Target Detection in Clutter/Interference Regions Based on Deep Feature Fusion for HFSWR

Maokai Wu, Ling Zhang , Jiong Niu, and Q. M. Jonathan Wu , *Senior Member, IEEE*

Abstract—High-frequency surface wave radar (HFSWR) is of great significance for maritime detection, but in the HFSWR echo signal, ship targets are often submerged in a variety of clutter and interference, making it difficult to detect vessels. In this paper, we propose an intelligent detection algorithm for targets concealed in strong clutter and complex interference environments. The algorithm has two stages: preprocessing and target detection. In the preprocessing stage, faster region-based convolutional neural networks Faster R-CNN are designed to identify and locate clutter and interference regions in the range Doppler spectrum; in the target detection stage, a two-level cascade algorithm is proposed. First, an extremum detection algorithm is proposed to identify suspicious target points in the clutter/interference regions, including real and false target points, to quickly obtain potential target positions. Second, in consideration of the characteristics of radar targets, two lightweight networks are designed to extract the CNN features and the stacked autoencoder features of the potential target locations. Then, fusion features are obtained and sent to an extreme learning machine that acts as a second-level classifier to distinguish between real and false target points. Experiments show that the proposed HFSWR target-detection algorithm has better performance for vessel detection in clutter/interference regions than the current mainstream detection algorithms.

Index Terms—Convolutional neural network (CNN), extremum detection, feature fusion, high-frequency surface wave radar (HFSWR), stacked autoencoder (SAE), target detection.

I. INTRODUCTION

HIGH-FREQUENCY surface wave radar (HFSWR) has over- the-horizon detection ability for real time monitoring of large areas of the exclusive economic zone and has been widely used for vessel target detection [1]. However, because HFSWR operates in the high-frequency band in complex electromagnetic environments, its echo signal contains not only target information, but also a large amount of clutter and interference, such as ground clutter, sea clutter, ionospheric clutter, and radio frequency interference (RFI). Therefore, it is important to accurately and effectively detect targets affected or even submerged by clutter and interference in complex environments [2].

The constant false-alarm rate (CFAR) [3] is a classical algorithm widely used in target detection. The common CFAR detection methods include mean, ordered statistics, and adaptive regression. The CFAR usually establishes a test statistic for each test cell of interest and compares it with a threshold value calculated from a reference cell in accordance with some criterion to obtain test results. The classical cell average CFAR (CA-CFAR) [4] uses the average value of the reference cell to obtain the detection threshold. The ordered statistic CFAR was proposed by He and Rohling [5] to sort the reference values near the test cell of interest and select one as the detection threshold, which removes the influence of wild values on noise estimation by sorting to ensure the accuracy of the threshold detection. The adaptive power regression CFAR algorithm proposed by Dzvonkovskaya and Rohling [6] set the maximum value of the upper confidence bounds of the range regression and Doppler regression curves at the cell of interest as the detection threshold. This method estimates the noise from a global perspective, which can effectively suppress a wide range of clutter and interference. Nowadays, due to the rapidity of CFAR detection, it is widely used in the preprocessing stage of the radar target detection field [7], [8].

On the basis of the CFAR algorithm, many ideas have been proposed for HFSWR target-detection algorithms. Jangal *et al.* [9], [10] proposed a wavelet-transform-based algorithm for target point detection in the range Doppler (RD) spectrum based on differences in the morphology and distribution of clutter and target points. This method used the multiscale characteristics of the wavelet transform to separate the clutter from the target points via the low-frequency and high-frequency components of the image. On the basis of this, Li *et al.* [11] proposed a detection algorithm based on discrete wavelet transforms, which can adaptively determine the scale of the wavelet transform and enhance the high-frequency coefficients to effectively eliminate clutter and noise while reducing the false-alarm rate. Similarly, on the basis of the variability of clutter and target morphological components in the RD spectrum, Grosdidier and Baussard [12] proposed a target-detection method based on morphological component analysis using multi-scale transformation and sparse expression for target extraction. Zhang *et al.* [13] introduced the concept of reverse phase transition in the Duffing oscillator and proposed a detection algorithm for HFSWR targets with a low signal-to-noise ratio (SNR). Cai *et al.* [14] proposed a target-detection method based on time-frequency analysis to obtain the time-frequency representation of the echo signal using

Maokai Wu, Ling Zhang, and Jiong Niu are with the College of Engineering, Ocean University of China, Qingdao 266100, China (e-mail: wumaokai97@163.com; zljoan@163.com; niujiong@ouc.edu.cn).

Q. M. Jonathan Wu is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B3P4, Canada (e-mail: jwu@uwindsor.ca).

a synchronous extraction transform algorithm. Although these algorithms can detect targets in certain circumstances, their detection performance is severely affected by human- determined factors, such as the scale of wavelet transform, complete dictionary, or *a priori* knowledge.

In recent years, some intelligent target-detection schemes have been applied in HFSWR target detection. Zhang *et al*. [15] proposed a scheme based on an error self-adjustment extreme learning machine (ES-ELM) and a cascade classifier. The Haar-like features of the potential target regions generated by the first-level classifier are extracted and sent into the ES-ELM for learning and classification. Subsequently, Zhang *et al*. [16] proposed an optimized ES-ELM (OES-ELM) and applied it to HFSWR target detection to further improve performance. After that, on the basis of characteristics of the target points in the RD spectrum, Li and Han [17] extracted and fused the textural and morphological features of the target and sent them to the classifier. These intelligent algorithms all require human selection of features and training of the classifier. The quality of the features selected will directly affect the results produced by the classifier and the quality of the target detection. Second, simple morphological features tend not to be very robust in complex environments with strong clutter and interference.

From the above analysis of the current HFSWR target- detection algorithm, several problems can be identified as follows.

1) Conventional algorithms often rely on artificial factors or conditions. It is difficult to set these artificial parameters in practice, but inaccurate parameters will cause a higher false-alarm rate in the detection process.
2) Some intelligent detection algorithms perform detection by extracting features of target points. The quality of the extracted features directly affects the performance of classification and target detection, while effective features require a lot of time and experience to design.
3) Most of the current algorithms have good detection in a stable environment, but drastically worse performance in strong clutter or interference due to the reduction of the SNR. In these circumstances, the extracted morphological features in the algorithms cannot effectively distinguish target and clutter.

To reduce the difficulty of target detection in complex environments affected by strong clutter and interference, we propose an intelligent algorithm based on cascade deep learning networks combined with ELM and feature fusion to effectively detect targets under the influence of strong clutter and interference, including targets at the edge of the clutter/interference region and partially submerged by the clutter/interference. Therefore, identification and localization of clutter and interference regions in the RD spectrum is the first step. Conventional detection of clutter or interference regions is achieved by image segmentation or classifier classification [18]–[20], and detection performance depends on many artificial factors. However, the deep-learning-based faster region-based convolutional neural network R-CNN detection framework [21], which can simultaneously localize and classify the target in less time through learning, has a good detection effect on normal-sized targets. Hence, faster R-CNN is designed for clutter and interference region detection in this article.

In clutter and interference regions, the morphology and energy of the target points vary. Conventional manually designed target-feature extraction operators, such as the Haar-like feature and local binary patterns, are often not very robust. Deep semantic features tend to perform better [22], [23]. Therefore, to avoid the trouble of designing preselected features manually, and to improve the adaptability of the extracted features in strong clutter or interference environments, we designed two lightweight machine-learning-based feature extractors, a CNN [24]–[26] and a stacked autoencoder (SAE) [27], to automatically extract features from different perspectives. The structure of the two networks is simplified, to avoid the time consumption of feature-extraction networks. Different types of features generally represent objects with different meanings. Multifeature fusion has been shown to bridge the gaps between different features and improve the accuracy of classification [28]–[30]. In view of this, the features extracted from two networks are sent to the multifeature fusion center to further improve discriminatory ability.

In this article, a two-level cascade target-detection algorithm for clutter and interference regions is proposed. Clutter and interference regions in the RD spectrum are detected by the faster R-CNN algorithm. The first level is an extremum detector based on grayscale values. In the first level, inspired by the CFAR algorithm, we use the average grayscale values of the reference cells around the cell under test to obtain the detection threshold, and then determine whether the cell under test is a suspicious target point. The purpose of the extremum detector we use in this level is to separate all suspicious target points from the clutter and interference region to improve the real time performance of the algorithm and to complete the target localization. The number of false-alarm targets in the first-level detection result is not in itself important, but does ensure that the detection rate is high enough to prevent the proposed algorithm from missing the target points in the first level. False targets will be removed in the next level of the algorithm. Next, lightweight CNN and SAE are combined to automatically extract potential region features from different perspectives. The features extracted from the two networks are processed and concatenated, and a robust sparse linear discriminant analysis (RSLDA) [31] algorithm is used to reduce redundancy and improve the discriminatory ability of the fusion features. Finally, the fusion features are sent to the second-level classifier, the ELM [32],[33], to classify true target points efficiently.

The article is organized as follows: Section II provides a brief overview of the general framework of our proposed HFSWR target-detection algorithm. Section III describes the clutter/ interference identification and localization method in the RD spectrum based on faster R-CNN. Section IV describes the proposed two-level cascade target-detection algorithm. Experimental results are shown in Section V. Finally, Section VI concludes the article.

