

Semisupervised heterogeneous ensemble for ship target discrimination in synthetic aperture radar images

Yongxu Li¹, Xudong Lai^{1*}, Mingwei Wang²

¹School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

²School of Computer Science, Hubei University of Technology, Wuhan 430068, China

Received 9 April 2021; accepted 27 December 2021

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Ship detection using synthetic aperture radar (SAR) plays an important role in marine applications. The existing methods are capable of quickly obtaining many candidate targets, but numerous non-ship objects may be wrongly detected in complex backgrounds. These non-ship false alarms can be excluded by training discriminators, and the desired accuracy is obtained with enough verified samples. However, the reliable verification of targets in large-scene SAR images still inevitably requires manual interpretation, which is difficult and time consuming. To address this issue, a semisupervised heterogeneous ensemble ship target discrimination method based on a tri-training scheme is proposed to take advantage of the plentiful candidate targets. Specifically, various features commonly used in SAR image target discrimination are extracted, and several acknowledged classification models and their classic variants are investigated. Multiple discriminators are constructed by dividing these features into different groups and pairing them with each model. Then, the performance of all the discriminators is tested, and better discriminators are selected for implementing the semisupervised training process. These strategies enhance the diversity and reliability of the discriminators, and their heterogeneous ensemble makes more correct judgments on candidate targets, which facilitates further positive training. Experimental results demonstrate that the proposed method outperforms traditional tri-training.

Key words: synthetic aperture radar, ship target discrimination, non-ship false alarms, semisupervised, heterogeneous ensemble

Citation: Li Yongxu, Lai Xudong, Wang Mingwei. 2022. Semisupervised heterogeneous ensemble for ship target discrimination in synthetic aperture radar images. *Acta Oceanologica Sinica*, 41(7): 180–192, doi: 10.1007/s13131-021-1980-2

1 Introduction

Synthetic aperture radar (SAR), as an active earth observation system that can operate day and night under all weather conditions and quickly capture high-resolution images with broad coverage, has shown great potential in marine applications, such as marine surveillance, marine environment, marine security, and fishery control. Ship detection and recognition is an important marine application and has been a topic of interest in recent decades (Aiello et al., 2019; Lang et al., 2020). The practical architecture of SAR automatic ship target recognition consists of three consecutive stages: detection, discrimination, and recognition/classification (Du et al., 2020). The classic constant false alarm rate (CFAR) detection approach has been widely employed to detect candidate targets from the sea background due to its adaptive ability and easy implementation (Ai et al., 2021). However, due to the complex interactions between the SAR system and the imaging scenes, a large number of false alarms may be generated because of sea clutter (Ao et al., 2018) and some typical phenomena of SAR images, such as azimuth ambiguity (Di Martino et al., 2014), sidelobes (Vespe and Greidanus, 2012), and residual image caused by azimuth shift (Li et al., 2019). These false alarms need to be eliminated via the discrimination stage that minimizes the interference to the subsequent high-complexity ship target recognition, reducing the cost and in-

creasing the effectiveness (Chen et al., 2020).

Ship target discrimination is commonly implemented by designing discriminators, including unsupervised learning methods, semisupervised learning (SSL) methods, and supervised learning methods. Unsupervised learning methods do not need labeled samples, but for complex scenes, their discrimination power is limited (Tello et al., 2009). Supervised learning methods such as the support vector machine (SVM) (Hwang and Jung, 2018; Ma et al., 2018; Lang et al., 2016; Falqueto et al., 2019), k-nearest neighbor (KNN) (Ma et al., 2018; Lang et al., 2016; Falqueto et al., 2019), decision tree (DT) (Lang et al., 2016; Falqueto et al., 2019), logistic regression (LR) (Falqueto et al., 2019), and discriminant analysis (DA) (He et al., 2018) for discriminating targets in SAR images have been researched, and the desired accuracy is achieved with enough labeled samples. Recently, owing to the exponential growth in computing power, algorithms developed with deep learning techniques have achieved automatic feature expression competence and higher detection accuracy by relying on more labeled samples (Kang et al., 2017; Chang et al., 2019). However, ship targets are usually labeled by matching the automatic identification system (AIS) information with large-scene SAR images, and manual visual interpretation is indispensable due to spatial registration errors and some ships not broadcasting AIS and is a difficult and time-consuming task

Foundation item: The National Natural Science Foundation of China under contract No. 61971455.

*Corresponding author, E-mail: laixudong@whu.edu.cn

(Pelich et al., 2015). In addition, the applicability of labeled samples in various imaging modes or different SAR systems is limited (Gao et al., 2019).

SSL techniques that mine information contained in unlabeled samples to compensate for the defects caused by insufficient information from labeled samples have attracted much attention (Belkin and Niyogi, 2004; Nigam et al., 2000; Camps-Valls et al., 2007; Liu et al., 2016). The key issue of SSL is selecting unlabeled samples with high reliability. Among SSL techniques, co-training and subsequent tri-training have been employed in many applications and are effective (Blum and Mitchell, 1998; Zhou and Li, 2005; Hua et al., 2017; Wang et al., 2020). Co-training first trains two base classifiers separately on two independent feature sets, and then each classifier predicts the unlabeled samples to augment the other classifier's training set. Standard co-training requires that the features be naturally partitioned into two sets, but such a requirement is difficult to meet in most machine learning problems. Tri-training is an extension of the co-training framework, and it is easy to implement since the only requirement is to train three base classifiers through different training sets created by bootstrap sampling the original labeled samples (OLSs). During the training process, the unlabeled samples are labeled by any classifier as long as the other two classifiers agree on their predictions. Tri-training has been subjected to rigorous theoretical derivation that if there are enough unlabeled samples, the classification noise rate will be mitigated; moreover, better generalizability is obtained by the ensemble of these three classifiers.

The base classifiers undergo further improvements through the training process only when appropriate unlabeled samples are selected (Wang et al., 2020). The classifier ensemble combines the outputs of multiple individual classifiers and often significantly improves the classification performance (Albukhanjer et al., 2017). Heterogeneous ensembles with different types of classifiers effectively increase the prediction accuracy of unlabeled samples compared to homogeneous approaches that use the same type of classifiers (Amozegar and Khorasani, 2016; Seijo-Pardo et al., 2017). Therefore, when there are too few available labeled samples to simulate the distribution of the entire data, the heterogeneous ensemble can make fewer generalization errors by maximizing their agreement on the unlabeled samples (Dasgupta et al., 2001). On this basis, further improvements can be achieved by employing multiple different classifiers and increasing the diversity among the classifiers (Zhou and Li, 2005; Wang et al., 2020). Multiple heterogeneous ship target discriminators are realized by extracting various SAR image features and employing different classification models. The SAR image feature is an important element of designing discriminators. Lincoln Laboratory carried out research on target discrimination earlier, and several classic features related to size, polarimetric properties, and power have been proven effective (Kreithen et al., 1993). After that, five spatial boundary attribute features were proposed to capture the changes in the spatial dispersion of the high-intensity pixels (Verbout et al., 1998). Based on these five features, three signal-to-noise ratio features were proposed by Gao (2011) to measure the contrast between target pixels and background pixels. Furthermore, ten shape-related features introduced by Bhanu and Lin (2003) were performed well. The differences between ship targets and false alarms are described by these features in terms of shape, size, texture, contrast, power, spatial distribution, etc., and are further enriched with different polarization SAR images. Another way to increase the diversity is to employ different models; the abovementioned models (SVM,

DT, DA, KNN, and LR) have a satisfactory classification ability and are competent for designing discriminators. Furthermore, some variants of models with reasonably set parameters and appropriate kernel functions such as the quadratic and Gaussian functions will cover a wide class of nonlinearities and lead to better performance (Haider et al., 2019; Baudat and Anouar, 2000), which are also adopted to achieve improvements.

In this paper, a semisupervised heterogeneous ensemble method based on a tri-training scheme is proposed to enhance the capability of ship target discrimination in SAR images with a limited number of labeled samples. First, various features commonly used in SAR image target discrimination are extracted and divided into different groups, and the diversity among the discriminators is effectively improved by randomly pairing those feature groups with different models and their classic variants. Second, if all the discriminators are directly employed without clear capacity in advance, incorrectly labeled samples may be added to the training sets to decrease the training efficiency. Hence, OLSs are used with 5-fold cross-validation to test their performance, from which better-performing ones with appropriate kernel functions and parameter settings along with suitable feature groups are determined as the initial discriminators. Then, these heterogeneous discriminators are initialized from training sets generated via bootstrap sampling the OLSs to implement the semisupervised training process. Moreover, when there are more than three initial discriminators, the prediction of unlabeled samples by the other discriminators may not be uniform; here, only the unlabeled samples that obtain the same predictions are selected. Compared with traditional tri-training, these strategies increase the reliability of the selected unlabeled samples, which in turn facilitates further positive training and improves the discrimination performance.

The remainder of this paper is organized as follows. Section 2 presents the methodology and implementation of the proposed method in detail. The experimental data and results are shown in Section 3. Finally, the conclusions are discussed in Section 4.

