


Peer-Reviewed Technical Communication

Segmentation of Sidescan Sonar Imagery Using Markov Random Fields and Extreme Learning Machine

Yan Song, Bo He , Ying Zhao, Guangliang Li, *Member, IEEE*, Qixin Sha, Yue Shen, Tianhong Yan, Rui Nian, and Amaury Lendasse

Abstract—As a widely used segmentation scheme, Markov random field (MRF) utilizes k -means clustering to calculate the initial model for sidescan sonar image segmentation. However, for the noise and intensity inhomogeneity nature of the sidescan sonar images, the segmentation results of k -means clustering have low accuracy, motivating us to use machine learning methods to initialize MRF. Meanwhile, an extreme learning machine (ELM), a supervised learning algorithm derived from the single-hidden-layer feedforward neural networks, learns faster than randomly generated hidden-layer parameters and is superior to a support vector machine (SVM). Therefore, in this paper, we proposed a novel method for sidescan sonar image segmentation based on MRF and ELM. The proposed method segments sidescan sonar images in object-highlight, object-shadow, and sea-bottom reverberation areas. Specifically, we intend to use an ELM to get an initial model for MRF. Moreover, to improve the stability of an ELM, a simple ensemble ELM (SE-ELM) based on an ensemble algorithm is utilized to obtain the prediction model. In an SE-ELM, we use an ensemble of ELMs and majority votes to determine the prediction of testing data sets. Then, the classification results of the SE-ELM are utilized to initialize MRF, termed as SE-ELM-MRF. With features consisting of pixels of small image patches, our experiments on real sonar data indicate that the SE-ELM performs better than other machine learning methods such as ELM, kernel-based extreme learning machine, SVM, and convolutional neural networks. Moreover, using SE-ELM as the initial method in the proposed SE-ELM-MRF, the segmentation results are smoother and the segmentation process converges faster than the traditional MRF.

Index Terms—Convolutional neural networks (CNNs), extreme learning machine (ELM), image segmentation, Markov random fields (MRFs), sidescan sonar, support vector machine (SVM).

I. INTRODUCTION

HOW to categorize sidescan sonar images efficiently and accurately is vital for doing oceanographic survey and realizing the autonomy of autonomous underwater vehicles (AUVs). Sidescan sonar is the main sensor for detecting the location and state of submarine objects [1] because it can obtain clear high-resolution images of the

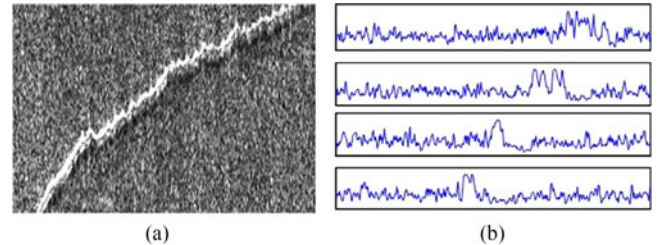


Fig. 1. Returned echo signal of sidescan sonar imagery.

sea bottom. The sidescan sonar works with two transducers emitting sound pulses at precise and regulated intervals. The system receives the echoes from the water column and seafloor shortly after a pulse is emitted. Fig. 1(a) shows a sidescan sonar image and Fig. 1(b) shows the gray level of some scan lines of the image shown in (a). From Fig. 1(b), we can see that the object-highlight area returns a higher signal than other echoes. Therefore, with appropriate methods, we could detect highlight objects from sidescan sonar images even under the influence of intensity inhomogeneity and noise.

Many researchers have put forward various methods to carry out sidescan sonar imagery segmentation, which can be put into two classes: supervised methods [2], [3] and unsupervised methods [4], [5]. For example, an active learning method based on model trees [2] was used to automatically label the new objects of sonar images which were not included in the training set. Stewart *et al.* [3] used a backpropagation neural network to classify three distinct geoaoustic provinces with several common texture features. The supervised methods could find out the encountered data which are consistent with the labeled training data. That is to say, it may not work well when the environment changes or the objects of interest are new to the classifier. The unsupervised methods such as density-based spatial clustering of applications with noise [6], Dirichlet process mixture modeling [7], and Gaussian mean shift [8] are usually based on clustering without requiring labeled data in advance. However, it may be time-consuming to tune the threshold of unsupervised methods if sonar images are obtained from a complex seafloor environment.

In addition, Mignotte *et al.* [9] proposed an unsupervised sonar segmentation algorithm based on Markov random field (MRF). MRF performed well in three-class sonar image segmentation in object-highlight, object-shadow, and sea-bottom reverberation areas with iterative conditional estimation (ICE) [10]. MRF even worked well with real sonar images [5], [9], [11]–[14]. Fig. 2(a) shows a sidescan sonar image whose MRF segmentation result is shown in Fig. 2(b). In Fig. 2(b), the white part denotes the object-highlight areas, the gray part denotes the sea-bottom reverberation areas, and the black part de-

Manuscript received July 15, 2016; revised October 19, 2017, November 11, 2017, January 24, 2018, and March 15, 2018; accepted March 20, 2018. This work was supported by the National Key Research and Development Program of China under Grant 2016YFC0301400. (Corresponding author: Bo He.)

Associate Editor: J. Cobb.

Y. Song, B. He, Y. Zhao, G. Li, Q. Sha, Y. Shen, and R. Nian are with the School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China (e-mail: songyan314@126.com; bhe@ouc.edu.cn; zhaoying@stu.ouc.edu.cn; guangliangli@ouc.edu.cn; sqxqx86@ouc.edu.cn; shenyue@ouc.edu.cn; nianrui_80@163.com).

T. Yan is with the School of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China (e-mail: thyan@163.com).

A. Lendasse is with the Department of Mechanical and Industrial Engineering and the Iowa Informatics Initiative, University of Iowa, Iowa City IA 52242-1527, USA and also with the Arcada University of Applied Sciences, Helsinki 00550, Finland (e-mail: amaury-lendasse@uiowa.edu).

Digital Object Identifier 10.1109/JOE.2018.2819278

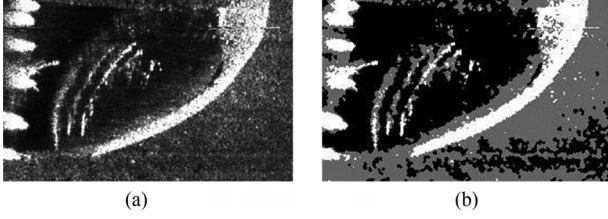


Fig. 2. (a) Raw sonar image. (b) Result of MRF segmentation on (a).

notes the object-shadow areas. MRF can segment most information correctly from the original sonar images, but it is sensitive to intensity inhomogeneity of the sonar image and takes about 48.21 s even when the size of the figure is only 170 pixels \times 270 pixels. (Details of the hardware will be given in Section IV.) One primary reason might be that MRF uses the segmentation results of k -means clustering to initialize its model parameters and k -means clustering has bad performance in segmentation on sidescan sonar images with much speckle noise and intensity inhomogeneity, which are the nature of the sidescan sonar images. This motivates us to find other methods to initialize MRF.

Recently, many algorithms based on an extreme learning machine (ELM) [15]–[18] have been proposed for image classification [19], [20] and segmentation [21], [22]. An ELM is derived from single-hidden-layer feedforward neural networks (SLFNs). The advantages of an ELM include that the hidden-layer parameters of the ELM can be generated randomly and it learns faster and can achieve higher accuracy [15], [23], [24] than the support vector machine (SVM) [25], [26]. Samat *et al.* [19] proposed a supervised classification method by combining ELM and active learning to carry out synthetic aperture radar (SAR) image classification. In [20], two ensemble ELMs based on bagging and Ada-boosting [27] were proposed to do hyperspectral image classification. In addition, an ELM was also employed to detect and segment liver tumors in a three-dimensional computed tomography [21]. Moreover, a fast benthic object recognition model [28] based on structured geometrical features and ELM was utilized to categorize real sonar imagery.

Therefore, in this paper, we propose an initialization method based on an ELM to get a better initial model for MRF. To improve the performance of an ELM, we proposed the simple ensemble extreme learning machine (SE-ELM) method based on the ensemble algorithm. An SE-ELM uses an ensemble of ELMs and a majority voting strategy to get the prediction results. Moreover, an SE-ELM is used to initialize MRF, termed as SE-ELM-MRF. We test our methods on real sonar images and our experimental results demonstrate that an SE-ELM is superior in sonar image segmentation compared with other machine learning methods, such as ELM, kernel-based extreme learning machine (KELM), SVM, and convolutional neural networks (CNN) [29], [30]. Moreover, SE-ELM-MRF has better segmentation performance and converges faster than MRF.

