Journal of Systems Engineering and Electronics
Vol. 34, No. 4, August 2023, pp.839 – 850

Range estimation of few-shot underwater sound source in shallow water based on transfer learning and residual CNN

YAO Qihai[1,2], WANG Yong[1,2,*], and YANG Yixin[1,2]

1. School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China;
2. Shaanxi Key Laboratory of Underwater Information Technology, Xi'an 710072, China

Abstract: Taking the real part and the imaginary part of complex sound pressure of the sound field as features, a transfer learning model is constructed. Based on the pre-training of a large amount of underwater acoustic data in the preselected sea area using the convolutional neural network (CNN), the few-shot underwater acoustic data in the test sea area are retrained to study the underwater sound source ranging problem. The S5 voyage data of SWellEX-96 experiment is used to verify the proposed method, realize the range estimation for the shallow source in the experiment, and compare the range estimation performance of the underwater target sound source of four methods: matched field processing (MFP), generalized regression neural network (GRNN), traditional CNN, and transfer learning. Experimental data processing results show that the transfer learning model based on residual CNN can effectively realize range estimation in few-shot scenes, and the estimation performance is remarkably better than that of other methods.

Keywords: transfer learning, residual convolutional neural network (CNN), few shot, vertical array, range estimation.

DOI: 10.23919/JSEE.2023.000095

1. Introduction

Underwater target location is a research hot spot in the field of underwater acoustic signal processing. Passive location technology is widely used in military and civil fields. Most of the traditional passive positioning methods are matched field processing (MFP) based on the underwater acoustic model. In 1976, Bucker proposed the linear matched field processor and established the actual environment model to realize passive positioning by calculating the fuzzy function of range and depth [1]. In 1981, Klemm proposed the generalized maximum entropy beamformer, which has a higher resolution and a better estimation performance than the linear matched field processor [2]. In 1988, Baggeroer established the MFP technology based on the horizontal layered marine environment waveguide model and obtained that increasing the bandwidth can effectively improve the accuracy of the matched field algorithm [3]. In 1996, Michalopoulou et al. superimposed the narrowband ambiguity function of each frequency point through the uncorrelated method and obtained that the uncorrelated method is feasible in the broadband MFP algorithm [4]. In 2003, Soares et al. designed normalized correlation processor and matched-phase correlation processor based on the correlation between various frequency points [5]. In 2006, Yang et al. designed a linear MFP algorithm in the scene of strong interference under the condition of environmental mismatch to suppress the interference [6]. The MFP method combines the acoustic propagation model with the array signal based on the characteristics of underwater acoustic channel to estimate the depth and range of underwater targets. However, its positioning performance depends heavily on parameters such as sea depth and sound velocity profile. In the scene of environmental mismatch, the accuracy of this method is seriously affected.

The data-driven model does not depend on the parameters of the marine environment. The data-driven method represented by the neural network has been widely used in the field of underwater acoustic passive positioning. In 1991, Steinberg et al. established a single-layer neural network model for point sound sources in homogeneous media to realize depth estimation [7]. However, at that time, the machine learning technology was relatively weak and not widely used, and the mainstream passive location algorithm was the matched field processing method. Therefore, in the field of underwater acoustic passive location, the development of machine learning

Manuscript received July 05, 2022.
*Corresponding author.
This work was supported by the National Natural Science Foundation of China (11974286;11904274) and the Shaanxi Young Science and Technology Star Program (2021KJXX-07), and the fundamental research funding for characteristic disciplines (G2022WD0235).

method was relatively slow for a long time. In 2017, Niu et al. used feedforward neural network (FNN) and support vector machine (SVM) to achieve effective localization of sound source under the condition that they only have approximate environmental prior parameter information [8]. In 2018, Wang et al. introduced the generalized regression neural network (GRNN) method in the field of underwater acoustic passive positioning. In the environment of high signal-to-noise ratio (SNR) in shallow water, they determined that this method can achieve effective target range estimation [9]. In the same year, Ferguson et al. extracted the features of cepstrum and generalized cross-correlation of sound source signal in shallow water waveguide environment, and realized target azimuth estimation based on convolutional neural network (CNN) [10]. In 2019, Liu et al. realized underwater target range estimation in the scene with roughly determined depth range based on one-dimensional (1D) CNN and integrated learning technology [11]. Under the condition of only approximate environmental parameters, Niu et al. generated a large amount of underwater acoustic data based on the sound propagation model, collected the sound pressure value received by a single hydrophone, and detected the sound source by using a 50-layer residual CNN [12]. Howarth et al. added Gaussian noise to underwater acoustic data to evaluate the performance of CNN in estimating target range and environment category under different SNRs [13]. In 2020, Komen et al. input the data obtained by modeling a variety of environmental parameters into the CNN model to realize target range estimation and environmental recognition. The network can effectively estimate the test set obtained by the sound field model, whereas the estimation performance of sea test data was relatively poor [14,15]. Ozanich et al. used the FNN for azimuth estimation and compared it with the SVM method. The results showed that the application of the deep FNN model in horizontal array passive positioning is feasible [16]. In the same year, Liu et al. proposed a multitask learning algorithm based on CNN, which can simultaneously estimate the range and depth of the sound source [17]. In 2021, Chen et al. proposed a CNN model whose training data are only composed of model simulation data. The results showed that the CNN model has a better estimation performance than the MFP method [18]. In the above research, CNN and SVM were used to estimate the position of a single underwater target, and satisfactory results were obtained.

Transfer learning is a machine learning method that uses existing knowledge to solve problems in different but related fields [19]. In the field of image processing, transfer learning has been widely used. In 2008, Dai et al. based on the transfer learning model and assisted image clustering with text data. This method can effectively improve the accuracy of image clustering [20]. In 2011, Zhu et al. took the label information on the image as a bridge for knowledge transfer between image and text, which helped improve the effect of image classification [21]. In 2019, Nakamura et al. proposed a fine-tuning method to make the model converge faster on small sample data sets and make the knowledge learned by the model more consistent with the target task [22]. Jang et al. proposed to use meta learning to learn the weights of transfer feature maps and transfer layers [23]. In 2022, Gerace et al. used the solvable model of synthetic data as a framework for modeling the correlation between data sets, and analyzed the generalization performance when transferring the trained feature map from the source task to the target task [24]. However, the application of transfer learning in underwater acoustic localization is less. In 2019, Wang et al. obtained the pre-training model by training a large amount of simulation data and then trained the measured data of a few shots. The research showed that this method can effectively realize the estimation of sound source range in the untested sea area [25]. In 2021, Cao et al. used the single-vector hydrophone simulation data generated by the sound propagation model for the pre-training of the model and used the sea trial data to realize the retraining of the model. The results showed that compared with the traditional CNN, the single-vector hydrophone azimuth estimation based on transfer learning is more robust [26].

