Applied Acoustics 211 (2023) 109513

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A deep-learning based high-gain method for underwater acoustic signal detection in intensity fluctuation environments

Hailun Chu [a,b,c], Chao Li [a,b], Haibin Wang [a,b,*], Jun Wang [a,b], Yupeng Tai [a,b], Yonglin Zhang [a,b], Fan Yang [c], Yannick Benezeth [c]

[a] State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] Laboratory ImViA, Universite Bourgogne Franche-Comte, Dijon 21078, France

ARTICLE INFO

Article history:
Received 14 February 2023
Received in revised form 12 June 2023
Accepted 24 June 2023
Available online 12 July 2023

Keywords:
Low signal-to-noise ratio
Intensity fluctuation
Deep learning
Active signal detection

ABSTRACT

To achieve efficient underwater acoustic (UWA) signal detection in low signal-to-noise ratio (SNR) scenarios, the transmitted signals are designed with a large time-bandwidth product to get a high detection gain. The linear correlator (LC) is considered as the maximum SNR detector, whose detection gain is proportional to the time-bandwidth product. However, the detection performance of LC degrades significantly in time-variant multipath UWA channel and non-Gaussian UWA ambient noise. In this study, we present a deep-learning based two-stage UWA signal detection method in intensity fluctuation environments. This method takes the advantages of the Conv-TasNet and Encoder-Decoder network, which utilizes an encoder module to extract signal features, a separation module to enhance the signal components and then another decoder module to reconstruct the transmitted signal. To demonstrate the performance of the proposed method, the datasets used for training and testing originated from the ASIAEX 2001 South China Sea (SCS) experiment. The experimental results show that our model outperforms the classical LC and channel estimation based LC (CE-LC) in constant false alarm rate (CFAR) detection and also surpasses the TCDAE and Conv-Tasnet models as evaluated by DSI-SNR1 and DSI-SNR2.

1. Introduction

The rapid development of marine resources has led to increased interest in active signal detection within complex underwater acoustic (UWA) environments. Active signal detection has a wide range of applications, including echo detection, target navigation, communication system wake-up, and time synchronization. To achieve high-gain UWA signal detection, a large time-bandwidth product signal is necessary. However, due to the strong absorption of high-frequency sound, the bandwidth available is limited. Therefore, the practical approach is to maximize the transmission signal duration to increase the gain in time domain.

Linear correlator (LC) is the most common detector in UWA detection system, which is known as the maximum signal-to-noise ratio (SNR) detector [1]. It assumes that the UWA ambient noise follows a Gaussian distribution, is uncorrelated, and the UWA channel is time-invariant with a single path. Obviously, these assumptions are unreasonable in real UWA environment. Besides, the UWA channel becomes more complex as the transmission signal duration increases, which results in the significant performance degradation of LC. The complexity is mainly reflected in three aspects: UWA ambient noise, multipath propagation and Doppler effect.

UWA ambient noise is of impulsive property, which can be caused by marine life [2] and human activities [3]. This means that large amplitude values occur frequently in time domain and manifests heavy tail property in probability density function (PDF) compared to the Gaussian noise. The performance of LC is negatively impacted by this property. To be precise, the symmetric alpha-stable ($S\alpha S$) distribution is adopted to represent the UWA noise [4]. However, the PDF of $S\alpha S$ has no closed form expression except for some cases, such as $\alpha = 2$ and $\alpha = 1$ corresponding to Gaussian and Cauchy cases, respectively.

Based on its heavy tail property, Chitre proposed the maximum-likelihood (ML) and locally optimal (LO) detectors with the detailed knowledge of noise probability distribution [4]. To reduce the computational complexity, Zozor used some particular properties of $S\alpha S$ to derive a parametric suboptimal detector without explicit PDF [5]. Chitre further compared the performance of different approximate functions of PDF [6]. Parametric detectors face

* Corresponding author at: State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: whb@mail.ioa.ac.cn (H. Wang).

https://doi.org/10.1016/j.apacoust.2023.109513
0003-682X/

challenges when applied to UWA noise due to its non-Gaussian and non-stationary properties, especially for long-duration signals, which can make it difficult to obtain a reliable noise distribution estimate [7]. To avoid the numerical integration computation of PDF, some non-parametric transfer functions are applied to LC, for example, the sign function which can improve its applicability in UWA noise [4, 8]. These detectors suppress the large amplitude signals to mitigate negative impacts of impulsive noise. It is clear that the performance of non-parametric detectors is inferior to that of parametric detectors.

Additionally, UWA channel is characterized as a double-selective fading channel [9], which refers to the multipath propagation and Doppler effect. The multipath propagation is caused by sea-surface and sea-bottom reflection. Due to the multipath propagation, signal energy is dispersed and pulse duration is extended. To concentrate the multipath energy, channel estimation methods are applied to acquire the multipath structures. Then, time reversal mirror (TRM) is used to achieve adaptive focusing [10–12]. Tian used robust orthogonal matching pursuit to get multipath information and improved the detection performance in multipath environment affected by impulsive noise [13]. With the help of channel estimation methods, the performance of LC is improved by utilizing multipath information in multipath environment. However, these channel estimation based methods may fail in low SNR scenes and lead to noise amplification.

Another challenge for signal detection is Doppler effect, which leads to the time-varying impulse response by platform motion and distortion of the propagation medium [14] [15]. As for linear platform motion, the Doppler shift can be corrected by moving target indicator (MTI) when there is no range migration [16]. When range migration occurs, it can be compensated using Keystone Transform [17] and Radon-Fourier Transform [18]. For more complex target motion model, Lin proposed a slow time reverse based method [19]. These methods are widely applied in radar system under the assumption of narrowband signal. Zhang formed a multi-pulse long-time integrated model and generalized it to the UWA wideband signal [20]. The prior knowledge of motion model is required in the above methods. When the motion model mismatches, the performance is deteriorated.

Another type of Doppler effect, resulting from propagation medium distortion, is typically less pronounced than those caused by platform motion, but it tends to be more erratic. The difficulty brought to signal detection is the decrease of temporal coherence. This physical phenomenon was observed and measured in shallow and deep water [21–24]. Phase and delay-locked loop structures are used to compensate this Doppler distortion with preliminary pilot synchronization in UWA communication [25], which is impossible in UWA active signal detection as the presence of interested signals is undetermined. Furthermore, the target motion model based methods mentioned previously seem not applicable for suppressing the random range migration as the motion model is unpredictable in this case.

To confront the channel fading and cope with impulsive noise, some transfer functions are utilized to enhance the received signals. In the presence of impulsive noise, the transfer functions are designed to suppress the large amplitude signals. For UWA multipath channel, the transfer functions are TRM with various channel estimation based methods. For Doppler effect, the transfer function is to compensate the range migration. However, the conventional detectors' performance suffers degradation in ASIAEX 2001 SCS environment due to the observed decrease in temporal coherence as reported in [23,24]. This decrease is especially severe on linear correlator.

In general, traditional techniques primarily concentrate on the present data, whereas deep learning methods can leverage historical data to learn. As a result, the performance of conventional methods may decline if the current samples do not align with the templates, such as the transmitted signal and motion model, particularly in a time-varying environment. However, deep learning based methods are considered to show promising performance in complex environment with its data-driven function expression. Upon completing the learning process, deep learning-based methods can compensate for template mismatch, which in turn leads to high-gain detection capabilities. In order to employ deep learning-based methods for active signal detection, the method must be capable of distinguishing between the signal and ambient noise, as well as dealing with channel fading. Convolutional neural networks (CNNs) have strong capability of extracting features. Stacked convolutional denoising Auto-Encoders are proposed for unsupervised training [26]. It can transform the raw data into high-level feature representations as well as data compression. Zhang utilizes the stacks of convolutional neural network with skip connections for UWA communication [27]. It has been shown that the stacked CNN can effectively extract promising features and achieve signal recovery. However, the temporal dependence in time domain is not efficiently captured for CNNs and stacked CNN. To get a more larger receptive field, a temporal convolutional neural network (TCNN) is proposed for real-time speech enhancement [28]. In many underwater application scenarios and early speech enhancement, deep learning based methods are implemented in the time–frequency domain. But the features in frequency domain are not enough for active signal detection. Conv-Tasnet is proposed to model the signal directly in time domain [29]. This model utilizes a 1-D convolutional layer and deconvolutional layer as the encoder and decoder modules to replace short-time Fourier transform (STFT) and inverse STFT, respectively. Then, masks for speaker separation are constructed by using TCNN which consists of several stacked dilated convolutional blocks. However, the feature extraction capability is not sufficient without considering the channel distortion.

In this study, we consider the active signal passing through UWA channels as the interfered signal with ocean noise, multipath propagation and Doppler effect. We resolve the detection problem in a two-stage way. At first, we apply a temporal convolutional separation network for signal reconstruction in time-domain. Afterwards, decide whether the signal is present or not by utilizing LC and channel estimation-based LC. The network consists of three modules: an encoder, a separator, and a decoder. The encoder is designed with stacked convolutional blocks used to extract higher-dimensional features and compress time-domain information. The separator comprises stacked dilated convolutional blocks aiming to form an efficient mask based on the compressed features. Then, the decoder can be applied to achieve time-domain signal reconstruction, which has an inverse architecture compared with the encoder. The loss function is the scale-invariant signal-to-noise ratio (SI-SNR) in this network.

The main contributions of this study are summarized as follows:

1 The deep-learning based separation network is introduced into UWA active signal detection. The signal detection problem is to separate the signal from ambient noise, which is fitting to the separation network.
2 The SI-SNR is chosen to replace the mean squared error (MSE) as the loss function in UWA active signal detection. The SI-SNR loss function is equivalent to correlation coefficient [30] which is our criteria for characterizing temporal coherence and signal recognition.
3 The deep-learning structure is utilized to suppress the random range migration caused by the distortion of the propagation medium. The detrimental influence, caused by this

random medium distortion, can be mitigated by compressing the time domain information.
4 These detectors are represented in a uniform form and the performance of these detectors is compared. These detectors differ in their corresponding transfer functions.

This paper is organized as follows. In Section 2, the conventional UWA active signal detection system is introduced. Section 3 presents the proposed deep-learning based two stage UWA active signal detection method. The performance of the proposed method is shown in Section 4. Finally, Section 5 concludes the paper.

2. UWA active signal detection system

In this section, the conventional UWA active signal detection system is introduced, and several conventional detectors are compared. The schematic is shown in Fig. 1. The transmitted signal propagates through the UWA channel to the receiving end with ambient noise. In the receiving end, the presence of signal or not is decided by the detector.

2.1. Underwater acoustic signal and channel model

The transmitted signal consists of M adjacent phase modulated signals, which can be defined as follows:

$$\begin{aligned} s_1(n) &= A\cos(\phi(n)) \quad 0 \leqslant n \leqslant N-1 \\ &= 0 \quad elsewhere \end{aligned} \tag{1}$$

$$s(n) = \sum_{m=0}^{M-1} s_1(n - mN)$$

where A is the magnitude, $\phi(n)$ is the instantaneous phase, and N is the signal duration in the samples for each phase modulated signal. For m-sequence, the instantaneous phase is defined as:

$$\phi(n) = 2\pi f_0 n T_s + \phi_0 m_{seq}(n) \tag{3}$$

where f_0, T_s ϕ_0 and m_{seq} are the center frequency, sample interval, phase angle, and m-sequence, respectively..