II. PROPOSED FRAMEWORK

Fig. 1 shows a typical RD spectrum of HFSWR with a large amount of RFI. The horizontal and vertical coordinates represent the Doppler frequency shift and the range cell, respectively. The change of color from red to blue indicates that the energy
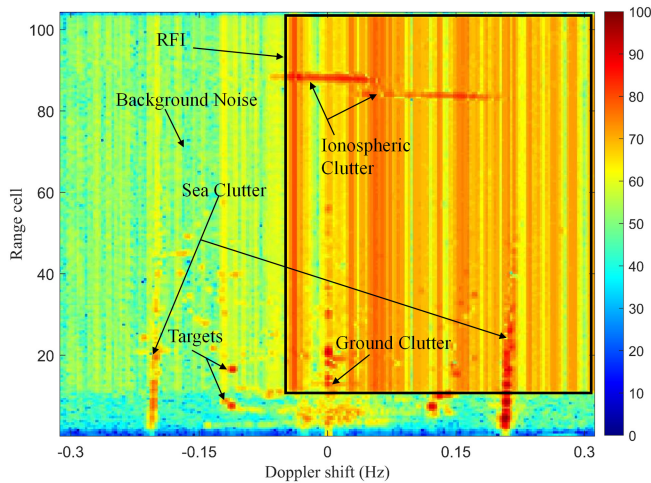
Fig. 1. Actual measured RD spectrum.

gradually decreases. As this figure shows, the RD spectrum contains the target points, sea clutter, ground clutter, ionospheric clutter, different levels of RFI, and background noise. Among these, the target points are shown as isolated peaks in the RD spectrum with a certain degree of expansion in both the Doppler and range directions. The ground clutter is located at zero Doppler frequency and has a certain degree of expansion in the Doppler direction and a ridge-like distribution with energy attenuation in the range direction. Sea clutter also has some expansion in the Doppler direction and a ridged distribution in the range direction. However, ionospheric clutter has some expansion in the range direction and a ridged distribution in the Doppler direction. RFI mainly shows expansion along the range direction in the RD spectrum, and the energy does not decrease with range. In addition, in the RD spectrum, many targets are seriously affected by clutter and RFI.

In this article, an intelligent detection method for vessel targets in strong clutter and complex interference environments is proposed. The framework of the proposed target-detection algorithm is shown in Fig. 2. The algorithm has two stages. The first stage is preprocessing, in which faster R-CNN is used to identify and locate the position of the clutter or interference in the RD spectrum. The details are illustrated in Section III. The second stage is target detection, which is based on a two-level cascade detection algorithm. The first level is an extremum detector for extracting potential target locations from clutter and interference regions. The detector is fast at detecting targets but hits a lot of false targets. The second level is an ELM that is responsible for discriminating between true and false targets among the candidates extracted by the first level, whose input is the fusion features of the potential target locations. The fusion features are obtained from the features extracted by the two feature extractors separately, and processed by feature processing. See section IV for details.

III. CLUTTER AND INTERFERENCE REGION DETECTION ALGORITHM

Because the morphology of targets outside the clutter and interference regions is obvious, traditional detection methods,

such as CFAR can deal with them easily. This paper focuses on the detection of targets within clutter and interference regions. In the preprocessing stage, we need to identify and localize the interference and clutter regions in the RD spectrum. On the basis of the ideas in [34], we use Faster R-CNN as a detection framework to identify and locate clutter and interference regions in the RD spectrum.

Faster R-CNN is a two-level, end-to-end target-detection framework that contains two subnetworks: one for the region proposal network to generate the region proposal box, and the other for Fast R-CNN, for target classification and localization. In this article, ResNet-101, which is pretrained on ImageNet, is used as the backbone of faster R-CNN. We select the actual measured RD spectrum to make a dataset and annotate sea clutter, RFI, and ionospheric clutter. Fig. 3 shows the results of clutter and interference detection using faster R-CNN: the framework can effectively identify and locate the clutter regions. Testing shows that faster R-CNN has a high detection rate for clutter and interference in the RD spectrum, and the mean average precision is close to 1. In terms of detection time, the algorithm used can complete the detection of an RD spectrum in 0.97s.

IV. PROPOSED TWO-LEVEL CASCADE DETECTION ALGORITHM

A. First-Level Detector—Extremum Detector

Fig. 4 shows the result of graying out an RD spectrum. The closer the color is to black, the higher the energy and the lower the grayscale value. The grayscale values of the target points, the clutter, and most of the RFI are lower than those of the background component in the clutter and interference regions. Fig. 5 shows the morphology of the target point in the grayed—out RD spectrum, with the vertical coordinate representing the grayscale values; from which it can be seen that the target point is morphologically represented as isolated extreme point. Therefore, we propose a grayscale feature-based extremum detector, which uses the average grayscale value of the reference cells around the cell under test as the evaluation criterion to quickly identify all the suspicious target points in the clutter and interference regions.

The first-level detector extracts all candidate targets using the following criterion:

$$f(x_{\text{test}}) = \begin{cases} 1, & h(x_{\text{test}}) \le \alpha k \\ 0, & h(x_{\text{test}}) > \alpha k \end{cases} \quad (1)$$

where α is the threshold factor, k is the average grayscale value of the reference cells, $h(x_{\text{test}})$ is the grayscale value of the test cell, $f(x_{\text{test}})$ is the category to which the cell belongs, 1 is the suspected target point, and 0 is the background component. The target points, clutter, and RFI have lower grayscale values than the background. Therefore, the points below the detection threshold are suspicious targets to be extracted. Statistically, almost all target sizes in our data are within the range of 15×15, so we select a 15×15 rectangular box around the cell under test as the reference cell. Here, we need to ensure that the detection rate P_d is high enough to detect all possible target points to avoid missing targets at the first level of the network, regardless of the
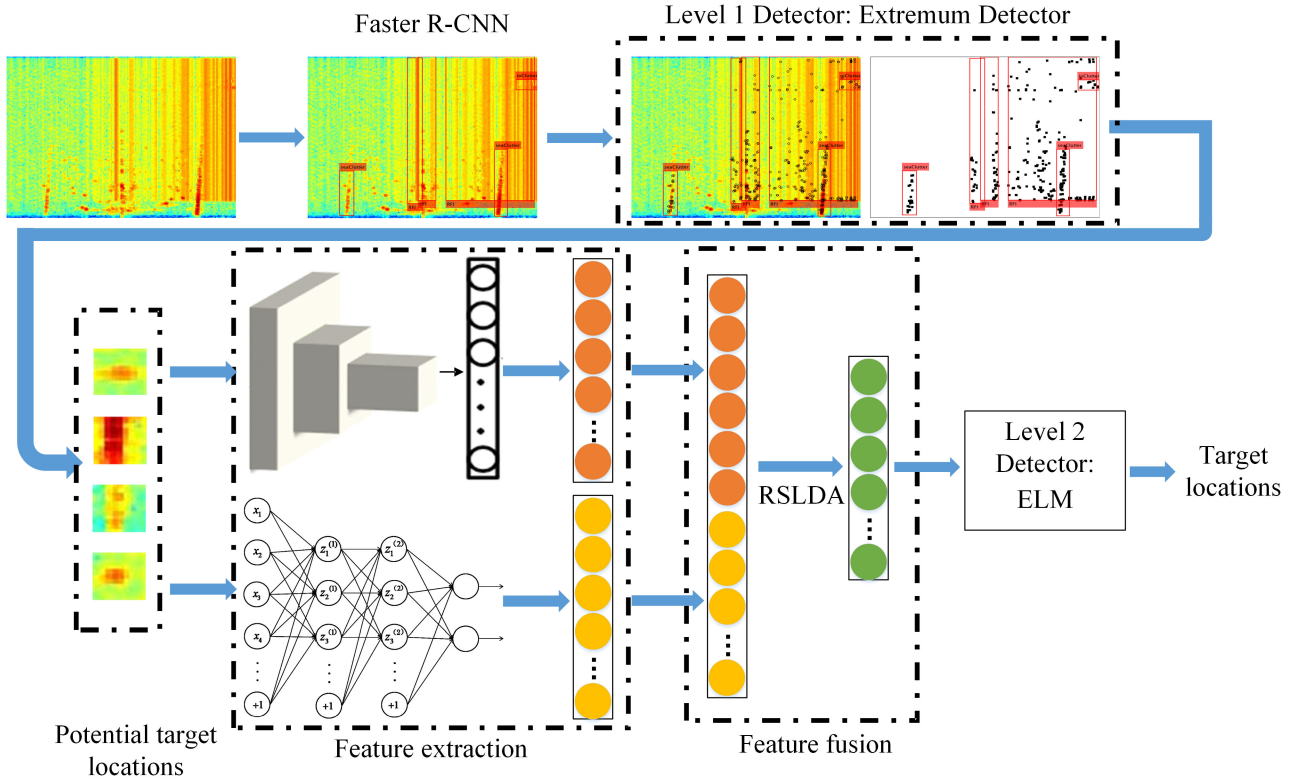
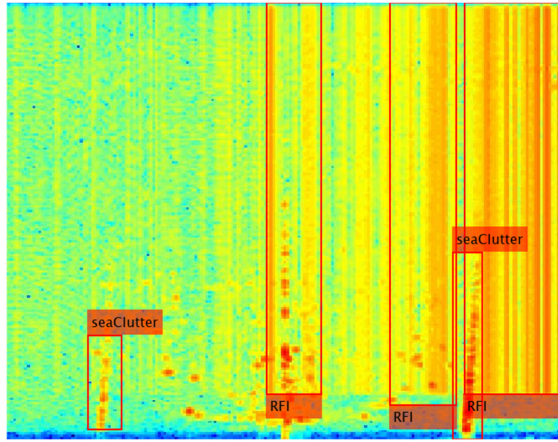Fig. 2. Framework of the proposed target-detection algorithm.

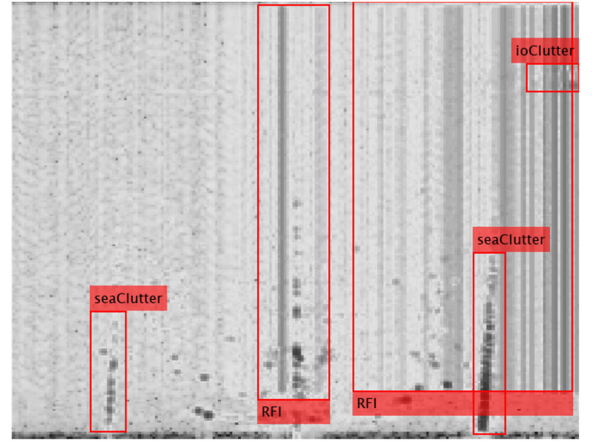Fig. 3. One example of the results of the clutter/interference detection.

