A Novel Approach for SAR Target Detection Based on Unsupervised Complex-Valued Extreme Learning Machine

Qinglong Hua,[1] Yun Zhang,[1] and Yicheng Jiang[1]
[1]*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, 150001, China*
Email: zhangyunhit@hit.edu.cn

Strong clutter seriously affects target-of-interest detection in synthetic aperture radar (SAR) images. This letter proposes an unsupervised target detection method (U-TDM) based on a complex-valued extreme learning machine (CV-ELM), the essence of which is to transform the problem of target detection into a pixel binary classification problem. The SAR image is first divided into several unlabeled patches, and fuzzy c-means (FCM) is used to construct the reference target patch set and the clutter patch set. Based on these two patch sets, CV-ELM is used to classify the neighboring patch of the pixel to be detected. Since the pixel intensity and distribution of target-of-interest and clutter are different, unsupervised pixel classification could be realized without ground-truth through U-TDM. Experimental results on GF-3 data and Sentinel-1 data show the efficiency of the proposed method in target detection with a heterogeneous clutter environment.

Introduction: Synthetic aperture radar (SAR) is an active microwave sensor that can perform all-time and all-weather imaging without being affected by light and climate [1]. SAR images are usually composed of target-of-interest and clutter. The strong sea clutter or ground clutter increases the difficulty of target detection. The quality of target detection dramatically affects the performance of subsequent SAR image processing. Accurate detection results are conducive to more accurately identifying the target's category and superstructure information.

After decades of development, many detection methods for SAR images have been proposed. The classic constant false alarm rate (CFAR) detector [2], [3] relies on contrast or brightness features to search for targets, which may result in a large number of clutter false alarms in the complex background of strong clutter. Moreover, since moving targets and stationary clutter have different Doppler parameter characteristics, spectral filter and time-frequency analysis could be used to detect moving targets in a single-channel SAR system. Huang *et al.* [4] proposed a time-Doppler chirp-varying (TDCV) filter for moving target detection and parameter estimation with the airborne system. The multi-channel SAR system could increase the spatial information of the echo and improve the ability to observe moving targets [5]. These methods have higher requirements for the consistency of system channels. The extended fractal (EF) method [6] is another representative research that fully integrates the intensity information and the spatial distribution information. It uses the spatial difference between the target and the clutter in reflected energy to detect the target. This EF-based method provides a good compromise between performance and speed.

Deep learning has shown excellent performance in SAR detection in recent years. Deep learning-based methods, such as ConvNets [7], SSD [8], and Faster-RCNN [9], are accomplished by using the gradient descent (GD) method to tune their large number of parameters, which have the advantage of high accuracy. Nevertheless, deep network training depends heavily on sufficient labeled samples, and the SAR images are time-consuming to annotate labels artificially. Some literature [10], [11] develops semi-supervised or unsupervised learning approaches to enhance the capacity of deep learning models. The extreme learning machine (ELM) [12] provides a faster convergence training process because the hidden layer parameters do not need to be tuned, and the output weights are analytically determined. ELM's low computational complexity and better generalization in regression tasks provide a new direction for our research.

Since SAR images belong to the complex domain, and ELM cannot process complex information, it is necessary to extend ELM to form complex-valued ELM (CV-ELM). This letter proposes an unsupervised target detection method (U-TDM) based on CV-ELM and transforms the target detection problem into a pixel binary classification problem. In U-TDM, fuzzy c-means (FCM) [13] is first used to construct the reference target and clutter patch set. Then target patches and clutter patches randomly selected from these two patch sets are classified by the CV-ELM together with the neighborhood patch of the pixel to be detected. Since the target and clutter are different in pixel intensity and distribution, unsupervised pixel classification could be achieved without actual labels or ground-truth. Experiments on GF-3 and Sentinel-1 data verify the effectiveness of the U-TDM.

U-TDM: The implementation details of the proposed unsupervised target detection method are shown in Fig. 1. Next, we elaborate on the U-TDM from three perspectives: reference patch sets, unsupervised strategy, and CV-ELM.

