



## Article

# A Joint Denoising Learning Model for Weight Update Space–Time Diversity Method <sup>†</sup>

Yu Zhang <sup>1</sup>, Dan Zhang <sup>2</sup>, Zhen Han <sup>1</sup> and Peng Jiang <sup>1,\*</sup>

<sup>1</sup> GNSS Research Center, Wuhan University, Wuhan 430079, China; zhangziju125@whu.edu.cn (Y.Z.); willhanzhen@whu.edu.cn (Z.H.)

<sup>2</sup> School of Computer Science, Wuhan University, Wuhan 430079, China; zhangxiaobei125@whu.edu.cn

\* Correspondence: jiangp@whu.edu.cn

<sup>†</sup> This paper is an extended version of our paper published in Zhang, Y.; Jiang, P.; Guo, W. An Underwater Array Localization Method Using Two-Stage Learning Model. *Geomat. Inf. Sci. Wuhan Univ.* **2021**, *46*, 1889–1899.

**Abstract:** Space–time diversity (STD) has been widely applied in underwater acoustic (UWA) communication due to its exceptional anti-multipath performance. However, underwater noise can seriously affect the processing results of STD. The conventional filtering algorithms cannot deal with the nonlinear components of underwater noise and may not work well for complex-type signals. This study proposes an improved STD method with a joint noise-reduction learning model for the above issues. We construct a noise-reduction learning model dedicated to complex-type UWA signals in the first stage. Complex-type features based on UWA data are extracted for pre-processing data, and a conditional generative adversarial network (CGAN) is used as the backbone network for noise-reduction. Residual learning is used to accomplish noise cancellation and yield noise-reduction estimates. In the second stage, an STD structure based on a weight update strategy is constructed. The STD structure can further constrain the weights of the signals from the main path, enhance the reception of the main path, and suppress the multi-access interference (MAI) caused by the spread spectrum communication. Finally, combining the signals on each path can improve the communication quality of the system based on the principle of the maximum signal-to-interference plus noise ratio (SINR). The simulation and experiments on a lake showed that the proposed method is more robust over the changing signal-to-noise ratio (SNR) and has a lower bit error rate (BER) than conventional methods.

**Keywords:** underwater acoustic communication; space–time diversity; noise-reduction; deep neural network; weight update strategy



**Citation:** Zhang, Y.; Zhang, D.; Han, Z.; Jiang, P. A Joint Denoising Learning Model for Weight Update Space–Time Diversity Method. *Remote Sens.* **2022**, *14*, 2430. <https://doi.org/10.3390/rs14102430>

Academic Editors: Yusuke Yokota and Takumi Matsuda

Received: 20 April 2022

Accepted: 17 May 2022

Published: 19 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



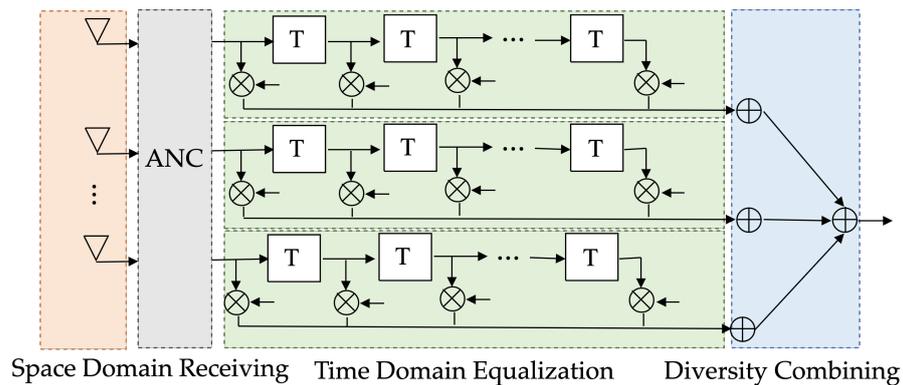
**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

High-quality UWA communication is essential for underwater data transmission, ocean resource exploration, and underwater vehicle information acquisition [1]. However, multipath effects can seriously disrupt UWA communication. The multipath effects contain multipath fading and inter-symbol interference (ISI) [2]. These phenomena are due to the fact that the acoustic signal can be reflected along multiple paths with different incidence angles and propagation time delays while superimposing with underwater noise on the receiver [3]. STD is widely used as a conventional method to suppress multipath effects in underwater communication [4,5]. In the STD structure, “S” indicates space-domain receiving, and “T” indicates time-domain equalizing. The underwater acoustic array receives the space-domain of the signal and decides the main path direction using beamforming technology [6–9]. Then, after equalization processing and diversity receiving in the time-domain, the multipath interference is suppressed and the BER of the communication is reduced [10]. However, the noise distributed in the underwater environment is a factor severely limiting high-speed and reliable underwater communication [11]. Noise can prevent the STD structure from accurately resolving the transmit signal and severely limit the

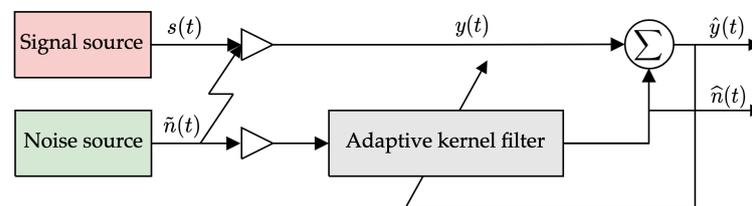
processing performance of STD [12]. Generally, underwater noise can be divided into water body noise and artificial noise [13]. The former is natural sound sources mainly including marine and biological noise. The latter is mechanical, especially noise from navigation and underwater operations. Vessel's underwater radiated noise (VURN) is recognized as a significant source of artificial noise [14]. It has been challenging to maintain high STD processing quality in an environment heavily affected by VURN.

To minimize the VURN effect, adaptive noise control (ANC) technology is commonly used as a noise-reduction method before the STD structure [15]. After ANC processing, the received signal will be diversity processed by the STD equalizer. Meanwhile, combined signals have higher communication gain [16]. A typical STD structure using ANC processing is shown in Figure 1.



**Figure 1.** A typical STD structure using ANC processing.

According to the object to be processed, ANC technology can be classified into linear and nonlinear methods. The linear method features simple computation and high real-time performance. However, it can only handle simple linear systems with limited applicability. The linear algorithms such as Wiener filtering [17], median filtering [18], and wavelet variation [19] may not work well for VURN, the nonlinear noise. In nonlinear signal processing, the kernel-based method (KBM) creates a nonlinear mapping in the input data and feature space to reduce noise, which can be considered a noise-reduction technique using a kernel learning model. KBM is widely used in image denoising and speech enhancement due to its excellent denoising performance [20]. In UWA array processing, KBM-based array noise cancellation is a typical application scenario, and its structure is shown in Figure 2.



**Figure 2.** Schematic diagram of the adaptive noise-reduction structure of the KBM.

In the structure of Figure 2, assume that the transmit signal is  $s(t)$ , the receive signal is  $y(t) = s(t) + n(t)$ , and  $n(t)$  is the noise. Define the reference noise as  $\tilde{n}(t)$ , which is the noise collected before communication, and  $\tilde{n}(t)$  can be considered as an infinite approximation to  $n(t)$ . The KBM-ANC uses online iterations of  $y(t)$  and  $\tilde{n}(t)$  to obtain the estimated noise  $\hat{n}(t)$ , and then, the noise-reduction signal  $\hat{y}(t)$  is obtained by the cancellation.  $\hat{y}(t)$  can be considered as a constant approximation fit to  $s(t)$ . The key part of the KBM-ANC is an adaptive filter. The adaptive filter of the KBM-ANC can be classified as a single-core function type and a multi-core function type [21]. The representative

algorithms are kernel recursive least squares (KRLS) [22], reproducing kernel Hilbert space (RKHS) [23], and the kernel affine projection algorithm (KAPA) [24]. It is worth noting that the KBM-ANC performs very effectively in acoustic array noise cancellation, but is not very suitable for UWA channels with complex-type characteristics. The above KBM-ANC algorithm is mostly applied to noise-reduction of real-type signals. At the same time, it has the minimal capability to handle complex-type signals.  $s(t)$  passes through the channel, resulting in  $y(t)$ .  $y(t)$  is a complex structure containing amplitude information in its real part and phase information in its imaginary part. Phase information is an essential part when performing STD processing [25]. If the phase calculation is inaccurate, it will hinder the time delay alignment of the beam formation in STD, affect diversity combining, and degrade the communication performance.

Structurally, the noise cancellation system of the KBM is already very close to the simplest form of the deep neural network (DNN), and both contain vital parts such as the kernel function filtering structure, learning mechanism, and online iteration [26]. Recently, the DNN has been widely used in image and speech noise-reduction and has made breakthroughs. DNN models have been the trend of conventional noise-reduction algorithms. The DNN has a more flexible hyperparameter approach and feature-rich modules [27]. However, there is little research on noise-reduction of UWA signals with the DNN. Researchers mainly focus on weak signal identification and signal type classification [28,29].

Using the direct sequence spread spectrum (DSSS) technique to optimize the transmission structure can significantly simplify the computational system of the equalizer in STD [30]. Code division multiple access (CDMA) is a specific DSSS technology application. CDMA signals have a delay extension in the UWA channel transmission, recognized as a signal re-transmission [31]. If the CDMA is used with the RAKE technique for transmitting and combining, the main path signal and the multipath signal can be summed according to a linear relationship. The above operation can simplify the structure of the equalizer and even omit the equalizer. However, the use of CDMA as the transmit signal introduces the problem of multi-access interference (MAI) [32]. CDMA applies different address codes to distinguish each transmit signal. It easily makes the received signal aliased in the time-domain and susceptible to co-channel interference (CCI) in the frequency-domain [33]. The STD structure using CDMA as the transmit signal does not effectively cope with MAI; although the computing efficiency is effectively improved, the communication effect is not improved.

In the STD structure of Figure 1, the space-domain receiving uses the beamforming method. Beamforming performs time delay estimation on the filtered  $y(t)$ , aligns each path of the signal, and then, performs time-domain processing. The time-domain equalization algorithm generally uses the least-mean-squares (LMS) method and the recursive least-squares (RLS) method [34,35]. The LMS method achieves the update of the adaptive filter weight parameters by the least mean squares, while the RLS method achieves the update of the adaptive filter weight parameters by the least squares and converges faster than the LMS method. Although both the LMS and RLS methods can effectively solve the multipath effect, an equalizer needs to be connected after each beamformer [36,37]. The computational complexity increases dramatically as the number of received array elements increases.

In summary, the KBM-based array noise cancellation technique can effectively cope with nonlinear noise environments in terms of noise-reduction. Still, it is less capable of handling underwater noise with complex-type characteristics. In the STD structure, the DSSS technique can optimize the transmit signal with the RAKE technique to simplify the equalization operation. However, it still leads to MAI. We propose an STD method based on a noise-reduction learning model to solve the above problems, which performs noise-reduction and diversity combining in a phased manner. The main work of this paper can be summarized as follows:

- To effectively suppress the nonlinear components contained in the VURN, we propose a CGAN-based residual cancellation learning model by referring to the structure of the KBM-ANC. The proposed model puts the received signal  $y(t)$  and the reference

noise  $\tilde{\mathbf{n}}(t)$  into the CGAN. A nonlinear mapping is created between them by performing “generative-adversarial”-based learning in CGAN to estimate noise  $\hat{\mathbf{n}}(t)$ . Then, the residual cancellation learning is performed by the RCL module to obtain the final noise-reduction estimate  $\hat{\mathbf{y}}(t)$ .

- To efficiently handle the complex-type characteristics of UWA signals, we propose a pre-processing structure for complex-type orthogonal compression. The proposed structure performs a complex convolution operation on the UWA signal to obtain separated real and imaginary features. The separated features are trained separately for noise-reduction to avoid the loss of complex-type data features. The separation of real and imaginary parts can lead to orthogonality corruption. To minimize the corruption, we normalize the orthogonal scale by compressing the real and imaginary features orthogonally. The proposed pre-processing method effectively fits the structure of the DM in the complex-CGAN, together with CReLU, and effectively improves the performance of the noise-reduction learning model for complex-type data.
- To effectively reduce the MAI due to CDMA signals, we propose an improved STD model. The model matches a weight matrix for each array element at the receiver. The model allows the main path signal to be enhanced and effectively reduces the influence of other multipath signals. Thus, the model reduces the correlation between signals and suppresses MAI in the system.

The rest of this paper is organized as follows. In Section 2, we compile typical literature and methods in noise-reduction and STD processing and summarize the advantages and disadvantages of each method to complete the summary of related works. Section 3 discusses the transmit–receive model of the underwater acoustic array, specifying the mathematical model of the transmit signal and noise. In Section 4, the structure of the proposed joint model is described in detail. It mainly includes the pre-processing data methods, the main components of the noise-reduction learning model, and the STD model based on the weight update. Finally, the evaluation and result analysis are shown in Section 5. In Section 6, conclusions are given.