Fig. 4. Typical grayed-out RD spectrum.

false-alarm rate P_f. P_d and P_f can be obtained as follows:

$$P_d = D_R/(D_F + D_N) \qquad (2)$$

$$P_f = D_F/(D_R + D_F) \qquad (3)$$

where D_R is the true target points detected, D_F is the false target points detected, and D_N is the undetected target points. The threshold factor α can change the detection rate and the false-alarm rate of the test results. To guarantee that all the targets are detected by the extremum detector, the appropriate value of α should be determined, which is easier compared to conventional CFAR. Fig. 6 shows the trend of P_d and P_f in the clutter and interference regions as α changes. As α increases, both P_d and

P_f increase rapidly. All targets in the clutter and interference regions are detected when $\alpha = 0.85$. Fig. 7 shows the detection result of a typical RD spectrum when $\alpha = 0.85$. It can be seen from the detection result that although the detection result is full of a large number of false targets, the first-level detector hits all the suspicious targets and avoids missing targets.

The extremum detection results in a set of candidate target points $\{(x_i, y_i) | i \in \{f(x_{test}) = 1\}\}$, where (x_i, y_i) is the ith central coordinate of the suspicious target point. Considering that a 15×15 area can contain most of the targets in our data, the potential target position is set to a 15×15 neighborhood of the central coordinates of each candidate target point.
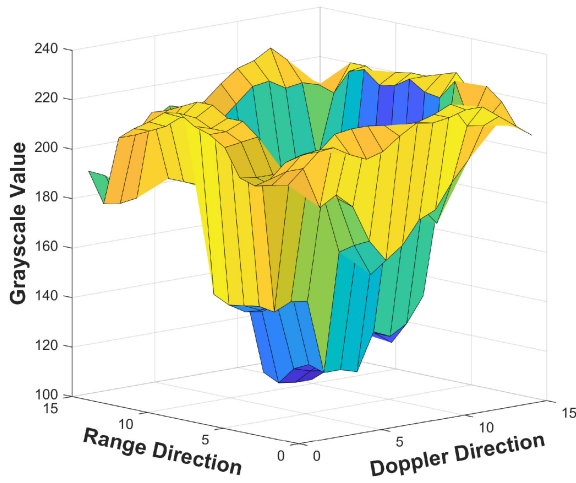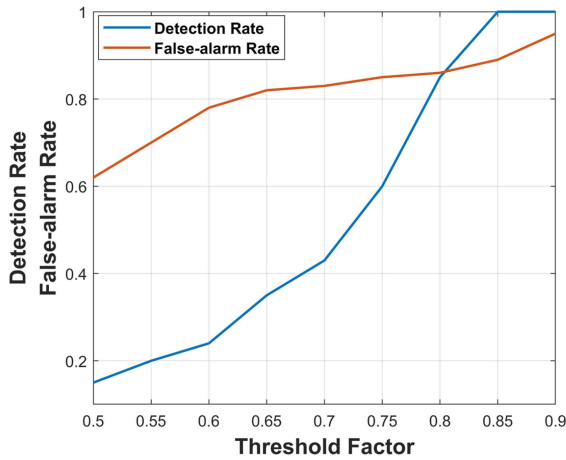
Fig. 5. Morphology of the target point.

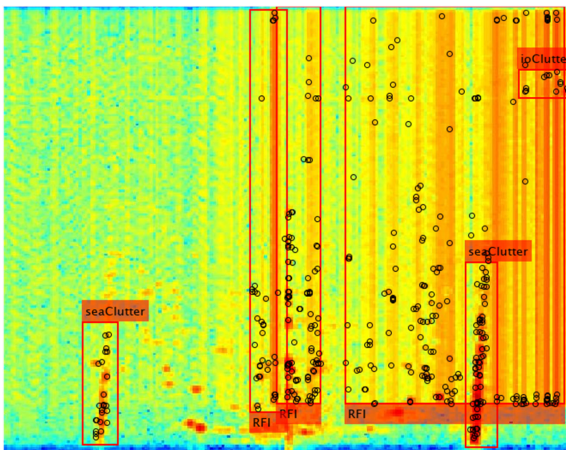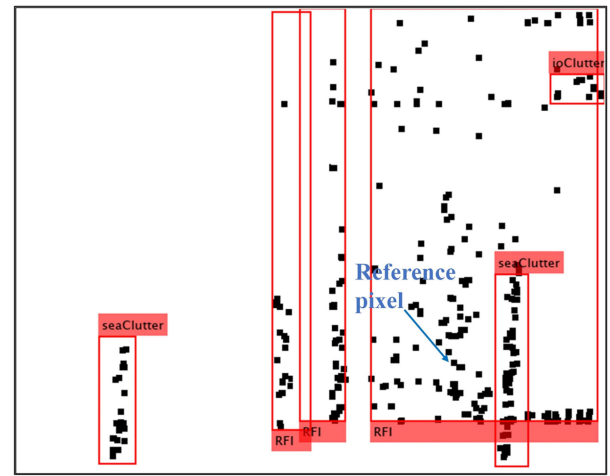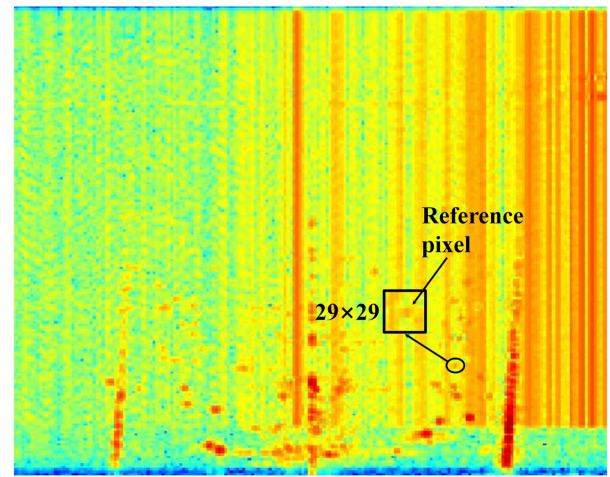Fig. 6. Target-detection rate and false-alarm rate.

Fig. 7. One example of the results of the first-level detector.

B. Feature Extraction

In this article, lightweight CNN and SAE are designed to automatically extract deep features of potential target positions by learning from different perspectives, to avoid the drawbacks

Fig. 8. Sliding window selection. (a) Potential target locations. (b) Original RD spectrum.

of manual design and preselected features. The RSLDA algorithm also aims to process the serial fusion feature to reduce redundancy in features while improving discriminatory ability.

1) Feature Extraction by CNN: In this article, CNN is chosen as a feature extractor to extract deep convolution features. CNN generally contains an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. In the convolutional layer, the input image is convolved with multiple kernels to generate multiple feature maps while preserving spatial information. The feature maps extracted by the convolutional layer are down-sampled by a pooling layer to reduce the dimensionality of the features and avoid overfitting. The convolutional and pooling layers make up the feature-extraction module. The fully connected layer, placed at the end of the network, can be considered the classifier of the network.

The potential target positions in the clutter and interference regions are obtained by the extremum detection algorithm. According to the potential target positions, we use the sliding window method to generate a certain size of sliding window for detection [7], [35]–[36]. As shown by the black regions in the red rectangular box in Fig. 8(a), each pixel in the black regions is used as a reference pixel to generate a sliding window of a certain

TABLE I
WHOLE CNN STRUCTURE

Layer	Layer Name	Specific Operation
1	imageInputLayer	$29 \times 29 \times 3$ images with zero-center normalization
2	Conv1	$8 \times 3 \times 3$ convolutions with stride [1 1] and Padding 'same'
3	Norm1	Batch normalization
4	ReLu1	ReLu
5	Maxpooling1	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
6	Conv2	$16 \times 3 \times 3$ convolutions with stride [1 1] and Padding 'same'
7	Norm2	Batch normalization
8	ReLu2	ReLu
9	Maxpooling2	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
10	Conv3	$32 \times 3 \times 3$ convolutions with stride [1 1] and Padding 'same'
11	Norm3	Batch normalization
12	ReLu3	ReLu
13	FC1	64 fully connected layers
14	ReLu4	ReLu
15	FC2	2 fully connected layers
16	Softmax	Softmax
17	Classification Output	Crossentropyex

TABLE II
PERFORMANCE COMPARISON OF ALEXNET AND THE LIGHTWEIGHT CNN

Network	Test Accuracy (%)	Times (s)
AlexNet	95.13	5.894
Our designed CNN	95.68	0.207

TABLE III
RADAR PARAMETERS

Parameters	Value
Transmit signal	FMICW
Operating frequency (MHz)	4.7
Sweep cycle (s)	0.125
Bandwidth (kHz)	30
Coherent integration time (s)	300
Antenna spacing (m)	14

TABLE IV
INFORMATION OF THE DATASET

Datasets	Number of window samples		Function
	Target	Non-target	
X_1	3500	3500	Pretrain the two feature-extraction networks
X_2	500	500	Evaluate the two feature-extraction networks
X_3	1000	1000	Train and evaluate the second-level detector ELM

size in the RD spectrum of the original RGB three-channel, which is sent to the CNN network to extract the CNN feature of that reference pixel.

Because the target points in the RD spectrum are usually tiny, it is difficult to get good convolution features with a small window. Statistically, almost all target sizes in our data are within the range of 15×15, combined with the criterion of the sliding window size selected for target detection in [7] and [35]. Therefore, we choose a 29×29 window centered on the reference pixel and twice the size of the target, as shown in Fig. 8(b), so that the target point is surrounded by some background information to let the CNN learn deep semantic features. At the same time, because the structure and parameters of CNN networks require extensive experience to debug, we here use a simple structure as much as possible while still ensuring the classification accuracy of the lightweight CNN networks. The CNN established in this article follows the classical AlexNet design idea and network structure, and the number of convolutional kernels increases with the depth of the network. To match the quantity of data and the number of classes while speeding up the network, the designed CNN network is shallower than that of AlexNet. The designed lightweight CNN structure includes three convolutional layers, two pooling layers, and two fully connected layers. The convolutional layers contain 8, 16, and 32 convolutional kernels respectively, each with a 3×3 kernel size. The two maximum pooling layers use a 2×2 window for down-sampling of the convolution features. Table I gives the structure of the whole designed CNN.