Construction of Reference Patch Sets: Before target detection in a SAR image $\mathbf{S}$, two reference patch set, target patch set $\aleph_{\mathbf{T}}$ and clutter patch set $\aleph_{\mathbf{C}}$, are necessary to be constructed. First, $\mathbf{S}$ is uniformly cropped into a series of patches of size $(2d+1)\times(2d+1)$, where d is the scale factor. These unlabeled patches are classified into target and clutter by FCM. The target patches with membership greater than the threshold T_{cof} constitute the target patch set $\aleph_{\mathbf{T}} = \{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_{N_t}\}$ and the clutter patches with membership greater than the threshold T_{cof} form the clutter patch set $\aleph_{\mathbf{C}} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_{N_c}\}$. N_t and N_c are the number of the target patch set and the clutter patch set, respectively. It should be noted that only background clutter is included in the clutter patch. The target patch contains both the target-of-interest and the background clutter. The reference patch sets could be regarded as the assisted knowledge used in U-TDM. Based on the reference patch sets, construct an unsupervised strategy in the next.

Unsupervised Strategy: Consider a complex SAR image $\mathbf{S} \in \mathbb{C}^{M\times N}$. M is the number of rows, and N is the number of columns. $s(m,n)$ represents the m-th row and n-th column of $\mathbf{S}$. Define $\mathbf{Q} \in \mathbb{C}^{M\times N}$ as the ground-truth of $\mathbf{S}$. Define $q(m,n)$ as the m-th row and n-th column of $\mathbf{Q}$. $q(m,n)$ also represents the category of $s(m,n)$. If the category is the target, then $q(m,n) = 1+1j$, otherwise $q(m,n) = 0$.

For the pixel $s(m,n)$ in the SAR image $\mathbf{S}$, it is difficult to effectively determine its category based on the information of a single pixel. Generally, the neighborhood of the pixel $s(m,n)$ has a high degree of correlation with $s(m,n)$, so it is feasible to use the neighborhood information of $s(m,n)$ to determine the category of $s(m,n)$. Define the neighborhood of $s(m,n)$ as $\mathbf{NH}(m,n)$.

$$\begin{aligned}&\mathbf{NH}(m,n)\\ &= \left\{s(m-d,n-d), \ldots s\left(x_j,y_j\right), \ldots, s(m+d,n+d)\right\}\end{aligned} \tag{1}$$

where $s\left(x_j,y_j\right)$ is the j-th pixel of $\mathbf{NH}(m,n)$, $1 < m < M$, $1 < n < N$. The number of pixels in $\mathbf{NH}(m,n)$ is $N_s = (2d+1)^2$.

Define two complex-valued functions $f^k(\cdot)$, $k = 1, 2$. When vectorized $\mathbf{NH}(m,n)$ is input, the output of $f^k(\cdot)$ is $q(m,n)$. When the vectorized reference target patch is input, the output of $f^k(\cdot)$ is $1+1j$. When the vectorized reference clutter patch is input, the output of $f^k(\cdot)$ is 0. Then the function $f^k(\cdot)$ could be expressed as

$$f^k_{m,n} = \min_{f^k(\cdot)} \left\|\mathbf{L}^k_i(m,n) - f^k\left(vec\left(\mathbf{P}^k_i(m,n)\right)\right)\right\|^2_F \tag{2}$$

where $\left\{\mathbf{P}^1_i \in \mathbb{C}^{(2d+1)\times(2d+1)}\right\}_{i=1}^{N_l+1} = \{\mathbf{NH}(m,n), \mathbf{T}_1, \ldots, \mathbf{T}_{N_l}\}$ is the input of $f^1_{m,n}(\cdot)$, and $\left\{\mathbf{L}^1_i \in \mathbb{C}^1\right\}_{i=1}^{N_l+1} = \{q(m,n), 1+1j, \ldots, 1+1j\}$ is the output of $f^1_{m,n}(\cdot)$. $\left\{\mathbf{P}^2_i\right\}_{i=1}^{N_l+1} = \{\mathbf{NH}(m,n), \mathbf{C}_1, \ldots, \mathbf{C}_{N_l}\}$ is the input of $f^2_{m,n}(\cdot)$, and $\left\{\mathbf{L}^2_i\right\}_{i=1}^{N_l+1} = \{q(m,n), 0, \ldots, 0\}$ is the output of $f^2_{m,n}(\cdot)$. $\|\cdot\|_F$ denotes the Frobenius norm. N_l represents the number of patches randomly selected from the reference patch sets. $vec\left(\mathbf{P}^k_i\right)$ represents vectorized $\mathbf{P}^k_i$. Each row of data in $\mathbf{P}^k_i$ is connected end to end to obtain a single row vector.

