Cognitive Computation (2019) 11:110–124
https://doi.org/10.1007/s12559-018-9606-5

A Novel Ship Target Detection Algorithm Based on Error Self-adjustment Extreme Learning Machine and Cascade Classifier

Wandong Zhang[1,2] · Qingzhong Li[1] · Q. M. Jonathan Wu[2] · Yimin Yang[3] · Ming Li[1]

Received: 30 June 2018 / Accepted: 10 October 2018 / Published online: 24 October 2018

Abstract

High-frequency surface wave radar (HFSWR) has a vital civilian and military significance for ship target detection and tracking because of its wide visual field and large sea area coverage. However, most of the existing ship target detection methods of HFSWR face two main difficulties: (1) the received radar signals are strongly polluted by clutter and noises, and (2) it is difficult to detect ship targets in real-time due to high computational complexity. This paper presents a ship target detection algorithm to overcome the problems above by using a two-stage cascade classification structure. Firstly, to quickly obtain the target candidate regions, a simple gray-scale feature and a linear classifier were applied. Then, a new error self-adjustment extreme learning machine (ES-ELM) with Haar-like input features was adopted to further identify the target precisely in each candidate region. The proposed ES-ELM includes two parts: initialization part and updating part. In the former stage, the L_1 regularizer process is adopted to find the sparse solution of output weights, to prune the useless neural nodes and to obtain the optimal number of hidden neurons. Also, to ensure an excellent generalization performance by the network, in the latter stage, the parameters of hidden layer are updated through several iterations using L_2 regularizer process with pulled back error matrix. This process yields appropriate output weights and the appropriate hidden weights. Experimental results show that (1) compared with standard ELM, our proposed ES-ELM has higher classification accuracy and training efficiency, and the generalization performance is not sensitive to regularization parameter, (2) the proposed ship target detection algorithm based on ES-ELM outperforms most of the state-of-the-art methods for detection accuracy and computational efficiency.

Keywords Extreme learning machine (ELM) · High-frequency surface wave radar (HFSWR) · Target detection · Range-Doppler (RD) image · Cascade classifier

Introduction

The high-frequency surface wave radar (HFSWR), which is based on electromagnetic wave propagation along the salty ocean surface due to wave's conductivity, provides a unique capability to observe far beyond the conventional microwave radar coverage [1, 2]. This property is most useful in many applications, such as ship detection and tracking objects that are beyond the line-of-sight. Consequently, HFSWR systems recently became an operational tool for coastal monitoring worldwide. However, as compared to the detection environment of other sensors, that of HFSWR is more complex because various clutters, such as sea clutter, ionospheric clutter, radio frequency interference, and background noises often appear together in one frame. Therefore, maritime target detection by HFSWR has become a challenging task.

Much work has been done on HFSWR ship target detection, which can be classified into two categories: constant false alarm rate (CFAR)-based methods and range Doppler (RD) image processing-based methods. For the CFAR-based methods, Hinz and Holters et al [3] and Liu and Lampropoulos [4] used the CFAR-based method to extract ships relying on the signal-to-noise ratio (SNR). Rohling [5] proposed an algorithm, which applied ordered statistic

✉ Qingzhong Li
liqingzhong@ouc.edu.cn

1 Department of Engineering, Ocean University of China, Qingdao, Shandong, China

2 Department of Electronic and Computer Engineering, University of Windsor, Windosr, ON, Canada

3 Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada

method to find the adaptive threshold, and it can be adopted in the cases where more than one target is presented in the reference window. In 2012, Gui [6] introduced a new two-dimensional detector, which can process the radar signal both in Doppler shift direction and range direction. Based on [6], Liang [7] utilized curve fitting as the preprocessing step to both weaken clutter and removed background noise, which significantly improved the detection results. However, these improved one-dimensional and two-dimensional CFAR-based methods still face challenges in radar applications due to the strong influences of clutter and noise.

For the RD image processing-based methods, Grosdidier and Baussard [8] proposed a ship target detection method based on morphological component analysis. In addition, Jangal et al [9, 10] analyzed the differences in geometrical feature and intensity distribution to separate targets from the background, and proposed a target detection approach by using one-dimensional wavelet transform. Based on Jangal's method [9], Li et al [11] proposed an automatic target detection method based on discrete wavelet transform (DWT), where the optimal scale of DWT can be automatically determined to separate ship targets from the others. In brief, the RD image processing-based methods are promising for weak targets detection with low SNR background, and can achieve better detection performance than CFAR-based methods. However, such methods concentrate too much on detection performance, thereby limiting their real-time practical applications due to the complexity of the algorithm.

In general, the shortcomings for these two categories are twofold: poor real-time performance and low detection accuracy. The goal of this paper is to develop a machine learning-based method that can provide both high detection performance and high real-time performance. To ensure the algorithm has better ship detection accuracy, a new error self-adjustment extreme learning machine (ES-ELM) network that acts as a refined classifier is applied to identify weak targets precisely in high-clutter harsh background. In addition, to improve the real-time detection performance, a two-stage cascade classification structure including linear classifier and ES-ELM network is proposed. To be more precise, the cascade strategy refines the roughly detected, which means that the linear classifier can extract candidate targets with high efficiency and the ship targets can be refined by ES-ELM network.

As is known to all, the machine learning methods have been adopted in many research areas [41, 42], and extreme learning machine proposed by Huang et al [12–16] is a famous machine learning-based artificial neural network, and has been adopted in subspace learning [17], feature learning and classification [18], online data learning [19], etc. Although the ELM has many advantages, the following two issues mainly limit its practical applications. (1) The first one is how to determine the minimum hidden nodes required without impacting the learning effectiveness. For this purpose, many approaches, such as enhanced incremental ELM (EI-ELM) [20], convex incremental ELM (CI-ELM) [21], optimal pruned ELM (OP-ELM) [22], and error-minimized ELM (EM-ELM) [23] were proposed. Nevertheless, these ELMs are usually time-consuming, since they find the suitable hidden neurons with several optimization processes. (2) The second one is how to select the proper regularization parameter C in ELM training. Although there are several selection methods [27, 28] that can be employed, whether the determined C is optimal remains unknown, so the trial-and-error method is usually adopted by most users.

Based on the above analysis, the contribution of this paper can be summarized as follows.

(1) A new network named ES-ELM is proposed. In the proposed ES-ELM, the L_1 regularizer process is applied to find the sparse solution of output weights, so that the optimal number of hidden neurons can be obtained. In addition, the hidden parameters of hidden layer are updated by several iterations with L_2 regularizer process and by pulled back error matrix, thereby obtaining not only the appropriate output weights but also the appropriate hidden weights. Moreover, the regularized parameter C in the structure of ES-ELM is not sensitive to generalization performance.
(2) A novel RD image ship target detection method is presented. The linear classifier is applied as the first-stage classifier, with which the prediction of this network is ambiguous, but shows higher classification efficiency, which means that this network is easy to design and can remove useless patterns with great speed. More importantly, the second-stage classifier in the proposed ES-ELM can extract ship targets from the ambiguous result precisely, so that the proposed method gives acceptable detection performance.

The rest of this paper is organized as follows. In the "Background" section, we have a brief review of ELM and provide a brief description of our problems. In the "Framework of the Proposed Ship Target Detection Algorithm" section, we have a diagram of the proposed algorithm. The first stage of the cascade classifier for ship target detection is described in "Linear Classifier." The Haar-like descriptor is shown in "Feature Extraction by Haar-Like Descriptor." The detail of the proposed ES-ELM structure is presented in "Proposed Error Self-adjustment Extreme Learning Machine Classifier." The "Experimental Verification" introduces the experimental results of the proposed ES-ELM and the proposed ship

detection method, respectively. Finally, the conclusion is given in "Conclusion."

Background

Review of ELM

ELM developed by Huang et al [12–15, 36] is a kind of SLFN, where only the output parameters are required to be trained while no training is required for the hidden layer parameters. In [37] and [38], Bartlett mentioned that the classifier has much smaller squared error on classification problems if the trained structure has smaller weights dimension. The attempt to keep the dimension of weight small leads to better generalization performance, and these theories correspond with ELM structure. Therefore, the goal of ELM is not only to reach the minimum training error, but also ensure the smallest norm of output weights.

$$\text{Minimize} : ||\boldsymbol{\beta}||^2 + C||\mathbf{H}\boldsymbol{\beta} - T||^2 \quad (1)$$

where H is the feature mapping matrix of hidden layer, $\boldsymbol{\beta}$ is the parameters of output layer. T is the expected training output, and C is a positive value, that is, regularized parameter.