To sum up, the traditional MFP method based on model driving has great limitations. It requires a large number of marine environment parameters and is prone to the problem of environmental mismatch. Compared with the machine learning method, it cannot extract the deep feature information in the underwater acoustic data, and the estimation performance is often poor. Although the traditional machine learning method can obtain the model similar to the estimation task through training, it often needs a large number of training data to support. The transfer learning method can be applied to new tasks by training similar tasks and solving the estimation problem with a small number of training samples. As for unfamiliar sea areas, neither enough underwater acoustic data nor environmental parameters generate a large amount of copy field data. Therefore, training an accurate deep neural network is impossible.

However, transfer learning can be used as a learning framework to apply the knowledge of known sea areas to the detection of unfamiliar sea areas. Therefore, the application of transfer learning in the field of underwater acoustic passive positioning has great research importance. This paper makes full use of a large amount of underwater acoustic data in the known sea area to train the corresponding pre-training model. Transfer learning is applied to the sea area target ranging with only a small amount of underwater acoustic data. The S5 data of SWellEX-96 experiment are used to verify the feasibility of this method.

2. MFP

Using the normal wave model, the sound pressure generated at (r,z) by a single frequency point sound source at $(0,z_s)$ [27] is expressed as

$$p(r,z) \approx \frac{i}{\rho(z_s)\sqrt{8\pi r}} \mathrm{e}^{-i\pi/4} \sum_{m=1}^{\infty} \Psi_m(z_s)\Psi_m(z) \frac{\mathrm{e}^{ik_{rm}r}}{\sqrt{k_{rm}}} \quad (1)$$

where i is the unit of imaginary number, ρ is the density of the medium, $\Psi_m(z_s)$ and k_{rm} are the mode depth function (characteristic function) and horizontal wave number of the mth normal mode, respectively.

The data received by the hydrophone array are matched with the copy vector generated through the normal wave model and environmental parameters, and the maximum value is found in the calculated blurred surface of range and depth. The corresponding position of this point is the estimated position [28]. The positioning ambiguity plane $\boldsymbol{B}$ of MFP is calculated from the cross-spectral density matrix $\boldsymbol{R}$ and copy vector $\boldsymbol{\omega}$:

$$\boldsymbol{B}(\varphi) = \boldsymbol{\omega}^{\mathrm{H}}(\phi)\boldsymbol{R}\boldsymbol{\omega}(\varphi), \quad (2)$$

$$\boldsymbol{R} = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{P}_l \boldsymbol{P}_l^{\mathrm{H}}, \quad (3)$$

$$\boldsymbol{\omega} = [p_1, p_2, \cdots, p_l, \cdots, p_L]^{\mathrm{T}}, \ l = 1,2,\cdots,L, \quad (4)$$

where $\boldsymbol{\omega}$ is the copy vector, p_l is the complex sound pressure received by the array under the copy field, ϕ is the position parameter of the sound source, $(\cdot)^{\mathrm{H}}$ is the conjugate transpose, L is the number of snapshots, and $\boldsymbol{P}_l$ is the frequency domain data vector of the array under the lth snapshot. The ambiguity plane of the broadband signal can be obtained by accumulating and averaging the ambiguity plane corresponding to each frequency.

3. CNN

3.1 Theoretical basis

The network structure of a typical CNN is shown in Fig. 1. In this network, each neuron is connected to the local receptive domain of the previous layer. Different levels of features in the original signal are obtained through convolution operation and nonlinear activation to realize the feature mapping of the previous layer [29].

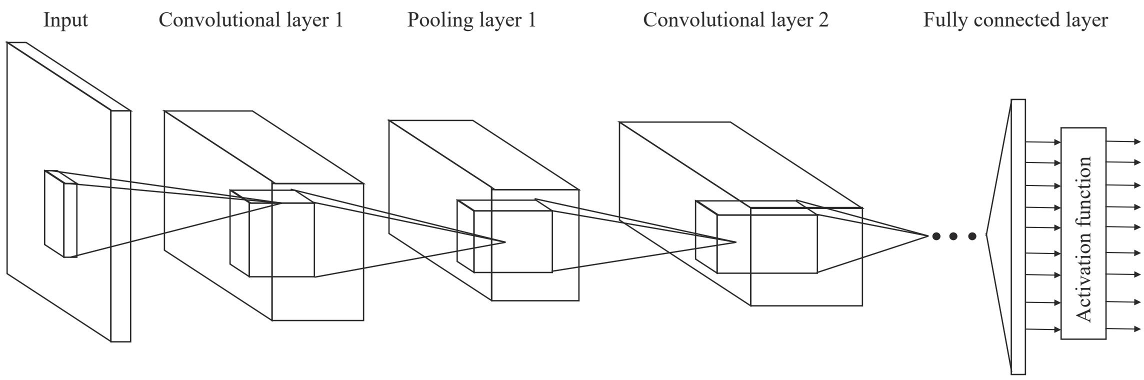


Fig. 1 Typical CNN model

The convolution can be expressed as

$$\boldsymbol{C}_l = f(\boldsymbol{W}_l * \boldsymbol{x}_{l-1} + \boldsymbol{b}_l) \quad (5)$$

where $*$ represents the convolution operation, $\boldsymbol{C}_l$ is the output of the current layer, $f(\cdot)$ represents the nonlinear activation function, $\boldsymbol{W}_l$ represents the weight of the lth layer, $\boldsymbol{x}_{l-1}$ represents the output of the previous layer, and $\boldsymbol{b}_l$ represents the deviation of the lth layer.

The pooling layer makes statistics on the overall characteristics of the nearby area at a certain location to reduce the diversity and dimensionality of feature selection, and effectively avoid the over fitting of the network while reducing the network parameters. In this paper, average pooling is adopted, which is expressed as

$$\boldsymbol{Z}_l = f(\boldsymbol{W}_l * \mathrm{mean}(\boldsymbol{x}_{l-1}) + \boldsymbol{b}_l) \quad (6)$$

where $\boldsymbol{Z}_l$ is the output mapping of the lth layer, and $\mathrm{mean}(\cdot)$ represents the average pooled sampling function.