2 Semisupervised heterogeneous ensemble ship target discrimination

2.1 SAR image features for discrimination

The SAR image features reflect the characteristics and differences between the ship targets and clutter, including geometric features, texture features, electromagnetic scattering features, etc. Relevant studies have confirmed that many features are able to stably distinguish targets and false alarms in SAR images under different polarizations and multiscale resolutions (Kreithen et al., 1993; Verbout et al., 1998; Gao, 2011; Bhanu and Lin, 2003). Specifically, under the Strategic Target Algorithm Research Contract, Lincoln Laboratory proposed three features for SAR image target discrimination, and features developed by the Environmental Research Institute of Michigan (ERIM), Rockwell International Corporation (Rockwell), and Loral Defense Systems (Loral) were also proven effective by Lincoln Laboratory. Among these features, some substantially overlapping features along with polarimetric features that require polarimetric SAR images are excluded in this paper. Therefore, only partial ERIM features, the contiguousness features of the Loral features, and the specific-entropy feature of the Rockwell features are chosen. The above features are collectively referred to as Old-Lincoln features. After that, Lincoln Laboratory proposed five spatial boundary attribute features, referred to as New-Lincoln features to distinguish them from the previous features. The University of California

proposed ten shape-related features, referred to as Bhanu features. Based on the New-Lincoln features, the National University of Defense Technology proposed three signal-to-noise ratio features referred to as Gao features. All the adopted features and their explanations are shown in Table 1.

2.2 The proposed discriminating process

The complete framework of the proposed semisupervised ship target discrimination method is shown in Fig.1. First, multiple heterogeneous discriminators are designed by extracting different SAR image features and employing different classification models. The 34 features mentioned above are extracted and divided into four groups. The Old-Lincoln features F_1-F_{16} are the first group. Considering that there are only a few New-Lincoln features $F_{17}-F_{21}$ and Gao features $F_{22}-F_{24}$, the sum of them $F_{17}-F_{24}$ form the second group. The Bhanu features $F_{25}-F_{34}$ comprise the third group. The last group contains all 34 features F_1-F_{34} . In addition, the five abovementioned models (i.e., the SVM, DT, DA, KNN, and LR models), along with their classic variants, are employed to increase diversity. The SVM model, which is a widely used model that requires less computational effort, separates the classes by a hyperplane based on support vector theory. The DT model, which is a basic regression and classification model with a clear structure and low computational com-

plexity, calculates the entropy of the samples to set criteria in each decision node and then splits unlabeled samples. As a traditional statistical method, DA projects the feature vectors to a lower-dimensional feature space to increase the separation and has proven successful on classification problems; it has the power to solve a series of problems. The KNN model is a popular technique that calculates the distance between unlabeled samples and their closest labeled samples for classification. This model performs well, and the results are easy to interpret. LR computes the probability of unlabeled samples corresponding to particular labels and has shown the potential for effective and efficient classification of different types of data. Second, as preliminary screening, all combinations by pairing each model and their variants with different feature groups are tested by the OLSs with 5-fold cross-validation to determine the initial discriminators with appropriate kernel functions, parameter setting and feature groups.

In the training process, five initial discriminators, denoted by $D1, D2, D3, D4,$ and $D5$ are initialized by different training sets ($L1, L2, L3, L4, L5$), which are from bootstrap sampling the OLSs. After that, each discriminator is iteratively refined by samples chosen from the unlabeled set, these samples are unanimously predicted by other discriminators and are considered to have high credibility. Taking $D1$ as an example, let $L1$ denote the ori-

Table 1. Adopted synthetic aperture radar image features for discrimination

	Feature name	Explanation	Feature Symbol
Old-Lincoln features	standard deviation	The standard deviation of all the pixels in a target-sized box.	F_1
	fractal dimension	The Hausdorff dimension of the spatial distribution of strong scatterers in the region of the target-sized box.	F_2
	weighted-rank fill ratio	The power of strong scatterers and normalizing by the total power of all pixels within the target-size box.	F_3
	mass	The number of pixels in the target-shaped blob.	F_4
	diameter	The length of the diagonal of the smallest rectangle that encloses the target-shaped blob.	F_5
	square-normalized rotational inertia	The second mechanical moment of the target-shaped blob around its center, normalized by the inertia of an equal mass square.	F_6
	maximum CFAR statistic	The maximum value in the CFAR image is contained within the target-shaped blob.	F_7
	mean CFAR statistic	The average value of the CFAR image is taken over the target-shaped blob.	F_8
	percent bright CFAR statistic	The percentage of pixels within the target-shaped blob that exceeds a certain CFAR value. The CFAR value is set as AvCS in this paper.	F_9
	specific-entropy	The number of pixels that exceed the threshold that is set to quantity corresponding to the 98th percentile of the surrounding clutter and normalize this value by the total number of pixels in a target-sized box.	F_{10}
New-Lincoln features	contiguousness	Segment each image (target-size box and CFAR image) into three separate images (shadow, background, and target) based on the amplitude of individual pixels, then computing numbers from each of these six regions of interest.	$F_{11}-F_{16}$
	threshold	The optimal threshold for an image chip is just greater than the clutter background pixel value and smaller than the target pixel value (active pixel).	F_{17}
	activation	The fraction of pixels that are activated in the optimally thresholded image.	F_{18}
	dispersion	The weighted average distance from the centroid of a high-intensity pixel on the object, where the weights are assigned in proportion to the mass at each pixel location.	F_{19}
	inflection	The rate of change of the mass dispersion statistic at the optimal threshold.	F_{20}
Gao features	acceleration	It measures the acceleration associated with the rate of change of the mass dispersion statistic at the optimal threshold.	F_{21}
	average signal-to-noise-ratio	The average contrast of the target or the false alarms to the background in a candidate chip.	F_{22}
	peak signal-to-noise-ratio	The peak contrast of the target or the false alarms to the background in a candidate chip.	F_{23}
Bhanu features	percentage of bright pixels	The percentage of the brightest pixels with contrast higher than $p\%$ of PSNR in all the "active" pixels and p is set to 50 according to the reference.	F_{24}
	projection	Project the potential target pixels on a horizontal line (or a vertical line, the major diagonal line, the minor diagonal line) and compute the maximum distance.	$F_{25}-F_{28}$
	distance	The minimum (or maximum, average) distance from each potential target pixel to the centroid.	$F_{29}-F_{31}$
	moment	The horizontal (or vertical, diagonal) second-order distance from each potential target pixel to the centroid.	$F_{32}-F_{34}$

Note: CFAR, constant false alarm rate; PSNR, peak signal-to-noise-ratio.

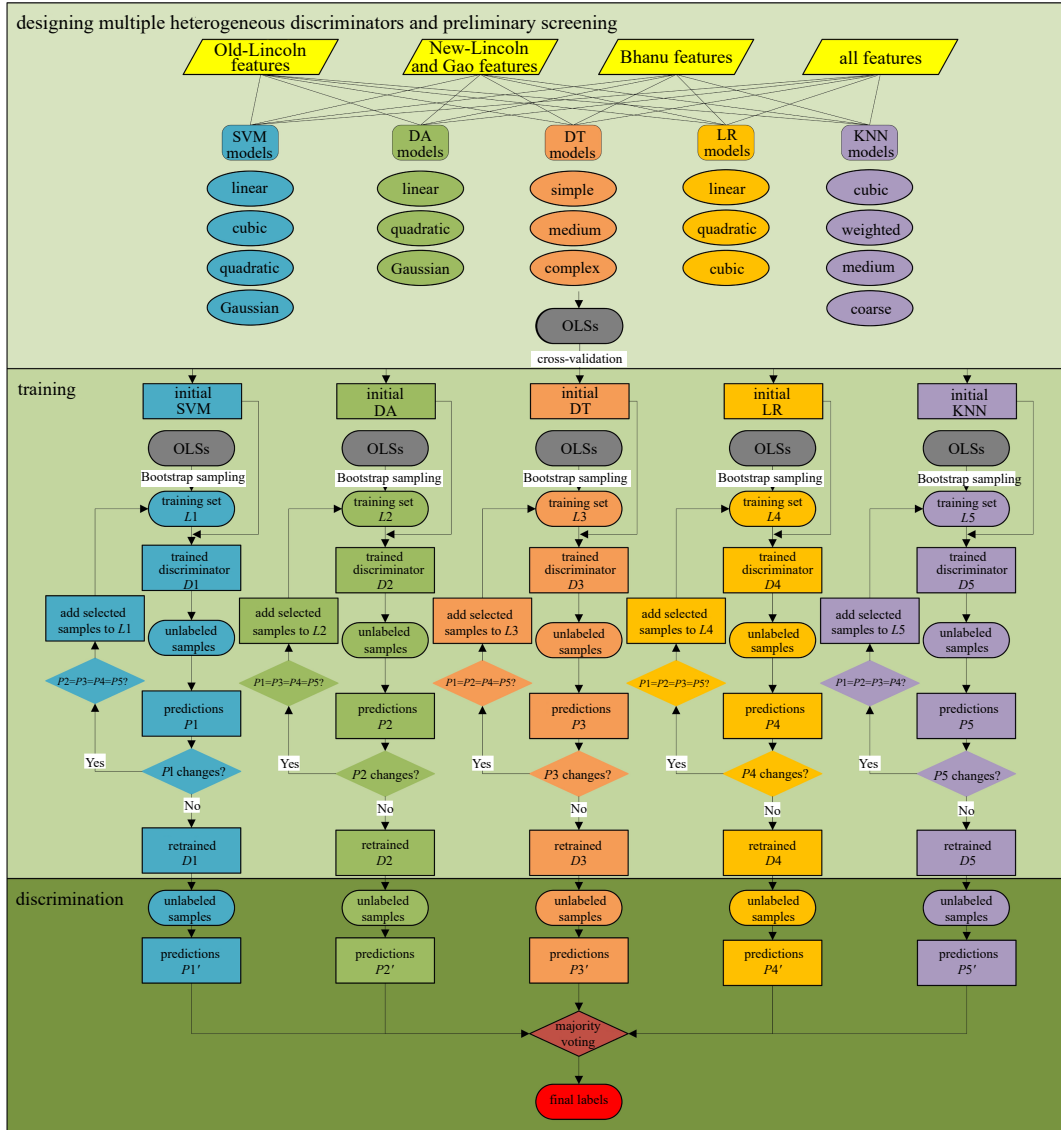


Fig. 1. The framework of the proposed method.

