

# Space-Borne SAR High-Resolution Inshore Ship Target Identification Based on a YOLO Recognition and Correction Approach

Yongchen Pan<sup>1</sup>, Graduate Student Member, IEEE, Hongcheng Zeng<sup>2</sup>, Member, IEEE, Yingjie Zhou, Wei Yang<sup>3</sup>, Member, IEEE, Yamin Wang<sup>4</sup>, Member, IEEE, Xinwei An<sup>5</sup>, Graduate Student Member, IEEE, Can Su<sup>6</sup>, Graduate Student Member, IEEE, Wei Xiong, and Wei Liu<sup>7</sup>, Senior Member, IEEE

**Abstract**—High-resolution space-borne synthetic aperture radar (SAR) images are important sources of information for marine monitoring and target recognition. For inshore port scenes, there is a wide variety of ship targets; moreover, as the scattering intensity of the land background is much higher than that of the sea surface, similar structures on land in SAR images can cause significant interference to the identification of ship targets, resulting in a high false alarm rate and false recognition rate. Therefore, the classification of ship targets in SAR images still requires professional interpretation and correction to eliminate possible false alarms and recognitions. In this article, by imitating the basic structural requirements for manual discrimination of target categories, a space-borne SAR high-resolution inshore ship classification and correction method of “Prerecognition—Fine Recognition—Matching Correction” is designed based on the YOLO recognition network. Taking three types of typical ship targets as examples, a high-resolution SAR inshore ship target dataset and a high-resolution SAR ship component architecture dataset are constructed to train the prerecognition network and the fine recognition network, respectively. The latter is based on the former to recognize the component composition of possible targets, and forms a component arrangement architecture vector. The final recognition results are obtained by matching and correcting the similarity distance with the three types of targets. This method can reduce the false alarm rate in ship target recognition in complex inshore port scenes to less than 10% and correct some false recognition results, thereby improving the accuracy of ship target identification and reducing the cost for manual correction.

**Index Terms**—High-resolution space-borne SAR inshore port images, High-resolution SAR inshore ship target

Received 12 May 2025; revised 11 July 2025; accepted 14 August 2025. Date of publication 19 August 2025; date of current version 3 September 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC3305901, and in part by the National Natural Science Foundation of China under Grant U23B2007. (Corresponding author: Hongcheng Zeng.)

Yongchen Pan, Hongcheng Zeng, Yingjie Zhou, Wei Yang, Yamin Wang, Xinwei An, and Can Su are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: pyc\_anais@buaa.edu.cn; zenghongcheng@buaa.edu.cn; zhouyingjie0601@163.com; yangweigigi@sina.com; wangyamin@buaa.edu.cn; anxinwei@buaa.edu.cn; Can\_Su@buaa.edu.cn).

Wei Xiong is with the Institute of Information Fusion, Naval Aviation University, Yantai 264001, China (e-mail: xiongwei@csif.org.cn).

Wei Liu is with the Department of Electrical and Electronic Engineering, Hong Kong Polytechnic University, Hong Kong 999077, China (e-mail: wei2.liu@polyu.edu.hk).

The relative code has been published at the URL: <https://github.com/anaisMissP/YOLORecognition-Correction.git>.

Digital Object Identifier 10.1109/JSTARS.2025.3599933

dataset, High-resolution SAR ship component architecture dataset, Recognition and correction method, Similarity distance matching correction.

## I. INTRODUCTION

INSHORE port scenes are transitional zones between land and sea, and given the complexity in the types and quantities of vessels navigating through these areas, which are key focal points for marine monitoring. Space-borne SAR could overcome the limitations of traditional optical remote sensing radar detecting, allowing all-weather and all-time imaging of the scene [1]. The extensive scene data provide rich resources of information for situational awareness of various marine targets [2].

Based on these abundant SAR data resources, the detection and recognition of ship targets becomes feasible. Traditional approaches such as HOG [3], SIFT [4], and LBP [5] have been investigated, successfully extracting specific features of ships. With the development of deep learning, more and more research efforts have been directed toward exploring SAR ship target recognition methods based on deep neural networks for higher efficiency and accuracy. Currently, convolutional neural networks (CNNs) [6], [7], [8] are the most widely employed deep learning-based approach for SAR ship target recognition. These methods mainly operate by extracting features from SAR images for classification and often incorporate target detection and segmentation techniques [9], [10], [11]. In addition to CNNs, network architectures such as VGG16 [12], ResNet50 [13], [14], [15], RetinaNet [16], and InceptionV3 [17] have been frequently utilized in subsequent studies.

However, these widely used methods encounter obstacles in detecting and recognizing targets in SAR imagery. The SAR signal is the result of multiple reflections, scattering, and interference, with rich interference noise and spurious echoes present in the resulting images [18]. In addition, targets in SAR images often exhibit a “speckled” or “blurred” appearance [18], with overlapping target information between adjacent pixels. This makes it difficult to accurately identify the boundaries and details of the target. Researchers attempt to improve the image quality through various sidelobe suppression methods [19] and help existing algorithms achieve better results, but it is hard to fix all the interference noise and motion blur. The low interpretability

of these images has limited the effects in ship target recognition [20].

To further improve the accuracy of algorithm-based recognition, researchers have delved deeply into target scattering information from various perspectives, which can be mainly categorized into amplitude information, phase information, and polarimetric scattering information.

Polarimetric scattering characteristics refer to the differences in scattering responses of a target under different polarization states [21]. They possess the ability to extract structural parameters of the target, particularly in distinguishing between isotropic and anisotropic targets [21], [22]. This feature exhibits significant advantages in differentiating targets with substantial structural differences, such as accurately identifying small ships on the sea surface under high sea conditions [23]. However, the scattering characteristics of ships with high structural similarity tend to be relatively close [23]. The amplitude and phase information of the SAR echo signals are complementary. Amplitude information generally reflects the overall scattering capability of the target, which is directly related to the structure, surface roughness, and dielectric constant of the target [24]. Phase information typically embodies the physical properties (material characteristics) and geometric properties (height measurement and distance measurement) of the target [24]. Due to its complexity and the lack of a direct correlation with target categories in most cases, phase information is often discarded in many studies, only considering amplitude information reflecting the pixel intensity of SAR images [24], [25]. Currently, some studies have explored the utilization of phase information to enhance image recognition accuracy and achieved certain progress in the complex image domain [25]. Nevertheless, it remains challenging to clarify the association between phase information and the target [26].

Most research still focuses on pixel-level SAR images based on amplitude information. To improve recognition accuracy, techniques such as multitask learning [27], [28], transfer learning [29], [30], [31], and data augmentation [32] have been combined by some researchers to improve the robustness and generalization capabilities of the model, and lessen the dependency of the network on specific datasets. Moreover, prior knowledge, such as port latitude and longitude coordinates, and common types of stationary ships [20], ship radiation noise features [18], and high-resolution range profiles (HRRP) [33] have been considered in the identification process. Some studies have explored the possibility of directly detecting and identifying ship targets in the signal domain [33], [34]. All the methods above can help enhance the recognition accuracy to some degree, but their performance remains rather limited in complex scenes with heavy interference.

One-stage solutions such as SSD [35] and YOLO [36], [37] are commonly used target detection and recognition algorithms. They can perform uniform and dense sampling across the entire scene, thus featuring high efficiency. However, these methods have a high probability of missed and incorrect detections in complex marine environments. The recent network architecture YOLOv8 has been significantly optimized in terms of algorithm accuracy, inference speed, and usability [38]. As a classic recognition network, YOLOv8 is a preferred choice in practice

due to its ease of transportation and verification [39]. Enhanced versions have been made by adding a fourfold downsampled feature map, introducing third-generation variable convolution and exponential moving average attention mechanisms, and replacing the loss function [40]. It can effectively improve accuracy for targets in SAR images, especially those with low information entropy, while still suffering a reduced accuracy in complex nearshore environments, as indicated by a relatively high false alarm rate [40], [41].

To sum up, the fundamental reason behind the challenge for ship recognition in complex scenarios such as inshore port areas is that SAR imaging is based on the electromagnetic scattering characteristics of targets [11], where structures with an approximately bisected angle tend to have stronger scattering echo energy than smoother and planar structures. Existing deep learning methods focus solely on pixel-level target feature extraction, making them prone to false detections in SAR images, especially in cluttered regions or areas with structurally similar strong scattering points [42], [43], [44]. In the nearshore background, the scattering intensity from the land background is much higher than that of the ocean, and buildings or industrial structures in ports may have similar scattering characteristics to ships, making it difficult to accurately segment the data using geographic coordinates alone. Deep learning networks struggle to differentiate between similar interference and targets, leading to high false alarm rates and recognition rates [42], [43], [45]. Although land-sea segmentation is an effective method to reduce the false alarm rate on land, for ship targets located at the land-sea junction, it is very likely to cause missed detection [46], [47]. Also, it is unable to completely cover similar structures along the shore, thus failing to completely solve the problems of false alarms and recognitions [46], [47].

To ensure the accuracy of ship target recognition, SAR image interpretation experts are required to further assess the network model's recognition results. This process is labor-intensive and necessitates updates to deep learning network models to address the high false alarm rates in ship recognition in nearshore scenes. In order to simulate the basic principles used by experts to identify the categories of ship targets, in this article we analyze the key and identifiable components of three ship categories—bulk carriers, container ships, and tankers, and construct a mathematical model of these key component structures to guide and design a fine-grained identification method as “Prerecognition—Fine Recognition—Matching Correction” for SAR inshore port ship targets. As a result, we construct a high-resolution SAR inshore ship target dataset to train the YOLOv8 ship recognition network, and a high-resolution SAR ship component architecture dataset to train the YOLOv8 component recognition network. Then, we add finely recognized component information to possible prerecognized ship target results. The components are further mapped back to the chopped target slices, yielding axial component arrangement architecture vectors. A similarity distance calculation with prior knowledge and target category probability estimation are performed, and the prerecognized ship target identification results are corrected to obtain the final recognition results.

The proposed method, which leverages key component architecture knowledge to guide YOLO recognition and correction,

significantly improves target recognition accuracy in complex environments. This approach reduces false alarms caused by land background interference and lessens the burden of manual verification. The main innovations of our work include the following:

- 1) Based on expert interpretation experience, a high-resolution SAR ship component architecture dataset is constructed, covering three common ship types (bulk carrier, container ship, tanker) and their corresponding components, providing a fine recognition data foundation.
- 2) A YOLO recognition and correction method is proposed, roughly described as “Prerecognition—Fine Recognition—Matching Correction,” addressing the limitations of single feature extraction recognition training methods that struggle with high- and low-scale features, reducing false alarm rates in complex scenes and improving target classification accuracy.
- 3) A matching correction algorithm is designed containing several definitions, such as ship target component arrangement architecture vector and similarity distance. The architecture vector transforms component distribution into a mathematical model. Then, we calculate similarity distances by comparing with standard vectors to correct the final recognition results.

The rest of this article is organized as follows: Section II covers the preprocessing work, including basic YOLOv8 network, knowledge of SAR ship target architecture, and construction of the two SAR high-resolution datasets used in this study. Section III details the YOLO recognition and correction method and its process for recognizing ship targets in complex nearshore SAR images. Section IV presents deployment network verification experiments, confirming that the proposed method improves the accuracy of ship target recognition and reduces high false alarm rate caused by land background interference. Section V discusses the advantages and limitations of the proposed method from accuracy and efficiency aspects. Finally, Section VI concludes this article and discusses future research directions.

## II. PREPROCESSING—NETWORK AND DATASET

### A. YOLOv8 Network—Pure Recognition Network

YOLOv8 is one of the important versions of the YOLO series, occupying a key position in the field of object recognition due to its efficient performance and flexible design [48]. Compared to previous versions, YOLOv8 has made significant progress in detection accuracy, inference speed, and model lightweighting. By introducing an anchor-free architecture, multiscale feature extraction mechanisms, and an improved path aggregation network (PANet), YOLOv8 can efficiently perform object detection tasks in a variety of scenarios, especially excelling in detecting small- and medium-sized targets [38], [49]. The YOLOv8 basic network is composed of backbone, neck, and head and involved in five scales of feature maps, as shown in Fig. 1. P1 to P5 represent five different sizes of feature maps. The convolution layer extracts smaller scale from the former map, while the upsample layer enriches the map to larger scale. The process

of feature map dimensional changes shows the basic image processing in YOLOv8 model.