The full connection layer integrates the high-dimensional information features after convolution and pooling. The layer uses the features corresponding to the linear equation to fit the input and then processes the information through the activation function. The model is

$$\boldsymbol{K} = f(\boldsymbol{w}_0 \cdot \boldsymbol{f}_v + \boldsymbol{b}_0) \tag{7}$$

where $\boldsymbol{f}_v$ is the eigenvector, $\boldsymbol{w}_0$ is the weight matrix, and $\boldsymbol{b}_0$ represents the offset matrix.

3.2 Residual connection model

The CNN model with residual connections can not only be used to build a deep network architecture, but also maintain the accuracy of the model [30]. The model adopts the network structure of jump connection to superimpose the shallow features and the deep features, which can effectively avoid the loss of shallow features during network training. The residual connection structure is shown in Fig. 2, in which x is the input of the current unit, and $F(x)$ is the mapping output of the current unit processed by the nonlinear transformation function. In the forward propagation of the CNN model, not only the mapping output result of each current unit is used as the input of the next unit, but also the input of the current unit is directly connected and added to the input of the next unit to realize the jump connection. Therefore, the input of the next unit is

$$H(x) = F(x) + x. \tag{8}$$

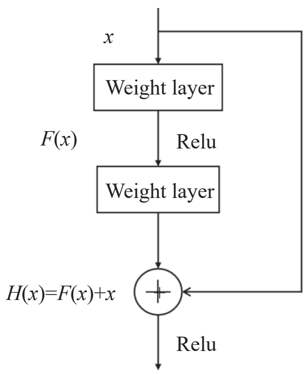


Fig. 2 Residual connection structure

Compared with traditional CNN, the most evident feature of the CNN model with residual connection is that many branches connect the input directly to the later layer. The residual CNN model only needs to obtain the difference information between input and output, which reduces the complexity of training objectives and the convergence time required for network model training.

3.3 CNN model construction

As a representative model in residual CNN, the ResNet18 model has an excellent recognition performance. In this paper, ResNet18 is used as the backbone network, and its framework is shown in Fig. 3. First, a convolution layer has a dimension of 7×7, followed by four ResBlocks, and a pool layer and a full connection layer at the rear. To adapt to the characteristics of underwater acoustic data, the CNN model adopted in this paper removes the pooling layer in the original ResNet18 model to retain more characteristic information in the input data, and changes the input layer, full connection layer, and output layer to the size suitable for this research task. The covariance matrix of the received data of the array is used for training, and the vertical array used contains 21 available elements. Therefore, the dimension of the covariance matrix is 21×21. To match the size of the convolution layer in the ResNet18 model, the input data need to be transformed into a 224×224 matrix. The problem of range estimation is transformed into a classification problem based on neural network, which has been widely used in the field of underwater sound source localization [8,12,18]. It can avoid searching in continuous space and effectively reduce the complexity and convergence time of network training. The corresponding grid interval needs to be set, which is similar to MFP. It can be expressed as

$$\text{grid} = \frac{r_{\max} - r_{\min}}{N} \tag{9}$$

where N is the number of grid intervals, and $r_{\max}$ and $r_{\min}$ represent the maximum range and the minimum range, respectively. The label of the ith sample used for training is generated according to

$$\text{label}_i = \frac{r_i - r_{\min}}{\text{grid}} \tag{10}$$

where r_i represents the range of the ith sample. The formula divides the maximum range and the minimum range into N categories, and assigns labels according to the target range. In practical applications, the setting of parameter N is related to the data set used. It should not be too large, otherwise, the grid interval will be less than the range resolution of training data, which will lead to the labels having no practical training significance. Because the velocity of the sound source in this study is 2.5 m/s, the grid interval corresponding to the label of 1 s-snapshot training data should not be less than 2.5 m. Theoretically, on the premise that the grid interval exceeds the range resolution, the estimation performance can be improved by increasing parameter N as much as possible. However, in the actual training process, only considering the increase of parameter N will lead to over fitting and poor generalization of the network, and the estimation performance of the model on the test data will be degraded. Based on the above considerations, the parameter N is set to 100, so the grid interval is 0.018 1 km, which meets the above grid interval conditions and the actual accuracy requirements of range estimation of underwater sound source.

The CNN model optimization algorithm is stochastic gradient descent with momentum (SGDM) algorithm. The number of batch training samples is 128 and the learning rate is 0.001. In order to avoid overfitting, the batch normalization layer is set after the convolution layer, the L2 normalization (weight decay) coefficient is set to 0.000 1, and the dropout layer with a ratio of 0.5 is also set.