ginal labeled training set of $D1$, let $|L1|$ denote the number of samples $L1$ contains, and let U denote the samples without labels, similarly, $|U|$ is the number of samples U contains. Let $L1_{t-1}$ and $L1_t$ denote the unlabeled samples in which other discriminators make the same prediction ($P2 = P3 = P4 = P5$) in the $(t-1)$ _{th} and the t _{th} iterations. Equation (1) is used to determine whether $L1_{t-1}$ and $L1_t$ are expanded to $L1$, which ensures that the possible noise contained in $L1_t$ is gradually reduced among each iteration and $D1$ undergoes further improvements. Then, in the $(t-1)$ _{th} and t _{th} iterations, the training sets of $D1$ are $L1 \cup L1_{t-1}$ and $L1 \cup L1_t$.

$$0 < \frac{e_t}{e_{t-1}} < \frac{|L1_{t-1}|}{|L1_t|} < 1, \quad (1)$$

where e_t represents the upper bound of the discrimination error rate of unlabeled samples in the t _{th} iteration, that is, the error rate generated when the other discriminators make the same prediction. Since e_t cannot be exactly estimated, it is approximated by the error rate of the combination of $D2$, $D3$, $D4$, and $D5$ predicting on the OLSs.

It is worth noting that when $e_t < e_{t-1}$ and $|L1_t| > |L1_{t-1}|$, since $|L1_t|$ may be much larger than $|L1_{t-1}|$, $e_t |L1_t|$ may not be less than $e_{t-1} |L1_{t-1}|$. If this happens, $|L1_t|$ can be randomly subsampled so that it still satisfies $e_t |L1_t| < e_{t-1} |L1_{t-1}|$. Let S denotes the size of $L1_t$ after subsampling; if Eq. (2) holds, $e_t |L1_t| < e_{t-1} |L1_{t-1}|$ is satisfied. In addition, $|L1_{t-1}|$ should satisfy Eq. (3) such that $|L1_t| > |L1_{t-1}|$ is still true after subsampling.

$$S = \frac{e_{t-1}}{e_t} |L1_{t-1}| - 1, \quad (2)$$

$$|L1_{t-1}| > \frac{e_t}{e_{t-1} - e_t}. \quad (3)$$

During the training process, the unlabeled samples labeled during one iteration only participate in that iteration and are still treated as unlabeled samples in the next iteration, which avoids the premature introduction of noise. Namely, at the t _{th} iteration, $L1_{t-1}$ will be put back into U so that $|U|$ remains unchanged in each iteration.

The above operation is performed on $D2$, $D3$, $D4$, and $D5$ to

complete one training iteration, and the iteration is terminated when the prediction on the unlabeled samples from each discriminator stops changing. Then, the training process is completed, and the refined discriminators are obtained.

At the final discrimination stage, the labels of the unlabeled samples are determined by majority voting via the refined discriminators.

2.3 Candidate targets detection

Before discrimination, a CFAR detector is employed to produce the candidate targets, which can maintain a given probability of false alarm p_{fa} by comparing pixels with an adaptive threshold T . T is obtained by solving Eq. (4) according to the amplitude probability density function $f(x)$ and preset p_{fa} . $f(x)$ is obtained by accurately modeling the sea clutter around the pixels of interest, and p_{fa} is set based on experience. The K distribution (while the amplitude follows the K -root distribution) is commonly used in the literature because of the compound formulation, which was introduced by Ward, enabling both small-scale and large-scale components of the sea clutter to be characterized (Ward et al., 2006). To set p_{fa} , a smaller value in combination with a selected distribution leads to a larger T that easily misses the ship; in contrast, it easily causes more false alarms. To ensure the highest detection rate as possible, in this paper, a smaller T is obtained by setting a larger p_{fa} to 10^{-1} . Although a large number of false alarms may be generated, they are used to verify the ability of discriminators. In addition, the size of the buffer cell and the background cell for the CFAR sliding window depends on the resolution of the image along with the maximum and minimum sizes of the ship target. In this paper, the buffer cell is defined as a circle with a radius equal to the world's longest ship, and the edge of the square background cell is $\sqrt{2}$ times that radius (Pe-

lich et al., 2015). After detection, a few labels are produced as the OLSs via the AIS information combined with manual interpretation.

$$p_{fa} = 1 - \int_0^T f(x) dx = \int_T^\infty f(x) dx. \quad (4)$$

3 Experimental results and analysis

3.1 Experimental data

The experimental platform is a PC with an AMD Ryzen 7-2700 CPU and 8 GB RAM, and the program of CFAR detector and all discriminators is implemented in MATLAB R2016a. In addition, the state-of-the-art deep learning detectors are employed for comparison, they are trained by using the PaddleDetection toolkit (PaddlePaddle Authors, 2021) on a single Tesla V100 GPU with 16 GB memory. Three SAR images captured by Sentinel-1 were obtained in three different regions of the East China Sea, as shown in Fig.2.

According to the predefined observation plan, these images are Level-1 Ground Range Detected high-resolution products manufactured in interferometric wide (IW) swath mode with VV and VH polarizations and a swath of approximately 250 km at a resolution of $20 \text{ m} \times 22 \text{ m}$ (Sentinel-1 Observation Scenario, 2020; Interferometric wide swath, 2020). Other relevant information is shown in Table 2. Therefore, each of the four feature groups mentioned above is expanded to the VV, VH, and dual-polarization groups, which are used to comparatively evaluate the ship discrimination performance under different polarizations; in total, 12 different feature groups are obtained.

Considering that in some sea areas, as ship targets are

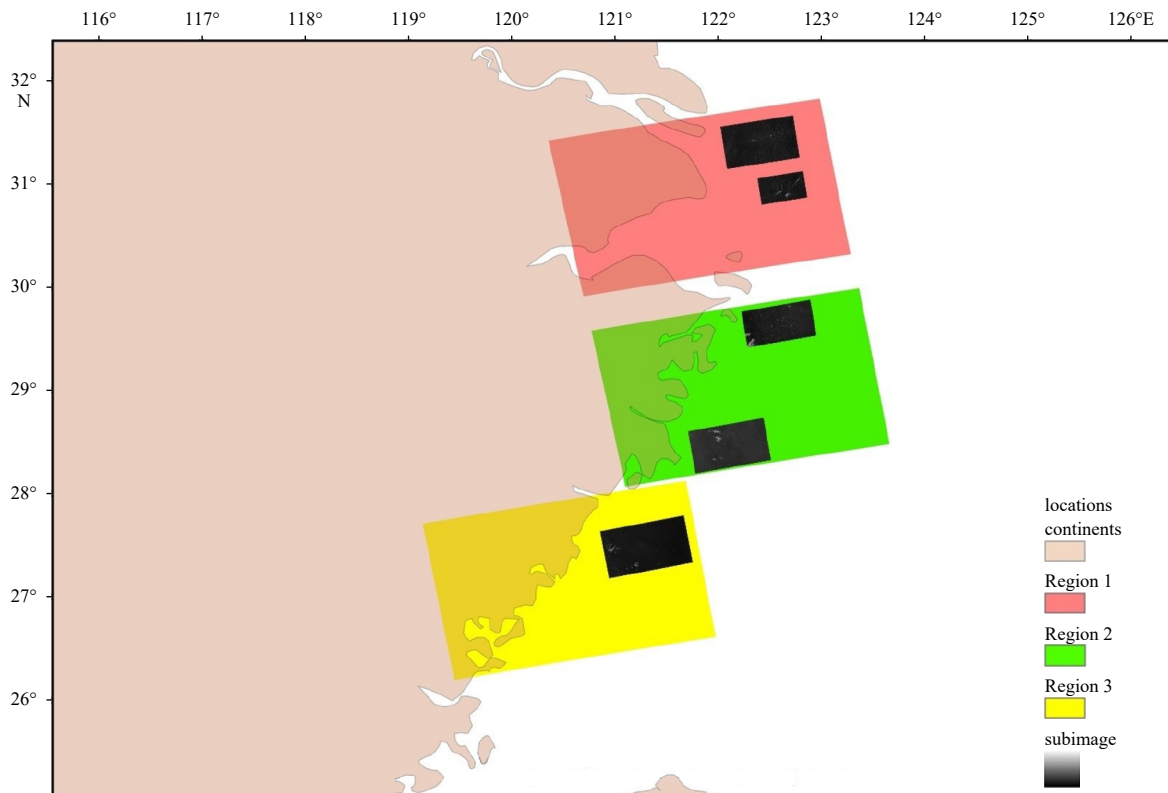


Fig. 2. Synthetic aperture radar images in three regions of the East China Sea.

