



Article Sea Clutter Amplitude Prediction via an Attention-Enhanced Seq2Seq Network

Qizhe Qu¹, Hao Chen², Zhenshuo Lei³, Binbin Li², Qinglei Du² and Yongliang Wang^{1,2,*}

- ¹ Electronic Information School, Wuhan University, Wuhan 430072, China; quqizhe@whu.edu.cn
- ² Air Force Early Warning Academy, Wuhan 430019, China
- ³ School of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China
- * Correspondence: ylwangkjld@163.com

Abstract: Sea clutter is a kind of ubiquitous interference in sea-detecting radars, which will definitely influence target detection. An accurate sea clutter prediction method is supposed to be beneficial while existing prediction methods are based on the one-step-ahead prediction. In this paper, a sea clutter prediction network (SCPNet) is proposed to achieve the *k*-step-ahead prediction based on the characteristics of sea clutter. The SCPNet takes a sequence-to-sequence (Seq2Seq) structure as the backbone, and a simple self-attention module is employed to enhance the ability of adaptive feature selections. The SCPNet takes the normalized amplitudes of sea clutter as inputs and is capable of predicting an output sequence with a length of *k*; the phase space reconstruction theory is also used to find the optimized input length of the sea clutter sequence. Results with the seadetecting radar data-sharing program (SDRDSP) database show the mean square error of the proposed method is 1.48×10^{-5} and 8.76×10^{-3} in the one-step-ahead prediction and the eight-step-ahead prediction, respectively. Compared with four existing methods, the proposed method achieves the best prediction performance.

Keywords: sea clutter prediction; seq2seq structure; gated recurrent unit (GRU); *k*-step-ahead predictions; neural-network applications

1. Introduction

Real-time and accurate marine surveillance plays a significant role in both military and civil fields, such as early warning and marine rescue [1]. Radar, which served as a full-weather and full-time sensor, is a ubiquitous solution to provide reliable range and Doppler resolutions. However, back-scattering echoes received by radar systems contain not only targets under test but also echoes from the sea surface. These echoes from the sea surface, i.e., sea clutter, are supposed to interfere with the observation of targets of interest. Once the sea wave becomes sufficiently wild, it will introduce the probability of a false alarm, which is almost inevitable in many kinds of radar systems [2,3]. In addition, owing to its physical property, the Doppler spectrum of sea clutter is prone to being wide. The wide Doppler spectrum then tends to cover targets under test [4], especially those dim targets with a relatively low velocity. Thus, it is of significance to suppress sea clutter while preserving the information of targets of interest.

There are lots of classical clutter-suppression algorithms based on transform domain cancellation, such as the moving target indicator (MTI) method and root loop cancellation. Nevertheless, targets with a low velocity are likely to be suppressed along with sea clutter by the MTI method. In addition, the spectrum center of sea clutter will not always be consistent with the zero Doppler frequency due to the relative motion of the sea surface, which is prone to performance degradation of these cancellation methods [5]. In addition, subspace projection methods also provide solutions to suppress clutter, such as the singular-value decomposition method [6]. Based on the multiresolution analysis and reconstruction



Citation: Qu, Q.; Chen, H.; Lei, Z.; Li, B.; Du, Q.; Wang, Y. Sea Clutter Amplitude Prediction via an Attention-Enhanced Seq2Seq Network. *Remote Sens.* **2023**, *15*, 3234. https://doi.org/10.3390/rs15133234

Academic Editor: Yukiharu Hisaki

Received: 31 March 2023 Revised: 8 June 2023 Accepted: 20 June 2023 Published: 22 June 2023



Copyright: © 2023 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/). theory, a series of clutter-suppression approaches are proposed via many kinds of wavelet transforms [7]. Empirical mode decomposition is also introduced into the sea clutter-suppression method [8]. Although these suppression methods have improved the clutter-suppression ability, sea clutter has obvious nonlinear, non-Gaussian, and nonstationary characteristics [9], which may cause performance degradation.

Except for the typical suppression methods mentioned above, another viable way to achieve robust clutter suppression is to predict the amplitude of sea clutter by learning its natural inner rules. Then, a simple cancellation in the time domain between received echoes and extrapolated clutter sequences can be employed to highlight targets of interest. A large number of references focused on clutter characteristics are proposed. For the characteristic of amplitude distribution, the Rayleigh distribution is designed for radar systems with low resolutions. The log-normal distribution is proposed for sea clutter with large dynamic ranges. The Weibull distribution and the K distribution are also widely used in clutter models. Later, more and more elaborate models are proposed to describe sea clutter [10].

As for the change rule with the time of sea clutter, it has been viewed as a stochastic process for a long time. In 1990, Haykin and Leung calculated and analyzed the correlation dimension of sea clutter, which first indicates that sea clutter is basically a chaotic process [11]. Furthermore, they have proposed a chaotic signal-processing method and a small-target detection method based on the chaotic characteristic [12,13]. Fortunately, the chaotic process of sea clutter actually provides a theoretical basis for clutter prediction. Leung has proposed a memory-based predictor to achieve nonlinear clutter cancellation [14]. A multimodel prediction approach is designed for sea clutter modeling in [15], and different radial basis neural networks are utilized to predict the peak of sea clutter.

With the development of machine learning, the support vector machine (SVM) is employed to predict real-life sea clutter data [16]. A sea clutter sequence prediction method based on a general-regression neural network (GRNN) and the particle-swarm optimization algorithm is proposed in [17]. Results with one dataset from the IPIX database [18] verify the feasibility of sea clutter prediction by a regression model, and the mean square error (MSE) of the proposed method is about 0.4754. In [19], the Volterra filter is employed to predict sea clutter and then detect low-flying targets and results with one dataset of the IPIX database indicating the MSE is about 3.24×10^{-3} . A weighted grey Markov model is constructed for quantitative prediction of sea clutter power in [20]. Results with measured data show the proposed model has better prediction accuracy.

Recently, modern neural networks (NNs) and deep learning (DL) have gained great popularity and improvement in lots of fields, such as computer vision and natural language processing. Hence, more and more researchers are making efforts to combine radar signal processing algorithms and DL-based methods [21–23]. As for the prediction of sea clutter, since the amplitude of sea clutter can be described as a sequence varying with time, the clutter prediction problem can be viewed as a sequence prediction problem in the natural language processing society. Recurrent neural networks (RNNs), equipped with hidden states, are capable of catching and preserving short-term information in sequences [24]. Nevertheless, plain RNNs are prone to the gradient exploding problem and the vanishing gradient problem.

Long short-term memory networks (LSTMs) are designed to deal with the above problems by introducing the gate mechanism [25]. LSTMs are consequently utilized to solve the clutter prediction problem. To obtain higher prediction accuracy at a farther distance, an LSTM model is proposed, and its hyperparameters are discussed in [26]. Simulations demonstrate that the optimized LSTM model has better prediction accuracy compared with the backpropagation network. Further, a clutter amplitude prediction system based on an LSTM is designed in [27]. Results from three real-life databases show that the mean MSE of the proposed system is lower than that of the artificial neural network (ANN) and radial basis function network.

