Hierarchical Sampling Representation Detector for Ship Detection in SAR Images

Ming Tong, Shenghua Fan, Jiu Jiang, and Chu He, *Senior Member, IEEE*

***Abstract*—Ship detection achieves great significance in remote sensing of synthetic aperture radar (SAR) and many efforts have been done in recent years. However, distinguishing ship targets precisely from the interference of multiplicative non-Gaussian coherent speckle is still a challenging task due to the discreteness, variability, and nonlinearity of ship scattering features. A detection framework based on hierarchical sampling representation is introduced to alleviate the phenomenon in this article. First, ships in SAR images exhibit multiplicative non-Gaussian coherent speckle, which introduces nonlinear characteristics under the imaging mechanism of SAR. Therefore, a statistical feature learning module is proposed with a learnable design to describe the nonlinear representations and expand the feature space. Second, our method designs a convex-hull representation to fit the irregular contours of ships represented by strong scattering points. Third, in order to supervise and optimize the regression of convex-hull representation, a sparse low-rank reassignment module is employed to evaluate the positive samples with SAR mechanism and reassign ones of high quality, which produces better results. Furthermore, experimental results on three authoritative SAR-oriented datasets for ship detection application present the comprehensive performance of our method.**

***Index Terms*—Convex-hull, ship detection, sparse and low-rank, statistical feature learning, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) serves as an active remote sensing technique that presents all-day observation with stability under all-weather circumstances. Typical target detection in SAR images has become a hot topic due to the rapid growth of high-quality data [1], [2], [3], [4], [5]. Ship detection reflects significance in the practical application of SAR remote sensing [6], [7], [8], [9], due to its key role in harbor management and maritime security.

Different from continuous linear texture features exhibited in optical remote sensing images, ship targets and backgrounds both exhibit multiplicative non-Gaussian coherent speckle with extremely low signal-to-noise ratio [10], which represents nonlinear characteristics under the unique imaging mechanism of SAR [11], causing two major problems in ship detection. First, in complex scenes, interference in the background may have similar backscatter intensity to ships [12], which leads to poor discrimination between ships and background utilizing first-order linear features such as grayscale intensity. Second, ship targets composed with speckles represent strong scattering points with irregular contours [13], indicating the presence of boundary uncertainty.

Taking the nonlinear characteristic of interference into account, researchers compensate for the first-order linear features such as grayscale intensity by learning differences in the statistical distribution of targets and backgrounds, which can be roughly divided into pixel-level statistical feature learning method and feature-level statistical feature learning method. Pixel-level statistical feature learning method [14], [15], [16] classifies pixels by estimating the local statistical distribution, which is guided through a constant false alarm rate (CFAR) on the decision criterion [17]. The paradigm performs sensitively to heuristic parameters of the statistical model, resulting in relatively low robustness. Feature-level statistical feature learning methods treat ship targets as one or more connected high-dimensional feature points to provide deeper statistical and structural information, among which the most representative one is superpixel segmentation [18], [19], [20], [21], [22], [23], [24]. The feature accuracy of these methods depends on the heuristic design of high-dimensional feature map generation procedure and decision vector. More recently, techniques based on convolutional neural network (CNN) have significantly boosted the state-of-the-art (SOTA) approaches for SAR object detection [25], [26], [27]. CNN-based methods cascade deep networks with conventional statistical feature learning methods at the feature or decision level to supplement nonlinear statistical feature descriptions [28], [29], [30]. However, the learning procedure of SAR image statistical characteristics by deep networks still relies on a simple cascade of heuristic designed statistical feature inputs. The network learning process adopts a large number of first-order linear transformations to extract multiple first-order moment features, which are performed to implicitly fit higher-order statistical features. As a result, it remains difficult to fit the nonlinear statistical features introduced by speckle phenomena, and the interpretability of the learning process is poor.

To ease the burden of boundary uncertainty and prediction for irregular contours, studies have focused on innovating SAR

Received 26 July 2024; revised 28 September 2024; accepted 21 October 2024. Date of publication 24 October 2024; date of current version 8 November 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0803000 and in part by the National Natural Science Foundation of China under Grant 41371342. *(Corresponding author: Chu He.)*

Ming Tong, Jiu Jiang, and Chu He are with the School of Electronic Information, Wuhan University, Wuhan 430079, China (e-mail: tongming@whu.edu.cn; jiangjiu@whu.edu.cn; chuhe@whu.edu.cn).

Shenghua Fan is with the School of Computer Science, Wuhan University, Wuhan 430079, China (e-mail: fanshenghua@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3485734

mechanism and developing mechanism detection models, which decompose SAR images or its feature maps into a pair of matrix components that represent ships with sparse characteristic and background with low-rank property. In other words, ship detection performs equivalently to application of sparse signal recovery [31]. These methods [32], [33], [34] can achieve strong stability and robustness with accurate prior knowledge such as sea clutter distribution prior, low-dimensional feature subspace vector prior, etc. CNN-based methods rely on the strong adaptability to implicit mechanisms through hierarchical learning without heuristic priors. A part of these methods introduces strong scattering points as the target sampling representation [13], [35], and explicitly supervises deep network learning using rectangular boxes composed of regression strong scattering point sets [36]. Another part implicitly guides feature learning direction through attention modules to optimize the accuracy of high-dimensional semantic features [37], [38], [39], [40]. Despite the fact that the above CNN-based methods achieve promising performance in SAR ship detection, fixed rectangular prediction boxes are still adopted to represent ship targets. The characterization for the mechanism of irregular target scattering points is not accurate, which makes it difficult to fit the feature of target scattering points. Besides, the above CNN-based methods rely on feature extraction and enhancement modules based on optical continuous texture to implicitly learn the mechanism of SAR ship targets, while ignoring the sparse low-rank mechanism unique to SAR targets and backgrounds.

In order to address the adverse effects of the absence of explicit statistics, mechanism learning, and fuzzy ship representation, a hierarchical ship detector is proposed in this article to fit SAR ship targets with statistical characteristics and unique mechanism, which is briefly presented in Fig. 1. Three modules are introduced to couple SAR characteristics and unique mechanism with CNN at three basic levels of detection network. First, for the level of feature extraction, a nonlinear description module with learnable ability is developed for the nonlinear features introduced by the multiplicative non-Gaussian speckle. The module is introduced to explicitly describe and fit the nonlinear statistical features introduced by the speckle effect through the network, which outputs statistical feature maps as the basis of the following procedure of representation regression. Second, for the level of representation regression, we introduce a fine-grained convex-hull representation learning module that fits the virtual contour of the space spanned by the internal texture through statistical features, to fit irregular scattering points of ships. Third, for the level of supervision, a reassignment module is introduced to supervise and optimize the above representation regression with explicitly modeling the sparse low-rank mechanism.

The contributions of this article can be summarized as follows.

1) A feature description module for point statistical learning named quadratic statistical learning module (QSLM) is designed for explicitly describing the nonlinear statistical features introduced by fitting the speckle effect, and expanding the feature space to compensate for the lack of nonlinear statistical feature learning with the convolution module.
2) A convex-hull generation module (CGM) with sparse low-rank convex-hull reassignment (SLCR) is proposed to optimize the fuzzy representation for rectangular areas of ship targets. By introducing a sparse low-rank mechanism between target sampling points and background, accurate regression of polygonal representation is supervised and optimized.
3) A hierarchical sampling representation framework is developed to adapt to the application of SAR ship detection. According to the SAR statistical characteristics and imaging mechanism, this framework describes the nonlinear statistical features introduced by the speckle effect and extracts the polygonal representation of ship targets with sparse low-rank mechanism learning.

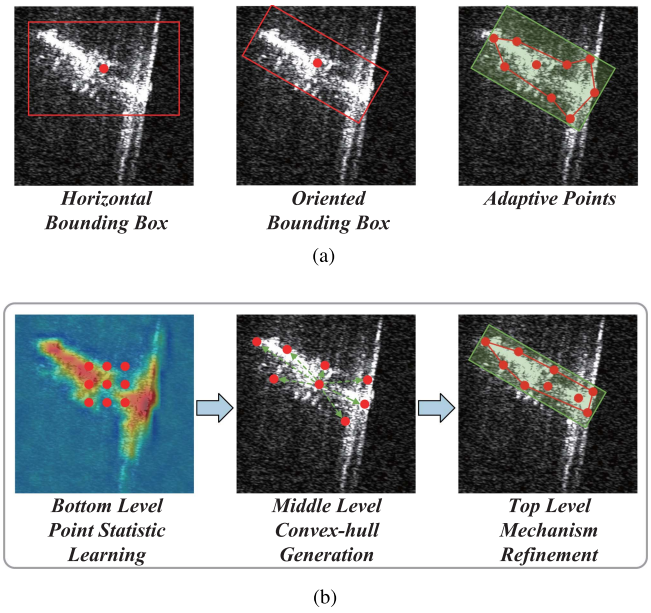


Fig. 1. Illustration of sampling representation in our method and comparisons of traditional representations. (a) Denotes the common sampling representations in object detection tasks, respectively. (b) Presents sampling representation in our framework. A precise representation explicitly fitted by statistical features and mechanism is proposed to optimize the representations designed for objects in optical images, which alleviates the lack of description for nonlinear features and the presence of boundary uncertainty caused by coherent speckle. (a) Traditional sampling representation. (b) Hierarchical sampling representation.

The rest of this article is organized as follows. Related works are introduced in Section II, and overall architecture with model details in our method is described in Section III. The following results of experiments and ablation analysis are presented in Section IV. Finally, Section V concludes this article.

II. Related Work

In this section, previous methods of SAR ship detection and related object detection in natural images are summarized. The descriptions of CNN-based statistic learning methods, sparse low-rank mechanism learning methods, and CNN-based object detection methods with convex-hull representation, are provided below, respectively.

A. Statistical Feature Learning of CNN-Based Object Detectors

CNN-based object detection methods remain challenging applied to SAR scenes on account of the domain mismatch from natural scene, which present unique characteristics and mechanism. One of the most representative characteristics of SAR image is the nonlinear statistical features introduced by fitting the speckle effect.

Recently, researchers have introduced various methods to couple the CNN framework with traditional statistical feature learning. Kang et al. [29] introduced the region proposed network in Faster R-CNN to the prediction of protected window of sea clutter statistics in CFAR ship detection. Khesali et al. [28] presented heuristic extraction of texture features as input to CNN detector. Guo et al. [30] performed GMM modeling to fit the distribution of strong scattering point areas extracted from scattering information with clustering and introduce enhanced strong scattering point area as a single channel input to CNN. Liu et al. [41] designed a single-layer network to learn statistical feature as input to ordinary CNN. However, the lack of effective strategy to integrate statistical characteristics and CNNs remains in the above methods. The learning procedure of statistical characteristics by deep networks relies on a simple cascade of heuristic designed statistical feature inputs with relatively poor interpretability. Therefore, a learnable nonlinear description CNN module is involved in this article.

B. Sparse Low-Rank Mechanism Learning

SAR ship detection task can be regarded as a process pf sparse signal recovery, in which ships present sparse characteristic with its pixels sparsely distributed, while sea clutter can be represented as low-rank feature matrix due to the random variety of its pixel values. Specifically, ship detectors based on sparse low-rank mechanism can be classified to global inference and partial window under the decision criterion whether a window sliding strategy is adopted in background clutter dictionary.

As for global inference research, Song et al. [31] developed the theoretical basis that ship detection can be equivalent to decomposition of SAR images into a matrix component that represents ship and another that represents background by introducing a probabilistic model. Zhang et al. [42] decomposed a circular SAR image sequence as low-rank static clutter component and sparse target component to detect moving targets. Li et al. [34] intercorporated the low-rank property and sparsity constrains into the SAR image reconstruction model with the goal of SAR image reconstruction and target extraction from undersampled data. As for partial window methods, Wang et al. [33] developed a superpixel detector for SAR ships, in which the Fisher vector of each superpixel is extracted to represent the feature of it, and judge the superpixel through the criteria that the linear combination of the dictionary for background feature vectors can formulate feature vector of any background superpixel. Although the above methods based on mechanism learning perform relatively acceptable in ship detection tasks of homogeneous background, they suffer the burden of numerous hyperparameters for background clutter modeling and sparse decomposition design to fit the practical changing scene. Therefore, a SLCR module is added to CNN framework for adaptively learning ship mechanism with high robustness.