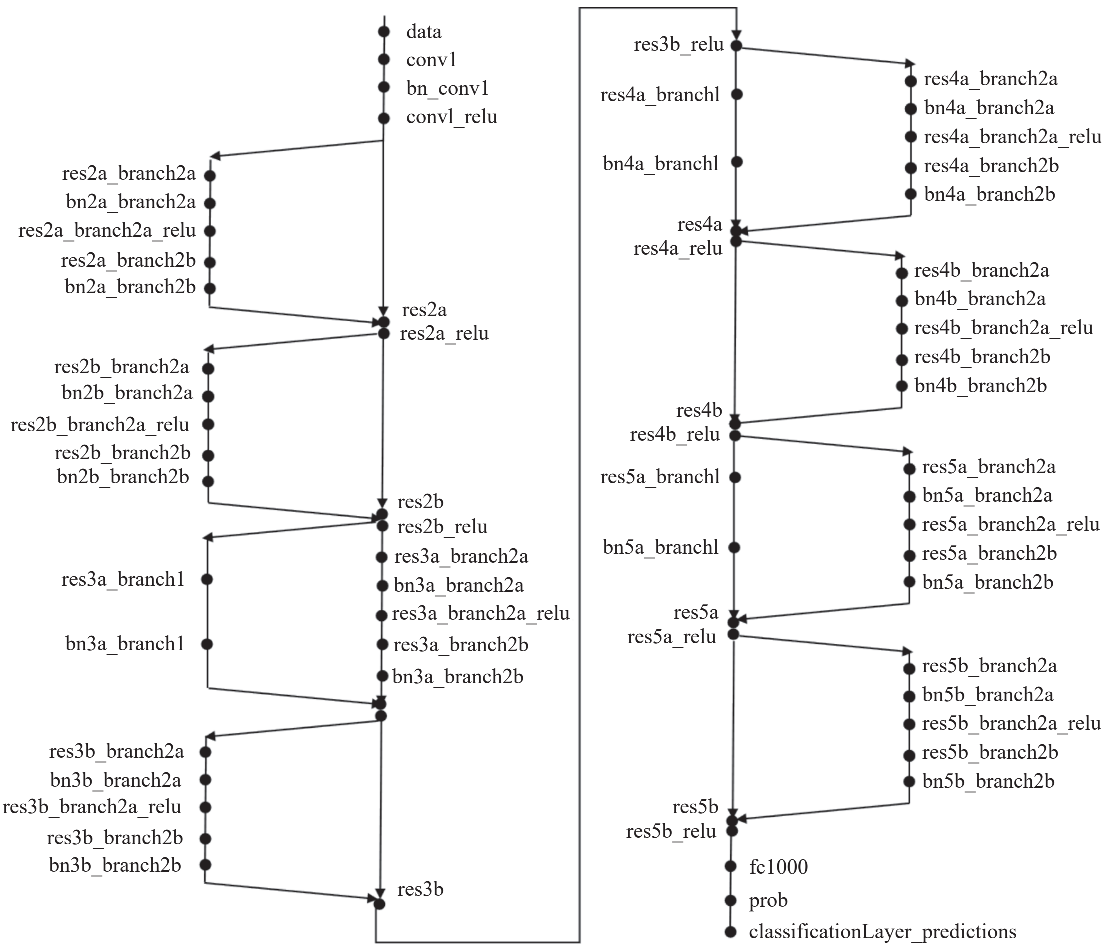


Fig. 3 ResNet18 model framework

4. Deep transfer learning and data preprocessing

Transfer learning for underwater ranging uses the knowledge learned from a scene (model sound field or historical environment sound field) to ranging the sound source of the new environment. Traditional machine learning models train data sets in different fields independently and cannot be directly applied in other environments. CNN is a kind of deep neural network and has been widely used in underwater acoustic target classification and location. Based on the traditional CNN, transfer learning can transfer a pre-training model to a new field. In this paper, the pre-training model is trained from the normalized experimental underwater acoustic data and transferred to another sea area for target range estimation with only few-shot underwater acoustic data. The transfer learning system is shown in Fig. 4.

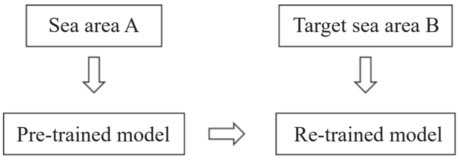


Fig. 4 Transfer learning system

4.1 Data preprocessing

To reduce the influence of sound source amplitude, the frequency domain complex sound pressure received by L-element array is normalized by

$$\tilde{\boldsymbol{P}}(f)=\frac{\boldsymbol{P}(f)}{\sqrt{\sum_{l=1}^{L}|p_l(f)|^2}}=\frac{\boldsymbol{P}(f)}{\|\boldsymbol{P}(f)\|_2}=\left[\tilde{p}_1(f),\tilde{p}_2(f),\cdots,\tilde{p}_L(f)\right]^{\mathrm{T}}. \tag{11}$$

The normalized covariance matrix is obtained by averaging the data of N_s snapshots. The calculation formula is

$$\boldsymbol{C}(f)=\frac{1}{N_s}\sum_{s=1}^{N_s}\tilde{\boldsymbol{P}}_s(f)\tilde{\boldsymbol{P}}_s^{\mathrm{H}}(f) \tag{12}$$

where $\tilde{\boldsymbol{P}}_s(f)$ is the complex sound pressure corresponding to the sth snapshot. For the covariance matrix, the real part and imaginary part are taken out, and a 21×21×2 input sample is obtained in parallel.

In this study, 1 s-snapshot is selected for signal feature extraction. Theoretically, shorter snapshot can be selected for training, so that parameter N of (9) can be set to a larger value and the number of samples is larger. However, in actual training process, the 1 s-snapshot contains obviously more feature information than the shorter snapshot. Therefore, selecting a shorter snapshot may not necessarily improve the estimation performance. For the above reasons, in the field of underwater sound source localization, the 1 s-snapshot is widely used for signal feature extraction.

4.2 Measurement standards

To compare the positioning performance of different sound source ranging methods, mean absolute percentage error (MAPE) is defined as

$$\mathrm{MAPE}=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{R_{gi}-R_{ti}}{R_{ti}}\right| \tag{13}$$

where R_{gi} is the neural network prediction data, and R_{ti} is the actual data.

4.3 Design of transfer learning model

The transfer learning model can make full use of a large number of underwater acoustic data sets T1 in the preselected sea area and a few-shot underwater acoustic data set T2 in the detection sea area to estimate the target range of the detection sea area test set T3, which is divided into the following steps.

(i) A large amount of underwater acoustic data are collected from the preselected sea area and a few shots of underwater acoustic data from the test sea area. For the collected time-domain sound pressure, the frequency-domain complex sound pressure is extracted by fast Fourier transform (FFT), and the norm is normalized.

(ii) A traditional CNN is built, which is used to pretrain the training set containing a large amount of underwater acoustic data in the preselected sea area. The characteristics $\tilde{\boldsymbol{P}}_L^{\mathrm{T1}}$ of T1 data set are input, and the pre-training model is trained by CNN.

(iii) A transfer learning model is built based on the traditional CNN, and the weights of the convolution layer and pooling layer of the pre training model are retained, that is, its convolution layer and pooling layer are frozen as the transfer layer. The weight of the full connection layer and the output layer are adjusted as the adjustment layer, and a new network is built from the transfer layer and the adjustment layer.

(iv) Using the newly built network, the data set with only few shots of underwater acoustic data in the detection sea area is retrained, and the characteristics $\tilde{\boldsymbol{P}}_L^{\mathrm{T2}}$ of T2 data set are input, so that the full connection layer and output layer obtain new weights.

(v) The trained transfer learning model is applied to the range estimation of the test sea area, the characteristics $\tilde{\boldsymbol{P}}_L^{\mathrm{T3}}$ of T3 data set are input, the estimated value of the output range is counted, and the estimation performance of the model is analyzed.