Given the doubly-selective characteristic of UWA channel, we can denote the UWA channel impulse response as

$$h(n;k) = \sum_{l=0}^{L-1} A_l(n)\delta(k - k_l(n)) \tag{4}$$

where L is the number of paths, n and k are the index of time and time delay, $A_l(n)$ and $k_l(n)$ denote the gain and time delay of the l-th path at sample n, respectively.

After passing through the UWA channel, the received signal can be expressed as

$$r(n) = \sum_{k=0}^{K-1} s(n-k)h(n;k) + w(n) \tag{5}$$

where K is the length of channel and $w(n)$ is the ambient noise.

Our aim is to determine whether the transmitted signal is received. Then, the detection problem can be formulated as a hypothesis-testing problem with two cases where hypothesis H_1 and H_0 mean that the transmitted signal is present or not.

$$r(n) = \begin{cases} w(n) & , if\ H_0 \\ \sum_{k=0}^{K-1} s(n-k)h(n;k) + w(n) & , if\ H_1 \end{cases} \tag{6}$$

2.2. Conventional underwater acoustic detectors

The test statistic for the LO detector can be given by [4]:

$$\Lambda(r) = \sum_{n=0}^{MN} g[r(n)]s(n) \tag{7}$$

where $g(\cdot)$ is the transfer function, which is determined by the distribution of ambient noise and also the UWA channel. When $g(\cdot)$ is a linear function, that is $g(x) = ax$, the test statistic can be simplified as LC detector, which is optimal in the presence of Gaussian noise. Obviously, the performance of LC detector will degrade in the presence of non-Gaussian noise. In the following, we will derive several detectors and explain their relationship.

To simplify the notation, the UWA channel impulse response is rewritten as $h(k)$ and zero-padded to length N, wherein the time-invariant channel is assumed. Then Eq. 5 can be expressed as

$$\boldsymbol{r} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{w} \tag{8}$$

where the signal matrix can be written as

$$\boldsymbol{S} = \begin{bmatrix} s(0) & s(N-1) & \cdots & s(1) \\ s(1) & s(0) & \ddots & s(2) \\ \vdots & \ddots & \ddots & \vdots \\ s(N-1) & s(N-2) & \cdots & s(0) \end{bmatrix} \tag{9}$$

Algorithm 1 OMP

Require: Received signal $\boldsymbol{r}$, dictionary matrix $\boldsymbol{S}$, iterations I
1: Initialization: iteration $i \leftarrow 0$ residual $\boldsymbol{e}_0 \leftarrow \boldsymbol{y}$ atom index set $\Lambda_0 \leftarrow \varnothing$
2: **while** $i < I$ **do**
3: Calculate inner product between residual and each atom, and then select the atom corresponding to the largest product, i.e., $j \leftarrow argmax|\boldsymbol{s}_j^T \boldsymbol{e}_i|$
4: Update the atom index set, i.e. $\Lambda_{i+1} \leftarrow \Lambda_i \cup j$
5: Calculate the channel estimation $\hat{\boldsymbol{h}} \leftarrow \boldsymbol{S}_{\Lambda_{i+1}}^{-1}\boldsymbol{r}$
6: Update the residual $\boldsymbol{e}_{i+1} \leftarrow \boldsymbol{y} - \boldsymbol{S}_{\Lambda_{i+1}}\hat{\boldsymbol{h}}$
7: $i \leftarrow i+1$
8: **end while**
9: Outputs: the channel estimation $\hat{\boldsymbol{h}}$ and the reconstructed signal $\boldsymbol{S}_{\Lambda_i}\hat{\boldsymbol{h}}$

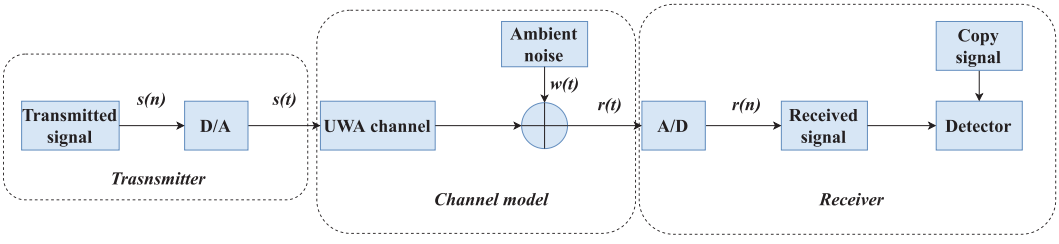

Fig. 1. The UWA active sonar detection system.

The output of LC can be expressed as

$$\boldsymbol{z}_{LC} = \boldsymbol{S}^T \boldsymbol{r} \tag{10}$$

where the vectors $\boldsymbol{r} = [r(0), \cdots, r(N-1)]^T, \boldsymbol{h} = [h(0), \cdots, h(N-1)]^T$, and $\boldsymbol{w} = [w(0), \cdots, w(N-1)]^T$.

The least-squares solution can be expressed as

$$\hat{\boldsymbol{h}}_{LS} = \boldsymbol{S}^{-1}\boldsymbol{r} = \left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\boldsymbol{S}^T\boldsymbol{r} \tag{11}$$

As the auto-correlation function of wideband signal is approximated to Dirac function [31], we can get

$$\hat{\boldsymbol{h}}_{LS} = \frac{1}{\sigma_s^2}\boldsymbol{S}^T\boldsymbol{r} \tag{12}$$

where σ_s^2 is the power of signal. From Eq. 10 and Eq. 12, we can see that LS detector is equivalent to LC detector.

For CE-LC detector, the output can be expressed as

$$\begin{aligned} \boldsymbol{z}_{CE-LC} &= \boldsymbol{r} * \left(\boldsymbol{s} \otimes \hat{\boldsymbol{h}}\right) \\ &= \boldsymbol{r} \otimes \left(\widetilde{\boldsymbol{s} \otimes \hat{\boldsymbol{h}}}\right) \\ &= \left(\boldsymbol{r} \otimes \tilde{\hat{\boldsymbol{h}}}\right) * \boldsymbol{s} \\ &= \boldsymbol{z}_{TRM-LC} \\ &= \boldsymbol{S}^T g_{CE}(\boldsymbol{r}) \end{aligned} \tag{13}$$

where $\tilde{\cdot}$ means time reversal operation, $\otimes$ means convolution operation, and $*$ means correlation operation. Obviously, using the same channel estimation, CE-LC detector is equivalent to TRM-LC detector. In this paper, the expression of CE-LC is adopted.

Given the sparsity of UWA channel, orthogonal matching pursuit (OMP) algorithm is used to estimate the channel. The procedure of sparse channel estimation and received signal reconstruction using OMP is described in Algorithm 1. The OMP algorithm uses partial columns of the dictionary matrix to achieve channel estimation, which can reduce the noise of the channel. In the following sections, we use OMP-LC to represent CE-LC.

According to the above descriptions, all detectors can be represented in a unified form with different transfer functions $g(\cdot)$. For LC detector, the transfer function is a linear function. As for CE-LC detector, the transfer function is a convolution function, which is related to the algorithms of channel estimation. The linear function is easy to implement but performs not well in non-Gaussian noise and double-selective fading channel. The convolution function utilizes the multipath information, but has a noise amplification issue. In the next section, we adopt deep-learning network to get a better transfer function.

3. Deep-learning based UWA active signal detection method

In this section, the architecture of deep-learning based detection method is proposed. As shown in Fig. 2, the proposed method has two stages. The first stage utilizes the deep-learning network to achieve signal enhancement, which consists of three modules-an Encoder, a Separator, and a Decoder. Furthermore, the LC and OMP-LC detector are applied to identify whether the signal is present or not in the second stage.

3.1. Encoder

The received signal is taken as the input of the Encoder. The Encoder is utilized to transform the discrete waveform into a single feature map, which is implemented by a 1-D convolutional block. The detailed architecture of the encoder is shown in Fig. 2(B).

The input is divided into F overlapped frames $\boldsymbol{x}_f \in \mathbf{R}^{L\times 1}$, where L is the length of each frame and $f = 1, 2, \cdots, F$ is the frame index. The stride of the first 1D convolutional is $L/2$, otherwise, the stride is 1. The kernel size is L. Zero padding is adopted in the 1-D convolutional block to ensure the number of frames to be the same. Therefore, the Encoder realizes the information compression and feature extraction. The feature map $\boldsymbol{w}$ can be formulated as:

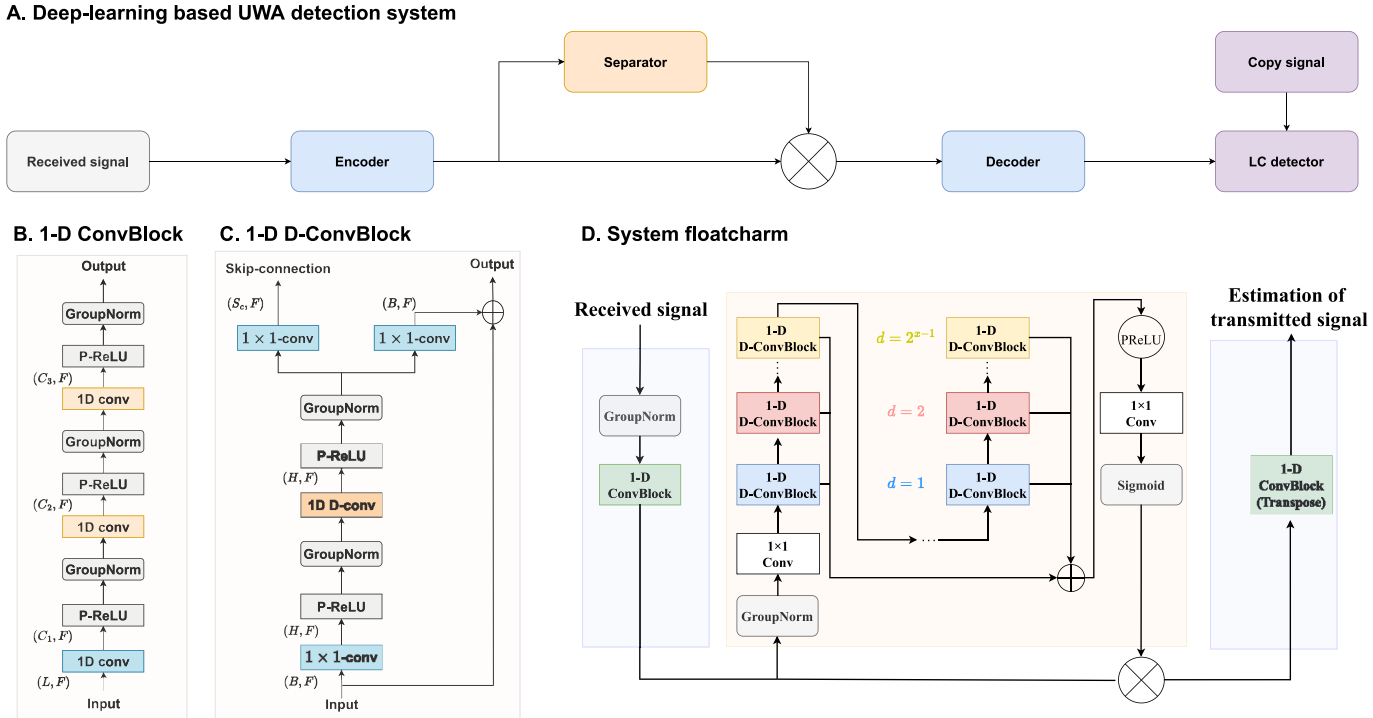

Fig. 2. The deep-learning based two-stage UWA active signal detection system.