C. Convex-Hull Representation

Orientation representation is introduced to classical object detectors to model the varying-oriented targets in optical scenes. Ding et al. [43] optimized the learning process of orientation representation through introducing an improved rotated RoI module, and R^3Det [44], ReDet [45], and Oriented RCNN [46] constantly improve the representation of oriented bounding box to achieve SOTA, subsequently. However, the representation is designed to fit the CNN feature grids that are axis-aligned, as a result of which it partly fails to adapt objects with complex geometric structure in dense scenes. To solve the problem, Guo et al. [47] introduced convex-hull representation composed of adaptive point set to dense object detection, which avoids feature aliasing via learnable feature adaption. Li et al. [48] optimized the learning process for adaptive points of convex-hull representation with an improved samples assignment strategy. There still remains gap of semantic representation between ship with speckle noise, which exhibits strong scattering point set with irregular layouts, and features on regular grids extracted by the existing rotation-robust detection methods. We model an adaptive ship representation to alleviate the fuzzy localization with SLCR strategy.

III. Methods

The overall framework of our method is illustrated by Fig. 2, and three-level modules are introduced in the proposed method: the point statistical learning module for extracting nonlinear statistic features, the CGM for generating ship representations, the convex-hull reassignment module to supervise the optimization process with sparse low-rank model.

The proposed method adopts a two-stage architecture. In feature extraction part, ResNet-50 backbone is introduced to extract convolution feature maps. Additionally, QSLM is incorporated to learn nonlinear statistical feature maps, which combine convolution feature maps to output features of feature extraction part. In the representation regression part, we introduce convex-hull generation module (CGM) to learn fine-grained convex-hull representation of ship targets, utilizing output features of feature extraction part. The fitting procedure of convex-hull representation adopts a two-stage regression of adaptive feature points to optimize the accuracy of localization. Finally, in the supervision part only applied in the training procedure, the sparse low-rank convex-hull reassignment (SLCR) module is introduced to reassign the positive samples of convex-hulls generated by representation regression part with SAR mechanism, which directs the learning of network to fit SAR characteristics and mechanisms. The proposed hierarchical modules are detailed in the following.

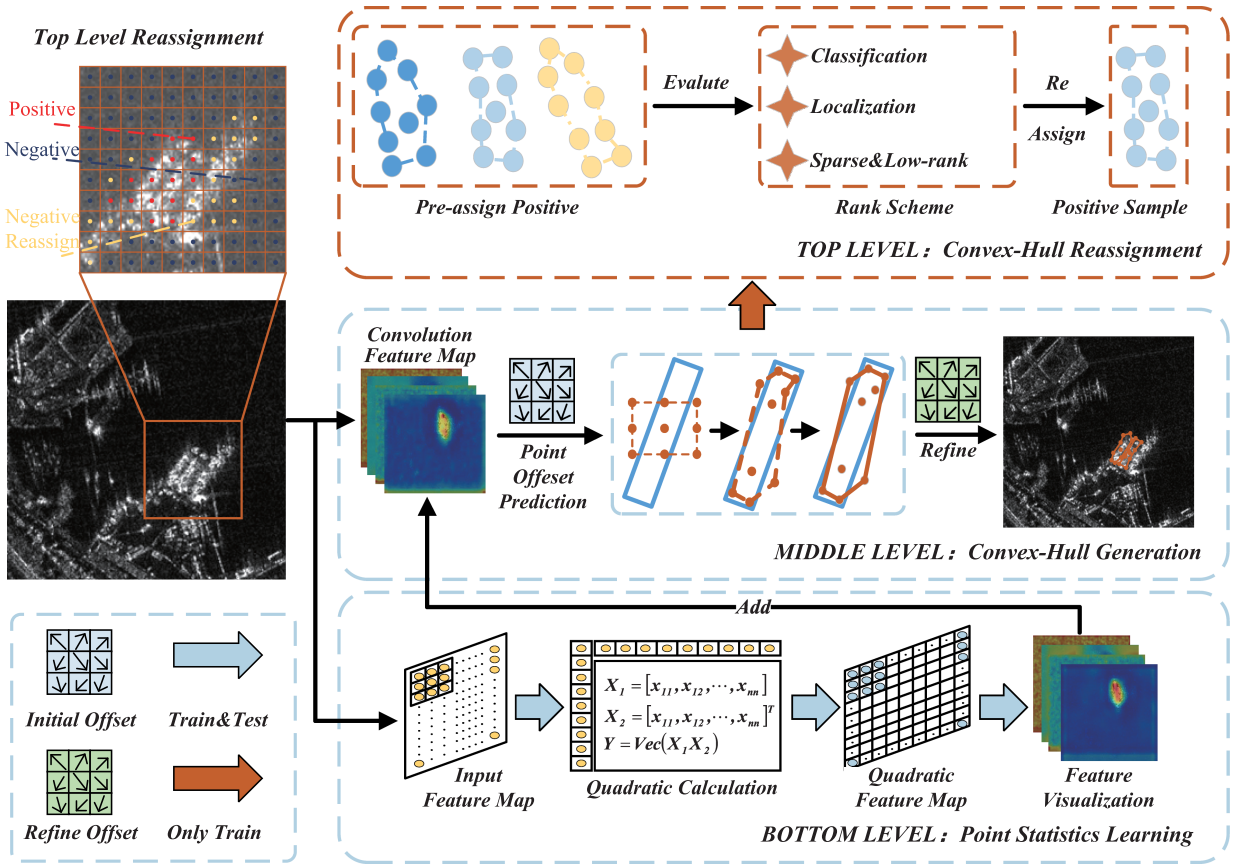


Fig. 2. Overall architecture of our method. The convex-hull representation is learned through an anchor-free framework, and the convolution feature map is attained through a backbone with FPN. The reassignment scheme of convex-hull samples is only applied to the proposed framework in the training phase.

A. Point Statistical Learning Module

Architecture of fully convolution in CNN-based object detection methods presents powerful capability of feature extraction and achieves stability in varying application scenes. However, the application of a large number of first-order linear transformations (convolution) supplemented by nonlinear dimensionality reduction (pooling) and quantitative highlighting (activation) to extract multiple first-order moment features, which are performed to implicitly fit higher order statistical features, remains difficult to fit the nonlinear statistical features introduced by speckle phenomena, and the interpretability of the learning process is poor.

Instead, in this article, we employ QSLM to mitigate this problem. Additionally, the module is essentially a nonlinear description module aimed at explicitly describing the nonlinear statistical features introduced by the fitting speckle effect, which compensates for the loss of linearly indivisible features.

As an active remote sensing radar, SAR antenna transmits electromagnetic waves and receives echo signals to produce SAR images [49]. The backscattering signals can be represented as the coherent sum of the large number of elementary scatterers within a resolution cell. Based on the above physical process, inherent variation arises as a result of interference among coherent echoes, which is called speckle [50]. The expression of SAR backscattering signal is formulated as follows:

$$S_{bs} = \sum_{k=1}^{N} A_k e^{j\varphi_k} \tag{1}$$

where S_{bs} denotes the SAR backscattering signal. N denotes the number of scatters located within the resolution cell, A_k and φ_k indicate amplitude and phase of the kth backscattering echoes, respectively. Speckle presents as fluctuations at pixel level among adjacent resolution cells in S_{bs}, causing the intrinsic randomness of SAR signals [49].

Although speckle is suppressed as noise to approximate the paradigm of optical image interpretation in some circumstances, characterization of the speckle statistics becomes a principal research domain in SAR interpretation to extract variation information [50], [51], [52], [53]. The construction of a specific mathematical model of the statistical analysis is extensively carried out in SAR domain as for the intrinsic randomness of SAR signals with speckle effect [54]. Numerous statistic distributions are introduced to approximate scattering environments [49], [55], [56], [57], [58], [59], as are presented in Fig. 3. Estimating parameters of statistical model is a significant method to define it, one of which is commonly used is the method of moments (MoM). MoM estimations of typical statistical models are presented in Table I.

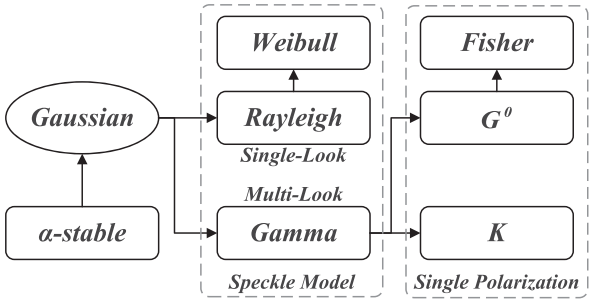


Fig. 3. Model and relationship of statistical distribution for SAR images.

TABLE I
MoM Estimation of Typical Distributions

Distribution model	Estimation parameters
Rayleigh [55]	$b = \sqrt{\frac{2}{\pi}} m_1$
Weibull [56]	$\frac{\Gamma(1+\frac{2}{c})}{\Gamma(1+\frac{1}{c})^2} = \frac{m_2}{m_1^2}$ $b = \frac{m_1}{\Gamma(1+\frac{1}{c})}$
Gamma [49]	$\mu = m_1$ $L = \frac{m_1^2}{m_2 - m1^2}$
$\mathcal{K}$ [57]	$\mu = m_1$ $v = \frac{(L+1)m_1^2}{L(m_2 - m_1^2) - m_1^2}$
G^0 [58]	$\alpha = -1 - \frac{Lm_2}{Lm_2 - (L+1)m_1^2}$ $\gamma = (-\alpha - 1)m_1$
Fisher [59]	$M = \frac{Lm_2}{L(m_2 - m_1^2) - m_1^2} + 1$ $\mu = (1 - \frac{1}{M})m_1$

m_1, m_1^2, and m_2 denote the first-order moment, the square of first-order moment and the second-order moment, respectively. L denotes the number of looks and $\Gamma(\cdot)$ denotes Gamma function, namely, $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$. As is shown in Table I, moment estimation is basically composed of the first-order moment with its square and the second-order moment as its elementary components. In other words, the statistical feature of SAR data can be described by fitting the above three components.

Considering the input feature map which can be denoted by a three-order tensor $X \in \mathbb{R}^{\mathrm{H} \times \mathrm{W} \times \mathrm{C}}$, as H and W indicate height and width in spatial dimension of it, respectively, while C reflects the channelwise number of feature maps, the vectorized form for a given channel can be represented as $\boldsymbol{x_i} = [x_1, x_2, \ldots, x_m]^T$, where $m = \mathrm{H} \times \mathrm{W}$ and $i = 1, \ldots, \mathrm{C}$. These moment values of $\boldsymbol{x_i}$ are formulated as follows:

$$m_1 = \frac{1}{m} \sum_{j=1}^{m} x_j \tag{2}$$

$$m_2 = \frac{1}{m} \sum_{j=1}^{m} x_j^2 \tag{3}$$

where m_1 and m_2 characterize the statistic distribution as first- and second-order moments, respectively.

As above, m_1 serves as an exceptional case of convolution operation with its kernel weights set to $\frac{1}{m}$. However, m_2 and ${m_1}^2$ cannot be achieved through convolution operations because of its quadratic terms x_j^2 and cross terms $x_j x_k$, As a result of which CNN detectors cannot fit the nonlinear statistical feature explicitly. QSLM is inspired by quadratic form in linear algebra theory, which is introduced to fit m_2 and m_1^2. Above quadratic terms and cross terms can be expressed by a quadratic polynomial with quadratic primitive introduced to CNN in this article, which can be formulated as

$$\begin{aligned} Q_1(x_1, x_2, \ldots, x_m) &= q_{11} x_1^2 + q_{12} x_1 x_2 + \ldots + \\ & q_{1m} x_1 x_m + q_{21} x_2 x_1 + q_{22} x_2^2 + \ldots + \\ & q_{2m} x_2 x_m + \ldots + q_{m1} x_m x_1 + q_{m2} x_m x_2 + q_{mm} x_m^2 \\ &= \sum_{j=1}^{m} q_{jj} x_j^2 + 2 \sum_{k=1}^{m-1} \sum_{l=k+1}^{m} q_{kl} x_k x_l \\ &= \omega_Q^T \mathrm{Vec}(\boldsymbol{x_i} \boldsymbol{x_i^T}) \end{aligned} \tag{4}$$

where $Q_1(x_1, x_2, \ldots, x_m)$ denotes the output feature map of quadratic primitive, q_{jj} and q_{kl} indicate the weights of quadratic and cross terms, both of which are introduced to fit m_2 and m_1^2. Vector expression above in (4) is abstracted to simplify the proposed quadratic primitive, where $\omega_Q = (q_{11}, q_{12}, \ldots, q_{mm})^T$, which can be learned by deep network as matrix of kernel weight equivalent to quadratic and cross weights, and $\mathrm{Vec}(\cdot)$ denotes the vectorization operation.