Table 2. Details of three synthetic aperture radar images

Regions	Acquire time	Incidence angle/(°)	Latitude	Longitude
1	2016-10-30	30.72-45.98	29.91°-31.82°N	120.36°-123.29°E
2	2020-03-01	30.79-46.08	28.07°-29.98°N	120.77°-123.66°E
3	2020-03-24	30.86-46.17	26.20°-28.12°N	119.13°-121.98°E

sparingly distributed and may not be present, directly using the entire SAR image with wide coverage for experiments is time consuming and inefficient. Therefore, we extracted five subimages by manual cutting from the three original images to conduct experiments, and there was no overlap between the subimages. Other relevant information is shown in Table 3.

3.2 Determining the initial discriminators

As mentioned above, preliminary screening is carried out first to determine the initial heterogeneous discriminators with good performance from well-paired combinations. For this purpose, two subimages 1 and 2, are extracted, and the verified CFAR detection results under VV polarization are shown in Figs 3a and b. Except for ship targets, which are used as positive samples, there are many false alarms caused by azimuth ambiguity, sidelobes, small islands, etc., which are used as negative samples. However, there are relatively few ship targets in Fig. 3a; hence, Fig. 3b is also extracted to supplement the number of positive samples. There are a total of 230 ships (indicated by red boxes) and 499 false alarms (indicated by green boxes), and these labeled

Table 3. Details of the subimages for experimental

No.	Regions	Ship targets	False alarms	Purpose
1	Region 1	68	369	OLSs
2	Region 3	162	130	OLSs
3	Region 1	27		test
4	Region 2	98		test
5	Region 2	70		test

samples are used as the OLSs.

In total, 204 candidate combinations were obtained by pairing 17 different classification models and 12 feature groups. They were tested by the OLSs with 5-fold cross-validation, and only the accuracy of the optimal variant of each model combined with different feature groups is listed in Table 4; the best performance is highlighted in bold.

The results reported in Table 4 show that most combinations performed well, with a maximum accuracy of approximately 98%, and the results can be further analyzed from three aspects. At the feature level, the differences among the feature groups are large, the Old-Lincoln feature groups and the groups including all features acquired the better results, most of which had an accuracy of more than 95%, and the highest average accuracy is 97.12% obtained by the groups including all features under VV polarization. The Bhanu features are slightly worse with accuracies mostly in the range of 90%–95%, which are significantly better than the combination of the New-Lincoln features and the Gao features with accuracies in the range of 76%–80%. At the polarization level, the difference between VV and VH is small, while the performance of dual-polarization with more features is unexpectedly worse than that of single polarization in some cases. It can be inferred that there is redundancy between the features of VV and VH polarization, which will affect the discrimination performance. At the classification model level, all five models can reach a high accuracy of approximately 97% with appropriate features, and the discriminators with the best performance are the Gaussian SVM, weighted KNN with Old-Lincoln features under VH polarization, the linear DA, quadratic LR combining the Old-Lincoln features under VV polarization, the complex DT combining all the features under dual-polarization. They are used as the initial discriminators to implement the proposed training process, and the order according to the accuracy is complex DT, Gaussian SVM, linear DA, weighted KNN, quadratic LR.

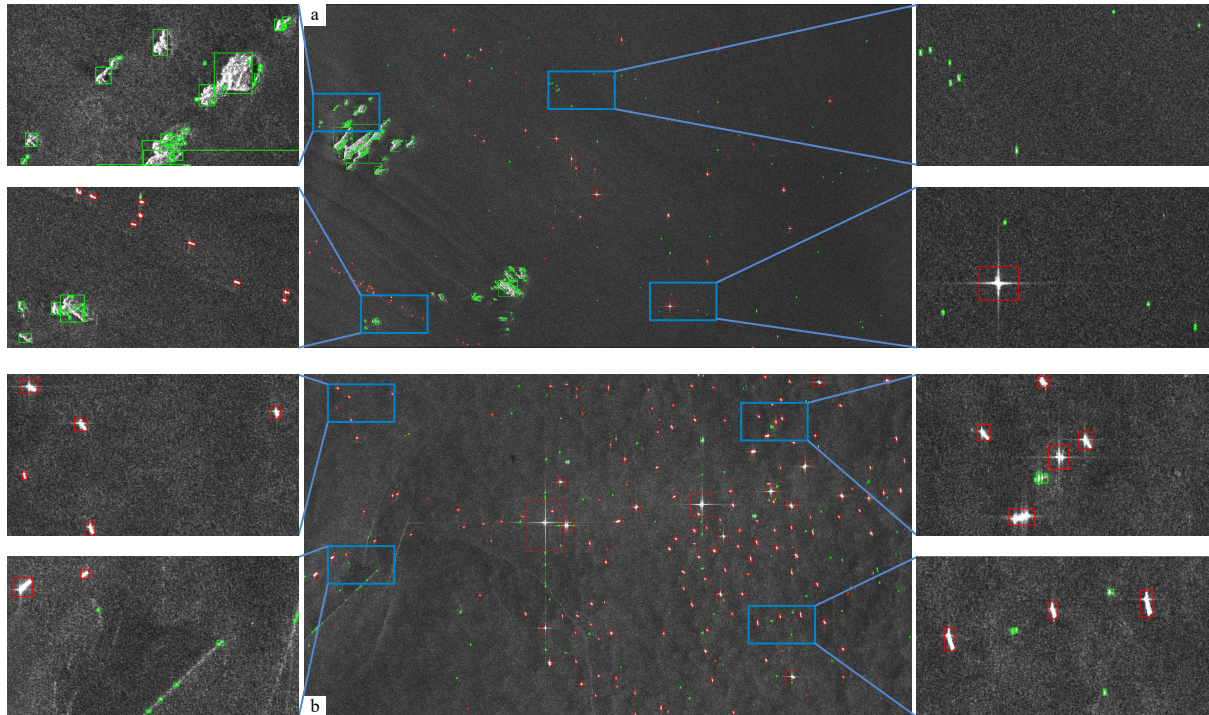


Fig. 3. The original labeled samples from subimages 1 and 2.

Table 4. Performance of different models pairing with different feature groups

No.	Feature groups	Gaussian SVM/%	Linear DA/%	Quadratic LR/%	Weighted KNN/%	Complex DT/%	Average/%
1	1_VV	96.80	97.50	96.80	96.80	97.50	97.08
2	1_VH	97.70	97.40	76.40	97.20	96.20	92.98
3	1_VV & VH	95.50	96.00	94.90	95.80	96.20	95.68
4	2_VV	76.60	78.40	82.80	77.40	81.00	79.24
5	2_VH	77.20	78.10	79.40	78.00	78.80	78.30
6	2_VV & VH	80.80	78.30	81.90	77.90	78.90	79.56
7	3_VV	88.40	87.30	95.60	96.10	95.60	92.60
8	3_VH	93.70	92.60	94.10	95.90	92.40	93.74
9	3_VV & VH	90.20	92.30	92.60	90.40	91.90	91.48
10	4_VV	97.20	97.20	96.40	97.10	97.70	97.12
11	4_VH	97.20	96.80	96.10	96.10	97.20	96.68
12	4_VV & VH	95.70	96.00	90.00	96.20	97.90	95.16
Average		90.58	90.66	89.75	91.24	91.78	

Note: 1_VV means the first feature group was obtained under VV polarization, and 1_VV & VH means the assembly of the first feature group was obtained under both VV and VH polarization. The best performance is highlighted in bold. SVM, support vector machine; DA, discriminant analysis; LR, logistic regression; KNN, knearest neighbor; DT, decision tree.

3.3 Comparative analysis of the discrimination results

3.3.1 State-of-the-art detection performance as comparison

Three subimages (Nos 3, 4, and 5) are used to carry out the contrast experiments, and as shown in Fig. 4, these scenes contain 195 verified ship targets, which are used as ground truth. In these subimages, the tiny islands are difficult to eliminate in the process of land masking, which is one of the main reasons for false alarms caused by the detectors. Besides, there are many strong ship targets accompanied by the typical phenomena of SAR images, such as azimuth ambiguity and sidelobes. These typical phenomena may not only mask weak ships but also be mistakenly detected as ships that cause false alarms. In addition, some relatively small ships are used to evaluate the ability of the discriminators to retain small ships.

The detection results of the CFAR detector and the state-of-the-art deep learning detector serve as a benchmark for verifying the performance of the discriminators. The CFAR detector under different p_{fa} settings from 10^{-1} to 10^{-10} is tested, and the best results are obtained by setting p_{fa} to 10^{-5} under VH polarization and 10^{-6} under VV polarization. Deep learning detectors include Faster R-CNN (Ren et al., 2017), Cascade R-CNN (Cai and Vasconcelos, 2021), Deformable ConvNets v2 (DCN) (Zhu et al., 2019), Deformable Transformers (DETR) (Zhu et al., 2021), PP-YOLOv2 (Huang et al., 2021) that have better target detection performance on ImageNet (Deng et al., 2009) have been selected. Then the LS-SSDD-v1.0 dataset (Zhang et al., 2020) is chosen to perform fine-tuning training with the same training set and test set (the first 6 000 chips as a training set and the remaining 3 000 chips as a test set). This dataset is constructed only from Sentinel-1 IW Level-1 GRDH images, contains 6 012 ships extracted from 15 large-scale images, which avoids possible influences from different SAR satellite parameters. These five detectors are trained with learning rate=0.001 25 and batch size=1. The detection performance of the fine-tuned deep learning detectors on the testing set of LS-SSDD-v1.0 is evaluated with mean average precision (mAP), and it is defined as follows:

$$mAP = \int_0^1 P(R) dR, \quad (5)$$

where P denotes precision, R denotes recall, and $P(R)$ denotes

the precision-recall curve.

$$\text{precision} = \frac{TP}{TP + FP}, \quad (6)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (7)$$

where TP denotes the number of true positives (correct detections), FN denotes that of false negatives (missed detections), and FP denotes that of false positives (false alarms).