Review of Sub-network ELM

Recently, a more interesting structure in which a hidden node itself can be substituted by a sub-network has been introduced. In that structure, each sub-network is constituted by several hidden nodes, and the sub-networks are added to neural network one-by-one based on the pulled back error matrix [24, 25]. The parameters for added sub-network node is calculated by least square method.

$$a_n = e_{n-1} \cdot x^T (C/I + \mathrm{x}\mathrm{x}^T)^{-1} \quad (2)$$

where x is the input data, a_n is the weight of new sub-network, e_{n-1} is the residual error of current network, $x^T(C/I + \mathrm{x}\mathrm{x}^T)^{-1}$ is the Moore-Penrose generalization inverse of the training samples.

Experiments showed that the learning speed of this method can be hundreds of times faster than the other ELMs, and the generalization performance better than that of the other learning methods that have hundreds of hidden nodes. In addition, this method has been explored further to construct the more complex stacked auto-encoder [26] and multi-layer network in feature representation [39].

In its structure, the number of the hidden neurons in each sub-network can be flexible and experiments indicate that sub-network ELM with m(m is equal to output dimension) hidden nodes give significantly better result in most data sets.

Problem Statement

The general ship target detection scenario of a HFSWR is illustrated as Fig. 1a. A typical HFSWR generated RD image is shown in Fig. 1b, which represents the power spectral density of the received signal according to the radar range and the Doppler frequency shift [1]. As seen from Fig. 1a, ship targets can be seen as isolated vessels moving in a predetermined direction, while the clutter has various returns decided by the spatial distribution of the sea waves which generate resonance scattering [2]. The receiving array detectors of a shore-based HFSWR system formed by omnidirectional antennae are a complicated device with many inductors and capacitors. These antennas line up in an array with several basic elements and are spaced

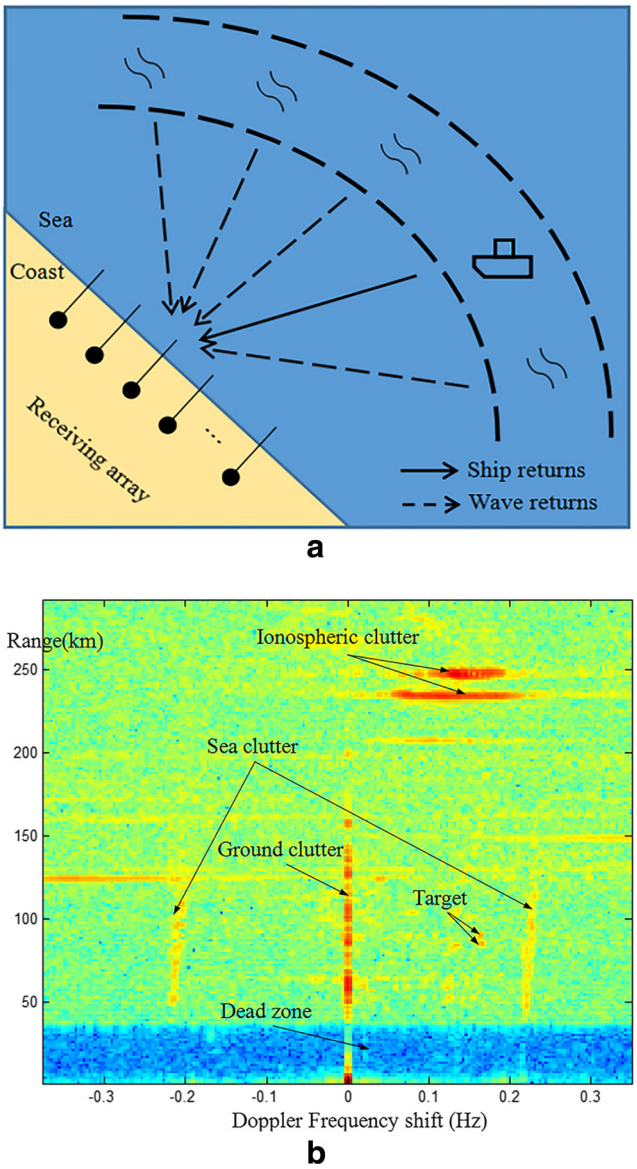

Fig. 1 **a** Sketch of detection scenario; **b** RD image for azimuth index 5

at homogeneity interval. In addition, the perpendicular direction to the receiving array is regarded as the azimuth reference.

The Doppler frequency shift, which describes the difference between the original transmitted frequency and the final received frequency, is a direct indication of the object's speed along the radial direction. Figure 1b is an RD image for azimuth index 5. We know that an RD image of a HFSWR generally contains not only the ship targets but also the interference elements, such as land clutter, sea clutter, ionospheric clutter (meteor clutter), and background noises. Since the ship target's radar signal has backscattering characteristic, the signature of targets in the RD image present as a kind of slightly extended point peak. On the contrary, the land clutter and sea clutter exhibit banded regions along the range direction, and the ionospheric clutter is ridged regions in the direction of Doppler frequency shift.

Framework of the Proposed Ship Target Detection Algorithm

Our goal is to extract and identify the ship targets from a complex background. Considering that ship targets and clutters have different morphological characteristics. In this paper, a machine learning-based method is proposed for fast detection of ship targets. All the original RD images are converted into gray images linearly, which are transformed from RD power level ranging from Pmin dB to Pmax dB to gray level from 0 to 255. The block diagram of the proposed framework is shown as Fig. 2.

As can be seen from the diagram, our proposed algorithm consists of two core models: a first-stage classifier and a second-stage classifier. The former is a linear classifier with one-dimensional gray-scale feature. The purpose of this classifier is to quickly remove the background area, to precisely maintain the highlighted ship targets area (non-background area) in an RD image. In other words, this linear classifier improves the real-time performance of the proposed RD image-based ship targets detection method to an acceptable level. The latter is an elaborate classifier, i.e., ES-ELM network. Its input feature is a 49-dimensional feature vector computed by Haar-like descriptor, and in its output, it adds labels to indicate whether this part is a ship target or not. The objective while designing ES-ELM network was to ensure our method detects ship targets accurately and reliably in the highlighted ship target's area.

Linear Classifier

Generally, as seen from Fig. 1b, the gray level of background of an RD image is of a much lower value than that of the targets and the clutter. That is, the gray levels with smaller values are associated with the background components, and the gray levels with larger values could be the targets regions or clutter areas. Therefore, a linear classifier with the one-dimensional RD power level feature

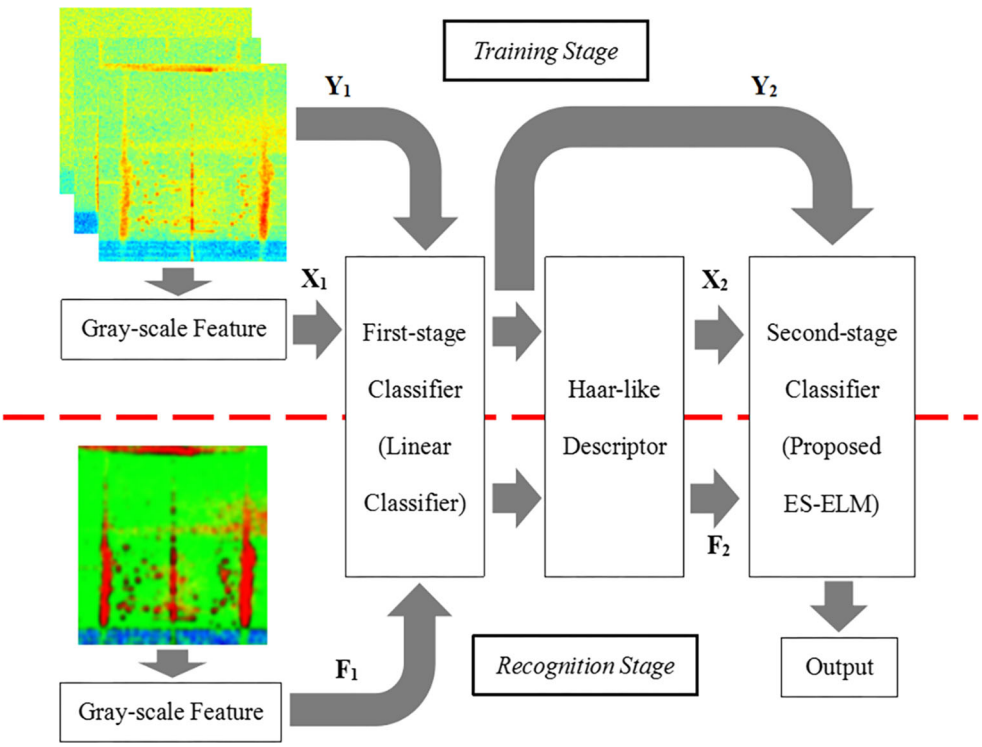

Fig. 2 Framework of the proposed ship target detection method

is applied for preliminary selection of the candidate target regions to eliminate the substantial background pixels and improve the detection efficiency.