Based on the reference patches, the classification problem of SAR image pixels is converted to solve a group of functions as formula (6). The classification result of the SAR image $\mathbf{S}$ could be obtained by integrating $\hat{\mathbf{Q}}^1$ and $\hat{\mathbf{Q}}^2$, that is, $\hat{\mathbf{Q}} = \hat{\mathbf{Q}}^1 \odot \hat{\mathbf{Q}}^2$. $\odot$ represents the Hadamard

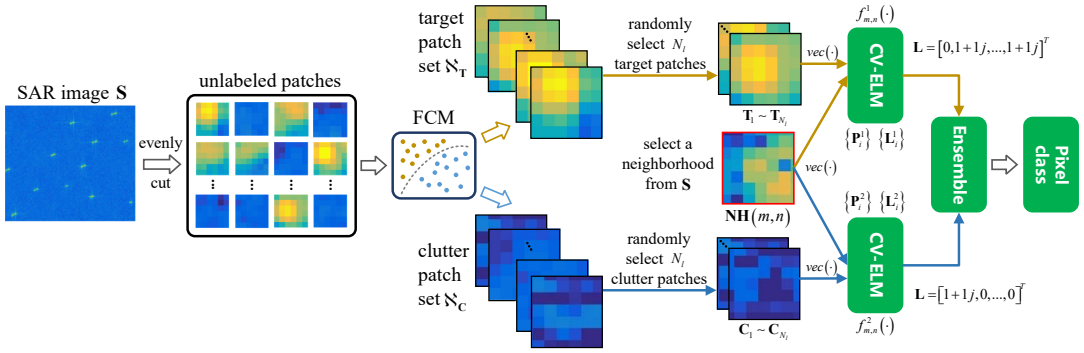

Fig 1 *The framework of target detection based on U-TDM.*

product.

$$\hat{\mathbf{Q}}^k = \begin{bmatrix} f^k_{1,1}\left(\mathbf{P}^k_i(1,1)\right) & \cdots & f^k_{1,N}\left(\mathbf{P}^k_i(1,N)\right) \\ \vdots & \ddots & \vdots \\ f^k_{M,1}\left(\mathbf{P}^k_i(M,1)\right) & \cdots & f^k_{M,N}\left(\mathbf{P}^k_i(M,N)\right) \end{bmatrix} \tag{3}$$

In supervised learning algorithms, knowing the value of $q(m,n)$ in advance is necessary before learning and training to solve the formula (6). But in the actual application, it isn't easy to obtain the value of $q(m,n)$ directly. Therefore, this letter adopts a unsupervised training strategy. In this strategy, whether the central point pixel $s(m,n)$ of $\mathbf{NH}(m,n)$ belongs to the target or the clutter, $q(m,n)$ in $\left\{\mathbf{L}^1_i\right\}$ is set to 0. $q(m,n)$ in $\left\{\mathbf{L}^2_i\right\}$ is set to $1+1j$. Next, the rationality of this strategy is explained by taking $f^2_{m,n}(\cdot)$, $\left\{\mathbf{P}^2_i\right\}$, and $\left\{\mathbf{L}^2_i\right\}$ as examples. If the central point pixel $s(m,n)$ of $\mathbf{NH}(m,n)$ belongs to the target-of-interest, the actual value of $q(m,n)$ in $\left\{\mathbf{L}^2_i\right\}$ is $1+1j$. Since the target and clutter have certain differences in pixel intensity and distribution, $f^2_{m,n}(\cdot)$ could classify $\mathbf{NH}(m,n)$ and reference clutter patches randomly selected from $\aleph_C$ well. If the central point pixel $s(m,n)$ of $\mathbf{NH}(m,n)$ belongs to the clutter, the actual value of $q(m,n)$ in $\left\{\mathbf{L}^2_i\right\}$ is 0. Although $q(m,n)$ is forcibly set to $1+1j$, When $f^2_{m,n}(\cdot)$ is limited in fitting ability and there are enough clutter patches, $f^2_{m,n}(\cdot)$ has difficulty separating $\mathbf{NH}(m,n)$ from reference clutter patches and would tend to classify $\mathbf{NH}(m,n)$ into the clutter category. Therefore, even if the label $q(m,n)$ is set to $1+1j$ when $s(m,n)$ belongs to the clutter, the correct classification result could be obtained. Similarly, setting $q(m,n)$ to 0 in $\left\{\mathbf{L}^1_i\right\}$ is also reasonable.