The rest of this paper begins with an introduction to ELM and KELM in Section II. Section III gives a detailed description of the presented method. Section IV shows and discusses the experimental results. Finally, Section V concludes this paper.

II. RELATED METHODS

A. Extreme Learning Machine

An ELM is derived from SLFNs. According to Huang *et al.* [15], [31], if the input weights and hidden-layer biases are set randomly and the hidden layer uses infinitely differentiable activation functions, the

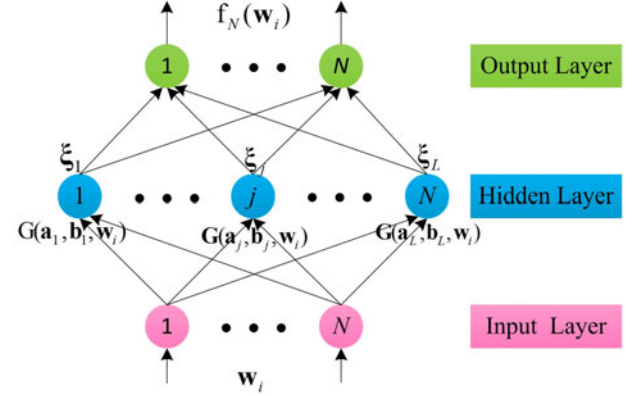


Fig. 3. Single-hidden-layer feedforward neural network.

SLFN classifier trained by an ELM is able to obtain an impressive generalization performance in multiclass classification cases.

Let the input observations of an ELM be $\{(\mathbf{w}_i, \mathbf{t}_i) | \mathbf{w}_i \in \mathbf{R}^{1 \times n}, \mathbf{t}_i \in \mathbf{R}^{1 \times m}, i = 1, \dots, N\}$. Here, \mathbf{w}_i is a $1 \times n$ input vector, \mathbf{t}_i is the class label of \mathbf{w}_i , and there are L hidden nodes in an ELM. If the class label of \mathbf{w}_i is p , the expected output vector of \mathbf{w}_i is $\mathbf{t}_i = [0, \dots, 1, \dots, 0]_m^T$, which means that only the p th element of $\mathbf{t}_i = [t_{i,1}, \dots, t_{i,p}, \dots, t_{i,m}]^T$ is one while the rest of the elements are equal to zero. The output function of an ELM is

$$f_N(\mathbf{w}_i) = \sum_{j=1}^L \xi_j h_j(\mathbf{w}_i) = \mathbf{h}(\mathbf{w}_i) \boldsymbol{\xi} \quad (1)$$

where $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_L]^T$ is the output weight between the hidden layer and the output layer, and $\mathbf{h}(\mathbf{w}_i) = [h_1(\mathbf{w}_i), h_2(\mathbf{w}_i), \dots, h_L(\mathbf{w}_i)]^T$ is the output vector of the hidden layer with respect to the input \mathbf{w}_i . In particular, in real applications, $h_j(\mathbf{w}_i)$ can be written as

$$h_j(\mathbf{w}_i) = G(\mathbf{a}_j \cdot \mathbf{w}_i + b_j), \quad \mathbf{a}_j \in \mathbf{R}^{1 \times n}, b_j \in \mathbf{R} \quad (2)$$

where $G(\cdot)$ is a nonlinear piecewise continuous function, $\{(\mathbf{a}_j, b_j)\}_{j=1}^L$ are randomly generated according to any continuous probability distribution, \mathbf{a}_j is the weight vector connecting the j th hidden node and the input nodes, b_j is the threshold of the j th hidden node, and $\mathbf{a}_j \cdot \mathbf{w}_i$ is the inner product of \mathbf{a}_j and \mathbf{w}_i . Given N training samples $\{(\mathbf{w}_i, \mathbf{t}_i)\}_{i=1}^N$, the ELM can resolve the following learning problem:

$$\|\mathbf{H}\hat{\boldsymbol{\xi}} - \mathbf{T}\| = \min \left(\sum_{i=1}^N \|f_N(\mathbf{w}_i) - \mathbf{t}_i\| \right) = \min_{\boldsymbol{\xi}} \|\mathbf{H}\boldsymbol{\xi} - \mathbf{T}\| \quad (3)$$

where

$$\mathbf{H} = \begin{bmatrix} G(\mathbf{a}_1 \cdot \mathbf{w}_1 + b_1) & \cdots & G(\mathbf{a}_L \cdot \mathbf{w}_1 + b_L) \\ \vdots & \ddots & \vdots \\ G(\mathbf{a}_1 \cdot \mathbf{w}_N + b_1) & \cdots & G(\mathbf{a}_L \cdot \mathbf{w}_N + b_L) \end{bmatrix}_{N \times L} \quad (4)$$

and $\mathbf{T} = [\mathbf{t}_1^T, \mathbf{t}_2^T, \dots, \mathbf{t}_N^T]^T$. According to [15], the minimum norm least squares solution of $\mathbf{H}\hat{\boldsymbol{\xi}} = \mathbf{T}$ is unique, which is

$$\hat{\boldsymbol{\xi}} = \mathbf{H}^\dagger \mathbf{T} \quad (5)$$

where \mathbf{H}^\dagger is the Moore–Penrose generalized inverse of matrix \mathbf{H} .

An ELM can be summarized in Fig. 3. In contrast to most of the existing approaches, an ELM only updates the output weights between

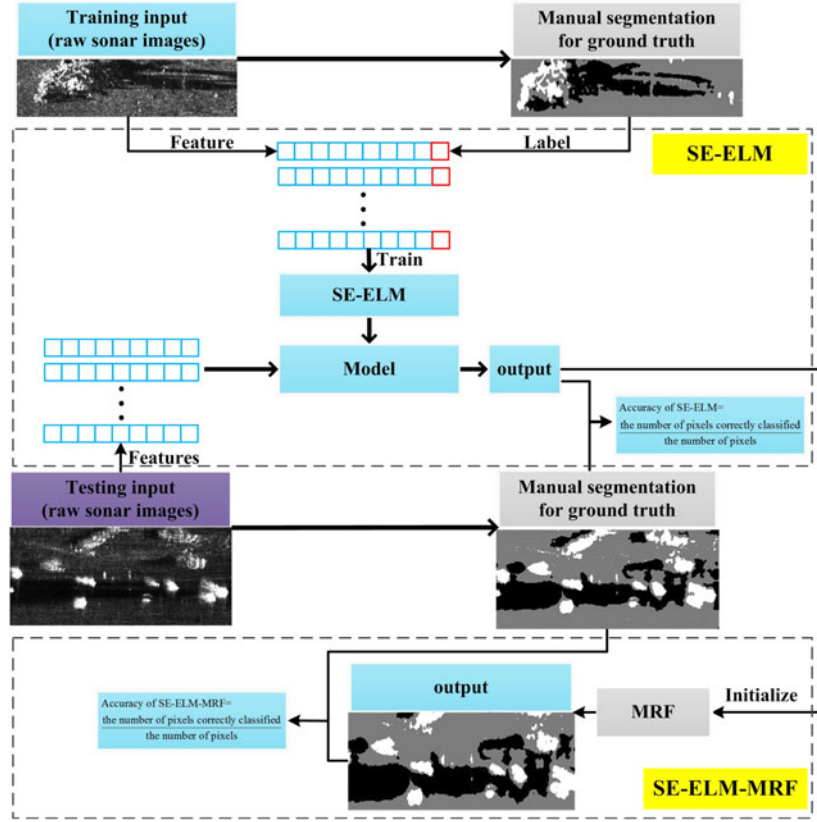


Fig. 4. Flowchart of the proposed SE-ELM-MRF method.

the hidden layer and the output layer, while the input weights and biases of the hidden layer are randomly generated. By adopting the squared loss on the prediction error, the training of output weights turns into a regularized least squares (or ridge regression) problem, which can be solved efficiently in the closed form. It has been shown that even without updating the parameters of the hidden layer, the SLFN with randomly generated hidden neurons and tunable output weights maintains its universal approximation capability [32].