In 2001, Viola and Jones first introduced AdaBoost method into human face detection [29], which represents a milestone in the real-time detection technology. Based on AdaBoost, the learning algorithm can select a small number of critical visual features and yields extremely efficient classifiers [30]. For each feature, a weak learning method is applied to determine the optimal threshold, and the minimum misclassified samples are performed.

Based on the concept of weak learning algorithm, a linear classifier is employed to remove the background region from an RD image. Given N distinct training samples, the designed linear classifier is shown as follows:

$$h(x_i) = \begin{cases} \text{Non-background}, g(x_i) \geqslant T \\ \text{Background}, g(x_i) < T \end{cases} \quad (3)$$

where x_i is the reference pixel, $g(x_i)$ is the gray-scale feature of the reference pixel, $h(x_i)$ is the output of linear classifier, and T is the threshold of the linear classifier obtained by the following training steps.

1) Obtaining the gray scale of each reference pixel, and sorting them in ascending order, denoted as t_i, $1 \leqslant i \leqslant N$.
2) Setting the optimal weight w_i of each sample, calculating the sum weight of background samples T^-. To retain the non-background component such as the targets, sea clutter, ground clutter to the extent possible, in our algorithm, the weights for non-background samples are set to 1, while the background weights are set to 0.01.

$$T^- = \sum_{i=1}^{N_b} t_i \times w_i \quad (4)$$

where N_b is the number of background samples.
3) For $i = 1, ..., N$, do
   a. Calculating the weighted sum of the non-background patterns and the background pattern prior to t_i, are denoted as S^{i+} and S^{i-}, respectively.

$$S^{i+} = \sum_{j=1}^{n_{nb}} t_j \times w_j, S^{i-} = \sum_{j=1}^{n_b} t_j \times w_j \quad (5)$$

   where n_{nb} is the non-background samples prior to t_i, while n_b is the background samples prior to t_i.
   b. Additionally, the current objective function e_i is evaluated with respect to S^{i+}, S^{i-} , T^-. To be more precise, the error function is mainly connected to the weighted sum of non-background patterns prior to t_i as well as the background patterns after t_i.

$$e_i = S^{i+} + (T^- - S^{i-}) \quad (6)$$

4) After step 3), a sequence $e_i (1 \leqslant i \leqslant N)$ is obtained. The optimal threshold T is given when the objective function reaches the least value.

$$\hat{i} = \arg\min_i \{S^{i+} + (T^- - S^{i-})\}$$
$$T = t_{\hat{i}} \quad (7)$$

Feature Extraction by Haar-Like Descriptor

The movement of ship targets in a coherence accumulation period will result in energy spreading in corresponding range of cells, thus exhibiting isolated peak regions in an RD image. Background noises are the superposition of external electromagnetic noise, cosmic noise, and internal noise. Thereby, they appear as isolated points of lesser intensity and variety than that of targets in an RD image. Since the magnitude of clutter varies slowly with regard to range axis or Doppler frequency shift axis, the information is a kind of banded region with slight change. Above all, the ship targets of HFSWR have different texture features that are different from other components of an RD image. Therefore, we employ Haar-like descriptor [31, 32] to extract local texture feature to identify ship targets precisely.

The concept of Haar-like feature descriptor was first proposed by Papageorgiou [33], and Haar-like feature was first adopted for object detection in clutter scene. Viola et al [34] and Lienhart et al [35] introduced a rapid detection scheme and designed a novel set of rotated Haar-like feature. Up to now, the Haar-like features can be classified into three types: edge features, line features, and center-surround features. Figure 3 gives the sketch of these three types.

In our algorithm, just the upright center-surround feature (i.e., as shown in Fig. 3a) is used to compose the

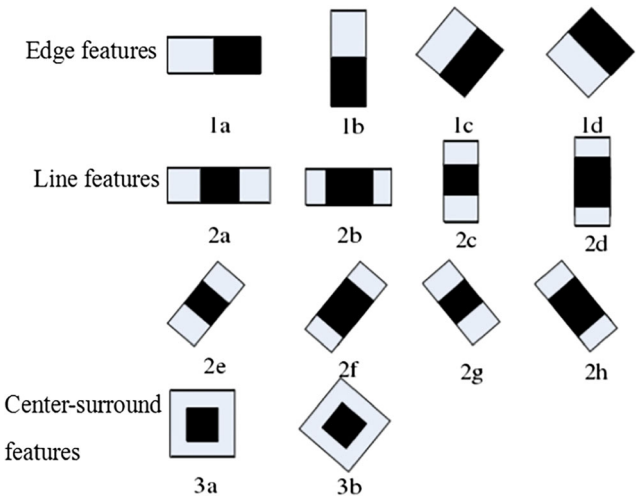

Fig. 3 Feature prototypes of haar-like features

Haar-like feature descriptor. For each reference pixel of target-highlighted RD image generated from the first-stage classifier, a 7×7 window centered at the reference pixel is used to extract the high level texture feature by Haar-like descriptor. The number of features generated from each window is 49.

Proposed Error Self-adjustment Extreme Learning Machine Classifier

After being processed by the proposed linear classifier, the RD image can be divided into two part, i.e., the background area and the highlighted ship targets area. The highlighted area is constituted by all kinds of clutters and ship targets, with almost all the background part removed. So the next task is to design a suitable network to detect all the ship targets from the highlighted ship targets area.

Inspired by [21, 24, 25, 39], we propose a new ES-ELM which would not only find the optimal number of hidden nodes, but also calculate the appropriate hidden parameters. The proposed ES-ELM is divided in two separate stages: the initialization stage and the updating stage. The aim of the initialization stage is to find a suitable feature mapping space including proper initial hidden parameters and hidden neuron number with L_1 regularizer process. The updating stage is to boost the generalization performance of our network through several updating processes based on L_2 regularization.

Initialization Stage

To find the appropriate feature mapping layer that can describes the input data in a simple expression, the sparse representation method and ELM network is combined in our proposed ES-ELM structure. In other words, the initialization stage enables the ES-ELM network to obtain the suitable feature space through the training data, and the sparse solution lends the trained network a lower computation complexity.

To be more precise, a hidden node will be useful for dealing with regression and classification problems if its connected weight coefficients are large. On the contrary, a hidden node will be weak or meaningless for the network if the connected weights are very small or equal to zero. Hence, in our ES-ELM, the L_1 regularizer process also known as lasso least square method is adopted to find the sparse solution of output weights, so that the weak neural nodes can be pruned and the optimal hidden neurons can be obtained. The L_1 regularizer can be defined as follows:

$$\text{Minimize} : ||\boldsymbol{\beta}||_1 + \lambda||\mathbf{H}\boldsymbol{\beta} - \mathbf{T}||_2^2 \tag{8}$$

where λ is a positive value.

Given the activation function $g(\cdot)$ and N arbitrary training samples, $\{(x_k, t_k)\}$, $1 \leqslant k \leqslant N$, $x_k \in R^n$, $t_k \in R^m$, which is sampled from a continuous system, $x \in R^{N\times n}$ is the training set, $t \in R^{N\times m}$ is the expected output, the main steps of initialization stage are described as follows.

Step 1: Initializing network with L hidden neurons (L equal to N), randomly generating the hidden parameters a and b. The feature matrix H can be obtained as follows:

$$H = g(x \cdot a + b) \tag{9}$$

Step 2: Calculating the output coefficients β by having L_1 regularizer, and ranking them in ascending order $\overline{\beta}(\overline{\beta_1} \leqslant \overline{\beta_2} \leqslant ... \leqslant \overline{\beta_L})$, the ratio of the first l hidden nodes to the sum nodes is as follows:

$$\lambda_l = \frac{\sum_{i=1}^{l} \overline{\beta_i}}{\sum_{j=1}^{L} \overline{\beta_j}} 1 \leqslant l \leqslant L \tag{10}$$

Step 3: Selecting a threshold λ, usually equal to 0.01. The appropriate number of hidden nodes d can be calculated by the following:

$$\begin{aligned} \hat{l} &= \min\{l|\lambda_l \leqslant \lambda\} \\ d &= L - \hat{l} \end{aligned} \tag{11}$$

Step 4: Pruning the useless hidden nodes and the hidden weights and bias can be denoted as $\hat{a}_f \in R^{n\times d}$ and $\hat{b}_f$.