To verify the effectiveness of our designed lightweight CNN, we introduced AlexNet based on transfer learning, fine-tuned on our training set, and compared with our designed CNN on a validation set. The dataset used will be described in detail in Section V. Table II gives the comparison results of the two networks. The classification accuracy and running time of the

two networks are evaluated on the same dataset and performed ten times respectively, with the results averaged.

From the comparison results, it can be seen that the proposed lightweight CNN is not very different from AlexNet based on transfer learning in classification accuracy, but the proposed network is significantly faster than AlexNet with larger parameters. Therefore, the proposed lightweight CNN is more suitable for our radar dataset. Here, we use the trained CNN model as the feature-extraction module. Marmanis *et al.* [37] explained that the shallow layers of the network are generic features, and the higher layers are less generic and more specific. The features contained in the last layers of the network are more important in the classification compared to the features in the previous layers. Therefore, we extract deep semantic features from the first fully connected layer for the feature fusion. The steps for extracting CNN features from potential target positions using the CNN networks are as follows.

1) Train CNN with a dataset with labels.
2) Fix the parameters of the pretrained CNN and truncate the model. The input layer to the first fully connected layer (FC1) is used as a convolution feature-extraction module.
3) Input the 29×29 RGB window centered on the reference pixel to the convolution feature-extraction module to obtain the 64-dimensional CNN feature $\mathbf{F}_{CNN}$ of that reference point.

2) Feature Extraction by SAE: After designing a lightweight CNN to extract deep semantic features from suspicious target positions, we designed a self-learning-based SAE as the second lightweight feature-extraction network from different perspectives. An SAE is a stack of automatic encoders (AEs). Fig. 9

TABLE V
COMPARISON WITH DIFFERENT FEATURE ON THE TEST SET

Feature	Dimension	ELM		
		Test accuracy (%)	Hidden neurons	Time (s)
$\mathbf{F}_{SAE}$	60	95.12	75	0.0009
$\mathbf{F}_{CNN}$	64	96.34	75	0.0038
$(\mathbf{F}_{SAE}, \mathbf{F}_{CNN})$	124	98.19	70	0.0050
$(\mathbf{F}_{SAE}, \mathbf{F}_{CNN}) - PCA$	50	98.12	50	0.0055
Our method	50	98.34	50	0.0052

TABLE VI
COMPARISON WITH DIFFERENT FUSION FEATURE REPRESENTATION ON THE TEST SET

Feature	Dimension	ELM		
		Test accuracy (%)	Hidden neurons	Time (s)
$\mathbf{F}_{Alex}$	330	93.42	455	0.0203
$(\mathbf{F}_{SAE}, \mathbf{F}_{Alex}) - RSLDA$	100	96.61	150	0.0210
$(\mathbf{F}_{CNN}, \mathbf{F}_{Alex}) - RSLDA$	100	96.43	170	0.0253
$(\mathbf{F}_{SAE}, \mathbf{F}_{CNN}, \mathbf{F}_{Alex}) - RSLDA$	100	98.14	150	0.0260
Our method	50	98.34	50	0.0052

TABLE VII
EVALUATION OF THE FIRST-LEVEL DETECTOR

Scope of detection	Evaluation index	Only second-level detector ELM	Proposed two-level cascade detector
Sea clutter/ RFI regions	P_d (%)	91.3	91
	P_f (%)	21.5	18.8
	M (%)	8.7	9
	E (%)	30.2	27.8
	Time (s)	51.9	18.2

Fig. 10.   Structure of the SAE.

that the output value is as close as possible to the input value. The mappings learned in the encoder of AE can then be used to extract features from the raw data. Fig. 10 shows the structure of a SAE, in which the output of the former layer is used as the input of the latter layer and the output layer of the SAE is a Softmax classifier. The SAE trains each AE using a greedy method, in which the hidden-layer output of the previous AE is used as the input of the next AE, and so on, to pretrain the entire network. Then, the network parameters are fine-tuned backwards on the basis of the output error of the classifier.

To extract more efficient features from the potential target position, similarly to the CNN feature extracted above, the SAE feature is extracted within the 29×29 neighborhood of the reference pixel. The 29×29 sliding window is then grayed out and collapsed into an 841-dimensional column vector as the input of the SAE we designed.
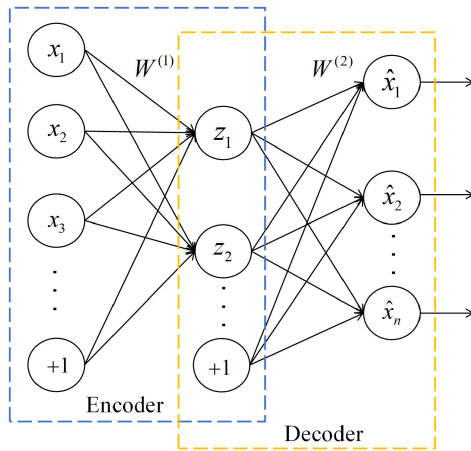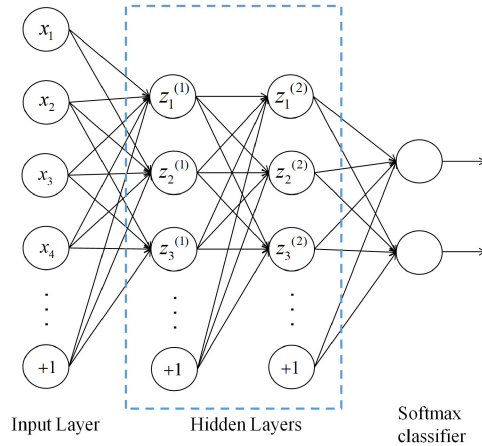

Fig. 9.   Structure of the AE.

shows the structure of an AE, consisting of an encoder and a decoder. The encoder maps the input signal x to the encoded signal z, and the decoder converts the encoded z to the output signal $\hat{x}$. AE trains and adjusts the weights of the network so

TABLE VIII
PERFORMANCE COMPARISON ON THE MEASURED RD SPECTRUM

Scope of detection	Evaluation index	Proposed algorithm	Modified CFAR	Wavelet-transform-based	OES-ELM-based
Sea clutter/ RFI regions	P_d (%)	91	85.6	86.7	82
	P_f (%)	18.8	41.3	31.8	63.8
	M (%)	9	14.4	13.3	18
	E (%)	27.8	55.7	45.1	81.8
	Time (s)	18.2	0.9	66	38.8

TABLE IX
DETECTION RATE RESULTS OF THE THREE ALGORITHMS

Scope of detection	Method	P_d (%)
Sea clutter/ RFI regions	Modified CFAR	87.23
	Wavelet-transform-based	82.60
	Proposed algorithm	93.25

TABLE X
DETECTION RATE RESULTS OF THE THREE ALGORITHMS FOR WEAK TARGETS

Scope of detection	Method	SCR (dB)		
		7.5	10	12.5
Sea clutter/ RFI regions	Modified CFAR (%)	3.47	20.13	56.38
	Wavelet-transform-based (%)	10.42	22.10	45.70
	Proposed algorithm (%)	16.67	67.53	80.40

Fig. 11. Influence of the number of neurons on classification accuracy.

1) Perform unsupervised layer-by-layer greedy training of all layers in the network, based on minimizing reconstruction errors.
2) *Train the Softmax Layer*: Use the labels of the training data to train the layer in a supervised manner.
3) Fine-tune the network by performing back propagation of the entire multilayer network on the basis of the training data in a supervised manner.

Based on the accuracy of the validation set, we chose the optimal hidden-layer neuron configuration. To guarantee accuracy, the training and testing were run ten times under each configuration and the results were averaged. Fig. 11 shows the test results for each hidden-layer neuron configuration on the validation set. It can be seen from the experimental results that the corresponding SAE achieved the best performance on the validation set when Num1 = 200 and Num2 = 60. Therefore, our final SAE structure has two hidden layers; the first hidden layer has 200 neurons and the second hidden layer has 60 neurons.

We chose the second hidden layer of SAE to extract features. The specific process is as follows.

1) Fix the parameters of the pretrained SAE and truncate the model. The input layer to the second hidden layer is used as the SAE feature-extraction module.
2) Gray out the 29×29 RGB window and collapse it into a column vector. Send it into the SAE feature-extraction module to obtain the 60-dimensional SAE feature $\mathbf{F}_{SAE}$ of that reference point.

SAE can extract deep features of the input data layer by layer under unsupervised learning, but it is pointed out in [28] that the more complex the network structure, as well as the more neurons in the network, the more likely the network is to be overfitted for a certain number of training samples. Kang *et al.* suggested reducing the number of SAE layers as much as possible under the premise of ensuring certain accuracy. Therefore, the SAE designed in this paper contains only two hidden layers.