In contrast, convolution primitive in CNNs can be formulated as

$$\begin{aligned} Q_2(x_1, x_2, \ldots, x_m) &= a_1 x_1 + a_2 x_2 + \ldots + a_m x_m \\ &= \sum_{j=1}^{m} a_j x_j = \omega_C^T \boldsymbol{x_i} \end{aligned} \tag{5}$$

where $Q_2(x_1, x_2, \ldots, x_m)$ denotes the output feature map of convolution primitive in CNNs and $\omega_C = (a_1, a_1, \ldots, a_m)^T$ denotes the learnable weights of convolution kernel.

To avoid discarding the learning of first-order linear feature, both output feature maps of convolution primitive and proposed quadratic primitive are combined in QSLM with elementwise addition followed by activation in backbone to characterize the distribution of features. The visualization of the calculation process for a 2×2 sliding window is shown in Fig. 4. Therefore, operation of QSLM can be formulated as

$$Q(x_1, x_2, \ldots, x_m) = \omega_C^T \boldsymbol{x_i} + \omega_Q^T \mathrm{Vec}(\boldsymbol{x_i} \boldsymbol{x_i^T}) \tag{6}$$

where $Q(x_1, x_2, \ldots, x_m)$ denotes the output feature map of QSLM in this article.

In details, the kernel of primitive in QSLM is designed to size of 3×3. Inspired by the convolution primitive, local feature map with equal size of above kernel is dimensionality reduced to 1-D feature vector, which performs autocorrelation subsequently to generate quadratic feature map $\boldsymbol{x_i} \boldsymbol{x_i^T}$ in (6).

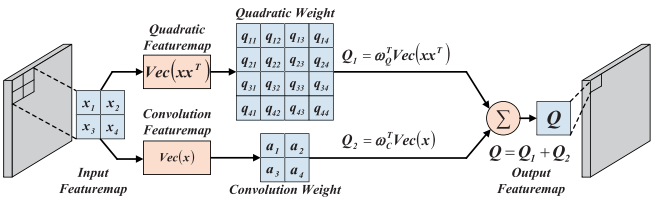


Fig. 4. Architecture of the proposed QSLM. Here, visualization of the calculation process for a 2×2 sliding window is shown, where the learnable quadratic weights and convolution weights are vectorized as $\boldsymbol{\omega_Q}$ and $\boldsymbol{\omega_C}$, respectively. Besides, output of QSLM maintains consistency in spatial dimension with input.

To further analyze the statistical learning of feature maps generated by cascade convolutions in CNN, we denote input SAR image with N_0 pixels as X^0. The i_0th pixel of N_0 can be represented as $x_{i_0}^0$. Then, the i_1th pixel of output feature map X^1 generated by the first convolution can be formulated as follows:

$$x_{i_1}^1 = \sum_{i_0=1}^{N_1} a_{i_0}^1 x_{i_0}^0 \tag{7}$$

where N_1 denotes the number of cells in the convolution kernel, and $a_{i_0}^1$ indicates the learning coefficient of convolution kernel at i_0th pixel. Analogously, the i_mth pixel of output feature map X^m generated by the mth convolution can be formulated as follows:

$$x_{i_m}^m = \sum_{i_{m-1}=1}^{N_m} a_{i_{m-1}}^m x_{i_{m-1}}^{m-1} \tag{8}$$

where N_m denotes the number of cells in the mth convolution kernel, and $a_{i_{m-1}}^m$ indicates the learning coefficient of mth convolution kernel at i_{m-1}th pixel of $(m-1)$th feature map. Taking X^m as input of quadratic primitive, the i_{m+1}th pixel $p_{i_{m+1}}^{m+1}$ of output feature map Q^{m+1} generated by the quadratic primitive can be calculated as follows:

$$p_{i_{m+1}}^{m+1} = \sum_{i_m=1}^{N_{m+1}} q_{i_m i_m} (x_{i_m}^m)^2 + 2 \sum_{k_m=1}^{N_{m+1}-1} \sum_{l_m=k_m+1}^{N_{m+1}} q_{k_m l_m} x_{k_m}^m x_{l_m}^m \tag{9}$$

where N_{m+1} denotes the number of cells in the $(m+1)$th quadratic kernel. $q_{i_m i_m}$ and $q_{k_m l_m}$ indicate the learning coefficients of quadratic and cross terms.

We can find out the fact that the module learns permutation and combination among multilayer coefficients for the cross terms of pixels within original images, when the above formula is expanded. The phenomenon theoretically represents the statistical feature learning of the module among neighborhoods of pixel spaces in the original image, which corresponds to the theory of feature-level statistical feature learning methods, especially superpixel ones, that are mentioned in Section I. In summary, QSLM fits deeper statistical and structural features through high-level feature maps that reflect statistical distribution among neighborhoods of pixel spaces in SAR images.

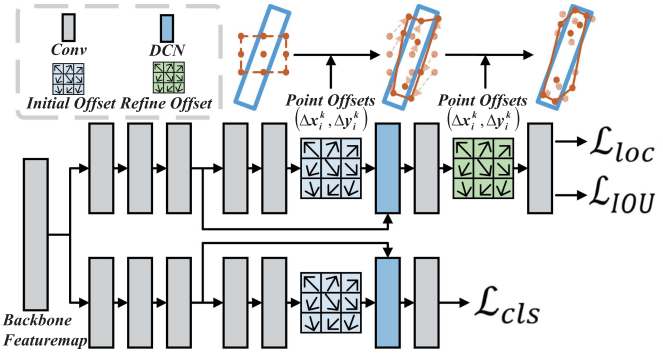


Fig. 5. Architecture of the CGM. The opaque dots reflect the adaptive representation points of the current stage, while the transparent dots represent ones of the previous stage. The transparent dashed lines reflect the 2-D directions of offsets. Besides, Conv and DCN denote convolution and deformable convolution, respectively. $\mathcal{L}_{\text{loc}}$, $\mathcal{L}_{\text{IOU}}$, and $\mathcal{L}_{\text{cls}}$ represent loss of localization, IOU, and classification, respectively.

B. Convex-Hull Generation Module

Most ships exhibit discrete set of strong scattering points composed of speckle noise in SAR images due to the physical property that the metal shells of ships possess strong scattering properties of electromagnetic, as a result of which boundary uncertainty of ship targets exists. Existing CNN-based methods still use fixed rectangular prediction boxes for ship representation to fit axis-aligned CNN feature grids, which is difficult to adapt features to cluster of strong scattering points with irregular layouts in SAR images.

Convex-hulls are introduced to represent contours of ships in the proposed method, each of which consists of learnable adaptive points to characterize the distribution of strong scattering points in SAR ship. The CGM is proposed to optimize the fuzzy representation of rectangular areas of ship targets and learn to delimit the contour of ship through adaptive feature points in this article.

Inspired by [47], the convex-hull representation of ship targets is defined as

$$R_i = \{(x_i^k, y_i^k)\}_i^{k=1,2,\ldots,9} \tag{10}$$

where R_i presents convex-hull representation for the ith location (x_i^k, y_i^k) on the feature maps $X \in \mathbb{R}^{H\times W\times C}$ consisting of 9 representative points, which are located as 3×3 grids at the initialization phase.

As shown in Fig. 5, proposed CGM consists of two stages following RepPoints [60] framework. At initialization stage, the backbone feature map is input to localization and classification subnets. Three 3×3 convolution layers followed by a 2-D offsets prediction head are introduced to the localization subnet. The initialized representative point sets are generated by object center point hypothesis (feature map bins) at the initialization stage, which is equivalent to predict an offset $(\Delta x_i^k, \Delta y_i^k)$ for each representative point from the center point with convolution operation. The classification subnet applies the same structure

with offsets prediction part of localization one, with a 3×3 deformable convolution layer followed.

At the refinement stage, a 3×3 deformable convolution layer is introduced to predict the refine offsets and updates the accurate set of representative points by refining from the first point set in the localization subnet. The Jarvis March algorithm [61] is introduced to generate a minimal convex-hull of ship targets on the refined point set above. Since the algorithm will determine whether each representation point serves as the outer contour of the minimum convex-hull, the number of representation points that compose the outer contour is less than or equal to 9.

In details, GIOU [62] is introduced to construct positive candidates between predictions and labels. Classification is conducted only on the first stage of CGM and for both stages, only positive convex-hull are assigned with localization targets in training.

To further elucidate the advantage of CGM representation, the elaborations of [60] and [47] are introduced in this part. [60] applies an initialization followed by refinement learning procedure for adaptive points to compose a representation of horizontal bounding box, which is designed for targets in fixed orientations of natural images. Guo et al. [47] reserved the learning procedure and improved the representation as convex to fit the targets in arbitrary orientations of optical remote sensing images. Meanwhile, an optimized module is introduced to refinement stage to fit densely packed objects.

In CGM, we innovatively introduce the convex-hull representation to fit the irregular contour of SAR ship granularly. However, the edge of the SAR ship target presents as virtual contour that serves as equivalent to space spanned by internal texture of target, as a result of which it fails to be learned by edge fitting. To optimize the domain gap mentioned above, as for feature extraction, statistical feature maps are introduced to learn convex representation in order to implicitly fit the internal texture of target, which is different from fitting procedure for representation of the above methods that are rooted in learning linear grayscale variation of object edge in optical images. As for the optimization, a reassignment module based on SAR mechanism is introduced to the refinement stage to explicitly supervise the fitting procedure of the virtual contour, which differs from the above methods in the supervision of refinement stage, as the natural fitting procedure of CNNs can already satisfy the representation of target edges in natural or optical remote sensing images.

C. SLCR Module

Although the proposed convex-hull representation facilitates localizing ships with arbitrary layouts, learning high-quality convex-hull points coincident with the mechanism of ship targets and background is essential to capture features with the unique SAR mechanism adaptively for ship localization. To this end, we proposed an assignment scheme to measure how the learned representation accords with sparse low-rank mechanism and optimize feature expression of positive samples at the training stage.

Similar to [48], a quality measure score S is defined to appraise the learned representative points of convex-hull representation, which is formulated according to three aspects, including ability of localization, classification, and adaptability to sparse low-rank mechanism. Thus, S can be denoted as follows:

$$S = \lambda_1 S_{\text{loc}} + \lambda_2 S_{\text{cls}} + \lambda_3 S_{\text{SR}} \tag{11}$$

where S_{loc}, S_{cls}, and S_{SR} indicate localization score, classification score, and sparse low-rank score, respectively. λ_1, λ_2, and λ_3 are the weighting factors for these scores, respectively.

As for the localization score S_{loc} and classification score S_{cls} of a convex-hull point set, we utilize the localization loss $\mathcal{L}_{\text{loc}}$ and classification loss $\mathcal{L}_{\text{cls}}$ corresponding to its classification confidence as evaluation of localization and classification ability, which does not incur extra computational load. Therefore, S_{loc} and S_{cls} can be defined as follows:

$$S_{\text{loc}}(R_i, g_j) = \mathcal{L}_{\text{loc}}(R_i^{\text{loc}}, g_j^{\text{loc}}) \tag{12}$$

$$S_{\text{cls}}(R_i, g_j) = \mathcal{L}_{\text{cls}}(R_i^{\text{cls}}, g_j^{\text{cls}}) \tag{13}$$

where g_j^{loc} and g_j^{cls} are points position and class label with the ground-truth g_j, R_i^{cls} and R_i^{loc} indicate the classification confidence and localization of convex-hull point set R_i, respectively. We expect compatibility of feature and position of the points assessed by the scores. Although the above scores measure spatial location distance and feature similarity between convex-hulls and labels, the compatibility of distribution with sparse low-rank mechanism is not evaluated as mentioned earlier.

SAR ship detection task can be regarded as a process pf sparse signal recovery, in which ships present sparse characteristic with its pixels sparsely distributed. Meanwhile, the feature vectors of sea clutter maintain relatively high correlation with each other and the statistical distribution of background clutter can be decomposed into low-dimensional subspaces composed of a few effective bases, which is called low-rank mechanism. The feature vectors of background can be represented as the linear combination of effective bases in background clutter dictionary (BCD). In contrast, the corresponding feature vectors of ship sample points can not be represented as it, which means low similarity between the feature vectors of ship and background. The representation of sparse and low-rank characteristic is shown in Fig. 6.