5. Verification of sea trial data

5.1 Experimental description

In this paper, the SWellEX-96 experimental S5 voyage is used. The environmental parameters of the sea trial are shown in Fig. 5. The map of the experimental S5 voyage is shown in Fig. 6. The blue line in the figure represents the track of the target sound source. The data received by S5 vertical array in this experiment were used for network training and testing, and to study the range estimation under the conditions of no strong interference and strong interference, respectively. The experimental ship towed two sound sources: deep (54 m) and shallow (9 m). The shallow sound source was selected. The shallow sound source emitted continuous wave (CW) signals at multiple frequency points between 109 Hz and 385 Hz. The speed of the experimental ship was about 2.5 m/s. A total of 75 min of data were collected in S5 voyage. The experiment used a vertical array with 22 hydrophones, the sampling frequency was 1.5 kHz, and the depth was 94.125−212.25 m. In the experiment, one hydrophone failed, so only the measurement data of the 21 other hydrophones can be used [31].

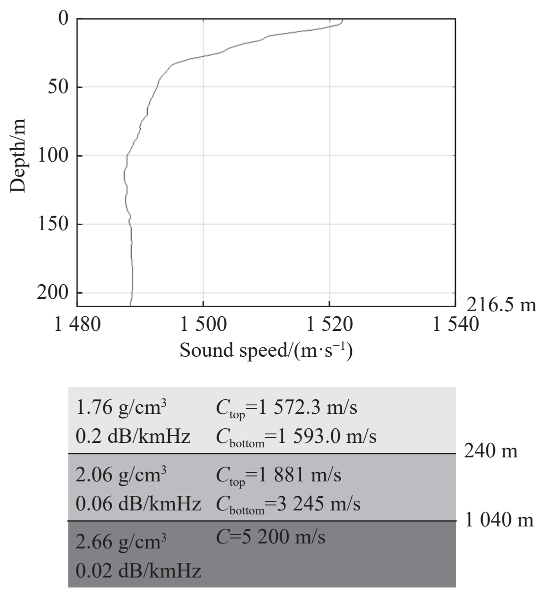


Fig. 5 SWellEX-96 experiment shallow water environmental parameter model [31]

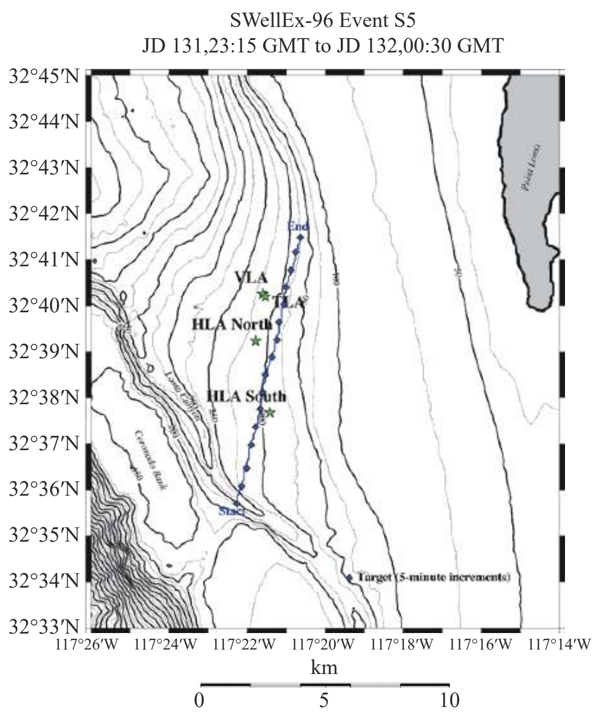


Fig. 6 SWellEX-96 experiment S5 voyage map [31]

5.2 Experimental data processing and result analysis

5.2.1 Selection of data set

In the SWellEX-96 experiment, 232 Hz and {163, 232, 385} Hz were selected as the narrowband and broadband frequencies, respectively. T1, T2, and T3 data sets were selected as pre-training data sets, few-shot data sets, and test sets, respectively. The distribution of each data set is shown in Fig. 6 and Fig. 7. In the 40−60 min of the experiment, 1 s-snapshot data of corresponding frequency were extracted every 1 s, and 1 200 groups of data were obtained. As a large number of underwater acoustic data sets were collected in known sea areas, they were used for pre-training and set as T1 data set. In the 60−75 min of the experiment, a group of 1 s-snapshot data was extracted every 29 s as a few-shot underwater acoustic data set of the test sea area for retraining. A total of 31 groups were set as T2 data set. During this period, a group of 1 s-snapshot data was extracted every 5 s (to avoid the leakage of the test set, the sample points coincident with the few-shot data set were removed) as a test set of the test sea area for testing the model. A total of 180 groups were set as T3 data set.

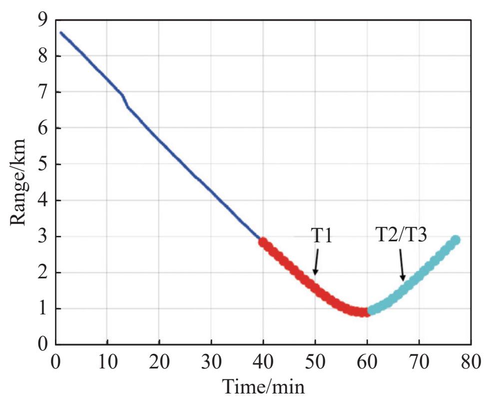
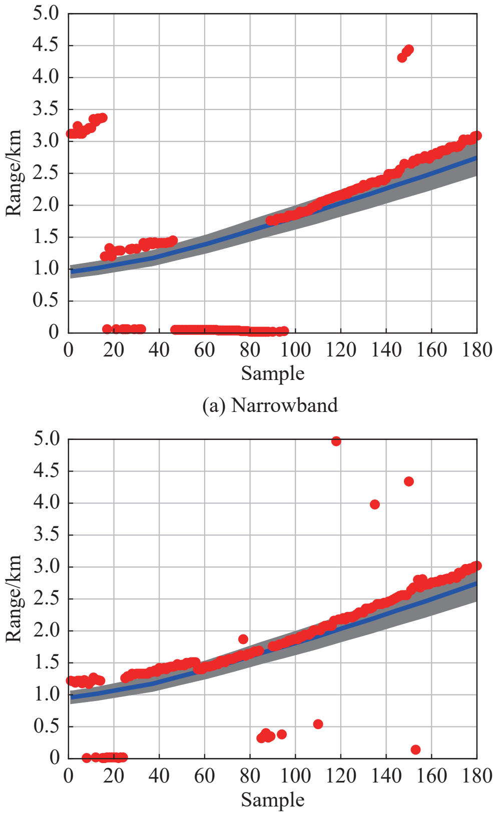


Fig. 7 Data set distribution

5.2.2 MFP

Fig. 8(a) and Fig. 8(b) show the MFP range estimation results of narrowband 232 Hz and broadband {163, 232, 385}Hz signals, respectively. The results show that the MFP method is prone to environmental mismatch. For narrowband signals, the test samples with a range of 1.5−2.0 km have a large error. For broadband signals, the range estimation can be roughly realized, but the estimation results of a large number of test samples below 1.5 km exceed the error limit.



