k) - \min[S(k)]}{\max[S(k)] - \min[S(k)]} \\ \hat{D}(k) &= \frac{D(k) - \min(D(k))}{\max[D(k)] - \min[D(k)]} \end{aligned} \tag{15}$$

where $S(k)$ and $D(k)$ are given by formula (2) and (6) respectively.

In order to effectively utilize the power spectrum information and DEMON spectrum information to maximize the recognition performance, it is necessary to fuse the two kinds of spectra to correspond to the original time-domain information and normalize the two kinds of spectra respectively, to make the effective amplitude of the spectrum consistent. Finally, the normalized fusion spectrum was obtained.

$$\hat{X}(k) = [\hat{S}(k), \hat{D}(k)] \tag{16}$$

At this time, $0 \leq \hat{X}(k) \leq 1$. $\hat{X}(k)$ meets the requirements of RBM for input data.

B. EXPERIMENT RESULT

In order to verify the performance of the proposed underwater acoustic target recognition system, the ShipsEar database is used in multi-target classification experiments. ShipsEar database contains underwater noise records generated by a variety of ships, as well as details of each record: ship type, hydrophone gain depth (H_G_D), Real-time weather, etc.

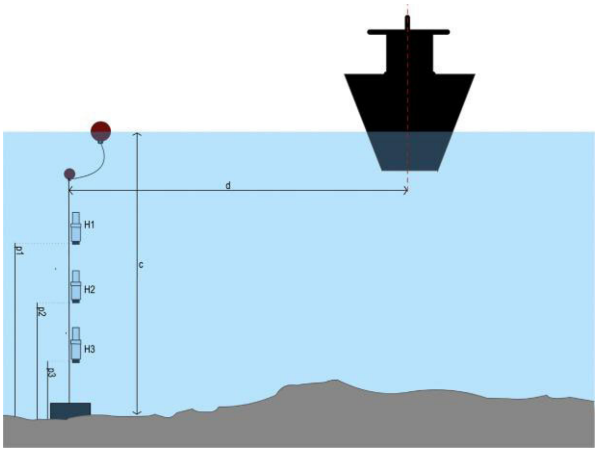


FIGURE 15. ShipsEar data collection diagram.

TABLE 2. Data classification in ShipsEar.

Category	Type of Vessel
Class A	fishing boats, trawlers, mussel boats, tugboats and dredgers
Class B	motorboats, pilot boats and sailboats
Class C	passenger, ferries
Class D	ocean liners and ro-ro vessels
Class E	background noise recordings

Figure 15 is a schematic diagram of the database acquisition. Three hydrophones are arranged vertically under the water to record the noise passing by the ship and conduct beamforming to ensure the maximum dynamic range. In very shallow areas (depths under 10 m), recordings were made with one or two hydrophones [26]. The amplifier used a 100 Hz high-pass filter to suppress marine background noise, the hydrophone sampling rate is 52,734 Hz, and the AD converter bit depth is 24 bits. According to the size of the ship, samples were divided into five categories, as shown in Table 2.

After preprocessing the underwater acoustic signal, original samples are obtained. There are 600 samples for each type of vessel, with a total of 3,000 samples. Among them, 6/7 served as training samples, and the rest served as test samples. Then, the RBM auto-encoder is used to encode and decode the training samples to obtain the reconstruction samples, and the reconstructed samples are added to the training set. In this experiment, the expanded sample set contains 6,000 samples. Then the training set is used to train the underwater acoustic target recognition system. Finally, the training system is used to test the multi-objective classification of the test set.

In the experiment, the adjusted Rand index (ARI) and the maximum value of the clustering rate (MVCR) were used to quantify the effect of target classification. ARI is used to quantifying the distribution similarity between the test set classification results and the actual tags, and MVCR is used to quantify the clustering accuracy of a single class.

Assuming that the dataset of N classes is clustered into K clusters, ARI is defined as

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{2ab}{N(N-1)}}{\frac{1}{2}(a+b) - \frac{2ab}{N(N-1)}} \quad (17)$$

where n_{ij} is defined as the number of samples that should belong to the ith clustered into the jth cluster. Higher ARI value means better clustering results.

MVCR analyzes the clustering effect of each class, which is defined as

$$MVCR(i) = \frac{\max(n_{ij}, j = 1, 2, \ldots, K, j \notin \Theta_i)}{\sum_j n_{ij}} \quad (18)$$

where Θ_i is the set of label that has been previously selected. n_{ij} is defined as the number of samples that should belong to the ith clustered into the jth cluster.