Except for the LSTM structure, a deep neural network (DNN) is employed to predict sea clutter characteristics via marine environmental factors in [28]. Results verify that the

DNN model can effectively predict important characteristics of sea clutter in the South China Sea. In [29], a gated feedback RNN (GF-RNN) is designed to predict sea clutter amplitude. Two measured databases, namely the IPIX database and a P-band database, are used to evaluate the prediction performance. Results show that the GF-RNN has better prediction ability than the SVM and ANN.

Although existing RNN-based methods have achieved high prediction accuracy or low MSE with several real-life databases, these methods all use one-step-ahead prediction to obtain corresponding results. That is, they usually predict only one amplitude value due to the one-step-ahead prediction, which may be inefficient and complex to get a required long prediction in practice. Thus, a *k*-step-ahead prediction is urgent for sea clutter prediction. However, existing plain RNN-based methods may be powerless to deal with a long output sequence.

Aiming at the above shortcoming of existing methods, a sea clutter prediction network, denoted as SCPNet is proposed in this paper to achieve the *k*-step-ahead prediction. The proposed SCPNet takes a sequence-to-sequence (Seq2Seq) structure as a backbone and is enhanced by a self-attention module. Hence, different from the existing methods based on one-step-ahead prediction, the purpose of this paper is to attempt to explore the prediction performance of the *k*-step-ahead prediction method. The difficulty and shortcomings are then analyzed and summarized through experiments with a real-life database. The main contributions of this paper are as follows:

- 1. Different from mainstream methods based on the one-step-ahead prediction, the proposed SCPNet is capable of predicting a long output sequence in one prediction via the *k*-step-ahead prediction, which is more helpful in practice and can be easily transformed into the one-step-ahead prediction when needed;
- A Seq2Seq model is introduced to deal with the problem of variable-length output sequences. The Seq2Seq model is capable of encoding input sequences and then decoding them as prediction sequences. Furthermore, a self-attention module is integrated to boost the learning and representation ability;
- 3. Although the IPIX database is a widely used benchmark, some physical parameters are relatively out-of-date owing to the limitations of the hardware. Instead, a newly released database [30], namely the sea-detecting radar data-sharing program (SDRDSP) database, is employed to assess the performance of each method.

The rest of this paper is organized as follows: Section 2 provides a problem formulation, the difference between one-step-ahead prediction and the *k*-step-ahead prediction, and an introduction to phase space reconstruction and the SDRDSP database. The details of the proposed method, the structure of the SCPNet, and the experiment setting are presented in Section 3. The results of the phase space reconstruction and comparisons of different prediction modes are given and analyzed in Section 4. The consistency analysis and the computational complexity are discussed in Section 5. Finally, the conclusions of this paper and future works are summarized in Section 6.

2. Materials

In this subsection, the sea clutter prediction problem is formulated to clarify the theoretical basis and contributions. Further, a brief introduction to the phase space reconstruction theory is provided. Some detailed physical parameters and experiment settings of the employed SDRDSP database are finally introduced.

2.1. Sea Clutter Properties and Problem Formulation

There are lots of research focused on the sea clutter model from different points of view, such as the amplitude distribution, the Doppler domain, and spatial and temporal correlations. The amplitude distribution is always used to describe the stochastic fluctuation of sea clutter through the probability density function (PDF). As mentioned in the introduction, some typical distribution models are employed to fit the amplitude distribution, the Rayleigh distribution, the log-normal distribution,

the Weibull distribution, and the K distribution. Furthermore, there are more and more elaborate models to describe actual sea clutter in recent years.

For echoes of sea clutter, the Bragg scatting returns are significant components, and they yield two distinct Doppler frequencies when the carrier frequency is lower than the L-band. When the carrier frequency is higher than the L-band, there is usually only one Bragg peak in the frequency spectrum [31]. These Doppler frequencies f_d can be calculated as [32]

$$f_d = \pm \sqrt{\frac{gf_c}{\pi c}},\tag{1}$$

where f_c is the carrier frequency, g is the gravity acceleration, and c is the speed of light. The analysis of temporal and spatial correlations can indicate a strong correlation in the range domain and the time domain, respectively. Mathematically, for a fully coherent radar overlooking a clutter-only ocean, the received sea clutter sequence with N sample points can be denoted as

$$\mathbf{X} = \{x(1), x(2), x(3), \dots, x(N)\}.$$
(2)

Then, the temporal correlation coefficient $\rho(t)$ and the spatial correlation coefficient $\rho(m)$ are defined as [33]

$$\rho(t) = \frac{R(t) - |E(\mathbf{X})|^2}{R(0) - |E(\mathbf{X})|^2}$$
(3)

$$\rho(m) = \frac{\frac{1}{N} \sum_{n=0}^{N-m} x(n) x^*(n+m) - |\overline{\mathbf{X}}|^2}{|\mathbf{X}|^2 - |\overline{\mathbf{X}}|^2}$$
(4)

where R(t) is the correlation function in the time domain and $E(\cdot)$ is the mathematical expectation. According to these properties, it is viable to learn inner features and predict the probable future fluctuation of sea clutter through NNs. Afterward, the extrapolated clutter sequence can be employed to detect an anomaly, especially the presence of a target by comparing it with the actual received echoes.

Since most existing NNs cannot process complex-value signals, the absolute value of the received clutter sequence then becomes an alternate, denoted as **Y**. Due to the complex nonlinear and chaotic characteristics of sea clutter, the absolute value sequence is usually normalized, which is also beneficial to training RNNs. Herein, a min-to-max normalization is employed as follows:

$$\overline{y}(n) = \frac{y(n) - y_{min}}{y_{max} - y_{min}}, n = 1, 2, 3, \dots, N,$$
(5)

where y_{min} and y_{max} are the minimal value and the maximal value, respectively. Then according to Takens' embedding theorem [34], the chaotic sequence \overline{Y} can be reconstructed as:

$$\mathbf{Z}(i,m) = \{\overline{y}(i), \overline{y}(i+\tau), \dots, \overline{y}(i+(m-1)\tau)\},\tag{6}$$

where $i = 1, 2, 3, ..., N - (m - 1)\tau$. τ and m denote the delay time and embedding dimension, respectively [35]. **Z**(*i*, *m*) is the *i*-th reconstructed sequence. Through the phase space reconstruction theory, a state transition function $f(\cdot)$ can be calculated as

$$\overline{y}(i+m\tau) = f(\mathbf{Z}(i,m)). \tag{7}$$

Apparently, the nonlinear transition function $f(\cdot)$ is the key to predicting the next value. However, how to find the function is a rather challenging task. Thanks to the universal approximation theorem, a full-connection RNN with adequate sigmoid-class hidden neurons is capable of fitting any nonlinear dynamic systems [36]. Thus, $f(\cdot)$ can be obtained through data-driven training. Equation (4) then is rewritten as

$$\overline{y}(i+m\tau) = \text{RNN}(\overline{y}(i), \overline{y}(i+\tau), \dots, \overline{y}(i+(m-1)\tau)),$$
(8)

where RNN(·) denotes a RNN and $\overline{y}(i + (m - 1)\tau + 1)$ is the prediction value of the RNN. As for the training of the RNN, supervised learning is employed to obtain the optimized weights. For a given training sample { \overline{Y}_t , L}, where \overline{Y}_t denote the training sequence and L denote the label sequence, the error of prediction is calculated via a loss function as follows

$$Error = \mathcal{L}(\mathbf{L}, \text{RNN}(\mathbf{Y}_t; \boldsymbol{\theta})), \tag{9}$$

where $\mathcal{L}(\cdot)$ is the loss function and θ is the learnable parameter in the RNN. Herein, the loss function is the MSE function that is widely used in the prediction problem. The MSE function is defined as

MSE =
$$\frac{1}{N_t} \sum_{i=1}^{N_t} (l(i) - \hat{y}(i))^2$$
, (10)

where N_t is the length of the prediction sequence; l(i) and $\hat{y}(i)$ denote the *i*-th true value and *i*-th prediction value, respectively.