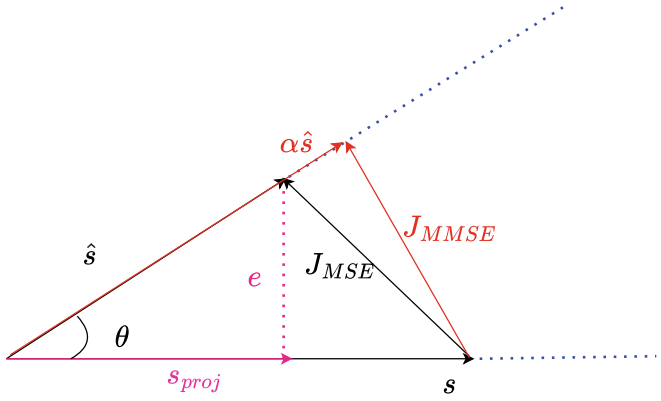

Fig. 3. The diagram of loss function.

$$\boldsymbol{w} = PReLU(\boldsymbol{U}_3 \cdot PReLU(\boldsymbol{U}_2 \cdot PReLU(\boldsymbol{U}_1 \cdot X)))$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_F] \in \mathbf{R}^{L\times F}$ contains all input frames, $\boldsymbol{U}_1 \in \mathbf{R}^{C_1\times L}, \boldsymbol{U}_2 \in \mathbf{R}^{C_2\times C_1}$, and $\boldsymbol{U}_3 \in \mathbf{R}^{C_3\times C_2}$ consist of $C_1, C_2 \times C_1$ and $C_3 \times C_2$ learnable kernels, respectively. Fig. 3.

3.2. Separator

The Separator is used to learn a non-negative mask $\boldsymbol{m}_{sep}$ from the feature map $\boldsymbol{w}$, which is the output of the Encoder. The estimated clean feature map of transmitted signal can be formed as:

$$\boldsymbol{w}_s = \boldsymbol{w} \odot \boldsymbol{m}_{sep} \tag{15}$$

The specific structure of the separator is shown in Fig. 2 (D). The Separator is implemented by the stacked 1-D dilated convolutional blocks (1-D ConvBlock). The number of 1-D Conv-Block is R, which has x layers. The kernel size is 3 in every 1-D ConvBlock. The dilation factors increase exponentially to ensure a sufficiently large reception field to incorporate with the long-range dependence temporal information. Besides, the $1 \times 1 - conv$ block is used at the beginning and the end of the module as a bottleneck layer to connect with the Encoder and Decoder. To keep the frame length the same, zero padding is used in every block whose structure is shown in Fig.2 (C). The residual path of each block serves as the input to the next block, and the skip-connection paths for all blocks are summed up and used as the output of the Separator.

3.3. Decoder

The Decoder is a symmetrical structure of the Encoder, which transforms the clean feature map $\boldsymbol{w}_s$ to a waveform, that is the estimation of transmitted signal. The overlapped frames of the estimation can be formulated as

$$\boldsymbol{Y} = PReLU(\boldsymbol{V}_3 \cdot PReLU(\boldsymbol{V}_2 \cdot PReLU(\boldsymbol{V}_1 \cdot \boldsymbol{w}_s))) \tag{16}$$

Then, we can get the estimation of the transmitted signal $\hat{\boldsymbol{s}}$ which is recovered by removing the overlapped frames.

3.4. Training objective

Here, we choose SI-SNR as our loss function, which has commonly been used in speech separation system. It is defined as:

$$\begin{cases} \hat{\boldsymbol{s}} = g_{DL}(\boldsymbol{r}) \\ \boldsymbol{s}_{proj} = \frac{\langle \hat{\boldsymbol{s}}, \boldsymbol{s} \rangle \boldsymbol{s}}{\|\boldsymbol{s}\|^2} \\ \boldsymbol{e} = \hat{\boldsymbol{s}} - \boldsymbol{s}_{proj} \\ SI - SNR = 10\log_{10} \frac{\|\boldsymbol{s}_{proj}\|^2}{\|\boldsymbol{e}\|^2} \end{cases} \tag{17}$$

where $\langle\cdot\rangle$ and $\|\cdot\|$ denote the inner product operation and power operation respectively. $g_{DL}(\cdot)$ denotes the deep learning network, $\hat{\boldsymbol{s}}$ denotes the signal estimation, $\boldsymbol{s}_{proj}$ denotes the projection of signal estimation onto the transmitted signal, and $\boldsymbol{e}$ is the residual between the signal estimation and the signal projection.

The mean square error (MSE) loss function is defined as:

$$J_{MSE} = \|\hat{\boldsymbol{s}} - s\|^2 \tag{18}$$

From Fig. 3, it is obvious that given the angle θ, SI-SNR is invariant with the scale α but MSE is dependent on the scale α. Besides, maximizing SI-SNR is equivalent to maximizing the correlation coefficient, which is exactly the purpose for improving temporal coherence. Furthermore, the minimum mean-squared error (MMSE) can be asymptotically approached by deep learning based methods [32,33] and is also equivalent to SI-SNR [30].

For the ASIAEX 2001 SCS experiment, internal waves are the major oceanic features in the exploration of acoustic volume interaction, resulting in sound-speed fluctuation and scattering. Due to the ocean sound field fluctuation, the experimental area is a time-varying environment and the amplitude gains of the received signal are variant at different time. Based on the above analysis, the SI-SNR function is more conductive for UWA active signal detection.

3.5. Time domain signal detection

Based on the deep-learning method, we can get the estimation of transmitted signal from the received signal in time domain.

$$\hat{\boldsymbol{s}} = g_{DL}(\boldsymbol{r}) \tag{19}$$

After passing through the LC and OMP-LC detector, we can get the output of the whole system

$$\boldsymbol{z}_{DL-LC} = \boldsymbol{S}^T g_{DL}(\boldsymbol{r}) \tag{20}$$

$$\begin{aligned} \boldsymbol{z}_{DL-OMP-LC} &= \boldsymbol{S}^T g_{OMP}(g_{DL}(\boldsymbol{r})) \\ &= \boldsymbol{S}^T g_{DL-OMP}(\boldsymbol{r}) \end{aligned} \tag{21}$$

We can see that the transfer function is replaced by the deep-learning network and can be combined with channel estimation methods.

4. Experiments

In this section, we evaluate the performance of the proposed method in the detection system. Several experiments were conducted to compare our method with other conventional and deep learning methods. In the following, we will introduce our datasets, then make comparison with deep neural network (DNN) based models and optimize the model's parameters. Finally, the implementation process with conventional methods is presented and the results are discussed.

4.1. Datasets preparation

The raw data used in this study was obtained by ASIAEX 2001 SCS experiment. The configuration of the ASIAEX 2001 SCS experiment is shown in Table 1 and the source transmission schedule is provided in Table 2. To form our datasets, the first 10 pulses and the top hydrophone are chosen. After bandpass filtering, the received signals are demodulated and resampled in 800 Hz. Our datasets contain about 40 h signal components and 60 h ambient noise. Under assumption H_1, each sample consists 10 pulses and its adjacent ambient noise. Under assumption H_0, each sample is composed of ambient noise. In our following experiments, the ratio of the training set to testing set was approximately 4:1 and cross

Table 1
Configuration of the experiment.

Parameters	Value
source	E400
source depth	99.7 m
distance	19 km
receiver	16-element vertical line array
receiver depth	42.75 m-121.5 m

Table 2
Source transmission schedule.

Parameters	Value
center frequency	400 Hz
bandwidth	100 Hz
m-sequence length	511
number of pulses	88
time length of a single pulse	5.11s
sample rate	3255.208252 Hz
resample rate	800 Hz

validation was chosen. To guarantee the generalization of our proposed method, the data in training set and testing set is from different experiment dates, which is considered to have different propagation medium fluctuations. Additionally, we calculate the correlation coefficient and SI-SNR matrix between the training and testing sets, which are shown in Fig. 4 (a) and (b). These matrices model the difference in channel characteristics and can be considered as a measure of distance between the training and testing sets. There are only 7 numbers greater than 0.7 in the correlation coefficient matrix and 5 numbers greater than 0 dB in the SI-SNR matrix. It can be seen that the correlation coefficients and SI-SNRs of the channel characteristics are low, indicating a difference in the environment between the testing set and the training set.

In the training set, the arrival time of pulses is random between [300,2000] sampling points and the SNR of each sample is random between [-15,0] dB. For the testing set, the SNR of each sample is from −25 dB to 0 dB. The SNR of the received signal is defined as

$$SNR = 10\log_{10}\frac{\left\|\boldsymbol{r}_{signal}\right\|^2 - \left\|\boldsymbol{r}_{noise}\right\|^2}{\left\|\boldsymbol{r}_{noise}\right\|^2}$$

where $\boldsymbol{r}_{signal}$ is the received signal under assumption H_1 and $\boldsymbol{r}_{noise}$ is the adjacent received signal under assumption H_0.

The time domain normalized waveform of the recorded underwater ambient noise (about 15 min) is shown in Fig. 5, where the impulsive property of the noise can be clearly observed. Compared with Gaussian distribution, it appears that the probability density for large noise amplitudes is significantly higher in Fig. 6.

Fig. 5. Time domain waveform of recorded underwater ambient noise.

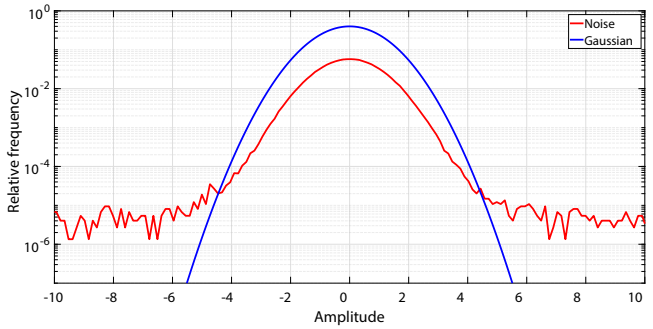

Fig. 6. Comparison of PDF between the underwater ambient noise and the Gaussian noise.

Here, we give two time-varying channels (SCS1 and SCS2) sampled from our testing sets, as well as their delay-Doppler spread functions (DDS) and power delay profiles (PDP) in Fig. 7, which are described in [14]. In Fig. 7 (a) and (b), the sub-pictures in the upper right corner are partial enlarged pictures from 1.2 s to 1.3 s. We can see that the two channels have several relatively stable paths but with some distortion,which means range migration. To provide additional evidence of the quality of the UWA channels, we employ empirical mode decomposition (EMD) as described in [34] to isolate the random component from the channel.

$$\boldsymbol{h} = \boldsymbol{h}_{\boldsymbol{d}} + \boldsymbol{h}_{\boldsymbol{w}} \tag{23}$$

where $\boldsymbol{h}$ represents the UWA channel. $\boldsymbol{h}_{\boldsymbol{d}}$ is known as the trend, representing the contribution of pseudo-deterministic physical phenomena to the channel's fluctuations. Additionally, $\boldsymbol{h}_{\boldsymbol{w}}$ represents channel fluctuations caused by scatterers that result in fast fading, which is the zero-mean wide-sense stationary uncorrelated scattering (WSS) ergodic random process.

