



Underwater Acoustic Target Recognition with Fusion Feature

Pengyuan Qi, Jianguo Sun, Yunfei Long, Ligu Zhang^(✉), and Tianye

College of Computer Science and Technology, Harbin Engineering University,
Harbin 150001, China
zhangliguo@hrbeu.edu.cn

Abstract. For the underwater acoustic targets recognition, it is a challenging task to provide good classification accuracy for underwater acoustic target using radiated acoustic signals. Generally, due to the complex and changeable underwater environment, when the difference between the two types of targets is not large in some sensitive characteristics, the classifier based on single feature training cannot output correct classification. In addition, the complex background noise of target will also lead to the degradation of feature data quality. Here, we present a feature fusion strategy to identify underwater acoustic targets with one-dimensional Convolutional Neural Network. This method mainly consists of three steps. Firstly, considering the phase spectrum information is usually ignored, the Long and Short-Term Memory (LSTM) network is adopted to extract phase features and frequency features of the acoustic signal in the real marine environment. Secondly, for leveraging the frequency-based features and phase-based features in a single model, we introduce a feature fusion method to fuse the different features. Finally, the newly formed fusion features are used as input data to train and validate the model. The results show the superiority of our algorithm, as compared with the only single feature data, which meets the intelligent requirements of underwater acoustic target recognition to a certain extent.

Keywords: Acoustic signal · Target recognition · Feature fusion

1 Introduction

Passive sonar is widely used in the underwater target recognition because of its excellent concealment and long working distance. It is usually designed to detect and identify targets from the ubiquitous clutter, a typical scenario shown in Fig. 1. During the passive sonar detection, pattern classification method is used to detect the underlying pattern or structures in the acoustic signal received by the front end. The sonar target classification system recognition process is illustrated in Fig. 2. The methods for underwater acoustic targets classification

Supported by Natural Science Foundation of Heilongjiang Province No. F2018006.

© Springer Nature Switzerland AG 2021

T. Mantoro et al. (Eds.): ICONIP 2021, LNCS 13108, pp. 609–620, 2021.

https://doi.org/10.1007/978-3-030-92185-9_50

are far from practical application, especially in a real-ocean environment. The reasons include the acoustic characteristics of different types of targets overlap, the complex and changeable ocean environment, the low signal-to-noise ratio of the receiving signal, high-quality data is rare and costly to obtain. These factors make the process of object classification a complicated problem. So far, the recognition of underwater acoustic signals has attracted widespread attention from scholars.

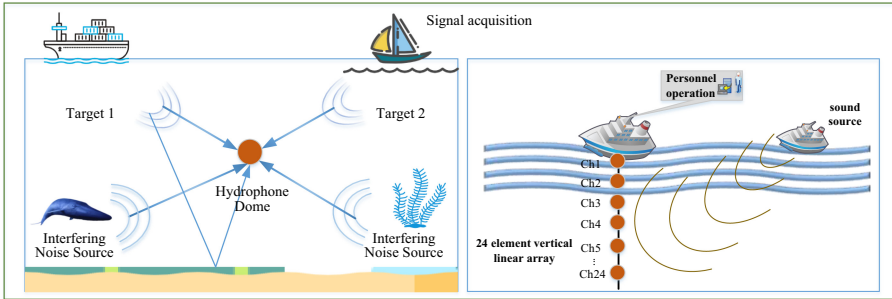


Fig. 1. Sonar working schematic diagram. An object can use the sonar equipment to analyze the underwater signals. The figure on the left shows the array element acquiring target information. The figure on the right shows a vertical line array sonar, the hydrophone array used is composed of 24 array elements, namely Ch1 to Ch24, and the array elements are equally spaced.

Aiming at the problem of underwater acoustic signal identification, a variety of identification methods are proposed. The characteristic parameters of time-domain waveforms and time-frequency analysis, nonlinear characteristic parameters [1,2] and spectrum analysis with the line spectrum characteristics, Low-Frequency Analysis and Recording(LOFAR), high-order spectra, Detection of Envelope Modulation on Noise(DEMON) are used commonly. The extracted auditory characteristic parameters commonly include Mel Frequency Cepstrum Coefficient(MFCC), Linear Predictive Cepstral Coefficient(LPCC) [3–5].