The backbone of YOLOv8 is based on the improved cross stage partial network (CSPNet), which enhances the computational efficiency and expressive power of the network through feature splitting and fusion. The input features are divided into two parts: one part extracts semantic information through deep convolutions, and the other part retains shallow features and fuses them with the deep features [48], [50]. This process can be expressed as follows:

$$F_{\text{CSP}} = \text{Concat}(F_{\text{shallow}}, F_{\text{deep}}) \quad (1)$$

where  $F_{\text{shallow}}$  and  $F_{\text{deep}}$  represent shallow and deep features, and Concat denotes the feature concatenation operation. Multiscale features are generated through downsampling operations (such as maxpooling or convolutions with stride 2) and provide the foundation for subsequent feature fusion and detection.

The neck structure uses the improved PANet, which efficiently fuses multiscale features [38], [48], [51]. PANet employs a bidirectional feature aggregation mechanism—both top-down and bottom-up—to combine high-resolution detail features with low-resolution semantic features. For given high-resolution features  $F_{i+1}$  and low-resolution features  $F_i$ , the core operation can be expressed as follows:

$$F_{\text{agg}} = W_{\text{up}} * \text{Upsample}(F_{i+1}) + W_{\text{down}} * F_i \quad (2)$$

where Upsample represents the upsampling operation.  $F_{i+1}$  denotes the weight parameters for the convolution between the upsampled result and the convolutional layer, and  $F_i$  denotes the weight parameters for the direct convolution. Both weights enable the feature maps of different layers to be fused with each other. By repeatedly performing upsampling and downsampling, PANet fully fuses the detail information of high-resolution features with the semantic information of low-resolution features, providing strong support for multiscale detection [38].

The detection head uses the anchor-free architecture, directly predicting the target box's center coordinates  $(x, y)$ , width  $w$ , height  $h$ , and class probability  $p(c)$ . The matching degree between the predicted box and the ground truth is optimized through a loss function that integrates position, shape, and class information [48], [51], [52]. The total loss function is defined as follows:

$$L = \lambda_{\text{coord}} L_{\text{IoU}} + \lambda_{\text{conf}} L_{\text{conf}} + \lambda_{\text{cls}} L_{\text{cls}} \quad (3)$$

where  $L_{\text{IoU}}$  is the intersection over union (IoU) loss, which measures the overlap between the predicted and ground truth boxes,  $L_{\text{conf}}$  is the confidence loss, which reflects the probability of a target existing in the box, and  $L_{\text{cls}}$  is the classification loss, which describes the difference between predicted and true classes. The weighting coefficients  $\lambda_{\text{coord}}$ ,  $\lambda_{\text{conf}}$ , and  $\lambda_{\text{cls}}$  balance the contribution of each part of the loss.

However, despite YOLOv8 emphasizing multiscale feature extraction, it is restricted when dealing with target categories that span large scales. While the backbone network has multiscale feature extraction capabilities, this design is more suitable for the fusion of adjacent-scale features [38]. Although the PANet neck structure enhances the interaction between semantic and detailed features, its fusion capability proves insufficient when

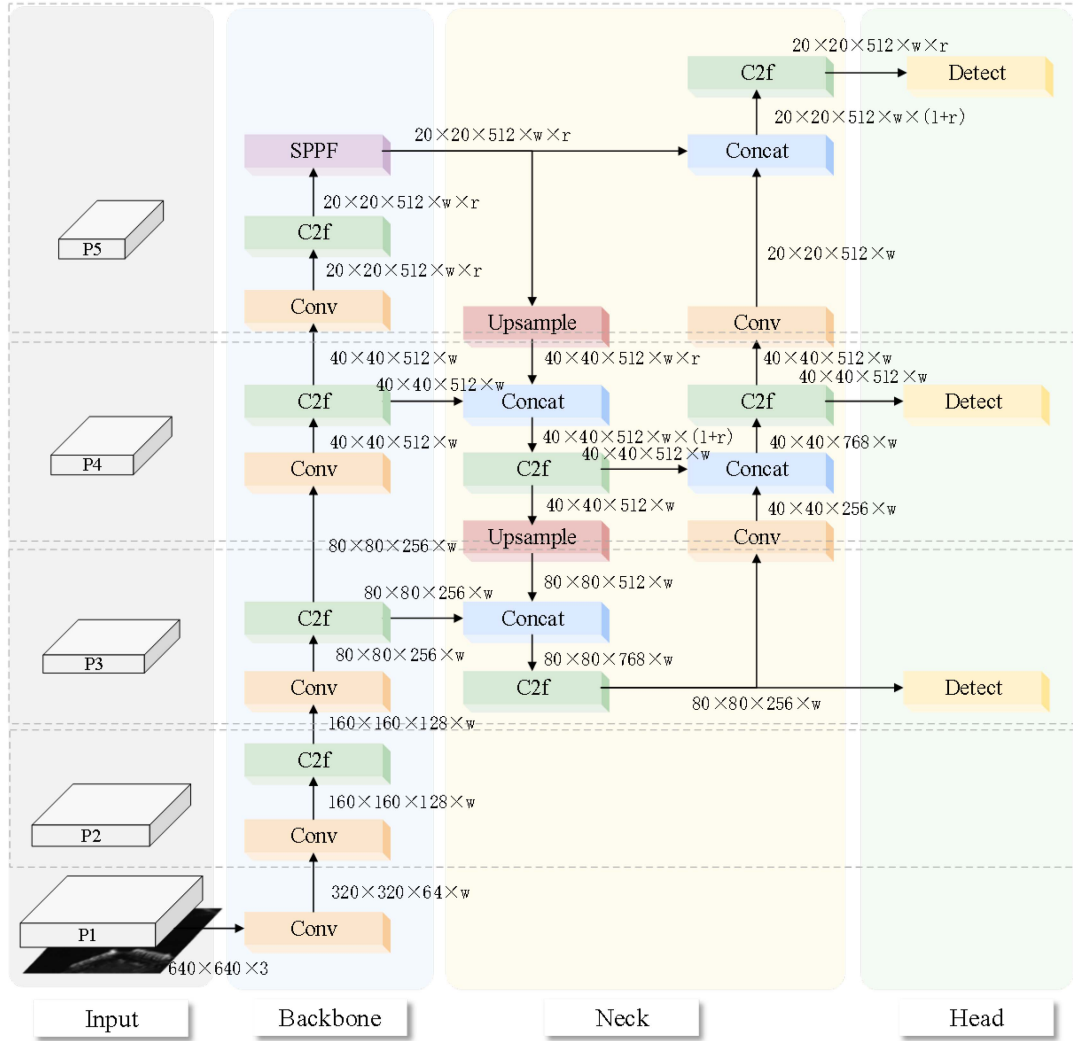


Fig. 1. YOLOv8 basic network architecture.

handling tasks involving extreme scale variations [38], [52]. This limitation progressively weakens valuable information during feature aggregation, increasing the model's susceptibility to semantic misinterpretation in cross-scale scenarios. Further complications arise with anchor-free architectures. While they simplify bounding box design, their accuracy for extremely large or small targets heavily depends on the adaptability of the feature map scale [48], [51]. In tasks with substantial scale differences, target box initialization becomes sensitive to feature map resolution, leading to suboptimal loss optimization. This is particularly problematic for ship recognition, where the spatial distribution of the entire ship vastly differs from its smaller components. A single detection head struggles to simultaneously address the distinct scale requirements of these two target types. Consequently, when learning ship target features, the YOLOv8 model often fails to effectively account for component-scale characteristics. This frequently results in misjudging structurally similar objects as specific ship classes, leading to persistent false alarms in nearshore environments and occasionally causing the false alarms and false recognitions of ship categories altogether, as illustrated by examples in Fig. 2.

Therefore, we consider further decomposing the ship recognition task. In inshore complex port scenarios, we can first recognize the ship target as prerecognition results, and then refine the key components' recognition and correct the original results through component architecture, better adapting to the characteristics of YOLOv8 and improving the false alarm and false recognition conditions.

### B. High-Resolution SAR Inshore Ship Target Recognition Dataset

This article uses high-resolution images collected by C-band and X-band SAR satellites as the original dataset, covering a diverse range of targets in nearshore complex scenarios. The dataset consists of 1190 images, including different types, angles, and scales of ship images. The collected images reach a resolution of 0.5 m and 1 m, providing strong resolution for fine details of targets. The slice size is set to  $1024 \times 1024$ . This setting is determined by taking into account both the optimal resolution and the maximum possible scale of the vessels, so as to avoid a situation where the area of a ship target within a single

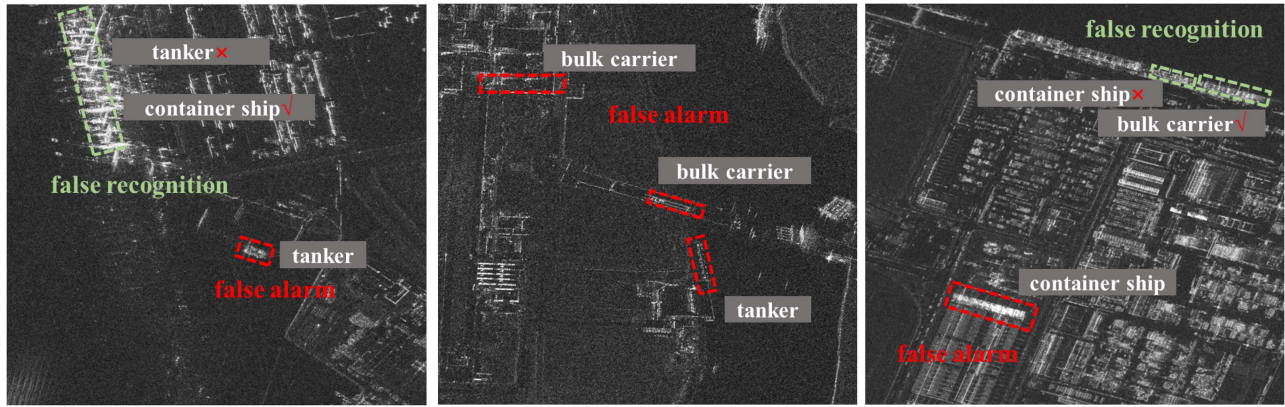


Fig. 2. Some examples for false alarm and false recognition in inshore scenarios: The target within the red box is a false alarm target, which means that the location does not correspond to any type of ship target; the target within the green box is a false recognition target and although there is a ship at this location, the category is incorrect.

TABLE I  
HIGH-RESOLUTION SAR INSHORE SHIP TARGET RECOGNITION DATASET  
INFORMATION

| Acquisition Device    | X-band AS01, C-band GaoFen-3                                            |     |
|-----------------------|-------------------------------------------------------------------------|-----|
| Number of Images      | 1190                                                                    |     |
| Average Image Size    | 1024 × 1024                                                             |     |
| Train/Val Split Ratio | 4:1                                                                     |     |
| Sea Resolution        | 0.5m, 1m                                                                |     |
| Location              | Yangpu Port, Tianjin Port, Keelung Port, Shanghai Port, Alexandria Port |     |
| Number of Ships       | bulk carriers                                                           | 493 |
|                       | Container ships                                                         | 369 |
|                       | Tankers                                                                 | 530 |

slice accounts for no more than 70% of its total area, which could lead to incomplete detailed information of the target. There are 2483 labeled ship targets in total, including 493 bulk carriers, 369 container ships, and 530 tankers. Overall, the dataset not only provides high resolution and diverse scene information, but also reflects the prominent features of the complex near-shore environment, providing important data support for research. Table I lists the relevant details of the dataset, and Fig. 3 shows some example images.