Based on these steps, the redundant hidden neurons including the meaningless nodes and the weak nodes can be removed, which means that the pruned hidden layer itself can limit the computation complexity of the training network.

Updating Stage

To ensure that the trained networks have excellent generalization performance, several iterations are performed by having L_2 regularizer process to directly update the appropriate hidden layer and output layer parameters.

Step 1: For d number of optimal hidden nodes, the feature mapping data H_1 can be obtained as follows:

$$H_1 = g(x \cdot \hat{a}_f + \hat{b}_f) \tag{12}$$

Step 2: For the desired output t, the parameters $\hat{a}_h$ ($\hat{a}_h \in R^{d\times m}$) of output layer can be obtained as:

$$\hat{a}_h = H_1^{-1} \cdot t \tag{13}$$

where $H_1^{-1} = (I/C + H_1^T H_1)^{-1} H_1^T$, C is a positive value.

Step 3: Obtaining the output error $e_{n-1}, e_{n-1} \in R^{N\times m}$, and we can get the error feedback data P_{n-1}. The initial iteration number of n is set to 2.

$$\begin{aligned} e_{n-1} &= t - H_{n-1}\cdot a_{\text{hn}-1} \\ P_{n-1} &= e_{n-1}\cdot (I/C + (\hat{a}_{\text{hn}-1})^T \\ &\quad \cdot\hat{a}_{\text{hn}-1})^{-1}\cdot(\hat{a}_{\text{hn}-1})^T \end{aligned} \tag{14}$$

Step 4: The desired feature mapping data H_n can be computed by as follows:

$$H_n = H_{n-1} + P_{n-1} \tag{15}$$

Step 5: Updating the hidden layer parameters $\hat{a}_{\text{fn}}$ and $\hat{b}_{\text{fn}}$ by:

$$\begin{aligned} \hat{a}_{\text{fn}} &= (I/C + x^T x)^{-1}x^T \cdot g^{-1}(u(H_n)) \\ \hat{b}_{\text{fn}} &= \text{sum}(x\cdot\hat{a}_{\text{fn}} - g^{-1}(u(H_n)))/N \end{aligned} \tag{16}$$

where u is the normalized function.

Step 6: Updating the output layer parameters $\hat{a}_{\text{hn}}$ ($\hat{a}_{\text{hn}} \in R^{d\times m}$) as follows:

$$\hat{a}_{\text{hn}} = [u^{-1}(g(x\cdot\hat{a}_{\text{fn}} + \hat{b}_{\text{fn}}))]^{-1}\cdot t \tag{17}$$

where u^{-1} is the inverse function of u.

Step 7: $n = n + 1$, repeating steps 3–6 until the output error between the two successive iterations is stable, i.e., $||e_{n-1}|| - ||e_n|| \leqslant \varepsilon$, where ε is an arbitrary small real number.

Remark 1 In step 2, in order to solve L_1 regularized problem, the $l1\_ls$ toolbox which is a MATLAB implementation adapting for lasso least square problem is applied to quickly obtain the expected parameters.

Remark 2 Based on Bartlett's theory proposed in [37] and [38], an efficient learning structure should not only have the smallest training error but also the smallest weights norm, which means that when training a SLFN, the goal should also be focused on reaching the smallest norm of weights including input and output weights. Therefore, the steps 3–6 are iterated to attain the smallest norms both the output weights and input weights across all solutions.

Proof of Proposed ES-ELM Structure

In this section, we will perform a brief proof of our proposed ES-ELM network.

Lemma 1 [40]*: Given a banded non-constant piecewise continuous activation function $g(\cdot)$, we have*

$$\lim_{(a,b)\to(a_0,b_0)} ||g(x\cdot a + b) - g(x\cdot a_0 + b_0)|| = 0 \tag{18}$$

Definition 1 Given an activation function $g(\cdot)$, N arbitrary samples: $\{(x_k, t_k)\}$, $1 \leqslant k \leqslant N$, for any continuous expected output t: we have $||t - (u^{-1}(g(x\cdot\hat{a}_{\text{fn}-1} + \hat{b}_{\text{fn}-1}))\cdot\hat{a}_{\text{hn}-1})|| - ||t - (u^{-1}(g(x\cdot\hat{a}_{\text{fn}} + \hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}})|| \geqslant 0$ holds with probability one if

$$\begin{aligned} \hat{a}_{\text{fn}} &= (I/C + x^T x)^{-1}x^T \cdot g^{-1}(u(H_n)) \\ \hat{b}_{\text{fn}} &= \text{sum}(x\cdot\hat{a}_{\text{fn}} - g^{-1}(u(H_n)))/N \end{aligned} \tag{19}$$

$$\begin{aligned} \hat{a}_{\text{hn}} &= g(x\cdot\hat{a}_{\text{fn}} + \hat{b}_{\text{fn}})^{-1}\cdot t \\ H_n &= H_{n-1} + P_{n-1} \end{aligned} \tag{20}$$

where H_n is the desired feature mapping data of the nth iteration, $(I/C + x^T x)^{-1}x^T$ is the Moore-Penrose generalization inverse of the training samples, C is the regularization coefficient, $g^{-1}(\cdot)$ is the activation function, u is the normalized function, $u(x) : R \to (0, 1]$, it can process the feature mapping data from its original range to range $(-1, 1)$, u^{-1} is the inverse function of u, which can process the feature mapping data from range $(-1, 1)$ to its original range.

The proposed definition shows that if the parameters of this ELM structure are updated by these equations, the output error will be decreasing and finally get stable.

Proof We proof that the sequence $||t - (u^{-1}(g(x\cdot\hat{a}_{\text{fn}-1} + \hat{b}_{\text{fn}-1}))\cdot\hat{a}_{\text{hn}-1})|| - ||t - (u^{-1}(g(x\cdot\hat{a}_{\text{fn}} + \hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}})||$ is decreasing and bounded.

Let e_{n-1} be the error matrix of output layer and P_{n-1} be the error feedback matrix, we get

$$P_{n-1} = e_{n-1}\cdot a^{-1}_{\text{hn}-1} \tag{21}$$

where $a^{-1}_{\text{hn}-1} = (I/C + a^T_{\text{hn}-1}\cdot a_{\text{hn}-1})^{-1}\cdot a^T_{\text{hn}-1}$ is the Moore-Penrose generalization inverse of the neural weights. Thus, we have the expected feature mapping matrix H_2

$$H_n = H_{n-1} + P_{n-1} \tag{22}$$

where H_{n-1} is the original feature mapping matrix. Let $\lambda_n = x\cdot a_{fn}$, and λ_n satisfies:

$$\begin{aligned} g(\lambda_n) &= u(H_n) \\ \lambda_n &= g^{-1}(u(H_n)) \end{aligned} \tag{23}$$

For a sine activation function, we can get the solution of the hidden layer£°

$$\hat{a}_{\text{fn}} = x^{-1}\arcsin(u(H_n)) \tag{24}$$

For a sigmoid activation function, we can get the following equation£°

$$\hat{a}_{\text{fn}} = -x^{-1}log(\frac{1}{u(H_n)} - 1) \tag{25}$$

where x^{-1} is the Moore-Penrose generalized inverse of the training samples. Then, we can obtain the following conclusions.

1) $\hat{a}_{\text{fn}}$ is one of the least-squares solutions of a general linear system, meaning that the minimum error can be obtained by the solution: $||x\cdot\hat{a}_{\text{fn}}-\lambda_n||=\min||x\cdot a_{\text{fn}}-\lambda_n||$.
2) The norm of error matrix can be further reduced by the introduction of $\hat{b}_{\text{fn}}$, while we can get the smallest error by Eqs. 24 and 25: $\hat{b}_{\text{fn}} = \text{sum}(x \cdot \hat{a}_f - g^{-1}(u(H_n)))/N$. For different activation equation, we can get the different $g^{-1}(\cdot)$.