In addition, since the number of hidden-layer neurons has a significant impact on network performance [38], we should properly configure the number of neurons in each hidden-layer in SAE to achieve better performance. We explore the impact of the hidden-layer neuron configuration on performance in the framework of two hidden layers. An experiment was performed to determine the number of neurons of each layer in SAE. We define the configuration of the hidden-layer neurons of the designed SAE as (Num1, Num2), where Num1 ∈ (100, 450) and Num2 ∈ (10, 100) refer to the number of the first and the second hidden-layer neurons, respectively. We will search for the optimal configuration in the range of Num1 and Num2. We set the step of Num1 to 25 and the step of Num2 to 10, so there are 150 hidden-layer configurations in total. We apply each configuration to construct the corresponding SAE, then train them on the same training set and test them on the same validation set. The specific training process is as follows.
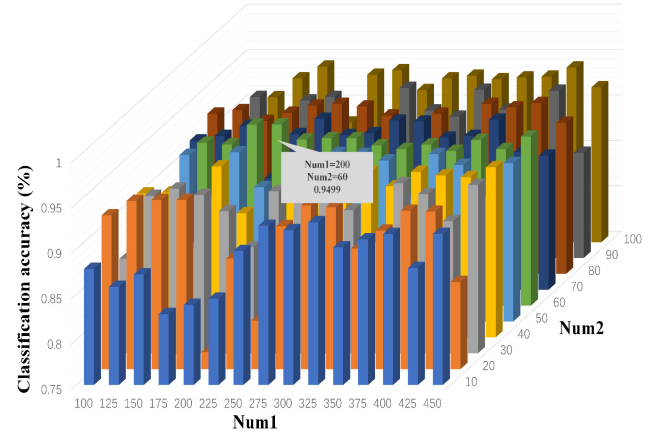
C. Multifeature Fusion

In this part, feature processing is performed by an RSLDA algorithm. The RSLDA algorithm can select and extract the most discriminating features with less discriminatory loss and is more robust to noise. Moreover, the algorithm is insensitive to the selection of feature dimensions. The RSLDA algorithm is briefly introduced below.

We define $X \in \mathbb{R}^{m \times n}$ as a dataset with n samples; each sample has m dimensions. The objective function of RSLDA can be written as follows:

$$\min_{P,D,E_r} \mathrm{Tr}(D^T(M_w - \varepsilon M_b)D + \sigma \|D\|_{2,1} + \lambda \|E_r\|_1)$$
$$\text{s.t.} \quad X = PD^T X + E, P^T P = I \tag{4}$$

where M_w and M_b are within-class and between-class scatter matrices, respectively, ε is a small positive parameter balancing the two scatter matrices, σ and λ are small positive trade-off parameters that determine the importance of the corresponding terms, E_r denotes the model's random noise, and $D \in R^{m \times d}(d < m)$ is the discriminant projection matrix. The constraint can be seen as a variant of principal component analysis (PCA) to retain energy. $P \in R^{m \times d}$ is an orthogonal reconstruction matrix.

According to the potential target positions detected by the extremum detector, a series of sliding windows is generated. CNN and SAE features are extracted from each window, and then feature fusion is performed in two steps. First, to avoid the impact of different scales, the two features are normalized separately and then concatenated to obtain the 124-dimensional serial fusion feature. Second, RSLDA is designed to further refine the obtained fusion feature, which reduces the redundancy of the feature while improving the discriminative ability of the fusion feature.

D. Second-Level Detector—ELM

After the fusion feature extraction in Section IV-C, we select the ELM to distinguish the true targets from among the suspected ones.

Supposing $R = \{(x^{(i)}, y^{(i)})|i = 1, 2, \ldots, q; x^{(i)} \in \mathbb{R}^n; y^{(i)} \in \mathbb{R}^2\}$ is a dataset with q samples, where $x^{(i)}$ is the n-dimensional fusion feature of the ith sliding window and $y^{(i)}$ is the desired output of the ith sliding window, representing whether it is the target or not, then the output of an ELM network with L hidden layer nodes is as follows:

$$\begin{bmatrix} g(w_1 x^{(1)} + b_1) & \cdots & g(w_L x^{(1)} + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 x^{(q)} + b_1) & \cdots & g(w_L x^{(q)} + b_L) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{bmatrix} = \begin{bmatrix} t^{(1)} \\ t^{(2)} \\ \vdots \\ t^{(q)} \end{bmatrix} \tag{5}$$

where $t^{(i)}$ is the actual output of the ith fusion feature, $g(\cdot)$ is the activation function, w_l is the input weight, b_l is the bias, and β_l is the output weight of the lth node. Equation (5) can then be simplified and written as follows:

$$H\beta = T \tag{6}$$

where H is the output of the hidden layer, $\beta = [\beta_1, \beta_2, \ldots, \beta_L]^T$, $T = [t^{(1)}, t^{(2)}, \ldots, t^{(q)}]^T$.

During the network training, the input weights and bias of the hidden layer are randomly determined. The aim of ELM
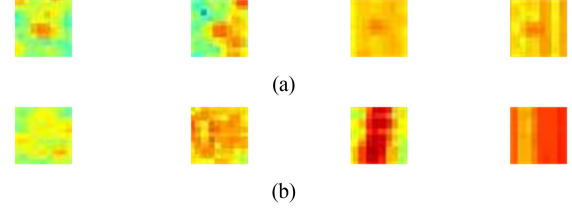

Fig. 12. Examples of target samples and nontarget samples. (a) Target samples, (b) Nontarget samples.

learning is to minimize the output error $\sum_{i=1}^{q} \|t^{(i)} - y^{(i)}\|$. Thus, the training procedure for the ELM is equivalent to solving the least-squares solution to (6), which can be expressed as follows:

$$\hat{\beta} = H^+ Y \tag{7}$$

where H^+ is the Moore-Penrose generalized inverse of H. Here, the fusion feature dataset is sent into the ELM for training, and the trained ELM can then be used as a second-level detector.

V. EXPERIMENT

Multiple experiments were performed to verify the effectiveness of the proposed cascade target-detection framework, especially for complex environments with strong clutter and interference. Our data come from the field HFSWR experiment conducted in Huanghai, China. The radar system parameters used are given in Table III. In Section V-A, we present the dataset used by the proposed algorithm for target detection; in Section V-B, we evaluate the discriminatory ability of the fusion feature on the test set; and in Section V-C, we perform tests on the actual measured RD spectrum and the RD spectrum of the simulated target points to evaluate the overall performance and target-detection performance of the proposed algorithm. The corresponding simulation experiments were performed using MATLAB 2019b with the hardware support of an Intel i5-9400 (2.90GHz) CPU and 16 GB memory.

A. Dataset for Target Detection

The proposed target-detection algorithm contains two feature-extraction networks to adaptively extract the features of the sliding window. We construct a dataset using measured RD spectra (875×656 pixels). The dataset consists of sliding windows of size 29×29, each centered on a reference pixel, and the labels of the dataset are "target" and "nontarget," where all targets are real target points. Because the proposed algorithm focuses on target detection in clutter and interference regions, most of the target points with positive labels are at the edge of, or inside, the clutter. The dataset contains 10 000 window samples, where the number of target samples is 5000 and the number of nontarget samples is 5000. Fig. 12 shows the examples of target samples and nontarget samples. We randomly divide the dataset, 7/10 of which is used as the training set X_1 to pretrain the two feature-extraction networks, 1/10 of which is used as the validation set X_2 to evaluate the two feature-extraction networks, and the remaining samples are used as the remaining set X_3 to train and evaluate the second-level detector ELM. Table IV gives the information of the dataset. When two feature-extraction

networks are trained and evaluated, for CNN, the window samples in X_1 and X_2 are directly used as inputs, while for SAE, the window samples in X_1 and X_2 need to be grayed out and collapsed into an 841-dimensional column vector as the input of the designed SAE. Fusion features are extracted from the X_3 in accordance with the scheme described in Section IV-C to train and evaluate the ELM-based target detector.

B. Discriminant Capacity Analysis of Fusion Feature

In this part, we verify the performance of the proposed two lightweight feature-extraction networks and the effectiveness of the proposed feature-fusion algorithm. The X_3 are divided into a training set and a test set; the ratio of training set to test set is 2:1. Our method here is the feature obtained through the feature-fusion scheme mentioned in Section IV-C. To evaluate the effectiveness of the proposed feature-fusion algorithm, a comparison has been made among the CNN feature $\mathbf{F}_{\text{CNN}}$, the SAE feature $\mathbf{F}_{\text{SAE}}$, the original serial fusion feature obtained after direct concatenation of the CNN and SAE features, and the feature obtained by processing the original serial fusion feature using PCA. The PCA algorithm is used to preserve 98% of the energy of the original serial fusion feature, in order to obtain a 50-dimensional fusion feature. For a fair comparison, we use the RSLDA algorithm to transform the original serial fusion feature to 50 dimensions.

Classification was performed using the ELM, and each group of experiments was performed 50 times and the final results averaged. Because the number of hidden-layer nodes has a large influence on the final classification performance, for each feature, multiple tests were performed to find the optimal number of hidden-layer nodes before the comparison experiment began. Table V gives the results for several features. The "time" here refers to the time consumed to extract features using different feature-extraction algorithms for a single sample. For a single feature, the discriminative ability of the CNN feature is slightly better than that of the SAE feature. This is due to the design of the CNN network, which uses multiple convolutional layers to extract more complex deep semantic features. Second, the original serial fusion feature has better performance than a single feature. This means that the extracted CNN feature and SAE feature are complementary, and the designed SAE can extract some information different from the deep convolution feature from various perspectives, which is a good way to describe the target point. After the PCA algorithm is used to process the serial fusion feature, there is little loss of accuracy, indicating that there is some redundancy affecting discriminatory ability. Furthermore, the fusion features obtained by the RSLDA algorithm achieve the best result, which indicates that the proposed feature-fusion strategy is effective.

We make use of t-distributed stochastic neighbor embedding [39] to visualize the original serial fusion feature, the fusion feature obtained using the PCA algorithm, and the fusion feature obtained by the RSLDA algorithm in Fig. 13. In comparison with the other two features, the fusion feature obtained using the RSLDA algorithm is clearly divided into two classes and is more aggregated within the classes.