(a) Narrowband

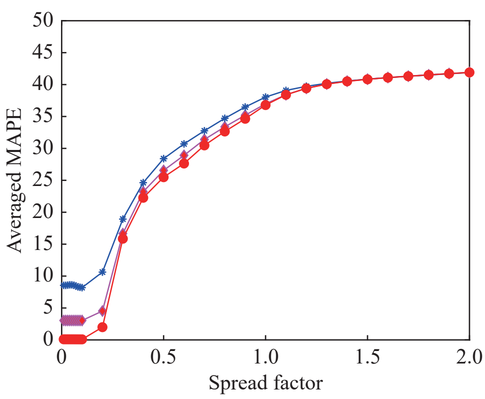
(b) Broadband

: True ranges; ● : Predicted ranges.

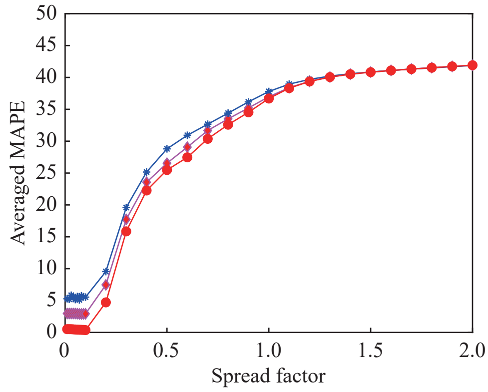
Fig. 8 Range estimation results of MFP method

5.2.3 GRNN

The GRNN method has only one network parameter, namely the spread factor. The selection range of the spread factor is set to $[0.01:0.01:0.1, 0.2:0.1:2.0]$. The 10-fold cross validation method is used to determine the optimal spread factor respectively [8]. T1, T2, and T1/T2 data sets are used as training sets. Fig. 9 shows the determination process of the optimal spread factor of the three training sets under narrowband and broadband signals. The GRNN model corresponding to the obtained optimal spread factor is used for the range estimation of the test set.



(a) Narrowband

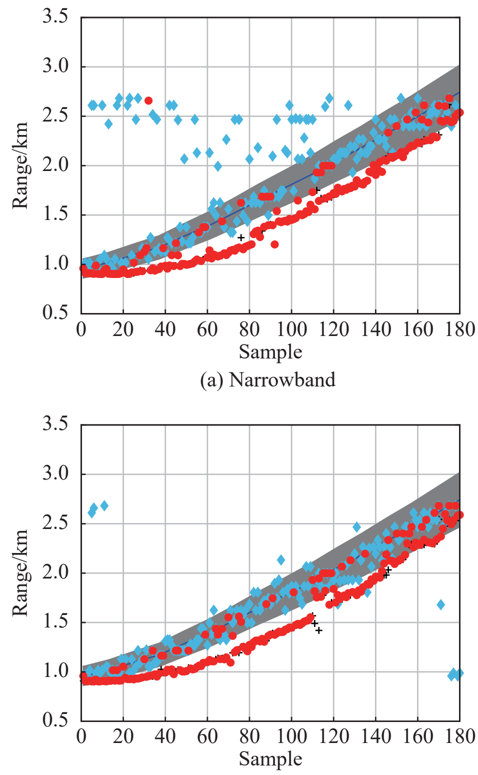


(b) Broadband

: T1; : T2; : T1/T2.

Fig. 9 Determination process of the optimal spread factor of the three training sets

Fig. 10 shows the range estimation results of GRNN method. When the training sets are T1 and T1/T2, the estimation results of test samples of narrowband and broadband signals mostly exceed the error limit, and compared with the training set T1, there are more estimation results of the training set T1/T2 in the error limit. When the training set is T2, the error of estimation results of many test samples for narrowband signals is large, and for broadband signals, the range estimation can be roughly realized, but there are still a small number of estimation results with large errors.



(a) Narrowband

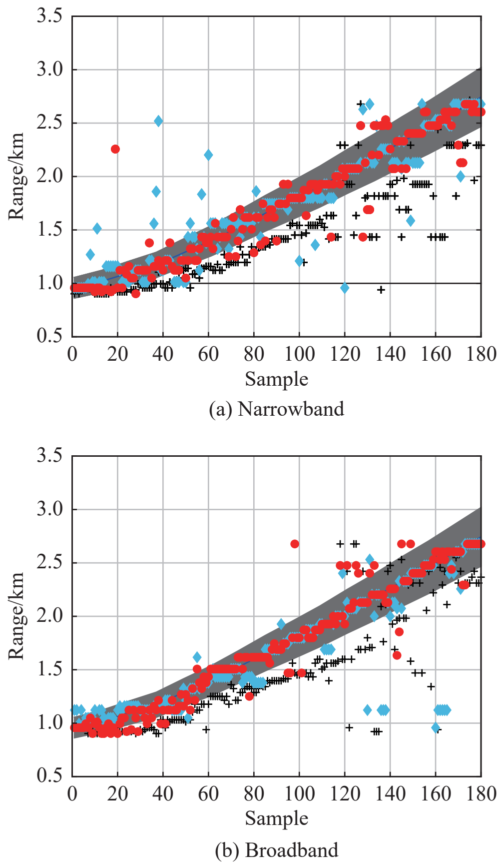
(b) Broadband

: True ranges; + : GRNN T1; ◆ : GRNN T2; ● : GRNN T1/T2.

Fig. 10 Range estimation results of GRNN method

5.2.4 Traditional CNN

Using the traditional CNN training method, when the training sets were T1 data set, T2 data set, and T1/T2 data set, Fig. 11 shows the estimation results of the T3 data set under the three training sets, respectively. In the environment without strong interference, when the training set was T1, the neural network model did not easily predict the test set due to the large range difference between the training set and the test set, so the effective range estimation cannot be realized in this scene. When the training set was T2, because the training set and the test set were in the same sea area, the range estimation can be realized to a certain extent, but due to the small amount of data in the training set, the error was still large. When the training set was the mixed data set of T1 and T2, because the mixed data set not only had a large amount of data but also contained the T2 data set in the same sea area as the test set, it can better realize the range estimation, but a certain error was still observed.