### C. Knowledge of SAR Ship Target Recognition

Target recognition in SAR remote sensing images differs significantly from that in optical images. Optical images can distinguish targets from their surroundings by color and contour but are highly affected by clouds and weather conditions. In contrast, SAR images capture the target's electromagnetic wave reflections and form images based on the intensity of reflected energy [1]. Although pseudocolor techniques are available, SAR images are typically grayscale, providing less information compared to RGB images [53]. The SAR imaging process makes image intensity more sensitive to dihedral structures with strong reflectance. The strong scattering points in the image represent structural information, which differs from human eye's usual judgment criteria [53]. Consequently, those unfamiliar with the SAR domain often struggle to analyze such images.

Although deep learning networks can extract target features from SAR imagery, ship recognition in nearshore environments presents unique challenges. The intense backscatter from complex coastal terrain—significantly stronger than sea surface reflections—interferes with neural networks' sliding-window detection. This results in elevated false alarm rates when identifying vessels in port areas [54], [55]. In practice, to ensure recognition accuracy, SAR ship recognition experts have to manually interpret the images, which takes considerable human resources [53]. Experts rely on extensive learning and work experience to establish a classification system for ship targets, identifying ships based on size, shape, and structural components. The knowledge system of professional SAR target recognition is complex and diverse. This article attempts to select and simplify the content related to the basic architecture of targets, which can play a guiding role in both recognition network learning and matching correction logics.

Therefore, we focus on three types of civilian ships: bulk carriers, container ships, and tankers, because they are common cargo ships in inshore ports. Comparing optical and SAR images of the same ship types, we analyze and model distinctive structural features for each type in SAR images, as illustrated in Fig. 4. In SAR images, the scattering of objects is formed by their own scattering intensity, speckle superposition, and noise [56]. Due to the randomness of scattering caused by speckle and noise, the intensity varies within a certain range. Therefore, we only discuss the intrinsic scattering characteristics, denoted as radar cross section [57], which can be theoretically analyzed through the target's own structure. There is no significant difference in material among the three types of ship targets. However, due to the different cargoes they transport, the corresponding ship components vary. In high-resolution SAR images, the components and surrounding ship structural features lead to different scattering characteristics. According to the characteristics of SAR imaging, under the general dihedral corner structure, most of the scattered echo energy can be retained after multiple reflections, so the corresponding image pixel brightness is high [56]; smooth surfaces approximate specular reflection, where most energy is difficult to reflect back in the emission

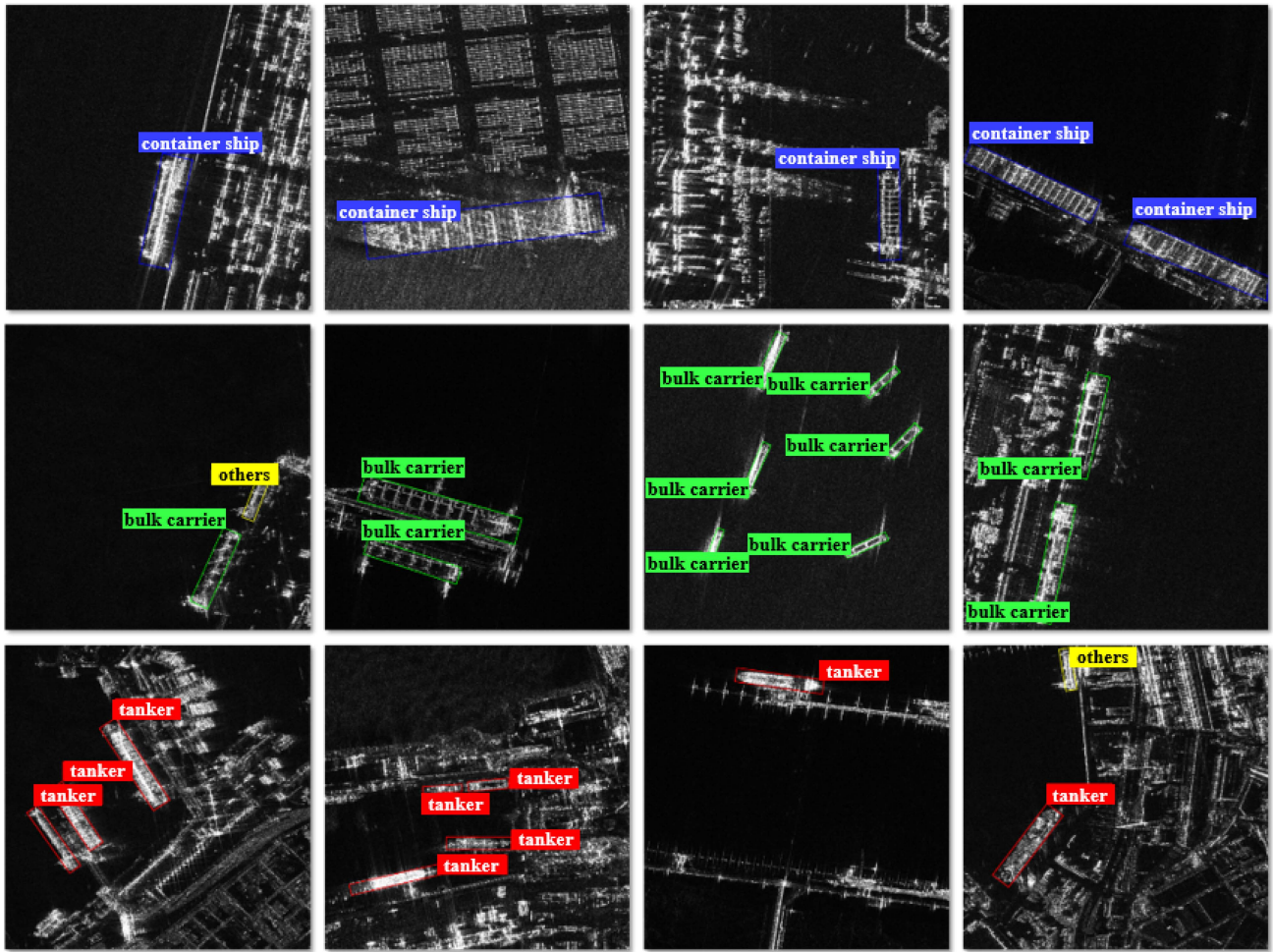


Fig. 3. Examples from high-resolution SAR inshore ship target recognition dataset: the targets in the blue box are container ships, in the green box are bulk carriers, in the red box are tankers, and in the yellow box are other types of ships.

direction, so the corresponding pixel brightness is low, while rough surfaces are the opposite [56], [57]. Based on these basic interpretation rules of SAR images, we analyze the component structures of the three types of ship targets and their corresponding SAR scattering characteristics.

1) *Bulk carriers*, as shown in Fig. 4(a), include two types of components: bridges and cargo holds [54].

a) *The bridge* is mostly located at the stern and is generally higher than other components on the ship. This is because the wheelhouse is at the upper end of the bridge, requiring a wide field of view. The bridge is equipped with many sensor devices, resulting in an uneven surface and complex radar signal reflection, presenting obvious bright spots in SAR images.

b) *Cargo holds* are the typical components of bulk carriers. There are generally 3 to 5 of them, possibly more, for loading various goods on board. They are rectangular parallelepipeds with very smooth surfaces, aiming to maximize cargo efficiency and simplify the cleaning process. The flat hold surface forms weak backscattering due to specular reflection, while the

edges of the hatch openings produce strong scattering bright lines due to edge diffraction. Moreover, there is often a certain distance between cargo holds. Therefore, in SAR images, they present rectangular structures with obvious and high-brightness edges and large low-gray areas in the middle.

2) *Container ships*, as shown in Fig. 4(b), include two types of components: bridges and containers [54].

a) *The bridge* is mostly located in the middle of the ship, leaning toward the stern. This is because container ships are relatively large in size, and a middle layout can significantly reduce the blind area of the bow view, avoid blocking the view by stacked high-rise containers, and facilitate the balance of the ship's center of gravity. The basic structure of the bridge is not much different from that of the bulk carrier; it is also higher than other components on the ship, with an uneven surface and strong scattering, presenting obvious rectangular bright spots in SAR images, and its width may exceed the hull.

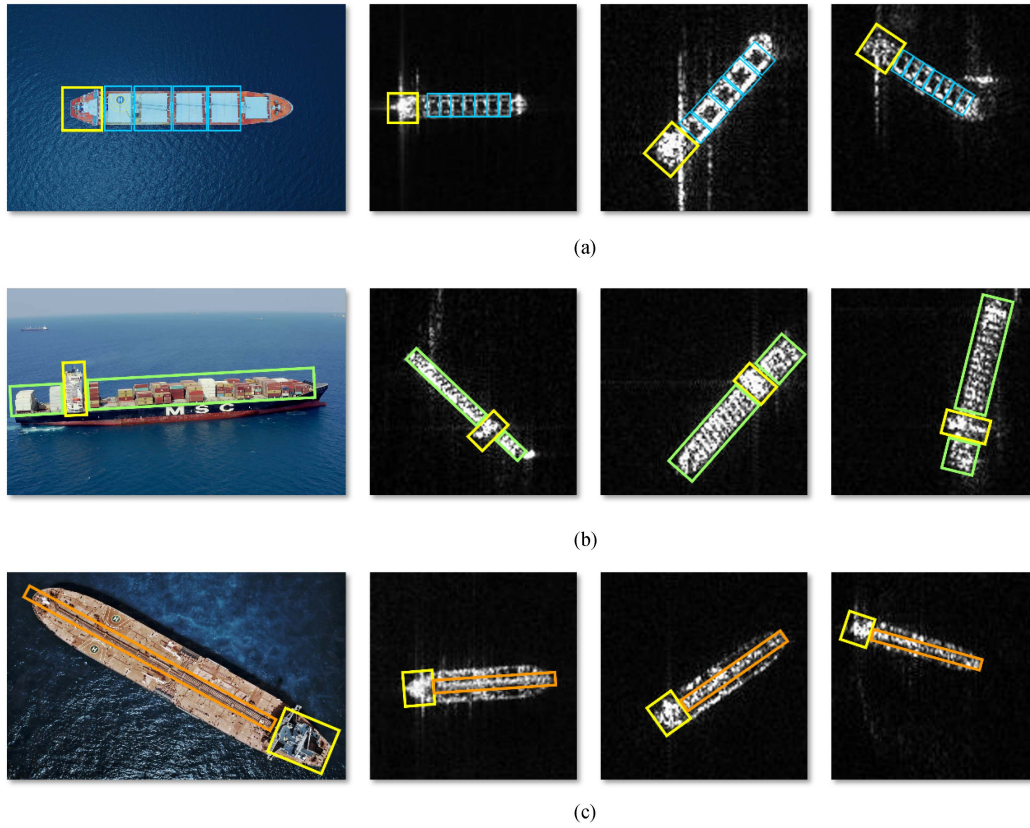


Fig. 4. Comparison of ship component structures in optical and SAR images: The yellow box represents the bridges of the three types of ships, the blue box represents the cargo holds, the green box represents the containers, and the orange box represents the oil pipelines; the key components of various ships in the optical image correspond one by one to those in the SAR images. (a) Bulk Carriers. (b) Container Ships. (c) Tankers.