## 2. Related Works

UWA communication has also been put forward for higher requirements with the continuous exploitation of global marine resources. The literature [38] pointed out that the two critical factors affecting UWA communication are the multipath effect and noise interference. The literature [39] introduced the diversity technique from base station communication into the underwater array to suppress the multipath effect. The BER is reduced to  $10^{-3}$  by increasing the receiver unit and using the maximal ratio combining (MRC) strategy to obtain the optimal diversity gain during the combining [40]. Due to the simple structure, the MRC method cannot work well in an environment with severe multipath effects. A differential channel method (DCM) using array communication was mentioned in the literature [41], where the estimation of the main path signal is obtained through beamforming. Then, the differential structure of the channels between multiple array elements is used to enhance the communication capability. However, it is difficult to improve the communication performance of the DCM when the number of array elements exceeds a threshold. With the continuous development of diversity technology, STD structures that combine spatial domain receiving and time-domain equalization processing have been applied to underwater communications [42]. STD can cooperate with the adaptive filter to complete the noise-reduction process and maintain the advantage of communication gain enhancement after the diversity combining. However, the complex operation seriously affects its operational efficiency. In the literature [43], a DS-CDMA signal based on the code filtering method was proposed to optimize the transmitting structure. The most significant advantage of this method is simplifying the equalization operation. The literature [44] used the 2D-RAKE technique to perform diversity combining. It pointed out that although the RAKE technique can define the energy of each branch by

weight assignment, the multi-access interference due to CDMA signals cannot be avoided and can seriously affect the quality of the communication.

In the earlier days, adaptive filtering was often applied in image noise-reduction. With the continuous development of adaptive filtering technology, UWA communication has started to apply this technology. The literature [45] proposed an RLS algorithm for linear noise-reduction processing of images. The literature [46] applied this algorithm to underwater communication and showed excellent results for simple noise. In [47], this literature considered VURN as the main reason affecting the UWA array communication, which can be viewed as a superposition of multiple linear spectra and continuous spectra in the frequency-domain. VURN generally exhibits convolutional properties, a nonlinear noise structure. The linear denoising methods, such as RLS, are not suitable. In [48], an array method was proposed for noise cancellation of sound signals, and adaptive filtering techniques were applied to nonlinear denoising structures. In the following years, ACN technology has developed rapidly, and the nonlinear algorithm represented by the KBM has shown excellent denoising performance in array communication. Typical KBM algorithms include RKHS [49] and KAPA [50]. The RKHS method accomplishes the estimation of errors in the Hilbert space by creating a nonlinear mapping of a high-dimensional space. KAPA transforms the affine projection algorithm into a kernel function to accomplish the filtering of noisy signals. RKHS was applied in the literature [51] to denoise hydroacoustic signals for hydroacoustic signal detection. However, unlike the STD processing of underwater array signals, the signal detection does not require phase solving, and the algorithm does not focus on the complex characteristics of the channel. A complex kernel least mean-squared (CKLMS) algorithm with a complex kernel function structure was mentioned in the literature [52]. However, it does not apply to radiation noise suppression with nonlinear characteristics. Structurally, RKHS and KAPA have converged very closely to the radial basis function neural network (RBFNN) [53], both possessing the core mechanisms for online learning. Although RBFNN is a simple form of the DNN, it can still be parameterized by three implicit layers. It can perform adaptive feature extraction on the data at the input side, which are features that KBM algorithms do not possess.

The literature [54] mentioned an improved RBFNN for the image filtering structure, and the results showed superiority over the linear adaptive noise-reduction algorithm. They further optimized the structure of the filtering process by replacing the radial basis kernel function with a convolutional kernel based on the literature [55]. A GAN-based controlled learning noise-reduction method was also proposed in the literature [56]. It trains its model by generators and adjudicators and is more effective than the CNN methods for noise-reduction. The literature [57] pointed out that the acoustic signal can be expanded into a two-dimensional image in both time–frequency dimensions. This theory is one of the bases for image noise-reduction methods for acoustic signals. Due to the complex-type nature of the UWA signal in the STD structure, the above methods are not applied to UWA array communication directly. The literature [58] proposed to characterize the acoustic signal by representing it as real and imaginary parts for speech enhancement. The literature [59] mentioned a DNN structure that can be used for complex-type input data by constructing complex convolutional modules. Although the above studies collectively showed that it is feasible to use the DNN method as a UWA array communication denoising and handle the complex-type characteristics of UWA signals, performing noise-reduction processing and MAI suppression in the STD structure is still challenging.

### 3. System Model for UWA Communication

In this section, the transmit–receive model of UWA array communication and the structural characteristics of VURN are described.

This research focuses on a single-input multiple-output (SIMO) system. A typical SIMO array communication model is shown in Figure 3. When the towing ship is close to the receiving array, the noise is mainly generated by the mechanical vibration of the power source and propeller friction [60]. These noises will be superimposed together with

the multipath to be received by the receiving array. The red line in Figure 3 indicates the main path, and the remaining black dashed lines are the multipath signals generated by reflections from the sea surface and the seafloor.

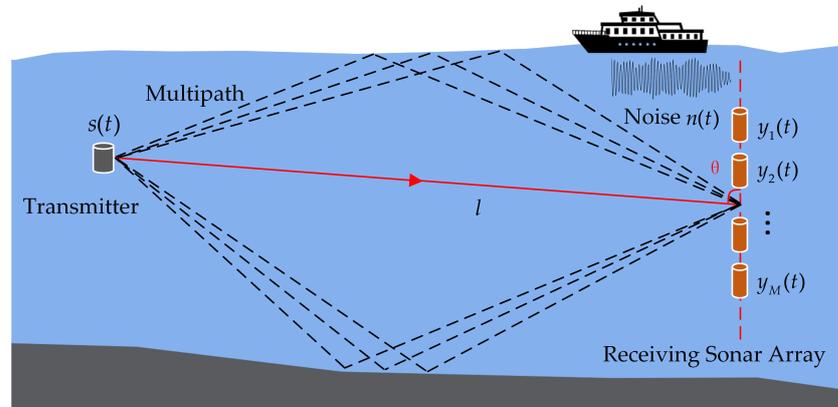


Figure 3. A UWA array transmit–receive structure of a SIMO system.

In the SIMO system, the baseband signal at the transmitter is  $\mathbf{s}(t)$  and the received signal is  $\mathbf{y}(t)$ .  $\mathbf{s}(t)$  is reflected to produce  $L$  multipath signals, and the main path is labeled as  $l$ . The angle between the principal path and the baseline is called the direction of arrival and is denoted as  $\theta$ .  $\mathbf{y}(t)$  can be expressed as:

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \\ \vdots \\ \mathbf{y}_M(t) \end{bmatrix} = \sum_{l=1}^L \delta_l h_l \mathbf{s}(t + \tau_l) + \mathbf{n}(t), \tag{1}$$

where  $\tau_l$  and  $\delta_l$  are the time delay and channel fading of the  $l$ th path signal to the receive array, respectively.  $A_l$  is the channel impact response of the  $l$ th path signal arriving at the receive array. The transmit signal uses CDMA, and  $\mathbf{s}(t)$  in Equation (1) can be expressed after spread spectrum processing as:

$$\mathbf{s}(t) = \sqrt{P} \mathbf{b}(t) \sum_{n=0}^{N_0-1} \mathbf{c}(k) \mathbf{p}_c[(t - N_0)T_c], \tag{2}$$

where  $P$  is the transmit power of the signal,  $\mathbf{b}(t)$  is the information bits,  $\mathbf{c}(k)$  is the spreading code sequence,  $T_c$  is the spreading code period,  $N_0$  is the spreading code length, and  $\mathbf{p}_c(t)$  is the pulse waveform of the code slice.

$\mathbf{n}(t)$  is the noise, i.e., VURN, which is the focus of this paper. When the tugboat is close to the array, the VURN can radiate through the seawater to the receiving array, causing a severe impact on the communication. The sources of VURN generation are the power system of the tugboat, the propeller blade friction, and hydrodynamic noise. VURN is generally considered as a superposition of linear spectra and continuous spectra [61], and  $\mathbf{n}(t)$  can be expressed as:

$$\mathbf{n}(t) = \mathbf{L}(t) + \mathbf{C}(t), \tag{3}$$

where  $\mathbf{L}(t)$  denotes the line spectrum component,  $\mathbf{C}(t)$  denotes the continuous spectrum component, and  $\mathbf{L}(t)$  is composed of multiple sinusoidal functions, which the following equation can obtain:

$$\mathbf{L}(t) = \sum_{i=1}^N A_i \sin(2\pi f_i t + \varphi_i), \tag{4}$$

where  $A_i, f_i, \varphi_i$  denote the amplitude, frequency, and initial phase, respectively;  $f_i$  satisfies  $f_i = n_b f_p$ ;  $n_b$  is the number of blades of the propeller;  $f_p$  is the shaft frequency of the propeller;  $N$  denotes the number of line spectra. Similarly,  $\mathbf{C}(t)$  can be obtained by the following equation:

$$\mathbf{C}(t) = \sum_{i=1}^M \frac{A_i}{\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma_i}} (t - iT_b), \tag{5}$$

$M$  denotes the number of continuous spectral duration periods;  $T_b$  is the time interval between continuous signals,  $T_b = 1/f_i$ ;  $\sigma_i$  denotes the degree of mixing between continuous signals. For the simulated system, noise  $\mathbf{n}(t)$  and reference noise  $\tilde{\mathbf{n}}(t)$  are the same data. However, the reference noise needs to be collected before the communication for the existing SIMO system. In shorter periods, we consider that the reference noise  $\tilde{\mathbf{n}}(t)$  and  $\mathbf{n}(t)$  are infinitely close to each other.

#### 4. A Joint Denoising Learning Model for Weight Update STD Method

In this section, firstly, the structure of the proposed method is presented. Secondly, the characteristics of UWA data are analyzed in detail, and a complex-type orthogonal compression pre-processing structure is designed. Then, a noise cancellation learning (NCL) model based on the complex-CGAN is constructed. Finally, an STD model based on a weight update strategy is proposed to suppress MAI. As Figure 4 shows the general structure of the proposed method, it can be considered staged processing.

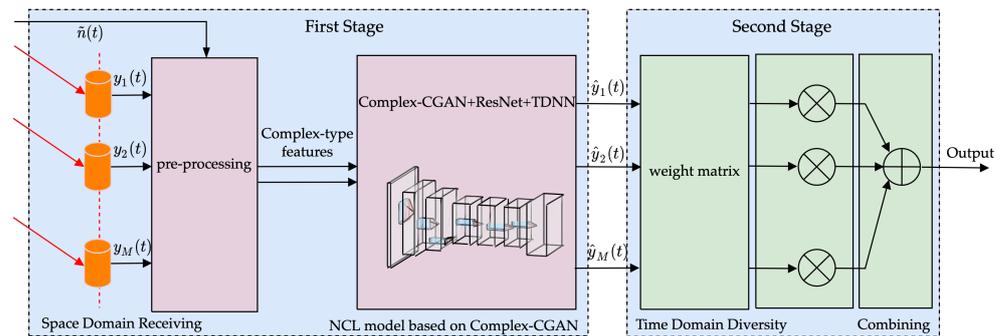


Figure 4. Two-stage structure of the proposed method.

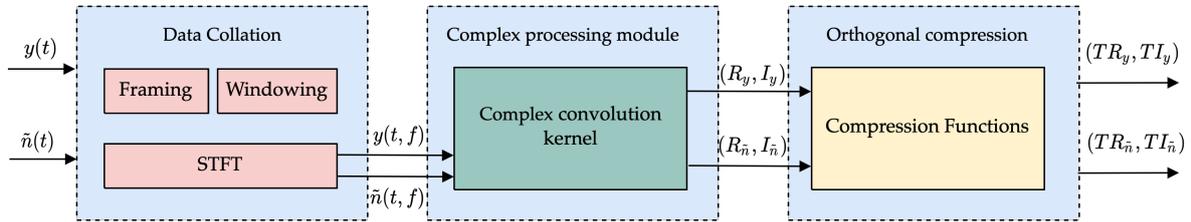
In the first stage, the UWA data are firstly received and collated in airspace. Next, the received signal  $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_M(t)]$  and the reference noise  $\tilde{\mathbf{n}}(t)$  are pre-processed to obtain the complex-type features that can be processed by the noise-reduction learning model. Then, the features are fed into the model, which mainly consists of complex-CGAN, ResNet, and TDNN, responsible for adaptive learning, noise cancellation, and time-series restoration, respectively. Finally, the output  $\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_M(t)]$  of the noise-reduction estimation is obtained.

In the second stage, the noise-reduction estimate  $\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_M(t)]$  is fed into the improved weight update model. The model collates the results of beamforming into a matrix of corresponding weights while maximizing the weighting of the main path signal and suppressing signals on other paths. In this way, it is possible to suppress the MAI caused by spread spectrum signals for transmitting signals. Finally, the output of the STD is obtained by the combining strategy, which has a lower BER.

##### 4.1. The UWA Signal Analysis and Pre-Processing

The signal transmitted through the UWA channel has the characteristics of a complex-type structure. At the same time, the UWA signal has short-time smoothness, i.e., it can behave as a quasi-steady-state process in a short period. In addition to having the characteristics of an acoustic signal, it can be represented in two dimensions by the time-

frequency relationship and has the structural characteristics of an image. Based on the above characteristics, we propose a complex-type orthogonal compression pre-processing structure, the composition of which is shown in Figure 5.



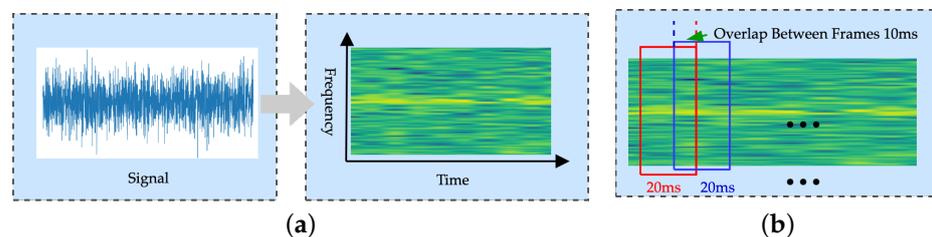
**Figure 5.** Complex-type orthogonal compression pre-processing structure.