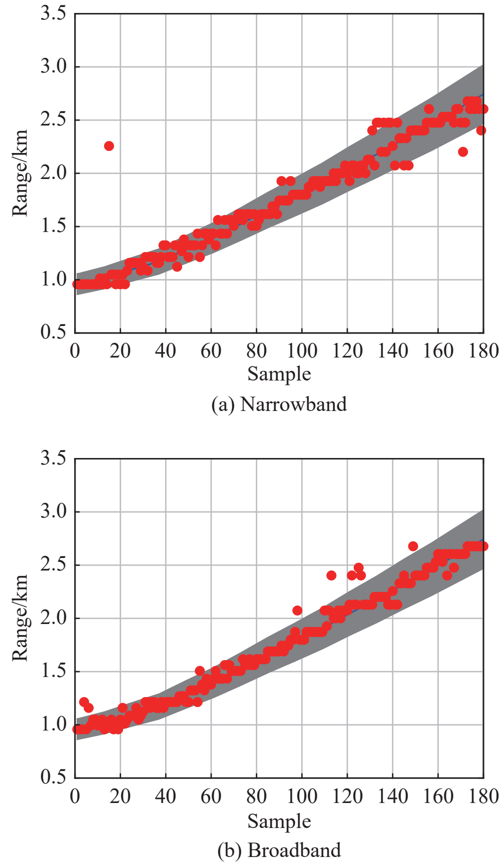


—— : True ranges; + : CNN T1; ◆ : CNN T2; ● : CNN T1/T2.

Fig. 11 Range estimation results of traditional CNN

5.2.5 Transfer learning

After the training of the transfer learning model, the range estimation results are shown in Fig. 12. Compared with MFP, GRNN, and traditional CNN methods, transfer learning can realize the target range estimation more accurately.



—— : True ranges; ● : Predicted ranges.

Fig. 12 Range estimation results of transfer learning

Table 1 shows the MAPE of the range estimation results of narrowband and broadband signals under various methods. According to the horizontal comparison, the MAPE of the range estimation results under the transfer learning method is much smaller than that of traditional CNN, GRNN, and MFP, and the estimation performance of transfer learning is substantially better than that of other methods. Among these methods, the estimation results of the MFP method has the largest error. The GRNN method can only realize approximate estimation of range, and the GRNN method can only realize approximate estimation of range, and the range estimation result of training set T1/T2 is better than that of training set T1 and T2. Compared with GRNN model, ResNet18 model can extract deeper feature information, so CNN method has better estimation performance. Compared with the traditional CNN whose training set is T1, the estimation performance of the traditional CNN using T2 data set is better. It can be obtained that the smaller the position difference between the training set and the test set, the better the estimation performance. If there are a large number of samples in the T2 data set, the traditional CNN can also obtain more accurate range estimation results for the test set in theory, but in the few-shot scenario, the effective range estimation cannot be realized through the traditional CNN. Compared with the traditional CNN method with only T1 or T2 training set, the traditional CNN method with T1/T2 training set has a large amount of data, so it can be trained more fully. Therefore, its estimation performance is the best among the traditional CNN methods with various training sets. The traditional CNN method with T1/T2 training set and the transfer learning method have the same amount of data used for training, but the former is to train and estimate in a large range, whereas the latter is to retrain the few-shot sea area under the condition of similar model weights that have been trained, that is, on the basis of making full use of the underwater acoustic data of the nearby sea area, focusing on training and accurate estimation in a small range for the few-shot sea area containing the test set. Therefore, compared with the traditional CNN method with T1/T2 training set, the estimation result of transfer learning method is better. According to the longitudinal comparison, compared with the narrowband signal, the wideband signal provides more characteristic information, so the range estimation result of the broadband signal is better.

Table 1 MAPE of range estimation results of narrowband and broadband signals using different methods

Method	MFP	GRNN			CNN			Transfer learning
		T1	T2	T1/T2	T1	T2	T1/T2	
Narrowband	55.6925	19.0144	12.9857	10.3732	17.5382	7.5285	5.8026	4.0886
Broadband	20.2007	17.4170	10.3911	8.1141	16.7749	6.9176	5.0352	2.9136

Taking broadband signal as an example, this study also analyzes the calculation time of different methods. The methods were developed on a workstation with the 11th Generation Intel(R) Core(TM) i7-1165G7 CPU*8. The calculation time of traditional MFP method is 9.9 s, and the training and testing time of different machine learning methods are shown in Table 2. The results show that the training time of CNN method is relatively long compared with GRNN in the training phase, but the difference of testing time of different methods is small in the testing phase. Although the training time of transfer learning method is long, it can achieve more accurate estimation, and for the trained model, it can achieve effective estimation of test samples in a short time.

Table 2 Training and testing time of different machine learning methods

Method	GRNN			CNN			Transfer learning
	T1	T2	T1/T2	T1	T2	T1/T2	
Training time/s	20.5	9.7	29.2	50.8	32.9	89.3	95.2
Testing time/s	2.3	2.1	2.4	2.8	2.5	3.1	2.9

6. Conclusions

The combination of machine learning in the field of underwater acoustic passive positioning is becoming closer. When only few shots of underwater acoustic data are in the test sea area, transfer learning can be used as a learning framework to apply the existing knowledge to the new environment. In this paper, a transfer learning model based on residual CNN is proposed, and the method is verified by using the S5 voyage data of SWellEX-96 experiment. The performances of underwater target range estimation of MFP, GRNN, traditional CNN, and transfer learning are compared. The experimental data processing results show that in the few-shot scenario, the MFP method is not applicable in the environment with a range of 0−2 km. GRNN method can only realize approximate estimation of range. The traditional CNN method has a good positioning only when the training set is T1/T2, whereas the transfer learning method can achieve robust range estimation. The transfer learning model has less research in the field of underwater acoustic passive positioning and has a large research space. The target range estimation under different types of interference and different depths of test set and training set needs to be further explored.

References

[1] BUCKER H P. Use of calculated sound fields and matched-field detection to locate sound sources in shallow water. The Journal of the Acoustical Society of America, 1976, 59(2): 368–373.