- b) *Containers* are located on both sides of the bridge, divided into foredeck containers and afterdeck containers. Containers are mostly rectangular parallelepipeds made of steel. To enhance bending strength and accommodate more goods, their surfaces are uneven, designed with grooved bulkheads. Therefore, the surface scattering intensity is high, and the edges also produce strong scattering bright lines. Both foredeck and afterdeck containers appear as bright spots in SAR images. Generally, foredeck containers are longer and larger in volume than afterdeck containers. Containers are stacked together without obvious gaps between them, their height does not exceed the bridge, and their width does not exceed the hull, presenting large-area rectangular regional bright spots in SAR images.
- 3) *Tankers*, as shown in Fig. 4(c), include two types of components: the bridge and oil pipelines [54].
  - a) *The bridge* is also located at the stern. To ensure the field of view, it is higher than other components on the ship, with an uneven surface, presenting bright spot features in SAR images. Due to the consistency of the bridge's function, there is no significant difference in the bridges among the three types of ships.
  - b) *Oil pipelines* are generally located in front of the bridge, along the centerline of the ship, almost a relatively thick pipe from the bow to the bridge, mainly for

the convenience of storing and transporting goods such as oil. The oil pipelines on the deck of the tanker are usually buried in deck trenches or corridors to avoid physical collisions and the risk of oil and gas leakage. The remaining areas of the deck remain flat without protrusions, not only to avoid the accumulation of oil-containing water but also to ensure the safe passage of crew members, while meeting the rigid requirements of fire isolation belts. Therefore, only the ship's edge eaves and the angle between the oil pipelines and the deck have a dihedral corner structure, presenting linear bright lines in SAR images, with low gray values on both sides of the deck until the edge of the ship, where the gray value is higher.

#### D. Key Component Architecture Dataset

To integrate structural information into deep learning networks, we construct a high-resolution SAR ship component architecture dataset as a training foundation based on the above analysis. Given the need for high-resolution images to analyze and annotate ship components, images from AS01 and GaoFen-3 satellite within 1 m resolution were used to build the dataset.

The dataset consists of 386 original single-target slices, including multiangle high-resolution SAR images of three

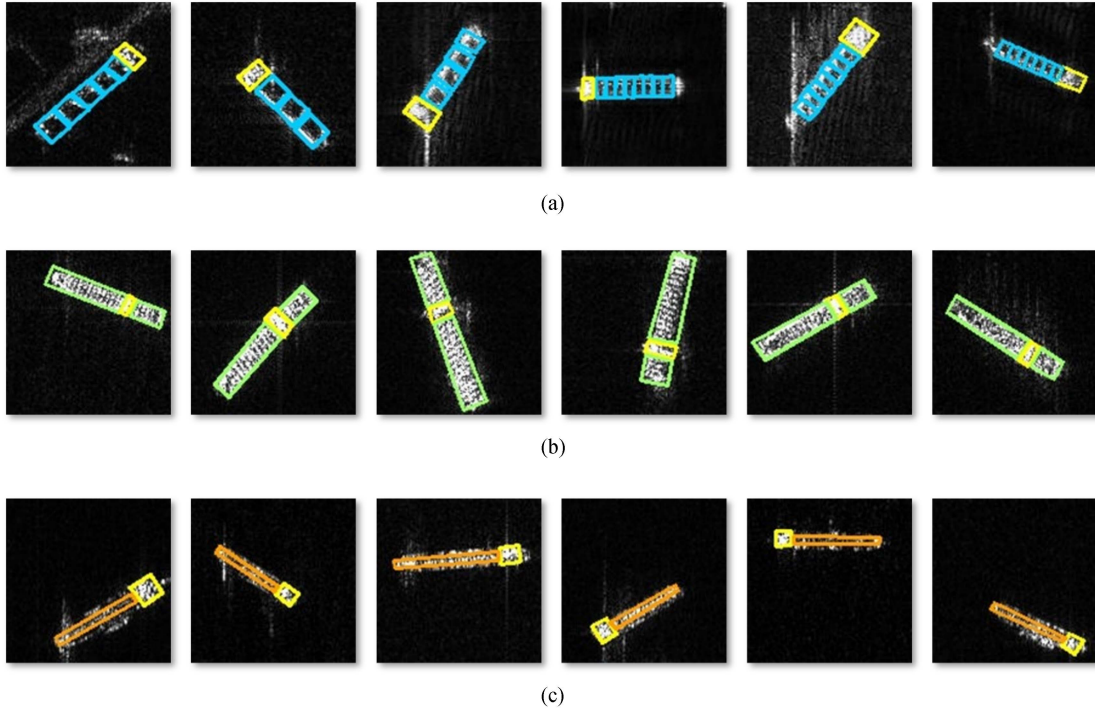


Fig. 5. Baseline examples from the high-resolution SAR component architecture dataset: The yellow box represents the bridges of the three types of ships, the blue box represents the cargo holds, the green box represents the containers, and the orange box represents the oil pipelines. (a) Bulk carriers. (b) Container Ships. (c) Tankers.

TABLE II  
HIGH-RESOLUTION SAR COMPONENT ARCHITECTURE DATASET INFORMATION

| Acquisition Device     | X-band AS01, C-band GaoFen-3 |      |
|------------------------|------------------------------|------|
| Number of Slices       | 1904                         |      |
| Average Slice Size     | $1024 \times 1024$           |      |
| Training-to-Test Ratio | 4:1                          |      |
| Resolution             | 1m, 0.5m                     |      |
| Number of Ships        | Bulk Carriers                | 748  |
|                        | Container Ships              | 528  |
|                        | Tankers                      | 628  |
| Number of Components   | Bridge                       | 1898 |
|                        | Cargo Hold                   | 4188 |
|                        | Container                    | 988  |
|                        | Oil Pipeline                 | 628  |

ship target types. The component recognition network learns component architectures, performing detailed analysis of small-scale components. Based on the 386 original image data at a size of  $128 \times 128$ , we can distinguish the different key component architectures corresponding to different types of vessels. However, in the inshore scenarios, the scale span of the three types of ships is relatively large. Therefore, we upsample the original image slice data to sizes of  $256 \times 256$ ,  $512 \times 512$ , and  $1024 \times 1024$ , and then complete all the slices into  $1024 \times 1024$  component fine-recognition slices. This approach expands the dataset and enhances its generalization ability. As a result, the dataset consists of 1904 images, including 748 bulk carriers, 528 container ships, and 628 tankers. Annotations were made for target categories and critical components, resulting in 1898 bridges, 4188 cargo holds, 988 containers, and 628 oil pipelines.

In general, the dataset features a wide variety of types and quantities of components, with distinct characteristics in the component architectures. After applying reasonable data augmentation operations, such as noise addition and rotation, the dataset effectively aids the YOLOv8 network in learning small-scale component features, providing essential data support for this research. Table II provides detailed information about the component architecture dataset, while Fig. 5 presents examples of annotated structures.

### III. METHODOLOGY

In response to the task of identifying three types of ship targets in complex inshore scenarios, we propose a “Prerecognition—Fine Recognition—Matching Correction” recognition and correction network guided by the key component architecture. Due to simplicity of the component structures, they are easily interfered with by strong scattering points in complex scenes. Using a single YOLOv8 network may result in a higher false alarm rate, and cross-scale feature learning could reduce the network’s learning performance. Therefore, we employ two networks to separately learn the features of ships and components, achieving more accurate results. The first YOLOv8 network is directly used for ship recognition in high-resolution SAR images of complex inshore scenes to obtain initial recognition results. Based on the results of ship prerecognition, target slices are cut from original images. The second YOLOv8 network finely recognizes the key components of the prerecognized targets. Based on the results of component fine recognition, we design a

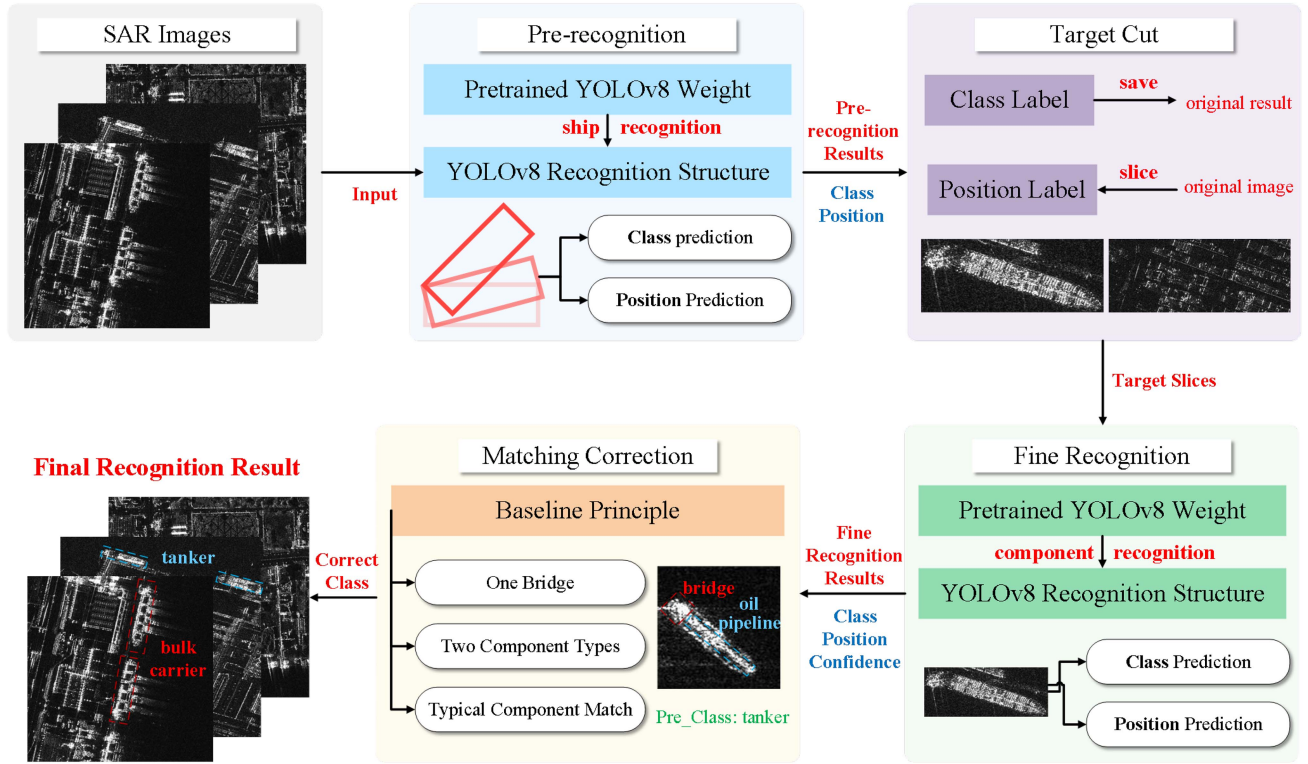


Fig. 6. Proposed “Prerecognition—Fine Recognition—Matching Correction” method.

matching correction algorithm through component arrangement architecture vector and similarity distance calculations to finally correct prerecognition results. Fig. 6 illustrates the main steps of the proposed YOLO recognition and correction method.

First, we input SAR images of inshore scenarios into the YOLOv8 ship prerecognition network to recognize ship targets. Based on the prerecognition results including the class and position of possible targets, the rotated bounding box images of the targets are sliced, completed, and quantized. We then use the YOLOv8 fine recognition network to recognize key components of possible ship targets from the quantized slices, and obtain the fine recognition results including the class, position, and confidence of components. Recognized components are filtered and formed into an architecture vector. Three base principles are set to determine whether the prerecognition result is convincing according to corresponding components, and correct ship category recognition results by matching fine recognition architectures with standard component architectures. The detailed method is stated as follows.

#### A. Ship Target Recognition in Inshore Scenarios

The high-resolution SAR inshore ship target recognition dataset comprises 1904 image slices at a size of  $1024 \times 1024$  with a resolution of 1 m and 0.5 m. These complex port scenes contain diverse ship types exhibiting significant size and shape variations. In particular, even within the same ship category and identical resolution, vessel dimensions differ substantially. For example, container ships range from under 200 m to over 400 m

in length, while other ship types show even smaller scales. In the proposed ship recognition method, since the subsequent fine recognition of components and matching correction are all based on the results of the previous steps, it is particularly important not to miss any targets during the prerecognition as much as possible. Facing the ship targets with a large scale span and high background interference under the  $1024 \times 1024$  slices, we choose to use the YOLOv8-p6 model architecture for training and prediction. The YOLOv8-p6 model is a branch of this series, mainly designed for deep learning with high-resolution input [48]. By introducing an additional detection head (P6 layer), a larger feature map output of  $40 \times 40$  is added in the deep layer of the network, which enhances the detection and recognition capabilities for large-scale objects and dense scenes [48].