According to Lemma 1, we have

$$\begin{aligned}
&||u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))-H_n||\\
&\geqslant \min_{a_{\text{fn}},b_{\text{fn}}}||u^{-1}(g(x\cdot a_{\text{fn}}+b_{\text{fn}})-H_n)||\\
&=||u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}})-H_n)||=||\sigma_n|| \qquad (26)
\end{aligned}$$

Let $\hat{H}_n = u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}})) = H_n \pm \sigma_n$. Because $||P_{n-1}||^2$, $||H_{n-1}||^2$, $||H_n||^2$, $||\hat{H}_n||^2 \geqslant ||\sigma_n||^2$, the residual error of hidden layer can be presented as follows:

$$\begin{aligned}
\Delta_h &= ||H_n - H_{n-1}||^2 - ||H_n - \hat{H}_n||^2\\
&= ||H_{n-1}||^2 - 2\langle H_{n-1}, H_n\rangle + 2\langle H_n, \hat{H}_{n-1}\rangle - ||\hat{H}_n||^2\\
&= ||H_{n-1}||^2 - 2\langle H_{n-1}, H_n\rangle + 2\langle H_n, H_n \pm \sigma_n\rangle\\
&\quad - ||H_n \pm \sigma_n||^2\\
&= ||H_{n-1}||^2 - 2\langle H_{n-1}, H_n\rangle + 2||H_n||^2 \pm 2\langle H_n, \sigma\rangle\\
&\quad - (H_n^2 \pm 2\langle H_n, \sigma_n\rangle + ||\sigma_n||^2)\\
&= ||H_n||^2 - 2\langle H_{n-1}, H_n\rangle + ||H_{n-1}||^2 - ||\sigma_n||^2\\
&= ||H_n - H_{n-1}||^2 - ||\sigma||^2 = ||P_n||^2 - ||\sigma_n||^2 \geqslant 0 \qquad (27)
\end{aligned}$$

And we consider the residual error of output layer as follows:

$$\begin{aligned}
\Delta_o &= ||t - H_{n-1}\cdot\hat{a}_{\text{hn}-1}||^2 - ||t - \hat{H}_n\cdot\hat{a}_{\text{hn}-1}||^2\\
&= ||H_n\cdot\hat{a}_{\text{hn}-1} - H_{n-1}\cdot\hat{a}_{\text{hn}-1}||^2\\
&\quad -||H_n\cdot\hat{a}_{\text{hn}-1} - \hat{H}_n\cdot\hat{a}_{\text{hn}-1}||^2\\
&= ||(H_n - H_{n-1})\cdot\hat{a}_{\text{hn}-1}||^2 - ||(H_n - \hat{H}_n)\cdot\hat{a}_{\text{hn}-1}||^2\\
&\geqslant ||(H_n - H_{n-1})||^2\cdot||\hat{a}_{\text{hn}-1}||^2 -\\
&\quad ||(H_n - \hat{H}_n)||^2\cdot||\hat{a}_{\text{hn}-1}||^2\\
&= (||(H_n - H_{n-1})||^2 - ||(H_n - \hat{H}_n)||^2)\cdot\hat{a}_{\text{hn}-1}||^2\\
&= (||H_n - H_{n-1}||^2 - ||\sigma||^2)\cdot||\hat{a}_{\text{hn}-1}||^2 \geqslant 0 \qquad (28)
\end{aligned}$$

So that we have $||t-u^{-1}(g(x\cdot\hat{a}_{\text{fn}-1}+\hat{b}_{\text{fn}-1}))\cdot\hat{a}_{\text{hn}-1}|| \geqslant ||t-u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}-1}||$ Because $\hat{a}_{\text{hn}} = (u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}})))^{-1}$ is one of the least-square solutions of the general linear system $u^{-1}(g(\mathbf{x}\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot a_{\text{hn}} = t$, we have

$$\begin{aligned}
&||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}-1}||\\
&\geqslant \min_{a_{\text{hn}}}||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot a_{\text{hn}}||\\
&=||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}}|| \qquad (29)
\end{aligned}$$

According to Eqs. 28 and 29, we have

$$\begin{aligned}
&||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}-1}+\hat{b}_{\text{fn}-1}))\cdot\hat{a}_{\text{hn}-1}||\\
&\geqslant \min_{a_{\text{hn}}}||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot a_{\text{hn}}||\\
&=||t - u^{-1}(g(x\cdot\hat{a}_{\text{fn}}+\hat{b}_{\text{fn}}))\cdot\hat{a}_{\text{hn}}|| \qquad (30)
\end{aligned}$$

So the sequence is decreasing and bounded. □

Experimental Verification

This section conducts numerical experiments to validate the effectiveness of the proposed algorithm. Two groups of experiments are conducted in MATLAB 2016b with 12 GB of memory and an i7-860 (2.8GHz) processor. To evaluate the performance of the proposed ES-ELM structure, one set of experiments was performed using classic data sets downloaded from the Internet, as presented in the "Comparison of Our ES-ELM with Other Neural Methods." To verify the effectiveness of the proposed ship targets detection method, two sets of experiments were conducted using, respectively, real HFSWR data and simulated data. These experiments are described in the following three sections, namely, "Datasets and Evaluation Indexes for Ship Target Detection," "Identification Results of Ship Targets in RD Images," and "Comparison of Overall Detection Performance."

Comparison of Our ES-ELM with Other Neural Methods

In this section, ten regression and classification problems are proposed to examine the performance of our proposed neural network (ES-ELM), including five regression problems and five classification problems. Table 1 shows the corresponding training and testing data sets of each benchmark problem. All these data sets are downloaded from UCI machine learning repository and LIBSVM data sets. The

Table 1 Specification of 10 benchmark data sets

Datasets	Type	#Attri	#Train	#test
Hill-Valley	Classification	101	606	606
Iris	Classification	4	60	90
BCW (original)	Classification	10	300	399
Covtype.binary	Classification	54	300000	281012
Wine	Classification	13	90	88
Forest fires	Regression	13	239	278
Wine quality	Regression	12	2898	2000
Abalone	Regression	8	3000	1477
CPUsmall	Regression	12	5000	4192
Facebook metrics	Regression	9	300	200

Table 2 Generalization performance comparison

Datasets	SVM		ELM			ES-ELM		
	Tr_accuracy	Te_accuracy	Tr_accuracy	Te_accuracy	nodes	Tr_accuracy	Te_accuracy	Nodes
Hill-Valley	56.93%	53.63%	81.02%	79.04%	300	*98.02%*	*97.03%*	275
Iris	*90.00%*	*85.56%*	81.67%	78.89%	30	88.33%	84.44%	42
BCW	85.67%	84.21%	87.67%	85.96%	300	*97.67%*	*96.49%*	163
Covtype.binary	75.89%	72.55%	*79.28%*	*76.37%*	500	78.15%	76.10%	867
Wine	85.56%	81.58%	86.67%	76.87%	300	*97.78%*	*92.22%*	11

input data was normalized into [− 1, 1], while the output data for regression were normalized into [0, 1]. The results obtained from ten tests were averaged, as shown in this table. We present our experimental results from two aspects.

Firstly, to demonstrate the advantage of the proposed method of classification problems, SVM, ELM, and our ES-ELM were first compared by testing on five classification problems: Hill-Valley, Iris, BCW, Covtype.binary, and Wine. Table 2 shows the comparison between the proposed method and two classic neural network methods. The better performance is given in italics. In Tables 2 and 3, Tr_accuracy and Tr_RMSE refer to training performance, Te_accuracy and Re_RMSE are with respect to testing performance.

When the sigmoid nodes were applied, the proposed ES-ELM achieved a much better classification performance than the basic ELM and SVM in almost all the test data sets. The results from Hill-Valley (large number of input-dimension and medium number of training samples), Iris (both small number of input-dimension and training samples), and Covtype.binary (medium number of input-dimension and large number of training samples) are summarized as follows:

1) For the Hill-Valley problem, the testing accuracy of our ES-ELM network is 40 and 20% better than SVM and ELM respectively. In addition, the training accuracy of our proposed method is also better than that of ELM and SVM.
2) For the Iris problem, both training and testing performance of ES-ELM is superior to the basic ELM, while the testing accuracy is lower than SVM.
3) For the Covtype.binary problem, both ELM and ES-ELM methods perform well in the training and testing accuracy and are superior to SVM.

Secondly, to demonstrate the advantage of the proposed method for regression problems, ES-ELM was compared with error back propagation (BP) and ELM in five regression problems, namely, Forest Fires, Wine Quality, Abalone, CPUsmall, and Facebook metrics. The comparison results of three methods are tabulated in Table 3. It turns out that in comparison with the other two methods, our proposed ES-ELM shows a higher testing accuracy in almost all the datasets. For the Facebook metrics problem, the test shows RMSE of ES-ELM to be almost one-tenth and a half of the basic ELM and BP methods.