Based on the above concept, the sparse low-rank score S_{SR} is proposed to measure compatibility of sample points distribution with sparse low-rank mechanism. Specifically, metrics of cosine similarity is introduced to calculate the mechanism between the feature vectors of representative points. Considering $v_{i,j}$ as the corresponding feature vector with the *j*th point of the *i*th set of convex-hull points, proposed S_{SR} of the *i*th convex-hull can be denoted as follows:

$$S_{\text{SR}} = 1 - \frac{1}{9} \sum_j \frac{v_{i,j}^* \cdot v_i^*}{\left\| v_{i,j}^* \right\| \times \left\| v_i^* \right\|} \tag{14}$$

where $v_{i,j}^*$ denotes normalized embedding feature vector for the *j*th point of the *i*th set of convex-hull points and v_i^* represents the mean of normalized embedding feature vectors from the *i*th convex-hull set.

As for the reassignment scheme, we model the proposed quality score of convex-hulls corresponding to certain annotation as a likelihood maximization for a probability distribution inspired by [63], as a result of which the probability of label assignment

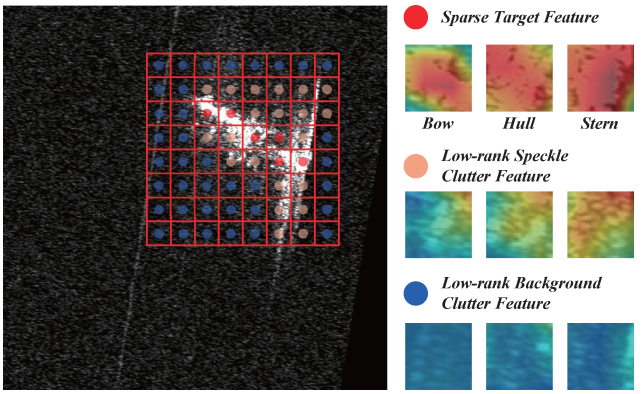


Fig. 6. Sketch of the representation for sparse and low-rank characteristic. Feature visualization of sparsely distributed target pixels from different parts, speckle clutter and background clutter dictionary are shown in the drawing in the right, respectively.

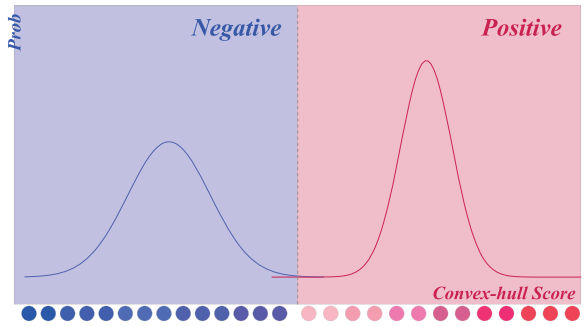


Fig. 7. Reassignment scheme of SLCR to separate samples using their probabilities. The decision criteria for assignment is abstracted to the division of two distributions from Gaussian mixture model.

can be calculated to reassign the convex-hull samples. Then, the Gaussian mixture model (GMM) of two modalities is introduced to fit the above samples with expectation-maximization (EM) algorithm applied to optimize. After EM estimation, the positive samples of reassignment can be determined with a heuristics separation boundary based on convex-hull probabilities, as shown in Fig. 7.

Noting that positive samples exhibit inadequate for modeling a GMM of two modalities in certain extreme situations of a small number of samples, dynamic k label assignment proposed by [48] is introduced to reassign the convex-hull samples to avoid training interruptions.

IV. Experiments

A. Dataset and Settings

Three ship datasets containing public SAR images and expert rotated annotations are introduced to construct the experiments: SAR ship detection dataset (SSDD+ [64]), the high-resolution SAR image dataset (HRSID [65]), and rotated ship detection dataset (RSDD [66]). Based on the following considerations, above datasets are chosen from existing public SAR ship datasets to perform the experiments. First, due to the fact that ship targets exhibit special characteristics with large aspect ratio, rotated annotation introduces less background noise and accurately characterizes the features of independent samples in locally dense scenes such as ports, etc., compared to vertical annotation. Second, the images in the above datasets cover complex scenarios with different sizes, which is appropriate for evaluating robustness and reflecting performance.

1) SSDD+: 1160 SAR images with 2540 rotated instances in various scenes compose the SSDD+, whose source varies in RadarSat-2, TerraSAR-X, and Sentinel-1. The images of SSDD+ possess a wide range of spatial resolutions, which vary from 1 to 15 m, and all polarization modes are included. SSDD+ is randomly reconfigured to a combination of training set and test set at 0.8 training rate, specifically 928 images in the former and the remaining 232 images in the latter.

2) HRSID: HRSID consists of 5604 images from Sentinel-1B, TerraSAR-X, and TanDEM-X with 16 951 expert annotation instances. The images with 800 $\times$ 800 pixels in HRSID covers various resolutions between 1 and 5 m corresponding to diverse scenes, which originate from 136 high-resolution SAR images. The official publisher of HRSID reconstruct it as a training part involving 1962 images coupled with a test part with 1962 images. We follow the division in the following experiments.

3) RSDD: RSDD includes 7000 SAR images originated from sources of GF-3 and TerraSAR-X satellites, which are cropped from 127 panoramic SAR images. It consists of 10 263 ship instances with multiple imaging modes, polarization modes, and resolutions, which are efficiently annotated through intelligent annotation with expert revision. Notably, the founders of RSDD assign 5000 images to training set and 2000 remaining images to test set. We follow the division in the following experiments.

B. Implementation Details

1) Data Preprocessing: For training, SSDD+ and RSDD images are resized to 512 $\times$ 512 in pixel dimension as the input of the network, while HRSID ones are processed to 800 $\times$ 800 pixels. Bilinear interpolation algorithm is applied as a resizing operation. Images are randomly filliped horizontal, vertical and diagonal with a 25% probability for data augmentation. Besides, SAR images with single channel in experimental datasets are expanded to three channels through self-replication.

2) Optimization: The stochastic gradient descent (SGD) algorithm [76] is introduced to instruct gradient descent in network training with its momentum set to 0.9 and a 0.0001 weight decay. Heuristic training parameters over a Nvidia TitanXP GPU are determined as 40 epochs with the learning rate initialized to 0.008 in which 0.1 step strategy adopted at 24, 32, and 38 nodes. The ResNet50 [77] backbone in detectors of experiments is pretrained on ImageNet dataset, and models in this article are trained based on the pretrained weight.

3) Inference: In the test stage, images are processed to the same size as trained ones for the input of the network. The NMS [71] is introduced to the inference results to remove duplicate bounding boxes.

C. Evaluation Metrics

We adopt four widely used and authoritative metric to quantitatively assess the detection performance, which are defined as follows.

For every predicted rotated bounding box R_{pre}, IoU is calculated with its corresponding annotation R_{gt}, which is calculated as follows:

$$\text{IoU} = \frac{\text{area}(R_{\text{pre}} \cap R_{gt})}{\text{area}(R_{\text{pre}} \cup R_{gt})}. \tag{15}$$

Additionally, precision and recall can be calculated as fundamental norms in estimate of the detection performance based on the concept of IoU, which can be formulated as follows:

$$P = \frac{N_{\text{tp}}}{N_{\text{tp}} + N_{\text{fp}}} \tag{16}$$

$$R = \frac{N_{\text{tp}}}{N_{\text{tp}} + N_{\text{fn}}} \tag{17}$$

where N_{tp} represents the number of true positives (TP). In SAR ship detection task, TP is defined as a detection result associated with certain annotation, where IoU exceeds the heuristic threshold. Instead, N_{fp} denotes the number of false positives (FP) and FP serves as a detection result with IoU below the same threshold. Otherwise, N_{fn} represents the number of false negatives (FN), which denotes the ground-truth whose IoU between itself and any detection results is below the threshold.

Besides, we leverage F1 to balance the reflection on detection performance of the above two metrics in order to evaluate models in a comprehensive manner, which is calculated as follows:

$$F1 = \frac{2 \times P \times R}{P + R}. \tag{18}$$

The precision-recall (PR) curve serves as a function fitted by pairs of precision and recall calculated in an equally spaced sequence of confidence threshold. Additionally, AP presents the comprehensive capacity of algorithms, which is calculated due to the area under the PR curve. Given the IoU of 0.5, according to [78], AP is formulated as

$$\text{AP} = \int_0^1 P(R)\, dR. \tag{19}$$

It is noticeable that mean average precision (mAP) serves equivalent to AP in this article, considering that ship detection task has single-class detection attributes.

D. Ablation Experiments

Multiple ablation experiments are implemented to analyze and evaluate the influence of each proposed component. In this section, the baseline network is coupled with the QSLM and SLCR for comparison under the same experiment settings, respectively. Taking oriented RepPoints as the base framework, mAP metric is selected to evaluate the module ablation in Table III.

1) Effect of QSLM: As shown in Table III, backbone with QSLM module gains 0.4% on mAP with SSDD+, 2.0% with HRSID and 1.0% with RSDD, which validates that the proposed component achieves comparable performance. Ship targets and backgrounds both exhibit multiplicative non-Gaussian coherent speckle with extremely low signal-to-noise ratio, which represents nonlinear characteristics under the unique imaging mechanism of SAR. Therefore, this module explicitly describes the nonlinear statistical features introduced by the speckle effect to enhance the learning ability of linearly indivisible features in SAR images. In conclusion, the proposed QSLM enhanced the feature extraction ability through learning of statistical features, and detectors with it prove gains on performance of mAP metric. Additionally, the proposed QSLM is a learnable primitives of feature extraction without extra supervision introduced to network.

To further analyze the impact of QSLM on feature extraction, experiments of QSLM located in various parts of the detection framework on HRSID are devised. As shown in Table IV, base framework of convex-hull representation without tricks is introduced as B0 with backbone of Resnet50 to eliminate interference influence. Among M01 to M05, B0 with QSLM located in different part are constructed. We choose three stages in backbone and two locations in detection head as comparative experiments. One can see that the network with QSLM located in stage4 achieves the highest performance. We assume that the statistical features extracted by shallow layers such as stage1 is diluted by subsequent linear feature extraction causing the decline detection performance, and the feature maps extracted by deep layers such as stage5 lose significant statistical information. Thus, Resnet50 with QSLM located in stage4 is used in our detectors.

In addition, a convolution control trial is devised to eliminate the impact of deepening the network, which is shown in Table V. B1 represents the baseline Oriented RepPoints, while M12 indicates the baseline with QSLM combined in stage4 of its backbone, and C12 represents B1 with a pair of convolution module combined to the same position of its backbone. One can see that the parallel convolution module added to the same location as QSLM has a slight impact on feature extraction, which demonstrate the effectiveness of proposed module.

Feature heatmaps of various scenes in Fig. 8 are introduced to intuitively visualize the effect of statistical feature learning in feature extraction, which are generated by the stage 4 of backbone networks. As shown in Fig. 8(b), the feature extracted by traditional convolutional backbone cannot represent semantics of ships precisely under the shape variety caused by coherent speckle. In contrast, Fig. 8(c) presents more integrated heatmaps of same ships, which reflects more precise feature attention. Meanwhile, as displayed in Fig. 8(b) and (c), due to the absence of statistical learning, features extracted by baseline model suffers low discrimination between targets and background, as a result of which false alarms appear in inshore scenes. Therefore, the supplement to linear features and promotion to feature extraction of statistical feature learning are demonstrated.