Next, our goal is to minimize the MSE between training samples and labels, and backpropagation through a time algorithm is used to train the RNN. However, it is hard to build a robust dependent relationship for a long range owing to the gradient exploding problem. So, a gradient clipping method is introduced during the training. Let g_t denote the current gradient; then the gradient clipping performs the following steps:

$$\begin{cases} g_t = g_t, \|g_t\|_2 \le b \\ g_t = \frac{b}{\|g_t\|_2} g_t, \|g_t\|_2 > b \end{cases}$$
(11)

where $\|\cdot\|_2$ indicates a 2-norm and *b* is a hyperparameter.

2.2. Phase Space Reconstruction

The length of an input clutter sequence is an important hyperparameter that, no doubt, will influence the prediction performance of RNNs. How to select the optimized length then becomes another difficulty. The length is manually selected by a series of experiments in lots of previous works, which, however, is a time-consuming job and cannot guarantee optimization. Thus, the phase space reconstruction theory is introduced in this paper to find the optimized length of the input sequence.

The received clutter sequence is actually a scalar that is more appropriate to analyze dynamic and chaotic characteristics in one-dimensional space. Thus, a phase space reconstruction is used to construct a higher dimensional space in the chaotic system for the clutter sequence, which can preserve the inner geometric properties of the original system. In addition, according to Takens' embedding theorem, when an appropriate embedding dimension *m* is selected, a phase space with the same meaning as the original system can be reconstructed by using the original time series, just as derived in Equation (3).

Under ideal conditions, the delay time τ and the embedding dimension m can be stochastically determined for a noise-free sequence with infinite length. However, for the finite clutter sequence with lots of noise, it is of significance to calculate τ and m precisely since inaccurate τ and m are supposed to interfere with reconstruction performance and follow-up analysis. Once the optimized τ and m are selected, the optimized input length is then selected. That is, the optimized length equals $m\tau$, just as indicated in Equation (4).

There are lots of algorithms to find the optimized τ , such as the autocorrelation algorithm and the mutual information algorithm. For a chaotic and nonlinear system, the mutual information algorithm is a better choice. Let $\mathbf{S} = \{s_1, s_2, s_3, ..., s_n\}$ and $\mathbf{Q} = \{q_1, q_2, q_3, ..., q_n\}$ denote two discrete systems, and $H(\mathbf{S})$ and $H(\mathbf{Q})$ are the information entropy of \mathbf{S} and \mathbf{Q} , respectively. Then the mutual information is defined as

$$I(\mathbf{Q}, \mathbf{S}) = H(\mathbf{Q}) - H(\mathbf{Q}|\mathbf{S}), \tag{12}$$

where $H(\mathbf{Q}|\mathbf{S})$ is the conditional probability density (CPD). For each element in two systems, $I(\mathbf{Q}, \mathbf{S})$ can be rewritten as

$$I(\mathbf{Q}, \mathbf{S}) = \sum_{i} \sum_{j} P_{sq}(s_i, q_j) \log\left[\frac{P_{sq}(s_i, q_j)}{P_s(s_i)P_q(j)}\right],$$
(13)

where $P_{sq}(s_i, q_j)$ is the joint probability density function. If we denote $[s, q] = [x(t), x(t + \tau)]$, that is, **S** represents the sequence x(t) and **Q** represents the sequence with the delay time $x(t + \tau)$. Then, $I(\mathbf{Q}, \mathbf{S})$ is a function with the delay time τ , denoted as $I(\tau)$. When $I(\tau) = 0$, there is no correlation between x(t) and $x(t + \tau)$. The first minimal value of $I(\tau)$ actually indicates the most unrelated relation between x(t) and $x(t + \tau)$, and the first minimal value consequently is the optimized delay time.

Once the delay time τ is selected, the Cao method [37] is employed to find the optimized *m*. In the phase space with *m* dimensions, each point vector $\mathbf{Z}_{m}(i) = \{z(i), z(i + \tau), \dots, z(i + m - 1)\tau)\}$ is supposed to have the nearest point $\mathbf{Z}_{m}^{N}(i)$ at a certain distance. When the dimension is m + 1, the distance will change and there will be the nearest point $\mathbf{Z}_{m+1}^{N}(i)$. Then, a variable is defined as

$$a(i,m) = \frac{\|\mathbf{Z}_{m+1}(i) - \mathbf{Z}_{m+1}^{N}(i)\|}{\|\mathbf{Z}_{m}(i) - \mathbf{Z}_{m}^{N}(i)\|},$$
(14)

where $\|\cdot\|$ denotes the Euclidean distance. Further, E(m) and $E_1(m)$ are defined as

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N - m\tau} a(i, m),$$
(15)

$$E_1(m) = \frac{E(m+1)}{E(m)}.$$
 (16)

If $E_1(m)$ does not change after *m* is greater than a certain value, the optimized embedding dimension *m* has been determined.

2.3. The One-Step-Ahead Prediction and the k-Step-Ahead Prediction

In the prediction problem, there are always two direct issues that remain to be dealt with, namely the length of the input sequence and the length of the output sequence in one prediction. Now that the former issue has been solved in the above subsection, it comes to the output length. Just as mentioned in the introduction part, existing methods almost depend on the one-step-ahead prediction, which means that they output only one value in one prediction. If a long prediction sequence is required, a recursive one-step-ahead prediction strategy is employed to achieve the prediction of a long sequence. When the length of an expected output sequence is *N*, the one-step-ahead prediction should be recursively repeated *N* times, just as shown in Figure 1a.

In Figure 1a, the boxes in each row represent a one-step-ahead prediction. The blue boxes mean the input sea clutter sequence in one prediction, and the dark blue box denotes the prediction value. After a series of predictions, a long prediction sequence is then obtained. In addition, a viable way to realize the *k*-step-ahead prediction via the one-step-ahead prediction is the recursive one-step-ahead prediction. That is, the current prediction value is supposed to be added to the next input sequence. Apparently, when the required length of the output sequence is relatively long, prediction errors of the recursive one-step-ahead prediction will be accumulated.

Different from the one-step-ahead prediction, a multiple-input multiple-output method is employed to achieve *k*-step-ahead prediction in this paper. Specifically, the length of the input sequence is the same as that of the one-step-ahead prediction, but it can output multiple values in one prediction, just as shown in Figure 1b. In one prediction, the *k*-step-ahead prediction directly outputs *k* values to arrange a prediction sequence. Appar-



ently, the number of predictions for the *k*-step-ahead prediction is much lower than that of the one-step-ahead prediction.

Figure 1. The difference between the one-step-ahead prediction and the *k*-step-ahead prediction. (a) the one-step-ahead prediction; (b) the *k*-step-ahead prediction.