More importantly, Table 4 shows us how reliable in comparison with others the proposed structure is. In Table 4, when regularization parameter C, training sets, and testing sets are all fixed, our proposed method possesses a much better and more stable performance than that of basic ELM. The different from other neural structures is that generalization performance is highly sensitive to parameter C, and the testing results of ES-ELM are not too reliant on parameter C. So, the user can choose regularization parameter C randomly without too much of impact on the generalization performance.

Table 3 Generalization performance comparison

Datasets	BP (node 200)		ELM			ES-ELM		
	Tr_RMSE	Te_RMSE	Tr_RMSE	Te_RMSE	nodes	Tr_RMSE	Te_RMSE	nodes
Forest fires	0.3633	0.4521	0.1958	0.2043	200	*0.1360*	*0.1935*	134
Wine quality	0.1915	0.1853	*0.1799*	*0.1785*	200	0.1941	0.1853	194
Abalone	0.0921	0.1727	0.0816	0.1498	200	*0.0633*	*0.0712*	98
CPUsmall	0.3285	0.4906	0.2497	0.2897	500	*0.2214*	*0.2108*	143
Facebook metrics	0.7517	0.9529	0.2185	0.2568	200	*0.0463*	*0.1029*	19

Table 4 Generalization performance comparison

C	Forest fires		Wine quality		Abalone		CPU small		Facebook metrics	
	ELM	Ours	ELM	Ours	ELM	Ours	ELM	Ours	ELM	Ours
$C = 2^{-2}$	0.1920	*0.1286*	0.1534	*0.2015*	0.1065	*0.0742*	0.2338	*0.2285*	0.0956	*0.0542*
$C = 2^{0}$	0.1957	*0.1204*	0.2148	*0.2159*	0.1895	*0.0638*	0.2398	*0.2227*	0.1386	*0.0499*
$C = 2^{2}$	0.1491	*0.1245*	0.1649	*0.1958*	0.1099	*0.0626*	0.2345	*0.2218*	0.1477	*0.0612*
$C = 2^{4}$	0.2048	*0.1367*	0.2493	*0.2226*	0.2201	*0.0628*	0.2561	*0.2214*	0.1602	*0.0418*

In short, our proposed ES-ELM shows better training and testing performance in almost all the data sets. It benefits from all the advantages of the first contribution, that is, the higher learning efficiency of the neural structure and the insensitivity to the regularized parameter C.

Datasets and Evaluation Indexes for Ship Target Detection

To examine the performance of our proposed ship detection methods for extracting ship targets, in this subsection, we describe the experimental data sets and evaluation indexes used in our test.

To demonstrate the performance of our proposed method for HFSWR ship targets detection, two training sets (data set S_1 and data set S_2) are employed for the first-stage classifier and the second-stage classifier respectively. In our experiment, the data was collected from an HFSWR located on the coast of Bohai Bay of China, where the carrier frequency of HFSWR is 4.7 MHz, and the coherent accumulative time is 291 ms. Thirty measured RD images (size: 256×256 pixels) are used in our experiments, in which the background consists of various clutters and background noise, and the targets are real ships.

1) For the training data set of the first-stage classifier, the input data set consisted gray scale of the selected pixels which were chosen from real-measured RD images. The output set was formed by the type of the selected pixels, that is, whether the chosen pixels belonged to highlighted ship target area or not.
2) For the training data set of the second-stage classifier, the input data set was formed by 49 dimensional features, and each of the 49 dimensional feature vectors was computed by using Haar-like descriptor for each small window (7×7 pixels) centered at each pixel within the highlighted ship target area. The output data set was formed whether this pixel is ship target or not.

The details of our training sets can be seen in Table 5.

The evaluation criteria are the average detection probability P_d, average false alarm probability P_f, missing ratio M_r, and error ratio E_r. These are defined as follows:

$$\begin{aligned} P_d &= \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \times 100\% \\ P_f &= \mathrm{FP}/(\mathrm{FP}+\mathrm{TP}) \times 100\% \\ M_r &= 100\% - P_d \\ E_r &= P_f + M_r \end{aligned} \tag{31}$$

where TP is the number of real targets detected, FN is the number of real targets that were not detected, TP +FN is the total number of real targets, and FP is the number of false targets detected.

Identification Results of Ship Targets in RD Images

In this section, a set of experiments was conducted by using real measured data to examine the performance of our proposed algorithm for ship target detection. The performance is evaluated with several algorithms as shown in Table 6. CH means Haar-like feature with centered descriptor. LH means several combined lined descriptors, and LBP means local binary pattern, which can describe the local texture feature.

As for the training part, training data sets S_1 and S_2 are used for training first-stage classifier and second-stage classifier. As for testing part, for each RD image, the well-trained first-stage classifier is used for capture the

Table 5 Specification of two training data sets

Dataset	Sample	Input		Output	
		Dimension	Feature	Dimension	Class
S_1	1274	1	Gray-scale	2	Background or not
S_2	576	49	Haar-like	2	Target or not

Table 6 Detail of comparison algorithms

Method	Meaning
CH+ES-ELM	Proposed method
CH+ELM(200)	Haar-like feature with centered descriptor, ELM with 200 hidden nodes as second-stage classifier
CH+BP(200)	Haar-like feature with centered descriptor, BP with 200 hidden nodes as second-stage classifier
LH+ES-ELM	Haar-like feature with lined descriptor(combined Fig. 3 2b, 2f, 2h),ES-ELM as second-stage classifier
LBP+ES-ELM	LBP feature, ES-ELM as second-stage classifier

target-highlighted image. Then, the Haar-like descriptor is exploited for high-level feature extraction. Finally, the ES-ELM is used for ship targets reorganization.

By using the proposed method, we carried out the ship target extraction experiments with twenty RD images, the average P_d, P_f, M_r, E_r under that RD image data set are recorded in Table 7, CH+ES-ELM, LH+ES-ELM, and LBP+ES-ELM are mainly used to compare the proposed algorithms with different feature extraction ways, and CH+ES-ELM, CH+ELM, and CH+BP are performed to evaluate the performance of our proposed ES-ELM network on RD image data set. It is noticed that all these methods have same linear classifier training process.

As the Table 7 shows, both CH+ES-ELM and LBP +ES-ELM methods perform well in this data set, and their detection performance is superior to that of LH+ ES-ELM. The reason can be considered that the center-surround Haar-like descriptor could exactly describe the isolated peak regions of an RD image, while the feature extracted by several lined descriptor(LH+ES-ELM) is redundant and not reliable. In addition, from the reason mentioned above, it is clear that the detection results from our proposed ES-ELM network(CH+ES-ELM) are much better than those from other two methods(CH+ELM, CH+BP).

Two typical results detected by our proposed method are shown in Fig. 4, where the background contains not only clutter regions but also some residual noises. Figure 4a, b has the two original RD images. Apparently, it is very difficult to recognize the ship targets from these complex images. Figure 4c, d shows the obtained target-highlighted RD images, in which the black areas are the non-target regions, and all the ship targets are captured by target regions. It should be noted that the final classifier ES-ELM just considers target regions of the target-highlighted RD image from the first-stage classifier. Figure 4e, f shows the final ship targets extraction results. It is seen that almost all the ship targets in these two images are recognized. More importantly, most of the ship targets infected by clutters are extracted. Therefore, the proposed method is suitable for extracting ship targets from complex background of RD images.

Table 7 Average detection results with different algorithms

Method	P_d	P_f	M_r	E_r
CH+ES-ELM	92%	6%	8%	15%
CH+ELM	86%	13%	14%	27%
CH+BP	84%	13%	16%	29%
LH+ES-ELM	89%	8%	11%	19%
LBP+ES-ELM	91%	6%	9%	15%

Comparison of Overall Detection Performance

In this section, aimed at comparing the effectiveness of our proposed ship target detection method with other detection methods, extensive comparative experiments were conducted on simulated data. The evaluation indexes used in this section are mentioned in "Datasets and Evaluation Indexes for Ship Target Detection." The CFAR algorithm proposed by Liang [7] and image processing methods by Li [11] are the forefront research in their own areas, so we compare our algorithm with these two methods. The wavelet transform-based method [11] uses discrete wavelet transform (DWT) to detects ship targets, in which the PSNR-based method is used to automatically select the optimal scale of range-wise and Doppler-wise one-dimension DWT, and the fuzzy set-based method is used to enhance high-frequency subbands to strengthen ship targets and depress background noises. The CFAR-based method [7] utilizes curve fitting method as preprocessing to weaken clutter and background noise, then a reference window of each reference cell is collected for evaluating whether it is the background noise or not. Finally, two-dimensional CFAR for each reference window is carried out for target detection. However, the CFAR-based method needs to determine the threshold factor correctly. For fair comparison, we use forty different values of threshold factor $\{0.05, 0.1, 0.15, ...1.95, 2\}$ to find the appropriate one.