In model design, the number of model layers and other super parameters affect the performance of the underwater acoustic target recognition system. The rationality of the selected hyperparameters can be verified by experiment and theoretical analysis. RBM auto-encoders usually use a three-hidden layer model to complete feature extraction, but there is no appropriate hyperparameter setting specification for data reconstruction using it. Therefore, this paper verifies the rationality of using independent hyperparameters in the data reconstruction system by designing a controlled experiment.

The RBM auto-encoder used for feature extraction uses three hidden layers with the number of neurons of 500, 200, and 50 respectively. This RBM auto-encoder can also be used for sample reconstruction at the same time. However, considering the quality of feature reconstruction samples and the independence from feature extraction, RBM auto-encoders of four hidden layers are used in this paper for sample reconstruction, and the number of neurons is 1000, 500, 250, 30 respectively. To verify the effectiveness of the reconstructed samples to the recognition system, this paper uses power spectrum data to reconstruct the samples of three-layer RBM and four-layer RBM respectively, and tests are carried out according to the structure shown in Fig. 13. The results are shown in Table 3.

TABLE 3. Experiment result of different level models.

Hidden layers	PWD accuracy	Reconstruction MSE
3	86.4%	1.533
4	87.0%	1.469

Table 3 shows the classification effect of the different level models after the reconstruction of the same data. It can be seen that the recognition rate and reconstruction mean square error obtained by using the four-hidden layer model are better than that of the three-hidden layer model. It shows that it is reasonable to choose an independent RBM reconstructor with 4 hidden layers in this paper.

To evaluate the spectrum types and the influence of spectrum reconstruction on the performance of underwater acoustic target recognition systems, this paper constructed a training set using the different spectrum of ShipsEar data set to test the above underwater acoustic target recognition system, to find the optimal performance of the spectrum input. The experimental results are shown in Table 4.

TABLE 4. Experiment result of different feature model.

Feature	Accuracy	ARI	Max MVCRs	Min MVCRs
Power Spectrum	85.6%	67.8%	97.9%	70.9%
DEMON Spectrum	57.2%	25.8%	80.8%	41.3%
Fusion Spectrum	91.4%	79.1%	**99.0%**	78.8%
Power spectrum*	87.0%	71.3%	98.0%	76.1%
DEMON Spectrum*	59.8%	27.3%	81.0%	51.1%
Fusion Spectrum*	**92.6%**	**82.4%**	**99.0%**	**84.1%**

* represents the experiment result of the model which is trained by expended sample set after data augmentation.

Bold data are the result of achieving the optimal performance of the test in various methods.

Table 4 shows that the classification result of the model which is trained by different features based on ShipsEar database. The recognition rate of the fusion spectrum was higher than that of the single spectrum, which shows that the fusion spectrum experiment has higher ARI and MVCR. The experiment of augmenting the training set with reconstructed data also verified the inhibitory effect of the proposed method on overfitting, and the original spectral results were improved slightly.

TABLE 5. Result of control experiment.

Method	Accuracy	Max MVCRs	Min MVCRs
GFCC+GMM	75.4%	100%	55.5%
Proposed method	92.6%	99.0%	84.1%

GFCC + GMM is the method adopted in [26]. Table 5 shows the performance comparison of the two classification methods. The proposed method is significantly superior to the GFCC + GMM method in terms of accuracy and Min MVCRs.

The method proposed in this paper is a cascading form, which has a low coupling degree between feature extraction module and classifier module and has good flexibility. Similarly, the proposed method can be extended to an end-to-end classification system. Figure 16 shows several extensions of the proposed method and a case where an end-to-end classifier is connected. Table 6 shows the results of direct classification using DBN and classification using the structure in Figure 16.

The classification performance of DBN on the original samples is obviously lower than that of the proposed method. By expanding the sample with RBM reconstructor, the classification performance of DBN has been improved, which is close to the performance of the proposed method. However, the proposed method is a cascading network with a low

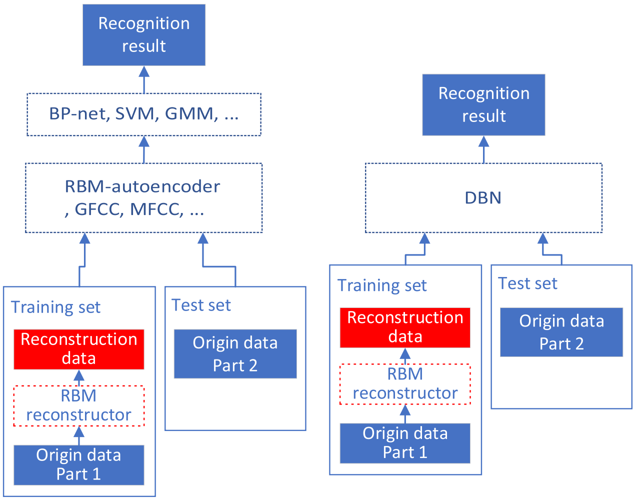


FIGURE 16. Extensibility schematic of the proposed method.