2.4. The SDRDSP Database

In spite of the fact that the IPIX database has been a classical benchmark in sea clutter prediction and floating target detection, it has been nearly 30 years since the first issue of the IPIX database was proposed. Particularly, with the rapid development of advanced radar systems, some physical parameters of the IPIX database, such as the range resolution of 30 m, are relatively out-of-date and may not adequately evaluate modern signal-processing methods.

Learning from the success of the IPIX database and other open-access databases, the SDRDSP database [30] is released by the Naval Aviation University to promote research on clutter characteristics, sea clutter suppression, and target detection. The employed X-band radar is a solid-state power amplifier radar for coastal surveillance and navigation, which has the advantages of high resolution, high reliability, and a small distance blind area. The detailed physical parameters are listed in Table 1, where "Transmitting signal" indicates the radar transmits single frequency pulses (denoted as T1) and linear frequency modulation (denoted as T2) pulses successively. Except for the measured data, the auxiliary material and documentation of experiments are provided in detail, which include the environment parameters of the observation ocean and the information of the automatic identification system (AIS) in cooperative targets.

From 2019 to 2022, authors have proposed 6 issues in the SDRDSP database, including the radar cross-section (RCS) calibration, the observation of sea clutter under different sea conditions, and marine target detection. Since we focus on the sea clutter prediction, the first issue of the SDRDSP database in 2020 is employed in this paper to train and assess each prediction method. During the experiment in this issue, the radar is placed at a bathing beach in Yantai City, the altitude of the experiment scenario is about 400 m, and the height is 80 m [38]. Under different sea conditions, the radar works in the staring mode to observe

sea clutter. The experimental scenario is shown in Figure 2, where the blue area indicates the area of observation.

Table 1. The detailed physical parameters of the employed radar in the SDRDSP database.

Name	Value		
Frequency	9.3~9.5 GHz		
Bandwidth	25 MHz		
Range resolution	6 m		
Pulse repetition frequency	1.6 kHz		
Peak power	50 W		
Antenna length	1.8 m		
Polarization mode	HH		
Operation mode	Staring or scan		
Horizontal beam width	1.2°		
Grazing angle	$1.2^{\circ} \sim 7^{\circ}$		
Transmitting signal	Single frequency pulse Linear frequency modulation pulse		



Figure 2. The diagram of the experimental scene in the Baidu map. The blue area indicates the area of observation.

To be specific, the clutter-only "20210106155330_01_staring" dataset is employed to construct the training and test samples. Herein, the sea condition is 3~4 and there are no moving targets in the observation area due to the high sea condition. In addition, there are 6940 pulses of both the T1 and the T2 in this dataset, and there are 2224 and 4346 range samples of the T1 and T2, respectively. The pulse widths of the T1 and T2 are 0.04 μ s and 3 μ s, respectively. Since linear frequency modulation signals are more commonly used in advanced radar systems, we pay more attention to the T2 mode in the dataset and only samples in the T2 mode are used in the following experiments. For a visual understanding, the time-domain and range–Doppler map are given in Figure 3.

To further analyze the properties of the employed SDRDSP database, the measured amplitude distribution and the frequency spectrum are provided in Figure 4a,b, respectively. The amplitude distribution reveals a trailing phenomenon and the frequency spectrum indicates a Bragg peak near the zero frequency. The temporal and spatial correlations are shown in Figure 5a,b, respectively. By comparing with the threshold 1/e (shown as the red dotted line), the strong temporal correlation and spatial correlation are about 3.53 ms and 13.5 m, respectively.



Figure 3. The time domain and range–Doppler map of the selected dataset. (**a**) the time domain of the T2; (**b**) the range–Doppler map of the T2.



Figure 4. The amplitude distribution and the frequency spectrum of the employed database. (a) the amplitude distribution; (b) the frequency spectrum.



Figure 5. The temporal correlation and the spatial correlation of the employed database. (a) the temporal correlation; (b) the spatial correlation.

3. Approach

In this section, a self-attention enhanced Seq2Seq structure is designed in this paper to tackle the *k*-step-ahead prediction problem. Considering that sea clutter is a chaotic time series, RNNs are still the backbone networks of the proposed structure inspired by the

success of the previous prediction methods. Thus, the basis of RNNs and the inner structure of the proposed SCPNet are introduced, and details of the experiments are provided.

3.1. The Basis of RNNs

In feedforward neural networks, information and features of the current input are transferred unidirectionally. That is, the output of these networks only relies on the current input, which makes it easy for networks to train and optimize. While the output not only relies on the current input but also the previous outputs in many actual tasks. For example, it is difficult for feedforward neural networks to process time series, such as speech and text.

Different from feedforward neural networks, RNNs have loop structures and shortterm memories, which indicates neurons in RNNs can receive information from both other neurons and self-neurons. Then RNNs are capable of processing time sequence data of any length by using neurons with self-feedback. For an input sequence $\mathbf{S} = (s_1, s_2, s_3, \dots, s_T)$, the hidden states \mathbf{H}_t in the corresponding hidden layer with self-feedback are updated as

$$\mathbf{H}_t = f(\mathbf{H}_{t-1}, s_t),\tag{17}$$

where \mathbf{H}_{t-1} denotes the last hidden state and $h_0 = 0$. $f(\cdot)$ is a nonlinear function or a feedforward neural network. That is, there are feedback connections between hidden states in RNNs, and these hidden states are capable of preserving memories and information from previous inputs.

Although plain RNNs can theoretically establish dependencies between states with long ranges, only short-term dependencies can be actually learned due to the problem of gradient explosion or vanishing gradient. To tackle this long-term dependency problem, the gating mechanism is introduced in RNNs to control the accumulating speed of information. The gating mechanism can selectively add new information and selectively forget previously accumulated information. Based on the gating mechanism, the LSTM and gated recurrent unit (GRU) are proposed. The GRU can be viewed as a slightly simplified LSTM cell, which provides less computational complexity [39].

Different from plain RNNs, there are two gates in a GRU, namely, the reset gate and the update gate. These gates are capable of controlling hidden states and the values of their learnable parameters are between 0 to 1. Mathematically, for the same input in Equation (14), the reset gate \mathbf{R}_t is calculated as

$$\mathbf{R}_t = \sigma(\mathbf{W}_{\mathrm{sr}}\mathbf{S} + \mathbf{W}_{\mathrm{hr}}\mathbf{H}_{t-1} + \mathbf{b}_r), \tag{18}$$

where W_{sr} and W_{hr} denote the weights for input sequences and last hidden states, respectively; \mathbf{b}_r is the bias term and $\sigma(\cdot)$ is the Sigmoid activation function. The reset gate actually controls the ratio of task-relevant previous states. Then the update gate \mathbf{U}_t is calculated as

$$\mathbf{U}_t = \sigma(\mathbf{W}_{\mathrm{su}}\mathbf{S} + \mathbf{W}_{\mathrm{hu}}\mathbf{H}_{\mathrm{t-1}} + \mathbf{b}_u),\tag{19}$$

where W_{su} , W_{hu} , and b_u denote the weights and the bias term, respectively.