Besides, as displayed in Fig. 8, the proposed QSLM in the first row presents relatively less effective on suppression of background features in comparison to those in the subsequent rows (second, third, and fourth), which reflects potential clarity constraints in the efficacy of the module. To address this

TABLE II
COMPARISON OF DIFFERENT ORIENTED CNN-BASED METHODS ON THREE DATASETS

Methods	SSDD+				HRSID				RSDD			
	mAP	Precision	Recall	F1	mAP	Precision	Recall	F1	mAP	Precision	Recall	F1
Oriented Object Detection Methods												
RetinaNet-OBB [67]	0.874	*0.934*	0.707	0.805	0.773	*0.896*	0.682	0.775	0.86	*0.945*	0.702	0.805
CSL [68]	0.786	0.914	0.707	0.797	0.762	0.889	0.679	0.77	0.791	0.94	0.696	0.8
S^2A-Net [69]	0.881	0.881	0.888	0.884	0.816	0.895	0.772	*0.829*	0.89	0.938	0.859	0.897
ReDet [45]	*0.898*	0.886	*0.901*	0.894	*0.832*	0.803	**0.849**	0.825	**0.9**	0.882	**0.923**	0.902
R^3Det [44]	0.844	0.883	0.812	0.846	0.758	0.882	0.696	0.778	0.804	0.905	0.806	0.853
FCOS-OBB [70]	0.604	**0.968**	0.247	0.393	0.619	**0.971**	0.465	0.629	0.707	0.961	0.51	0.667
RepPoints-OBB [60]	0.789	0.885	0.643	0.745	0.692	0.855	0.544	0.664	0.775	0.885	0.598	0.714
Faster RCNN-OBB [71]	0.855	0.884	0.872	0.878	0.795	0.776	0.803	0.79	0.884	0.861	0.887	0.874
Oriented R-CNN [46]	0.884	0.904	0.888	*0.896*	0.823	0.825	0.826	0.826	*0.899*	0.887	*0.92*	0.903
RoI-Trans [43]	0.878	0.884	0.886	0.885	*0.832*	0.823	*0.831*	0.827	0.898	0.902	0.904	*0.903*
Gliding Vertex [72]	0.838	0.879	0.856	0.867	0.795	0.803	0.805	0.804	0.892	0.879	0.898	0.888
Oriented SAR Ship Detection Methods												
O-BBAVs [73]	0.777	0.856	0.663	0.747	0.725	0.882	0.636	0.739	0.805	0.948	0.785	0.859
DCMSNN [74]	0.853	0.881	0.871	0.875	0.796	0.779	0.803	0.791	0.883	0.861	0.889	0.875
DrBoxv2 [75]	0.301	0.548	0.471	0.506	0.151	0.362	0.289	0.322	0.186	0.479	0.366	0.415
Convex-hull Object Detection Methods												
CFA [47]	0.806	0.895	0.613	0.728	0.814	0.841	0.791	0.816	0.887	0.927	0.878	0.902
Oriented Rep [48]	0.888	0.894	0.878	0.886	0.812	0.853	0.767	0.807	0.886	0.876	0.917	0.896
Proposed Method	**0.914**	0.904	**0.902**	**0.903**	**0.84**	0.866	0.796	**0.83**	*0.899*	**0.946**	0.882	**0.913**

The optimum and suboptimal outcomes for four indexes on three datasets are marked in bold and italics, respectively.

TABLE III
ABLATION STUDY OF THE PROPOSED METHOD (MAP)

QSLM	SLCR	SSDD+	HRSID	RSDD
×	×	0.888	0.812	0.886
✓	×	0.892	0.832	0.896
×	✓	0.896	0.819	0.89
✓	✓	**0.914**	**0.84**	**0.899**

The optimum outcomes are marked in bold.

TABLE IV
INFLUENCE OF DIFFERENT QSLM POSITIONS

Methods	Backbone			Detector		mAP
	stage_1	stage_4	stage_5	Input	DCN	
B0						0.704
M01	✓					0.688
M02		✓				**0.736**
M03			✓			0.692
M04				✓		0.71
M05					✓	0.693

The optimum outcomes are marked in bold.

phenomenon, RaRes (radiometric resolution) is introduced to evaluate the quality of background clarity to describe optimal range of applicability for the module, which is calculated as follows:

$$\mathrm{RaRes} = 10 \lg \left(\frac{\sigma}{\mu} + 1 \right) \tag{20}$$

where σ and μ denotes the standard deviation and mean of SAR image, respectively. RaRes serves as an indicator for measuring the grayscale clarity of SAR images, which distinguishes the backscattering coefficient of targets by describing the radiation quality of each pixel.

After analyzing above metric, we find out that features of foreground ships can be reinforced explicitly and background features be suppressed under the conditions where RaRes > 2.5. On the contrary, although features of foreground ships can be refined by statistical features introduced by QSLM, the suppression of background features suffers relatively decline under the conditions where RaRes ≤ 2.5. We assume that the discriminative features of background are difficult to fit in scenes with extremely low quality of clarity, as the statistical distributions of background is disturbed by the imaging quality.

2) Effect of SLCR: As displayed in Table III, the introduction of SLCR module in training phase raises promotion to mAP metric with 0.8 % on SSDD+, 0.7% on HRSID, and 0.4% on RSDD, respectively, which indicates the necessity of reassignment scheme. As mentioned in Section III-C, ship targets

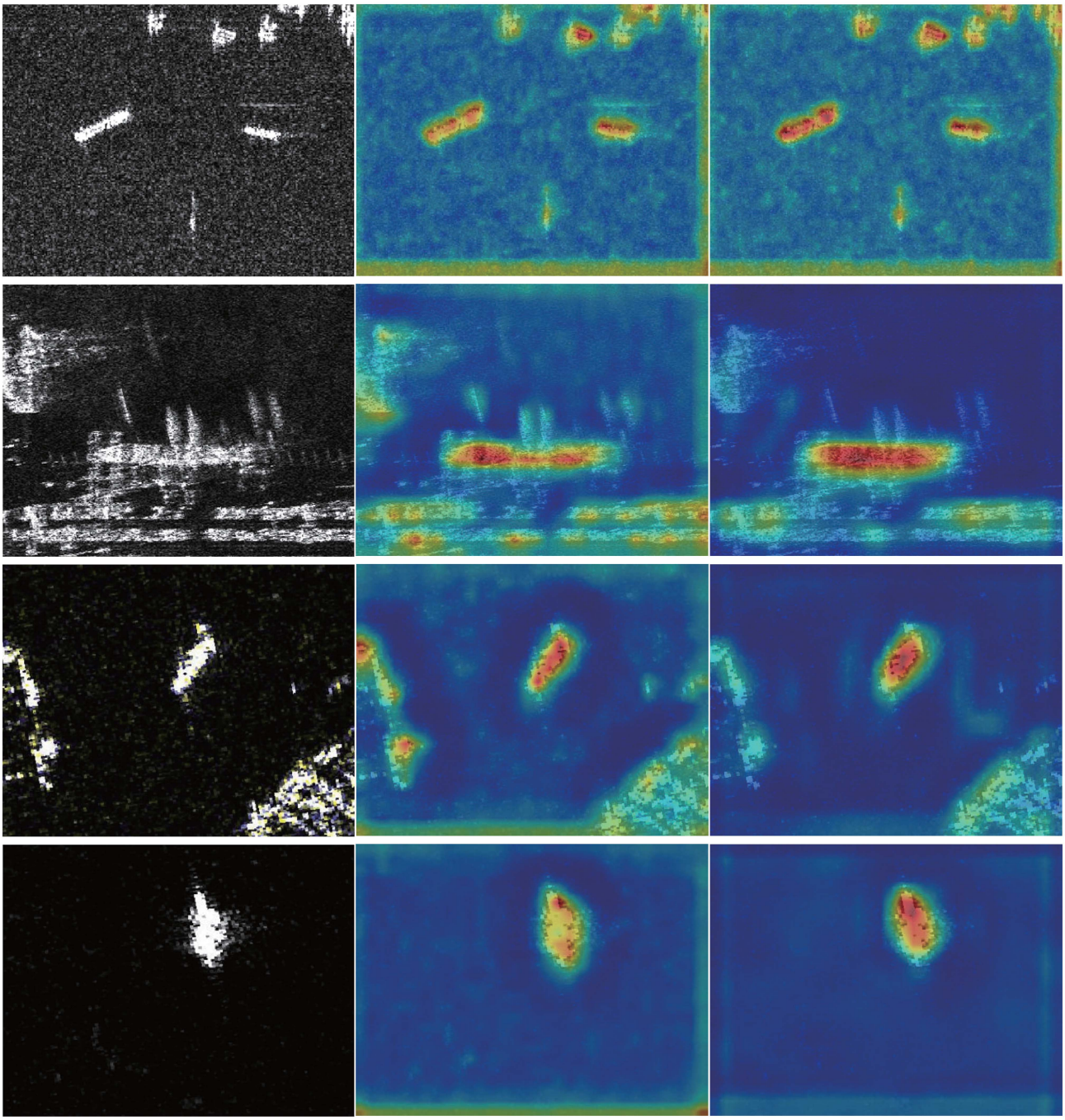


Fig. 8. Visualization of extracted ship features. (a) Original SAR images. (b) Heatmaps of features extracted by baseline. (c) Heatmaps of features extracted by baseline with QSLM.

TABLE V
PERFORMANCE OF CONVOLUTION CONTROL EXPERIMENT

Methods	mAP		
	SSDD+	HRSID	RSDD
B1	0.888	0.812	0.886
C12	0.868	0.829	0.888
M12	**0.892**	**0.832**	**0.896**

The optimum outcomes are marked in bold.

composed with speckle noise represents strong scattering points with irregular contours, indicating the presence of more positive sample points with low quality compared to optical images. As the sparse low-rank mechanism can model the distribution of strong scattering points in ships, this model can evaluate the quality of positive sample points learned by the network precisely and suppress sample points with low quality, which may result in false alarms and vague localization.

TABLE VI
INFLUENCE OF EACH QUALITY SCORE IN SLCR (MAP)

Q_{cls}&Q_{loc}	Q_{SR}	SSDD+	HRSID	RSDD
×	×	0.888	0.812	0.886
✓	×	0.879	0.809	0.879
×	✓	0.502	0.333	0.46
✓	✓	**0.896**	**0.819**	**0.89**

The optimum outcomes are marked in bold.

TABLE VII
EFFECT OF DIFFERENT WEIGHTING FACTORS OF Q_{SR}

λ_3	Precision	Recall	F1	mAP
0.1	**0.904**	**0.902**	**0.903**	**0.914**
0.2	0.88	0.876	0.878	0.880
0.3	0.899	0.882	0.890	0.900
0.4	0.874	0.871	0.872	0.879
0.5	0.877	0.882	0.879	0.886

The optimum outcomes are marked in bold.

Each component of the quality measure is introduced to the reassignment scheme term-by-term in order to analyze performance. Base framework of convex-hull representation without reassignment module is introduced as baseline. Table VI presents that reassignment scheme with modeling quality score of classification and localization suffers a decline on detection performance, which proves the fact that traditional estimation parameters cannot represents the statistical distribution of SAR ships completely. The sparse low-rank mechanism is proved to be effective in evaluation of samples for SAR ships due to the fact that the comprehensive quality score reaches top in detection performance. Yet notably, the sparse low-rank scoring is a supplement to basic sample assignment schemes with classification and localization in detection framework on SAR images, which can explain the phenomenon of huge decline on AP with only sparse low-rank score introduced to model the positive samples.

The weighting factors λ_3 of sparse low-rank score Q_{SR} in the quality measurement component also played key roles. We select five different weighting factors for comparative study on SSDD+. The weighting factor 0.1 reaches top in all metrics in Table VII. We assume that the quality of localization and classification mainly dominate the processing of network learning, as the features extracted by the network mainly consist of first-order moment features. Thus, an appropriate weighting factor can be set to balance the influence among scores of localization, classification, and sparse low-rank mechanism to optimize the scheme.

E. Performance Analysis of Oriented Object Detectors

Typical oriented object detectors from related domains are introduced to construct the comparison experiments, which can be detailed as: 1) recently published methods with oriented object detectors for optical remote sensing images, including oriented-based RetinaNet [67] (RetinaNet-OBB), CSL [68], S2ANet [69], ReDet [45], R3Det [44], oriented-based FCOS [70] (FCOS-OBB), Rotated-based RepPoints [60] (RepPoints-OBB), oriented-based Faster RCNN [71] (Faster RCNN-OBB), Oriented R-CNN [46], ROI Transformer [43] and Gliding Vertex [72]; 2) recently published methods with oriented object detectors for SAR images, including oriented Gaussian function-based BBAVs [73] (O-BBAVs), DCMSNN [74], DRBOX-v2 [75]; 3) oriented detectors with convex-hull representations, including CFA [47] and Oriented RepPoints [48]. All methods of comparison experiments are implemented on PyTorch framework, except for DRBOX-v2, which is coded with TensorFlow. Besides, ResNet-50 is introduced as the backbone of them.