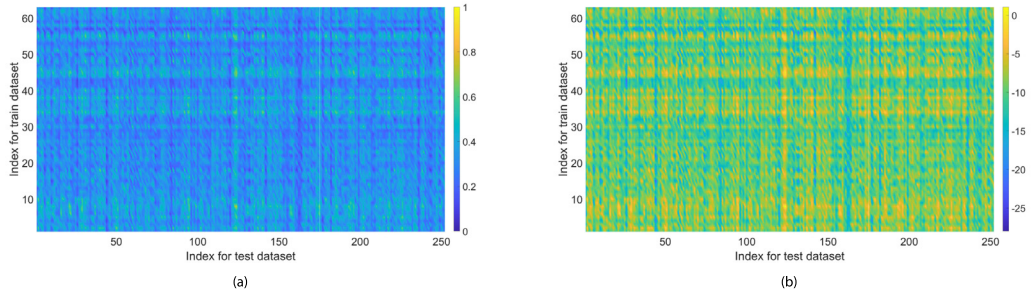

Fig. 4. (a) The correlation coefficient matrix between the training and testing sets; (b) The SI-SNR matrix between the training and testing sets.

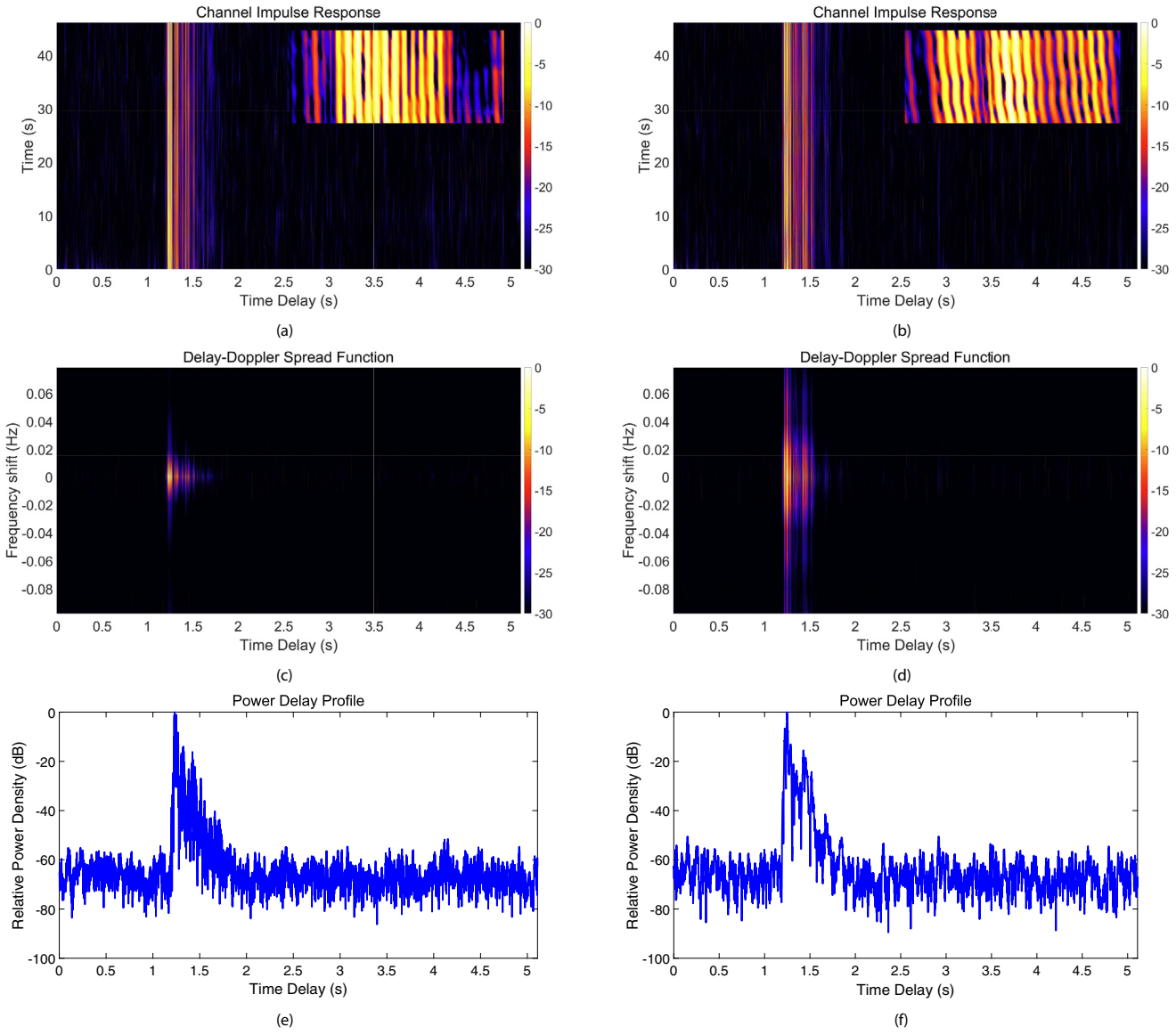

Fig. 7. Channel impulse response: (a) SCS1; (b) SCS2. Delay-Doppler spread function: (c) SCS1; (d) SCS2. Power delay profile: (e) SCS1; (f) SCS2.

Then the average fade rate (AFR) defined in [27] can be applied to evaluate the UWA channels.

$$AFR = 10\log\frac{Pow(\boldsymbol{h_w})}{Pow(\boldsymbol{h})} \tag{24}$$

The AFR values of the SCS1 and SCS2 channels are 0.0505 and 0.6312, respectively. We can see that the SCS1 channel with small AFR value is of higher quality. This is because its stable paths' energy is more concentrated and it has smaller Doppler frequency shift.

4.2. Comparison with baselines

In this section, the advantages of our model over previous models are analyzed without considering the multipath information. As a comparison, the LC detector is used as the conventional baseline model. In addition, the TCDAE [35] and Conv-Tasnet [29] are used as the representative DNN-based baseline models, which are briefly introduced as follows: TCDAE is one-dimensional time-domain denoise approach with skip connections between encoder and decoder layers. Conv-Tasnet is a fully-convolutional time-domain audio separation network for end-to-end time-domain speech separation. All the above models are trained on the same datasets. Our proposed model was implemented on the advanced deep-learning framework *PyTorch* and trained using NVIDIA RTX 4090 GPU. The parameters and symbols of our proposal are listed in Table 3.

Table 3
Summary of system parameters and symbols.

Parameter	Value	Symbol	Value
Optimizer	Adam	L	255
Learning rate	1e-3	F	355
Batch number	128	C_1	128
Epoch number	200	C_2	256
Training SNR	−15 dB - 0 dB	C_3	512
Training TOA	300–2000 sample points	B/H/S_c	128
Test SNR	−25 dB–0 dB	x	8
Test TOA	300–2000 sample points	R	3

Table 4
Quantitative comparisons with other UWA signal detection methods on our dataset.

SNR(dB)	Methods	DSI-SNR1(dB)	DSI-SNR2(dB)
−20	LC	3.49	−5.62
	TCDAE	6.58	−3.13
	Conv-Tasnet	10.53	−0.50
	Our model	**16.05**	**7.90**
−18	LC	5.08	−3.97
	TCDAE	8.42	−1.29
	Conv-Tasnet	15.53	5.95
	Our model	**24.83**	**16.69**
−15	LC	7.72	−1.33
	TCDAE	11.16	1.45
	Conv-Tasnet	19.58	6.01
	Our model	**28.76**	**20.61**
Ave.	LC	5.43	−3.64
	TCDAE	8.72	0.99
	Conv-Tasnet	15.21	3.82
	Our model	**23.21**	**15.07**

In audio separation task, the signal-to-distortion ratio (SDR) [36] and SI-SNR are used as objective metrics. To be pointed out, the SDR is equivalent to SI-SNR for one target signal. Unlike audio tasks, the aim of UWA signal detection task is to magnify the difference between the two scenarios with and without the presence of the target signal. Therefore, we define two differential SI-SNR (DSI-SNR) loss functions as our objective metrics:

$$\begin{aligned} DSI-SNR1 &= \frac{1}{N_{cv}}\sum_{i=0}^{N_{cv}}\left(SI-SNR\left(\hat{\boldsymbol{s}}_{i|H_1},\boldsymbol{s}\right)\right) \\ &\quad -\frac{1}{N_{cv}}\sum_{i=0}^{N_{cv}}\left(SI-SNR\left(\hat{\boldsymbol{s}}_{i|H_0},\boldsymbol{s}\right)\right) \end{aligned} \tag{25}$$

$$\begin{aligned} DSI-SNR2 &= \frac{1}{N_{cv}}\sum_{i=0}^{N_{cv}}\left(SI-SNR\left(\hat{\boldsymbol{s}}_{i|H_1},\boldsymbol{s}\right)\right) \\ &\quad -max\left(SI-SNR\left(\hat{\boldsymbol{s}}_{i|H_0},\boldsymbol{s}\right)\right) \end{aligned} \tag{26}$$

where $\hat{\boldsymbol{s}}_{i|H_1}$ and $\hat{\boldsymbol{s}}_{i|H_0}$ represent model outputs when the i-th noisy signal sample and i-th noise sample are inputted, respectively. N_{cv} is the number of signals in the test dataset. The DSI-SNR1 and DSI-SNR2 both characterize the improvement in SI-SNR. The former measures the average improvement, while the latter focuses on the improvement at low probability of false alarm (P_{FA}).

The results for DSI-SNR1 and DSI-SNR2 are presented in Table 4 at three different SNRs. It is evident that the conventional LC detector performs poorly in detecting UWA signals with low SNRs, while the DNN-based baselines outperform the conventional methods across all evaluation metrics. Moreover, our proposed method outperforms other DNN-based techniques on our datasets at different SNRs, achieving the highest evaluation scores. Our method provides a 17.78 dB gain in DSI-SNR1 and an 18.71 dB gain in DSI-SNR2, which indicates its effectiveness even at low SNRs. These results demonstrate that our proposed method is an efficient solution for detecting UWA signals, particularly in low SNR conditions.

4.3. Parameter optimization

In this section, the parameters including the kernel size L in the encoder and decoder, the number of filters H in the separator and the number of 1-D D-ConvBlocks in each repeat x are optimized with $SNR = -18$ dB and the results are shown in Table 5. We can see that for different H values, DSI-SNR1 results are similar. However, the methods achieved the best results for DSI-SNR2 when we set $H = 128$. Meanwhile, increasing the H in the separation module greatly increases the model size. Additionally, when we reduce the L and x values, there is significant performance degradation in both DSI-SNR1 and DSI-SNR2. This is particularly evident for the L parameter.

4.4. Evaluation on unseen datasets

This section further explores the generalization performance of our model on unseen datasets. For this purpose, we utilize two

Table 5
Tuning the hyper-parameters.

L	H	x	DSI-SNR1 (dB)	DSI-SNR2 (dB)
127	128	8	19.34	6.87
255	128	8	**24.83**	**16.69**
511	128	8	23.54	11.15
255	128	7	21.47	12.77
255	256	8	24.16	13.80
255	512	8	**25.57**	15.50

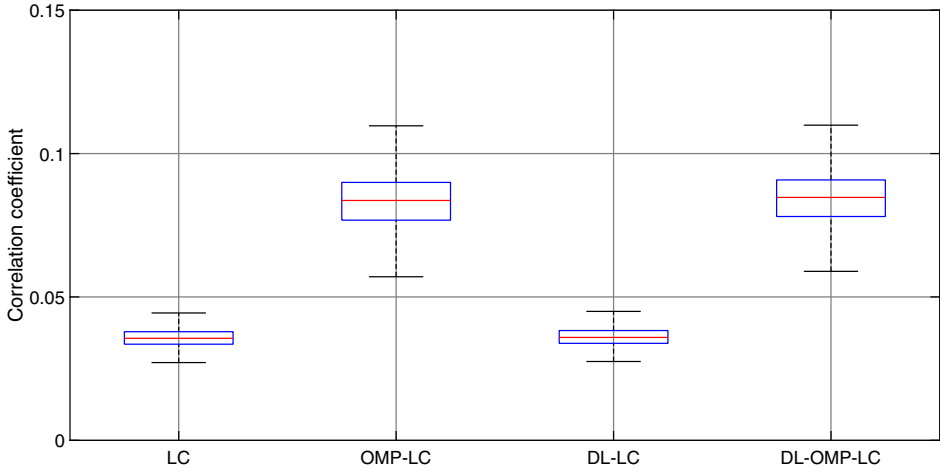

Fig. 8. The output noise levels of LC, OMP-LC and DL-LC.