B. Kernel-Based Extreme Learning Machine

Huang *et al.* proposed a KELM [31] based on kernel methods to improve the generalization ability of the original ELM. Specifically, a regularization coefficient ϖ was added for calculating the output weights ξ' and ϖ and could be chosen according to the optimum validation results. ξ' can be calculated as follows:

$$\xi' = \mathbf{H}^T \left(\frac{\mathbf{I}}{\varpi} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T}, \quad \text{if } N \leq L \quad (6)$$

$$\xi' = \left(\frac{\mathbf{I}}{\varpi} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T}, \quad \text{if } N > L. \quad (7)$$

III. PROPOSED SEGMENTATION APPROACH

This section describes the proposed segmentation method. Fig. 4 shows the flowchart of the whole method, which consists of three main parts: manual segmentation for ground truth [33], [34], SE-ELM, and SE-ELM-MRF. First, we utilize manual segmentation to get the ground truth of sidescan sonar images. Then the SE-ELM is trained with the labeled data set. Finally, the segmentation results of the

SE-ELM are used to initialize MRF to get the final segmentation results. Accordingly, this section is organized as follows: Section III-A presents the details of an SE-ELM and Section III-B provides the details of SE-ELM-MRF.

A. Simple Ensemble Extreme Learning Machine

In this paper, we use an SE-ELM to perform sonar image classification. The results of manual segmentation are used as ground truth images to train the supervised method SE-ELM. An SE-ELM can reduce the randomness caused by randomly generated hidden-layer parameters and improve the performance of the ELM, which make it more stable than the original ELM. In an SE-ELM, several ELMs work in parallel, the k -fold cross validation [35] is utilized to select the optimal number of hidden nodes for each ELM, and the majority voting [36] is employed for the prediction of the testing data. The flowchart of the SE-ELM is shown in Fig. 5. An SE-ELM is carried out with the following three steps.

- 1) Split the original data set \mathbf{V} into k subsets $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ randomly and each subset consists of M samples. A sample is a vector combining the label vector and its input vector. The number of samples in \mathbf{V} is n_v . If M is divided exactly by n_v , then $k = n_v/M$; if n_v divided by M gives out a quotient of u with a remainder, then $k = u + 1$.
- 2) Use the j th ($1 \leq j \leq k$) reunion sets $\mathbf{V}_{j,\text{train}} = \mathbf{V} - \mathbf{v}_j$ as the training data set and the j th subset \mathbf{v}_j as the validation data set to train the j th classifier ELM_j of the SE-ELM to get the output parameter \mathbf{c}_j . Then we get all output parameters $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ for all ELMs.

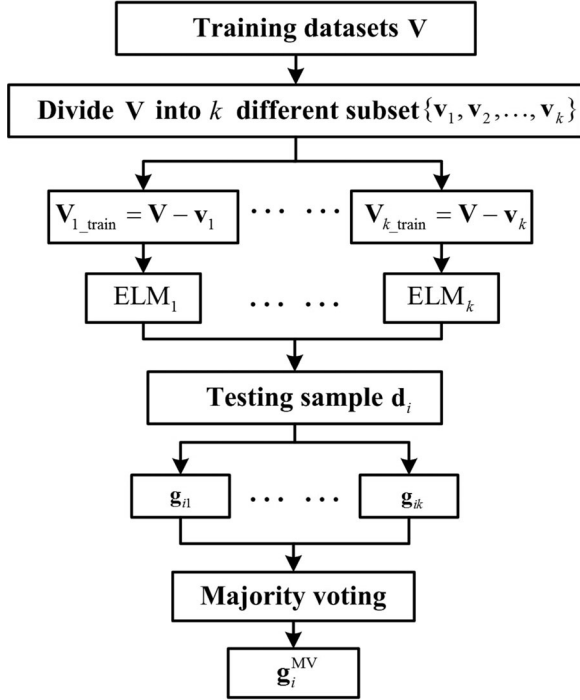


Fig. 5. Flowchart of an SE-ELM.

TABLE I
SIZE OF RAW SIDESCAN SONAR IMAGES AND THE COMPUTATION COST OF
MRF SEGMENTATION

Data Sets	Size (pixels)	Time for MRF segmentation (seconds)
Image 1	130 × 400	44.37
Image 2	70 × 500	37.56
Image 3	170 × 270	48.21
Image 4	150 × 380	51.23
Image 5	140 × 370	43.71
Image 6	200 × 420	80.08
Image 7	600 × 360	186.43
Image 8	860 × 660	602.21
Image 9	250 × 300	79.25

- 3) For testing data set \mathbf{D} , the number of samples of \mathbf{D} is s and $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_s\}$. Calculate the output \mathbf{g}_{ij} of \mathbf{d}_i using the output parameter \mathbf{c}_j of classifier ELM_j . After performing majority voting to $\mathbf{g}_i = [\mathbf{g}_{i1}^T, \dots, \mathbf{g}_{ij}^T, \dots, \mathbf{g}_{ik}^T]$, we get the output \mathbf{g}_i^{MV} of \mathbf{d}_i . Therefore, the output of \mathbf{D} is $\{\mathbf{g}_1^{\text{MV}}, \mathbf{g}_2^{\text{MV}}, \dots, \mathbf{g}_s^{\text{MV}}\}$.

The first two steps of an SE-ELM are based on k -fold cross validation while the third step is based on majority voting. The pseudocode of an SE-ELM is shown in Algorithm 1. Specifically, the original data set \mathbf{V} is divided into k subsets, of which $(k-1)$ subsets are randomly chosen to form a training data set. Then, we have k training data sets and each of them consists of M samples.

Let the k different classifiers be $\{\text{ELM}_1, \dots, \text{ELM}_j, \dots, \text{ELM}_k\}$ and the corresponding k different output parameters be $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_j, \dots, \mathbf{c}_k\}$. The output of sample \mathbf{d}_i of testing

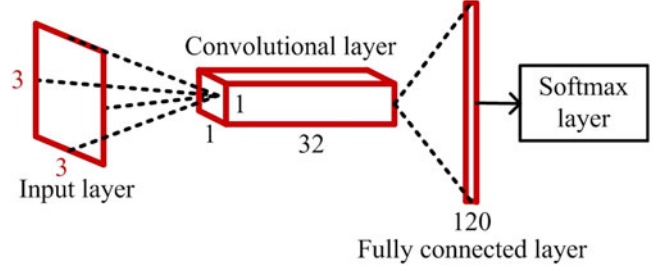


Fig. 6. Architecture of the CNN used in this paper.

Algorithm 1: Simple Ensemble Extreme Learning Machine.

Input: training data set \mathbf{V} , testing data set \mathbf{D} , the number of samples of \mathbf{D} is s , $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_s\}$, activation function $G(\cdot)$ of ELM.

Steps:

- 1: Divide the training set \mathbf{V} into k subsets: $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$.
 - 2: **for** $j = 1 : k$ **do**
 - 3: Set $\mathbf{V}_{j,\text{train}} = \mathbf{V} - \mathbf{v}_j$. Use $\mathbf{V}_{j,\text{train}}$ as the training data set and \mathbf{v}_j as the validation data set to get the optimal number of hidden nodes N_j of the classifier ELM_j .
 - 4: Train the classifier ELM_j to get output parameter \mathbf{c}_j .
 - 5: **end for**
 - 6: Save all output parameters $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$.
 - 7: **for** $i = 1 : s$ **do**
 - 8: **for** $j = 1 : k$ **do**
 - 9: Calculate the output \mathbf{g}_{ij} of \mathbf{d}_i using the output parameter \mathbf{c}_j of classifier ELM_j .
 - 10: **end for**
 - 11: Calculate the output \mathbf{g}_i^{MV} of \mathbf{d}_i by doing majority voting to $\mathbf{g}_i = [\mathbf{g}_{i1}^T, \dots, \mathbf{g}_{ij}^T, \dots, \mathbf{g}_{ik}^T]$.
 - 12: **end for**
- Output:** output of \mathbf{D} : $\{\mathbf{g}_1^{\text{MV}}, \mathbf{g}_2^{\text{MV}}, \dots, \mathbf{g}_s^{\text{MV}}\}$.
-

data set \mathbf{D} given by the j th classifier ELM_j is \mathbf{g}_{ij} . So the output of testing data \mathbf{d}_i given by the SE-ELM is $\mathbf{g}_i = [\mathbf{g}_{i1}^T, \dots, \mathbf{g}_{ij}^T, \dots, \mathbf{g}_{ik}^T]$. In this paper, we use a simple majority voting principle [36] to decide the final output of \mathbf{D} . For binary classification

$$\mathbf{g}_i^{\text{MV}} = \begin{cases} [1, 0]^T, & \text{if } \sum_{j=1}^k \mathbf{g}_i(1, j) \geq \frac{k}{2} \\ [0, 1]^T, & \text{if } \sum_{j=1}^k \mathbf{g}_i(2, j) > \frac{k}{2} \end{cases} \quad (8)$$

where \mathbf{g}_i^{MV} is the result of \mathbf{g}_i using majority voting, and $\mathbf{g}_i(1, j)$ is the j th element in the first row of the matrix \mathbf{g}_i . This principle can be extended to multiclass classification. For example, for the three-class classification, the output of the j th ($1 \leq j \leq k$) classifier ELM_j can be $[1, 0, 0]^T$, $[0, 1, 0]^T$, or $[0, 0, 1]^T$. If the number of the outputs $[0, 0, 1]^T$ is more than that of $[0, 1, 0]^T$ and that of $[1, 0, 0]^T$, then the output of the majority voting is $[0, 0, 1]^T$.