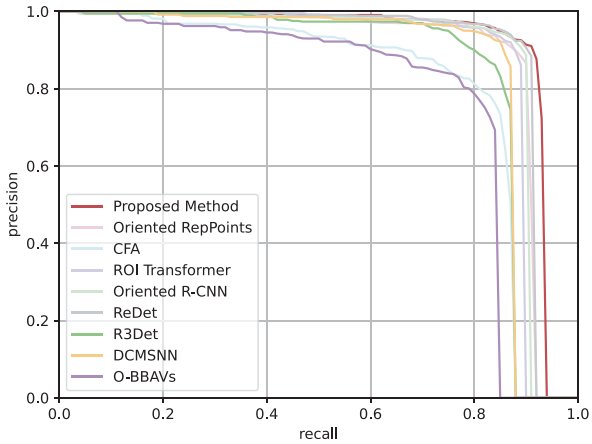


Fig. 9. PR curves of typical methods on SSDD+.

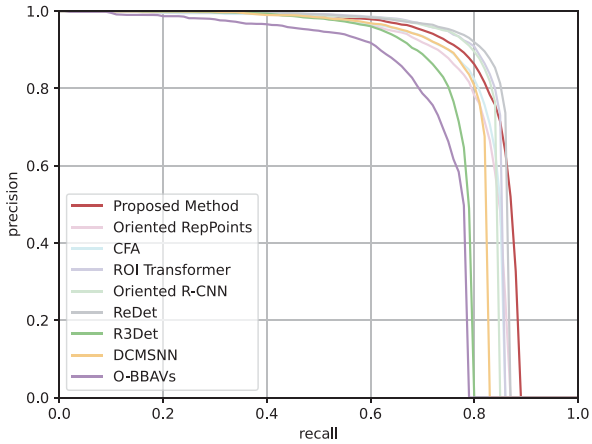


Fig. 10. PR curves of typical methods on HRSID.

1) Experiments on SSDD+: Table II displays the quantitative comparison of the above typical detectors and ours on SSDD+, and the detailed measure curves for the above methods are introduced in Fig. 9 to provide a comprehensive comparison. The optimum and suboptimal outcomes for four indexes on three datasets are marked in bold and italics, respectively. Based on the results of comparison, our approach achieves optimum results of both mAP and F1 metric among the detectors, in which the effectiveness of the network with statistical features and mechanism of SAR combined is verified. Different from the above methods

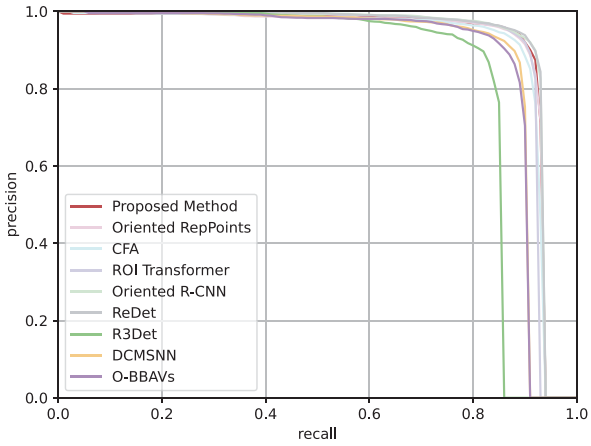


Fig. 11. PR curves of typical methods on RSDD.

for optical scenes, the detectors devised for SAR ones, such as O-BBAVs, DCMSNN, and DRBOX-v2, perform unsatisfactory. The major cause serves the lack of explicit learning for statistical features and sample representation of traditional anchors, which is not fit for the characteristics of strong scattering points with irregular contours. Meanwhile, the proposed detector outperforms the detectors with convex-hull representations designed for optical scenes due to the introduction of sparse and low-rank mechanisms, which affect sample quality.

2) Experiments on HRSID: Table II displays the quantitative comparison of the above typical object detectors and ours on HRSID, and the detailed measure curves for the above methods are introduced in Fig. 10. Proposed detector achieves the best performance on both mAP and F1 metrics, as reported in Table II, while ReDet and Oriented R-CNN achieve the suboptimal mAP of 83.2%. Notably, the powerful oriented optical-devised detectors reach effective performance relatively. We assume that the spatial resolution of images under HRSID varies within a narrow range of 0.5–3 m, the imaging quality of which is much higher than one under SSDD+. As a result, the contour of the ship tends to be fixed and regular, while the characterization of the scattering cluster corresponds in a way to the continuous texture features extracted by ships on optical images.

3) Experiments on RSDD: The proposed method reaches suboptimal performance among detectors listed and outperforms detectors devised for SAR ships and ones with convex-hull samples in Table II, while ReDet achieves the highest mAP of 90%. Fig. 11 presents the detailed comparison of performance by PR curves. The F1 index obtained in the proposed prediction achieves the highest by exceeding RoI-Trans, which represents the frontier for oriented-based detectors. Thus, the proposed methods can effectively reach equilibrium in various indictors, which represents the comprehensive performance of our work on extracting features of ships in complicated scenes, as the RSDD possesses a wide variety of actual scenes. It is noticeable that the introduction of SAR characteristic learning still raises a 1.3% promotion on mAP compared to baseline and reaches a relatively acceptable performance within 0.1% mAP behind the SOTAs of oriented object detectors.

4) Analysis of PRC: Extra PR curves are introduced to further analyze the comparison results in Fig. 12 following [79] and [80], which include AP_{50} curve, AP_{75} curve, localization error curve, and false alarm curve. PR curves with an IoU thresh of 0.5 and 0.75 are presented as C50 and C75, and the areas under them denote AP_{50} and AP_{75} metrics, respectively. Similarly, a PR curve with IoU thresh of 0.1 is present as Loc. Loc curve is introduced to visually display the impact of localization error, which is defined as misaligned detection boxes with a substandard overlap (IoU $\in [0.1, 0.5)$) to the ground truth. BG represents the PR curve with FP removed under IoU thresh of 0.1, and FN curve fits the performance without FP and FN.

As a result, the areas of several colors in Fig. 12 explains different physical signification. The white area is equivalent to AP_{75} metric as mentioned above. The gray area represents the AP metric on the IoU thresh of $(0.5, 0.75)$. The blue, purple, and orange areas reflect the decline of detection performance due to the inferior-quality predictions with fuzzy localization, false alarms in background, and missing ships, respectively.

The analysis exhibited in Fig. 12 reflects that the proposed method also achieves optimal performance of resistance to false alarms and missing detections. As for the effect of Loc, BG, and FN, corresponding to the blue area of Fig. 12(a), the mAP is increased by 0.07, which reflects the certain effect of localization vagueness for predictions. Then, as displayed in the purple area, the proposed detector suffers slightly from false alarms, as mAP has a small increase of 0.006 after removing them. The missing ships also have a slight effect on the detector with a mAP decrease of 0.01, as presented by the orange area.

5) Visualization Analysis: Visualization of performance in two typical scenes of ship detection application from SSDD+ test data is introduced to present qualitative predictions intuitively in Fig. 13. The detection performance on dense distribution of open sea region is displayed in Fig. 13(a). ReDet, R^3Det, and Oriented RepPoints suffer a relatively high false alarm rate, while CFA and O-BBAVs miss some densely distributed ships. Compared with the above methods, ROI Transformer, Oriented R-CNN, and DCMSNN can keep a balance between high recall and relatively few false alarms, while some low localization accuracy boxes still exist. Obviously, our method gets the highest accuracy of ship detection with the lowest false alarm rate in the scene. As for complex scenes, Fig. 13(b) is introduced to reflect anti-interference capability of detectors in the inshore region. Most methods, specially R^3Det, suffer relatively poor capability to distinguish ships from inshore facilities with similar backscatter intensity, while O-BBAVs and CFA miss some inshore ships. Due to the higher order statistics features and sparse low-rank mechanism introduced by QSLM and SLCR, respectively, the proposed method extracts more discriminative features to perform adaptable to complex scattering environments in contrast.

In order to qualitatively present the reflection of predictions on HRSID and further underline the superiority of the proposed detector, we chose two images with more complex ship detection application scenes from these data, which include the large-scene inshore region and the inshore region with dense distribution. The correctly detected ships and detection errors are displayed in Fig. 14 in the same way as Fig. 13. One can see that

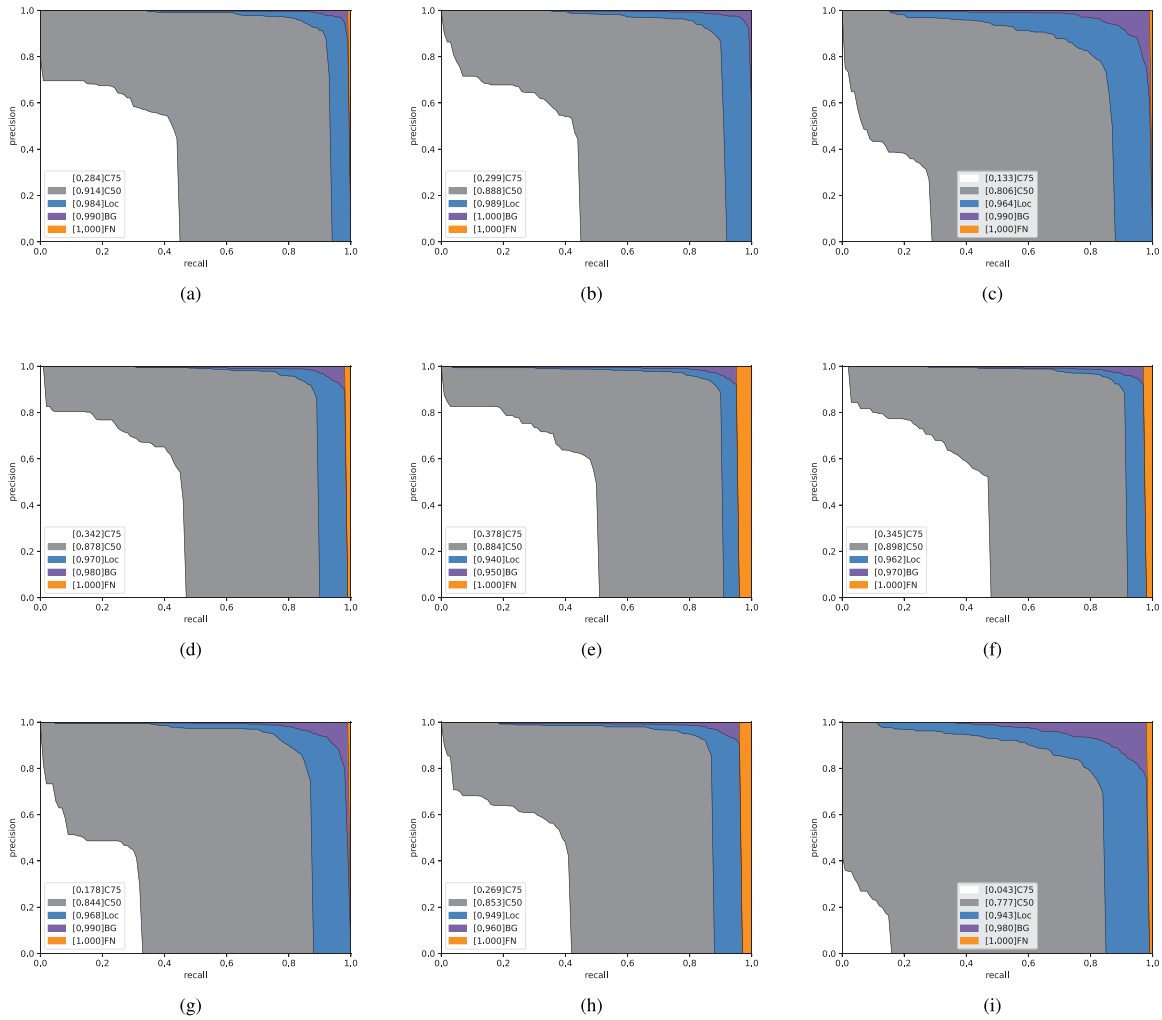


Fig. 12. Detailed PRC for typical methods on SSDD+. Detailed metrics are represented as legends in various colors calculated by integral between adjacent curves. (a) Proposed method. (b) Oriented RepPoints. (c) CFA. (d) ROI Transformer. (e) Oriented R-CNN. (f) ReDet. (g) R^3Det. (h) DCMSNN. (i) O-BBAVs.