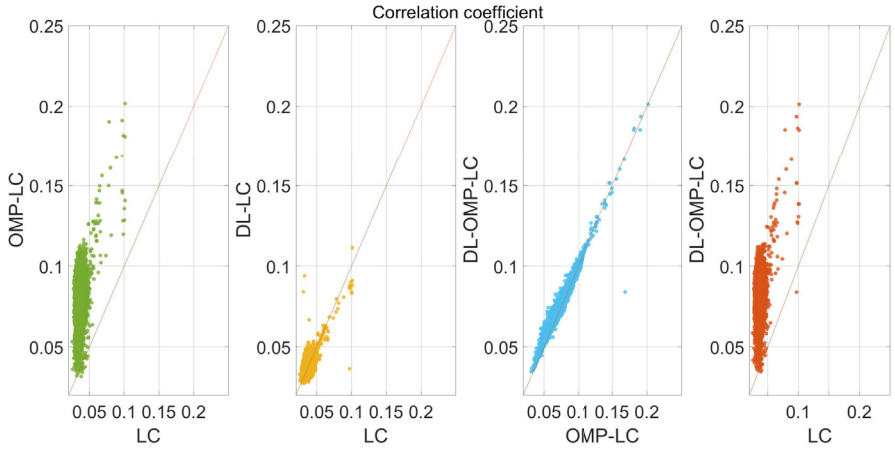

Fig. 9. The comparisons of output noise levels among LC, OMP-LC and DL-LC.

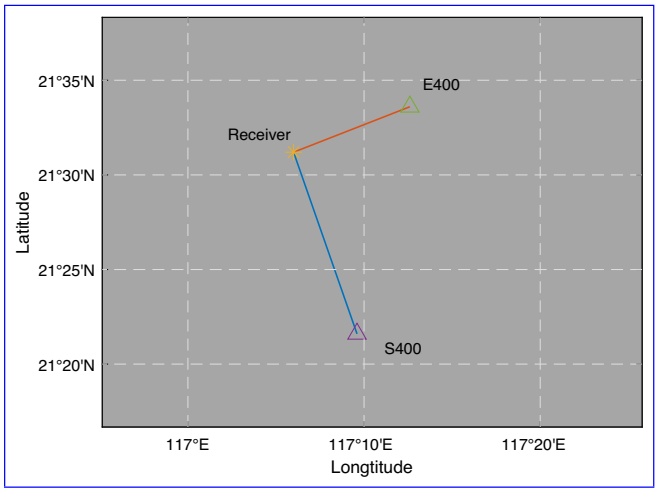

Fig. 10. AXIAEX 2001 SCS experiment area.

different unseen scenarios to test our method. The first scenario is from the unseen source S400 in the AXIAEX 2001 SCS experiment and its geographic coordinates are shown in Fig. 10. The distance between S400 source and recever is about 30 km. To be pointed out, the E400 and S400 sources transmit the same frequency band signal but with different modulation sequences. We demodulate the S400 source transmitted signal and then modulate it by the E400 source's m-sequence. After that, this test dataset utilizes the same template signal. To simplify it, we call it the S400 dataset. The second is the unseen noise dataset, which is from ShipsEar dataset [?] and conducted in Cortegada. In this scenario, we form our test dataset by mixing the noise signal in ShipsEar dataset with the source E400 signals in the AXIAEX 2001 SCS experiment. The results are shown in Fig. 12.

For the S400 dataset (Fig. 12(a)), we observe that the LC detector has 1.8 dB performance improvement in DSI-SNR1 and DSI-SNR2. Meanwhile, our model achieves the similar DSI-SNR1 and DSI-SNR2, which means that the detection performance of our model is not affected and our model has a good generalization performance in different sound propagation sea area. Besides, for the unseen noise dataset (ShipsEar), it can be seen from Fig. 12 (b) that

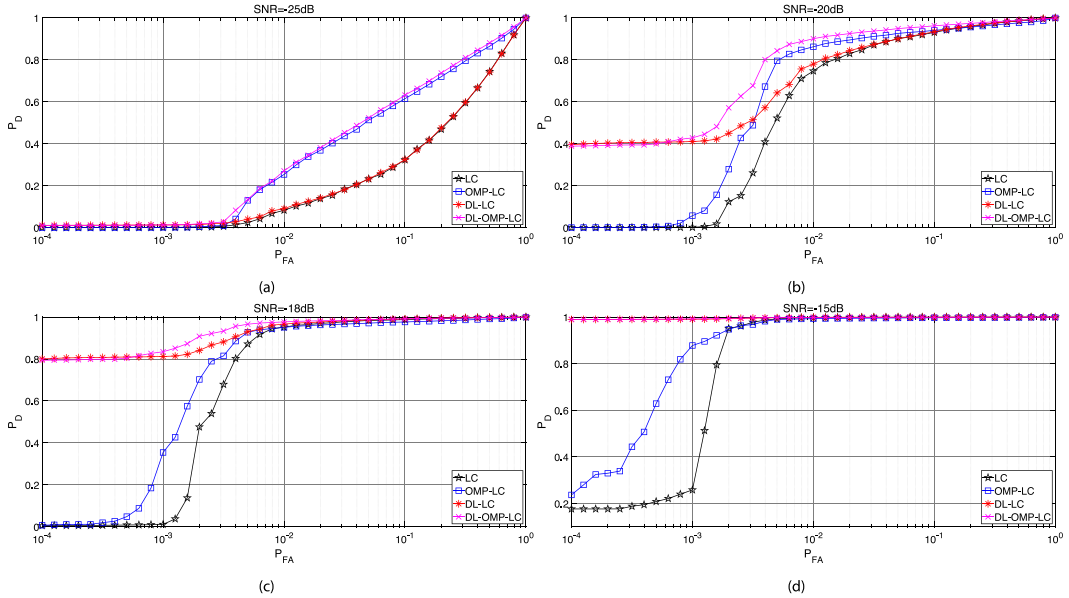

Fig. 11. ROC comparisons for different detectors in different SNR: (a) SNR = −25 dB; (b) SNR = −20 dB; (c) SNR = −18 dB; (d) SNR = −15 dB.

the LC detector has about 0.9 dB degradation in DSI-SNR1 and has about 6 dB gain in DSI-SNR2. Our model has about 2 dB gain in DSI-SNR1 and DSI-SNR2. The noise samples (rain, flow and wind) in the ShipsEar dataset are collected from the sea surface. The E400 and S400 noise samples are collected from the underwater environment. It can be seen that our model exhibits a robust performance in different noise background environment. Furthermore, the performance gain of our model over LC detector is relatively stable in both unseen datasets. All the above results demonstrate that our proposed model is effective to deal with unseen datasets and has a good generalization performance.

4.5. Detailed analysis with conventional methods

As the advantages of our model are analyzed in previous sections. In this section, the detailed analysis is implemented to demonstrate the performance of our proposed detection methods. The detectors in Eq. 20 and 21 are experimentally compared with two other approaches:

1 Normalized LC based detection system. This method is widely used for various detection systems because it offers stable performance with low complexity.
2 OMP based detection system. Fisrt, this method uses OMP to acquire a prior knowledge of the UWA channel, and then reconstructs the copy signal or deconvolves the received signal. Finally, normalized LC is applied to detect if the object signal exists. It improves the performance of detection system when the OMP method is valid.

As we discussed before, our aim is to get a higher probability of detection (P_D) while maintaining a low P_{FA}. In this section, we will

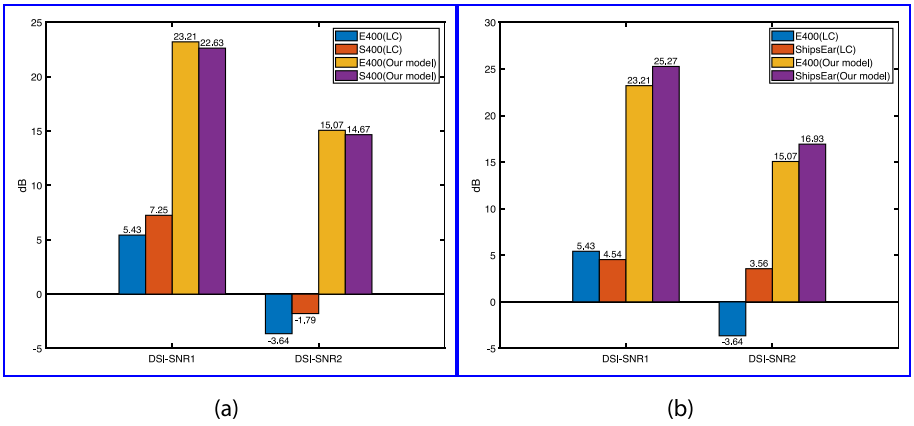

Fig. 12. Evaluation of our model for unseen datasets in terms of DSI-SNR1 and DSI-SNR2: (a) S400 dataset; (b) ShipsEar dataset.

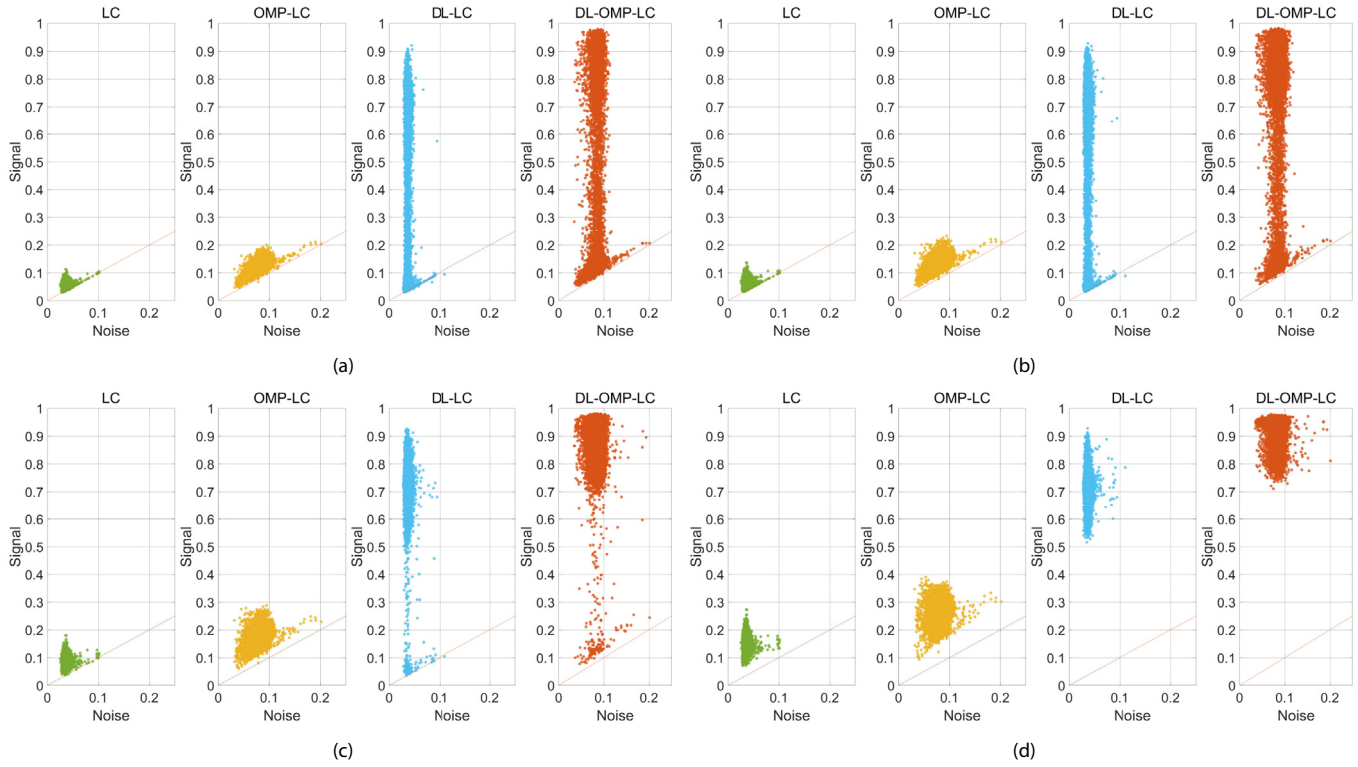

Fig. 13. The output correlation coefficients of LC, OMP-LC, DL-LC and DL-OMP-LC in different SNR: (a) SNR = −20 dB; (b) SNR = −18 dB; (c) SNR = −15 dB; (d) SNR = −10 dB.