Majority voting is a very useful method and can usually obtain good results. But this method may go to a tie in binary classification or multiclass classification. In this paper, we solve the tie situations between different classes with the traditional k -nearest-neighbor method [37].

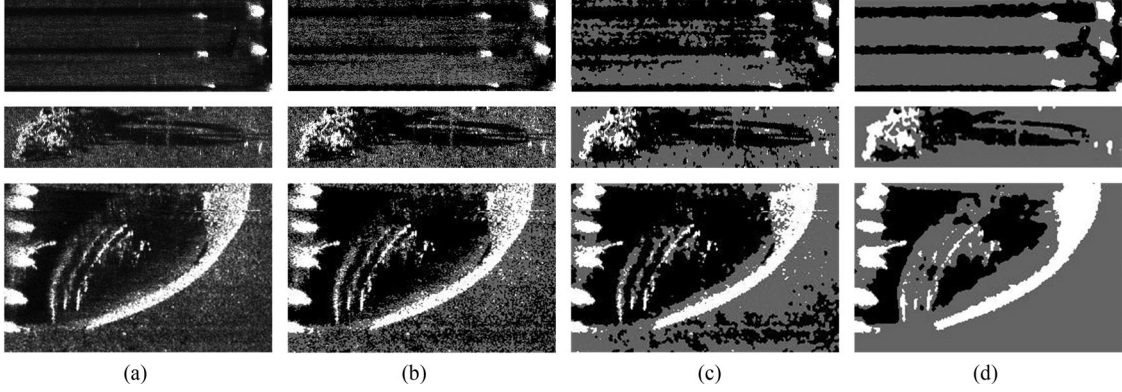


Fig. 7. (a) Raw sidescan sonar images. (b) Segmentation results of k -means clustering. (c) Segmentation results of MRF. (d) Corresponding ground truth images obtained using manual segmentation.

B. SE-ELM-Based Markov Random Field

As a supervised method, an SE-ELM can classify each pixel of sidescan sonar images with the learned model. Its performance depends mostly on the quality of training data sets and the generalization ability of an SE-ELM. Therefore, the ground truth of the training data is obtained using manual segmentation to reserve the details of the image and to reduce the influence of intensity inhomogeneity and noise. With such training data sets, the well-trained SE-ELM is less sensitive to noise and intensity inhomogeneity than k -means clustering. Therefore, we use the results of the SE-ELM to calculate the initialization parameters for MRF, which forms our proposed method SE-ELM-MRF. The final results are developed by ICE.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we describe the experimental setup and present the experimental results and analysis.

A. Data Sets for Performance Evaluation

Nine sidescan sonar images are used to evaluate the performance of the proposed method. The first seven images [as shown in Fig. 7(a) and the first four rows of Fig. 9(a)] are all obtained at the Yarkand River, Xinjiang Province, China. The eighth image [as shown in the fifth row of Fig. 9(a)] is obtained at Yantai, China. And the ninth image [as shown in the last row of Fig. 9(a)] is obtained at Qingdao, China. Each image is segmented in object-highlight, object-shadow, and background areas. The size of these images and the computation cost of MRF segmentation are shown in Table I.

Specifically, the first three images are used to form the training data set, while the rest six images are used to form the six testing data sets (data sets 4–9).

B. Parameters Setup

The experiments are carried out on the nine real sidescan sonar images using the Matlab software on a Windows 7 64-b system with Intel(R) Core(TM) i5-4210U CPU and 64 GB RAM. The number of classifiers in an SE-ELM is set to 10 and a tenfold cross-validation strategy is used to select the optimal number of hidden nodes ranging from 20 to 500.

The performance of the proposed classification method SE-ELM is compared with that of ELM, KELM [31], SVM [25], [26], and CNN. The optimal number of hidden nodes L of an ELM is 300 and the

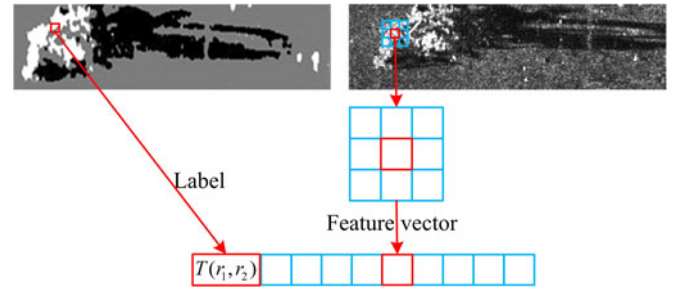


Fig. 8. Flowchart shows the way to get the training data set.

TABLE II
SEGMENTATION ACCURACY (%) OF k -MEANS CLUSTERING AND MRF

Data Sets for testing	k -means clustering	MRF
Data Set 4	69.52	74.77
Data Set 5	45.88	48.76
Data Set 6	60.35	61.63
Data Set 7	59.24	62.17
Data Set 8	79.21	84.85
Data Set 9	43.88	44.72

activation function $G(\cdot)$ of ELM, KELM, and SE-ELM is sigmoid. The regularization coefficient ϖ of the KELM is set to 1.

The architecture of the CNN is shown in Fig. 6. For fair comparison, the input images of the CNN are of size 3×3 . The output of the CNN is in the class of the central pixel of each input. As is shown in Fig. 6, this CNN includes one convolutional layer, one fully connected layer, and one softmax layer. The convolutional layer consists of 32 feature maps. Each feature map is connected to the input through filters of size 3×3 . The generated feature maps are of size 1×1 . The fully connected layer has 120 units, and each unit is connected to all units of its previous layer. The CNN is trained with data sets formed by the image patches from Fig. 7(a) and the corresponding manual segmentation ground truth shown in Fig. 7(d). There are 129 832 training