The pre-processing structure consists of three parts: data collation, complex-type data processing, and orthogonal compression. In the pre-processing stage,  $\mathbf{y}(t)$  and  $\tilde{\mathbf{n}}(t)$  will be separated and compressed into  $(\mathbf{TR}_y, \mathbf{TI}_y)$  and  $(\mathbf{TR}_{\tilde{n}}, \mathbf{TI}_{\tilde{n}})$  after being processed in the above three steps.  $(\mathbf{TR}_y, \mathbf{TI}_y)$  and  $(\mathbf{TR}_{\tilde{n}}, \mathbf{TI}_{\tilde{n}})$  are used as input features for the complex-CGAN, which makes the noise-reduction learning model cope with complex-type data with better performance. The specific pre-processing steps are as follows:

(1) Data collation. Since the UWA signal has a short-time stable characteristic, we need to crop the UWA data to be within the short-time smooth interval [62]. The UWA signal can show relatively smooth amplitude characteristics at 20–30 ms, suitable for feature extraction. The data can be cropped in frames using windowing, and we set the frame length of each data to 20 ms and the overlap length between frames to 10 ms. A window function was used in the framing to reduce the spectral energy leakage, and the Hamming window was used for the window function [63], which takes the value of 0.53836, and  $N$  is the window length; the window function can be expressed as:

$$\text{window}(n) = \alpha_0 - (1 - \alpha_0) \cos\left(\frac{2\pi n}{N - 1}\right), \quad (6)$$

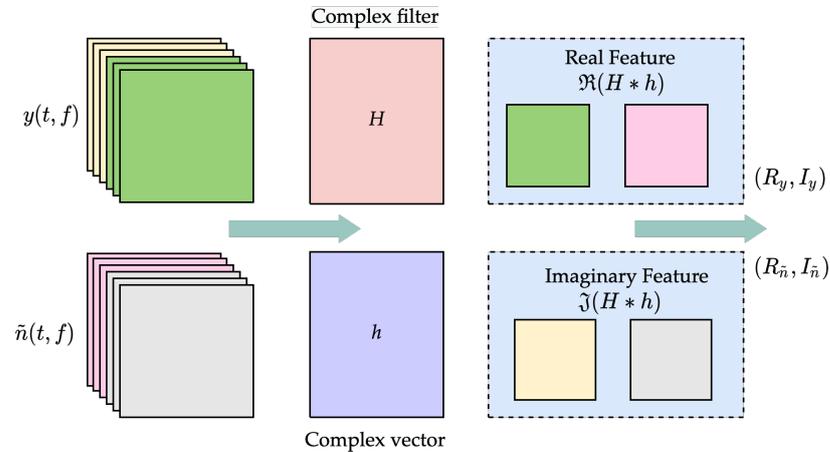
Short-time Fourier transform (STFT) is applied to each frame to obtain the time–frequency two-dimensionalization matrix. For  $\mathbf{y}(t)$  and  $\tilde{\mathbf{n}}(t)$ , the corresponding STFT results are  $\mathbf{y}(t) \rightarrow \mathbf{y}(t, f)$  and  $\tilde{\mathbf{n}}(t) \rightarrow \tilde{\mathbf{n}}(t, f)$ . The time–frequency two-dimensional form of the received signal  $\mathbf{y}(t)$  is shown in Figure 6a. Figure 6b shows the data framing strategy for the pre-processing stage.



**Figure 6.** Data collation of UWA signals: (a) the time–frequency two-dimensional form of the received signal; (b) the data framing strategy for the pre-processing stage.

(2) Complex processing module. Since the noise-reduction learning model needs to process complex-type data, we designed a complex processing module in the pre-processing stage, which is mainly used to extract complex features. For the received signal  $\mathbf{y}(t)$ ,  $\mathbf{y}(t) = \mathbf{R} + \mathbf{I} * j$ , its real part  $\mathbf{R}$  contains the amplitude information of the signal and the imaginary part  $\mathbf{I}$  contains the phase information. For this purpose, we used multiple complex convolution kernels in the complex processing module to perform feature extraction of complex-type data to obtain features with real and imaginary parts separated. Suppose

the complex vector of a target signal is set as  $\mathbf{h}$ . In that case, the complex filter matrix is  $\mathbf{H}$ ,  $\Re(\mathbf{H} * \mathbf{h})$  denotes the real part of the convolution operation, and  $\Im(\mathbf{H} * \mathbf{h})$  denotes the imaginary part; Figure 7 shows the convolution structure of the complex processing module.



**Figure 7.** Convolutional structure of complex processing module.

(3) Orthogonal compression. After processing by the complex processing module, the features with real and imaginary parts separated are obtained, and this separation can destroy the orthogonality of the signal. To minimize this damage, we need to perform orthogonality compression to ensure the uniformity of feature scales. The hyperbolic tangent function is used to compress the real and imaginary components, respectively, and the corresponding Tanh-compressed real component (TR) and Tanh-compressed imaginary component (TI) are obtained as the normalized input features:

$$\mathbf{TR}(t, f) = \frac{\beta(1 - e^{-\alpha Z_R})}{1 + e^{-\alpha Z_R}}, \quad (7)$$

$$\mathbf{TI}(t, f) = \frac{\beta(1 - e^{-\alpha Z_I})}{1 + e^{-\alpha Z_I}}, \quad (8)$$

where  $\alpha$  and  $\beta$  are the restriction factors, taking values of 0.5 and 10, respectively;  $Z$  denotes the objects to be compressed, i.e.,  $(\mathbf{R}_y, \mathbf{I}_y)$  and  $(\mathbf{R}_{\tilde{n}}, \mathbf{I}_{\tilde{n}})$ . After compression and normalization, the features change to  $(\mathbf{TR}_y, \mathbf{TI}_y)$  and  $(\mathbf{TR}_{\tilde{n}}, \mathbf{TI}_{\tilde{n}})$ .

#### 4.2. Noise-Reduction Learning Models Based on Complex-CGAN

As one of the KBM adaptive filters, the algorithm represented by KRLS cannot effectively cope with the UWA signal with complex structural characteristics. The core of the KRLS algorithm is an iterative learning mechanism by which a nonlinear mapping is created between the received signal and the acquisition noise. A radial basis function (RBF) neural network, similar to KRLS, is a single hidden layer feedforward neural network that accomplishes online learning by using RBF as the activation function. RBFNN has more adjustable parameters than the KRLS algorithm. RBFNN, with the addition of complex convolutional units, can handle complex-type data, but still has the problems of simple structure and easy structure loss of feedback information. Although CNNs have been successfully applied in image noise-reduction processing, there is a drawback of unstable convergence for UWA data because such network models do not have a judgment feedback mechanism. In recent years, GAN models have improved noise-reduction performance for nonlinear and nonstationary signals by introducing the structure of generation-discrimination. The GAN consists of a generative model (GM) and a discriminative model (DM), alternately optimized and trained until they reach a steady equilibrium state [64]. The CGAN is a derivative upgrade from the GAN that can add conditional attributes to the samples, further increasing the accuracy of the parameters in the feedback mechanism.

Inspired by the above algorithm structure, this paper proposes a noise-reduction learning model dedicated to UWA data features. Specifically, the model is a complex-CGAN noise-reduction network based on residual cancellation learning (RCL). The structural composition of the noise-reduction model is shown in Figure 8. In the following, we illustrate the operation process of the noise-reduction model with the learning process of the real part in the complex-type signal.

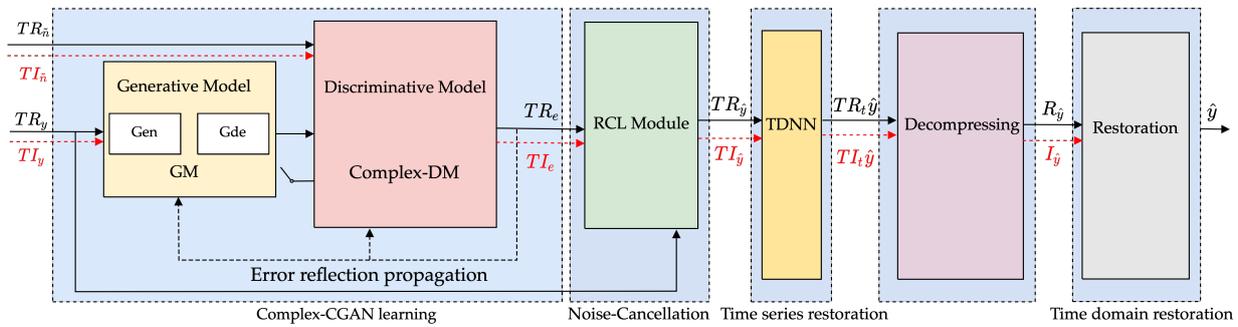


Figure 8. Structural components of the noise-reduction learning model.

We added a complex convolutional kernel unit to the original CGAN, which can turn the DM into the complex-DM. The complex-DM is combined with the complex activation function CReLU to form the basic structure of the complex-CGAN, which is capable of handling complex-type data. The GM can be considered a feature generation network by which noise-reducing features are generated that are realistic. The DM is a discriminative network that constantly feeds back to correct the generation quality of the GM by discriminating the differences between noise-bearing features and denoised reference features. The model is alternatively optimized and trained until a Nash equilibrium is reached and the final noise-reduced estimate is output. Specifically, for the complex-CGAN, its input is the pre-processing result  $TR_y$ . After  $TR_y$  enters the GM for the first time, it generates an initialization value, which will be constantly compared and fed back to  $TR_{\tilde{n}}$  for learning. The GM contains the encoding part Gen and the decoding part Gde [65]. Gen gradually discards the redundant information in the original high-dimensional data by the dimensionality reduction of  $TR_y$ , outputs the valuable information in the signal as a low-dimensional vector, and then, reconstructs the high-dimensional features by Gde. Figure 9 shows the detailed structural setup of the GM in the complex-CGAN.

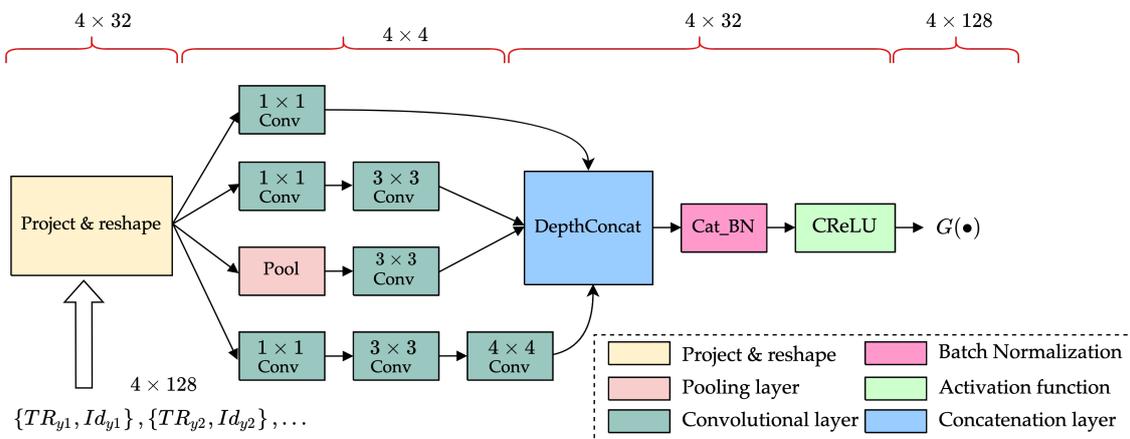
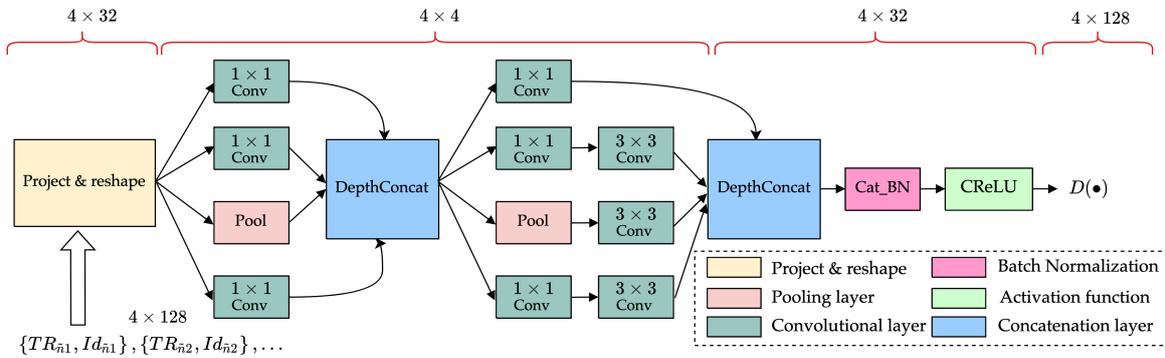


Figure 9. The detailed structural setup of the GM in the complex-CGAN.

Then,  $TR_y$  and  $TR_{\tilde{n}}$  are sequentially input into the DM for determination and identification until a steady-state mapping distribution is reached. By feedback learning through

error reflection propagation between  $\mathbf{TR}_y$  and  $\mathbf{TR}_{\tilde{n}}$ , the accuracy of adversarial training is continuously enhanced [66]. After the above feedback learning, we can assume that the generative ability of the GM and the discriminative ability of the DM will be improved. The purpose of this stage is to estimate a noise component  $\mathbf{TR}_e$  by learning the mapping so that it infinitely approximates  $\mathbf{TR}_{\tilde{n}}$ . Figure 10 shows the detailed structural setup of the DM in the complex-CGAN.