In addition, to evaluate the effectiveness of the proposed two lightweight feature-extraction networks, we introduced the
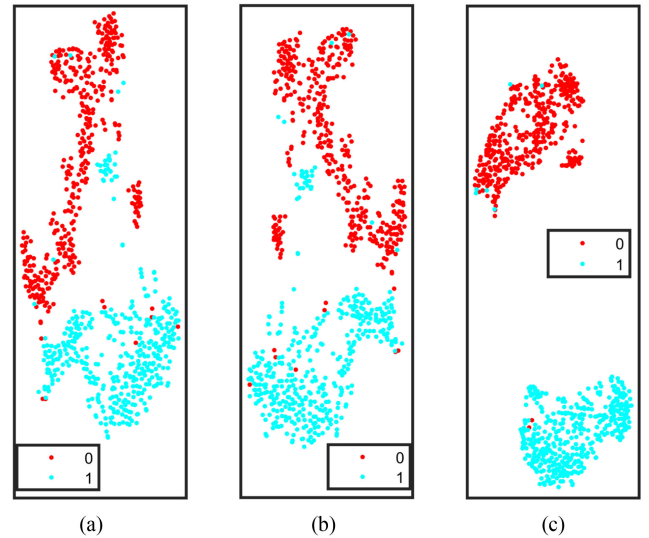

Fig. 13. Visualization of (a) the original serial fusion feature. (b) Fusion feature obtained using the PCA algorithm. (c) Fusion feature obtained using the RSLDA algorithm.

transfer-learning-based AlexNet network as the standard feature extractor [23]. We extracted deep features from the first fully connected layer. Since the extracted features are up to 4096 dimensions, we reduce the dimension of the final AlexNet feature $\mathbf{F}_{\text{Alex}}$ to 330 by the PCA algorithm while retaining 98% of the original feature energy. We used different fusion feature representation and evaluated them separately. The specific experimental setup is the same as that of Table V. Table VI gives that the transfer-learning-based AlexNet feature representation is much worse than the two lightweight feature extractors we designed specifically for radar target detection. However, feature fusion of $\mathbf{F}_{\text{Alex}}$ with $\mathbf{F}_{\text{SAE}}$ and $\mathbf{F}_{\text{CNN}}$, respectively, reveals that both classification results have different degrees of improvement over single features, showing that the fusion feature contains more information in characterization of the target. In addition, we find that the fusion of $\mathbf{F}_{\text{Alex}}$ with $\mathbf{F}_{\text{SAE}}$ outperforms the fusion with $\mathbf{F}_{\text{CNN}}$, which demonstrates that the designed SAE can extract more effective features, also given in Table V. The experimental results show that the three-feature-fusion scheme is a significant improvement compared to the single-feature. Moreover, in terms of detection time, due to the deeper layers and large number of parameters in AlexNet, the detection time for a single sample is 0.0203s and the three-feature-fusion scheme is even longer than 0.026s, while our algorithm takes only 0.0052s.

C. Detection Results for Vessels in the RD Spectrum

In the previous section, we briefly evaluated the proposed algorithm on a test set. Next, we verify the performance of the proposed algorithm on the RD spectrum. In this paper, target-detection rate P_d, false-alarm rate P_f, error rate E, and target-omission rate M are used as the evaluation indexes of detection performance. P_d and P_f are defined in (2) and (3); M and E are defined as follows:

$$M = 1 - P_d \tag{8}$$

$$E = P_f + M. \tag{9}$$

1) Evaluation of the First-Level Detector: We introduce a grayscale feature-based extremum detector in the first level of the two-level cascade detector to realize the rough extraction of suspicious targets and improve the efficiency of the algorithm. To verify the necessity of introducing the first-level detector, a comparison experiment was performed. We removed the first-level detector from the proposed two-level cascade detector and used only the ELM directly to detect targets in the clutter and interference regions. At this point, since the first-level detector is removed, we use each pixel in the clutter and interference regions as a reference pixel to generate the sliding windows, after which all processes are the same as those in the proposed algorithm.

We selected 30 RD spectra of known ship locations for detection, and all ship information was obtained through automatic identification system (AIS). We counted the detection results in clutter and interference regions identified during the preprocessing stage. A performance comparison with the proposed cascade detector in terms of time and accuracy is also provided. Table VII gives the detection performance of the two algorithms.

From the comparison results, we can see that the P_d of the two algorithms are basically the same, which indicates that the proposed first-level detector can extract all suspicious target points without missing targets. The P_f of the two-level cascade detector is lower than that of only the second-level detector ELM. This is because the proposed first-level detector can effectively filter some clutter windows misclassified by the second-level detector. In terms of detection time, the proposed algorithm is significantly more efficient than that of only the second-level detector ELM. This is because the proposed algorithm can roughly extract all the suspicious target points, which saves detection time by filtering out the invalid windows for the subsequent algorithm.

From the above comparison experiment, we come to a conclusion that the proposed two-level cascade target-detection algorithm can improve the detection efficiency and accuracy compared with using only the ELM network.

2) Target Detection by Field RD Spectrum: In this part, the effectiveness of the proposed detection framework is evaluated by comparison with current mainstream HFSWR target-detection algorithms: the modified CFAR algorithm; the wavelet-transform-based algorithm [11] proposed by Li *et al.*; and the OES-ELM-based algorithm [16] proposed by Zhang *et al*.

The modified CFAR algorithm is mainly divided into three steps. First, curve fitting is used to suppress clutter and background noise. Then, a two-dimensional CA-CFAR is performed in the Doppler direction and range direction, respectively. Finally, a global peak detection is performed, and an "AND" operation is performed on the results of the three detections to obtain the final detection results.

The wavelet-transform-based algorithm [11] uses a peak SNR-based algorithm to determine the scale of the wavelet transform, and then a fuzzy-set-based algorithm is used to enhance the high-frequency coefficients. Finally, reconstruct the high-frequency coefficients to suppress clutter and background noise for target detection.

The OES-ELM-based algorithm [16] uses a two-stage cascade detector for target detection. The first stage uses a linear classifier based on grayscale features to determine the location of suspicious target points. The second stage sends the extracted Haar-like features at suspicious target locations to the OES-ELM for real target classification.

According to the principal analysis, the modified CFAR algorithm and the wavelet-transform-based algorithm implement target detection after suppressing the clutter, while our algorithm and the OES-ELM-based algorithm perform the target detection directly on the RD spectrum. We selected 30 RD spectra of known ship locations for detection, and all ship information was obtained through AIS. Table VIII compares the performances of the four algorithms. We count the detection results in clutter and interference regions identified during the preprocessing stage. In detection time, the modified CFAR algorithm, wavelet-transform-based algorithm, and OES-ELM-based algorithm count the time from the input RD spectrum to the output detection results. Our proposed algorithm counts the test time of the two-level cascade detector.

Table VIII gives that the proposed algorithm has better detection performance than the other three. The proposed algorithm has the lowest false-alarm rate and the highest detection rate. This indicates that the extracted fusion features can better characterize the target points and are more robust. Although the wavelet-transform-based algorithm can remove sea clutter, ground clutter, ionospheric clutter, and RFI regions and retain only the high-frequency components in the RD spectrum, its detection performance is poor due to the variable morphology and energy of the clutter regions. In particular, the OES-ELM-based target detection algorithm extracts Haar-like features of the potential target regions and classifies the real target using the OES-ELM, which can detect targets in the smooth region of the RD spectrum. However, in our experiments, the algorithm performed poorly in the clutter and interference regions, with a high false-alarm rate, indicating that the simple central Haar-like features alone are not sufficient for detection in clutter and interference regions. Due to the complexity of the environment in the clutter and interference regions, morphological features cannot effectively distinguish targets from clutter, while the proposed deep feature representation can better distinguish between targets and clutter.

For the same RD spectrum, the modified CFAR algorithm had the fastest average testing time, and the wavelet-transform-based algorithm was the slowest. The proposed algorithm is faster on average than the OES-ELM-based algorithm. The main reason is that the first-level classifier of the OES-ELM-based algorithm uses a linear classifier based on grayscale values to obtain potential target regions. This treats almost all clutter and interference regions as potential target regions and consumes a lot of time. However, the proposed algorithm is more efficient by using a simple extremum detector to extract potential targets.

Fig. 14 shows the detection results for two actual ship targets for 15 min. The RD spectrum of the current moment and the distance and velocity of the two ship targets are acquired every 3 min. Fig. 14(a) shows the RD spectra of the sea clutter, including the negative Doppler-frequency shift at the moment of acquisition. The two specific ship targets, t1 and t2, are provided by the AIS installed on the vessels. The horizontal coordinate is the target speed in kilometers per hour, and the vertical coordinate
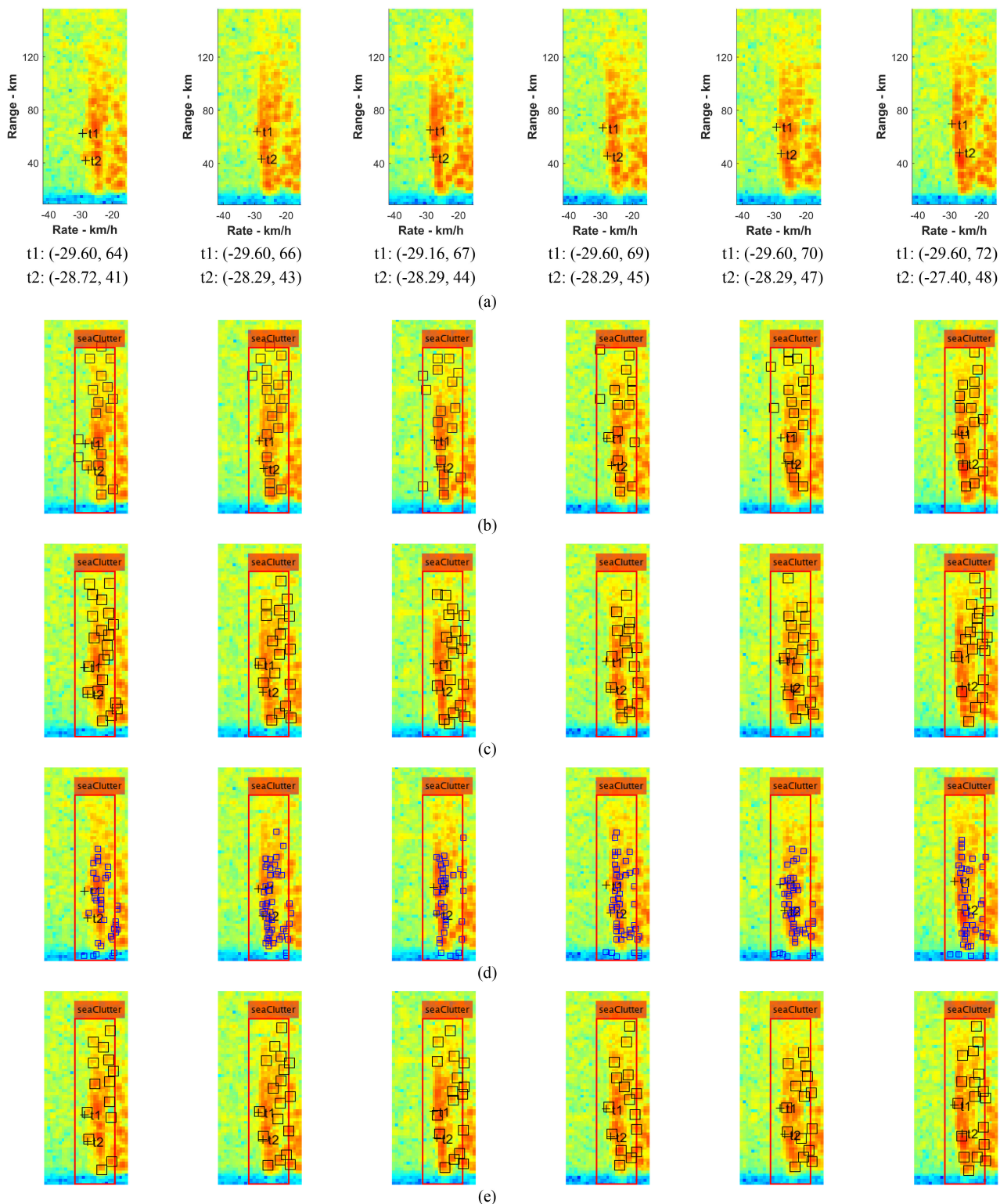
Fig. 14.   (a) Partial RD spectrum containing two typical targets. (b) Detection results of modified CFAR algorithm. (c) Detection results of wavelet-transform-based algorithm. (d) Detection results of OES-ELM-based algorithm. (e) Detection results of proposed algorithm.