The network was successfully trained for 400 epochs, yielding the theoretically optimal weight file for feature parameters. When identifying ship targets in complex coastal scenarios, the backbone network is configured using the parameter settings in the weight file to extract features, as shown in Fig. 7. The dataset labels target location information using four vertices of rotated bounding boxes. During network processing, these four points are first converted into the center position  $(x_c, y_c)$ , box width  $w$ , height  $h$ , and rotation angle  $\theta$  to represent the rotated bounding box. At the final output stage, the coordinates are reverted to the original four-point representation. This transformation facilitates the network in assigning prior rotated bounding boxes across feature maps at each position and orientation. The convolutional layers then output the target confidence score as (4),

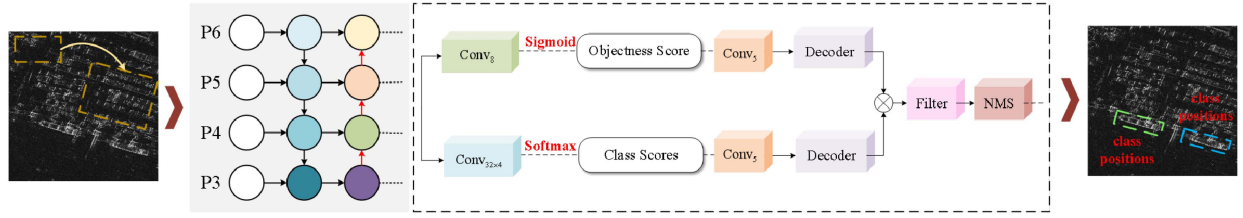


Fig. 7. Prerecognition: The YOLOv8-p6 ship target recognition diagram, where the yellow box represents rotated bounding box and at each position, it attempts to iterate through different sizes and rotation angles to check if there are situations similar to the three types of ship targets, i.e., having high values of Objectness Scores  $\times$  Class Scores.

which is passed through a sigmoid function to map the score to a probability range of 0 to 1, representing the likelihood of a target existing at the given position.

$$\begin{aligned} \text{Objectness Score}(x_c, y_c, w, h, \theta) \\ = \text{Sigmoid}(\text{Conv}_{80}(x_c, y_c, w, h, \theta)) \end{aligned} \quad (4)$$

For each rotated bounding box, the convolutional layers output raw scores for four categories, which are then normalized into a probability distribution using the Softmax function as follows:

$$\begin{aligned} \text{Class Scores}(x_c, y_c, w, h, \theta, c) \\ = \text{Softmax}(\text{Conv}_{32 \times 4}(x_c, y_c, w, h, \theta)). \end{aligned} \quad (5)$$

The convolutional layers further output bounding box refinement parameters, which are decoded to produce accurate rotated bounding box coordinates as follows:

$$\begin{aligned} \text{Predicted OBB Boxes}(x_c, y_c, w, h, \theta, c) \\ = \text{Decoder}(\text{Conv}_5(x_c, y_c, w, h, \theta)) \end{aligned} \quad (6)$$

The target confidence scores are multiplied by the category probabilities to obtain composite scores for each category at every position and rotated prior box. Unreasonable or low-scoring candidate boxes are filtered out, and a nonmaximum suppression (NMS) algorithm is applied to eliminate redundant rotated bounding boxes with high overlap and relatively lower confidence scores. This process results in the final rotated bounding box detection as follows:

$$\begin{aligned} \text{Final Detection} \\ = \text{NMS}(\text{Filtered}(\text{Objectness Scores} \times \text{Class Scores})). \end{aligned} \quad (7)$$

The output coordinates  $(x_c \pm \frac{w}{2} \times \cos \theta, y_c \pm \frac{h}{2} \times \sin \theta)$  of the rotated bounding box are converted back to normalized four-point coordinates  $(\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_3, \hat{y}_3, \hat{x}_4, \hat{y}_4)$  and exported as a text file, including the prerecognized ship category, coordinates, and confidence score.

### B. Prerecognized Target Slice Processing

Based on the recognition results of ship targets, we obtain the ship category, coordinates (the horizontal and vertical coordinates of the four points of the rotated bounding box), and the category confidence. As mentioned, this article mainly focuses on three types of ship targets and the other ship targets recognized can be directly considered as background because they are generally smaller fishing vessels. For each ship target, we only

retain the category result with the highest confidence level as the prerecognition result.

Since the architecture of the target key component is relatively simple and it is easily interfered by the scattering background, we need to copy the identified target from its original position and create new slices, magnifying the details and reducing the interference, which would facilitate fine recognition of the component. Therefore, we propose a dynamic magnification strategy to enhance small-scale ship details. For targets with original dimensions  $\leq 200 \times 200$  pixels, we apply  $4\times$  upsampling using bilinear interpolation to amplify subtle structural features; for targets where either width or height  $\leq 200$  pixels, a  $2\times$  magnification is performed. This adaptive scaling ensures that critical details in small targets are preserved while avoiding excessive interpolation artifacts in larger targets. The magnified slices are then embedded into a standardized  $1024 \times 1024$  canvas through zero-padding, centered to maintain spatial consistency.

Specifically, based on the four-point position results obtained from the recognition, we retain the target within the smallest positive rectangular area, and replace the remaining part with pure black color. The magnification process operates on the cropped region  $I_{\text{slice}}$  prior to padding. Let  $w_s$  and  $h_s$  denote the width and height of  $I_{\text{slice}}$ , respectively. The scaling factor  $\lambda$  is determined as follows:

$$\lambda = \begin{cases} 4 & \text{if } w_s \leq 200 \wedge h_s \leq 200 \\ 2 & \text{if } w_s \leq 200 \vee h_s \leq 200 \\ 1 & \text{otherwise.} \end{cases} \quad (8)$$

The resized slice  $\tilde{I}_{\text{slice}}$  is then computed through bilinear interpolation as follows:

$$\tilde{I}_{\text{slice}} = \mathcal{R}(I_{\text{slice}}, [\lambda w_s], [\lambda h_s]) \quad (9)$$

where  $\mathcal{R}(\cdot)$  represents the resizing operator. This adaptive magnification enhances the resolvability of ship components while maintaining aspect ratios. We subsequently pad  $\tilde{I}_{\text{slice}}$  with zeros to form a standardized  $1024 \times 1024$  slice  $L$ , ensuring compatibility with downstream recognition networks. The final padded slice preserves radiometric fidelity as 16-bit unsigned integer TIFF format, crucial for retaining SAR intensity dynamics. The complete processing flow is shown in Fig. 8.

The prerecognition position of each target is a normalized result as  $(\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_3, \hat{y}_3, \hat{x}_4, \hat{y}_4)$ . Multiplied with  $(w, h)$ , the width and height of the corresponding SAR image  $I_0$  accordingly, we obtain the four-point coordinates of the prerecognition position  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$  on

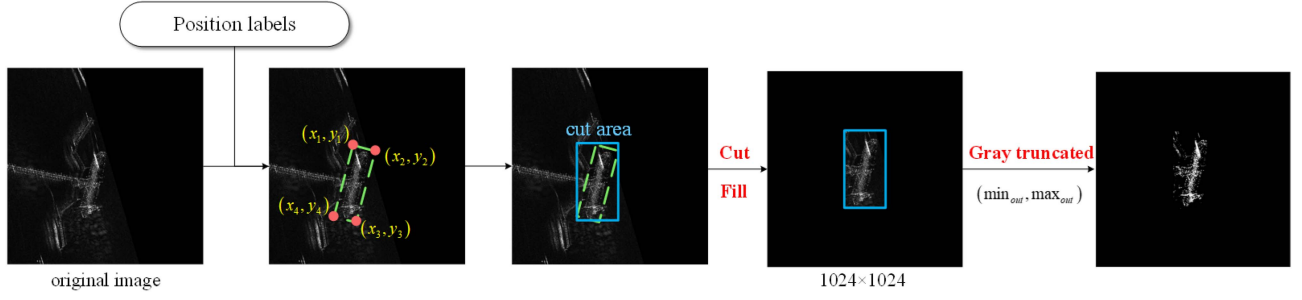


Fig. 8. Target cut diagram: Based on the results of the four-point coordinates of the prerecognition positions, we obtain the minimum bounding upright rectangle outside the rotated bounding box; then, we cut the upright rectangular slice into a prerecognition slice of the ship; taking this slice as the center, we complete the background and quantize it to highlight the details of the prerecognized target.

the image. According to the maximum and minimum values of the horizontal coordinates as  $x_{\max} = \max(x_1, x_2, x_3, x_4)$ ,  $x_{\min} = \min(x_1, x_2, x_3, x_4)$ , and the maximum and minimum of the vertical coordinates as  $y_{\max} = \max(y_1, y_2, y_3, y_4)$ ,  $y_{\min} = \min(y_1, y_2, y_3, y_4)$ , a rectangular image is cropped from the original image as the prerecognized target slice as follows:

$$I_{\text{slice}} = I_0(x_{\min} : x_{\max}, y_{\min} : y_{\max}). \quad (10)$$

Thus, we can obtain the width of the slice  $w_s = x_{\max} - x_{\min}$ , and the height  $h_s = y_{\max} - y_{\min}$ . Take this slice as the center of a new image, and pad it with zeros around to form a  $1024 \times 1024$  slice, denoted as  $L = 1024$ . Equation (11) shown of the bottom of this page defines the new standard slice as follows:

### C. Component Fine Recognition Network

Based on the original data of 1544 images, we can distinguish the different key component architectures corresponding to different types of ships. Since the key components of ships are relatively small in size compared to the target itself, some components such as cargo holds and bridges may only occupy dozens of pixels. The general YOLOv8 model does not have the ability to recognize tiny objects. Therefore, we adopt the YOLOv8-p2 model for component training and prediction. The YOLOv8-p2 model is also a branch of this series, which is mainly designed for the detection of small-sized targets. By introducing a higher resolution feature layer (P2 layer), the resolution of the feature map can reach up to  $160 \times 160$ , retaining more detailed information [48]. It is suitable for scenarios with dense small objects whose size is larger than  $5 \times 5$  pixels and can also adapt to image inputs of  $1024 \times 1024$  size [48]. The specific method for fine-grained component recognition is shown in Fig. 9.

The network was successfully trained for 500 epochs, yielding a weight file with theoretically optimal feature parameters. During component identification, the backbone network is configured using the parameters in the weight file to extract features.

Compared to the YOLOv8 network for target recognition, this network uses a different backbone and detection head, which will not be elaborated here. The differences lie in the component categories and feature scales. The final output includes the following:

- 1) Component categories  $\text{Class}_{\text{component}}$ ;
- 2) Four-point position coordinates  $\text{Position}_{\text{component}}$ ;
- 3) Confidence scores  $\text{Confidence}_{\text{component}}$ .

### D. Component Architecture Matching and Correction

We design a component architecture matching and correction method, as shown in Fig. 10, based on knowledge of SAR ship recognition analyzed in Section II.