The reset gate and last hidden states are then combined to calculate the current candidate hidden states $\tilde{\mathbf{H}}_t$ as

$$\mathbf{H}_{t} = \tanh(\mathbf{W}_{sh}\mathbf{S} + \mathbf{W}_{hh}(\mathbf{R}_{t} \odot \mathbf{H}_{t-1}) + \mathbf{b}_{h}), \tag{20}$$

where \mathbf{W}_{sh} , \mathbf{W}_{hh} , and \mathbf{b}_h denote the weights and the bias term, respectively. The " \odot " is the Hadamard product. To combine the update gate \mathbf{U}_t and candidate hidden states, the current hidden states \mathbf{H}_t can be calculated as

$$\mathbf{H}_{t} = \mathbf{U}_{t} \odot \mathbf{H}_{t-1} + (1 - \mathbf{U}_{t}) \odot \tilde{\mathbf{H}}_{t}$$
(21)

For clarity, the reset gate is beneficial to catch the short-term dependency and the update gate is helpful to build the long-term dependency in sequences. The flowchart of a GRU is shown in Figure 6.



Figure 6. The flowchart of a GRU.

3.2. The Structure of the Proposed SCPNet

The proposed SCPNet consists of a Seq2Seq structure and a simple self-attention module, where the Seq2Seq structure is employed to tackle the *k*-step-ahead prediction and the self-attention module is capable of boosting the adaptive feature selection. First, the key to the *k*-step-ahead prediction is to construct a network where its input and output are sequences, and the length of the input sequence and output sequence can be different. For the input sequence $\mathbf{S} = (s_1, s_2, s_3, \dots, s_T)$ and an output sequence $\mathbf{Y} = (y_1, y_2, y_3, \dots, y_N)$, the goal of the Seq2Seq is to approximate the CPD as

$$p(\mathbf{Y}_{1:N}|\mathbf{S}_{1:T}) = \prod_{n=1}^{N} p(y_n | \mathbf{Y}_{1:(n-1)}, \mathbf{S}_{1:T}),$$
(22)

where the subscript means the start and end of the sequence. Then, the maximum likelihood estimation can be used to find the optimization weight $\hat{\theta}$ as follows,

$$\hat{\theta} = \arg\max_{\theta} \sum_{n=1}^{N} \log p(\mathbf{Y}_{1:N} | \mathbf{S}_{1:T}).$$
(23)

Taking the time-series characteristics of sea clutter into consideration, two RNNs are employed in the SCPNet to construct a Seq2Seq model, where one RNN is used as an encoder and another is used as a decoder. For the encoder, the RNN can encode the input sequence **S** to a matrix **E** with a fixed dimension, and the vector usually is the hidden state at the last time step. That is,

$$\mathbf{E} = \mathrm{RNN}_{en} \left(\mathbf{H}_{T-1}^{en}, \theta_{en} \right), \tag{24}$$

where \mathbf{H}_{T-1}^{en} is the hidden state in the previous time step and θ_{en} is the learnable parameter in the encoder RNN.

Another RNN is then employed to encode and output the prediction sequence, where the initial hidden state is the hidden state in the last time step. That is,

$$\mathbf{H}_{0}^{de} = \mathbf{E},\tag{25}$$

at the *t*-th time step of the decoder, the current hidden state is

$$\mathbf{H}_{t}^{de} = \text{RNN}_{de} \left(\mathbf{H}_{t-1}^{de}, \theta_{de} \right), \tag{26}$$

where \mathbf{H}_{t-1}^{de} is the hidden state in the previous time step and θ_{de} is the learnable parameter in the decoder RNN. Finally, the last hidden state of the decoder RNN is employed to obtain the prediction sequence via a linear layer.

Although the GRU and the Seq2Seq model improve the ability to build a long-term dependency, they are still prone to losing information from early stages if the input sequence is too long. Furthermore, the physical and chaotic characteristics of sea clutter are complex. It may be satisfactory for existing LSTM-based methods to achieve the one-step-ahead prediction, while the proposed SCPNet is required to not only build long-term dependency of the input sequence but also find time-series dependency of the prediction sequence with the length of *k* in the *k*-step-ahead prediction.

Thus, to further boost the feature adaptive selection ability, a simple self-attention module is introduced. The self-attention module is capable of constructing an external score to evaluate the significance of the output of each neuron at each moment in the GRU. The attention module aims to adaptively find task-relevant outputs and suppress useless features through the score function. A standard attention module always consists of a query, a key, and a value [24]. Mathematically, let **A** denote the input of the attention module, the query matrix **Q**, the key matrix **K**, and the value matrix **V** are calculated as:

$$\begin{cases} \mathbf{Q} = W_q \mathbf{A} \\ \mathbf{K} = W_k \mathbf{A} \\ \mathbf{V} = W_v \mathbf{A} \end{cases}$$
(27)

where $\mathbf{Q} = \{q_1, q_2, q_3, ..., q_N\}$, $\mathbf{K} = \{k_1, k_2, k_3, ..., k_N\}$, and $\mathbf{V} = \{v_1, v_2, v_3, ..., v_N\}$. For each $q_{n'}$ the score S_c is then calculated as:

$$S_c = \sum_{j=1}^{N} \operatorname{SoftMax}\left(s_f(\boldsymbol{k}_j, \boldsymbol{q}_n)\right) v_j$$
(28)

where SoftMax(·) denotes the SoftMax function and $s_f(\cdot)$ is the score function. In the proposed SCPNet, the score function is achieved via a layer perceptron where the inputs are the outputs of the GRU in the decoder. The outputs of the GRU are then scaled by calculated scores.

In the final layer of the proposed SCPNet, a linear layer is employed to map its inputs to the prediction sequence. To enrich the feature map, the scaled outputs of the decoder and the original outputs of the decoder are concatenated to serve as the inputs of the final linear layer. The detailed setting and structure of the proposed SCPNet are provided in Table 2, where "GRU 1 (Encoder)" means a GRU in the encoder part, the hidden size indicates the size of the hidden state, and "# Layers" means the number of layers. The flowchart of the proposed method is also shown in Figure 7, where "Concatenate" denotes the vector concatenation operation and "Attention" represents the attention module.

Table 2. The detailed structure of the proposed SCPNet.

Layer	Input Size	Output Size	Hidden Size	Activation	# Layers
GRU 1 (Encoder)	1	32	32	Sigmoid Tanh	1
GRU 2 (Encoder)	32	32	32	Sigmoid Tanh	1
GRU 3 (Decoder)	32	32	32	Sigmoid Tanh	1
GRU 4 (Decoder)	32	32	32	Sigmoid Tanh	1
Linear 1 (Attention)	32	32	-	SoftMax	1
Linear 2 (Output)	64	32	-	-	1



Figure 7. The flowchart of the proposed method.

3.3. Details of Experiments

As mentioned above, the clutter-only "20210106155330_01_staring" dataset in the SDRDSP database is employed to construct the training and test samples. Since the T2 mode, namely, the linear frequency modulation mode, is much more commonly used than the single frequency mode in modern radar systems. The measured signals in the T2 mode are used and their absolute values are normalized to serve as the inputs of each method. In addition, due to the relatively weak amplitude of sea clutter far from the coast, only the first 1280 range samples of each pulse are selected for the sake of computational burden.

During the training of each method, a 2-norm gradient clipping method is employed to avoid the gradient exploding problem. The Adam [40] optimization algorithm is used to find the optimized weight of each method. Meanwhile, a cosine annealing warm restart is used for cooperative work with Adam to improve efficiency. As for the hyperparameters, the learning rate is 0.01, the batch size is 2048, and the number of epochs is 200. For all experiments, a server, equipped with a CPU of Intel Xeon Gold 6226R with a RAM of 256 G and a GPU of Nvidia Quadro RTX 6000 with a video memory of 24 G, is used. The software platform includes Python 3.8.5, PyTorch 1.7.1, and CUDA 11.0.