**Figure 10.** The detailed structural setup of the DM in the complex-CGAN.

The input  $\mathbf{TR}_y$  of the GM is collapsed into  $\{\mathbf{TR}_{y1}, Id_{y1}\}, \dots, \{\mathbf{TR}_{yi}, Id_{yi}\}$ , where  $Id$  is to label the input features, and the input feature size is  $4 \times 32$ . Similarly, the set of features of DM is obtained as  $\{\mathbf{TR}_{\tilde{n}1}, Id_{\tilde{n}1}\}, \dots, \{\mathbf{TR}_{\tilde{n}i}, Id_{\tilde{n}i}\}$ . In the GM, the feature set is adjusted to a  $4 \times 32$  vector of data features after reshaping. Then, after the pooling layer, it will be adjusted to a  $4 \times 4$  downsampling matrix. In the DM, although there are two pooling layers, they will still be adjusted to a  $4 \times 4$  feature matrix. We used categorical conditional batch normalization (CatBN) as a criterion in batch normalization, which can further reduce orthogonality loss. The activation function CReLU is a complex-type ReLU that satisfies the Cauchy–Riemann equation [67]. In this paper, a loss function with the  $L1$  loss term  $\zeta_{L1}(G)$  is used to optimize the network [68] with the following objective function:

$$G^* = \arg \min_{G, D} V_{CGAN}(G, D) + \lambda \zeta_{L1}(G), \quad (9)$$

$$\zeta_{L1}(G) = E_{\tilde{n}, y} \left[ \|G(\mathbf{TR}_{yi}, \mathbf{TR}_{\tilde{n}i})\|_1 \right], \quad (10)$$

where  $\lambda$  is the weight of the  $L1$  loss term and  $G^*$  denotes the optimized GM.  $E[\cdot]$  denotes the estimated value after each round of calculation, and for  $\mathbf{TR}_y$ , the estimated value is  $\mathbf{TR}_e$ .  $G[\cdot]$  and  $D[\cdot]$  denote the outputs of the GM and DM [69], respectively, which are both upsampled to a matrix of size  $4 \times 128$ . The GM has 3 convolutional layers of  $1 \times 1$ , 3 convolutional layers of  $3 \times 3$ , and 1 convolutional layer of  $4 \times 4$ . The DM has 6 convolutional layers of  $1 \times 1$  and 3 convolutional layers of  $3 \times 3$ .  $V_{CGAN}(G, D)$  denotes the standard CGAN loss function:

$$\min_G \max_D V_{CGAN}(D, G) = E_{\tilde{n}, y} [\log D(\mathbf{TR}_{yi}, \mathbf{TR}_{\tilde{n}i})] + E_{\tilde{n}, y} [\log(1 - D(\mathbf{TR}_{yi}, \mathbf{TR}_{\tilde{n}i}))], \quad (11)$$

When Equation (9) reaches the Nash equilibrium, the GM and DM stop their operations, and the estimated value of the last round is obtained as  $\mathbf{TR}_e$ . Similarly, for the imaginary part input feature  $\mathbf{TI}_y$ ,  $\mathbf{TI}_e$  is also obtained after the training of this model.

The RCL module references the traditional ANC noise cancellation technique for residual learning, and it is worth noting that RCL is a dynamic process. When Equation (9) reaches the Nash equilibrium,  $\mathbf{TR}_y$  and  $\mathbf{TR}_e$  are processed by RCL, and the output obtains  $\mathbf{TR}_{\hat{y}}$  with a noise-reduction effect. The proposed model uses ResNet as the residual network and designs it as a combined inception and residual structure. The residual learning

module is also equipped with the same arithmetic function for complex-type data. The RCL structure used is defined in Figure 11. In the RCL module, for the real part, the inputs are  $\mathbf{TR}_y$  and  $\mathbf{TR}_e$ , both of which are  $4 \times 128$  matrices. After RCL processing,  $\mathbf{TR}_{\hat{y}}$  is obtained. Similarly, for inputs  $\mathbf{TI}_y$  and  $\mathbf{TI}_e$ ,  $\mathbf{TI}_{\hat{y}}$  is obtained.

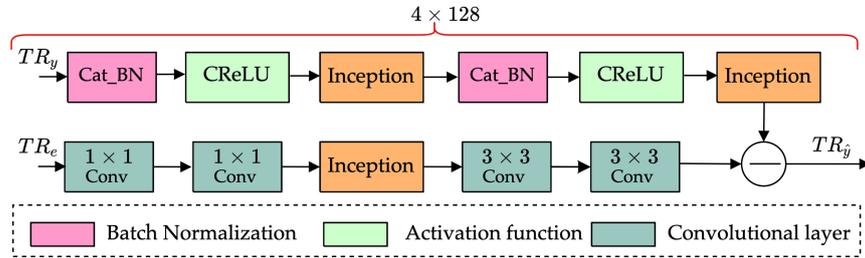


Figure 11. Structural composition of the RCL module.

Since the input to the system is a UWA signal with time-series characteristics, the time delay between the received signals needs to be obtained when STD processing is performed on the noise-reduction results. However, the time-series characteristics between signals are destroyed after using the noise-reduction learning process. To reduce the loss of temporal characteristics and make the estimation of time delay in the STD operation more accurate, the TDNN is used after the CGAN noise-reduction process. The TDNN is a deep learning network dedicated to time-series correlation processing, which can effectively restore the temporal characteristics of the data [70]. The TDNN stores the above temporal information through a time delay neuron (TDN). The network structure of the TDNN is shown in Figure 12.

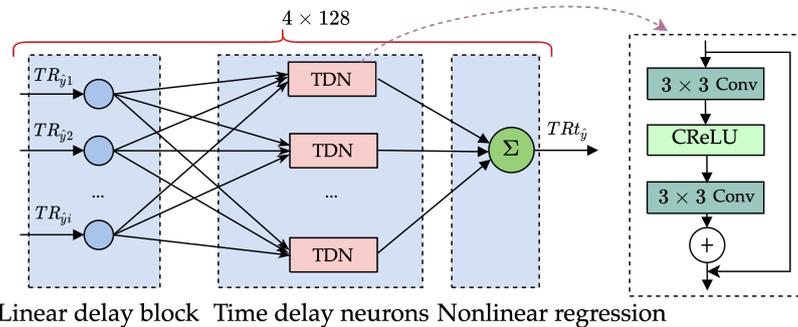


Figure 12. Network structure composition of TDNN.

The TDNN is a DL model with a simple structural form; the proposed model is structured as a three-layer network with input, hidden, and output layers. The input is the data after RCL processing, i.e.,  $\mathbf{TR}_{\hat{y}}$ , and the output is the time-series-adjusted  $\mathbf{TR}_{t\hat{y}}$ .  $\mathbf{TR}_{\hat{y}}$  and  $\mathbf{TR}_{t\hat{y}}$  are  $4 \times 128$  matrices. The input layer can be thought of as a time delay block for adjusting the time delay. The hidden layer consists of 64 TDNs; each TDN contains two convolutional layers and an activation function, and the convolutional layer size is  $3 \times 3$ . The TDNN has 128 convolutional kernels for convolutional operations in the hidden layer. The output layer is used to implement a nonlinear regression with time delay. Take  $\mathbf{TR}_{\hat{y}}$  as an example. When it is input into the TDNN structure, its current moment and the previous  $N$  moments are stored in the TDNN. The output of each TDN can be represented as Equation (12).

$$O_{TDN}(t) = fun \left( \sum_{i=1}^M \left[ \sum_{d=1}^N \mathbf{TR}_{\hat{y}_i}(t-d) \text{weight}_{id} + b_i \right] \right), \tag{12}$$

$O_{TDN}(t)$  is the output;  $M$  denotes the number of received signals;  $N$  represents the  $N$ th moment.  $\mathbf{TR}_{\hat{y}_i}$  indicates the  $i$ th input;  $weight$  is the TDN weight, where  $\mathbf{TR}_{\hat{y}_i}(t-d)$ ,

$d = 1, \dots, N$  is the  $i$ th input layer node and  $b_i$  is the bias of the  $i$ th input.  $fun$  is the excitation function. From Equation (12), it can be seen that the neuron output is determined by the current moment of each input and its data from the last  $N$  moments together. Therefore, the TDNN can effectively handle nonlinear dynamic timing problems. After the time-series adjustment, the output of each neuron in the output layer of the TDNN is shown in Equation (13):

$$O^r(t) = fun\left(\sum_{j=1}^M \sum_{d=1}^N \mathbf{H}^i(t-d) \times v_{jd}^r + c_j^r\right), r = 1, 2, \dots, R, \tag{13}$$

where  $r$  is the output layer node number and  $R$  is the number of output layers.  $v_{jd}^r$  is the TDNN weight of the connection between the  $r$ th output layer node and the implied layer node.  $v_j^r$  is the bias of the  $r$ th output layer node.  $\mathbf{H}^i$  indicates the input of the  $i$ th hidden layer.  $O^r(t)$  is the output, which is the result of each node being restored by the time-series, and this result is  $\mathbf{TR}_{t\hat{y}}$ . Similarly, when the input is  $\mathbf{TI}_{t\hat{y}}$ , the output is  $\mathbf{TI}_{t\hat{y}}$ .

Because compression is performed in the pre-processing stage, decompression is required during signal reduction, which the following equation can achieve:

$$\hat{\mathbf{R}}(t, f) = -\frac{1}{\alpha} \log\left(\frac{\beta - \mathbf{TR}_{t\hat{y}}}{\beta + \mathbf{TR}_{t\hat{y}}}\right), \tag{14}$$

$$\hat{\mathbf{I}}(t, f) = -\frac{1}{\alpha} \log\left(\frac{\beta - \mathbf{TI}_{t\hat{y}}}{\beta + \mathbf{TI}_{t\hat{y}}}\right), \tag{15}$$

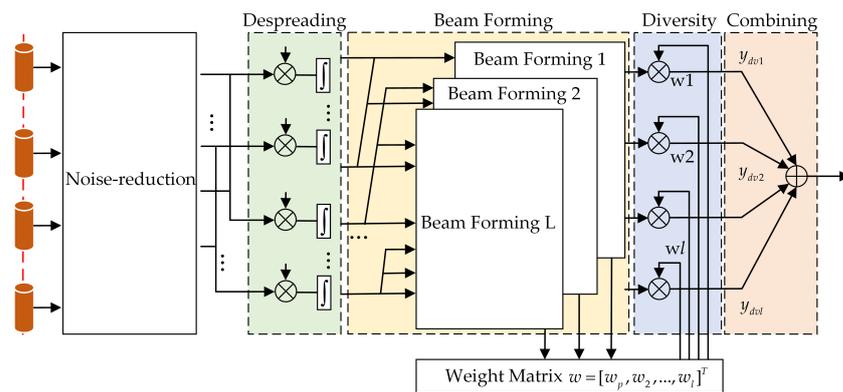
$\hat{\mathbf{R}}(t, f)$  and  $\hat{\mathbf{I}}(t, f)$  represent the real and imaginary parts after time-series restoration and decompression with the noise-reduction effect.  $\hat{\mathbf{R}}(t, f)$  and  $\hat{\mathbf{I}}(t, f)$  are restored to the time-domain by the inverse Fourier transform and combined to obtain Equation (16), where  $j$  denotes the imaginary part.

$$\hat{\mathbf{y}}(t) = \hat{\mathbf{R}}(t) + \hat{\mathbf{I}}(t) \times j, \tag{16}$$

$\hat{\mathbf{y}}(t)$  is the result of the first stage of noise-reduction about  $\mathbf{y}(t)$ .

### 4.3. The STD Model Based on Weight Update Strategy

The use of CDMA for the transmit signal effectively simplifies the time-domain equalization operation in the STD structure. Still, MAI cannot be avoided by the system, and this is because there is a certain correlation between the signals in each path, while the received signals are mixed in the time- and frequency-domains, resulting in a severe deterioration of the SNR, which causes MAI. The conventional STD model only targets the primary path signal of the desired user, while other path signals are suppressed as interference and cannot effectively suppress MAI. For this purpose, we propose an improved STD model in the second stage, matching a weight matrix for each array element at the receiver end. The model enables the main path signal to be enhanced and effectively reduces the influence of other multipath signals, thus reducing the correlation between signals and suppressing MAI in the system. The received signals of each path are then combined by the maximum SNR principle, which significantly improves the BER of the system. Figure 13 shows the structural composition of the STD model based on the weight update strategy.