With the development of big data technology and the improvement of computer computing power, Machine Learning(ML), especially Deep Learning(DL) has been widely used in related application fields, e.g., Support Vector Machines, Back Propagation Neural Networks, K-Nearest Neighbor is employed for underwater acoustic signal recognition [6–10]. However, with the increase of the amount of data, ML can hardly meet the needs of existing recognition tasks. DL showed strong data processing and feature learning capabilities by the commonly used of Denoising Auto-Encoder (DAE) [11], LSTM [12], Deep Convolutional Neural Networks (DCNN) [13]. More and more scholars have begun to apply DL to the underwater acoustic target recognition, e.g., Z. Xiangyang et al. proposed a method of transforming the original one-dimensional underwater acoustic signal into a multi-dimensional acoustic spectrogram [14]. H. Yang

et al. proposed an LSTM-based DAE collaborative network [15]. H. Yang et al. proposed an end-to-end deep neural network based on auditory perception-Deep Convolutional Neural Network (ADCNN) [16], Y. Gao et al. proposed the combination of Deep Convolutional Generative Adversarial Network (DCGAN) and Densely Connected Convolutional Networks (DenseNet), which extracts deep features for underwater acoustic targets [17]. J. Chen et al. proposed a LOFAR spectrum enhancement (LSE)-based underwater target recognition scheme [18]. Considering the relative scarcity of underwater acoustic data sets for training, G. Jin et al. presented a novel framework that applied the LOFAR spectrum for preprocessing to retain key features and utilized Generative Adversarial Networks (GAN) for the expansion of samples to improve the performance classification [19]. The above works show that deep network has powerful modeling ability for complex functions with high dimensional input.

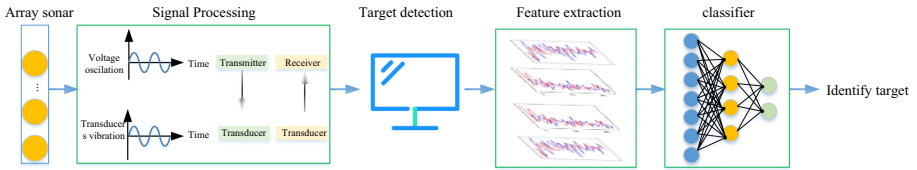


Fig. 2. The sonar target recognition process. Sonar equipment can detect objects through electro-acoustic conversion and information processing.

The deep network models rely only on a single spectral feature, such as the STFT feature [20] and the LPS feature [21], some important characteristics of radiated noise from underwater targets may be lost. In this paper, by extracting these two kinds of features, the advantages of LSTM system in complex underwater acoustic signal modeling are further studied. In addition, the audio signal has timing characteristics, the LSTM network is usually more excellent than other networks for processing the timing information. Influenced by this, we do this by exploring two different properties of the radiated noise training data set: the frequency spectrum and the phase spectrum in low-frequency band. The framework of the proposed underwater target classification model is described in Fig. 3. Our experimental results show that the proposed method performs significantly better than the single feature in terms of recognition accuracy. The contributions of this paper are summarised as follows:

- (1) The model is used to automatically learn the effective feature representation of complex target signals, and it can greatly improve the performance of pattern recognition system compared with the previous manual feature extraction.
- (2) We construct a joint feature for the depth model based on the spectrum and phase spectrum information, and make full use of the advantages of the depth structure to achieve feature complementarity and reduce the impact of the inherent defects of a single feature.

- (3) Our method is tested on the underwater acoustic signals which is different from the previous work under simulation conditions and achieves outstanding performance compared with the single method.

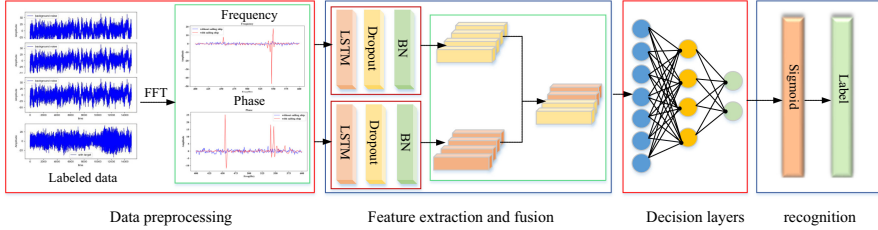


Fig. 3. The proposed frequency-phase spectrum identify model.