[2] KLEMM R. Range and depth estimation by line arrays in

shallow water. Signal Processing, 1981, 3(4): 333–344.
[3] BAGGEROER A B. Matched field processing: source localization in correlated noise as an optimum parameter estimation problem. The Journal of the Acoustical Society of America, 1988, 83(2): 571–587.
[4] MICHALOPOULOU Z H, PORTER M B. Matched-field processing for broad-band source localization. IEEE Journal of Oceanic Engineering, 1996, 21(4): 384–392.
[5] SOARES C, JESUS S M. Broadband matched-field processing: coherent and incoherent approaches. The Journal of the Acoustical Society of America, 2003, 113(5): 2587–2598.
[6] YANG K D, MA Y L, ZOU S X, et al. Linear matched field processing based on environmental perturbation. Acta Acustica, 2006, 31(6): 496–505. (in Chinese)
[7] STEINBERG B Z, BERAN M J, CHIN S H, et al. A neural network approach to source localization. The Journal of the Acoustical Society of America, 1991, 90(4): 2081–2090.
[8] NIU H Q, OZANICH E, GERSTOFT P. Ship localization in Santa Barbara Channel using machine learning classifiers. The Journal of the Acoustical Society of America, 2017, 142(5): EL455–EL460.
[9] WANG Y, PENG H. Underwater acoustic source localization using generalized regression neural network. The Journal of the Acoustical Society of America, 2018, 143(4): 2321–2331.
[10] FERGUSON E L, WILLIAMS S B, JIN C T. Sound source localization in a multipath environment using convolutional neural networks. Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2018: 2386−2390.
[11] LIU Y N, NIU H Q, LI Z L. Source ranging using ensemble convolutional networks in the direct zone of deep water. Chinese Physics Letters, 2019, 36(4): 044302.
[12] NIU H Q, GONG Z X, OZANICH E, et al. Deep-learning source localization using multi-frequency magnitude-only data. The Journal of the Acoustical Society of America, 2019, 146(1): 211–222.
[13] HOWARTH K, KOMEN D, NEILSEN T B, et al. Effect of signal to noise ratio on a convolutional neural network for source ranging and environmental classification. The Journal of the Acoustical Society of America, 2019, 146(4): 2961–2962.
[14] KOMEN D, NEILSEN T B, KNOBLES D P, et al. A convolutional neural network applied to measured time series for source range and ocean seabed classification. The Journal of the Acoustical Society of America, 2019, 146(4): 2930.
[15] KOMEN D, NEILSEN T B, HOWARTH K, et al. Seabed and range estimation of impulsive time series using a convolutional neural network. The Journal of the Acoustical Society of America, 2020, 147(5): EL403–EL408.
[16] OZANICH E, GERSTOFT P, NIU H Q. A feedforward neural network for direction-of-arrival estimation. The Journal of the Acoustical Society of America, 2020, 147(3): 2035–2048.
[17] LIU Y N, NIU H Q, LI Z L. A multi-task learning convolutional neural network for source localization in deep occan. The Journal of the Acoustical Society of America, 2020, 148(2): 873–883.
[18] CHEN R, SCHMIDT H. Model-based convolutional neural network approach to underwater source-range estimation. The Journal of the Acoustical Society of America, 2021, 149(1): 405–420.
[19] WEISS K, KHOSHGOFTAAR T M, WANG D D. A survey of transfer learning. Journal of Big Data, 2016, 3(1): 1–40.
[20] DAI W Y, CHEN Y Q, XUE G R, et al. Translated learning: transfer learning across different feature spaces. Proc. of the Advances in Neural Information Processing Systems, 2008: 353−360.
[21] ZHU Y, CHEN Y Q, LU Z Q, et al. Heterogeneous transfer learning for image classification. Proc. of the 25th Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, 2011: 1304−1309.
[22] NAKAMURA A, HARADA T. Revisiting fine-tuning for few-shot learning. https://arxiv.org/abs/1910.00216.
[23] JANG Y H, LEE H K, HWANG S J, et al. Learning what and where to transfer. Proc. of the International Conference on Machine Learning, 2019: 3030–3039.
[24] GERACE F, SAGLIETTU L, MANNELLI S S, et al. Probing transfer learning with a model of synthetic correlated datasets. Machine Learning: Science and Technology, 2022, 3(1): 015030.
[25] WANG W B, NI H Y, SU L, et al. Deep transfer learning for source ranging: deep-sea experiment results. The Journal of the Acoustical Society of America, 2019, 146(4): EL317–EL322.
[26] CAO H G, WANG W B, SU L, et al. Deep transfer learning for underwater direction of arrival using one vector sensor. The Journal of the Acoustical Society of America, 2021, 149(3): 1699–1711.
[27] PORTER M B. The KRAKEN normal mode program. Washington D.C.: Naval Research Lab, 1992.
[28] YANG K D. The matched field processing of underwater acoustic array signals. Xi'an: Northwestern Polytechnical University Press, 2008. (in Chinese)
[29] FUKUSHIMA K, MIYAKE S, ITO T. Neocognitron: a neural network model for a mechanism of visual pattern recognition. IEEE Trans. on Systems, Man, and Cybernetics, 1982, 13(5): 826–834.
[30] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770−778.
[31] MUARRY J, ENSBERG D. The swellex-96 experiment. http://www.swellex96.ucsd.edu/.

Biographies

YAO Qihai was born in 1997. He received his Master's degree in ship and ocean engineering from Northwestern Polytechnical University in 2022. He is pursuing his Ph.D. degree at Northwestern Polytechnical University. In 2020, he completed an exchange project of the Oxford prospects program of the University of Oxford, UK, and obtained an excellent graduation certificate. His main research direction is the application of machine learning in array signal processing.

E-mail: 2019260659@mail.nwpu.edu.cn

WANG Yong was born in 1987. He received his B.S. degree in applied electronic engineering and M.S. and Ph.D. degrees in underwater acoustic engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2009, 2011, and 2015, respectively. From 2016 to 2017, he was a post-doctoral fellow with Xi'an Jiaotong University, Xi'an, China. He is currently an associate professor with the School of Marine Science and Technology, NPU. His research interests include array signal processing, parameter estimation, and sonar signal processing.
E-mail: yongwang@nwpu.edu.cn

YANG Yixin was born in 1975. He received his B.S. degree in applied electronic engineering and M.S. and Ph.D. degrees in underwater acoustic engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 1997, 1999, and 2002, respectively. From June 2002 to June 2004, he was a research fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Since July 2004, he has been with the School of Marine Science and Technology, NPU, where he became a professor in 2006. He is currently the Vice President of NPU, the Chair of the Underwater Acoustics Committee of the Acoustical Society of China, and a member of the Acoustical Society of America. His research interests include acoustic array signal processing, spectral estimation, and their applications.
E-mail: yxyang@nwpu.edu.cn