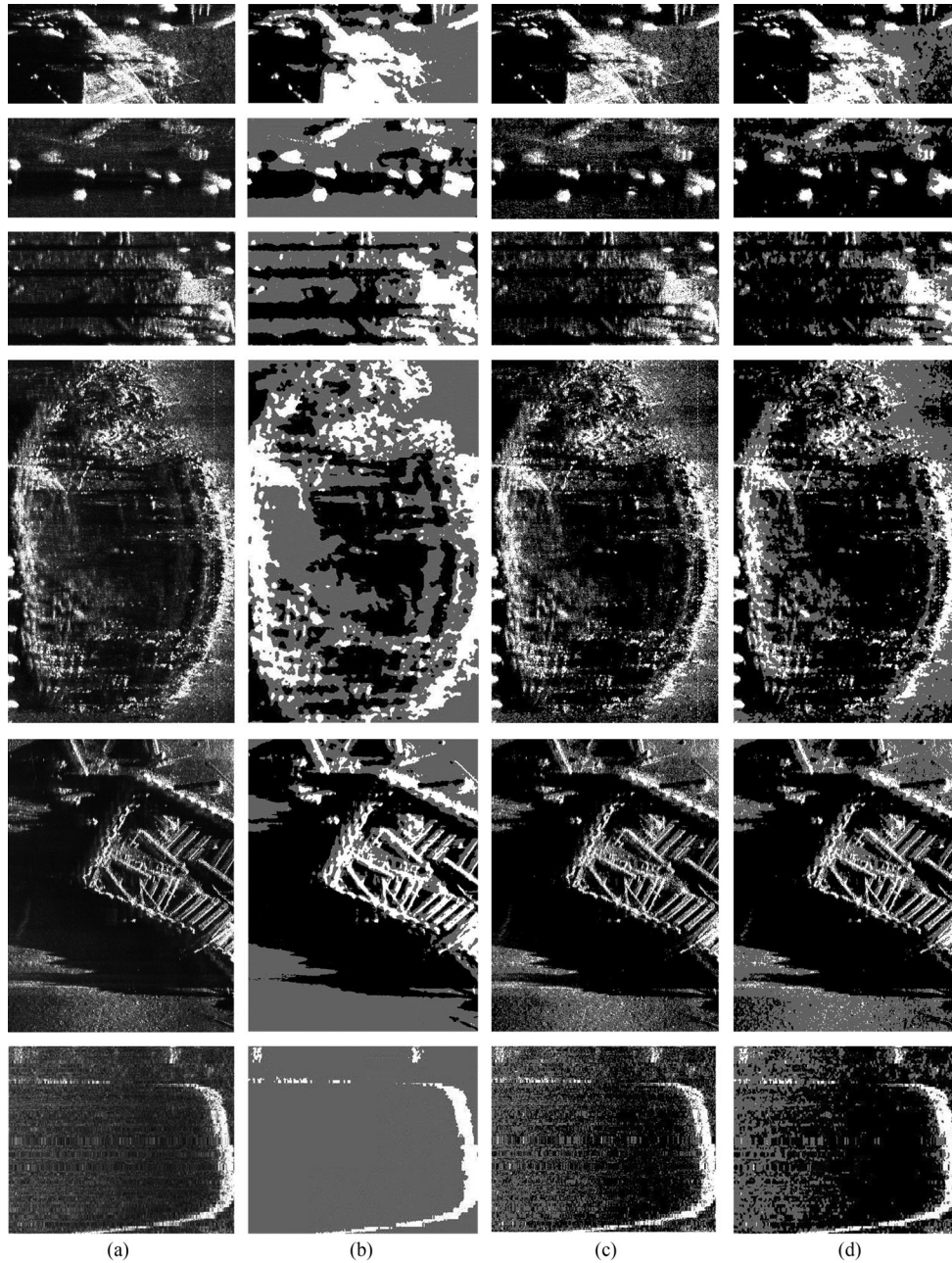


Fig. 9. Comparison of the segmentation results of k -means clustering and MRF on the six testing data sets. (a) Raw sidescan sonar images. (b) Ground truth. (c) Segmentation results of k -means clustering. (d) Segmentation results of MRF.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT MACHINE LEARNING METHODS
ON THE TRAINING DATA SET

Methods	Training accuracy	Training time (seconds)
ELM	90.42%	3.35
KELM	92.59%	742.06
SVM	90.63%	110.25
CNN	90.75%	5.90
SE-ELM	90.88%	13.00

samples $[(130 - 3) \times (400 - 3) + (70 - 3) \times (500 - 3) + (170 - 3) \times (270 - 3) = 129\,832]$ for CNN.

C. Experimental Results

The raw sonar images are shown in Fig. 7(a), the segmentation results of k -means clustering are shown in (b), and those of MRF are shown in (c). Fig. 7(d) presents the corresponding ground truth images.

In this paper, each feature vector consists of nine pixel values. As shown in Fig. 8, let the location of target pixel $T(r_1, r_2)$ in Fig. 7(d) be (r_1, r_2) , the pixels of local image patch around (r_1, r_2) in Fig. 7(a) of size 3×3 are selected as features. Then, the local image patch is reshaped into a row feature vector. The number of training samples used in this paper is roughly equal to the number of training image pixels.

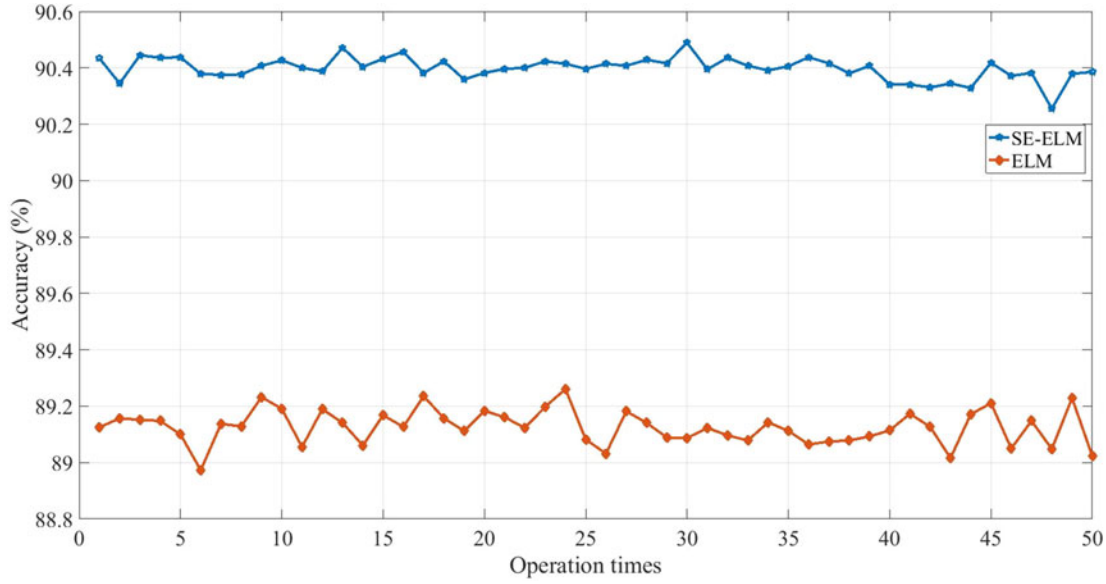


Fig. 10. Performance comparison of ELM and SE-ELM when they are operated on the validation data set 50 times, respectively.

TABLE IV
COMPARISON OF ELM, KELM, SVM, CNN, AND SE-ELM ON THE SIX TESTING DATA SETS

	Data Sets for testing	ELM	KELM	SVM	CNN	SE-ELM
Testing accuracy	Data Set 4	89.39%	85.13%	89.43%	89.32%	90.74%
	Data Set 5	87.42%	87.79%	86.20%	86.49%	91.19%
	Data Set 6	88.09%	85.35%	87.88%	87.95%	91.12%
	Data Set 7	85.84%	79.43%	86.56%	86.50%	89.09%
	Data Set 8	88.48%	73.79%	88.92%	88.88%	91.09%
	Data Set 9	98.03%	88.14%	97.69%	97.62%	98.40%
Testing time (seconds)	Data Set 4	0.23	73.29	87.82	0.73	0.84
	Data Set 5	0.22	66.17	85.42	0.60	0.73
	Data Set 6	0.36	110.08	100.61	1.01	1.14
	Data Set 7	0.85	283.21	332.04	2.88	3.00
	Data Set 8	0.98	572.30	356.81	6.49	7.46
	Data Set 9	0.18	96.67	163.35	0.86	0.97

Therefore, there are 129 832 training samples $[(130 - 2) \times (400 - 2) + (70 - 2) \times (500 - 2) + (170 - 2) \times (270 - 2) = 129\,832]$ for ELM, KELM, SVM, and SE-ELM. And the number of samples of each testing data set is roughly equal to the number of the pixels of each testing image. The six testing data sets consist of 55 944, 50 784, 82 764, 214 084, 564 564, and 73 904 samples, respectively.

1) *Segmentation Results of MRF*: Fig. 9 shows the comparisons of the segmentation results of k -means clustering and MRF on the six testing data sets. The ground truth is shown in Fig. 9(b) for comparison. Table II shows the segmentation accuracy of k -means clustering and MRF. We can see from Fig. 9 and Table II that the results of k -means clustering are noisy, which may result in high computation cost and

bad performance of MRF segmentation. So it is essential to find another initialization method that can provide a segmentation result more similar to the ICE convergence than k -means clustering.