CV-ELM: For the above mapping function $f^k_{m,n}(\cdot)$, CV-ELM proposed in this letter is used to solve it, which is a complex-valued version of ELM. ELM is one of the outstanding models in the field of pattern recognition. Its network structure is similar to traditional neural networks, and its learning efficiency is better. ELM provides a faster training convergence and does not require repeated iterations to adjust weights. It has the advantages of fast learning speed and good generalization performance. However, ELM could only process real-valued data. For complex SAR images with phase information, it is more appropriate to use CV-ELM for processing. Next, take solving $f^2_{m,n}(\cdot)$ as an example to introduce the principle of CV-ELM.

Given data $\left\{\mathbf{P}^2_i\right\}_{i=1}^{N_l+1} = \{\mathbf{NH}(m,n), \mathbf{C}_1, \ldots, \mathbf{C}_{N_l}\}$ and label $\left\{\mathbf{L}^2_i\right\}_{i=1}^{N_l+1} = \{1+1j, 0, \ldots, 0\}$. $\mathbf{X}_i = vec\left(\mathbf{P}^2_i(m,n)\right) = [x_1, \ldots, x_u, \ldots, x_{N_s}]$ is the input with dimension N_s. In CV-ELM, weight $\left\{\mathbf{W} \in \mathbb{C}^{N_s \times L}\right\} = [\mathbf{W}_1, \ldots, \mathbf{W}_v, \ldots, \mathbf{W}_L]$ and bias $\left\{\mathbf{b} \in \mathbb{C}^L\right\} = [b_1, \ldots, b_v, \ldots b_L]^T$ are randomly generated, L is the node number of the hidden layer, $\mathbf{W}_v = [w_1, \ldots, w_u, \ldots w_{N_s}]^T$, then the output o_i of CV-ELM could be expressed as

$$o_i = \sum_{v=1}^{L} \beta_v \sigma\left(\mathbf{X}_i \mathbf{W}_v + b_v\right) \tag{4}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_v, \ldots \beta_L]^T$ denotes output weight of the hidden layer. $\sigma(\cdot)$ is the complex-valued activation function [14], which adopts the tanh activation function for the amplitude and keeps the phase unchanged. For the complex variable z, the activation process of $\sigma(\cdot)$ is as follows.

$$\sigma(z) = \tanh(|z|)\exp(j\arg(z)) \tag{5}$$

The loss function of CV-ELM is the quadratic norm of the difference between the expectation and actual output, expressed as follows.

$$\sum_{i=1}^{N_l+1} \| o_i - \hat{y}_i \| = 0 \tag{6}$$

where $\hat{y}_i = vec\left(\mathbf{L}^2_i\right)$ is the expectation or label of dimension 1. According to the theory of CV-ELM, there exists $\boldsymbol{\beta}$ that makes the above formula (9) true. Use matrixes to express the formula (7) as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{L} \tag{7}$$

where $\mathbf{L}$ is the expectation or label set by unsupervised strategy. For $\left\{\mathbf{L}^1_i\right\}$, $\mathbf{L} = [0, 1+1j, \ldots, 1+1j]^T$. For $\left\{\mathbf{L}^2_i\right\}$, $\mathbf{L} = [1+1j, 0, \ldots, 0]^T$. $\mathbf{H} \in \mathbb{C}^{(N_l+1)\times L}$ represents the output matrix of the hidden layer. Let $N_L = N_l + 1$, then $\mathbf{H}$ could be expanded into the following form.

$$\mathbf{H} = \begin{bmatrix} \sigma(\mathbf{X}_1\mathbf{W}_1 + b_1) & \cdots & \sigma(\mathbf{X}_1\mathbf{W}_L + b_L) \\ \vdots & \ddots & \vdots \\ \sigma(\mathbf{X}_{N_L}\mathbf{W}_1 + b_1) & \cdots & \sigma(\mathbf{X}_{N_L}\mathbf{W}_L + b_L) \end{bmatrix} \tag{8}$$

Then the output weight $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{L} \tag{9}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse.