We only retain component recognition results with confidence scores greater than 0.1 and establish the following three baseline principles:

- 1) *Each ship target must have a bridge*: The bridge is a critical component that every ship target must have. If no bridge is recognized, the target is corrected as background. According to the practical engineering condition, the situations that do not meet this principle probably include
  - a) If the highest-confidence component is a cargo hold, it might correspond to a land-based bulk cargo area;
  - b) If the highest-confidence component is a container, it might correspond to a container yard;
  - c) If the highest-confidence component is an oil pipeline, it might correspond to similar land-based structures.
- 2) *Ship targets consist of at least two types of components, including the bridge*: For the three types of ship targets analyzed in this study, the key components include a common bridge and another component unique to each type. If only the bridge is present, the target may be a small vessel or a land-based strong scattering region, and it is classified as background. If two or more types of components are present, the next principle must be taken into consideration.

$$I_{1024}(i, j) = \begin{cases} 0 & \text{if } i < \frac{L-w_s}{2} \text{ or } i \geq L - w_s \text{ or } j < \frac{L-h_s}{2} \text{ or } j \geq L - h_s \\ I_{\text{slice}} \left( i - \frac{L-w_s}{2}, j - \frac{L-h_s}{2} \right) & \text{if } \frac{L-w_s}{2} \leq i < L - w_s \text{ and } \frac{L-h_s}{2} \leq j < L - h_s. \end{cases} \quad (11)$$

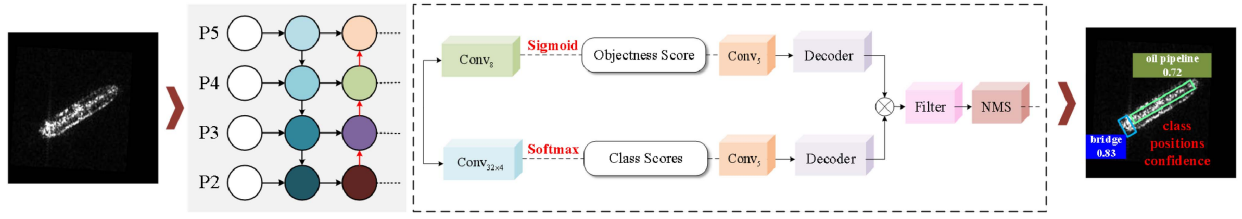


Fig. 9. Fine recognition: YOLOv8-p2 component architecture recognition diagram, where the basic recognition method is consistent with that of the prerecognition network.

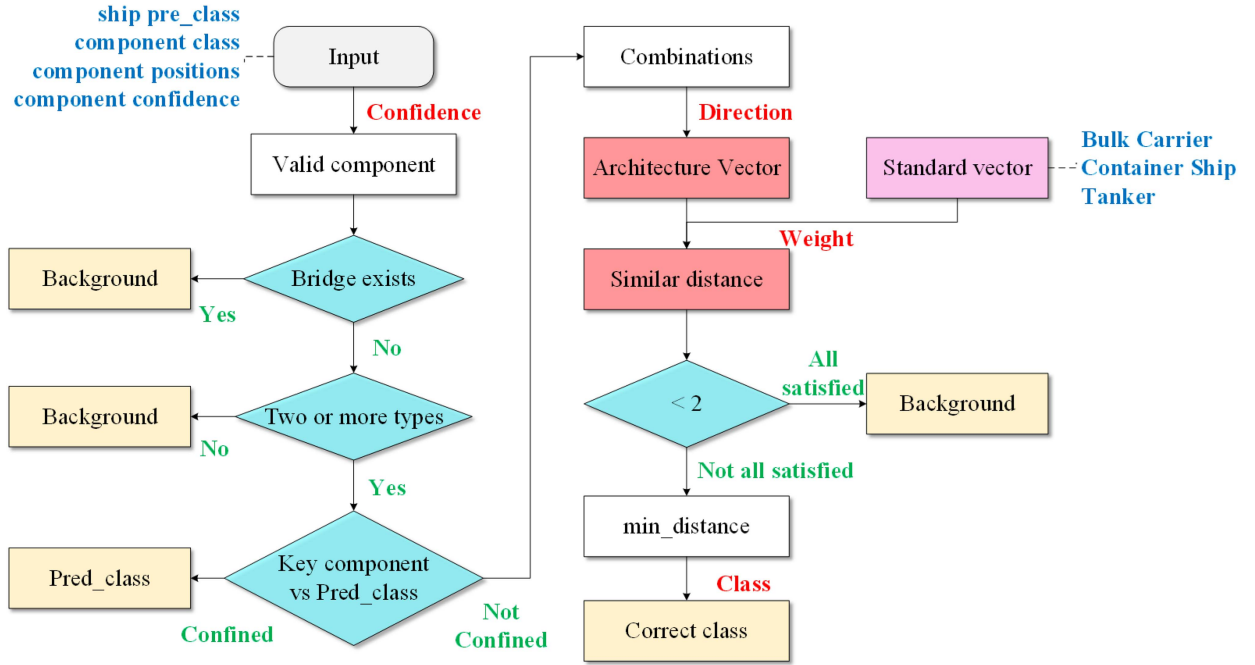


Fig. 10. Flowchart of the proposed component architecture matching and correction method.

- 3) *The ship target category corresponds to the highest-confidence key component type, excluding the bridge:* Among valid components, excluding the bridge, the highest-confidence key component should match the ship target category's characteristic component. If the highest-confidence component contradicts the target category, this may indicate a classification error or interference from strong land-based signals. Further evaluation is then needed through the following component architecture matching and correction method.

Component recognition results satisfying all three principles retain their prerecognized target categories. Otherwise, corrections are applied. In a complex inshore port scenario where valid components include the bridge but the highest-confidence component does not match the target type, we design a similarity distance matching scheme to correct the final recognition results.

Assume that the valid components of a target are characterized by their categories, positions, and confidence scores. The four types of components are encoded as follows: the bridge is 0, the cargo hold is 1, the container is 2, and the oil pipeline is 3. The four points of a bounding box are defined in clockwise order starting from the top-left point. Assume the number of

component types, excluding the bridge, is  $N$ . Because each ship has exactly one bridge, the direction of the component architecture is defined as follows:

direction =

$$\begin{cases} \max x_2 - \min x_4 \geq \max y_1 - \min y_3, \vec{x} \\ \max x_2 - \min x_4 < \max y_1 - \min y_3, \vec{y}. \end{cases} \quad (12)$$

For each valid bridge, the bridge and other valid components are sorted according to this arrangement direction to form the *component recognition vector* ( $\text{cls}_1, \text{cls}_2, \dots, \text{cls}_n$ ). The *standard component architecture vectors* for the three ship types are defined as follows:

$$\begin{aligned} \text{bulk carrier} &: \{0, 1, 1, \dots\} \\ \text{container ship} &: \{\dots, 2, 2, 0, 2, 2, \dots\} \\ \text{tanker} &: \{0, 3, 3, \dots\}. \end{aligned} \quad (13)$$

The similarity distance is defined as follows:

$$d_{\text{sim}}^C = \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - C| \quad (14)$$

where

- 1)  $\text{con}_k$ : The confidence score of the  $k$ th component;
- 2)  $\text{cls}_k$ : The category code of the  $k$ th component;
- 3)  $w_k$ : The weight corresponding to the  $k$ th component;
- 4)  $C$ : The number representing the standard component.  $C \in \{0, 1, 2\}$ .

The weight vector is designed such that components closer to the bridge are assigned higher weights. Therefore, the weight vector is defined using an exponential function as follows:

$$\vec{w} = \left( \frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^k}, \dots \right), k = 1, 2, \dots, n \quad (15)$$

where  $k$  represents the position relative to the bridge, and  $n$  denotes the total number of components excluding the bridge. When the number of components approaches infinity, the weight formula ensures convergence to  $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2^i} = 1$ . For container ships, as the bridge is located in the middle, the weights for components on both sides have to be calculated. Thus, the overall weight vector is further multiplied by  $\frac{1}{2}$ .

In addition, the mean distances of the three key components (excluding the bridge) are not uniform. The mean distance for containers is  $\frac{2}{3}$ , while for the other two types, the mean is 1. Consequently, when calculating the similarity distance for container ships, the weight vector is multiplied by  $\frac{3}{2}$ . Combined with the above  $\frac{1}{2}$ , the final scaling factor is  $\frac{3}{4}$ .

Using the bridge as the coordinate origin, the standard component architecture vectors for each ship type are padded to match the length of the recognized component vector by filling missing positions with zeros. The similarity distances  $d_{\text{sim}}^{1,2,3}$  are computed as follows:

$$\begin{aligned} \text{bulk carrier} : d_{\text{sim}}^1 &= \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - 1| \\ \text{container ship} : d_{\text{sim}}^2 &= \frac{3}{4} \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - 2| \\ \text{tanker} : d_{\text{sim}}^3 &= \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - 3|. \end{aligned} \quad (16)$$

The final classification decision is made as follows:

- 1) If all three similarity distances ( $(d_{\text{sim}}^1, d_{\text{sim}}^2, d_{\text{sim}}^3)$ ) are greater than setting threshold  $\text{det}$ , classify the target as *background*. The threshold is the biggest calculation when the nearest component is recognized incorrectly. Through systematic parameter optimization experiments analyzing misidentification rates of critical components, the threshold is empirically set to  $\text{det} = 2$ . If all three distances are larger than  $\text{det}$ , we consider that none of the categories fit the standard component architecture, which is defined as a false alarm.
- 2) Otherwise, choose the ship type corresponding to the smallest similarity distance as *corrected ship class*.

The threshold is set on the basis of minimum tolerance. We establish the similarity distance threshold based on the minimum tolerance for similarity distances, which quantifies the permissible discrepancy between the identified component

architecture and the true component architecture corresponding to the vessel category. Specifically, for a given target category, we allow up to 2 components to be misidentified. According to the definition of similarity distance, components closer to the bridge carry higher weights. Therefore, we calculate the maximum value of the similarity distance threshold using an example where the two closest components are misidentified while all other components are correctly identified. For the three vessel categories in this study, the maximum threshold value is 3. In practical applications, the threshold should be set slightly lower than this maximum value.

#### E. Component Architecture Similarity Distance Promotion

Assume that the cargo hold is encoded as  $a$ , the container is  $b$ , and the oil pipeline is  $c$ . The calculation method of the normalization parameters could be updated as follows: Since the bridge of the container ship is located in the middle, the components on both sides need to be calculated, therefore the overall weight vector should be multiplied by  $\frac{1}{2}$ . Meanwhile, to ensure the uniformity of the mean distances of the three types of key components except the bridge, we calculate the mean value of every category of ships, the mean value of the bulk carrier  $E_1 = \frac{\|a-a\| + \|a-b\| + \|a-c\|}{3}$ , the container ship  $E_2 = \frac{\|b-a\| + \|b-b\| + \|b-c\|}{3}$ , and the tanker  $E_3 = \frac{\|c-a\| + \|c-b\| + \|c-c\|}{3}$ . Therefore, when calculating the similarity distance with the standard bulk carrier, it is necessary to multiply by  $\frac{1}{E_1}$ ; when calculating the similarity distance with the container ship, multiply by  $\frac{1}{E_2}$ , and comprehensively, it is necessary to multiply by  $\frac{1}{2E_2}$ ; the similarity distance with the oil tanker is to multiply by  $\frac{1}{E_3}$ .