2 Method

2.1 Model Overview

In the first phase, we need to extract the low-level features of different domains based on LSTM network and the multi-domain feature vectors are spliced into joint feature inputs suitable for model training. The joint feature is composed of the frequency spectrum feature and phase spectrum feature. In this paper, the feature subsets of frequency and phase are fused directly in the series form to form multi-category fusion features. In the classification stage, CNN was used to classify and identify the targets. The design of the framework based on CNN is described in Fig. 4. In the prediction classification stage, the above process was repeated to obtain the fusion feature subset of the test samples, and the trained classifier was used to identify the target category.

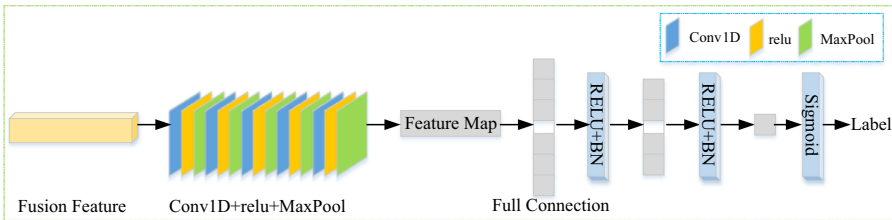


Fig. 4. The framework based on Convolutional Neural Network.

2.2 Frequency-Phase Spectrum Analysis

In the actual marine environment, the underwater acoustic signal is commonly affected by the following two aspects: 1) environmental noise; 2) experimental platform noise. Figure 5 displays the time-domain waveform of the original underwater acoustic signal, which is part of the underwater target in the data set. The strong background noise caused that the time-domain waveform of the original underwater acoustic signal shows noise-like characteristics. In order to verify the effectiveness of multidimensional feature fusion method proposed in this paper, we choose to analyze the frequency spectrum and the phase spectrum of signals by Fourier Transform on the time-domain waveform, Fourier Transform is shown in Eq. (1). Figure 6(a) and Fig. 6(b) are respectively the frequency spectrum and phase spectrum of the underwater acoustic signal, in which the red line represents the signal with sailing ship target and the blue line represents the signal without sailing ship target (background noise).

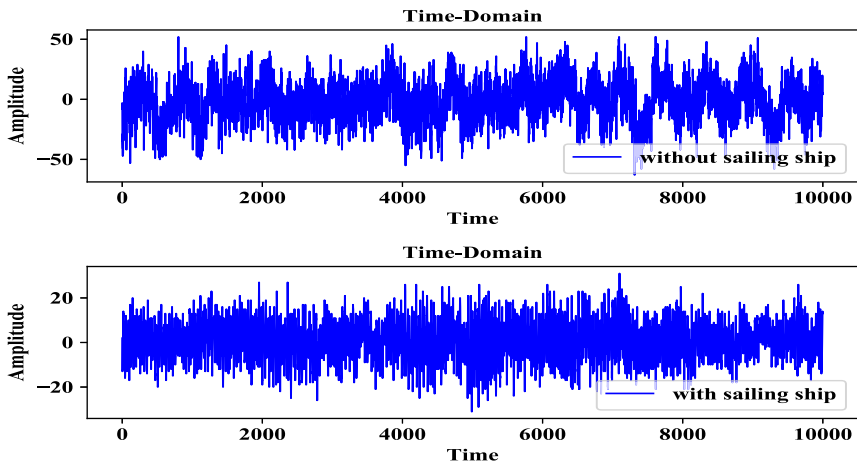


Fig. 5. Time-domain waveform.

Continuous spectrum and line spectrum make up the frequency spectrum of the underwater acoustic signal commonly. The ship-radiated noise when sailing includes three kinds of low-frequency line spectrum, which all in 100 Hz–600 Hz. Therefore, in Fig. 6(a) the frequency spectrum comparison chart, the peak-to-peak value of the signal with the sailing ship is significantly higher than the background noise 420 Hz–460 Hz and 520 Hz–560 Hz because of line spectrum. In the phase spectrum comparison chart, the difference of peak-to-peak value is equally obvious within the aforementioned frequency range. If the features of the underwater acoustic signal are analyzed only from the frequency spectrum, that will lose part information of the signal. Taking the frequency spectrum and

phase spectrum of the underwater acoustic signal as the input of the recognition model can effectively compensate for the lack of underwater acoustic signal characteristics.

$$F(\omega) = \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (1)$$

Where, $f(t)$ refers to the time-domain data of original underwater acoustic signal.