As shown in Table 5, the Cascade R-CNN, Faster R-CNN, and DCN are significantly better (the mAP of them are 83.25%, 81.57%, and 82.58%, respectively), and it is higher than the optimal mAP of about 75% from the paper of LS-SSDD-v1.0. Hence, the trained Cascade R-CNN detector with better performance is applied to carry out the contrast experiments.

The detection results of the CFAR detector and Cascade R-CNN under VV polarization are shown in Figs 4 and 5. In such a complex background, there are a large number of false alarms caused by the CFAR detector. Fortunately, it detects almost all ship targets. The false alarms generated by Cascade R-CNN are obviously less, but there are still some small islands, a few azimuth ambiguities, and sidelobes, etc., that are wrongly detected as ship targets.

The detection results over these three subimages obtained by the CFAR detector and Cascade R-CNN are quantitatively evaluated by the following metrics, as shown in Table 6.

$$FOM = \frac{TP}{GT + FP}, \quad (8)$$

$$PoD = \frac{TP}{GT}, \quad (9)$$

$$FAR = \frac{FP}{GT}, \quad (10)$$

where FOM indicates the figure of merit, PoD indicates the probability of detection, and FAR indicates the false alarm rate. TP denotes the number of true positives (correct detections), GT denotes that of ground truth, and FP denotes that of false positives (false alarms).

From the above results, the CFAR detector is able to effect-

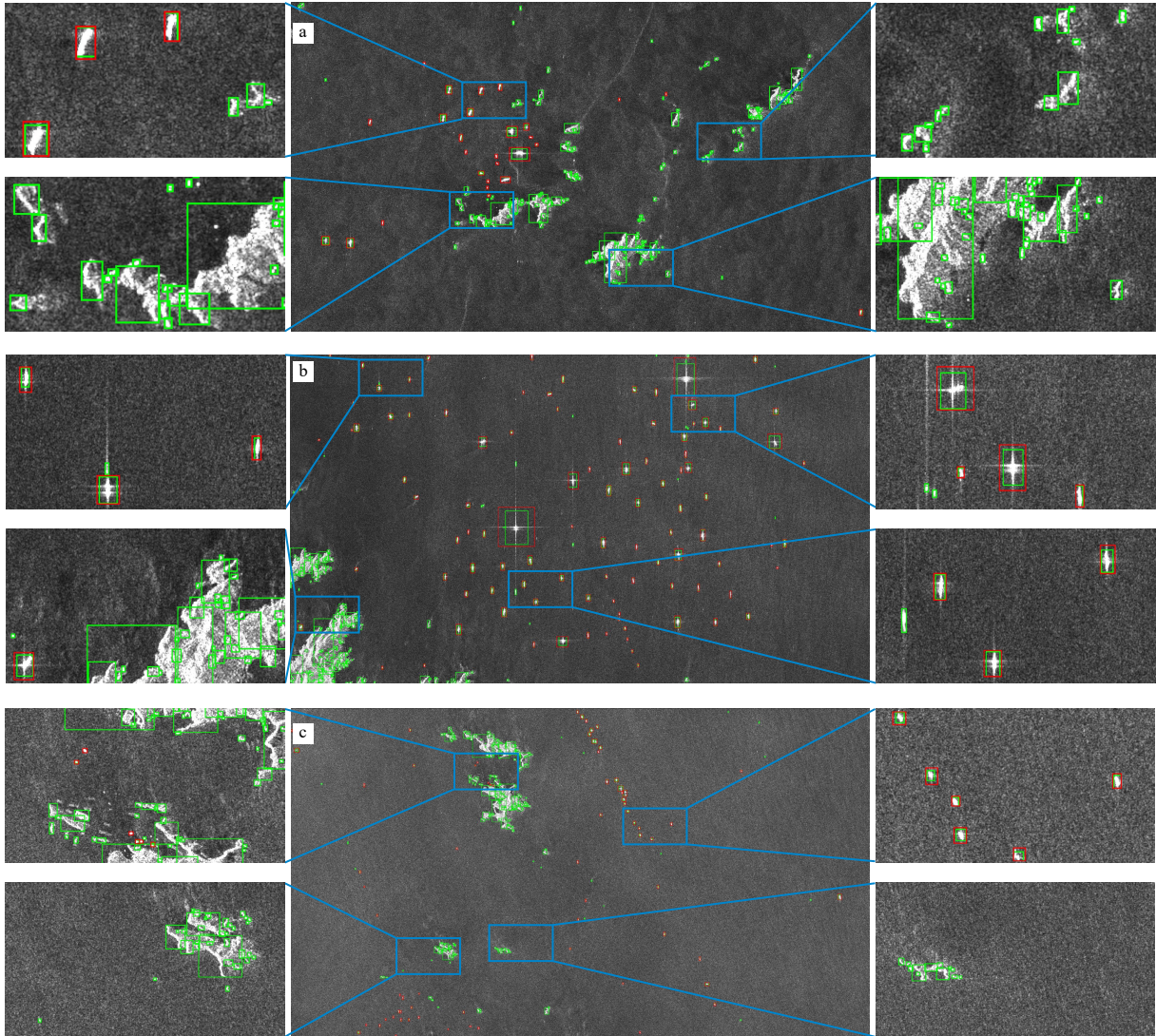


Fig. 4. Three subimages and the detection results by the constant false alarm rate detector. The green boxes denote the detection results; the red boxes denote the ground truth.

Table 5. The performance of five deep learning detectors on the testing set of LS-SSDD-v1.0

Model	Backbone	Epoch	mAP/%
Cascade R-CNN	ResNet50-vd-SSLDv2-FPN	12	83.25
Faster R-CNN	ResNet50-vd-SSLDv2-FPN	12	81.57
DCN	ResNet50-vd-FPN	12	82.58
DETR	ResNet50	300	58.87
PP-YOLO v2	ResNet50_vd	300	49.70

Note: In different backbone, vd means employing ResNet with version D (He et al., 2019), SSLD means employing Simple Semi-supervised Label Distillation (Cui et al., 2021), and FPN means employing Feature Pyramid Networks (Lin et al., 2017). DCN, Deformable ConvNets v2; DETR, Deformable Transformers. The best performance is highlighted in bold.

ively detect ship targets and ensure a high detection rate after setting an appropriate p_{fa} . However, a large number of false alarms are generated that are even more numerous than ship targets. Cascade R-CNN is comparable to the high detection rate of the CFAR detector, and the FAR is significantly lower. It is worth noting that, Cascade R-CNN has a good performance on the test set of LS-SSDD-v1.0. However, in terms of the practical application

capability, the FAR under VV polarization is close to 60% and under VH polarization is around 30%, which will still seriously affect subsequent applications, and the discrimination process is required.

3.3.2 Discrimination results

Multiple comparative experiments are carried out to comparatively analyze the effectiveness of the proposed method. The three classification models and three feature groups with the best performance (referring to the average accuracy in Table 4) are selected to implement the tri-training process, namely, the complex DT, weighted KNN, and linear DA models and the feature groups 1, 10, and 11. Specifically, traditional tri-training is denoted by TT, which is based on these three models and the No. 10 feature group. TT is improved by enhancing the diversity via randomly combining these three models with three feature groups is denoted by D-TT. The tri-training process is implemented by the top three discriminators obtained by preliminary screening and is denoted by PS-TT. In contrast, the proposed method employs all five initial discriminators.