is the target range. Fig. 14(b)–(e) shows the detection results of the modified CFAR algorithm, the wavelet-transform- based algorithm, the OES-ELM-based algorithm, and the proposed algorithm at the corresponding moments. It can be seen that the two selected target points are affected by the clutter at most of the moments, and are even partially submerged by the clutter for a time, but the proposed algorithm can detect both consistently. The other three algorithms have obvious failings. Moreover, the OES-ELM-based algorithm cannot work properly in these circumstances and has a very high false-alarm rate. Hence, in
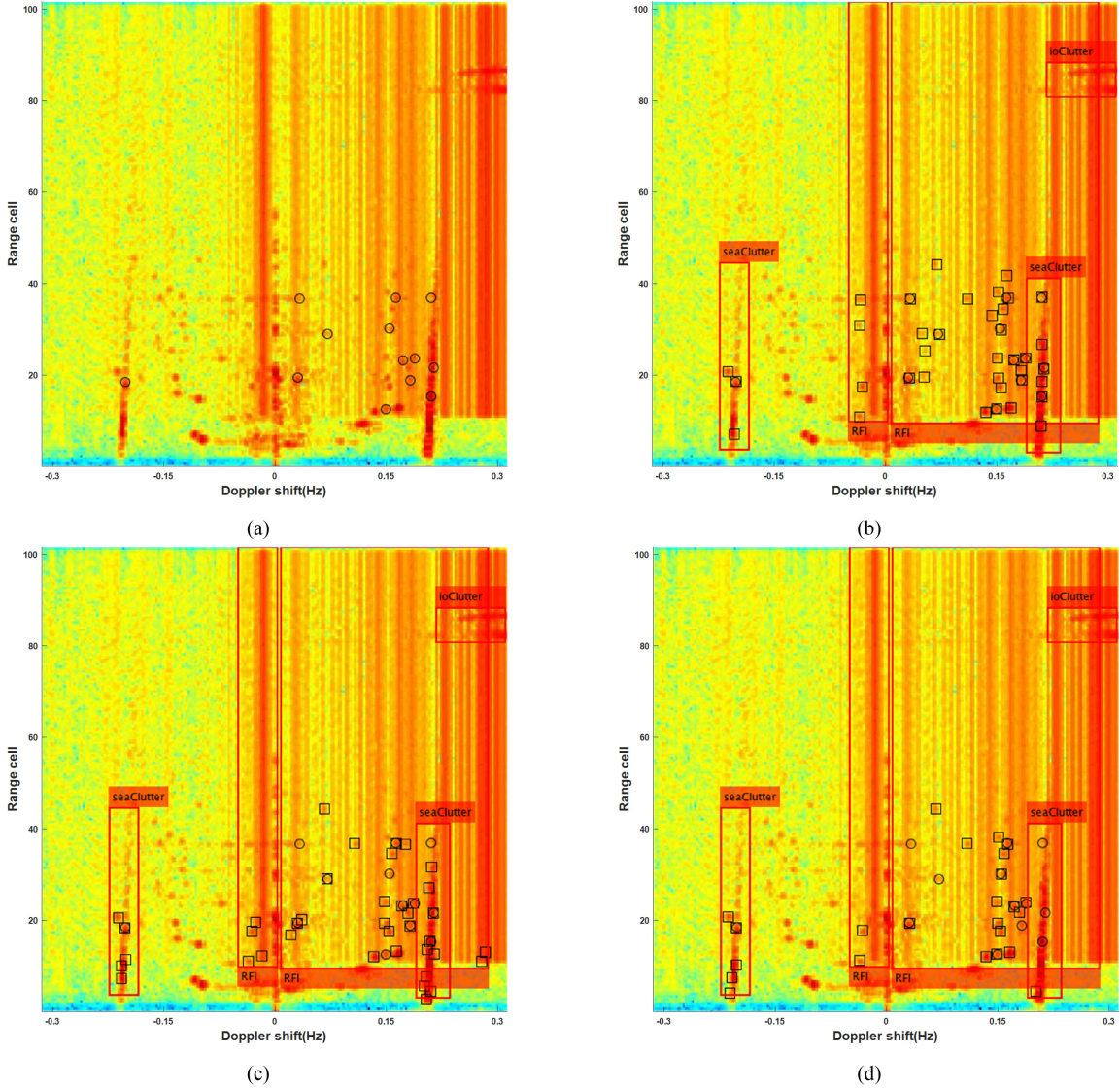
Fig. 15.   (a) Original RD spectrum. (b) Detection results of proposed algorithm. (c) Detection results of modified CFAR algorithm. (d) Detection results of wavelet-transform-based algorithm.

subsequent experiments, the OES-ELM-based algorithm will not be used as a comparison algorithm.

3) Comparison of Target Detection Performance: For HF-SWR target detection, the detection rate P_d is more important than the false-alarm rate P_f, as false targets can be removed by subsequent schemes such as multi-frame correlation. Therefore, in this part, target-detection experiments were performed in the clutter and interference regions to evaluate the performance of the proposed algorithm. On the one hand, to completely measure the P_d of each algorithm, a large number of RD spectra are required. On the other hand, as the number and location of artificially added simulated target points are determined, and their signal-to-clutter ratio (SCR) can be set artificially, it is easy for us to precisely perform comparison experiments. For the above considerations, we added simulated target points in the clutter and interference regions of the RD spectrum to obtain a large number of experimental samples to evaluate the P_d of

each algorithm. Table IX gives the detection results of the three algorithms for 60 RD spectra with simulated target points added. The detection rate is defined as the ratio of the number of detected targets to the total number of targets.

Table IX gives the algorithm proposed in this article has better detection performance than the other two in the interference and clutter regions. The proposed algorithm has the highest target detection rate P_d, which is attributed to the use of two deep networks to extract features, and the fusion feature better characterizes the target point. The detection rates of the wavelet-transform-based algorithm and the modified CFAR algorithm are not much different. Fig. 15(a) shows a typical RD spectrum with randomly added simulated target points, which contains a variety of clutter and a large amount of RFI. Fig. 15(b)–(d) shows the detection results of the proposed algorithm, the modified CFAR algorithm, and the wavelet-transform-based algorithm, respectively. The circles mark the simulated targets, and the
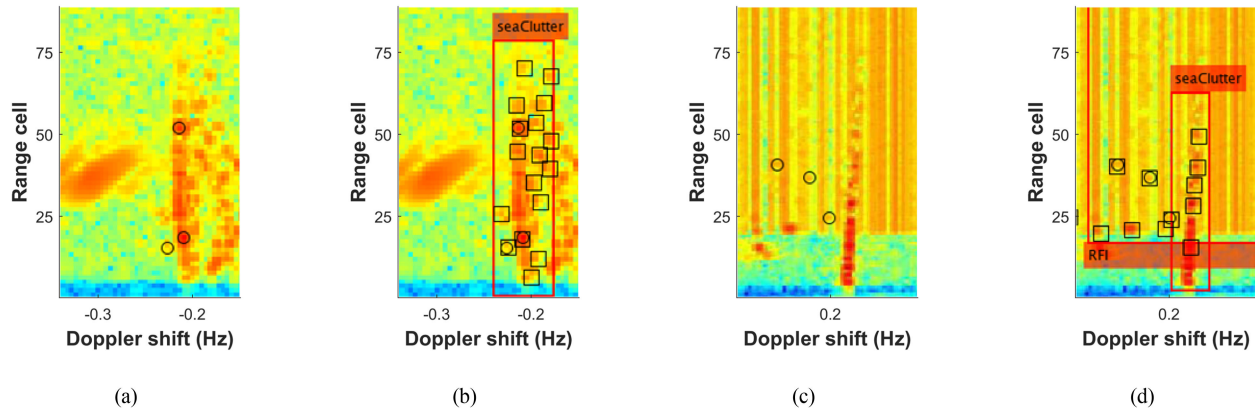
Fig. 16. Detection of weak targets with SCR of 12.5dB. (a), (c) Original RD spectrum. (b), (d) Detection results of proposed algorithm.

squares mark the targets detected by each algorithm. The proposed algorithm can accurately detect all the added targets in the clutter/interference regions, but the other two miss a lot of them. To further evaluate the performance of the proposed algorithm in detecting weak targets affected by clutter and interference, we added lower-energy targets in the clutter and interference regions. Here, the energy of the target is defined as the SCR, as follows:

$$\text{SCR} = 10 \cdot \log_{10}\left(\frac{P_s}{P_{ac}}\right) \tag{10}$$

where P_s is the amplitude of the target point and P_{ac} is the average amplitude of the clutter around the target point. The SCR of the added targets ranges from 7.5 to 12.5 dB. Three sets of tests are performed, each of which is a statistical result of 45 RD spectra. Table X compares the results of the three algorithms. When the SCR of the simulated target point is less than 10 dB, the morphology of the target point is almost invisible. However, because the algorithm in this paper uses two deep learning networks to extract features, which has better robustness, it still works better than the other two algorithms in the low-SCR case. Fig. 16 shows two typical detection results when the simulated target energy is 12.5 dB. The circles in Fig. 16 mark the location of the simulated target points, showing that the proposed algorithm can effectively detect the edge of the clutter and interference as well as the submerged target point.