**Figure 13.** The structural composition of the STD model based on the weight update strategy.

In the above structure, how to obtain the weight matrix accurately is the key to the algorithm. The weight matrix is used as a constraint that maximizes the signal’s weight direction. At the same time, the signals from other multipath directions are suppressed as much as possible. Considering the path  $p$  that first arrives at the receiver unit as the desired signal, assume the existence of a weight update matrix:

$$\mathbf{w} = [\mathbf{w}_p, \mathbf{w}_2, \dots, \mathbf{w}_L]^T, \tag{17}$$

Then, the DOA of the desired direction is obtained by the array processing method, denoted as  $\theta_p$ . To ensure that the desired signal from direction  $\theta_p$  is received correctly and to maximize the suppression of the other  $l-1$  branches’ interference, the constraint of  $\mathbf{w}_p$  can be written as Equation (18).

$$\mathbf{z} = \begin{cases} \mathbf{w}_p \mathbf{A}_l(\theta_p) = 1 \\ \mathbf{w}_i \mathbf{A}_l(\theta_l) = 0; i = 1, \dots, L - 1 \end{cases} \tag{18}$$

$L$  indicates the number of beamformers. By weighting and summing the outputs of each array element, the array beam is directed in one direction to achieve directional reception of the signal, thereby enhancing the signal and suppressing interference. The output power  $P_{\text{out}}$  of the received signal is denoted as:

$$P_{\text{out}} = \mathbf{E}\{|\mathbf{z}(t)|^2\} = \mathbf{E}\left\{\left[\mathbf{w}^H \mathbf{y}(t)\right] \left[\mathbf{y}^H(t) \mathbf{w}\right]\right\} = \mathbf{w}^H \mathbf{R}_{yy} \mathbf{w}, \tag{19}$$

In the above Equation (19),  $\mathbf{R}_{yy}$  is the covariance matrix of the received signal. In the algorithm of this paper, the linearly constrained minimum variance (LCMV) beamformer is used. The optimal weights are based on the beamformer output power being the minimum. It is also guaranteed that the output of the desired signal is 1 in Equation (18). The weight vector of each receiving array element can be found by Equation (20)

$$\mathbf{w}_p = \arg \min_{\mathbf{w}} P_{\text{out}} \text{ s.t. } \mathbf{w}_p^H \mathbf{A}(\theta_p) = 1, \tag{20}$$

According to Equation (20), the output signal of each receiving array element, i.e., the output  $\mathbf{y}_{dv}(k)$  of each diversity receive branch, is

$$\mathbf{y}_{dv}(k) = \mathbf{w}_p^H \mathbf{y}(k) = \mathbf{w}_p^H \left( \sum_{l=1}^L \delta_l \mathbf{A}_l \mathbf{s}_l(k) \right) + \mathbf{w}_p^H \left( \sum_{l=1}^L \mathbf{A}_l \mathbf{s}(k) + \mathbf{n}(k) \right), \tag{21}$$

$\mathbf{y}(k)$  is a discretized representation of  $\mathbf{y}(t)$ , where the continuous-time variables  $t = kT_b, p = 1, 2, \dots, L$ .

STD combining follows the space-domain receiving and time-domain weighting processing. It combines the received signals of each diversity branch according to the corresponding strategy. The received signals are combined according to the maximum

SINR. The power of the transmitted signal is denoted as  $D[s(t)] = E_b$ . The transmitting symbols are independent of each other, then the output power of the transmitting signal in the  $p$ th diversity branch is

$$N_p = D \left[ \mathbf{w}_p^H \left( \sum_{l \neq p}^L \delta_l \mathbf{A}_l \mathbf{s}_l(t - \tau_p) \right) + \mathbf{w}_p^H \left( \sum_{l=1}^L \mathbf{A}_l \mathbf{s}_l(t - \tau_p) + \mathbf{n}(t) \right) \right], \quad (22)$$

Then, the SINR of the  $p$ th diversity receive branch is

$$\text{SINR}_p = \frac{\sigma_p}{N_p}, \quad (23)$$

The purpose of the STD combining is to maximize the value of  $\text{SINR}_p$  taken. The signal received by each diversity branch is multiplied by the gain factor  $\sqrt{N_n/N_p}$ , so that it has the same noise output power  $N_n$ .  $\sigma_p$  is the output power of the  $p$ th path signal, space domain then multiplied by the weighting factor  $\eta_p$ .  $\eta_p$  is the eigenvalue of the weight matrix  $\mathbf{w}_p$ . The combined signal is denoted as:

$$\mathbf{y}(k) = \sum_{p=1}^L \eta_p \sqrt{\frac{N_n}{N_p}} \mathbf{w}_p^H \mathbf{s}(t - \tau_p), \quad (24)$$

The power of the useful signal in the STD combining signal is:

$$E_c = \left( \sum_{p=1}^L |\eta_p| \sqrt{\frac{N_n}{N_p}} \cdot \sigma_p \right)^2 \quad (25)$$

The power of interference and noise in the STD combining signal is:

$$N_c = \sum_{p=1}^L |\eta_p|^2 N_n \quad (26)$$

According to Equation (23), the SINR of the combined STD signal is obtained using Schwarz's inequality as:

$$\text{SINR}_c = \frac{E_c}{N_c} = \frac{\left( \sum_{p=1}^L |\eta_p| \sqrt{\frac{\sigma_p}{N_p}} \right)^2}{\sum_{p=1}^L |\eta_p|^2} \leq \frac{\left( \sum_{p=1}^L |\eta_p|^2 \right) \sum_{p=1}^L \left( \frac{\sigma_p}{N_p} \right)}{\sum_{p=1}^L |\eta_p|^2} = \sum_{p=1}^L \left( \frac{\sigma_p}{N_p} \right) \quad (27)$$

When  $|\eta_p| = \sqrt{\sigma_p/N_p}$ , the SINR can reach the maximum value  $\sum_{p=1}^L (\sigma_p/N_p)$ . This results in a combined output with the best communication effect.

After a thorough analysis of the underwater environment, this paper argues that the UWA signal can refer to the characteristics of the sound and image when applying the DL model. After that, the proposed noise-reduction learning model is applied to complete the noise-reduction process for the received signal. Then, the optimal combining of the noise-reduced signals is achieved using the proposed STD model based on the weight update strategy. An improved STD model effectively suppresses MAI. The final combined signal with a lower SNR is obtained; theoretically, the received signal has a lower BER. We conducted simulation tests and on-lake tests to verify the effectiveness of the proposed method.

### 5. Evaluation and Result Analysis

Numerical simulations and a lake trial were conducted to evaluate the performance of the proposed algorithms. The transmit signal is designed for data framing according to the structure shown in Figure 14. Under a low SNR, the time-bandwidth product of the linear frequency modulation (LFM) signal should be large enough for frame synchronization

detection. Thus, the detection probability of the received signal is improved by obtaining a more significant gain of the correlation operation. The first LFM signal is used in the correlation operation of the received signal to detect the implementation of frame synchronization and the time delay estimation of the multipath. The latter LFM signal is used to perform a coarse estimate of the Doppler frequency offset. In addition to the transmitted information bits, the modulated signal in the data frame carries a training sequence used to train the channel estimator to guarantee the convergence of its tap coefficients.



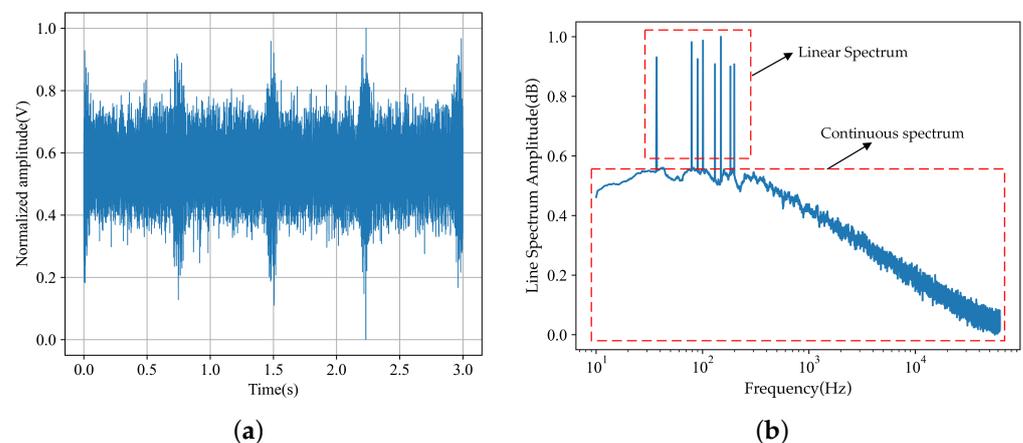
**Figure 14.** Design of transmitting signal structure.

The parameter settings of DS-CDMA in the simulations and lake trial are shown in Table 1. In the simulation system, the specifications of our computer were as follows: Intel(R) Core(TM) i5-6400 CPU 2.7 GHz (4 cores), 16 GB RAM, 1 TB memory with Windows and MATLAB installed.

**Table 1.** The parameter settings of DS-CDMA in the simulations and lake trial.

Parameters	Values of the Simulations and Lake Trial
Bandwidth	5 kHz
Carrier frequency	10 kHz
Spread spectrum code	7-order m-sequence
Filter roll-off factor	0.25
Bit rate	26.5 bps
Frame synchronization signal pulse width	LFM (50 ms)
Carrier synchronization	2-order phase-locked loop
Channel coding	1/2 convolutional codes

We simulated the noise with the amplitude  $A$  of the linear part of the spectrum in VURN as a normalized value,  $A = 0.3$ ; the number of propeller blades is  $n_b = 3$ ; the shaft frequency of the propeller is  $f_p = 100$  Hz; the number of line spectra is  $N = 8$ ; the initial phase is  $\varphi_i = 45$ . In the continuous spectrum, the duration period is  $M = 10$ ; the degree of mixing between consecutive signals is  $\sigma_i$ , which takes the value of 0.4. To make the simulation effect as close as possible to the lake test, we set the SNR as the broadband SNR, taking the value of  $[-20, 10]$  dB. Figure 15a shows the time-domain visualization results of VURN, and Figure 15b shows the spectral characteristics of VURN, which contains both linear spectra and continuous spectra.



**Figure 15.** Visualization results of the time-domain and frequency spectrum of VURN: (a) the time-domain visualization results of VURN; (b) the spectral characteristics of VURN.

In the simulation experiment, the input data are received signal  $\mathbf{y}(t)$  and reference noise  $\tilde{\mathbf{n}}(t)$ . After pre-processing, 4000 frames  $(\mathbf{TR}_y, \mathbf{TI}_y)$  and 4000 frames  $(\mathbf{TR}_{\tilde{n}}, \mathbf{TI}_{\tilde{n}})$  were obtained.  $(\mathbf{TR}_y, \mathbf{TI}_y)$  and  $(\mathbf{TR}_{\tilde{n}}, \mathbf{TI}_{\tilde{n}})$  denote the orthogonal compression features of  $\mathbf{y}(t)$  and  $\tilde{\mathbf{n}}(t)$ . The 3000 frames of data are the training dataset, and 1000 frames of data were used for the testing dataset. During the training of the noise-reduction learning model, the learning rate was set to 0.001, the learning step was 2, and the batch size was 16, and the training was complete after 400 epochs of data iterations. In the structure of the complex-CGAN, the input feature size of the GM was  $4 \times 32$ , which was downsampled into  $4 \times 4$  feature data after convolution and pooling layers. Then, processed by depthconcat and the activation function to form data of size  $4 \times 32$ , the final output upsampling result was  $4 \times 128$ . Similarly, the DM has an input size of  $4 \times 32$  and an output of size  $4 \times 128$ . In the RCL module, the data feature size was always kept at  $4 \times 128$ . In the TDNN operation, the data feature size was also  $4 \times 128$ . Since the input feature size is  $4 \times 128$  and there is eight input channels, the number of convolution channels of  $1 \times 1$  is 32; the number of convolution channels of  $3 \times 3$  is 64; the number of convolution channels of  $4 \times 4$  is 64; the convolution kernel parameters are  $4 \times 128 \times 8 + 1 \times 1 \times 32 + 3 \times 3 \times 64 + 4 \times 4 \times 64 = 5728$ .

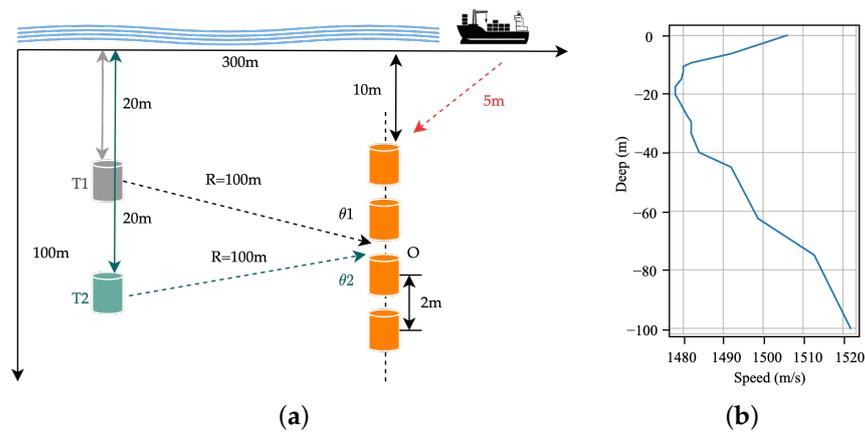
Table 2 shows the specific parameter settings of the proposed model in the simulation phase, which mainly contains the parameters of the complex-CGAN, RCL module, and TDNN.