2) *Segmentation Results of SE-ELM*: To verify the effectiveness of the SE-ELM, 10% of the samples in the original training data set are utilized as validation data set, while the rest is used as training data set. Fig. 10 shows the experimental results of an ELM and SE-ELM when they are operated 50 \times respectively on the validation data set. We can see from Fig. 10 that the classification result of an SE-ELM is more stable and better than that of an ELM. Then, an SE-ELM is compared with other machine learning methods on the training data set (as shown in Table III) and the six testing data sets (as shown in Table IV). We can

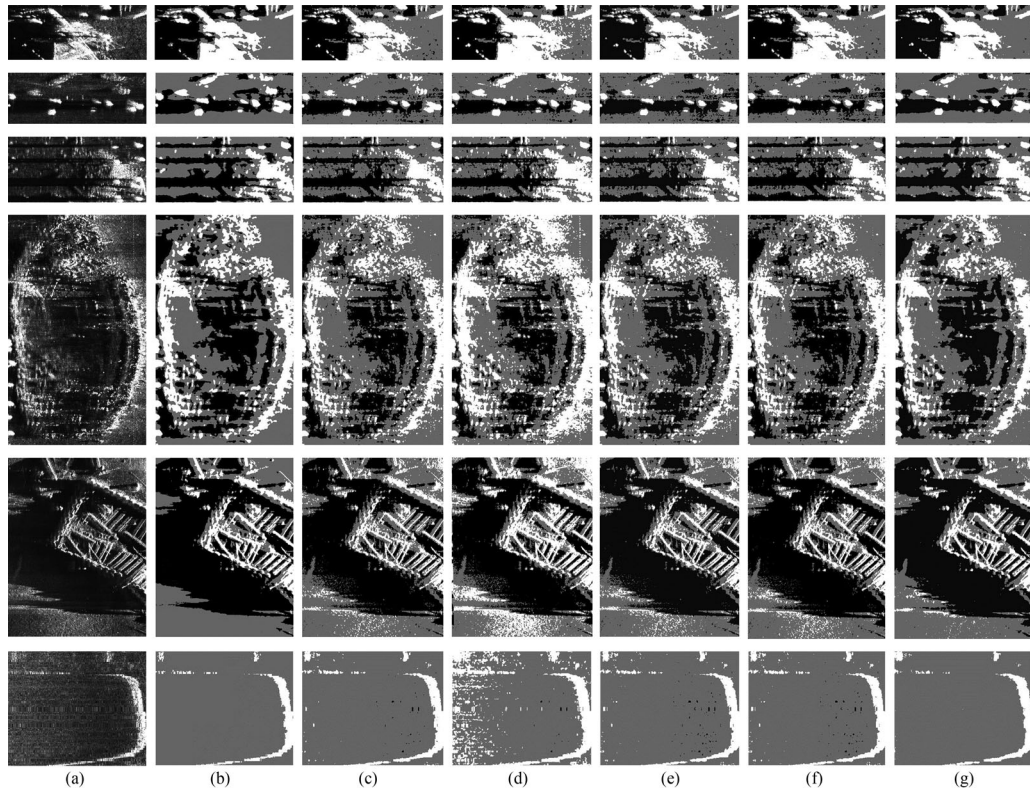


Fig. 11. Comparison of the segmentation results of ELM, KELM, SVM, CNN, and SE-ELM. (a) Raw sidescan sonar images. (b) Ground truth. (c) Segmentation results of ELM. (d) Segmentation results of KELM. (e) Segmentation results of SVM. (f) Segmentation results of CNN. (g) Segmentation results of SE-ELM.

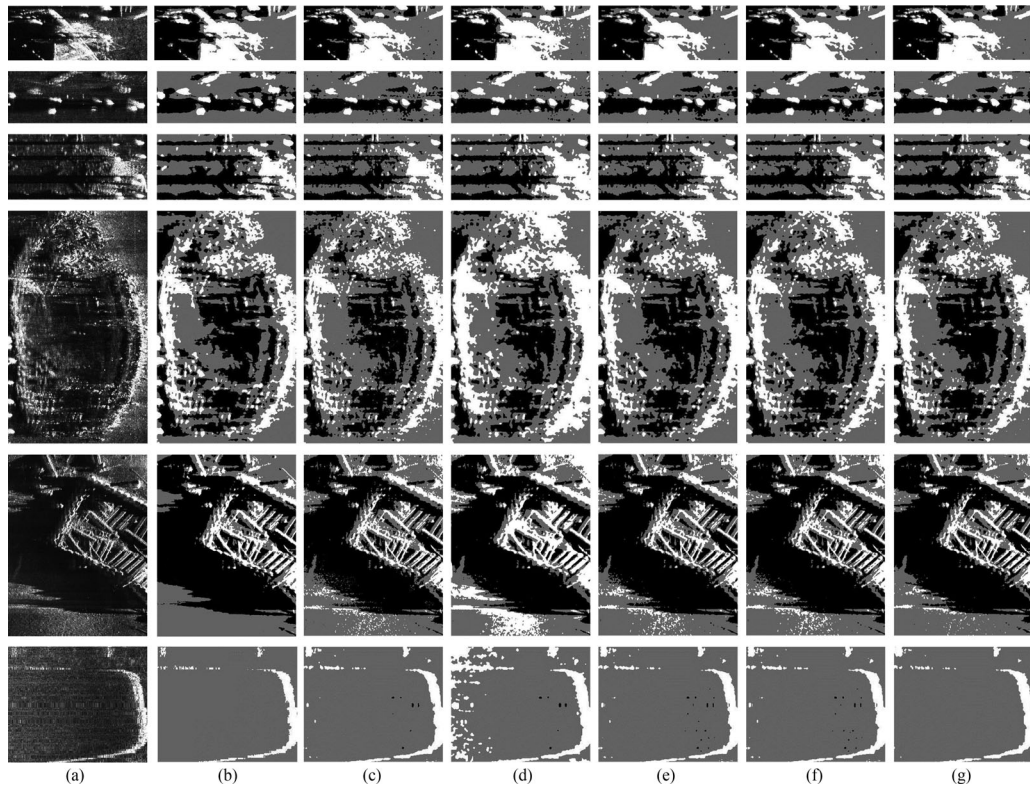


Fig. 12. Comparison of the methods using MRF postprocessing. (a) Raw sidescan sonar images. (b) Ground truth. (c) Segmentation results of ELM and MRF. (d) Segmentation results of KELM and MRF. (e) Segmentation results of SVM and MRF. (f) Segmentation results of CNN and MRF. (g) Segmentation results of SE-ELM and MRF.

TABLE V
COMPARISON OF METHODS COMBINING MRF WITH ELM, KELM, SVM, CNN, AND SE-ELM, RESPECTIVELY, ON THE SIX TESTING DATA SETS

	Data Sets for testing	ELM and MRF	KELM and MRF	SVM and MRF	CNN and MRF	SE-ELM and MRF
Testing accuracy	Data Set 4	90.15%	86.36%	90.02%	90.13%	91.72%
	Data Set 5	89.30%	88.41%	89.36%	89.50%	91.99%
	Data Set 6	89.75%	86.52%	90.14%	90.17%	91.88%
	Data Set 7	87.85%	82.16%	88.60%	88.68%	89.90%
	Data Set 8	89.48%	81.16%	89.95%	89.94%	91.53%
	Data Set 9	98.13%	90.64%	97.88%	97.80%	98.52%
Testing time (seconds)	Data Set 4	1.07	74.19	88.65	1.65	1.67
	Data Set 5	1.06	66.99	86.30	1.48	1.56
	Data Set 6	1.46	111.08	101.63	2.06	2.11
	Data Set 7	2.64	285.04	333.86	4.66	4.82
	Data Set 8	5.02	657.30	360.92	10.63	11.59
	Data Set 9	1.11	97.64	164.33	1.91	1.95

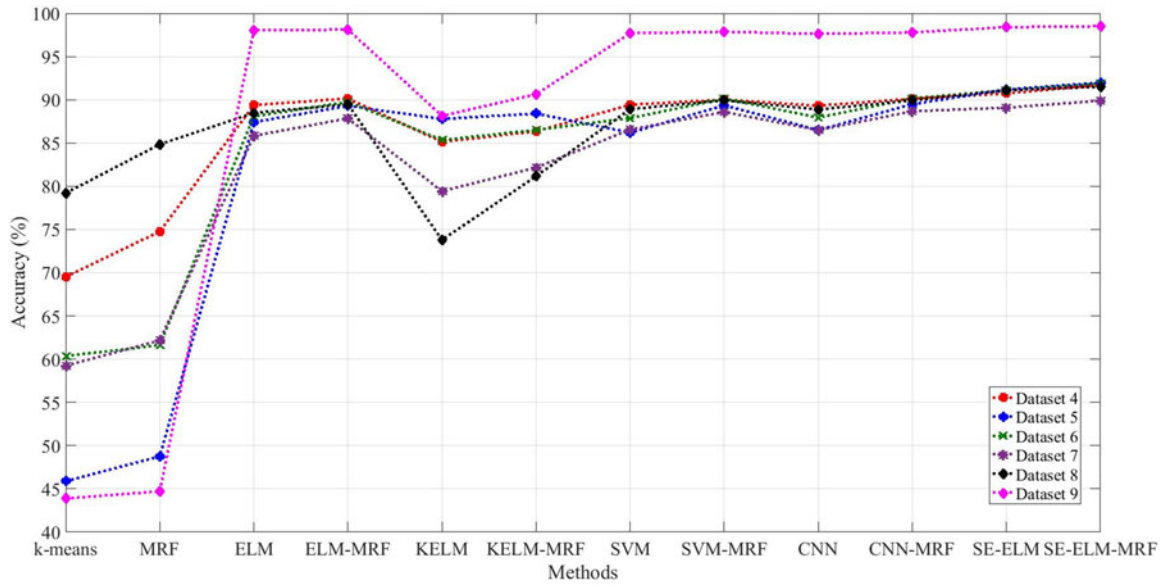


Fig. 13. Comparison of the methods used in the experiments.

see from Table IV that an SE-ELM performs better than other algorithms in terms of segmentation accuracy on each testing data set. And the computational costs of an SE-ELM on the testing data sets (from data set 4 to data set 9) are much lower than those of SVM and KELM. We also present the segmentation results of SE-ELM, ELM, KELM, SVM, and CNN, as shown in Fig. 11. Fig. 11 indicates that an SE-ELM can segment sidescan sonar images better when compared with other methods.