The similarity distances  $d_{\text{sim}}^{a,b,c}$  are computed as follows:

$$\begin{aligned} \text{bulk carrier} : d_{\text{sim}}^a &= \frac{1}{E_1} \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - a| \\ \text{container ship} : d_{\text{sim}}^b &= \frac{1}{2E_2} \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - b| \\ \text{tanker} : d_{\text{sim}}^c &= \frac{1}{E_3} \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - c|. \end{aligned} \quad (17)$$

Suppose, the number of target types increases to  $K$ , and the component structure of each ship target is composed of a bridge and another special component. The type of the special component is  $(A_1, A_2, \dots, A_K)$ , and the special component type corresponding to the  $M_{\text{th}}$  ship is  $A_M$ . For targets similar to container ships where special components are distributed on both sides of the bridge, the overall weight should be multiplied by  $\frac{1}{2}$ . At the same time, to ensure the uniformity of the mean distance of key components except the bridge, the mean distance of key components corresponding to each type of target is calculated as  $E_M = \frac{\sum_{j=1}^K \|A_M - A_j\|}{K}$ , and it is estimated that when calculating the similarity distance with the target, it is necessary to multiply by  $\frac{1}{E_M}$ . Multiplying the normalization parameters corresponding to the target types can obtain the final

TABLE III  
HARDWARE CONFIGURATION FOR EXPERIMENTS

| Hardware      | Configuration               |
|---------------|-----------------------------|
| GPU           | NVIDIA Quadro RTX 6000/8000 |
| Video Memory  | 24G                         |
| CPU           | Intel Xeon Gold 6226R       |
| CPU Frequency | 2.9GHz                      |
| Memory        | 503G                        |

TABLE IV  
TRAINING RESULTS OF THE YOLOv8 SHIP RECOGNITION NETWORK

| Ship           | Number | Precision | Recall Rate | mAP @50 | mAP @50-95 |
|----------------|--------|-----------|-------------|---------|------------|
| Bulk Carrier   | 105    | 0.971     | 0.961       | 0.984   | 0.954      |
| Container Ship | 74     | 0.939     | 0.932       | 0.927   | 0.872      |
| Tanker         | 119    | 0.961     | 0.924       | 0.942   | 0.890      |
| Others         | 189    | 0.976     | 0.894       | 0.933   | 0.897      |
| All            | 487    | 0.962     | 0.928       | 0.947   | 0.903      |

normalization parameters as follows:

$$\xi_M = \begin{cases} \frac{1}{E_M}, & \text{if bridge is on the edge} \\ \frac{1}{2E_M}, & \text{if bridge is in the middle.} \end{cases} \quad (18)$$

In the same way, under the assumption that the number of target types increases to  $K$ , the similarity distance formula is encoded as follows:

$$d_{\text{sim}}^M = \xi_M \sum_{k=1}^n \frac{\omega_k}{\text{con}_k} \times |\text{cls}_k - M|. \quad (19)$$

The maximum of similarity distance threshold  $\text{det}$  should also be adjusted as follows:

$$\text{det}_{\text{max}} = \xi_{M_{\text{max}}} \times \frac{3(M-1)}{4} \|A_i - A_j\|_{\text{max}, i, j=1, 2, \dots, K}. \quad (20)$$

#### IV. EXPERIMENTS

##### A. Experimental Setup

Table III lists the hardware configurations used in all the experiments below. The ship target recognition and component recognition network were trained using the two constructed datasets. We validate the “Prerecognition—Fine Recognition—Matching Correction” method with 21 real-scene inshore high-resolution SAR images (higher than 1 m), from ports such as Shanghai Port, Zhoushan Port, Port Elizabeth, and many other port areas domestically and internationally. The experiments demonstrated the method’s effectiveness in improving recognition accuracy, especially in suppressing false alarm rates and false recognition rates in inshore target recognition task.

##### B. Experiments on Recognition Networks

We use the validation set of high-resolution SAR inshore ship recognition dataset to test the trained YOLOv8-p6 ship recognition network, with results presented as Table IV. Also, we use the validation set of high-resolution SAR component

TABLE V  
TRAINING RESULTS OF THE YOLOv8 COMPONENT RECOGNITION NETWORK

| Component    | Number | Precision | Recall Rate | mAP @50 | mAP @50-95 |
|--------------|--------|-----------|-------------|---------|------------|
| Bridge       | 373    | 0.964     | 0.952       | 0.985   | 0.649      |
| Cargo Hold   | 942    | 0.951     | 0.975       | 0.979   | 0.718      |
| Container    | 160    | 0.989     | 0.993       | 0.993   | 0.845      |
| Oil Pipeline | 114    | 0.997     | 0.968       | 0.995   | 0.704      |
| All          | 1589   | 0.966     | 0.974       | 0.985   | 0.729      |

architecture dataset to test the trained YOLOv8-p2 component recognition network, with results presented as Table V. Both tables show that both networks have been trained properly, as each model is suitable for its corresponding task. It is limited for one single YOLOv8 recognition network to achieve high precision in complex inshore scenarios, while it becomes easier to recognize smaller scale components without the interference from strong-scattered background.

##### C. Experiments on YOLO Recognition and Correction Method

We applied inshore SAR images of a size larger than  $10000 \times 10000$ . Using identical training weights, we compared ship recognition results before and after adding the component recognition correction step. As shown in Fig. 11, the prerecognition results are employed to recognize component structure, match standard structure and correct to the final results from left to right. The figure highlights background falsely identified as ships, ship targets falsely recognized in the prerecognition step, as well as targets correctly recognized, which are corrected back to background, to the right category, or remain the initial result after fine recognition and matching correction.

Figs. 12 and 13 are, respectively, the results of YOLO recognition and correction method for two nearshore scenes with a 0.5 m resolution. Both figures indicate that employing the YOLO component recognition and matching correction method can reduce false alarms and false recognitions in single YOLOv8 recognition environments and correct these erroneous recognition results.

As can be seen in Figs. 12 and 13, of all the prerecognition targets in inshore scenes, a large number of background areas are recognized as corresponding targets. After the proposed YOLO recognition and correction method performs component recognition and matching correction on the prerecognition results, it successfully removes some false alarms and corrects some falsely recognized targets. However, there are still some false alarm targets that have not been corrected, because the architectures are too similar from SAR images, with an inference that it may be a scenario with strong interference, such as a shipyard or a cargo loading area nearshore. In Fig. 13, there is a typical shore-side tower crane blocking the sea surface. This may not only prevent the target from being detected and recognized, but also change the characteristics of the ship target itself and its component architecture. Such common inshore devices, including bulk cargo areas, container areas, and tank areas, cause great interference to ship recognition.

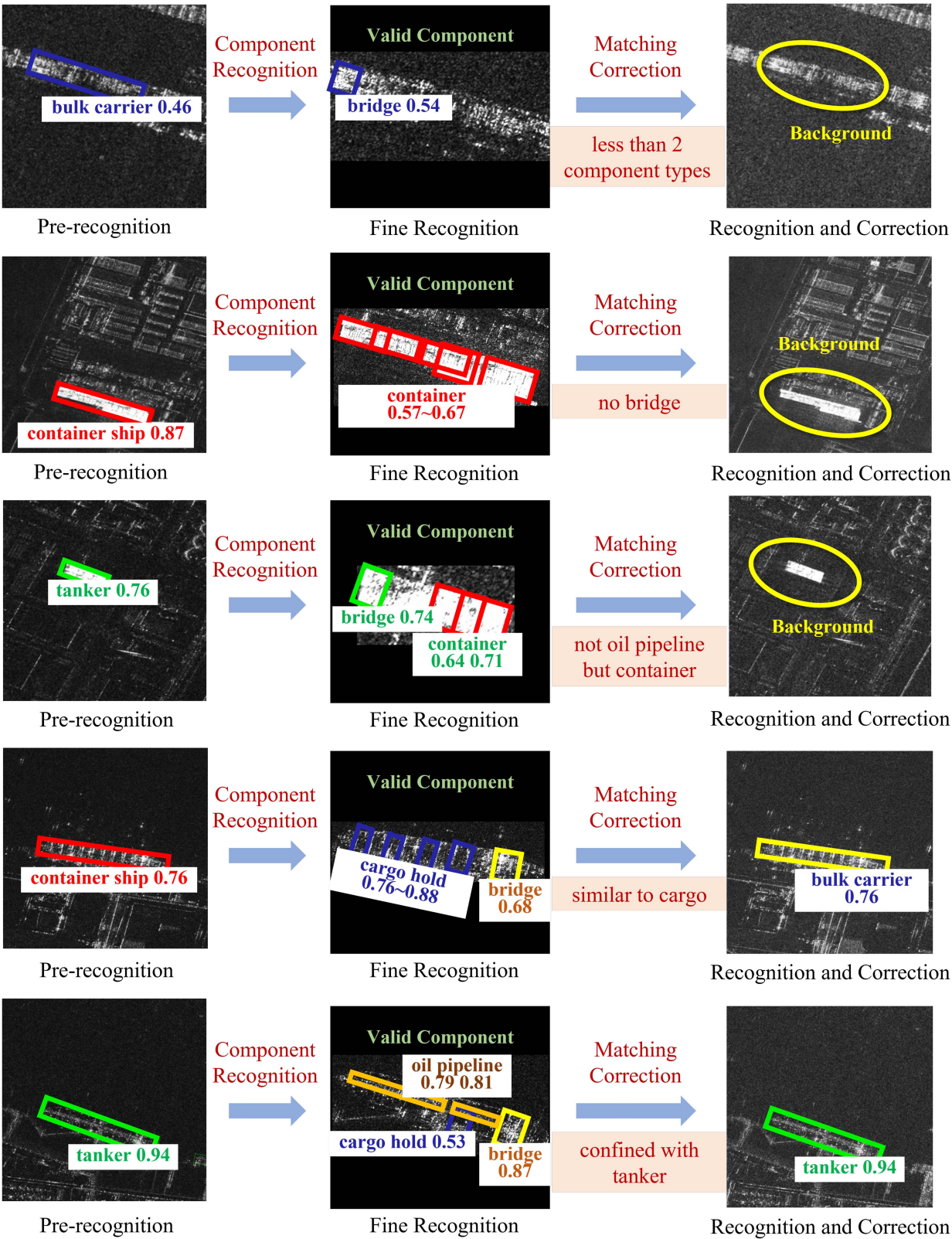


Fig. 11. Comparison results of the YOLO ship recognition network without and with fine recognition and correction method: From left to right are the prerecognition, fine recognition, and recognition correction results of the slices; the first three rows show the cases where the three types of prerecognized targets are judged as false alarms; the fourth row is the correction of the falsely recognized ship targets, and the fifth row is the verification of the correct recognition results.

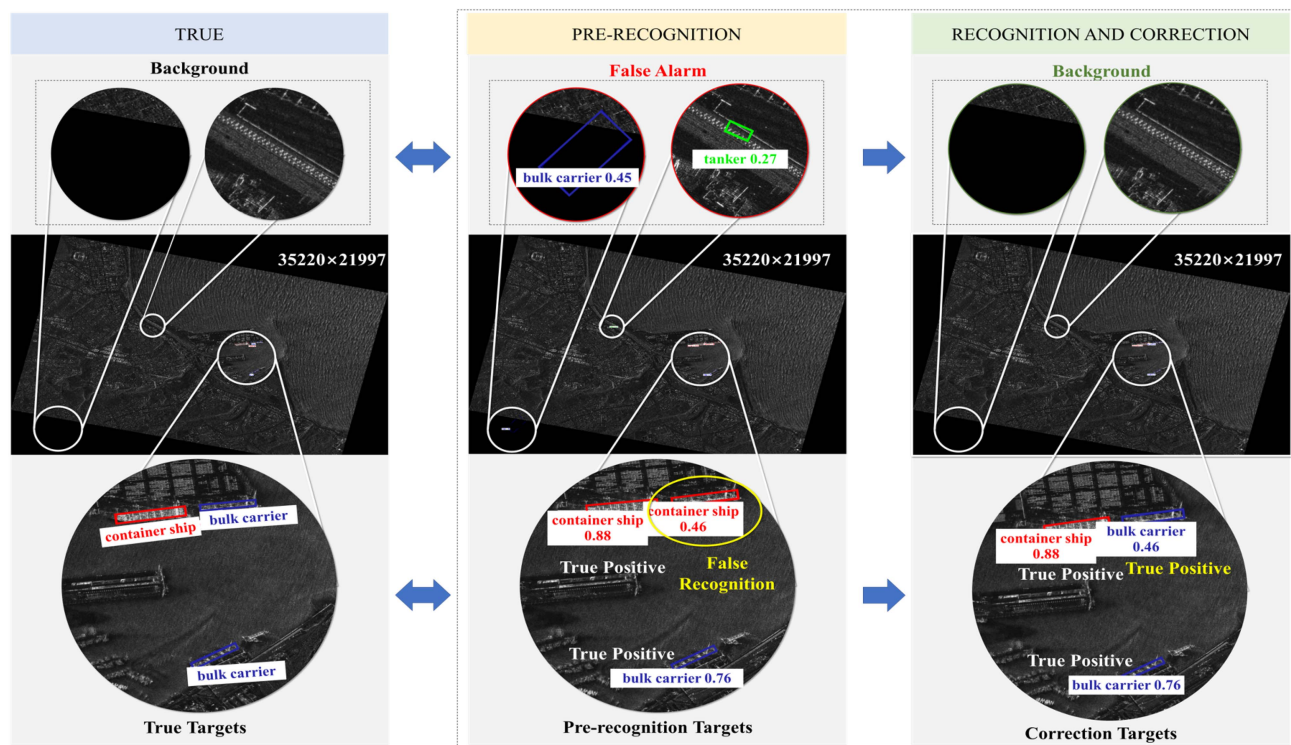


Fig. 12. Results of YOLO ship recognition and correction method on Scene I (Port Elizabeth): The ground truth is presented in the left column, which is compared with the recognition results in the two columns on the right, where the prerecognition results are in the middle column.