The results over these three subimages obtained by the initial

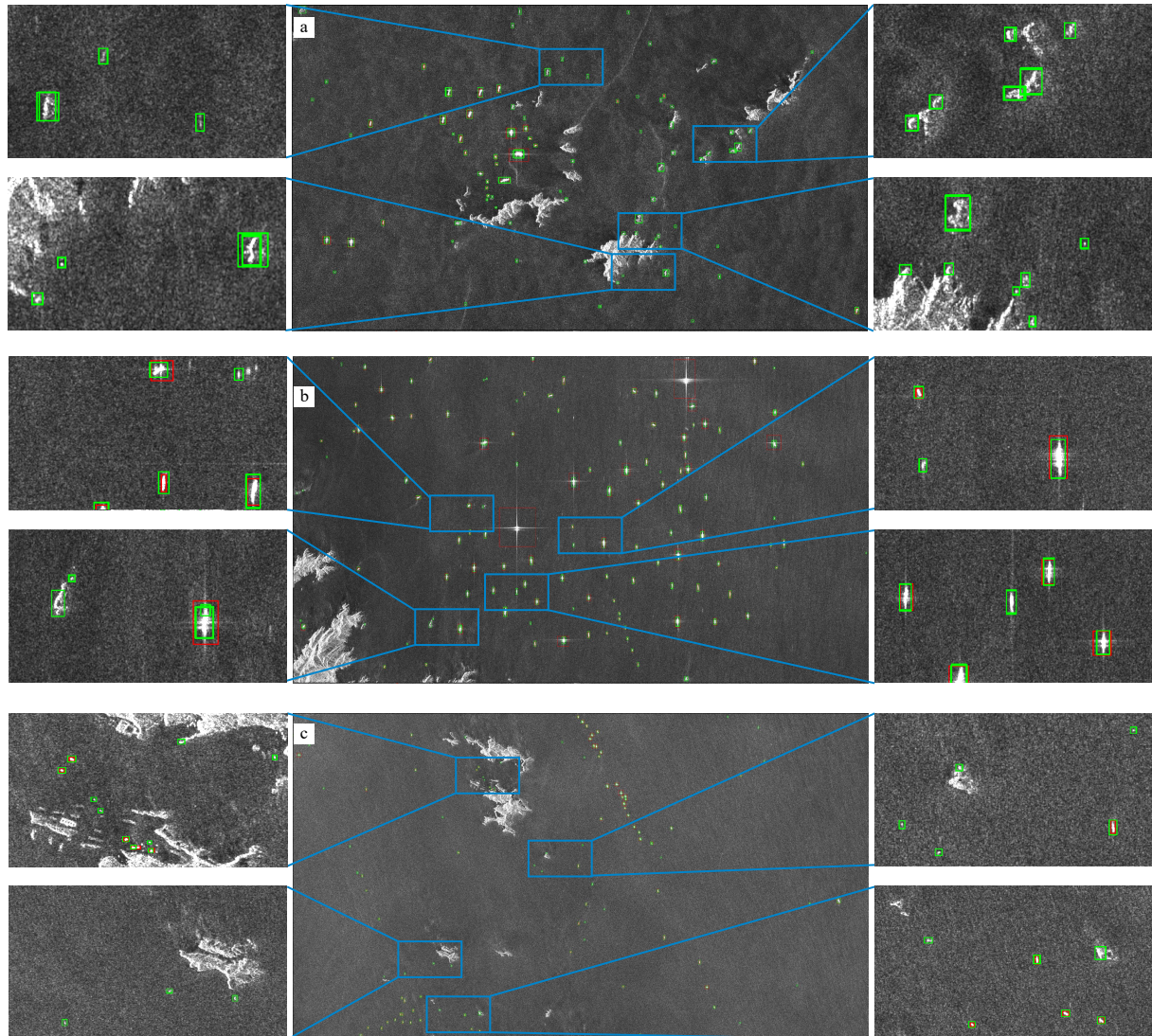


Fig. 5. Three subimages and the detection results by the Cascade R-CNN. The green boxes denote the detection results; the red boxes denote the ground truth.

Table 6. The results of the constant false alarm rate detector and Cascade R-CNN over three subimages

Methods	Missed ships	False alarms	PoD/%	FAR/%	FOM/%
K-CFAR_VV	4	183	97.95	93.85	50.53
K-CFAR_VH	1	257	99.49	131.79	42.92
Cascade R-CNN_VV	2	114	98.97	58.46	62.46
Cascade R-CNN_VH	5	64	97.44	32.82	73.36

Note: PoD indicates the probability of detection, FAR indicates the false alarm rate, and FOM indicates the figure of merit.

discriminators and the discriminators refined by the proposed method are shown in Table 7. In contrast, all the initial discriminators can greatly reduce false alarms while ensuring a high detection rate of more than 90%; among them, the complex DT is superior to the others. The FOM of each initial discriminator has been improved by the proposed training process as more ships are correctly identified and more false alarms are eliminated; specifically, the FAR dropped to less than 7%, and the PoD remained above 96%. The most obvious improvement is the weighted KNN with a FOM that increased more than 12%, and the refined quadratic LR model is more accurate than other dis-

Table 7. The discrimination results of the initial discriminators and refined discriminators

	Discriminators	Missed ships	False alarms	PoD/%	FAR/%	FOM/%
Initial	complex Tree	11	5	94.36	2.56	92.00
	linear DA	10	12	94.87	6.15	89.37
	weighted KNN	11	31	94.36	15.90	81.42
	Gaussian SVM	19	7	90.26	3.59	87.13
	quadratic LR	20	10	89.74	5.13	85.37
Refined	complex Tree	9	5	97.44	4.62	93.14
	linear DA	12	3	98.46	6.15	92.75
	weighted KNN	8	4	97.95	4.10	94.09
	Gaussian SVM	4	6	96.92	2.05	94.97
	quadratic LR	6	3	98.46	3.08	95.52

Note: PoD indicates the probability of detection, FAR indicates the false alarm rate, and FOM indicates the figure of merit. DA, discriminant analysis; KNN, knearest neighbor; SVM, support vector machine; LR, logistic regression.

criminators.

The discrimination results obtained by TT, D-TT, PS-TT, and the proposed method are shown in Table 8. The FOM of D-TT is

Table 8. The discrimination results of the different semisupervised methods

Methods	Missed ships	False alarms	PoD/%	FAR/%	FOM/%
TT	10	10	94.87	5.13	90.24
D-TT	10	5	94.87	2.56	92.50
PS-TT	5	8	97.44	4.10	93.63
Proposed method	2	3	98.97	1.54	97.47

Note: PoD indicates the probability of detection, FAR indicates the false alarm rate, and FOM indicates the figure of merit.

slightly higher than that of TT since the diversity of D-TT has been increased by the other two feature groups. PS-TT outper-

forms D-TT, showing the effectiveness of the preliminary screening for determining the initial discriminators so that fewer incorrect labels are selected in the training process. It is observed that the differences between TT, D-TT, and PS-TT are not obvious, as they are all constructed by the discriminators selected from [Table 4](#) with better performance and relatively small gaps. Compared with PS-TT, the results show that the proposed method can greatly reduce the number of false alarms while maintaining a high PoD, which will save considerable time in subsequent applications. The discrimination result images of TT, D-TT, PS-TT, and the proposed method are given in [Figs 6–8](#). All the targets are marked by bounding boxes; the correct detections are marked

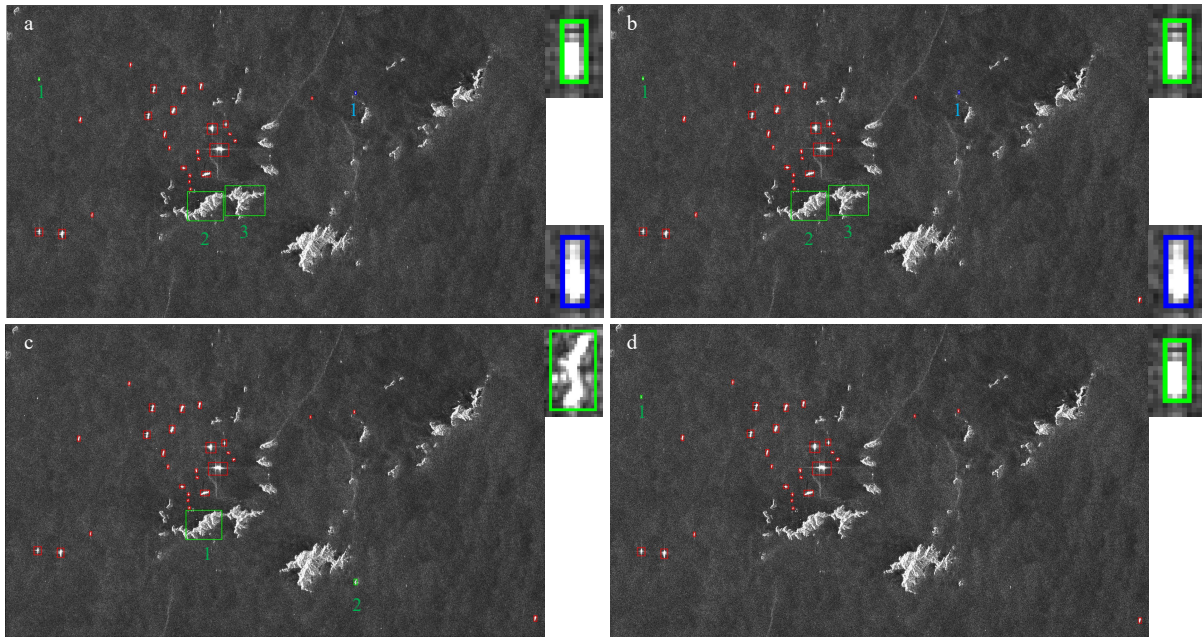


Fig. 6. Ship target discrimination results of subimage No. 3. a. Results by TT; b. results by D-TT; c. results by PS-TT; d. results by the proposed method. Correct detections are marked with red boxes, false alarms are marked with green boxes and are numbered, the missed ships are marked with blue boxes and are numbered.