Experimental Results: The proposed method is compared with two representative methods, CFAR and EF. A GF-3 SAR image containing sea clutter and a SAR dataset containing 207 ship targets are used to test the effectiveness of the proposed method in different scenes. Three indicators of missed-detection probability P_m [15], recall P_r [16], and precision P_p [16] are used to measure the performance of the proposed method. P_m represents the proportion of target pixels that are incorrectly classified as clutter pixels to all the target pixels. P_r is the proportion of target pixels correctly classified to all target pixels. P_p represents the proportion of target pixels that are correctly classified to all the pixels classified as target pixels. For the experimental parameters, the node number of the hidden layer in CV-ELM is 51, the size of $\mathbf{NH}(m,n)$ is 3×3, and N_l is 480, T_{cof} is 0.8. Conventionally, the degree of freedom (DoF) of CV-ELM is twice that of a counterpart ELM. In order to make a fair comparison, an ELM is designed to have the same DoF as the CV-ELM. The node number of the hidden layer in ELM is 102.

Performance Comparison on A GF-3 SAR Image: The processing results of different methods for a GF-3 image are shown in Fig. 2. The GF-3 image has one ship target and strong and heterogeneous sea clutter. The P_m, P_r, and P_p performance comparisons are shown in Table 1. According to the qualitative analysis in Fig. 2, the traditional method of CFAR could detect most of the pixels in ship targets. Still, it also incorrectly detects some strong clutter as targets. By comparing the original image in Fig. 2(a) and the detection result of the EF in Fig. 2(d), strong

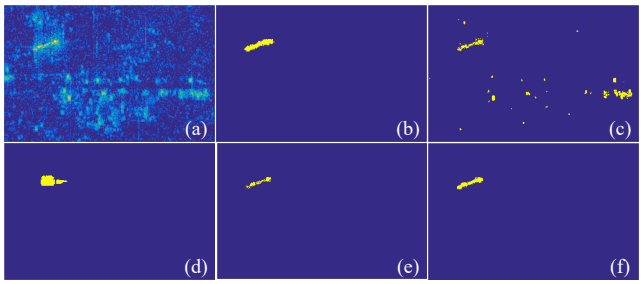

Fig 2 *Results of target detection for GF-3 image. (a) Original image. (b) Ground-truth. (c) CFAR processed image. (d) EF processed image. (e) U-TDM with ELM processed image. (f) U-TDM with CV-ELM processed image.*

Table 1. Performance Comparison on GF-3 SAR Image (%)

	CFAR	EF	U-TDM (ELM)	U-TDM (CV-ELM)
P_m	49.72	47.73	7.10	**5.68**
P_r	50.28	52.27	92.90	**94.32**
P_p	25.00	41.16	68.27	**69.60**

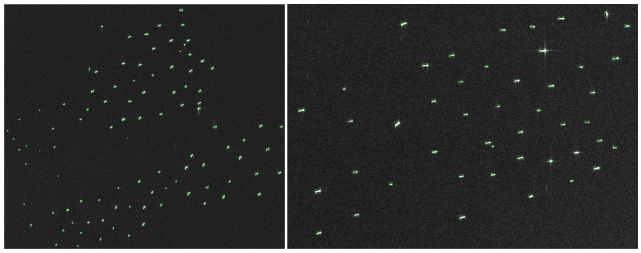

Fig 3 *Result of ship detection for Sentinel-1 SAR images. The correctly detected targets, the false alarms, and the missed alarmsand are indicated by the green, red, and blue rectangle boxes, respectively.*

clutter does not affect the EF. Still, the detection result of the EF is not very accurate due to the characteristics of the EF, and the detected target area is larger than the actual target area. The U-TDM using ELM has a better target detection effect, but a part of the clutter area is incorrectly detected as the target. The U-TDM using CV-ELM is not affected by strong clutter, and the detection result is relatively accurate.

From the quantitative analysis of Table 1, the missed-detection probability of the U-TDM using CV-ELM is the lowest, which is 5.68%, and the missed-detection probability of the CFAR is the highest, which is 49.72%. The missed-detection probability of the U-TDM is about one-tenth of that of CFAR. The U-TDM also achieved the highest recall of 94.32%. The precision of the CFAR is the lowest, which is 25.00%, and the U-TDM using CV-ELM is the highest, reaching 69.60%. Based on the three indicators, the performance of the U-TDM using CV-ELM is the best, and the U-TDM using ELM is second, which fully demonstrates the effectiveness of the proposed knowledge-assisted method. It should be noted that for the nearshore area, sea-land segmentation should be carried out before target detection to avoid the influence of strong land clutter.