4. Results

In this subsection, the input length is selected through the results of the phase space reconstruction. Then, comparisons of the one-step-ahead prediction of the proposed method and the state-of-the-art methods are provided and analyzed. Moreover, comparisons of the one-step-ahead prediction and the computational complexity of each method are shown and discussed.

4.1. Results of the Phase Space Reconstruction

For a prediction method based on RNNs, the input length is an important hyperparameter that is supposed to influence the prediction performance without a doubt. In many previous works focused on the sea clutter prediction problem, the optimized input length is selected manually by a series of experiments, which is inefficient and exhausting in practice. Fortunately, based on the phase space reconstruction theory, the optimized input length can be found due to the chaotic characteristics.

As mentioned in Section 2, the optimized length equals $m\tau$, just as indicated in Equation (4). The delay time τ is calculated through the mutual information algorithm, and the result is shown in Figure 8a. When the delay time τ is greater than eight, the mutual information is almost stable and no longer decreasing. Hence, the optimized delay time τ is eight. As for the optimized embedding dimension *m*, it is selected by the Cao method and its result is shown in Figure 8b. Similarly, when the embedding dimension is greater than eight, the *E*₁ is no longer decreasing. Then the optimized embedding dimension *m* is 8, which indicates the optimized input length is 64.



Figure 8. The results of the phase space reconstruction. (a) the delay time τ ; (b) the embedding dimension *m*.

4.2. Comparisons on the One-Step-Ahead Prediction

Since the existing prediction methods are based on the one-step-ahead prediction, results with the one-step-ahead prediction of the proposed method and several state-of-the-art methods, including the GRNN, the DNN, the LSTM, and the GF-RNN, are compared and analyzed. Although the proposed method is designed for the *k*-step-ahead prediction, the proposed method can be easily converted to the one-step-ahead prediction. To be specific, if the input sequence and the interval of each input sequence are the same as the existing methods, the proposed method can be viewed as the one-step-ahead prediction by only using the first prediction value in each prediction.

As for the metrics of comparisons, except for the MSE defined in Equation (7), the mean absolute error (MAE) is also employed to evaluate the performance, which is defined as

$$MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} |l(i) - \hat{y}(i)|,$$
(29)

where $|\cdot|$ denotes the absolute value function. The MSE and MAE of each method are provided in Table 3, where results are in the format of average \pm standard deviation.

	SCPNet (Proposed)	GRNN	DNN	LSTM	GF-RNN
MSE	$egin{array}{ll} 1.48 imes 10^{-5} \ \pm 5.2 imes 10^{-6} \end{array}$	$\begin{array}{c} 1.87 \times 10^{-2} \\ \pm 5.1 \times 10^{-3} \end{array}$	$1.02 imes 10^{-3} \ \pm 2.6 imes 10^{-4}$	$3.55 imes 10^{-5} \ \pm 6.2 imes 10^{-6}$	$1.72 imes 10^{-5} \ \pm 5.1 imes 10^{-6}$
MAE	$2.25 imes 10^{-3} \ \pm 3.6 imes 10^{-4}$	$9.9 imes 10^{-2} \ \pm 1.4 imes 10^{-2}$	$2.02 \times 10^{-2} \\ \pm 3.3 \times 10^{-3}$	$\begin{array}{c} 3.71 \times 10^{-3} \\ \pm 3.8 \times 10^{-4} \end{array}$	$2.85 imes 10^{-3} \ \pm 3.5 imes 10^{-4}$

Table 3. The MSE and MAE of each method.

For the one-step-ahead prediction, the MSE and the MAE of the SCPNet are about 1.48×10^{-5} and 2.25×10^{-3} , respectively. That is, the proposed SCPNet is capable of achieving the smallest MSE and MAE, which indicates the SCPNet can still have the best prediction performance compared with the four existing methods. In addition, compared with the GRNN and the DNN, the RNN-based networks, namely, the SCPNet, the

LSTM, and the GF-RNN, obviously gain better prediction performance, which verifies the superiority of RNNs.

For a visible comparison, the prediction result of each method is shown in Figure 9, where "Raw" denotes the ground truth of sea clutter and red points are the prediction results of each method. Due to the poor prediction performance of the GRNN, its result is not provided herein. Overall, four methods are capable of fitting the variation trend of sea clutter, but there are a lot of prediction errors of the DNN and the LSTM at the bottom of the sea clutter. The prediction performance of the proposed SCPNet is slightly better than that of the GF-RNN in this clutter sample. To summarize, it is not a difficult task to achieve the one-step-ahead prediction of existing RNN-based methods, and the proposed SCPNet has the best prediction performance.



Figure 9. The prediction result of each method on the one-step-ahead prediction. (**a**) the DNN; (**b**) the LSTM; (**c**) the GF-RNN; (**d**) the proposed SCPNet.

4.3. Comparisons on the k-Step-Ahead Prediction

In the *k*-step-ahead prediction, the prediction method is required to predict the *k* value in one prediction. For those existing methods, it is a viable way to use the recursive one-step-ahead prediction and then achieve the *k*-step-ahead prediction indirectly. The proposed method uses the Seq2Seq structure to directly achieve the *k*-step-ahead prediction. Then, the results on the *k*-step-ahead prediction are analyzed and discussed. Table 4 shows the MSE of each method with different *k* in the *k*-step-ahead prediction.

For clarity, only three RNN-based methods are compared in Table 4. When *k* is 1, it then becomes the one-step-ahead prediction. When *k* equals 2, the MSE of the proposed SCPNet is 1.35×10^{-4} which is about 0.56×10^{-4} and 1.4×10^{-4} lower than that of the GF-RNN and the LSTM, respectively. Compared with the MSE of the one-step-ahead prediction, the MSE of each method is remarkably decreased when *k* = 2. Even for the proposed method, the MSE is about 1.2×10^{-4} higher compared with the one-step-ahead prediction. When *k* is gradually increasing, the MSE of each method is also increasing.

However, no matter what k is, the MSE of the proposed SCPNet is the lowest compared to that of the GF-RNN and the LSTM. It is of note that the MSE of the LSTM is dramatically decreased when k = 32, which is because prediction errors are supposed to be accumulated in the recursive one-step-ahead prediction.

	k = 1	<i>k</i> = 2	k = 4	<i>k</i> = 8	k = 16	<i>k</i> = 32
SCPNet (Proposed)	$1.48 imes 10^{-5}$	$1.35 imes 10^{-4}$	$1.39 imes10^{-3}$	$8.76 imes 10^{-3}$	0.016	0.021
LSTM	$3.55 imes 10^{-5}$	$2.75 imes10^{-4}$	$2.76 imes 10^{-3}$	0.015	0.058	0.313
GF-RNN	$1.72 imes 10^{-5}$	$1.91 imes 10^{-4}$	$2.23 imes 10^{-3}$	0.011	0.024	0.035

Table 4. The MSE of each method with different *k*.