2.3 Frequency-Phase Feature Fusion Recognition

Figure 3 describes the process of expressing feature extraction. The frequency feature and phase feature can be obtained from the spectrogram. The process can finally extract two-dimensional features and form new feature vectors. The new feature vectors \vec{N} can be expressed as:

$$\vec{N}_i = \{F_i(t), P_i(t)\} \quad (2)$$

where t is time series, $F_i(t)$ is the frequency characteristic value at time i , $P_i(t)$ is the phase characteristic value at time i .

In this paper, the joint feature input N is build for deep learning network to identify underwater acoustic signal. In Sect. 2.2, we analyze feature of the signal. The peak-to-peak value of frequency spectrum and phase spectrum is obviously different in 100 Hz–600 Hz. Therefore, when preprocessing the underwater acoustic signal, we need to obtain the frequency spectrum and the phase spectrum of the underwater acoustic signal by Fourier transform in 100 Hz–600 Hz and normalize them by the Deviation Standard method. Take processed frequency spectrum and phase spectrum as the input of model, and feature learning from them through LSTM network. The dimensions of the feature are 50. Then, obtain fusion feature by concatenating. The dimensions of the fused feature are 100. Finally, the recognition result can be implemented by the FC layers and Sigmoid function. The specific Algorithm 1 of the proposed multi-dimensional fusion feature is as follows:

Algorithm 1. Multi-dimensional Fusion Feature.

Require: original underwater acoustic signals $s(t)$;

Ensure: fusion feature vector;

- 1: Initialization: LSTM, Dropout(rate=0.25),BatchNormalization(BN),P,F,N;
 - 2: Let $x(t) \leftarrow s(t)$;
 - 3: Calculate $x(k)$ using Equation (1);
 - 4: Decompose $x(k)$ into $freq(k)$ and $phase(k)$;
 - 5: Calculate Frequency_ feature F using $BN(Dropout(LSTM(freq(k))))$;
 - 6: Calculate Phase_ feature P using $BN(Dropout(LSTM(phase(k))))$;
 - 7: Calculate output_feature $N \leftarrow F + P$;
 - 8: **return** N ;
-

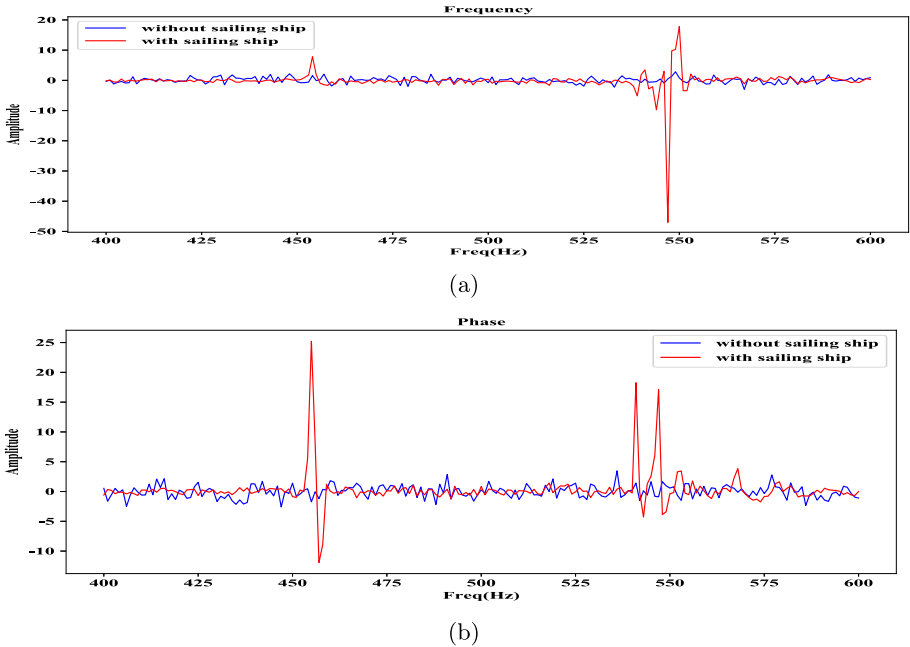


Fig. 6. The frequency and phase comparative chart by Fourier transform of the time-domain waveform, where the first row shows a frequency spectrum of underwater target. Wherein, the ordinate is the spectrum value. The second row shows a phase spectrum of underwater target. Wherein, the ordinate is the phase value.

3 Experiments

In this section, we introduce the implementation details and quality assessment criteria. Finally, the experimental results are given which proves the superiority of the method proposed in this paper.