Performance on SAR Ship Dataset: A SAR ship dataset is used to test the effect of the proposed method. This dataset contains 4 large-scene Sentinel-1 SAR images with a total of 207 ship targets. The labels are annotated manually by experts via the software named LabelImg. Fig. 3 shows the detection result of U-TDM using CV-ELM for one large-scene Sentinel-1 SAR image. A detailed comparison is presented in Table 2. The IoU threshold is 0.3. From Table 2, the proposed method achieves the highest number of correctly detected targets and the lowest number of false and missing alarms and achieves good detection results even for small targets.

Table 2. Performance Comparison on the SAR Ship Dataset

	Correctly detected target	False alarm	Missing alarm
CFAR	153	64	54
EF	120	84	87
U-TDM (ELM)	195	15	12
U-TDM (CV-ELM)	**198**	**7**	**9**

Conclusion: In this letter, we propose a U-TDM based on an unsupervised strategy and CV-ELM for SAR target detection. CV-ELM converts the target detection problem into a pixel binary classification problem, which could fully exploit the characteristic differences in amplitude and phase between the target and clutter. In U-TDM, the weights of CV-ELM are solved by an unsupervised strategy based on reference patch sets, which avoids the acquisition difficulty of ground-truth. Experiments show that the proposed method could improve target detection performance compared with traditional methods. In the future study, the proposed method will be used to detect targets in a radio-frequency interference background.

References

1. Pelich, R., et al.: Ais-based evaluation of target detectors and sar sensors characteristics for maritime surveillance. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8(8), 3892–3901 (2014)
2. Pappas, O., et al.: Superpixel-level cfar detectors for ship detection in sar imagery. IEEE Geosci. Remote Sens. Lett. 15(9), 1397–1401 (2018)
3. Wang, Z., et al.: Adaptive cfar detectors for mismatched signal in compound gaussian sea clutter with inverse gaussian texture. IEEE Geosci. Remote Sens. Lett. 19, 1–5 (2021)
4. Huang, Y., et al.: Gmti and parameter estimation via time-doppler chirp-varying approach for sngle-channel airborne sar system. IEEE Trans. Geosci. Remote Sens. 55(8), 4367–4383 (2017)
5. Suwa, K., et al.: Image-based target detection and radial velocity estimation methods for multichannel sar-gmti. IEEE Trans. Geosci. Remote Sens. 55(3), 1325–1338 (2017)
6. Kaplan, L.M., et al.: Improved sar target detection via extended fractal features. IEEE Trans. Aerosp. Electron. Syst. 37(2), 436–451 (2001)
7. Chen, S., et al.: Target classification using the deep convolutional networks for sar images. IEEE Trans. Geosci. Remote Sens. 54(8), 4806–4817 (2016)
8. Du, L., et al.: Saliency-guided single shot multibox detector for target detection in sar images. IEEE Trans. Geosci. Remote Sens. 58(5), 3366–3376 (2020)
9. Lin, Z., et al.: Squeeze and excitation rank faster r-cnn for ship detection in sar images. IEEE Geosci. Remote Sens. Lett. 16(5), 751–755 (2019)
10. Chendeb El Rai, M., et al.: Semisegsar: A semi-supervised segmentation algorithm for ship sar images. IEEE Geosci. Remote Sens. Lett. 19, 1–5 (2022)
11. Du, L., et al.: Sar target detection network via semi-supervised learning. J. Electron. Inf. Technol. 42(1), 154–163 (2020)
12. Huang, G., et al.: Trends in extreme learning machines: A review. Neural Networks 61, 32–48 (2015)
13. Havens, T.C., et al.: Fuzzy c-means algorithms for very large data. IEEE Trans. Fuzzy Syst. 20(6), 1130–1146 (2012)
14. Hirose, A.: Complex-valued neural networks. Springer Science and Business Media (2012)
15. Wang, Y., et al.: Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 sar images. Remote Sens. Lett. 9(8), 780–788 (2018)
16. Schilling, H., et al.: Object-based detection of vehicles using combined optical and elevation data. ISPRS-J. Photogramm. Remote Sens. 136, 85–105 (2018)