TABLE 6. Experiment result of dbn.

Method	Accuracy	Max MVCRs	Min MVCRs
DBN	89.3%	98.0%	81.5%
DBN*	92.7%	99.0%	82.3%

* represents the experiment result of the model which is trained by expended sample set after data augmentation.

TABLE 7. Training time.

Process	Preprocess	Feature extraction	Data expansion	BP network
Time (s)	56	27.8	473	1107

coupling degree between the feature extraction module and classifier module, and it has better openness than the DBN network.

Table 7 shows the training time of the proposed method. The proposed method is tested on a workstation with an 8-core CPU (I7 9700K) and 16GB RAM. The method is simulated by MATLAB software. The result shows that the training time is acceptable.

The results show that: (1) The features extracted by RBM auto-encoding have better classification performance than those extracted by traditional methods. (2) Under the condition of comprehensive utilization of power spectrum and demodulation spectrum information, the performance of the recognition system has been significantly improved. (3) The performance of the recognition system is further improved after the data augmentation processing based on the RBM auto-encoder.

V. CONCLUSION

This paper proposes an underwater acoustic target recognition method based on RBM auto-encoder and BP neural network. The proposed method has the following characteristics. (1) The power spectrum and demodulation spectrum of ship radiated noise signals are used as the input of feature extractor, which avoids the loss of important rhythm characteristics of ship radiated noise when only using the power spectrum in the traditional recognition methods. (2) The RBM autoencoder is used for unsupervised feature extraction of the combination data of the power spectrum and demodulation spectrum, which has better feature extraction performance than the conventional feature extraction methods based on the presupposed models. (3) The RBM auto-encoder is used to enlarge the data samples and improve the performance of the recognition system.

The experimental results show that the proposed method has better performance than the traditional method, which provides good technical support for the target classification and recognition function of the SONAR system.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their useful comments.

REFERENCES

[1] P. T. Arveson and D. J. Vendittis, "Radiated noise characteristics of a modern cargo ship," *J. Acoust. Soc. Amer.*, vol. 107, no. 1, pp. 118–129, Jan. 2000, doi: 10.1121/1.428344.

[2] F. Bao, C. Li, X. Wang, Q. Wang, and S. Du, "Ship classification using nonlinear features of radiated sound: An approach based on empirical mode decomposition," *J. Acoust. Soc. Amer.*, vol. 128, no. 1, pp. 206–214, Jul. 2010, doi: 10.1121/1.3436543.

[3] Y. S. Cheng, J. X. Qiu, Z. Liu, and H. T. Li, "Challenges and prospects of underwater acoustic passive target recognition technology," *J. Appl. Acoust.*, vol. 38, no. 4, pp. 653–659, 2019.

[4] X. Cao, X. Zhang, R. Togneri, and Y. Yu, "Underwater target classification at greater depths using deep neural network with joint multiple-domain feature," *IET Radar, Sonar Navigat.*, vol. 13, no. 3, pp. 484–491, Mar. 2019, doi: 10.1049/iet-rsn.2018.5279.

[5] 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, Oct. 2018, Art. no. 1214301, doi: 10.1155/2018/1214301.

[6] J. Zhang and Y. Ding, "Underwater target recognition based on spectrum learning with convolutional neural network," in *Proc. IEEE 5th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Chongqing, China, Jun. 2020, pp. 1520–1523, doi: 10.1109/ITOEC49072.2020.9141661.

[7] L. M. Gray and D. S. Greeley, "Source level model for propeller blade rate radiation for the world's merchant fleet," *J. Acoust. Soc. Amer.*, vol. 67, no. 2, pp. 516–522, Apr. 1997.

[8] W. K. McDowell, W. B. Mikhael, and A. P. Berg, "Efficiency of the KLT on voiced & unvoiced speech as a function of segment size," in *Proc. IEEE Southeastcon*, Orlando, FL, USA, Mar. 2012, pp. 1–5.