In practical HFSWR monitoring, not all the RD images know the exact number and position of the ship targets. Although we can use automatic identification system (AIS)

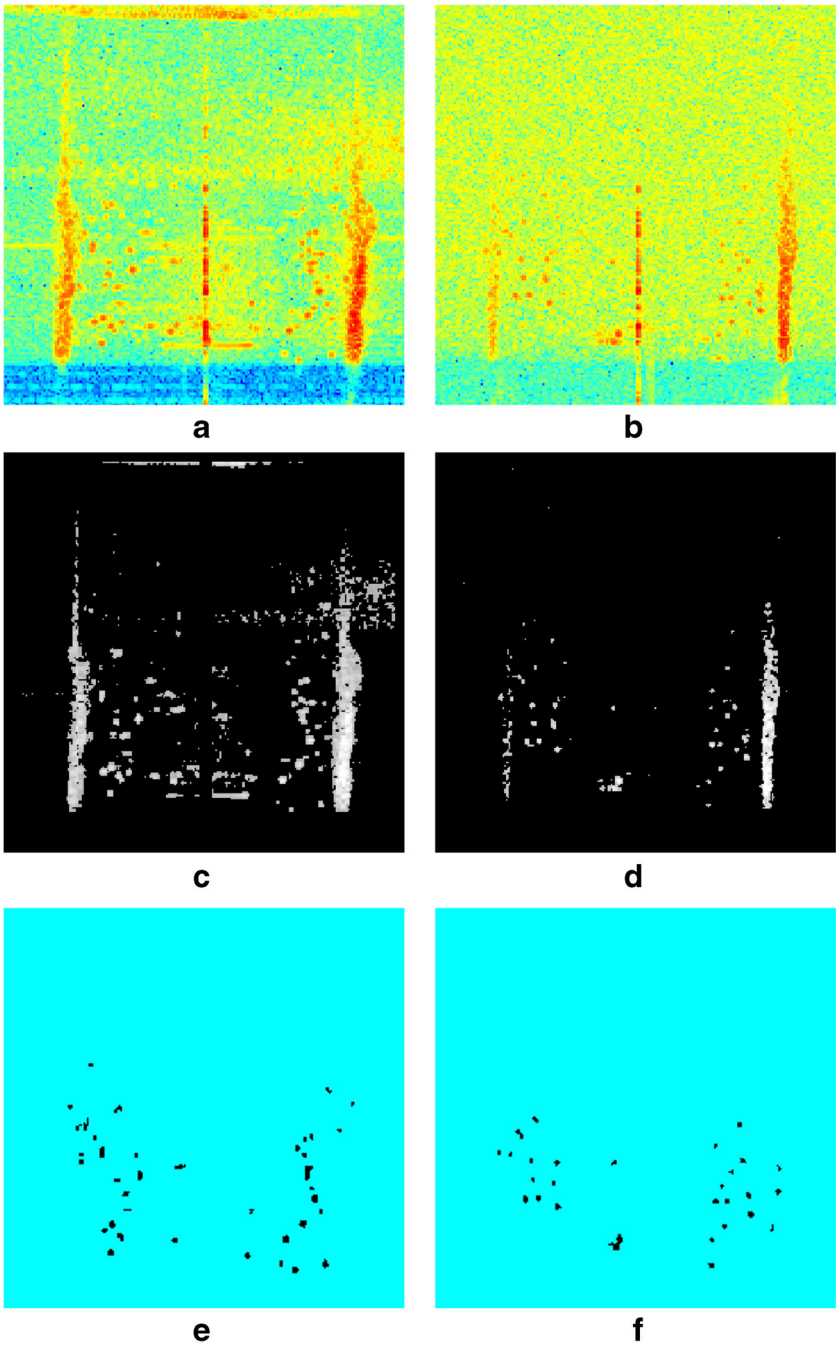

Fig. 4 Two typical ship target extraction results **a**, **b** two original RD images; **c**, **d** target-highlighted RD images; **e**, **f** ship target detection results

Table 8 Detection results of three methods (time: aaverage testing time (s))

Method	P_d	P_f	M_r	E_r	Time
Ours	*92%*	*5%*	*8%*	*13%*	*3.1784*
CFAR-based	85%	13%	15%	28%	4.9010
Wavelet-based	90%	8%	10%	18%	6.1398

Table 9 Comparison results of two algorithm for targets located near clutter

Dataset	SNR of simulated ship targets				
	20 dB	25 dB	30 dB	35 dB	40 dB
CFAR-based	0%	3%	5%	18%	32%
Wavelet-based	3%	10%	45%	83%	93%
Ours	5%	57%	78%	93%	92%

data to receive the information such as unique identification, position, and speed. AIS on some ships may not work because the installation of AIS is not a mandatory. Therefore, the detection performance of our proposed method and other methods are compared by simulated RD data, where the ship targets are added on the measured RD background randomly, and the SNR of the added targets ranges from 20 to 40 dB. Since the number and position of the added ship targets are known, it is easy for us to perform comparison experiments. In the first experiment, 200 RD images are chosen as test sample, each of these images has a number of added ship targets.

Table 8 shows the results of the comparison of these three algorithms, along with the comparison of the evaluation indexes mentioned above, and the testing time of the three methods, where the better testing results shown in italics. In our comparison experiments, the wavelet-based method considers the time spent on discrete wavelet transform (DWT) with the adaptive threshold method of image segmentation. In contrast, the CFAR-based method just calculates the time taken from the beginning of curve fitting to the end of ship detection. However, the time counted in our proposed method simply covers the testing stage while the training stage is not considered. Owing to the introduction of multistage classifier, most of the background regions and background noises are recognized and eliminated at the first-stage classifier, so the testing time of our proposed method is 30% less than those of CFAR-based method and wavelet transform-based method. This advantage will be even more striking when higher resolution RD images are used.

As shown in the first four columns of Table 8, our proposed method achieves a better detection performance, where the average P_d of our proposed algorithm exceeds those of CFAR-based method and wavelet-based method by 2 and 7% respectively. In addition, P_f, M_r, and E_r of the proposed method are all lower than those of the other two methods. It is worth noting that our method has a much lower Er than the other two methods. The reason for the significant reduction in these three indexes is that in our algorithm, the high levels of feature representation of clutter and targets by the use of Haar-like descriptor are very robust under variant conditions. More importantly, the proposed ES-ELM classifier has a better classification performance than other classifiers.

To further evaluate the capability of the three methods in detecting weak targets which are infected by sea clutter, five sets of tests were performed. Each test employs fifty RD images, and the SNR of the simulated targets ranges from 20 to 40 dB. As shown in Table 9, compared with CFAR-based method, our proposed method shows a significant improvement in weak target detection in every test. As compared with wavelet-based method, our method has a much better target detection performance if the SNR of the simulated targets ranges from 20 to 35 dB. Therefore, some infected and weaker targets are more easily detected by our method.

Figure 5 is a typical detection result by using wavelet-based method and our proposed method. The CFAR-based method is not presented because this method cannot detect any simulated target. Three simulated targets added in sea clutter region are marked with red circles. Figure 5a is the original RD image, Fig. 5b is the target detection result of the wavelet-based algorithm, and Fig. 5c is the target detection result of our method. It can be seen that the wavelet-based algorithm can just detect two of them, and our algorithm can detect all the three targets.

In conclusion, our proposed ship detection method shows a better performance in weak target detection on account of the advantage arising from the second contribution, namely, the use of multistage classifier and ES-ELM structure.