3) *Segmentation Results of MRF Postprocessing*: Fig. 12 shows the segmentation results of the methods combining ELM,

KELM, SVM, CNN, and SE-ELM with MRF, respectively. Fig. 13 presents the performance comparison of the methods used in our experiments, and Table V provides the segmentation accuracy and time consumption of the methods combining ELM, KELM, SVM, CNN, and SE-ELM with MRF, respectively. Fig. 14 gives the accuracy improvement of each method after and before MRF postprocessing. In this paper, SE-ELM-MRF stands for the method using an SE-ELM to initialize MRF, which means that the segmentation results of an SE-ELM provide initial models for MRF. It can be clearly seen from Fig. 12 that the results of SE-ELM-MRF have more smooth

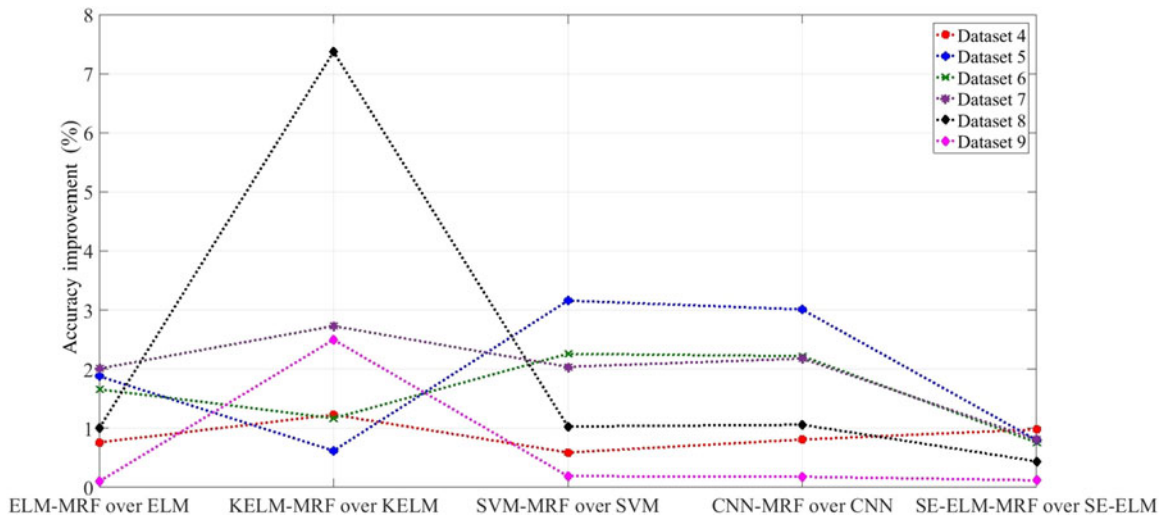


Fig. 14. Accuracy improvement after and before MRF postprocessing. “ELM-MRF over ELM” stands for the performance improvement of ELM-MRF when it is compared with ELM.

edges and correct details than those of other methods. Moreover, Table V and Fig. 13 indicate that SE-ELM-MRF has better segmentation performance than other methods and converges faster than MRF. Furthermore, Fig. 14 shows that MRF postprocessing gives performance improvement for each method.

4) *Generalization Performance Analysis of SE-ELM and SE-ELM-Based MRF*: To prove the generalization performance of an SE-ELM and SE-ELM-MRF, we use sidescan sonar images obtained from different seabeds (data sets 8 and 9) to test the SE-ELM and SE-ELM-MRF and compare their results, as shown in Figs. 11–14 and Table V. These two sonar images contain much noise and intensity inhomogeneity, as shown in Fig. 11(a). Qualitative results and quantitative results are given in Figs. 12 and 13, respectively. We can see from Fig. 12 and Table V that even on data sets obtained from different seabeds, an SE-ELM and SE-ELM-MRF can provide better outcomes in terms of accuracy than other methods. Therefore, an SE-ELM and SE-ELM-MRF have superior generalization performance and are robust against different seabed types.

V. CONCLUSION

In this paper, a novel method SE-ELM-MRF based on an SE-ELM and MRF is presented for sidescan sonar image segmentation. To validate the performance of the SE-ELM and SE-ELM-MRF, nine real sidescan sonar images collected by Marine Sonic, Yorktown, VA, USA, are used in our experiments. Furthermore, to prove the generalization performance of SE-ELM-MRF, six sidescan sonar images obtained from three different seabeds are used. Experimental results indicate that an SE-ELM shows better performance than ELM, KELM, SVM, and CNN in sidescan sonar segmentation. Besides, using an SE-ELM as the initial segmentation algorithm instead of k -means clustering, SE-ELM-MRF outperforms MRF in segmentation accuracy and convergence speed. The generalization performance of SE-ELM-MRF are verified through the good segmentation results on sonar images of different places.

Future work includes using sidescan sonar images with more intensity inhomogeneity than the images used in this paper to train SE-ELM and test whether it can be more robust. Another direction of future work might be to use the segmentation results to aid the navigation and path planning of AUVs.

REFERENCES

- [1] M. Klein, “Side scan sonar,” in *International Handbook of Underwater Archaeology*. New York, NY, USA: Springer-Verlag, 2002, pp. 667–678.
- [2] E. Kriminger, J. T. Cobb, and J. C. Principe, “Online active learning for automatic target recognition,” *IEEE J. Ocean. Eng.*, vol. 40, no. 3, pp. 583–591, Jul. 2015.
- [3] W. K. Stewart, M. Jiang, and M. Marra, “A neural network approach to classification of sidescan sonar imagery from a midocean ridge area,” *IEEE J. Ocean. Eng.*, vol. 19, no. 2, pp. 214–224, Apr. 1994.
- [4] M. Lianantonakis and Y. R. Petillot, “Sidescan sonar segmentation using texture descriptors and active contours,” *IEEE J. Ocean. Eng.*, vol. 32, no. 3, pp. 744–752, Jul. 2007.
- [5] S. Reed, Y. Petillot, and J. Bell, “An automatic approach to the detection and extraction of mine features in sidescan sonar,” *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, Jan. 2003.
- [6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 226–231.
- [7] M. D. Escobar and M. West, “Bayesian density estimation and inference using mixtures,” *J. Amer. Stat. Assoc.*, vol. 90, no. 430, pp. 577–588, 1995.
- [8] Y. Cheng, “Mean shift, mode seeking, and clustering,” *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
- [9] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, “Three-class Markovian segmentation of high-resolution sonar images,” *Comput. Vis. Image Understanding*, vol. 76, no. 3, pp. 191–204, 1999.
- [10] F. Salzenstein and W. Pieczynski, “Unsupervised Bayesian segmentation using hidden Markovian fields,” in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 1995, vol. 4, pp. 2411–2414.
- [11] S. Reed, I. T. Ruiz, C. Capus, and Y. Petillot, “The fusion of large scale classified side-scan sonar image mosaics,” *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2049–2060, Jul. 2006.
- [12] S. Reed, Y. Petillot, and J. M. Bell, “Unsupervised mine detection and analysis in side-scan sonar: A comparison of Markov random fields and statistical snakes,” *IEEE J. Quantum Electron.*, vol. 40, no. 4, pp. 354–363, 2001.
- [13] M. Jiang, W. Stewart, and M. Marra, “Segmentation of seafloor sidescan imagery using Markov random fields and neural networks,” in *Proc. OCEANS, Eng. Harmony Ocean*, 1993, pp. III456–III461.
- [14] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, “Markov random field and fuzzy logic modeling in sonar imagery: Application to the classification of underwater floor,” *Comput. Vis. Image Understanding*, vol. 79, no. 1, pp. 4–24, 2000.
- [15] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [16] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks,” *IEEE Trans. Neural Networks*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.