3.1 Dataset and Experiment Platform

The method is verified by two kinds of signal data: is there a sailing ship. Each signal in this paper comes from a passive sonar in the marine and a sampling rate of 25600 Hz. The total number of samples of the model is 1.8×10^4 , in which the number of the signal with a sailing ship is 1.0×10^4 , the number of the signal without a sailing ship is 0.8×10^4 . In order to ensure the validity of the verification results, the paper randomly selected 1.72×10^4 from the sample library to form a number of training sets, the remaining samples as the test set. We train our model on the NVIDIA TITAN XP by CUDA 9.0 and Keras.

3.2 Implementation Details

Comparative Experiments. For fairness, in this paper, we use homologous underwater acoustic signals. By the data processing, we obtain the frequency spectrum, the phase spectrum, and the MFCC feature. The dimensions of the overall MFCC feature are 96, which includes MFCC parameter, first-order difference MFCC, and second-order difference MFCC. Our model uses the frequency spectrum, and the phase spectrum as the input data. Comparative experiments use frequency spectrum, phase spectrum, and MFCC feature as single input.

Table 1. LSTM network parameters

Input data	Input Nodes	LSTM NO	Layers	Feature Dimension	FC Layer
Frequency	500	1	1	50	2
Phase	500	1	1	50	2
MFCC	96	1	1	50	2
Feature fusion	1000	2	1	100	2

Training Setup. The model has trained with 1.72×10^4 samples, the learning rate is 5×10^{-6} by 700 epochs. The LSTM layer params of experiments in Table 1. A dropout layer is inserted after the LSTM layer with the dropout rate of 0.25.

Quality Assessment Criteria. This paper mainly focuses on the classification problem of the 2 types of acoustic signals. In order to evaluate the proposed method, we used the data samples from real marine data, F1 score and accuracy rate as evaluation indexes. We used the true negative rate, true positive rate, false positive rate, and false negative rate from ML and use F1-score which is the harmonic mean recall of Recall rate (R) and Precision rate (P) to evaluate the model's recognition effect of underwater targets in the test set. The F1-score calculation formula is shown in Eq. (3), and the Accuracy Rate is shown in Eq. (4).

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

Where P is Precision rate, R is Recall rate.

$$\text{Accuracy Rate} = \frac{A_{\text{acc}}}{T_{\text{total}}} \times 100\% \quad (4)$$

Where A is the total number of objects that can be correctly identified, T is the total number of the two targets.

3.3 Experimental Results

In this paper, we compare the performance of the proposed method with other methods, i.e., MFCC, frequency, and phase on the validation set. In this paper,

Table 2. Confusion matrix of experimental results

Recognition result	Method		Prediction real label	Prediction goalless label	In total
Real targeted lable	Sigle feature	Frequency	341	77	418
		Phase	302	116	
		MFCC	349	69	
	Fusion feature	Our method	381	37	
Real goalless label	Sigle feature	Frequency	307	75	382
		Phase	322	60	
		MFCC	329	53	
	Fusion feature	Our method	337	45	

the F1 score and accuracy are adopted as the evaluation indexes. As can be seen from Table 2, the F1-score in frequency, phase and MFCC is 64%, 57.9% and 63.1% respectively, the calculated F1-Score is 72.1% with our method by Eq. (3), compared with frequency feature and phase feature, the fusion feature can improve the performance of recognition precision.

To simulate practical applications of recognition for ship-radiated noise, the classification accuracy of each acoustic event is used to measure the classification performance of the model, which is defined as the percentage of all acoustic events that are correctly classified. The classification accuracy of the proposed method and the comparison method is shown in Table 3. As we can see from