make a comparison about the output noise levels of these methods under assumption H_0, which are corresponding to the detection threshold (Th) in CFAR detection. The box plots of the above detectors are shown in Fig. 8. The comparisons of correlation coefficients are shown in Fig. 9, where the straight lines are noise level contours. If the scatters are above these lines, then the noise level is amplified. We can see that LC has a lowest output noise level. The noise level of DL-LC and DL-OMP-LC are equivalent to LC and OMP-LC, respectively. The noise level of OMP-LC and DL-OMP-LC is doubled at least compared to LC. Besides, OMP-LC has more outliers than DL-LC. It means that the DL-LC method is more stable with the underwater ambient noise and DL does not has the issue of noise amplification, which makes the high P_D with a low P_{FA} possible.

The ROCs are one of the most important graphs showing the performance of detection from the relationship between P_D and P_{FA}. The ROCs of different detectors in different SNR are shown in Fig. 11, and the following observations can be made. (1) The performance of LC is not good in low SNR and is inferior to other proposed methods. (2) OMP-LC is unable to get a good performance when P_{FA} is low. This is because OMP-LC is likely to amplify the noise level, which leads to a higher detection threshold when keeping P_{FA} low. (3) OMP-LC is also not efficient in low SNR since OMP algorithm fails under this condition. (4) DL-LC performs well when P_{FA} is below 10^{-3}. Given the $P_{FA} = 10^{-4}$, when SNR is -20 dB, -18 dB and -15 dB, the P_D can reach 0.4, 0.8 and 1. (5) When P_{FA} is beyond 10^{-2}, DL-OMP-LC has the best performance and OMP-LC shows a better performance than DL-LC. However, this area is not in our interest. If P_{FA} is very high, the detection system will always find targets and brings much burden to the operators. (6) The DL signal enhancement method augments the received signal when the object signal exists. Before the processing of LC and OMP-LC, DL is applied to enhance the signal. It can be seen that there is a great performance improvement with both LC and OMP-LC detetors.

In the following, we will analyze how the DL network enhances the signal in detail. Take the SNR of -20 dB, -18 dB, -15 dB and -10 dB as examples, the output correlation coefficients of the proposed methods are illustrated in Fig. 13. The horizontal and vertical axis represents the output correlation coefficients of noise and noised signal at a given SNR. It can be seen that OMP-LC and DL-OMP-LC can enhance the signal for most samples, but LC and DL-LC enhance it selectively when SNR = -20 dB. With the increase of SNR, the correlation coefficients of LC and OMP-LC improve slowly. On the contrary, the correlation coefficients of DL-LC and DL-OMP-LC improve rapidly. It means that on the one hand, noised signals are not enhanced if DL network considers it as noise, and on

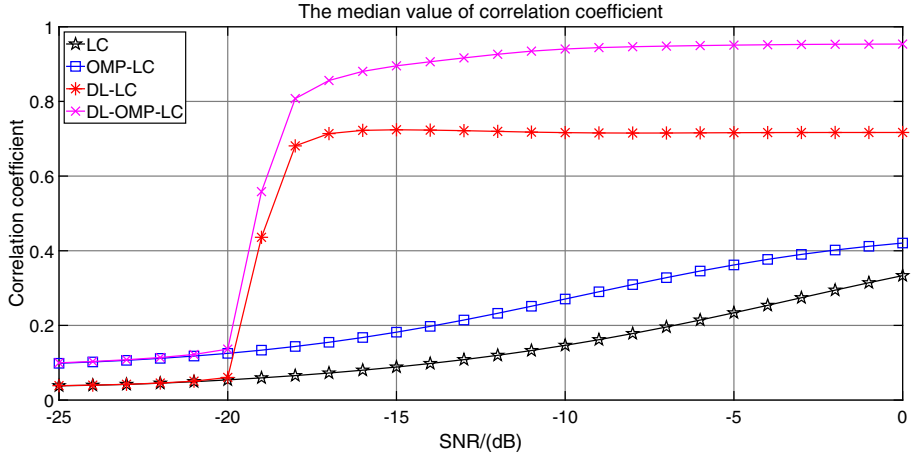

Fig. 14. Dependency of median values of correlation coefficients versus SNR.

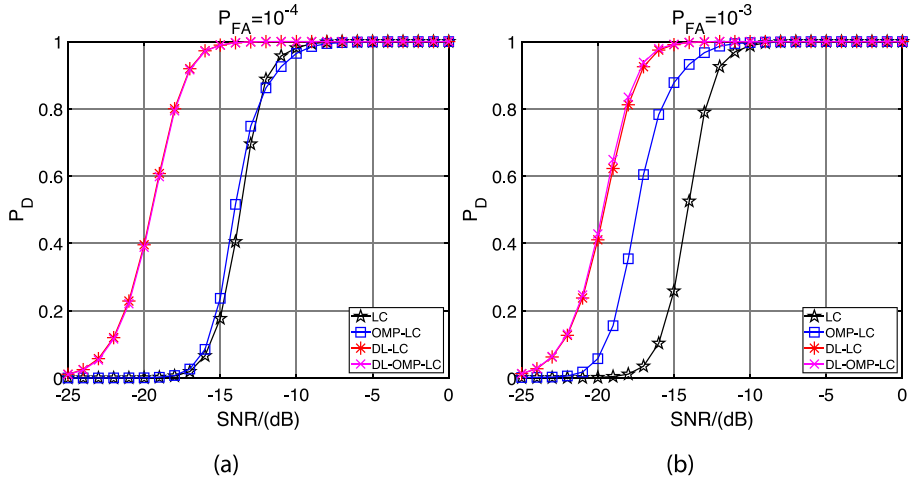

Fig. 15. Dependency of P_D versus SNR in CFAR detection: (a) $P_{FA} = 10^{-4}$; (b) $P_{FA} = 10^{-3}$;.

the other hand, the DL network tends to output a very clean signal if the features of signals are captured.

To further illustrate the enhancement capability of the DL network, we test the median values of correlation coefficients with the above four methods versus SNR, which are shown in Fig. 14. It is necessary to point out that the detection is more reliable with more higher coefficient and the correlation coefficient is equivalent to the SNR gain. We can see that the median values of correlation coefficients of DL-LC and DL-OMP-LC increase rapidly to 0.7 and 0.9, respectively. For LC and OMP-LC, their correlation coefficients increase slowly. When SNR = −15 dB, the correlation coefficients of LC and OMP-LC only have about 0.1 and 0.19. When SNR exceeds

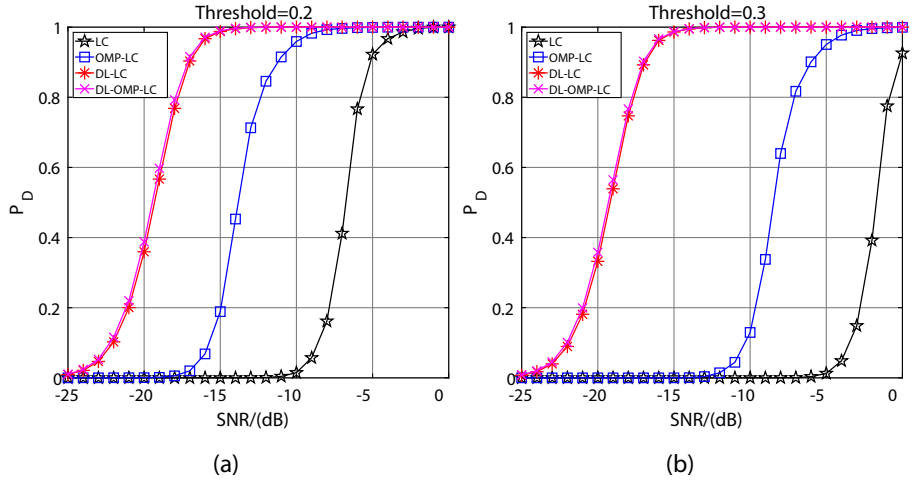

Fig. 16. Dependency of P_D versus SNR in different detection thresholds: (a)Th = 0.2; (b)Th = 0.3.

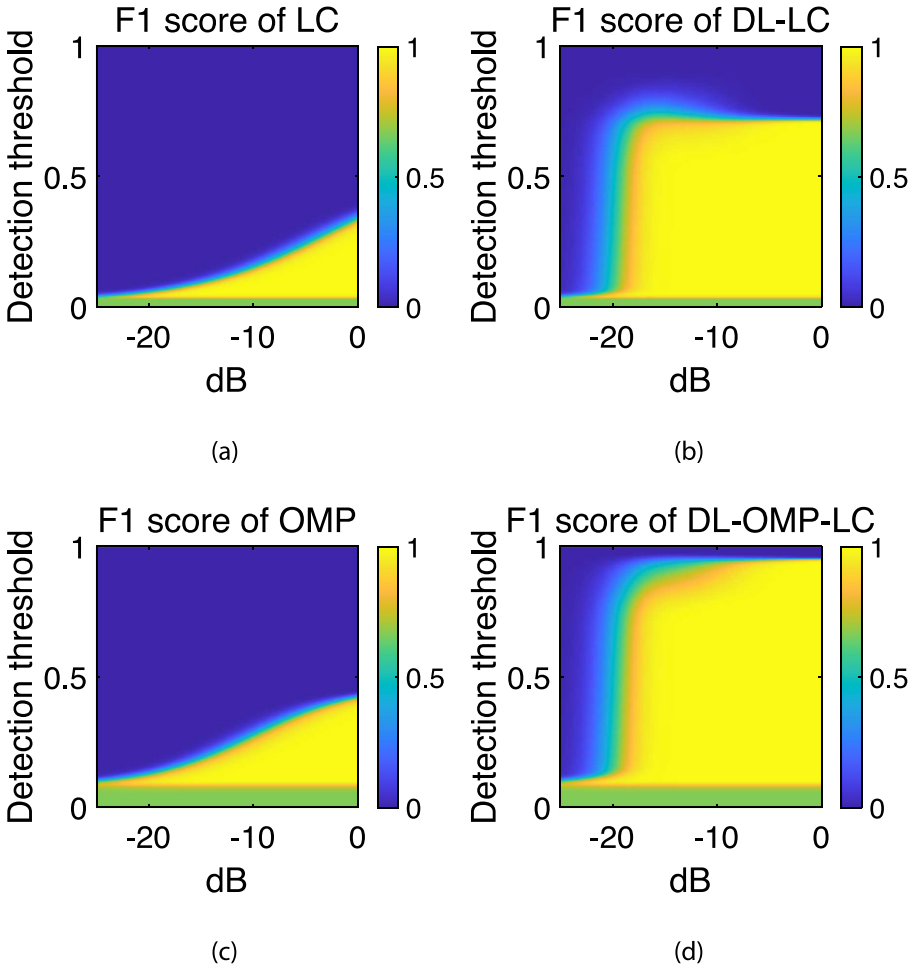

Fig. 17. F1 score of different detectors: (a) LC; (b) DL-LC; (c) OMP-LC; (d) DL-OMP-LC.