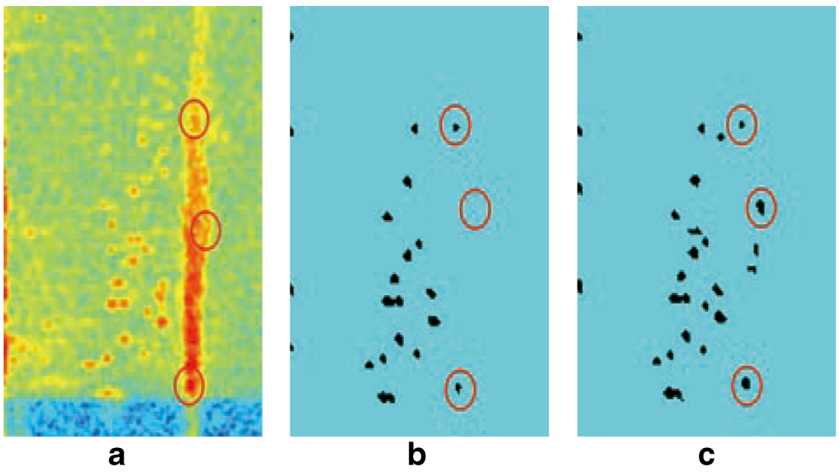

Fig. 5 Ship target detection result: **a** original RD image; **b** target detection result of wavelet transform-based algorithm; **c** target detection result of our algorithm

Conclusion

This paper proposes an efficient ship target detection method as well as a novel improved ELM network which has been rigorously proven. Here are several interesting features inside.

1) In the proposed ES-ELM, the optimal number of hidden neurons can be obtained by using the L_1 regularizer process. In addition, the parameters of output layer and hidden layer can be appropriately calculated by iteration steps which are mainly based on L_2 regularizer with pulled back output error. Moreover, it has been verified that the generalization performance of our network is not sensitive to regularized parameter C.
2) Through the use of cascade classification strategy, the proposed RD image target detection algorithm can impressively boost the detection efficiency, and paves the way for real-time detection of ship targets. In addition, by using our ES-ELM structure as the final classifier with Haar-like input feature, the detection algorithm achieves similar or much better detection performance than the previous methods. Therefore, this paper presents a useful targets detection method for RD images. Our future work will focus on further applying the proposed method for ship target tracking with sequential RD frames.

Funding Information This work was funded by the National Nature Science Foundation of China (grant nos. 41506114 and 61132005) and by the National Marine Technology Program for Public Welfare (grant no. 201505002).

Compliance with Ethical Standards

Conflict of Interests The author declares that she has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Grosdidier S, Baussard A, Khenchaf A. HFSW Radar model: simulation and measurement. IEEE Trans Geosci Remote Sens. 2010;48(9):3539–49.
2. Huang W, Gill E, Wu X, et al. Measurement of sea surface wind direction using bistatic high-frequency radar. IEEE Trans Geosci Remote Sens. 2012;50(10):4117–22.
3. Hinz JO, Holters M, Zolzer U, et al. Presegmentation-based adaptive CFAR detection for HFSWR. In: IEEE radar conference; 2012. p. 665–670.
4. Liu T, Lampropoulos GA, Fei C. CFAR ship detection system using polarimetric data. In: IEEE radar conference; 2008. p. 1–4.
5. Rohling H. Radar CFAR thresholding in clutter and multiple target situations. IEEE Trans Aerosp Electron Syst. 2007;AES-19(4):608–21.
6. Gui R. Detection target located in nonstationary background based on two-dimensions constant false alarm rate. Geomatics Inf Sci Wuhan University. 2012;37(3):354–7.
7. Liang J. Target CFAR detection method and software implementation with two-dimension data for HFSWR Qingdao: Ocean University of China. 2014.
8. Grosdidier S, Baussard A. Ship detection based on morphological component analysis of high-frequency surface wave radar images. Iet Radar Sonar and Navigation. 2012;6(9):813–21.
9. Jangal F, Saillant S, Helier M. Wavelet contribution to remote sensing of the sea and target detection for a high-frequency surface wave radar. IEEE Geosci Remote Sens Lett. 2008;5(3):552–6.
10. Jangal F, Saillant S, Helier M. Wavelets: a versatile tool for the high frequency surface wave radar. In: IEEE radar conference; 2007. p. 497–502.
11. Li Q, Zhang W, Li M, et al. Automatic detection of ship targets based on wavelet transform for HF surface wavelet radar. IEEE Geosci Remote Sens Lett. 2017;14(5):714–8.
12. Wang T, Cao J, Lai X, Chen B. Deep weighted extreme learning machine. Cogn Comput. https://doi.org/10.1007/s12559-018-9602-9. 2018.
13. Huang GB, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern. 2012;42(2):513–29.
14. Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006;70(1):489–501.
15. Huang G, Huang GB, Song S, et al. Trends in extreme learning machines: a review. Neural Netw. 2015;61(C):32–48.
16. Cao J, Zhang K, Luo M, et al. Extreme learning machine and adaptive sparse representation for image classification. Neural Netw. 2016;81:91–102.
17. Liu Y, Zhang L, Deng P, et al. Common subspace learning via cross-domain extreme learning machine. Cogn Comput. 2017;9(3):1–9.
18. Duan L, Bao M, Cui S, et al. Motor imagery EEG classification based on kernel hierarchical extreme learning machine. Cogn Comput. 2017;9(6):1–8.
19. Mao W, Jiang M, Wang J, et al. Online extreme learning machine with hybrid sampling strategy for sequential imbalanced data. Cogn Comput. 2017;9(7):1–21.
20. Huang GB, Chen L. Enhanced random search based incremental extreme learning machine. Neurocomputing. 2008;71(16):3460–8.
21. Huang GB, Chen L. Convex incremental extreme learning machine. Neurocomputing. 2007;70(16):3056–62.
22. Miche Y, Sorjamaa A, Bas P, et al. OP-ELM: optimally pruned extreme learning machine. IEEE Trans Neural Netw. 2010;21(1):158–62.
23. Feng G, Huang GB, Lin Q, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning. IEEE Trans Neural Netw. 2009;20(8):1352–7.
24. Yang Y, Wang Y, Yuan X. Bidirectional extreme learning machine for regression problem and its learning effectiveness. IEEE Trans Neural Netw. 2012;23(9):1498–1505.
25. Yang Y, Wu QM. Extreme learning machine with subnetwork hidden nodes for regression and classification. IEEE Trans Cybern. 2016;46(12):2885–98.
26. Yang Y, Wu QM, Wang Y. Autoencoder with invertible functions for dimension reduction and image reconstruction. IEEE Trans Syst Man Cybern Syst. 2016;PP(99):1–15.
27. Dienstfrey A, Hale PD. Colored noise and regularization parameter selection for waveform metrology. IEEE Trans Instrum Meas. 2014;63(7):1769–78.

28. Bauer F, Lukas MA. Comparing parameter choice methods for regularization of ill-posed problems. Math Comput Simul. 2011;81(9):1795–841.
29. Viola P. Robust real-time object detection. International workshop on statistical and computational theories of vision modeling. Learning, Computing, and Sampling. 2001;57(2):87.
30. Ding X, Ma Z. Real-time face detection with self-adaptive cost sensitive AdaBoost. In: IEEE conference on industrial electronics and applications; 2008. p. 1980–1982.
31. Ma S, Bai L. A face detection algorithm based on Adaboost and new Ha ar-like feature. In: IEEE international conference on software engineering and service science; 2017. p. 651–654.
32. Doungmala P, Klubsuwan K. Helmet wearing detection in Thailand using Haar like feature and circle hough transform on image processing. In: IEEE international conference on computer and information technology; 2017. p. 611–614.
33. Papageorgiou CP, Oren M, Poggio T. A general framework for object detection. In: IEEE international conference on computer vision; 2002. p. 555–562.
34. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: IEEE conference on computer vision and pattern recognition; 2001. p. 511.
35. Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection. In: IEEE international conference on image processing; 2002. p. 900–903.
36. Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. Int J Mach Learn Cybern. 2011;2(2):107–22.
37. Bartlett PL. For valid generalization, the size of the weights is more important than the size of the network. In: Advances in neural information processing systems; 1997. p. 134–140.
38. Bartlett PL. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. IEEE Trans Inf Theory. 1998;44(2):525–36.
39. Yang Y, Wu QMJ. Multilayer extreme learning machine with subnetwork nodes for representation learning. IEEE Trans Cybern. 2015;46(11):2570–83.
40. Huang GB, Chen L. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw. 2006;17(4):879–92.
41. Qing T, Chen SC. Cross-heterogeneous-database age estimation through correlation representation learning. Neurocomputing. 2017;238(17):286–95.
42. Fu ZJ, Huang FX, Ren K, et al. Privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data. IEEE Trans Inf Forensics Secur. 2017;12(8):1874–84.