[9] X. Zeng, Q. Wang, C. Zhang, and H. Cai, "Feature selection based on ReliefF and PCA for underwater sound classification," in *Proc. 3rd Int. Conf. Comput. Sci. Netw. Technol.*, Dalian, China, Oct. 2013, pp. 442–445.

[10] Z. B. Lu, X. H. Zhang, and J. Zhu, "Feature extraction of ship-radiated noise based on Mel frequency cepstrum coeffcients," in *Proc. Conf. Ship Sci. Technol.*, Feb. 2004, pp. 51–54.

[11] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba, "Learning with hierarchical-deep models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1958–1971, Aug. 2013, doi: 10.1109/tpami.2012.269.

[12] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, 2019, doi: 10.1109/access.2019.2901066.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[14] G. Jin, F. Liu, H. Wu, and Q. Song, "Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal," *J. Experim. Theor. Artif. Intell.*, vol. 32, no. 2, pp. 205–218, Mar. 2020, doi: 10.1080/0952813X.2019.1647560.

[15] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Comput.*, vol. 24, no. 8, pp. 1967–2006, Aug. 2012, doi: 10.1162/NECO_a_00311.

[16] Q. Wang, L. Wang, X. Zeng, and L. Zhao, "An improved deep clustering model for underwater acoustical targets," *Neural Process. Lett.*, vol. 48, no. 3, pp. 1633–1644, Dec. 2018, doi: 10.1007/s11063-017-9755-7.

[17] C.-C. Chang and C.-J. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.

[18] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 719–725, Jul. 2000, doi: 10.1109/34.865189.

[19] J. W. Xie, "Research on underwater sound source separation technology based on deep learning," M.A. dissertation, Univ. Electron. Sci. Technol. China, Chengdu, China, 2019.

[20] X. Luo and Y. Feng, "An underwater acoustic target recognition method based on restricted Boltzmann machine," *Sensors*, vol. 20, no. 18, p. 5399, Sep. 2020, doi: 10.3390/s20185399.

[21] S. L. Fang, S. P. Du, X. W. Luo, N. Han, and X. N. Xu, "Feature analysis and recognition technology of underwater acoustic targets," *Bull. Chin. Acad. Sci.*, vol. 2019, 34, pp. 297–305, doi: 10.16418/j.issn.1000-3045.2019.03.007.

[22] D. Ross and W. A. Kuperman, "Mechanics of underwater noise," *J. Acoust. Soc. Amer.*, vol. 86, no. 4, p. 1626, Oct. 1989, doi: 10.1121/1.398685.

[23] Y. Wen and M. Henry, "Time frequency characteristics of the vibroacoustic signal of hydrodynamic cavitation," *J. Vib. Acoust.*, vol. 124, no. Oct. 2002, pp. 1–19, 2002.

[24] P. Clark, I. Kirsteins, and L. Atlas, "Multiband analysis for colored amplitude-modulated ship noise," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3970–3973, doi: 10.1109/ICASSP.2010.5495776.

[25] D. Hanson, J. Antoni, G. Brown, and R. Emslie, "Cyclostationarity for passive underwater detection of propeller craft: A development of DEMON processing," in *Proc. Acoust.*, Nov. 2008, pp. 1–6.

[26] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Appl. Acoust.*, vol. 113, pp. 64–69, Dec. 2016, doi: 10.1016/j.apacoust.2016.06.008.

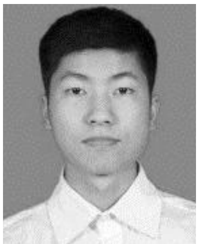
[27] F. Liu, Q. Song, and G. Jin, "Expansion of restricted sample for underwater acoustic signal based on generative adversarial networks," in *Proc. 10th Int. Conf. Graph. Image Process. (ICGIP)*, May 2019, Art. no. 1106948.

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[29] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2924–2932.

XINWEI LUO (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Southeast University, Nanjing, China, in 2001, 2006, and 2013, respectively. He is currently an Associate Professor with the Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University. His research interests include acoustic signal processing, target detection, parameter estimation, and underwater target classification.

YULIN FENG (Student Member, IEEE) received the B.S. degree from Northeast University, Shenyang, China, in 2019. He is currently pursuing the master's degree with the Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University. His research interests include acoustic signal processing, array signal processing, parameter estimation, and signal detection.



MINGHONG ZHANG received the B.S. degree from Southeast University, Nanjing, China, in 2020. He is currently pursuing the master's degree with the Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University. His research interests include acoustic signal processing, signal detection, and signal classification.