**Table 2.** Specific network parameter settings for the noise-reduction learning model.

Network	Components, Kernel Size	Component Number
Complex-CGAN	Conv, $1 \times 1$	GM 3; DM 6.
	Conv, $3 \times 3$	GM 3; DM 3.
	Conv, $4 \times 4$	GM 1.
	Pool, $3 \times 3$	GM 1; DM 2.
	CatBN	GM 1; DM 1.
	CRReLU	GM 1; DM 1.
RCL module	Conv, $1 \times 1$	2
	Conv, $3 \times 3$	3
	Inception	1
	Pool, $3 \times 3$	2
	CRReLU	2
TDNN	Conv, $3 \times 3$	2
	CRReLU	1

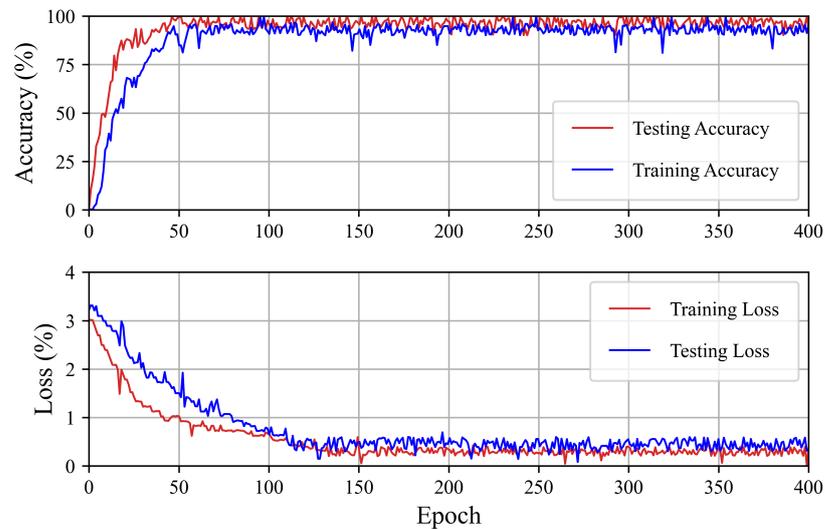
### 5.1. Simulation Results

The simulation experiment used Bellhop [71] to simulate the transmit–receive scenario of the underwater array. Bellhop is a toolbox dedicated to underwater environment simulation, supporting a wide range of parameter settings. By editing the environment file *\*.env*, we can implement the shock response simulation for the UWA channel. The transmitter is a single acoustic source, and its position will change twice, corresponding to angles of  $\theta_1 = 30^\circ$  and  $\theta_2 = 60^\circ$ . Distance  $R = 100$  m between the transmitting end and the center O. The receiver is a four-element uniform linear array (ULA). The sound field is limited to  $100 \text{ m} \times 300 \text{ m} \times 150 \text{ m}$  in the water. To simulate the tugboat’s situation close to the communication array, we set the distance between the tugboat and the array to 5 m when the VURN is the main source of the noise. The transmit and receive settings for the simulation environment are shown in Figure 16a. The sound velocity profile is shown in Figure 16b.



**Figure 16.** The setup of the simulation test: (a) the transmit and receive settings for the simulation environment; (b) the sound velocity profile.

Accuracy and loss function are important metrics that describe the DL model during training and testing. Higher accuracy represents better performance, and faster convergence of the loss function represents better training and testing results. Figure 17 shows the accuracy and loss function statistics on the training and test datasets in the simulation experiments. Figure 17 shows that the noise-reduced learning model maintains a better processing advantage on both the training and test datasets.



**Figure 17.** Comparison of accuracy and loss function statistics for training and test datasets.

In the simulation experiments, we compared the performance of the proposed method with more typical algorithms in UWA signal processing. These algorithms are the Wiener filtering method, the KBM-ANC method, and the CNN method. The Wiener filtering algorithm uses finite-length Wiener filtering with time-domain processing. The KBM-ANC method uses the nonlinear-KLMS algorithm, which uses a Gaussian kernel function to compute the mean-squared error and can be considered a dictionary-based nonlinear sparse approximation problem. The definition of nonlinear-KLMS is as follows:

$$\mu = \max_{i \neq j} \left| \kappa(\mathbf{u}_{\omega_i}, \mathbf{u}_{\omega_j}) \right| \tag{28}$$

where  $\kappa(\cdot)$  is a normalized kernel function and  $\mathbf{u}$  is a dictionary element. The coherence coefficient criterion only allows the input  $\kappa(\mathbf{u}_{\omega_i}, \mathbf{u}_{\omega_j})$  to be added to the dictionary when it is below the threshold  $\mu_0$ . The value of  $\mu_0$  is taken in the interval of  $[0,1]$ , which controls the sparsity of the dictionary and was taken as 0.2 in the simulation experiment. The dictionary can be constructed by Equation (29), and  $M$  denotes the length of the dictionary, which is taken to be equal to the length of the data frame in the experiment.

$$\max_{m=1, \dots, M} |\kappa(\mathbf{u}_m, \mathbf{u}_{\omega_m})| \leq \mu_0 \tag{29}$$

The CNN-based noise-reduction method refers to the structure in the literature [x]. In this case, the network is divided into three parts: the first part is Conv+ReLU (1 layer); the second part is Conv+BN+ReLU (3 layers); the third part is Conv (1 layer). The learning objective is the residuals between signals, and the loss function uses the mean-squared error (MSE) criterion. The input of the CNN noise-reduction method and the proposed method are the same data, and the difference is that the CNN method does not have the same pre-processing process as the proposed method.

We compared the recognition rate and root-mean-squared error (RMSE) and the proposed algorithm for the received signals. The above algorithm comparisons were made in the case of  $\theta_1 = 30^\circ$ , and the compared methods include Wiener filtering, the KBM-ANC, the CNN, and the proposed method, where “Ours1” denotes the noise-reduction processing model of the first stage in the proposed method. The above experiments were repeated 200 times, and the SNR varied from  $-20$  to  $10$  dB. The statistical signal recognition rate is the probability that the noise-reduced signal  $\hat{\mathbf{y}}(t)$  can be judged as the transmit signal  $\mathbf{s}(t)$ . The signal recognition rate reflects the noise-reduction effect of each algorithm, and the RMSE reflects the stability of each of the above algorithms. Figure 18a shows the signal recognition rate comparison of each method. Figure 18b shows the RMSE comparison of each method.

As can be seen in Figure 18a, the proposed algorithm shows a more excellent signal recognition rate when dealing with noise-reduction processing, which indicates that the proposed algorithm has the best noise-reduction effect. It is rather obvious that the other three methods have extremely poor noise-reduction at lower SNRs because none of the above three can handle complex-type data. However, it is worth noting that the CNN method significantly improves noise-reduction compared to KBM-ANC. The Wiener algorithm has the worst noise-reduction effect, and this is because the simulation environment we designed is a nonlinear system, which causes it to be not applicable. Figure 18b shows that the proposed algorithm also maintains the same stability advantage in the RMSE.

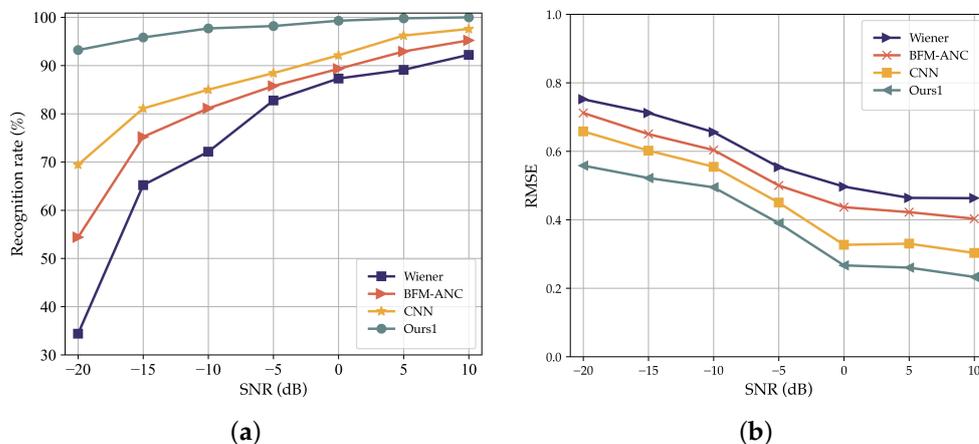
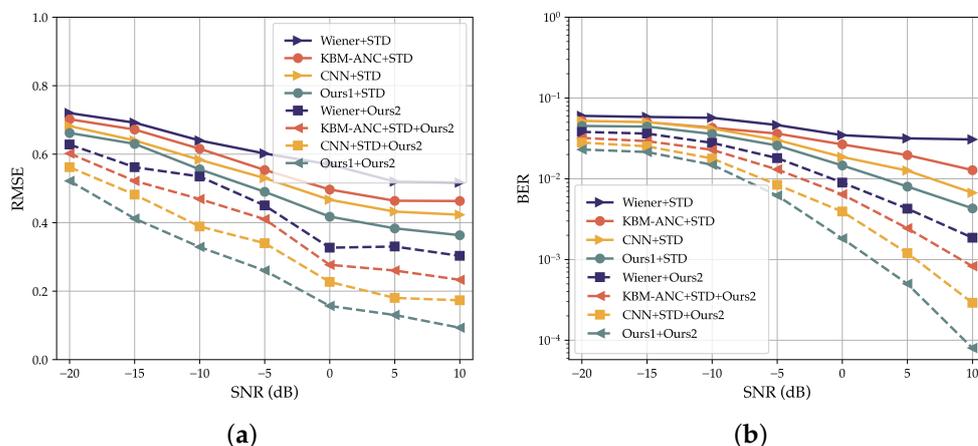


Figure 18. Comparison of signal recognition rate and RMSE of each algorithm: (a) the signal recognition rate comparison of each method; (b) the RMSE comparison of each method.

Theoretically, after the noise-reduction in the first stage, the BER of STD diversity processing in the second stage will decrease. Because the STD’s beamforming performance will be improved when the effect of noise is reduced, the system’s estimation of the main path angle will also be more accurate, resulting in a lower BER of the communication. The ability to cope with the effect of MAI is also reflected in the performance of the BER. The lower the BER, the better the system suppresses MAI. To verify the suppression effect of each type of method on MAI, in Figure 19a, we compare the RMSR of the above noise-reduction methods jointly processed by STD and the RMSE of the above noise-reduction methods jointly processed by STD and weight-based update. In Figure 19b, we also performed the same BER comparison, which is used to illustrate the effect on the communication quality improvement. “Wiener+STD” means that the Wiener method is used for noise-reduction in the first stage, and the conventional STD method is used for diversity combining in the second stage. Joint methods such as “KBM-ANC+STD”, “CNN+STD”, and “Ours1+STD” also exist. “Ours1+STD” indicates that the proposed noise-reduction learning model is used in the first stage, and the traditional STD method is used in the second stage. Similarly, these noise-reduction methods can be jointly processed with the proposed weight-based update STD. Thus, methods such as “Wiener+Ours2”, “KBM-ANC+Ours2”, “CNN+Ours2”, and “Ours1+Ours2” also exist. “Ours1+Ours2” denotes the two-stage processing method proposed in this paper. Each of the above experiments was repeated 200 times at  $\theta_1 = 30^\circ$  and an SNR of  $[-20, 10]$  dB.

Figure 19a shows that the stability of resistance to MAI increases after replacing the second stage with the proposed STD method, and the best results are obtained with the two-stage method proposed in this paper. The same advantage is also reflected in the BER comparison in Figure 19b. The high or low BER value directly reflects MAI’s good or bad suppression. The lower BER value indicates better communication quality and a lower impact on MAI. The above results show that the two-stage method proposed in this paper has better resistance to MAI.

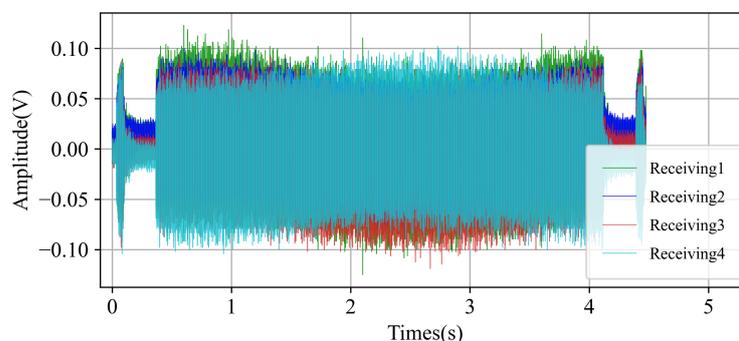


**Figure 19.** Comparison of RMSE and BER performance between various joint algorithms: (a) the RMSE performance comparison between various joint algorithms; (b) the BER performance comparison between various joint algorithms.

### 5.2. Lake Experimental Results

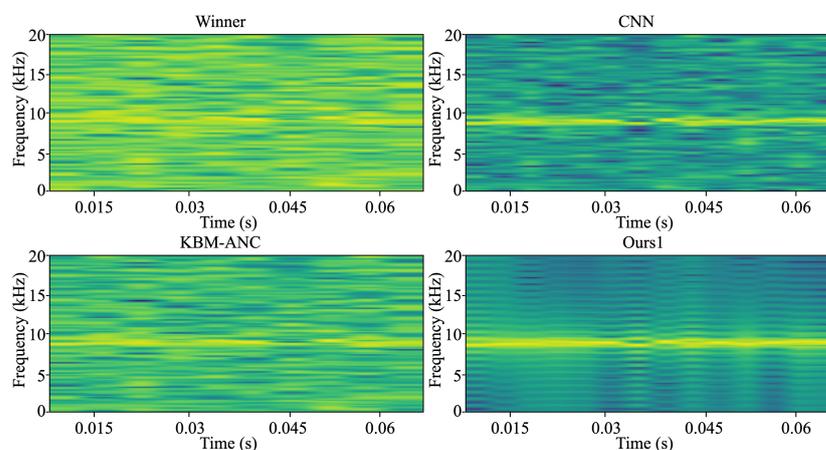
To further verify the proposed algorithms, a real lake experiment was performed at Huating Lake, Anhui Province, China, in July 2021. The transmit signal was consistent with the simulation test. The receiving array was 5 m away from the tugboat, and the transmit angle was 30 and 60. The receiving array was a four-element ULA with a 2 m array spacing, and the transmitting end was 100 m from the receiving array. Before the on-lake communication test, we performed channel measurements with LFM signals and measured the on-lake SNR to about 5 dB. In the experiment on the lake, the tugboat was at

a standstill, but the engine was always in working condition; the effect of noise was always present. The received signal, as shown in Figure 20, was obtained from four array elements.