For a visible comparison, the prediction result when k is 8 of each method is shown in Figure 10. The prediction sequence of the DNN can hardly fit the raw clutter sequence, even the prominent peak has not been predicted by the DNN. As for the LSTM, the prediction peak is also much lower than the raw peak and there are a lot of prediction errors at the bottom of the clutter sequence. The prediction peak of the GR-RNN, nevertheless, is higher than the raw clutter sequence. Moreover, there are a large number of false peaks around fluctuations of the raw clutter sequence, which means there will be lots of false high values in the prediction sequence. As for the proposed SCPNet, it can fit most peaks and fluctuations of the raw clutter sequence and there are only a few false high values in the prediction sequence, which verifies the superiority of the proposed method in the *k*-step-ahead prediction.



Figure 10. The prediction result of each method on the 8-step-ahead prediction. (**a**) the DNN; (**b**) the LSTM; (**c**) the GF-RNN; (**d**) the proposed SCPNet.

5. Discussions

5.1. The Consistency Analysis

Furthermore, to evaluate the prediction results in depth, the consistency between the predictions and raw clutter sequence is analyzed and shown in Figure 11, where the stateof-the-art GF-RNN is compared. Herein, the red dotted line denotes the prediction results equal to the raw clutter sequence. Hence, fewer outliers away from the red dotted line indicate better prediction performance. The prediction results of both of the two methods are more located under the normalized amplitude of 0.5. In the area of the real normalized amplitude of 0 to 0.3, the blue points of the SCPNet are more concentrated around the red dotted line than the orange points of the GF-RNN.



Figure 11. The consistency between prediction methods and the raw clutter sequence.

Most outliers of the SCPNet are located at the low right part and under the red dotted line, which means false predictions of the SCPNet tend to be lower than the real values. On the contrary, most outliers of the GF-RNN are located in the upper left part, which indicates false predictions of the GF-RNN tend to be higher than the real values. Taking the time-domain cancellation into consideration, if the prediction sequence is higher than the real clutter, the amplitude of the target of interest would also be weakened falsely, which is prone to losing targets in the following detection step.

To conclude, it is more challenging to achieve the *k*-step-ahead prediction compared with the one-step-ahead prediction. When the *k* is gradually becoming greater, prediction errors of each method are also increasing. Especially for the recursive one-step-ahead pre-diction, prediction errors are supposed to be accumulated, which causes remarkable performance degradation. On the other hand, the proposed SCPNet achieves better prediction performance on the *k*-step-ahead prediction compared with existing methods.

Since the characteristics of input sea clutter under kinds of sea conditions are different as well. To be specific, amplitude distributions of sea clutter are supposed to change with different sea conditions and observation areas, and some environmental parameters, including the wind velocity, will also have an impact on the spectrum of sea clutter. On the other hand, the proposed method based on supervised learning requires that the training dataset and the test dataset should conform to the same distribution so that the learning method can achieve better performance. Sea clutter with different characteristics obviously does not belong to the same distribution. Thus, the accuracy of prediction is supposed to be related to the distributions and the spectrum of sea clutter. For the sake of further applications, the proposed method can be improved via transfer learning under different distributions and spectrums of clutter.

5.2. The Performance of Target Detection

As pointed out in the introduction, the aim of the proposed method is to use extrapolated clutter sequence to detect the presence of a target through time cancellation. In the radar community, the MTI is a well-known and widely used clutter-suppression method based on time cancellation. To demonstrate the performance of clutter suppression and target detection, a target-containing dataset, namely, the "20210106160919_01_staring" dataset is employed to verify the performance of both the proposed method and the MTI method. The raw echo sequence and detection results are shown in Figure 12. Herein, the MTI indicates a three-pulse-cancellation MTI method.



Figure 12. The raw echo sequence and detection results. (**a**) the raw sequence; (**b**) the detection result of the MTI method; (**c**) the detection result of the proposed method.

In Figure 12a, the target of interest is partly covered by the wild clutter and it is very challenging to detect the target while keeping a satisfactory probability of false alarm. Although wild clutter amplitudes are greatly decreased after the MTI method, the target of interest is drastically suppressed as well due to the low speed of the target. As for the result of the proposed method, most wild clutter amplitudes are suppressed while the target is retained. Compared with the raw sequence, the target of the proposed method is much more distinguished, which is supposed to be beneficial to target detection.

5.3. The Computational Complexity

For the sake of practical applications, it is also necessary to analyze the computational complexity of each method. Herein, three metrics are employed to measure the computational complexity, including the inference time, the floating-point operations (FLOPs), and the learnable parameters. The inference time is the consuming time of one prediction on the above-mentioned hardware platform, the FLOPs are widely used to measure the

complexity of a neural network, and the learnable parameters indicate the number of learnable data in a neural network.

The computational complexity of each method is listed in Table 5. The inference time of the proposed SCPNet is about 9.8 ms, the FLOPs are 1685 K, and the SCPNet has about 23 K learnable parameters. Although the DNN has the lowest computational complexity in Table 5, the prediction performance of the DNN in the k-step-ahead prediction is the worst. The computational complexity of the GF-RNN and the LSTM is also lower than that of the proposed SCPNet, while the inference time of the SCPNet is only 2 ms more than that of the GF-RNN. More importantly, the computational complexity of the SCPNet is actually based on a 32-step-ahead prediction. That is, based on the computational complexity listed in Table 5, the SCPNet can predict 32 values in one prediction while the rest of the methods only predict one value. Hence, the proposed method is still more efficient than the GF-RNN in practice.

	SCPNet (Proposed)	GF-RNN	LSTM	DNN
Time (ms)	9.8	7.8	3	0.6
FLOPs (K)	1685	246	305	1.3
Parameter (K)	23	3.5	4.5	1.2

Table 5. The computational complexity of each method.

6. Conclusions

In this paper, an attention-enhanced Seq2Seq structure is designed to achieve both the one-step-ahead prediction and the *k*-step-ahead prediction. The proposed SCPNet takes full advantage of the powerful ability of the Seq2Seq model in the time-series processing and the adaptive feature selection of the self-attention module. In addition, the phase space reconstruction theory is employed to find the optimized input sequence. Results with the SDRDSP database demonstrate that the proposed SCPNet is capable of achieving the lowest MSE in both the one-step-ahead prediction and the *k*-step-ahead prediction, which verifies that the proposed method outperforms the GRNN, the DNN, the LSTM, and the GF-RNN. The analysis of the computational complexity shows the proposed method takes relatively more FLOPs and parameters while the proposed method is still more efficient in practice. The promising results of the proposed method also underscore the potential of the Seq2Seq structure in the *k*-step-ahead prediction of sea clutter.

Author Contributions: Conceptualization, Q.Q. and H.C.; methodology, Q.Q.; software, Q.Q. and Z.L.; validation, B.L., Q.D. and Y.W.; writing—original draft preparation, Q.Q.; writing—review and editing, H.C., Z.L. and Y.W.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grants 62001510.

Data Availability Statement: Not applicable.