- [17] Y. Lan, Y. C. Soh, and G.-B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing*, vol. 72, no. 13, pp. 3391–3395, 2009.
- [18] X. Wang and M. Han, "Online sequential extreme learning machine with kernels for nonstationary time series prediction," *Neurocomputing*, vol. 145, pp. 90–97, 2014.
- [19] A. Samat, P. Gamba, P. Du, and J. Luo, "Active extreme learning machines for quad-polarimetric SAR imagery classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 35, pp. 305–319, 2015.
- [20] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E²LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
- [21] W. Huang *et al.*, "Liver tumor detection and segmentation using kernel-based extreme learning machine," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 3662–3665.
- [22] W. Huang *et al.*, "A semi-automatic approach to the segmentation of liver parenchyma from 3d CT images with extreme learning machine," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2012, pp. 3752–3755.
- [23] G.-B. Huang, Z. Bai, L. L. C. Kasun, and C. M. Vong, "Local receptive fields based extreme learning machine," *IEEE Comput. Intell. Mag.*, vol. 10, no. 2, pp. 18–29, May 2015.
- [24] Y. Zeng, X. Xu, and Y. Fang *et al.*, "Traffic sign recognition using extreme learning classifier with deep convolutional features," *Int. Conf. Intell. Sci. Big Data Eng. (ISIDE 2015)*, Suzhou, China, vol. 9242, pp. 272–280, 2015.
- [25] C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [26] J. Weston and C. Watkins, "Multi-class support vector machines," Roy. Holloway Univ. London, Surrey, U.K., Tech. Rep. CSD-TR-98-04, 1998.
- [27] R. Maclin and D. Opitz, "An empirical evaluation of bagging and boosting," in *Proc. 14th Nat. Conf. Artif. Intell./9th Conf. Innovative Appl. Artif. Intell.*, 1997, pp. 546–551.
- [28] W. Cai, R. Nian, B. He, and A. Lendasse, "A fast sonar-based benthic object recognition model via extreme learning machine," in *Proc. OCEANS*, 2015, pp. 1–4.
- [29] Y. LeCun and T. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361. Boston, MA, USA: MIT Press, 1995.
- [30] 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.
- [31] G.-B. 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.
- [32] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014.
- [33] T. Celik and T. Tjahjadi, "A novel method for sidescan sonar image segmentation," *IEEE J. Ocean. Eng.*, vol. 36, no. 2, pp. 186–194, Apr. 2011.
- [34] G. Huo, S. X. Yang, Q. Li, and Y. Zhou, "A robust and fast method for sidescan sonar image segmentation using nonlocal despeckling and active contour model," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 855–872, Apr. 2017.
- [35] J. G. Moreno-Torres, J. A. Sáez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k -fold cross-validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, Aug. 2012.
- [36] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Inf. Fusion*, vol. 6, no. 1, pp. 63–81, 2005.
- [37] H. Joutsijoki and M. Juhola, "Kernel selection in multi-class support vector machines and its consequence to the number of ties in majority voting method," *Artif. Intell. Rev.*, vol. 40, no. 3, pp. 213–230, 2013.



Yan Song received the M.S. degree in electronic engineering from Ocean University of China, Qingdao, China, in 2015, where she is currently working toward the Ph.D. degree.

Her research interests include machine learning and sonar image processing.



Bo He received the M.S. degree in inertial technology and navigation equipment and the Ph.D. degree in control theory and control engineering from Harbin Institute of Technology, Harbin, China, in 1996 and 1999, respectively.

From 2000 to 2003, he was a Postdoctoral Fellow at Nanyang Technological University, Singapore, where he worked on mobile robots and unmanned vehicles, including his research works on precise navigation, control, and communication. In 2004, he joined the Ocean University of China, Qingdao, China, where he is currently a full Professor. His research interests include AUV design and applications, AUV SLAM, AUV control, and machine learning.



Ying Zhao received the B.S. degree in electronic engineering from Ocean University of China, Qingdao, China, in 2014, where she is currently working toward the M.S. degree.

Her research interests include image processing and machine learning.



Guangliang Li (M'14) received the B.S. and M.Sc. degrees from the School of Control Science and Engineering, Shandong University, Jinan, China, in 2008 and 2011, respectively, and the Ph.D. degree in computer science from the University of Amsterdam, Amsterdam, The Netherlands, in 2016.

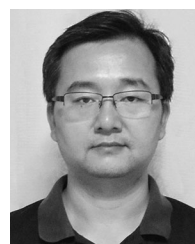
He was a Visiting Researcher at the Delft University of Technology, Delft, The Netherlands, and a Research Intern with Honda Research Institute Japan, Co., Ltd., Japan. He is currently a Lecturer at the Ocean University of China, Qingdao, China. His research interests include reinforcement learning, human agent/robot interaction, and robotics.

search interests include reinforcement learning, human agent/robot interaction, and robotics.



Qixin Sha received the M.S. degree in electronic engineering from Ocean University of China, Qingdao, China, in 2010.

He is currently an Engineer at the Department of Electronic Engineering, Ocean University of China. His research interests include software development, navigation, and control of AUV.



Yue Shen received the M.S. degree in electrical engineering from the Luoyang Institute of Technology, Luoyang, China, in 1998 and the Ph.D. degree in mechanical and electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2003.

He is currently an Associate Professor at the College of Information Science and Engineering, Ocean University of China, Qingdao, China.



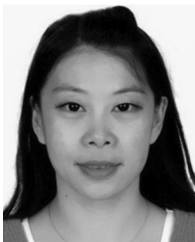
Tianhong Yan received the Ph.D. degree in aerospace engineering and mechanics from the School of Aeronautics and Astronautics, Harbin Institute of Technology, Harbin, China, in March 1999.

He was with the Institute of Micro-Satellite Engineering, Chinese Academy of Sciences, Beijing, China, working on the system design, mechanics analysis, and optimization of large flexible structures used in aerospace engineering. From 2000 to 2004, he was working on dynamics and control for high-speed high-precision mechatronics systems, e.g., hard disk drives, back-end semiconductor equipment, with the Nanyang Technological University of Singapore and the R&D Centre of ASM Pacific, Singapore. From 2004 to 2008, he was the Senior System Engineer and an Assistant Manager at the China National Lithography Tool Engineering Centre, Shanghai Micro Electronics Equipment, Co., Ltd., working on front-end semiconductor equipment, i.e., lithography tool. He is the main scientist for the national key grand projects of the National 863 plan and the 10th five years plan. Since 2009, he has been with China Jiliang University, Hangzhou, China, as a full Professor. His research interests include dynamics and servo control of high-speed and high-precision mechatronics systems in large or complex equipment, and the autonomous underwater vehicles.



Amaury Lendasse was born in 1972 in Belgium. He received the M.S. degree in mechanical engineering, the M.S. degree in control, and the Ph.D. degree in applied mathematics, all from the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1996, 1997, and 2003, respectively.

In 2003, he was a Postdoctoral Researcher at the Computational Neurodynamics Lab, University of Memphis. From 2004 to 2014, he was a Senior Researcher and an Adjunct Professor with the Adaptive Informatics Research Centre, School of Science, Aalto University (better known as the Helsinki University of Technology), Espoo, Finland. He has created and lead the Environmental and Industrial Machine Learning Group at Aalto. He is currently an Associate Professor with The University of Iowa, Iowa City, IA, USA, and a Visiting Professor with the Arcada University of Applied Sciences, Helsinki, Finland. He was the Chairman of the Annual European Symposium on Time Series Prediction Conference and a member of the editorial board and program committee of several journals and conferences on machine learning. He is the author or coauthor of more than 200 scientific papers in international journals, books, or communications to conferences with reviewing committee. His research interests include Big Data, time-series prediction, chemometrics, variable selection, noise variance estimation, determination of missing values in temporal databases, nonlinear approximation in financial problems, and functional neural networks and classification.



Rui Nian received the B.S. degree in signal and information processing from Ocean University of China (OUC), Qingdao, China, in 2002 and the Ph.D. degree in marine information detection and processing and electronic technology from both OUC and the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 2007 and 2009, respectively.

She is an Assistant Professor at the Information Science and Engineering College, OUC. Her primary research interests include pattern recognition, computer vision, image processing, cognitive science, machine learning, and high-dimensional space analysis and her major interest focusses on ocean information detection and processing.