VI. CONCLUSION

We propose a new intelligent algorithm for target detection in clutter and interference regions for HFSWR, based on a cascade deep learning network combined with ELM and feature fusion. The clutter/interference region in the RD spectrum is quickly identified by faster R-CNN. Then, a two-level cascade detector is proposed to effectively detect real target points in the clutter and interference regions. In the proposed two-level cascade detector, the extremum detector, as the first-level detector, is used for extraction of suspicious target positions. To overcome the disadvantages of conventional feature-extraction operators, which are difficult to design and have poor robustness in clutter and interference regions, two trained networks are also designed to extract the deep features of potential target positions from different perspectives. Then, the features are fused and sent to the ELM

for accurate classification. The effectiveness of the fusion feature was verified in the feature-comparison experiment. Furthermore, to show the detection performance of the proposed algorithm in the clutter/interference regions, we compared it with the current mainstream detection algorithms. The experiments for actual vessel targets show that the proposed algorithm has better detection performance than the conventional algorithms. The proposed algorithm is optimal in both target-detection rate and false-alarm rate. In the target-detection experiments, it has the highest detection rate for weak targets with low SCR concealed in the clutter and interference regions. Hence, the proposed algorithm is usable for the intelligent detection of HFSWR vessels affected by strong clutter and interference.

REFERENCES

[1] A. M. Ponsford and J. Wang, "A review of high frequency surface wave radar for detection and tracking of ships," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 18, no. 3, pp. 409–428, 2010.

[2] Y. Wang, J. Zhang, Y. Ji, and X. Mao, "Progress in research on HFSWR target detection under a background of sea clutter," *Marine Sci.*, vol. 40, no. 9, pp. 140–144, 2016.

[3] H. Rohling, "Ordered statistic CFAR technique-an overview," in *Proc. Int. Radar Symp.*, 2011, pp. 631–638.

[4] H. M. Finn and R. S. Johnson, "Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates," *RCA Rev.*, vol. 29, no. 3, pp. 414–464, 1968

[5] Y. He and H. Rohling, "A new CFAR processor based on ordered statistoc," in *Proc. Syst. Eng. Electron.*, 1994, pp. 17–23.

[6] A. L. Dzvonkovskaya and H. Rohling, "Target detection with adaptive power regression thresholding for HF radar," in *Proc. Radar Conf.*, 2006, pp. 1–4.

[7] L. Zhang, Q. Li, and Q. M. Jonathan Wu, "Target detection for HFSWR based on an S³D algorithm," *IEEE Access*, vol. 11, pp. 224825–224836, 2020.

[8] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship classification in SAR images using a new hybrid CNN–MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, pp. 551–562, 2019.

[9] F. Jangal, S. Saillant, and M. Helier, "Wavelets: A versatile tool for the high frequency surface wave radar," in *Proc. IEEE Radar. Conf.*, 2007, pp. 497–502.

[10] F. Jangal, S. Saillant, and M. Helier, "Wavelet contribution to remote sensing of the sea and target detection for a high-frequency surface wave radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 552–556, Jul. 2008.

[11] Q. Li, W. Zhang, M. Li, J. Niu, and Q. M. Jonathan Wu, "Automatic detection of ship targets based on wavelet transform for HF surface wavelet radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 714–718, May 2017.

[12] S. Grosdidier and A. Baussard, "Ship detection based on morphological component analysis of high-frequency surface wave radar images," *IET Radar Sonar Navigation,* vol. 6, no. 9, pp. 813–821, 2012.

[13] L. Zhang, L. Zeng, M. Li, and H. Wang, "Weak target detection based on complex duffing oscillator for HFSWR," in *Proc. 35th Chin. Control Conf.*, 2016, pp. 4982–4987.

[14] J. Cai, H. Zhou, W. Huang, and B. Wen, "Ship detection and direction finding based on time-frequency analysis for compact HF radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 72–76, Jan. 2020.

[15] W. Zhang, Q. Li, Q. M. Jonathan Wu, Y. Yang, and M. Li, "A novel ship target detection algorithm based on error self-adjustment extreme learning machine and cascade classifier," *Cogn. Comput.*, vol. 11, no. 1, pp. 110–124, 2019.

[16] W. Zhang, Q. Li, Q. M. J. Wu, and M. Li, "Sea surface target detection for RD images of HFSWR based on optimized error self-adjustment extreme learning machine," *Acta Automatica Sinica*, vol. 47, no. 1, pp. 108–120, 2019.

[17] Q. Li, and Y. Han, "Sea surface target detection algorithm for RD images of HFSWR," *Modern Radar*, vol. 42, no. 8, pp. 6–12, 2020.

[18] Z. Jin, Q. Pan, Y. Liang, Y. Cheng, and W. Zhou, "SVM-based land/sea clutter identification with multi-features," in *Proc. 31th Chin. Control Conf.*, 2012, pp. 3903–3908.

[19] Y. Li, W. Zeng, N. Zhang, and W. Tang, "Gabor feature based ionospheric clutter region extraction in range-Doppler map," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2014, pp. 269–270.

[20] Y. Li, M. He, and N. Zhang, "An ionospheric clutter recognition method based on machine learning," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2017, pp. 1637–1638.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[22] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 512–519.

[23] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, and E. Tzeng, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.

[24] Y. Bengio, *Learning Deep Architectures for AI*. Delft, The Netherlands: Now Publishers, 2009.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[27] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[28] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, 2017, Art. no. 192.

[29] Y. Wang, B. Song, P. Zhang, N. Xin, and G. Cao, "A fast feature fusion algorithm in image classification for cyber physical systems," *IEEE Access*, vol. 5, pp. 9089–9098, 2017.

[30] L. Gao, X. Zhang, J. Gao, and S. You, "Fusion image based radar signal feature extraction and modulation recognition," *IEEE Access*, vol. 7, pp. 13135–13148, 2019.

[31] J. Wen *et al.*, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.

[32] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[33] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1/3, pp. 489–501, 2006.

[34] L. Zhang, W. You, Q. M. J. Wu, S. Qi, and Y. Ji, "Deep learning-based automatic clutter/interference detection for HFSWR," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1517.

[35] Y. Xie, J. Tang, and L. Wang, "Radar target detection using convolutional neutral network in clutter," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2019, pp. 1–6.

[36] L. Wang, J. Tang, and Q. Liao, "A study on radar target detection based on deep neural networks," *IEEE Sensors Lett.*, vol. 3, no. 3, Mar. 2019, Art. no. 7000504.

[37] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using imagenet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.

[38] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, Nov. 2015.

[39] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, 2014.

Maokai Wu was born in Qingdao, China, in 1997. He received the B.S. degree in automation from the University of Jinan, Jinan, China.

He is with the College of Engineering, Ocean University of China, Qingdao, China, majoring in control engineering. His research interests include high-frequency surface wave radar and applications, and machine learning.

Ling Zhang received the B.S. degree in automation and the M.S. degree in signal and information processing from Ocean University of China, Qingdao, China, in 2000 and 2003, respectively, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2007.

She is currently an Associate Professor with the College of Engineering, Ocean University of China. Her research interests include intelligent information processing, communication signal processing, radar signal processing and machine learning.
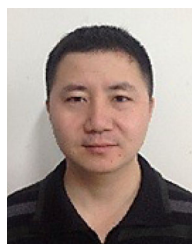
Jiong Niu received the M.S. degree in control theory and control engineering from the Ocean University of China, Qingdao, China, in 2008.

He is currently an Engineer with the College of Control Theory and Control Engineering, Ocean University of China, Qingdao, China. His research interests include ocean remote sensing using high-frequency surface wave radar, Array signal processing and intelligent instrument and meter.

Q. M. Jonathan Wu (Senior Member, IEEE) received the M. S. degree from Coventry Polytechnic and the Ph.D. degree from the University of Wales, U.K.

He is a Professor of Electrical and Computer Engineering and a Tier 1 Canada Research Chair in Automotive Sensors and Information Systems since 2005. He is the Founding Director of the Computer Vision and Sensing Systems Laboratory, University of Windsor, Windsor, ON, Canada. Prior to joining the University, he was a Senior Research Officer and a Group Leader with the National Research Council of Canada. He has authored or co-authored one book in the area of three-dimensional vision and more than 350 peer-reviewed papers (including 200 journal publications) in areas of computer vision, multimedia information processing, and intelligent systems

Dr. Wu is an Associate Editor for IEEE TRANSACTION ON CYBERNETICS, IEEE TRANSACTION CSVT, NEUROCOMPUTING, and *Cognitive Computation*. He has also served on editorial board for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTION SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and *International Journal of Robotics and Automation*. He is a fellow of the Canadian Academy of Engineering.