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

References

- Chen, X.; Su, N.; Huang, Y.; Guan, J. False-alarm-controllable radar detection for marine target based on multi features fusion via CNNs. *IEEE Sens. J.* 2021, 21, 9099–9111. [CrossRef]
- Lei, Z.; Chen, H.; Zhang, Z.; Dou, G.; Wang, Y. A cognitive beamforming method via range-Doppler map feature for skywave radar. *Remote Sens.* 2022, 14, 2879. [CrossRef]
- Zhang, T.; Wang, Z.; Xing, M.; Zhang, S.; Wang, Y. Research on multi-domain dimensionality reduction joint adaptive processing method for range ambiguous clutter of FDA-Phase-MIMO space-based early warning radar. *Remote Sens.* 2022, 14, 5536. [CrossRef]
- 4. Lv, M.; Zhou, C. Study on sea clutter suppression methods based on a realistic radar dataset. *Remote Sens.* 2019, 11, 2721. [CrossRef]

- Xie, H. Sea Clutter Suppression Method Based on Complex-Valued Neural Network. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2022.
- 6. Poon, M.W.Y. A Clutter Suppression Scheme for High Frequency (HF) Radar. Master's Thesis, Memorial University of Newfoundland, St. John's, NL, Canada, 1991.
- Jangal, F.; Saillant, S.; Hélier, M. Ionospheric clutter cancellation and wavelet analysis. In Proceedings of the European Conference on Antennas and Propagation, Nice, France, 6–10 November 2006; pp. 1–6.
- Huang, N.E.; Zheng, S.; Steven, R.I. A new view of nonlinear water waves: The Hilbert spectrum. Annu. Rev. Fluid Mech 1999, 31, 417–457. [CrossRef]
- Qu, Q.; Wang, Y.-L.; Liu, W.; Li, B. A false alarm controllable detection method based on CNN for sea-surface small targets. IEEE Geosci. Remote Sens. Lett. 2022, 19, 1–5. [CrossRef]
- Rosenberg, L.; Watts, S. Radar Sea Clutter: Modelling and Target Detection; Institution of Engineering and Technology: London, UK, 2022; pp. 32–38.
- Haykin, S.; Leung, H. Model reconstruction of chaotic dynamics: First preliminary radar results. In Proceedings of the ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, CA, USA, 23–26 March 1992; pp. 125–128. [CrossRef]
- 12. Leung, H.; Lo, T. Chaotic radar signal processing over the sea. IEEE J. Ocean. Eng. 1993, 18, 287–295. [CrossRef]
- Li, B.; Haykin, S. Chaotic detection of small target in sea clutter. In Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; pp. 237–240. [CrossRef]
- 14. Leung, H. Nonlinear clutter cancellation and detection using a memory-based predictor. *IEEE Trans. Aerosp. Electron. Syst.* **1996**, 32, 1249–1256. [CrossRef]
- Xie, N.; Leung, H.; Chan, H. A multiple-model prediction approach for sea clutter modeling. *IEEE Trans. Geosci. Remote Sens.* 2003, 41, 1491–1502. [CrossRef]
- 16. Jiang, B.; Wang, H.; Fu, Y.; Li, X.; Guo, G.R. Prediction of sea clutter chaotic time series using LS-SVM. *Prog. Nat. Sci.* 2007, 17, 415–420. (In Chinese)
- Gao, Z.; Chen, L. Sea clutter sequences regression prediction based on PSO-GRNN method. In Proceedings of the 2015 8th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 12–13 December 2015; pp. 72–75. [CrossRef]
- 18. Cognitive Systems Laboratory, McMaster University, Canada. The IPIX Radar Database. Available online: http://soma.mcmaster.ca//ipix.php (accessed on 28 February 2022).
- 19. Xing, H.; Yan, Y. Detection of low-flying target under the sea clutter background based on Volterra filter. *Complexity* **2018**, *12*, 591. [CrossRef]
- 20. Chen, Z.; Tian, B.; Zhang, S.; Xu, Q. quantitative prediction of sea clutter power based on improved grey Markov model. *Atmosphere* **2022**, *13*, 1085. [CrossRef]
- Qu, Q.; Wei, S.; Liu, S.; Liang, J.; Shi, J. JRNet: Jamming recognition networks for radar compound suppression jamming signals. IEEE Trans. Veh. Technol. 2020, 69, 15035–15045. [CrossRef]
- 22. Wei, S.; Qu, Q.; Zeng, X.; Liang, J.; Shi, J.; Zhang, X. Self-Attention Bi-LSTM networks for radar signal modulation recognition. *IEEE Trans. Microw. Theory Tech.* **2021**, *69*, 5160–5172. [CrossRef]
- Qu, Q.; Wang, Y.-L.; Liu, W.; Wei, S.; Du, Q. IRNet: Interference recognition networks for automotive radars via autocorrelation features. *IEEE Trans. Microw. Theory Tech.* 2022, 70, 2762–2774. [CrossRef]
- 24. Qiu, X. Neural Networks and Deep Learning; Machine Press: Beijing, China, 2020; pp. 131–135.
- 25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
- Zhao, J.; Wu, J.; Guo, X.; Han, J.; Yang, K.; Wang, H. Prediction of radar sea clutter based on LSTM. J. Ambient. Intell. Human Comput. 2019, 335, 4. [CrossRef]
- Ma, L.; Wu, J.; Zhang, J.; Wu, Z.; Jeon, G.; Tan, M.; Zhang, Y. Sea clutter amplitude prediction using a long short-term memory neural network. *Remote Sens.* 2019, 11, 2826. [CrossRef]
- Chen, X.; Wu, J.; Guo, X. Prediction of sea clutter characteristics by deep neural networks using marine environmental factors. *Environ. Dev. Sus.* 2022, 26, 1–20. [CrossRef]
- Ma, L.; Zhang, J.; Wu, J.; Zhang, Y.; Zhao, P.; Xia, X. Prediction of sea clutter using gated feedback recurrent neural network. *Chin. J. Radio Sci.* 2020, 35, 257–263. (In Chinese)
- 30. Liu, N.; Dong, Y.; Wang, G.; Ding, H.; Huang, Y.; Guan, J.; Chen, X.L.; He, Y. Sea-detecting X-band radar and data acquisition program. *J. Radars* **2019**, *8*, 656–667. [CrossRef]
- 31. Ding, H.; Dong, Y.; Liu, N.; Wang, G.; Guan, J. Overview and prospects of research on sea clutter property cognition. *J. Radars* **2016**, *5*, 499–516. [CrossRef]
- 32. Khan, R.H. Ocean-clutter model for high-frequency radar. IEEE J. Ocean. Eng. 1991, 16, 181–188. [CrossRef]
- Ma, L. Research on Sea Clutter Characteristics Based on Deep Learning and Marine Environmental Parameters. Ph.D.'s Thesis, Xidian University, Shanxi, China, 2020.
- 34. Takens, F. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics;* Morel, J.M., Teissier, B., Eds.; Springer: Cham, Switzerland, 1981; pp. 366–381.

- 35. He, K. Research on Sea Clutter Suppression Method of HFSWR Based on RBF Neural Network. Master's Thesis, Jiangsu University of Science and Technology, Zhenjiang, China, 2021.
- 36. Hornik, K.; Stinchcombe, M.; White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
- 37. Cao, L. Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series. *Phys. Rev.* **1997**, *110*, 43–50. [CrossRef]
- Liu, N.; Ding, H.; Huang, Y.; Dong, Y.L.; Wang, G.Q.; Dong, K. Annual progress of the sea-detecting X-band radar and data acquisition program. J. Radars 2021, 10, 173–182. [CrossRef]
- Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Gated feedback recurrent neural networks. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015; pp. 2067–2075.
- Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.