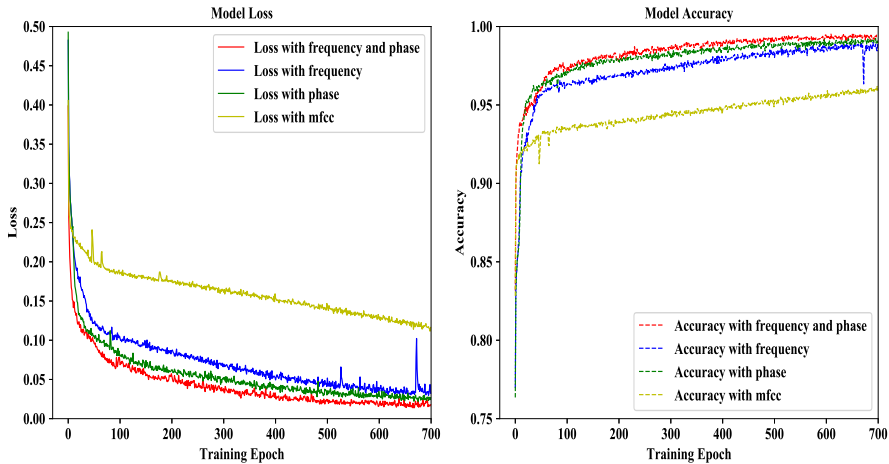


Fig. 7. Loss and accuracy of the model, where the left shows the classification loss of the proposed method with frequency, phase and MFCC, the right shows the classification accuracy of the proposed method with frequency, phase and MFCC.

Table 3, compared with a single feature of underwater acoustic target recognition methods, the proposed fusion method effectively improves the classification accuracy of the underwater acoustic target.

To further show the effectiveness of our proposed model, the recognition performance results on the validation set are illustrated in Fig. 7, which details the classification accuracy improvement of feature fusion relative to phase spectrum, frequency spectrum, and MFCC in each class. As shown in Fig. 7, in the process of model training, there is no over-fitting or under-fitting phenomenon, and there is no gradient disappearance or gradient explosion. By testing the model with measured data, with the number of model training steps increases, the proposed method can achieve a higher recognition accuracy on the validation set. We provide a confusion matrix for the recognition result of the proposed model, as shown in Fig. 8. Each row of the confusion matrix correspond to the real label and each column corresponds to the predicted label.

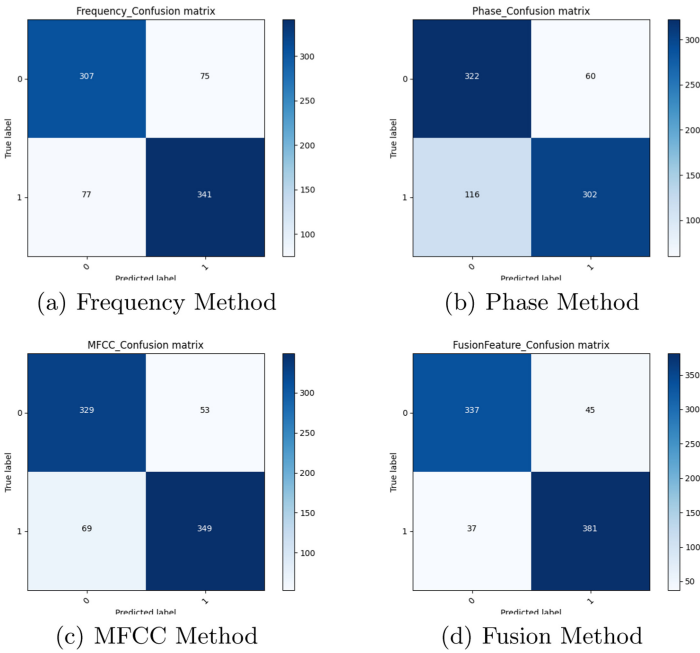


Fig. 8. The confusion matrix of the proposed model obtained from testing data.

Table 3. The classification results of proposed method and compared methods

Method	Model	Accuracy
Frequency	One-DCNN	81.00%
Phase	One-DCNN	78.00%
MFCC	One-DCNN	84.75%
Fusion-Feature	One-DCNN	89.75%

4 Conclusion

This paper focuses on how to introduce the acoustic feature of the frequency-phase spectrum and two types of feature fusion model into the passive recognition problem. In order to alleviate the identify difficulty in the actual marine, the recognition method based on frequency-phase spectrum analysis is proposed. In this method, the LSTM is used for multi-class feature extract. By analyzing the target and background noise, two kinds of target data are obtained. Experiments show that the frequency-phase spectrum recognition method proposed can effectively distinguish the above two types of target, and the recognition results of the two types of feature fusion are better than other cases. In addition, our method strengthens the interpretability of the features extracted compared to deep learning technology.

In this paper, only the two discriminant methods were studied and introduced. However, due to the lack of relevant research on optimization selection, the comparison of the discriminant effects after optimization between the two methods needs to be further studied and discussed in the future.