ROI Transformer, Oriented R-CNN, and DCMSNN generate relatively enormous false alarm object in Fig. 14(a). R^3Det, O-BBAVs, CFA, and Oriented RepPoints keep higher false alarm rate, and they have the lowest detection rate. Compared with the above methods, ReDet performs relatively satisfactorily with few detection errors, while the proposed method presents preferable detection performance in large-scale inshore scenes. As for the challenging inshore scene with dense distribution in Fig. 14(b), our algorithm surpasses others on detection accuracy, which reflects its high robustness. It is noticeable that missing tiny targets and false alarms on land still appears, though the proposed network has a satisfactory detection performance. We assume that tiny targets with extremely limited pixels lack discriminative features, which confuse feature extractor even with the support of statistical feature learning. Further improvement to optimize the predictions for our detectors is required in the future.

6) Computational Complexity Analysis: To comprehensively evaluate performance across multiple dimensions, experiments related to computational complexity are conducted on SSDD+ in this section as supplements to accuracy metrics. The experimental results of typical methods are detailed in

TABLE VIII
COMPARISION OF TYPICAL METHODS ON COMPUTATIONAL COMPLEXITY

Methods	Memory (GB)	Inference Efficiency(FPS)	FLOPs (GFLOPs)
RoI-Trans	2.419	27.8	77.15
Oriented R-CNN	2.111	30.5	63.28
ReDet	2.451	25.1	40.88
R^3Det	3.729	30.4	82.17
DCMSNN	2.307	35.0	63.25
O-BBAVs	1.979	25.3	**27.88**
RepPoints-OBB	**1.797**	**36.6**	48.56
CFA	**1.797**	35.1	48.56
Oriented Reppoints	**1.797**	30.8	48.56
Proposed Method	3.175	28.8	58.23

The optimum outcomes are marked in bold.

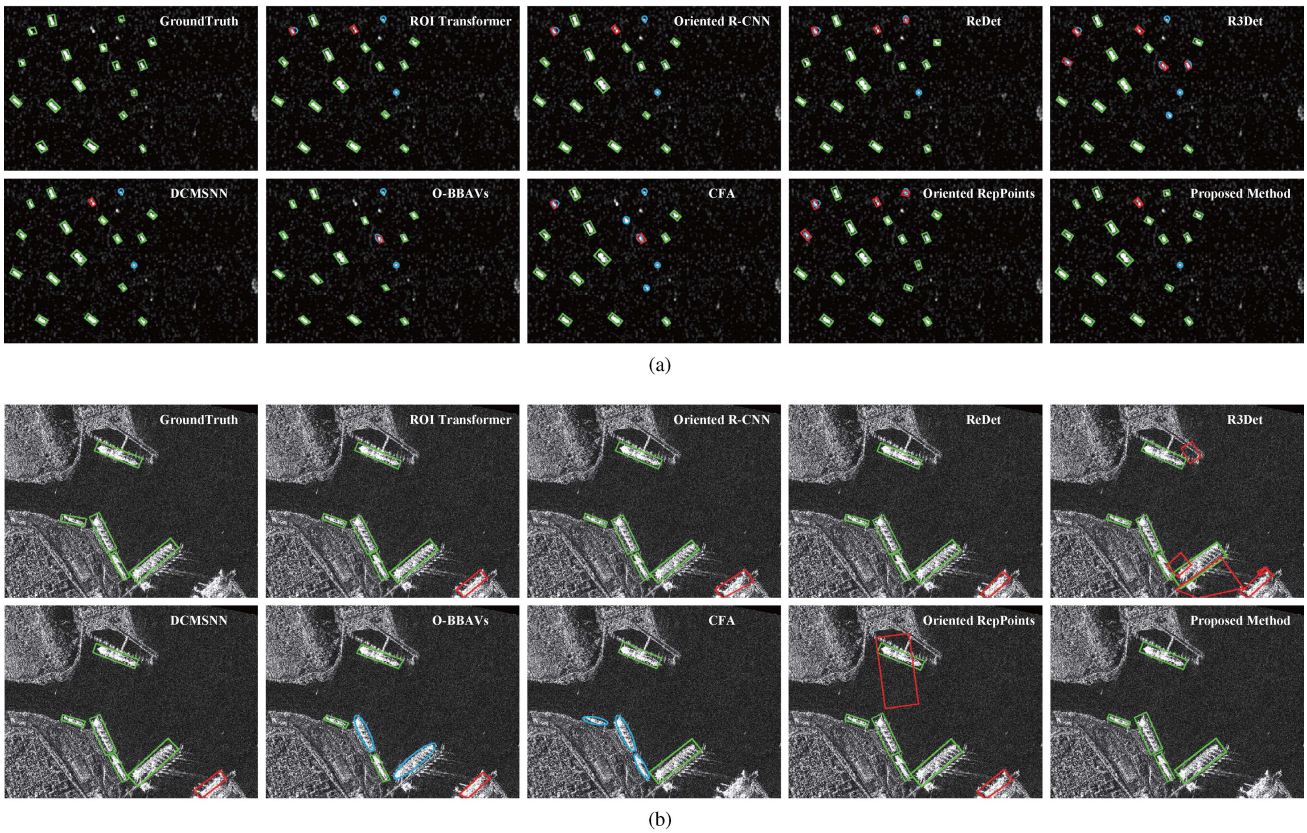


Fig. 13. Visualization for detection performance of typical methods in different scenes of SSDD+. The green boxes, red boxes, and blue circles indicate the correctly detected ships, false alarms, and the missing ships, respectively. (a) Results on dense distribution. (b) Results on inshore region.

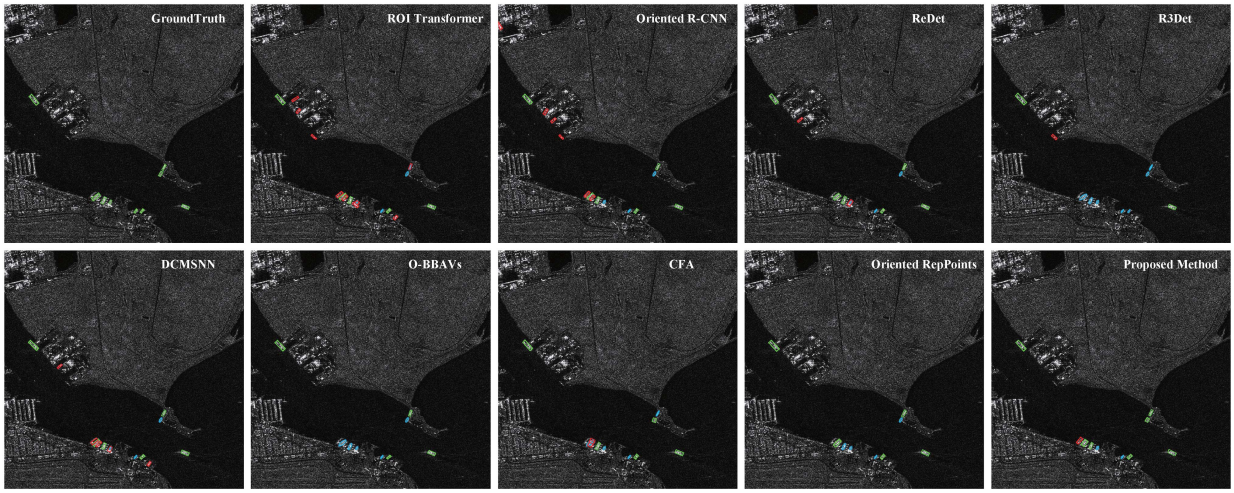
Table VIII. Metric memory is defined as GPU memory consumption during inference, which is introduced to evaluate the memory consumption of methods. Inference efficiency represents the number of images inferenced within a second, which is introduced to reflect computational efficiency. Floating point operations (FLOPs) is defined as the total number of floating-point operations in the model, which reflects the model complexity.

As is shown in Table VIII, O-BBAVs possesses the lowest model complexity owing to its single-stage anchor-free framework, while the calculations of gaussian heatmaps burden the memory consumption and inference efficiency. The two-stage anchor-free methods, such as Rotated RepPoints, CFA, and so on, carry the same memory consumption and model complexity, as they share the same model structure including backbone, neck, and two-stage heads. The inference efficiency reflects the differences in representation learning and sample judgment of them. Although the proposed method achieves progress on accuracy, the applications of statistical feature learning and mechanism supervision suffer a relatively acceptable decline in inference efficiency compared to other methods based on convex-hull representation. Meanwhile, extra memory consumption is introduced by the statistical learning module as the statistical feature maps are conducted and calculated on the GPU. Yet it is worth noting that the proposed method still possesses higher inference efficiency and lower model complexity compared to the state-of-arts with two-stage anchor-based framework, which reflects certain practicality and scalability. Moreover, we will continuously optimize limitations such as memory consumption and strive to strike a balance between efficiency and performance to expand model scalability and practicality in the future work.

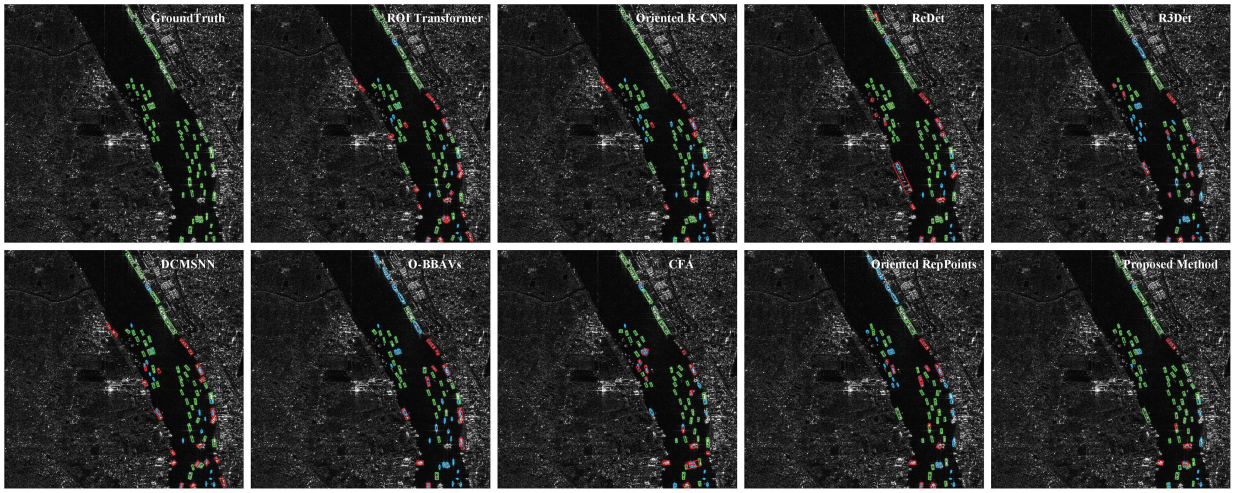
F. Discussion

Although our method presents relatively progressive performance on three datasets of SAR ship detection, potential constraints still exist in extreme application scenes. First, slight missing tiny targets exit though the proposed network has a satisfactory detection performance on various scenes. We assume that the extreme lack of discriminative features leads to the confusing feature extraction even with the support of statistical learning. Second, the quality of SAR image clarity partly affects the accuracy of background feature fitting. The main reason may lie in the fact that the statistical distributions of background are severely disturbed by the imaging quality. Third, the computational complexity, especially memory consumption, increases in the proposed method, which approaches those of two-stage anchor-based methods. We assume that the calculation of statistical feature map with spatial-scale matrix multiplication operation on GPU mainly causes the increment in memory consumption. This partly limits the model's practicality and scalability.

In future research, we will focus on collecting and analyzing missing targets and false alarms in the above extreme scenes



(a)



(b)

Fig. 14. Visualization for detection performance of typical methods in different scenes of HRSID. (a) Results on large-scene inshore region. (b) Results on inshore region with dense distribution.

to explore an optimized manner of coupling the SAR characteristics and mechanism with CNN. Besides, we will continue to develop optimization strategies of computational complexity to pursue a balance between detection accuracy and efficiency, which will expand practicality and scalability of our method.