−15 dB, our proposed methods have a relatively stable and higher correlation coefficient values. From the point of view of median values, our proposed methods are always better than the conventional methods. The increase in the median values is quite significant, which means a high gain detection.

In addition to the ROC curves, in many cases, the dependence of P_D versus SNR at small P_{FA} is of higher interest, which is investigated in Fig. 15. It can be seen that the proposed detectors are better than conventional detectors regardless of SNR. Besides, the proposed detectors show a promising performance in very low SNR. When SNR is below −15 dB, LC and OMP-LC are almost undetectable at $P_{FA} = 10^{-4}$, but P_d of the proposed methods is able to get a 0.4 at −20 dB and almost 1 at −15 dB. If we relax P_{FA} to 10^{-3}, the performance of OMP-LC has been improved, but has at least 2 dB performance gap with the proposed methods. The detection performance of the proposed methods is at the same level when P_{FA} is low. This shows that the performance improvement is due to the DL network in this area.

To compare the detection performance in fixed thresholds, we increase the detection threshold to 0.2 and 0.3, which are normally used in real engineering applications. The dependence of P_D versus SNR is shown in Fig. 16. When the threshold equals 0.2, there is about 6 dB and 13 dB performance improvement compared to OMP-LC and LC. When the threshold equals 0.3, LC is completely undetectable and OMP-LC is able to get a 0.6 P_D at −8 dB. Our proposed detectors DL-LC and DL-OMP-LC dose not loss much performance. It means that our proposed detectors are able to get a high detection gain.

To verify the tolerance of the proposed methods in threshold selection, we show the relationship between F1 score, SNR and threshold in Fig. 17. It can be seen that the maximum detection thresholds of LC, OMP-LC, DL-LC and DL-OMP-LC are 0.3, 0.4, 0.7 and 0.85, respectively. For conventional detectors, OMP-LC is more tolerant than LC. This result also holds for the DL based detectors, that is, DL-OMP-LC is more tolerant than DL-LC. Meanwhile, the proposed detectors are more tolerant than the conventional detectors, which also means that the proposed detectors are of higher detection gain.

The above analysis is from the perspective of statistics. Next, we will analyze from the perspective of samples, which means time-domain correlation coefficient, channel impulsive response, DDS and PDP. Time-domain correlation coefficient and PDP are shown in Fig. 18–19 when SNR is −15 dB. We find that the peak SNR gain is about 55 dB with DL network and the main lobe width is more narrower with OMP. The peak SNR gain is from the background noise suppression and peak augmentation. As shown in Fig. 20, the background noise levels of the proposed detectors are very low. Besides, the time-delay and Doppler spread are quite small.

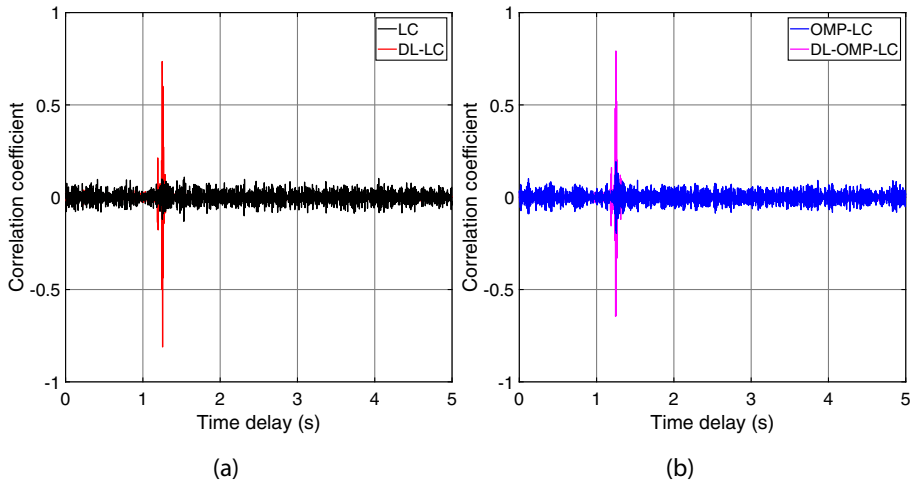

Fig. 18. Correlation coefficient of different detectors in time domain: (a) LC and DL-LC; (b) OMP and DL-OMP-LC.

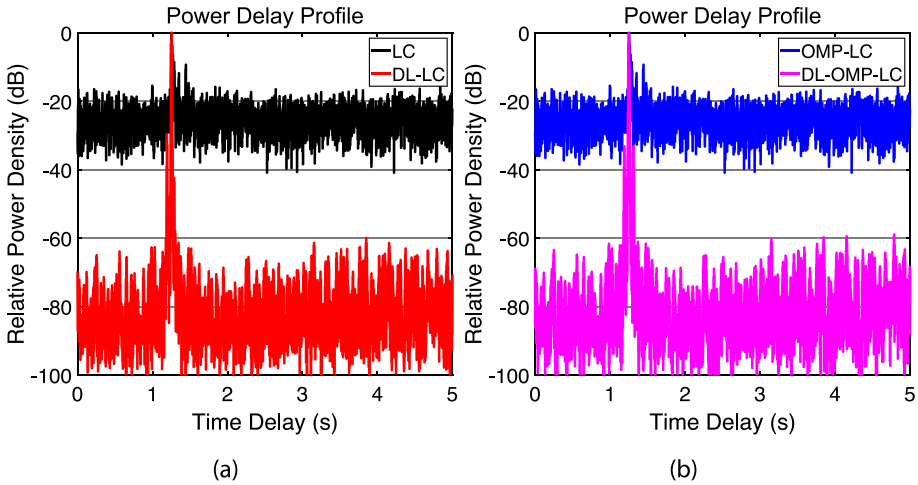

Fig. 19. PDP of different detectors: (a) LC and DL-LC; (b) OMP and DL-OMP-LC.

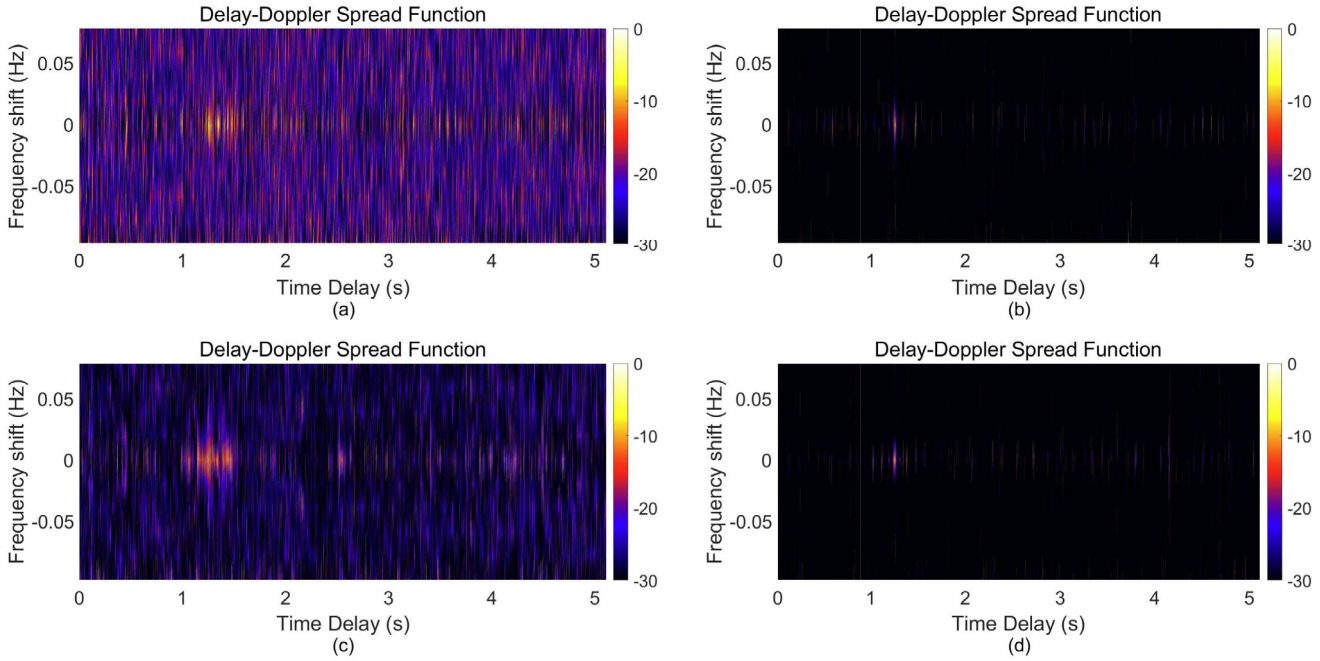

Fig. 20. DDS of different detectors: (a) LC; (b) DL-LC; (c) OMP-LC; (d) DL-OMP-LC.

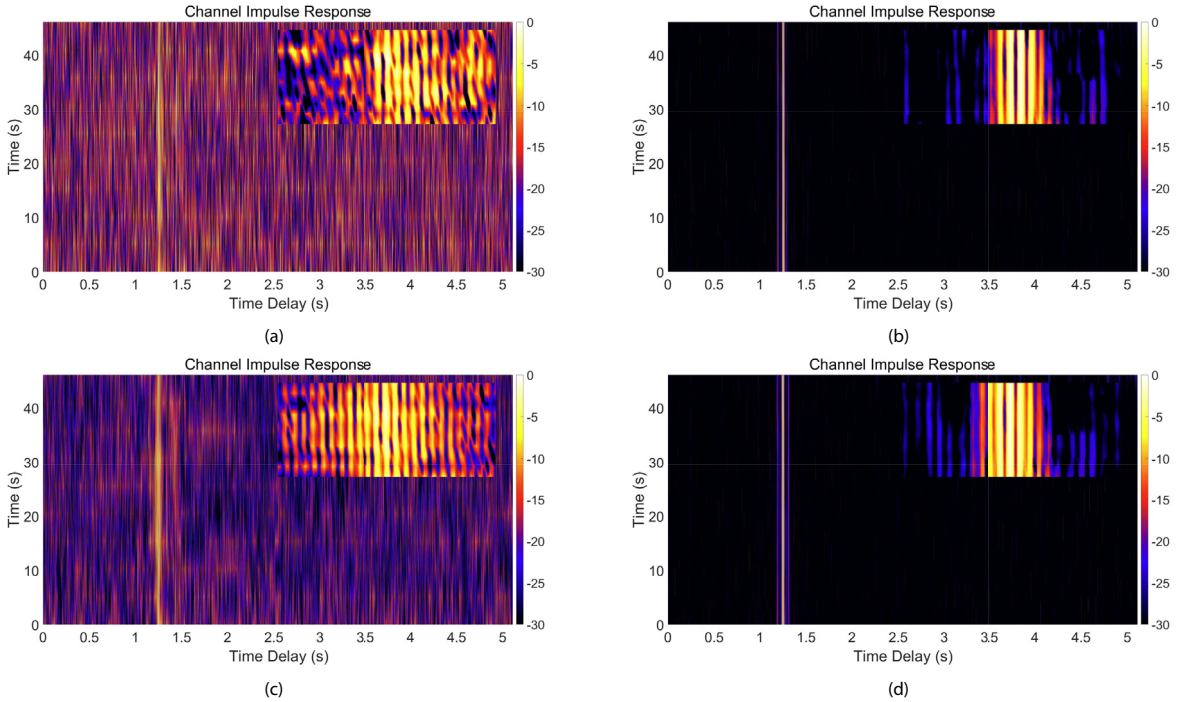

Fig. 21. Channel impulsive response of different detectors: (a) LC; (b) DL-LC; (C) OMP-LC; (D) DL-OMP-LC.