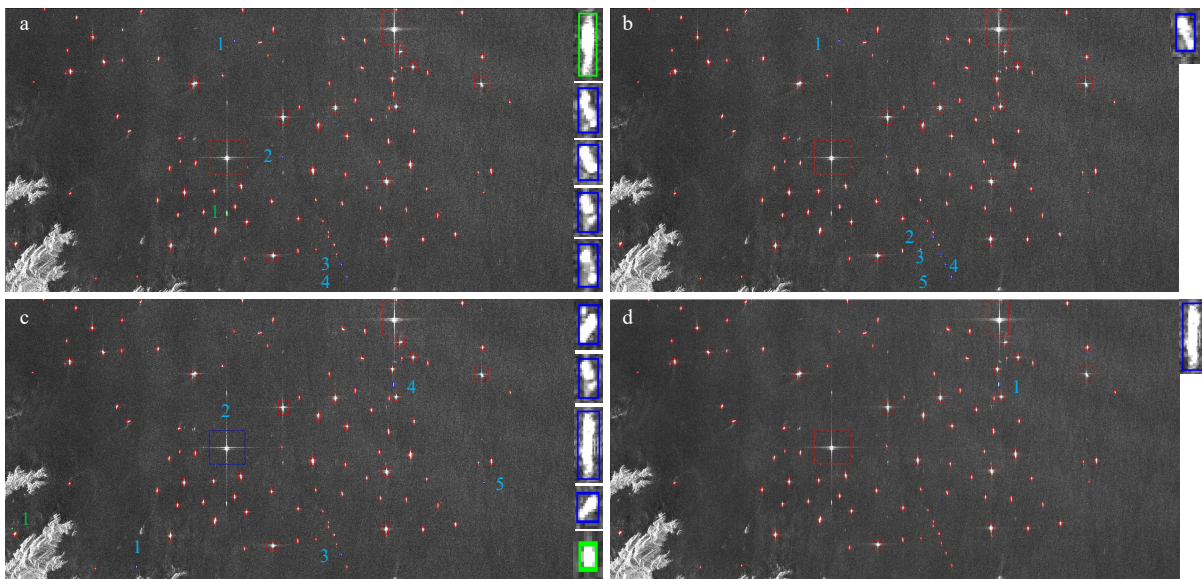


Fig. 7. Ship target discrimination results of subimage No. 4. a. Results by TT; b. results by D-TT; c. results by PS-TT; d. results by the proposed method. Correct detections are marked with red boxes, false alarms are marked with green boxes and are numbered, the missed ships are marked with blue boxes and are numbered.

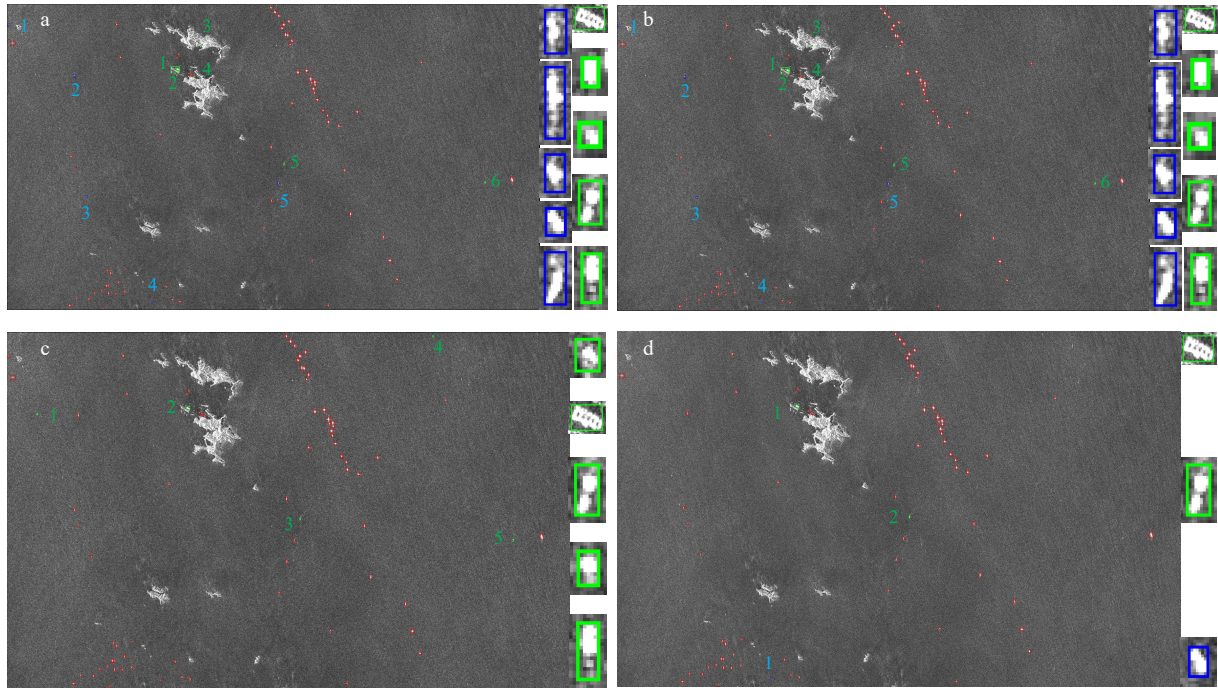


Fig. 8. Ship target discrimination results of subimage No. 5. a. Results by TT; b. results by D-TT; c. results by PS-TT; d. results by the proposed method. Correct detections are marked with red boxes, false alarms are marked with green boxes and are numbered, the missed ships are marked with blue boxes and are numbered.

with red boxes, the false alarms are marked with green boxes and are numbered, the missed ships are marked with blue boxes and are also numbered. At the same time, considering that some incorrectly discriminated targets are too small in these large scenes, these targets are extracted and listed on the right side of the resulting images.

As shown in Figs 6–8, all the methods show good discrimination competence and are able to select ships from most of the candidate targets and avoid false alarms. However, there are still a few misidentified targets, mainly due to missed small ships and false alarms caused by objects that have geometric and texture features similar to those of the ships. We noticed that all the methods can deal with false alarms caused by azimuth ambiguity, sidelobes, and the residual image caused by azimuth shift. In Figs 6 and 8, a small number of islands are misidentified as ships by TT, D-TT, and PS-TT. The proposed method is able to reduce such misjudgments. As seen in Figs 7 and 8, there are many small ships with only a few pixels in these images, some of which are affected by sidelobes or azimuth ambiguity of other strong targets around them that affect the prediction results, causing missed detections. TT, D-TT, and PS-TT all produced some missed detections, and the proposed method improved this situation. By comprehensive analysis, the results of these four SSL methods confirm that increasing the diversity and determining the appropriate initial discriminators can improve the discrimination performance. In addition, further improvements are achieved by adopting more heterogeneous discriminators. The test data support that the proposed method can stably improve the discrimination performance and has superior generalizability compared with the traditional tri-training.

4 Conclusions

In this paper, a semisupervised heterogeneous ensemble method is proposed for improving the performance of ship tar-

get discrimination in SAR images by mining the information from unlabeled samples. First, the discrimination ability of four kinds of SAR image features and five traditional classification models with their classic variants are investigated. Based on this analysis, compared with traditional tri-training, further improvements are achieved by increasing the diversity of the discriminators, determining the appropriate initial discriminators, and employing multiple heterogeneous discriminators. Experiments carried out with Sentinel-1 SAR images show that the optimal result of the CFAR detector is a FOM of 50.53%, and the Cascade R-CNN detector is a FOM of 73.36%. In contrast, the initial discriminators can greatly reduce false alarms while ensuring a high detection rate, and the highest FOM is 92.00%. Their performance improved after the training process, and the most obvious improvement is the weighted KNN with the FOM increasing by more than 12%. The more reliable result comes from the proposed method with a FOM of 97.47%, which increased more than 7% compared with traditional tri-training. In addition, the results show that the Old-Lincoln features contribute the most to the discrimination, and blindly employing all the features may lead to degradation due to feature redundancy. Therefore, future work will utilize feature manipulation to improve discrimination performance.

Acknowledgements

The authors are very grateful to the European Space Agency for providing the experimental dataset.

References

- Ai Jiaqiu, Pei Zhilin, Yao Baidong, et al. 2021. AIS data aided Rayleigh CFAR ship detection algorithm of multiple-target environment in SAR images. *IEEE Transactions on Aerospace and Electronic Systems*, doi: [10.1109/TAES.2021.3111849](https://doi.org/10.1109/TAES.2021.3111849)
- Aiello M, Vezzoli R, Gianinetto M. 2019. Object-based image analysis approach for vessel detection on optical and radar images. *Journal of Applied Remote Sensing*, 13(1): 014502