V. Conclusion

In this article, a hierarchical framework of SAR ship detector coupled with sparse low-rank mechanism and statistical characteristics is proposed. First, the proposed method extracts statistical features from SAR images with a point statistical learning module in the bottom level. Second, CGM is designed to effectively represent the characteristics of SAR ship to optimize the location reflection in the middle level. Third, the SLCR module is designed to improve the quality of convex-hull representations with sparse and low-rank mechanism introduced to rank the positive samples in the top level. Finally, comprehensive experiments with qualitative and quantitative evaluations on the SSDD+, HRSID, and RSDD confirm the performance superiority in various scenes. For the future work, the mechanism and characteristics of SAR will be studied continuously on architecture that effectively couples with deep networks. Besides, we will focus on complicated application scenes such as tiny ships with dense distribution to further improve our ship detector.

References

[1] Z. Wang, L. Du, J. Mao, B. Liu, and D. Yang, "SAR target detection based on SSD with data augmentation and transfer learning," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 1, pp. 150–154, Jan. 2019.

[2] L. Du, L. Li, D. Wei, and J. Mao, "Saliency-guided single shot multibox detector for target detection in SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 58, no. 5, pp. 3366–3376, May 2020.

[3] J. Zhang et al., "Water body detection in high-resolution SAR images with cascaded fully-convolutional network and variable focal loss," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 1, pp. 316–332, Jan. 2021.

[4] Y. Zhang et al., "SAR object detection encounters deformed complex scenes and aliased scattered power distribution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4482–4495, 2022.

[5] T. Zhao et al., "Artificial intelligence for geoscience: Progress, challenges and perspectives," *Innovation*, vol. 5, no. 9, 2024, Art. no. 100691.

[6] J. Cui, H. Jia, H. Wang, and F. Xu, "A fast threshold neural network for ship detection in large-scene SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6016–6032, 2022.

[7] L. Huang et al., "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote. Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.

[8] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1042–1056, 2023.

[9] C. Zhao, X. Fu, J. Dong, R. Qin, J. Chang, and P. Lang, "SAR ship detection based on end-to-end morphological feature pyramid network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4599–4611, 2022.

[10] C. Deledalle, L. Denis, S. Tabti, and F. Tupin, "Mulog, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389–4403, Sep. 2017.

[11] C. He, B. He, X. Liu, C. Kang, and M. Liao, "Statistics learning network based on the quadratic form for SAR image classification," *Remote. Sens.*, vol. 11, no. 3, 2019, Art. no. 282.

[12] C. He, Y. Ye, L. Tian, G. Yang, and D. Chen, "A statistical distribution texton feature for synthetic aperture radar image classification," *Front. Inf. Technol. Electron. Eng.*, vol. 18, no. 10, pp. 1614–1623, 2017.

[13] Y. Kang, Z. Wang, J. Fu, X. Sun, and K. Fu, "SFR-Net: Scattering feature relation network for aircraft detection in complex SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 60, 2022, Art. no. 5218317.

[14] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.

[15] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.

[16] Z. Cui, H. Quan, Z. Cao, S. Xu, C. Ding, and J. Wu, "SAR target CFAR detection via GPU parallel operation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote. Sens.*, vol. 11, no. 12, pp. 4884–4894, Dec. 2018.

[17] P. Iervolino et al., "Ship detection in SAR imagery: A comparison study," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Fort Worth, TX, USA, Jul. 23–28, 2017, pp. 2050–2053.

[18] S. Liu, Z. Cao, and H. Yang, "Information theory-based target detection for high-resolution SAR image," *IEEE Geosci. Remote. Sens. Lett.*, vol. 13, no. 3, pp. 404–408, Mar. 2016.

[19] W. Yu, Y. Wang, H. Liu, and J. He, "Superpixel-based CFAR target detection for high-resolution SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 13, no. 5, pp. 730–734, May 2016.

[20] T. Li, Z. Liu, L. Ran, and R. Xie, "Target detection by exploiting superpixel-level statistical dissimilarity for SAR imagery," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 4, pp. 562–566, Apr. 2018.

[21] O. A. Pappas, A. Achim, and D. R. Bull, "Superpixel-level CFAR detectors for ship detection in SAR imagery," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 9, pp. 1397–1401, Sep. 2018.

[22] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote. Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2018.

[23] X. Wang, C. Chen, Z. Pan, and Z. Pan, "Superpixel-based LCM detector for faint ships hidden in strong noise background SAR imagery," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 3, pp. 417–421, Mar. 2019.

[24] H. Lin, C. Hang, K. Jin, L. Zeng, and J. Yang, "Ship detection with superpixel-level fisher vector in high-resolution SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 17, no. 2, pp. 247–251, Feb. 2020.

[25] J. Zhao, Z. Zhang, W. Yu, and T. Truong, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708, 2018.

[26] Z. Deng, H. Sun, S. Zhou, and J. Zhao, "Learning deep ship detector in SAR images from scratch," *IEEE Trans. Geosci. Remote. Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019.

[27] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.

[28] E. Khesali, H. Enayati, M. Modiri, and M. Mohseni Aref, "Automatic ship detection in single-pol SAR images using texture features in artificial neural networks," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 40, pp. 395–399, 2015.

[29] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, 2017, pp. 1–4.

[30] Q. Guo, H. Wang, and F. Xu, "Scattering enhanced attention pyramid network for aircraft detection in SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 9, pp. 7570–7587, Sep. 2021.

[31] S. Song, B. Xu, Z. Li, and J. Yang, "Ship detection in SAR imagery via variational Bayesian inference," *IEEE Geosci. Remote. Sens. Lett.*, vol. 13, no. 3, pp. 319–323, Mar. 2016.

[32] Y. Huang, G. Liao, Z. Zhang, Y. Xiang, J. Li, and A. Nehorai, "SAR automatic target recognition using joint low-rank and sparse multiview denoising," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 10, pp. 1570–1574, Oct. 2018.

[33] X. Wang, G. Li, A. Plaza, and Y. He, "Ship detection in SAR images via enhanced nonnegative sparse locality-representation of fisher vectors," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 11, pp. 9424–9438, Nov. 2021.

[34] M. Li, W. Huo, Z. Li, J. Wu, and J. Yang, "SAR image reconstruction and target extraction with under-sampled data via low-rank and sparsity matrix decomposition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Brussels, Belgium, Jul. 11–16, 2021, pp. 4564–4567.

[35] Y. Zhao, L. Zhao, Z. Liu, D. Hu, G. Kuang, and L. Liu, "Attentional feature refinement and alignment network for aircraft detection in SAR imagery," *IEEE Trans. Geosci. Remote. Sens.*, vol. 60, 2022, Art. no. 5220616.

[36] Y. Sun, X. Sun, Z. Wang, and K. Fu, "Oriented ship detection based on strong scattering points network in large-scale SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 60, 2022, Art. no. 5218018.

[37] Y. Zhao, L. Zhao, C. Li, and G. Kuang, "Pyramid attention dilated network for aircraft detection in SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 18, no. 4, pp. 662–666, Apr. 2021.

[38] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.

[39] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.

[40] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, 2020.

[41] X. Liu, C. He, Q. Zhang, and M. Liao, "Statistical convolutional neural network for land-cover classification from SAR images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 17, no. 9, pp. 1548–1552, Sep. 2020.

[42] Z. Zhang, W. Shen, X. Yue, Y. Lin, and W. Hong, "Single channel CSAR moving target detection algorithm based on low-rank sparse decomposition," in *Proc. CIE Int. Conf. Radar*, 2021, pp. 929–932.

[43] J. Ding, N. Xue, Y. Long, G. Xia, and Q. Lu, "Learning RoI transformer for oriented object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 16–20, 2019, pp. 2849–2858.

[44] X. Yang, J. Yan, Z. Feng, and T. He, "R3Det: Refined single-stage detector with feature refinement for rotating object," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 4, pp. 3163–3171.

[45] J. Han, J. Ding, N. Xue, and G.-S. Xia, "ReDet: A rotation-equivariant detector for aerial object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2786–2795.

[46] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3520–3529.

[47] Z. Guo, C. Liu, X. Zhang, J. Jiao, X. Ji, and Q. Ye, "Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8792–8801.

[48] W. Li, Y. Chen, K. Hu, and J. Zhu, "Oriented reppoints for aerial object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1829–1838.

[49] H.-C. Li, W. Hong, Y.-R. Wu, and P.-Z. Fan, "On the empirical-statistical modeling of SAR images with generalized gamma distribution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 386–397, Jun. 2011.

[50] Y. Wang, T. L. Ainsworth, and J.-S. Lee, "On characterizing high-resolution SAR imagery using kernel-based mixture speckle models," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 968–972, May 2015.

[51] M. Migliaccio, L. Huang, and A. Buono, "SAR speckle dependence on ocean surface wind field," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5447–5455, Aug. 2019.

[52] T. Esch, A. Schenk, T. Ullmann, M. Thiel, A. Roth, and S. Dech, "Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1911–1925, Jun. 2011.
[53] A. Verlanti, F. Nunziata, M. Migliaccio, M. Melis, C. Bignami, and M. Musacchio, "Speckle-related parameters to monitor water area coverage of salty ponds," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, 2024, pp. 4815–4818.
[54] X. Deng, C. López-Martínez, J. Chen, and P. Han, "Statistical modeling of polarimetric SAR data: A survey and challenges," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 348.
[55] E. E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, Apr. 2004.
[56] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Rijeka, Croatia: SciTech, 2004.
[57] I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K distribution parameters for SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 5, pp. 989–999, Sep. 1993.
[58] A. C. Frery, H.-J. Muller, C. d. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.
[59] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.
[60] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "RepPoints: Point set representation for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9657–9666.
[61] R. A. Jarvis, "On the identification of the convex hull of a finite set of points in the plane," *Inf. Process. Lett.*, vol. 2, no. 1, pp. 18–21, 1973.
[62] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 658–666.
[63] K. Kim and H. S. Lee, "Probabilistic anchor assignment with IoU prediction for object detection," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K.: Springer, Aug. 23–28, 2020, pp. 355–371.
[64] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era: Models, Methods Appl.*, 2017, pp. 1–6.
[65] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[66] X. Congan et al., "RSDD-SAR: Rotated ship detection dataset in SAR images," *J. Radars*, vol. 11, no. 4, pp. 581–599, 2022.
[67] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[68] X. Yang and J. Yan, "Arbitrary-oriented object detection with circular smooth label," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K.: Springer, Aug. 23–28, 2020, pp. 677–694.
[69] J. Han, J. Ding, J. Li, and G.-S. Xia, "Align deep features for oriented object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5602511.
[70] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9627–9636.
[71] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[72] Y. Xu et al., "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.
[73] J. Zhang, M. Xing, G.-C. Sun, and N. Li, "Oriented Gaussian function-based box boundary-aware vectors for oriented ship detection in multiresolution SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5211015.
[74] J. Jiao et al., "A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection," *IEEE Access*, vol. 6, pp. 20881–20892, 2018.
[75] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.
[76] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[78] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, 2010.
[79] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 6–12, 2014, pp. 740–755.
[80] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 340–353.

Ming Tong received the B.S. degree in electronic information engineering from the School of Electronic Information, Wuhan University, Wuhan, China, in 2018, where he is currently working toward the Ph.D. degree in communication and information system.

His research interests include object detection and machine learning.

Shenghua Fan received B.S. degree in computer science and technology from the School of Computer Science, Jianghan University, Wuhan, China, in 2018, and M.S. degree in computer science and technology from the School of Computer Science, Wuhan University, Wuhan, in 2020, where he is currently working toward the Ph.D. degree in software engineering.

His research interests include deep learning target tracking, artificial intelligence, and embedded image processing.

Jiu Jiang received the B.S. degree in measurement and control technology and instrument from the School of Electronic Information, Wuhan University, Wuhan, China, in 2019, where he is currently working toward the Ph.D. degree in communication and information system.

His research interests include image processing and computer vision.

Chu He (Senior Member, IEEE) received B.S. degree in electronics and information systems, M.S., and Ph.D. degrees in communication and information system from Wuhan University, Wuhan, China, in 1996, 1999, and 2007, respectively.

Since 1999, he has been with the School of Electronic Information, Wuhan University, where he has been a Full Professor, since 2014. From 2007 to 2008, he was a Visiting Scholar with the Machine Vision Group, University of Oulu, Finland. He has authored more than 70 articles in related journals and proceedings. He was selected for the talent program of Wuhan Yellow Crane Talent Program (Science and Technology), in 2015. His research interests include image processing and artificial intelligence.