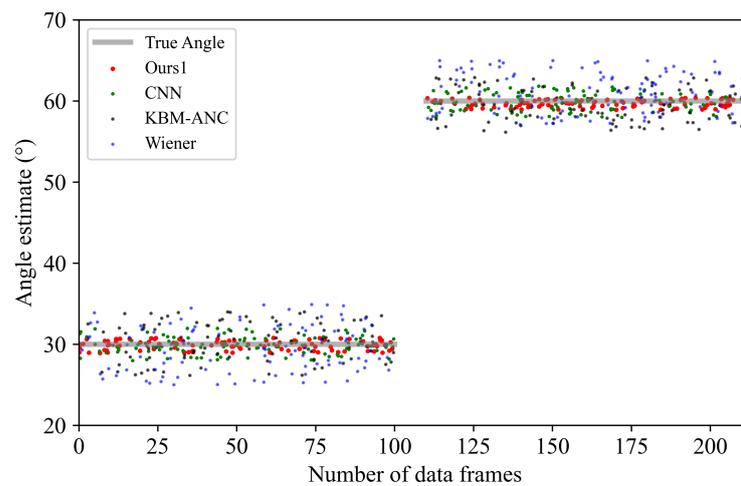
**Figure 20.** Time-domain visualization of the received signal in the on-lake experiment.

To verify the noise-reduction capability of the proposed method for the experimental data on the lake, we performed a time–frequency-based visualization comparison of the noise-reduction results. The UWA signal has both the characteristics of an acoustic signal and the two-dimensional properties that can be represented as an image. A two-dimensional visual comparison of the noise-reduction signal can reflect the effect of noise-reduction. Figure 21 shows the two-dimensional visualization comparison of the four noise-reduction methods after processing. The better the noise-reduction, the clearer the two-dimensional visualization. The visualization results in Figure 21 show that the proposed noise-reduction learning model maintains the same advantage in the on-lake experiments.



**Figure 21.** The two-dimensional visualization comparison of the four noise-reduction methods after processing.

Noise-reduction of the received signal will give a more accurate angle of beamforming estimation in the STD. To verify the improvement of STD performance by noise-reduction methods, we compared the statistics of STD-estimated angles after processing by various noise-reduction methods. We performed 100 operations under each angle for the statistics. Figure 22 shows the statistics of angle estimation. The results show that Wiener filtering and the KBM-ANC can hardly improve the accuracy of angle estimation; although the convergence of the angle estimates of the CNN method is better, the accuracy is still low; the proposed method has the best convergence for angle estimation and has a more significant accuracy of angle estimation. The above comparison results show that the noise-reduction process has performance improvement for STD, which is reflected in the accurate estimation of the main diameter angle.



**Figure 22.** Statistics of angular estimates of beam formation after using the four noise-reduction methods.

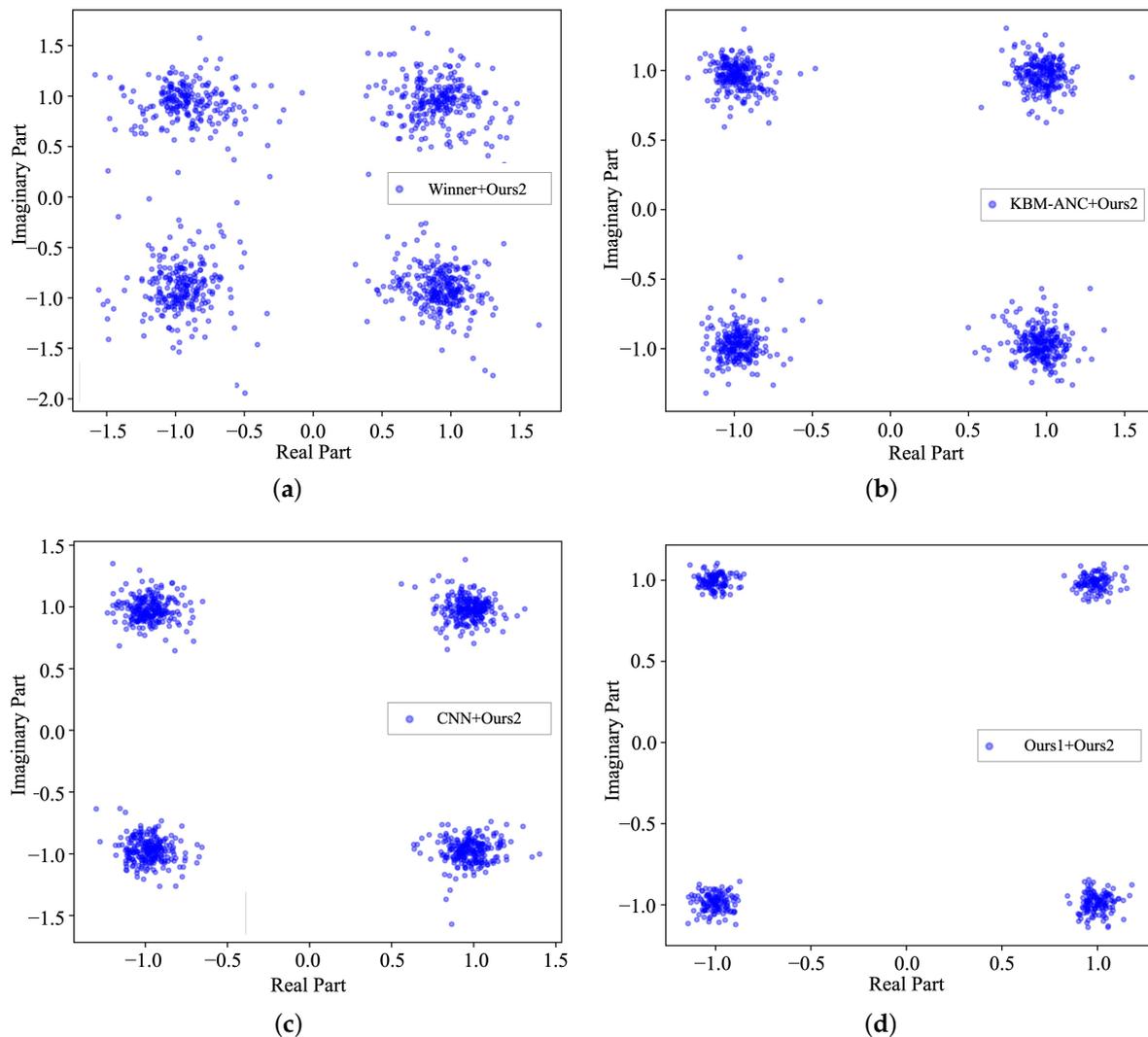
Both the noise-reduction effect and the performance improvement of noise-reduction on STD showed the effectiveness of the proposed method. To further verify the inhibitory ability of the proposed method on MAI, a two-stage comparison of the joint method was carried out in the on-lake experiment. The statistical results are shown in Table 3 and include the RMSE and BER for the two emission angles  $\theta_1 = 30^\circ$  and  $\theta_2 = 60^\circ$ . The proposed method maintains a better suppression of MAI from the statistical results in both perspectives. In the on-lake experiments, the proposed method effectively resists MAI due to the adoption of CDMA signals and improves the communication quality.

**Table 3.** RMSE and BER statistics for each joint method in the on-lake experiment.

Angle	Method	RMSE	BER
$\theta_1 = 30^\circ$	Wiener + STD	0.5520	$0.81 \times 10^{-2}$
	KBM-ANC + STD	0.4638	$0.60 \times 10^{-2}$
	CNN + STD	0.4527	$0.52 \times 10^{-2}$
	Ours1 + STD	0.4104	$0.32 \times 10^{-2}$
	Wiener + Ours2	0.3698	$0.77 \times 10^{-3}$
	KBM-ANC + Ours2	0.2239	$0.55 \times 10^{-3}$
	CNN + Ours2	0.2013	$0.19 \times 10^{-3}$
	Ours1 + Ours2	0.1804	$0.45 \times 10^{-4}$
$\theta_2 = 60^\circ$	Wiener + STD	0.5201	$0.79 \times 10^{-2}$
	KBM-ANC + STD	0.4723	$0.62 \times 10^{-2}$
	CNN + STD	0.4387	$0.49 \times 10^{-2}$
	Ours1 + STD	0.3912	$0.38 \times 10^{-2}$
	Wiener + Ours2	0.3741	$0.74 \times 10^{-3}$
	KBM-ANC + Ours2	0.2187	$0.52 \times 10^{-3}$
	CNN + Ours2	0.1922	$0.21 \times 10^{-3}$
	Ours1 + Ours2	0.1732	$0.49 \times 10^{-4}$

To more intuitively reflect the suppression effect of the proposed method on MAI, we also compared the constellation plots. A constellation plot is a visualization used to describe improvements to communication quality. The more convergent the constellation plot is, the better the communication is indicated in this paper, i.e., the better the suppression of MAI. Since the MAI we used to improve the system was mainly reflected in the second stage, we focused on comparing the noise-reduction method with the joint treatment of the second stage. In the constellation plots, the Wiener filter has the highest communication BER, which is reflected in the worst convergence in Figure 23a. Although it seems that the convergence effect of Figure 23b,c is not very obvious, the CNN noise-reduction is improved

in terms of BER. Figure 23d shows that the proposed method has a better suppression of MAI, which is reflected by the better convergence of the constellation diagram and the lower BER. Since we used an STD structure based on a weight update strategy in the second stage, the structure uses weights to constrain each system receiver. This improved STD structure also simplifies the computational complexity because it does not need to be calculated for each receiver like the equalizer structure. Using the trained network model also improves the efficiency of computing. Using the proposed two-stage model structure can enhance the system's performance requirements for real-time.



**Figure 23.** The comparison of constellation diagrams of different methods: (a) the constellation diagrams for “Wiener+Ours2”; (b) the constellation diagrams for “KBM-ANC+Ours2”; (c) the constellation diagrams for “CNN+Ours2”; (d) the constellation diagrams for the proposed joint processing method.

## 6. Conclusions

This paper proposed an improved STD structure with a joint noise-reduction learning model to process UWA data with complex-type characteristics for noise-reduction, suppress MAI, and improve communication quality. Specifically, a staged model was constructed to minimize the effect of nonlinear noise during STD processing. The model pre-processes the UWA data by complex-type orthogonal compression in the first stage. Then, the CGAN is adapted to become a complex-CGAN dedicated to UWA data processing so that the

learning model can perform noise-reduction for complex-type data. In the second phase, an improved STD structure is constructed that constrains the main diameter and other interference paths through weight updates. It can effectively suppress MAI that came with the use of CDMA as a transmitted signal.

The proposed method effectively solves the UWA communication problem in a non-linear noisy environment and effectively improves handling complex-type data while suppressing MAI. However, the UWA channel is a changing structure, and the proposed method does not adapt to the more drastic channel changes. It is worth investigating how to maintain the advantages of STD processing with the complexity of channel changes in future work. For example, it can be explored to add the extraction of channel features in the pre-processing stage and dig into the channel's variation characteristics as much as possible during the network model construction. For the STD structure, we can also introduce the space–time coding technique to train the receiving process by rule coding to improve the communication performance of the sender and receiver.

**Author Contributions:** Conceptualization, Y.Z. and P.J.; data curation, Y.Z.; formal analysis, Y.Z.; funding acquisition, P.J.; investigation, Y.Z.; methodology, Y.Z. and Z.H.; project administration, P.J.; resources, P.J.; software, D.Z.; validation, Y.Z. and D.Z.; visualization, Y.Z.; writing—original draft, Y.Z.; writing—review and editing, D.Z. and P.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (2020YFB0505803) and the National Key Research and Development Program of China (2016YFB0501800).

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Climent, S.; Sanchez, A.; Capella, J.; Meratnia, N.; Serrano Martín, J. Underwater Acoustic Wireless Sensor Networks: Advances and Future Trends in Physical, MAC and Routing Layers. *Sensors* **2014**, *14*, 795–833. [[CrossRef](#)] [[PubMed](#)]
2. Stojanovic, M.; Preisig, J. Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Commun. Mag.* **2009**, *47*, 84–89. [[CrossRef](#)]
3. Andrew, C.; Jill, K.; Suleyman, S. Signal processing for underwater acoustic communications. *IEEE Commun. Mag.* **2009**, *47*, 90–96.
4. Wan, L.; Zhou, H.; Xu, X.; Huang, Y.; Zhou, S.; Shi, Z.; Cui, J. Adaptive Modulation and Coding for Underwater Acoustic OFDM. *IEEE J. Ocean. Eng.* **2015**, *40*, 327–336. [[CrossRef](#)]
5. Song, A.; Mohsen, B.; Song, H.C.; Hodgkiss, W.S.; Porter, M.B. Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx). *J. Acoust. Soc. Am.* **2008**, *123*, 856–865. [[CrossRef](#)]
6. Wong, K.T.; Zoltowski, M.D. Self-initiating MUSIC-based direction finding in underwater acoustic particle velocity-field beamspace. *IEEE J. Ocean. Eng.* **2000**, *25*, 262–273. [[CrossRef](#)]
7. Yang, T.C. Deconvolved Conventional Beamforming for a Horizontal Line Array. *IEEE J. Ocean. Eng.* **2018**, *43*, 160–172. [[CrossRef](#)]
8. Palmese, M.; Trucco, A. An Efficient Digital CZT Beamforming Design for Near-Field 3-D Sonar Imaging. *IEEE J. Ocean. Eng.* **2010**, *35*, 584–594. [[CrossRef](#)]
9. Edwards, J.R.; Schmidt, H.; LePage, K.D. Bistatic synthetic aperture target detection and imaging with an AUV. *IEEE J. Ocean. Eng.* **2001**, *26*, 690–699. [[CrossRef](#)]
10. Yahong, R.Z.; Jing, X.W.; Cheng, S.X. Turbo equalization for single-carrier underwater acoustic communications. *IEEE Commun. Mag.* **2015**, *53*, 79–87.
11. Preisig, J.C.; Deane, G.B. Surface wave focusing and acoustic communications in the surf zone. *J. Acoust. Soc. Am.* **2004**, *116*, 2067–2080. [[CrossRef](#)]
12. Guo, S.Z.; He, F.D. Spatial diversity in multichannel processing for underwater acoustic communications. *Ocean Eng.* **2011**, *38*, 1611–1623.
13. Dahl, H.; Dall, O.; David, R. Range-Dependent Inversion for Seabed Parameters Using Vector Acoustic Measurements of Underwater Ship Noise. *IEEE J. Ocean. Eng.* **2021**, *1*, 1–10. [[CrossRef](#)]
14. He, C.; Huo, S.Y.; Zhang, Q.F.; Wang, H.; Huang, J.G. Multi-channel iterative FDE for single carrier block transmission over underwater acoustic channels. *China Commun.* **2015**, *12*, 55–61. [[CrossRef](#)]
15. Cao, Y.; Shi, W.; Sun, L.J.; Fu, X.M. Frequency Diversity Based Underwater Acoustic Passive Localization. *IEEE Internet Things J.* **2021**, *1*, 1. [[CrossRef](#)]
16. Ian, F.A.; Dario, P.; Tommaso, M. Underwater acoustic sensor networks: Research challenges. *Ad Hoc Netw.* **2005**, *3*, 257–279.