- Albukhanajer W A, Jin Yaochu, Briffa J A. 2017. Classifier ensembles for image identification using multi-objective Pareto features. *Neurocomputing*, 238: 316–327, doi: [10.1016/j.neucom.2017.01.067](https://doi.org/10.1016/j.neucom.2017.01.067)
- Amozegar M, Khorasani K. 2016. An ensemble of dynamic neural network identifiers for fault detection and isolation of gas turbine engines. *Neural Networks*, 76: 106–121, doi: [10.1016/j.neunet.2016.01.003](https://doi.org/10.1016/j.neunet.2016.01.003)
- Ao Wei, Xu Feng, Li Yongchen, et al. 2018. Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(2): 536–550, doi: [10.1109/JSTARS.2017.2787573](https://doi.org/10.1109/JSTARS.2017.2787573)
- Baudat G, Anouar F. 2000. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10): 2385–2404, doi: [10.1162/089976600300014980](https://doi.org/10.1162/089976600300014980)
- Belkin M, Niyogi P. 2004. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56(1): 209–239
- Bhanu B, Lin Yingqiang. 2003. Genetic algorithm based feature selection for target detection in SAR images. *Image and Vision Computing*, 21(7): 591–608, doi: [10.1016/S0262-8856\(03\)00057-X](https://doi.org/10.1016/S0262-8856(03)00057-X)
- Blum A, Mitchell T. 1998. Combining labeled and unlabeled data with co-training. In: *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*. Madison, WI: ACM, 92–100
- Cai Zhaowei, Vasconcelos N. 2021. Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5): 1483–1498, doi: [10.1109/TPAMI.2019.2956516](https://doi.org/10.1109/TPAMI.2019.2956516)
- Camps-Valls G, Marsheva T V B, Zhou Dengyong. 2007. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10): 3044–3054, doi: [10.1109/TGRS.2007.895416](https://doi.org/10.1109/TGRS.2007.895416)
- Chang Yanglang, Anagaw A, Chang Lena, et al. 2019. Ship detection based on YOLOv2 for SAR imagery. *Remote Sensing*, 11(7): 786, doi: [10.3390/rs11070786](https://doi.org/10.3390/rs11070786)
- Chen Shiyuan, Li Xiaojiang, Chi Shaoquan, et al. 2020. Ship target discrimination in SAR images based on BOW model with multiple features and spatial pyramid matching. *IEEE Access*, 8: 166071–166082, doi: [10.1109/ACCESS.2020.3022642](https://doi.org/10.1109/ACCESS.2020.3022642)
- Cui Cheng, Guo Ruoyu, Du Yuning, et al. 2021. Beyond self-supervision: A simple yet effective network distillation alternative to improve backbones. *arXiv preprint*, arXiv: 2103.05959
- Dasgupta S, Littman M L, McAllester D. 2001. PAC generalization bounds for co-training. In: *Proceedings of the 14th International Conference on Neural Information Processing Systems*. Vancouver, British Columbia: MIT Press, 375–382
- Deng Jia, Dong Wei, Socher R, et al. 2009. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE, 248–255
- Di Martino G, Iodice A, Riccio D, et al. 2014. Filtering of azimuth ambiguity in stripmap synthetic aperture radar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(9): 3967–3978, doi: [10.1109/JSTARS.2014.2320155](https://doi.org/10.1109/JSTARS.2014.2320155)
- Du Lan, Dai Hui, Wang Yan, et al. 2020. Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1): 461–472, doi: [10.1109/TGRS.2019.2937175](https://doi.org/10.1109/TGRS.2019.2937175)
- Falqueto L E, Sá J A S, Paes R L, et al. 2019. Oil rig recognition using convolutional neural network on Sentinel-1 SAR images. *IEEE Geoscience and Remote Sensing Letters*, 16(8): 1329–1333, doi: [10.1109/LGRS.2019.2894845](https://doi.org/10.1109/LGRS.2019.2894845)
- Gao Gui. 2011. An improved scheme for target discrimination in high-resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1): 277–294, doi: [10.1109/TGRS.2010.2052623](https://doi.org/10.1109/TGRS.2010.2052623)
- Gao Fei, Shi Wei, Wang Jun, et al. 2019. Enhanced feature extraction for ship detection from multi-resolution and multi-scene synthetic aperture radar (SAR) images. *Remote Sensing*, 11(22): 2694, doi: [10.3390/rs11222694](https://doi.org/10.3390/rs11222694)
- Haider N S, Singh B K, Periyasamy R, et al. 2019. Respiratory sound based classification of chronic obstructive pulmonary disease: A risk stratification approach in machine learning paradigm. *Journal of Medical Systems*, 43(8): 255, doi: [10.1007/s10916-019-1388-0](https://doi.org/10.1007/s10916-019-1388-0)
- He Jinglu, Wang Yinghua, Liu Hongwei, et al. 2018. A novel automatic PolSAR ship detection method based on superpixel-level local information measurement. *IEEE Geoscience and Remote Sensing Letters*, 15(3): 384–388, doi: [10.1109/LGRS.2017.2789204](https://doi.org/10.1109/LGRS.2017.2789204)
- He Tong, Zhang Zhi, Zhang Hang, et al. 2019. Bag of tricks for image classification with convolutional neural networks. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA: IEEE, 558–567
- Hua Wenqiang, Wang Shuang, Liu Hongying, et al. 2017. Semisupervised PolSAR image classification based on improved cotraining. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11): 4971–4986, doi: [10.1109/JSTARS.2017.2728067](https://doi.org/10.1109/JSTARS.2017.2728067)
- Huang Xin, Wang Xinxin, Lv Wenyu, et al. 2021. PP-YOLOv2: A practical object detector. *arXiv preprint*, arXiv: 2104.10419
- Hwang J I, Jung H S. 2018. Automatic ship detection using the artificial neural network and support vector machine from X-band SAR satellite images. *Remote Sensing*, 10(11): 1799, doi: [10.3390/rs10111799](https://doi.org/10.3390/rs10111799)
- Interferometric wide swath. 2020. [https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath\[2020-9-29\]](https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath[2020-9-29])
- Kang Miao, Ji Kefeng, Leng Xiangguang, et al. 2017. Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sensing*, 9(8): 860, doi: [10.3390/rs9080860](https://doi.org/10.3390/rs9080860)
- Kreithen D E, Halversen S D, Owirka G J. 1993. Discriminating targets from clutter. *The Lincoln Laboratory Journal*, 6(1): 25–52
- Lang Haitao, Tao Yunhong, Niu Lihui, et al. 2020. A new scattering similarity based metric for ship detection in polarimetric synthetic aperture radar image. *Acta Oceanologica Sinica*, 39(5): 145–150, doi: [10.1007/s13131-020-1563-7](https://doi.org/10.1007/s13131-020-1563-7)
- Lang Haitao, Zhang Jie, Zhang Xi, et al. 2016. Ship classification in SAR image by joint feature and classifier selection. *IEEE Geoscience and Remote Sensing Letters*, 13(2): 212–216, doi: [10.1109/LGRS.2015.2506570](https://doi.org/10.1109/LGRS.2015.2506570)
- Li Yongxu, Lai Xudong, Zhang Xi, et al. 2019. Comparative study of sea clutter distribution and ship detectors' performance for Sentinel-1 synthetic aperture radar image. *Journal of Applied Remote Sensing*, 13(4): 044506
- Lin T Y, Dollár P, Girshick R, et al. 2017. Feature pyramid networks for object detection. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE, 936–944
- Liu Hongying, Zhu Dexiang, Yang Shuyuan, et al. 2016. Semisupervised feature extraction with neighborhood constraints for polarimetric SAR classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7): 3001–3015, doi: [10.1109/JSTARS.2016.2532922](https://doi.org/10.1109/JSTARS.2016.2532922)
- Ma Liyong, Tang Lidan, Xie Wei, et al. 2018. Ship detection in SAR using extreme learning machine. In: *International Conference on Machine Learning and Intelligent Communications*. Heidelberg: Springer, 558–568
- Nigam K, McCallum A K, Thrun S, et al. 2000. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2): 103–134
- PaddlePaddle Authors. 2021. PaddlePaddle/Paddledetection: object detection and instance segmentation toolkit based on PaddlePaddle. <https://github.com/PaddlePaddle/PaddleDetection> [2021-10-20]
- Pelich R, Longépé N, Mercier G, et al. 2015. AIS-based evaluation of target detectors and SAR sensors characteristics for maritime surveillance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8): 3892–3901, doi: [10.1109/JSTARS.2015.2506570](https://doi.org/10.1109/JSTARS.2015.2506570)

- [10.1109/JSTARS.2014.2319195](https://doi.org/10.1109/JSTARS.2014.2319195)
- Ren Shaoqing, He Kaiming, Girshick R, et al. 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)
- Seijo-Pardo B, Porto-Díaz I, Bolón-Canedo V, et al. 2017. Ensemble feature selection: homogeneous and heterogeneous approaches. *Knowledge-Based Systems*, 118: 124–139, doi: [10.1016/j.knosys.2016.11.017](https://doi.org/10.1016/j.knosys.2016.11.017)
- Sentinel-1 Observation Scenario. 2020. [https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario\[2020-09-29\]](https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario[2020-09-29])
- Tello M, López-Martínez C, Mallorquí J J, et al. 2009. Advances in unsupervised ship detection with multiscale techniques. In: 2009 IEEE International Geoscience and Remote Sensing Symposium. Cape Town: IEEE, IV-979–IV-982
- Verbout S M, Weaver A L, Novak L M. 1998. New image features for discriminating targets from clutter. In: Proceedings Volume 3395, Radar Sensor Technology III. Orlando, FL: SPIE, 120–137
- Vespe M, Greidanus H. 2012. SAR image quality assessment and indicators for vessel and oil spill detection. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11): 4726–4734, doi: [10.1109/TGRS.2012.2190293](https://doi.org/10.1109/TGRS.2012.2190293)
- Wang Shuang, Guo Yanhe, Hua Wenqiang, et al. 2020. Semi-supervised PolSAR image classification based on improved tri-training with a minimum spanning tree. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12): 8583–8597, doi: [10.1109/TGRS.2020.2988982](https://doi.org/10.1109/TGRS.2020.2988982)
- Ward K D, Tough R J A, Watts S. 2006. *Sea Clutter: Scattering the K Distribution and Radar Performance*. London: The Institution of Engineering and Technology
- Zhang Tianwen, Zhang Xiaoling, Ke Xiao, et al. 2020. LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. *Remote Sensing*, 2020, 12(18): 2997
- Zhou Zhihua, Li Ming. 2005. Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on Knowledge and Data Engineering*, 17(11): 1529–1541, doi: [10.1109/TKDE.2005.186](https://doi.org/10.1109/TKDE.2005.186)
- Zhu Xizhou, Hu Han, Lin S, et al. 2019. Deformable ConvNets V2: More deformable, better results. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA: IEEE, 9300–9308
- Zhu Xizhou, Su Weijie, Lu Lewei, et al. 2021. Deformable DETR: Deformable transformers for end-to-end object detection. arXiv: 2010.04159