Channel impulsive responses of these detectors are investigated in Fig. 21, it can be seen that the distortion in multipath (i.e. range migration) is migrated by the proposed detectors. Based on the above results, it is obvious that our proposed methods are able to denoise the signal and compensate the channel fading both in multipath and Doppler.

To further illustrate the source of detection gain, a simulation was conducted in the random distortion environment with range migration. In time-invariant UWA channel, the relative time delay for adjacent pulses is a constant, i.e., the time length of a single pulse. In this simulation, we model the range migration as the abrupt time of arrivals between adjacent pulses, which leads to a random time delay shift for each pulse. Here, we ignore the multipath propagation and set the number of multipath to 1. Then, the random time delay shift for each pulse is set from 0 to 62.5 ms. Besides, we assume that there is no ambient noise in this simulation.

Monte Carlo simulations were used to evaluate the performance, assuming that the total number of Monte Carlo run is 10 000. The median values of correlation coefficients versus different

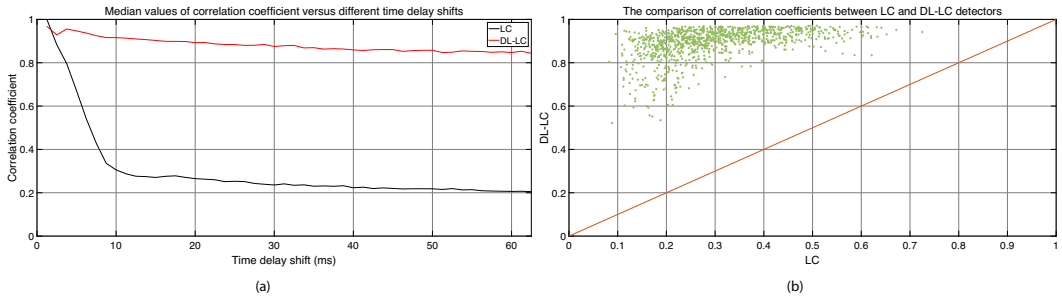

Fig. 22. (a) Median values of correlation coefficient versus different time delay shifts; (b) The comparison of correlation coefficients between LC and DL-LC detectors when the time delay shift is 10 ms.

time delay shifts are shown in Fig. 22 (a). We can see that the LC detector is more sensitive to the time delay shift and its performance degrades significantly with the increase of time delay shift. On the contrary, the DL-LC detector is able to suppress the random range migration. When the time delay shift is 62.5 ms, the median value of correlation coefficient for DL-LC is above 0.8, in which the LC detector only has 0.2. Then we make the comparison of correlation coefficients between LC and DL-LC detectors when the time delay shift is 10 ms. As shown in Fig. 22 (b), the correlation coefficients of DL-LC are always higher than that of LC. Under this condition, the correlation coefficient of LC is hard to get 0.8 but the median value of DL-LC is above 0.8.

From this simulation, we show that the deep-learning structure is able to suppress the random range migration. When the range migration occurs, the performance of LC detector degrades significantly. As for the DL-LC detector, the correlation coefficients are pretty high even with a large range migration. It means that the time gain is acquired by the DL network.

5. Conclusion

In this study, we proposed a deep-learning based two-stage method for UWA active signal detection, in which the encoder was applied to extract latent features from the received signals, the separator was designed to obtain the clear object signal map and the decoder was used to reconstruct signal from the masked feature map. In the first stage, we utilized deep-learning network to enhance the received signal. In the second stage, two types of conventional detectors were developed to decide if the object signal is present or not. Moreover, a novel SI-SNR function was introduced to UWA active signal detection, the benefits and equivalence of which were explained. Furthermore, these detectors were unified into one form with their own transfer functions, which are related to their detection performance. In particular, a simulation was conducted to illustrate the source of detection gain and show the ability of suppressing the random range migration. Finally, the real-world experimental results demonstrate that the proposed detectors outperform the conventional detectors.

Data availability

The data that has been used is confidential.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported in part by the Key Technologies Research and Development Program under Grant No. 2021YFC3101403, National Natural Science Foundation of China under Grant No. 62171440 and the China Scholarship Council under Grant No. 202110580001, CAS Specific Research Assistant Funding Program.

References

[1] Turin G. An introduction to matched filters. IRE Trans Inform Theory 1960;6(3):311–29.
[2] Au WWL, Banks K. The acoustics of the snapping shrimp synalpheus parneomeris in kaneohe bay. J Acoust Soc Am 1998;103(1):41–7.
[3] Lee KM, McNeese AR, Wochner MS, Wilson PS. Reduction of underwater sound from continuous and impulsive noise sources using tethered encapsulated bubbles. J Acoust Soc Am 2012;132(3):2062.
[4] Chitre M, Potter J, Ong S-H. Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise. IEEE J Oceanic Eng 2006;31:497–503.
[5] Zozor S, Brossier J-M, Amblard P-O. A parametric approach to suboptimal signal detection in α)stable noise. IEEE Trans Signal Process 2006;54:4497–509.
[6] Mahmood A, Chitre M. Optimal and near-optimal detection in bursty impulsive noise. IEEE J Oceanic Eng 2017;42(3):639–53.
[7] P. Brockett, M. Hinich, G. Wilson, Nonlinear and non-gaussian ocean noise, Acoust Soc Am J 82.
[8] S.A. Razavi, E. Ollila, V. Koivunen, Robust greedy algorithms for compressed sensing, in: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 969–973.
[9] Eggen T, Baggeroer A, Preisig J. Communication over doppler spread channels. part i: Channel and receiver presentation. IEEE J Oceanic Eng 2000;25:62–71.
[10] Zhang Z, Wang H, Yao H. Pulse ranging method based on active virtual time reversal in underwater multi-path channel. J Mar Sci Eng 2020;8:883.
[11] Pan X, Ding Z, Jiang J, Gong X. Robust time-reversal is combined with distributed multiple-input multiple-output sonar for detection of small targets in shallow water environments. Appl Acoust 2018;133:157–67.
[12] Zhang G, Hovem JM, Dong H, Liu L. Coherent underwater communication using passive time reversal over multipath channels. Appl Acoust 2011;72:412–9.
[13] Y. nan Tian, X. Han, S.A. Vorobyov, J. wei Yin, Q. yu Liu, G. Qiao, Wideband signal detection in multipath environment affected by impulsive noise, J Acoust Soc Am 152 (2022) 445–455.
[14] van Walree PA. Propagation and scattering effects in underwater acoustic communication channels. IEEE J Oceanic Eng 2013;38:614–31.
[15] Qarabaqi P, Stojanovic M. Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels. IEEE J Oceanic Eng 2013;38:701–17.
[16] Tao R, Zhang N, Wang Y. Analysing and compensating the effects of range and doppler frequency migrations in linear frequency modulation pulse compression radar. IET Radar Sonar Navigat 2011;5:12.
[17] Y. Li, T. Zeng, T. Long, Z. Wang, Range migration compensation and doppler ambiguity resolution by keystone transform, 2006 CIE International Conference on Radar.
[18] Xu J, Yu J, Peng Y-N, Xia X-G. Radon-fourier transform for radar target detection, i: Generalized doppler filter bank. IEEE Trans Aerosp Electron Syst 2011;47:1186–202.
[19] Lin L, Sun G, Cheng Z, He Z. Long time coherent integration for maneuvering target detection based on itrt-mrft. IEEE Sens J 2020;20:3718–31.
[20] Zhang Y, Xi'an F. Multiplication-based pulse integration for detecting underwater target in impulsive noise environment. IEEE Access 2016;4:6894–900.

[21] Yang TC. Measurements of temporal coherence of sound transmissions through shallow water. J Acoust Soc Am 2006;120:3220.
[22] Yang TC. Temporal coherence of sound transmissions in deep water revisited. J Acoust Soc Am 2008;124(1):113–27.
[23] L. Wu, Y. Ren, H. Niu, J. Wang, J.F. Lynch, T.F. Duda, Statistic characteristics of the acoustic field in the asiaex 2001 south china sea experiment, AIP Conference Proceedings.
[24] Liu A, Ramp S, Zhao Y, Tang T. A case study of internal solitary wave propagation during asiaex 2001. IEEE J Oceanic Eng 2004;29:1144–56.
[25] Stojanovic M, Catipovic J, Proakis J. Phase-coherent digital communications for underwater acoustic channels. IEEE J Oceanic Eng 1994;19(1):100–11.
[26] Du B, Xiong W, Wu J, Zhang L, Zhang L, Tao D. Stacked convolutional denoising auto-encoders for feature representation. IEEE Trans Cybern 2017;47:1017–27.
[27] Zhang Y, Li C, Wang H, Wang J, Yang F, Meriaudeau F. Deep learning aided ofdm receiver for underwater acoustic communications. Appl Acoust 2022;187:108515.
[28] A. Pandey, D. Wang, Tcnn: Temporal convolutional neural network for real-time speech enhancement in the time domain, ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[29] Luo Y, Mesgarani N. Conv-tasnet: Surpassing ideal time-frequency magnitude masking for speech separation. IEEE/ACM Trans Audio, Speech, Language Process 2019;27:1256–66.
[30] Rugini L, Banelli P. On the equivalence of maximum snr and mmse estimation: Applications to additive non-gaussian channels and quantized observations. IEEE Trans Signal Process 2016;64(23):6190–9.
[31] Zeng W-J, Jiang X, Li X-L, Zhang X-D. Deconvolution of sparse underwater acoustic multipath channel with a large time-delay spread. J Acoust Soc Am 2010;127:909–19.
[32] Hu Q, Gao F, Zhang H, Jin S, Li GY. Deep learning for channel estimation: Interpretation, performance, and comparison. IEEE Trans Wireless Commun 2021;20(4):2398–412.
[33] Zhang Y, Wang H, Li C, Chen X, Meriaudeau F. On the performance of deep neural network aided channel estimation for underwater acoustic ofdm communications. Ocean Eng 2022;259:111518.
[34] Socheleau F-X, Laot C, Passerieux J-M. Stochastic replay of non-wssus underwater acoustic communication channels recorded at sea. IEEE Trans Signal Process 2011;59:4838–49.
[35] Tawara N, Kobayashi T, Ogawa T. Multi-channel speech enhancement using time-domain convolutional denoising autoencoder. in: Interspeech; 2019.
[36] Vincent E, Gribonval R, Fevotte C. Performance measurement in blind audio source separation. IEEE Trans Audio, Speech, Language Process 2006;14(4):1462–9.