17. Peng, L.S. Image denoising algorithm via doubly local Wiener filtering with directional windows in wavelet domain. *IEEE Signal Process. Lett.* **2005**, *12*, 681–684. [[CrossRef](#)]
18. Astola, J.; Haavisto, P.; Neuvo, Y. Vector median filters. *Proc. IEEE Inst. Electr. Electron. Eng.* **1990**, *78*, 678–689. [[CrossRef](#)]
19. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [[CrossRef](#)]
20. Rout, D.K.; Subudhi, B.N.; Veerakumar, T.; Chaudhury, S. Walsh–Hadamard-Kernel-Based Features in Particle Filter Framework for Underwater Object Tracking. *IEEE Trans. Industr. Inform.* **2020**, *16*, 5712–5722. [[CrossRef](#)]
21. Comaniciu, D.; Ramesh, V.; Meer, P. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 564–577. [[CrossRef](#)]
22. Ferwerda, J.; Hainmueller, J.; Hazlett, C.J. Kernel-based regularized least squares in R (KRLS) and Stata (krls). *J. Stat. Softw.* **2017**, *79*, 1–26. [[CrossRef](#)]
23. Kailath, T. RKHS approach to detection and estimation problems—I: Deterministic signals in Gaussian noise. *IEEE Trans. Inf. Theory* **1971**, *17*, 530–549. [[CrossRef](#)]
24. Liu, W.; Principe, J.C. Kernel affine projection algorithms. *Eurasip J. Adv. Signal Process.* **2008**, *2008*, 1–12. [[CrossRef](#)]
25. Jackson, D.R.; Dowling, D.R. Phase conjugation in underwater acoustics. *J. Acoust. Soc. Am.* **1991**, *89*, 171–181. [[CrossRef](#)]
26. Muller, K.R.; Mika, S.; Ratsch, G.; Tsuda, K.; Scholkopf, B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw. Learn Syst.* **2001**, *12*, 181–201. [[CrossRef](#)]
27. Wang, D.; Zou, Y.X.; Wang, W.W. Learning soft mask with DNN and DNN-SVM for multi-speaker DOA estimation using an acoustic vector sensor. *J. Franklin Inst.* **2018**, *355*, 1692–1709. [[CrossRef](#)]
28. Ridao, P.; Tiano, A.; El-Fakdi, A.; Carreras, M.; Zirilli, A. On the identification of nonlinear models of unmanned underwater vehicles. *Control Eng. Pract.* **2004**, *12*, 1483–1499. [[CrossRef](#)]
29. Li, Y.; Lu, H.; Li, J.; Li, X.; Li, Y.; Serikawa, S. Underwater image de-scattering and classification by deep neural network. *Comput. Electr. Eng.* **2016**, *54*, 68–77. [[CrossRef](#)]
30. Tazebay, M.V.; Akansu, A.N. Adaptive subband transforms in time–frequency excisers for DSSS communications systems. *IEEE Trans. Signal Process.* **1995**, *43*, 2776–2782. [[CrossRef](#)]
31. Hara, S.; Prasad, R. Overview of multicarrier CDMA. *IEEE Commun. Mag.* **1997**, *35*, 126–133. [[CrossRef](#)]
32. Zhang, J.; Chong, E.K.; Tse, D.N.C. Output MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems. *IEEE Trans. Inf. Theory* **2001**, *47*, 1128–1144. [[CrossRef](#)]
33. Nagatsuka, M.; Kohno, R. A spatially and temporally optimal multi-user receiver using an array antenna for DS/CDMA. *IEICE Trans. Commun.* **1995**, *78*, 1489–1497.
34. Sun, J.; Li, X.; Chen, K.; Cui, W.; Chu, M. A novel CMA+DDLMS blind equalization algorithm for underwater acoustic communication. *Comput. J.* **2020**, *63*, 974–981. [[CrossRef](#)]
35. Xiao, Y.; Yin, F.L. Blind equalization based on RLS algorithm using adaptive forgetting factor for underwater acoustic channel. *China Ocean Eng.* **2014**, *28*, 401–408. [[CrossRef](#)]
36. Shi, Y.; Konar, A.; Sidiropoulos, N.D.; Mao, X.P.; Liu, Y.T. Learning to beamform for minimum outage. *IEEE Trans. Signal Process.* **2018**, *66*, 5180–5193. [[CrossRef](#)]
37. Chen, J.C.; Yao, K.; Hudson, R.E. Source localization and beamforming. *IEEE Signal Process. Mag.* **2002**, *66*, 30–39. [[CrossRef](#)]
38. Vukadin, P.; Hudec, G. Acoustic telemetry system for underwater control. *IEEE J. Ocean. Eng.* **1991**, *16*, 142–145. [[CrossRef](#)]
39. Wen, Q.; Ritcey, J.A. Spatial diversity equalization applied to underwater communications. *IEEE J. Ocean. Eng.* **1994**, *19*, 227–241.
40. Li, S.; Smith, P.J.; Dmochowski, P.A.; Yin, J. Analysis of analog and digital MRC for distributed and centralized MU-MIMO systems. *IEEE Trans. Veh. Technol.* **2018**, *68*, 1948–1952. [[CrossRef](#)]
41. Qu, F.; Wang, Z.; Yang, L. Differential orthogonal space–time block coding modulation for time-variant underwater acoustic channels. *IEEE J. Ocean. Eng.* **2016**, *42*, 188–198. [[CrossRef](#)]
42. Mitran, P.; Ochiai, H.; Tarokh, V. Space-time diversity enhancements using collaborative communications. *IEEE Trans. Inf. Theory* **2005**, *51*, 2041–2057. [[CrossRef](#)]
43. Stojanovic, M.; Freitag, L. Multichannel detection for wideband underwater acoustic CDMA communications. *IEEE J. Ocean. Eng.* **2006**, *31*, 685–695. [[CrossRef](#)]
44. Chen, Y.F.; Zoltowski, M.D.; Ramos, J.; Chatterjee, C.; Roychowdhury, V.P. Reduced-dimension blind space–time 2-D RAKE receivers for DS-CDMA communication systems. *IEEE Trans. Signal Process.* **2000**, *48*, 1521–1536. [[CrossRef](#)]
45. Elad, M.; Feuer, A. Superresolution restoration of an image sequence: Adaptive filtering approach. *IEEE Trans. Image Process.* **1999**, *8*, 387–395. [[CrossRef](#)]
46. Eleftheriou, E.; Falconer, D. Tracking properties and steady-state performance of RLS adaptive filter algorithms. *IEEE Trans. Acoust. Speech Signal Process.* **1986**, *34*, 1097–1110. [[CrossRef](#)]
47. Malinowski, S.J.; Gloza, I. Underwater noise characteristics of small ships. *Acta Acust. United Acust.* **2002**, *88*, 718–721.
48. Goel, A.; Vetteth, A.; Rao, K.R.; Sridhar, V. Active cancellation of acoustic noise using a self-tuned filter. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2004**, *51*, 2148–2156. [[CrossRef](#)]
49. Sriperumbudur, B.K.; Fukumizu, K.; Lanckriet, G.R. Universality, Characteristic Kernels and RKHS Embedding of Measures. *J. Mach. Learn. Res.* **2011**, *12*, 2389–2410.

50. Wu, Q.; Li, Y.; Zakharov, Y.V.; Xue, W.; Shi, W. A kernel affine projection-like algorithm in reproducing kernel hilbert space. *IEEE Trans. Circuits Syst. II Express Briefs* **2019**, *67*, 2249–2253. [[CrossRef](#)]
51. Mitra, V.; Wang, C.J.; Banerjee, S. Lidar detection of underwater objects using a neuro-SVM-based architecture. *IEEE Trans. Neural Netw.* **2006**, *17*, 717–731. [[CrossRef](#)] [[PubMed](#)]
52. Boloix, R.; Murillo, J.J.; Tsiftaris, S.A. The generalized complex kernel least-mean-square algorithm. *IEEE Trans. Signal Process.* **2019**, *67*, 5213–5222. [[CrossRef](#)]
53. Dhanalakshmi, P.; Palanivel, S.; Ramalingam, V. Classification of audio signals using SVM and RBFNN. *Expert Syst. Appl.* **2009**, *36*, 6069–6075. [[CrossRef](#)]
54. Verma, N.K.; Singh, S. Image sequence prediction using ANN and RBFNN. *Int. J. Image Graph* **2013**, *13*, 1340006. [[CrossRef](#)]
55. Stoffel, M.; Gulakala, R.; Bamer, F.; Markert, B. Artificial neural networks in structural dynamics: A new modular radial basis function approach vs. convolutional and feedforward topologies. *Comput. Methods Appl. Mech. Eng.* **2020**, *364*, 112989. [[CrossRef](#)]
56. Petroni, S.; Tripoli, G.; Combi, C.; Vigna, B.; De Vittorio, M.; Todaro, M.T.; Passaseo, A. Noise reduction in GaN-based radio frequency surface acoustic wave filters. *Appl. Phys. Lett.* **2004**, *85*, 1039–1041. [[CrossRef](#)]
57. Xiao, C.; Wang, X.; Dou, H.; Li, H.; Lv, R.; Wu, Y.; Song, G.; Wang, W.; Zhai, R. Non-Uniform Synthetic Aperture Radiometer Image Reconstruction Based on Deep Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 2359. [[CrossRef](#)]
58. Chandrakala, S.; Rajeswari, N. Representation Learning Based Speech Assistive System for Persons With Dysarthria. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1510–1517. [[CrossRef](#)]
59. Zhou, W.; Zhu, Z. A novel BNMF-DNN based speech reconstruction method for speech quality evaluation under complex environments. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 959–972. [[CrossRef](#)]
60. He, C.B.; Huang, J.G.; Ding, Z. A Variable-Rate Spread-Spectrum System for Underwater Acoustic Communications. *IEEE J. Ocean. Eng.* **2009**, *34*, 624–633.
61. Rahmati, M.; Pompili, D. UNISec: Inspection, Separation, and Classification of Underwater Acoustic Noise Point Sources. *IEEE J. Ocean. Eng.* **2018**, *43*, 777–791. [[CrossRef](#)]
62. Wang, N.Y.; Wang, H.S.; Wang, T.W.; Fu, S.W.; Lu, X.; Wang, H.M.; Tsao, Yu. Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *29*, 184–195. [[CrossRef](#)] [[PubMed](#)]
63. George, E.B.; Smith, M.J. Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. *IEEE Trans. Speech Audio Process.* **1997**, *5*, 389–406. [[CrossRef](#)]
64. Li, Y.; Fu, R.; Meng, X.C.; Jin, W.; Shao, F. A SAR-to-Optical Image Translation Method Based on Conditional Generation Adversarial Network (cGAN). *IEEE Access* **2020**, *8*, 60338–60343. [[CrossRef](#)]
65. Guo, X.; Nie, R.; Cao, J.; Zhou, D.; Mei, L.; He, K. FuseGAN: Learning to fuse multi-focus image via conditional generative adversarial network. *IEEE Trans. Multimed.* **2019**, *21*, 1982–1996. [[CrossRef](#)]
66. Shao, Z.; Lu, Z.; Ran, M.; Fang, L.; Zhou, J.; Zhang, Y. Residual encoder–decoder conditional generative adversarial network for pansharpening. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *17*, 1573–1577. [[CrossRef](#)]
67. Liu, X.; Gao, Z.; Chen, B.M. MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *17*, 1488–1492. [[CrossRef](#)]
68. Kim, M. ML/CGAN: Network Attack Analysis Using CGAN as Meta-Learning. *IEEE Commun. Lett.* **2021**, *25*, 499–502. [[CrossRef](#)]
69. Deng, J.; Pang, G.Y.; Zhang, Z.Y.; Pang, Z.B.; Yang, H.Y.; Yang, G. cGAN Based Facial Expression Recognition for Human-Robot Interaction. *IEEE Access* **2019**, *7*, 9848–9859. [[CrossRef](#)]
70. Rossi, R.; Murari, A.; Gaudio, P. On the Potential of Time Delay Neural Networks to Detect Indirect Coupling between Time Series. *Remote Sens.* **2020**, *22*, 584. [[CrossRef](#)]
71. Miron, M.; Barclay, D.R.; Bousquet, J.F. The Oceanographic Sensitivity of the Acoustic Channel in Shallow Water. *IEEE J. Ocean. Eng.* **2021**, *46*, 675–686. [[CrossRef](#)]