References

1. Oswald, J.N., Au, W.W., Duennebie, F.: Minke whale (*balaenoptera acutorostrata*) boings detected at the station aloha cabled observatory. *J. Acoust. Soc. Am.* **129**(5), 3353–3360 (2011)
2. Esfahanian, M., Zhuang, H., Erdol, N.: Using local binary patterns as features for classification of dolphin calls. *J. Acoust. Soc. Am.* **134**(1), EL105–EL111 (2013)
3. Chinchu, M., Supriya, M.: Real time target recognition using labview. In: International Symposium on Ocean Electronics (SYMPOL), pp. 1–9. IEEE (2015)
4. Wang, W., Li, S., Yang, J., Liu, Z., Zhou, W.: Feature extraction of underwater target in auditory sensation area based on MFCC. In: IEEE/OES China Ocean Acoustics (COA), pp. 1–6 (2016)
5. Zhang, L., Wu, D., Han, X., Zhu, Z.: Feature extraction of underwater target signal using mel frequency cepstrum coefficients based on acoustic vector sensor. *J. Sens.* **11–17** (2016)
6. Liu, H., Wang, W., Yang, J.-A., Zhen, L.: A novel research on feature extraction of acoustic targets based on manifold learning. In: International Conference on Computer Science and Applications (CSA), pp. 227–231. IEEE (2015)
7. Sun, L., Kudo, M., Kimura, K.: Reader semi-supervised multi-label dimension reduction. *IEICE Trans. Inf. Syst.* **100**(10), 2597–2604 (2017)
8. Sherin, B., Supriya, M.: Sos based selection and parameter optimization for underwater target classification. In: OCEANS MTS/IEEE Monterey, pp. 1–4. IEEE (2016)
9. Li, H., Cheng, Y., Dai, W., Li, Z.: A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition. In: International Conference on Signal Processing (ICSP), pp. 2169–2173. IEEE (2014)
10. Barngrover, C., Althoff, A., DeGuzman, P., Kastner, R.: A brain-computer interface (BCI) for the detection of mine-like objects in sidescan sonar imagery. *IEEE J. Oceanic Eng.* **41**(1), 123–138 (2015)

11. Verwimp, L., Pelemans, J., Wambacq, P., et al.: Character-word LSTM language models, arXiv preprint [arXiv:1704.02813](https://arxiv.org/abs/1704.02813) (2017)
12. Mimura, M., Sakai, S., Kawahara, T.: Speech dereverberation using long short-term memory. In: Sixteenth Annual Conference of the International Speech Communication Association (2015)
13. Wang, P., Peng, Y.: Research on feature extraction and recognition method of underwater acoustic target based on deep convolutional network. In: International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 863–868. IEEE (2020)
14. Xiangyang, Z., Jiaruo, H., Lixiang, M.: Image representation of acoustic features for the automatic recognition of underwater noise targets. In: Third Global Congress on Intelligent Systems, pp. 144–147. IEEE (2012)
15. Yang, H., Xu, G., Yi, S., Li, Y.: A new cooperative deep learning method for underwater acoustic target recognition. In: OCEANS 2019-Marseille, pp. 1–4. IEEE (2019)
16. Yang, H., Li, J., Shen, S., Xu, G.: A deep convolutional neural network inspired by auditory perception for underwater acoustic target recognition. *Sensors* **19**(5), 1104 (2019)
17. Gao, Y., Chen, Y., Wang, F., He, Y.: Recognition method for underwater acoustic target based on DCGAN and DenseNet. In: 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC), pp. 215–221. IEEE (2020)
18. Chen, J., Liu, J., Liu, C., Zhang, J., Han, B.: Underwater target recognition based on multi-decision lofar spectrum enhancement: a deep learning approach. arXiv preprint [arXiv:2104.12362](https://arxiv.org/abs/2104.12362) (2021)
19. Jin, G., Liu, F., Wu, H., Song, Q.: Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. *J. Exp. Theoret. Artif. Intell.* **32**(2), 205–218 (2020)
20. Kamal, S., Mohammed, S.K., Pillai, P.S., Supriya, M.: Deep learning architectures for underwater target recognition. In: 2013 Ocean Electronics (SYMPOL), pp. 48–54. IEEE (2013)
21. Cao, X., Zhang, X., Yu, Y., Niu, L.: Deep learning-based recognition of underwater target. In: 2016 IEEE International Conference on Digital Signal Processing (DSP), pp. 89–93. IEEE (2016)