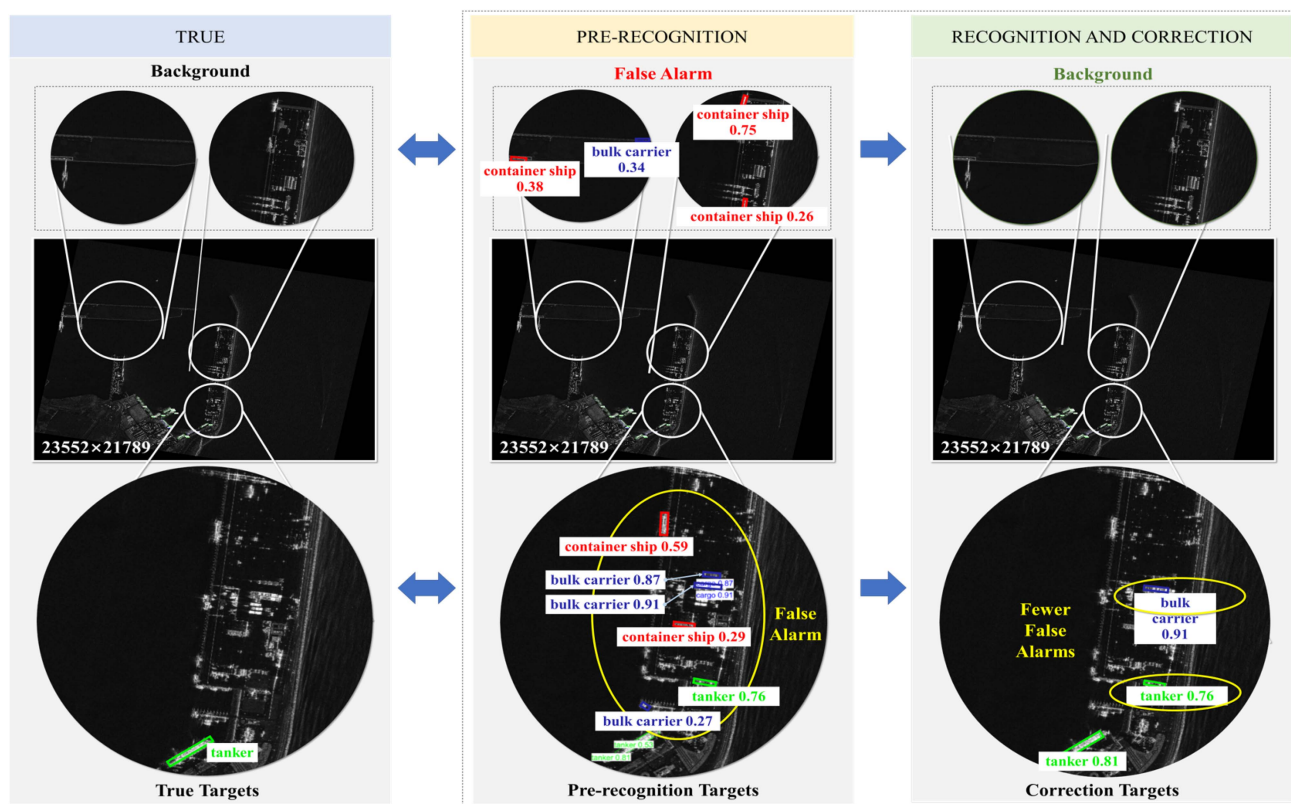


Fig. 13. Results of YOLO ship recognition and correction method on Scene II (Port of Duqm): The ground truth is presented in the left column, which is compared with the recognition results in the two columns on the right, where the prerecognition results are shown in the middle column.

TABLE VI  
COMPARISON OF RECOGNITION FRAMEWORKS

|                | Model           | Total Number | TP  | Recall Rate | FN | Missing Target Rate | M  | False Recognition | FP | False Alarm Rate | Precision   |
|----------------|-----------------|--------------|-----|-------------|----|---------------------|----|-------------------|----|------------------|-------------|
| Bulk Carrier   | YOLOv8(OBB)     | 349          | 232 | 0.66        | 64 | 0.18                | 53 | 0.16              | 38 | 0.12             | 0.88        |
|                | Add correction  | 349          | 300 | <b>0.86</b> | 31 | <b>0.09</b>         | 18 | <b>0.05</b>       | 16 | <b>0.05</b>      | <b>0.95</b> |
|                | YOLOv8-p6       | 349          | 248 | 0.71        | 36 | 0.10                | 65 | 0.18              | 55 | 0.15             | 0.85        |
|                | Proposed method | 349          | 313 | <b>0.90</b> | 22 | <b>0.06</b>         | 14 | <b>0.04</b>       | 12 | <b>0.04</b>      | <b>0.96</b> |
| Container Ship | YOLOv8(OBB)     | 347          | 217 | 0.63        | 60 | 0.17                | 70 | 0.21              | 53 | 0.16             | 0.84        |
|                | Add correction  | 347          | 295 | <b>85</b>   | 21 | <b>0.06</b>         | 31 | <b>0.09</b>       | 17 | <b>0.05</b>      | <b>0.95</b> |
|                | YOLOv8-p6       | 347          | 226 | 0.65        | 44 | 0.13                | 77 | 0.21              | 56 | 0.16             | 0.84        |
|                | Proposed method | 347          | 316 | <b>0.91</b> | 15 | <b>0.04</b>         | 16 | <b>0.05</b>       | 14 | <b>0.04</b>      | <b>0.96</b> |
| Tanker         | YOLOv8(OBB)     | 287          | 201 | 0.70        | 60 | 0.21                | 27 | 0.10              | 51 | 0.18             | 0.82        |
|                | Add correction  | 287          | 244 | <b>0.85</b> | 25 | <b>0.09</b>         | 18 | <b>0.06</b>       | 23 | <b>0.08</b>      | <b>0.92</b> |
|                | YOLOv8-p6       | 287          | 233 | 0.81        | 27 | 0.09                | 27 | 0.09              | 42 | 0.14             | 0.86        |
|                | Proposed method | 287          | 250 | <b>0.87</b> | 19 | <b>0.07</b>         | 15 | <b>0.05</b>       | 11 | 0.04             | <b>0.96</b> |
| All            | Proposed method | 983          | 879 | <b>0.89</b> | 56 | <b>0.06</b>         | 45 | <b>0.05</b>       | 37 | 0.04             | <b>0.96</b> |

The bold entities are the results of our proposed method. Baseline is YOLOv8(OBB). If we simply add the component recognition and matching correction steps based on YOLOv8(OBB), the corresponding bold entities show that post-correction results are better than pre-correction ones. In the same way, the results of Proposed Method are both better than those of baseline and simple YOLOv8-p6.

## V. DISCUSSION

### A. Accuracy Analysis

Next we compared the proposed YOLO recognition and correction method with similar recognition frameworks using the same dataset, identical environment, and training iterations. Table VI lists the recognition results for three types of ship targets on 21 real-scene SAR inshore images, including the true number of ship targets in each category  $Total_{Number}$ , the number of correctly predicted targets TP (true positive), the recall rate (the coverage of the true targets by the method, calculated as  $TP/Total_{Number}$ ), the number of undetected targets FN (false negative), the missing target rate (the proportion of undetected targets, calculated as  $FN/Total_{Number}$ ), the number of false recognitions  $M$ , the false recognition rate [the measurement of the classification ability, calculated as  $M/(TP+FP+M)$ ], the number of false alarms FP (False Positive), the false alarm rate [the proportion of the background in predictions, calculated as  $FP/(TP+FP+M)$ ] and the precision rate [the reliability of the recognition results, calculated as  $TP/(TP+FP+M)$ ] [58], [59].

From the above results, it can be seen that the proposed “Prerecognition—Fine Recognition—Matching Correction” method can effectively recognize all three types of ship targets at a recall rate of 85% and a precision of 90%. Furthermore, the correction stage has been shown to significantly suppress false positives that are attributable to background clutter, which was initially misclassified as “Others.” By exchanging the prerecognition network structure, it can also prove that the YOLO recognition and correction method can successfully reduce the high false alarm rate from a single YOLO recognition network, and correct some of the false recognition results and improve accuracy under the same training conditions.

However, it should be noted that although the YOLO recognition and correction method can effectively reduce false alarms, there may still be cases of incorrect discrimination.

- 1) Prerecognition results are the basis of our proposed method. We control the missed detection rate by choosing YOLOv8-p6 network and setting a low threshold of confidence to keep as many potential ship targets as possible. Some missing targets are “detected” afterward because they are detected but falsely recognized as “others” in

the first place. There are certain circumstances such as the nearshore infrastructures, resulting in severe target occupation and unavoidable missed detections. It should be noted that YOLOv8-p6 is not the only applicable network for ship target recognition. It is easy to replace with many other networks and we could correct based on the obtained prerecognition results.

- 2) The accuracy of component recognition depends on the quality of the image data, specifically whether the details are clear or not. If the SAR images have high interference or poor focusing, it will possibly lead to inaccurate component recognition. Component is a variety of small-scale simple-structure target for recognition and we have to make sure every component remains at least 10 pixels in the images to reflect the distinguishable scattering characteristics. Therefore, we use C-band and X-band SAR images with a resolution of 1 m and 0.5 m for accurate component recognition.

In future study, we plan to integrate component recognition and correction with ship recognition, add more auxiliary discrimination information, and use the guidance of prior knowledge to improve the accuracy of ship target recognition in nearshore port scene. Moreover, we aim to further suppress the false alarm rate, missing target rate, and false recognition rate in port target recognition tasks by adding more support information and designing suitable network models.

### B. Efficiency Analysis

We describe the proposed method as a cascaded “Prerecognition—Fine Recognition—Matching Correction” workflow. While this approach improves recognition precision via YOLO recognition and correction, it inherently sacrifices computational efficiency due to sequential dependencies. Specifically, the pipeline introduces additional overhead from three key operations: Ship slice segmentation, component recognition, and matching correction procedures.

Complexity measurements averaged over multiple SAR image ( $13970 \times 11898$ ) tests demonstrate a critical time distribution pattern: The prerecognition stage dominates computational

TABLE VII  
EFFICIENCY ANALYSIS

| Step                | Time(s)       |
|---------------------|---------------|
| Pre-recognition     | <b>145.87</b> |
| Target slicing      | 2.44          |
| Fine recognition    | 73.36         |
| Matching Correction | 0.26          |
| Total               | <b>221.93</b> |

The bold entites are the time cost in pre-recognition and total algorithm.

costs when processing 69 suspected targets. Subsequent segmentation and correction phases exhibit relatively low overhead since their time consumption is primarily governed by network inference rather than algorithm-level processing.

Empirical results confirm that although these added steps reduce efficiency, they achieve essential benefits including enhanced accuracy alongside suppressed false alarms and misclassifications. The overall time consumption remains within acceptable operational limits as shown in Table VII. To address this efficiency trade-off, future work will integrate component structural priors directly into the recognition network, thereby reducing cascaded complexity while maintaining performance gains.

## VI. CONCLUSION

In this work, a YOLO recognition and correction method has been proposed based on key component structural knowledge to suppress the high false alarm rate and false recognition rate for ship target recognition task, especially in complex inshore SAR images. The method could be described as “Prerecognition—Fine Recognition—Matching Correction.” The prerecognition results for the three ship classes first undergo refinement through cropping and quantitative enhancement. This processed data then enables detailed identification of critical component architecture. Based on three core correction principles, we develop a matching correction algorithm, which evaluates the relationship between the observed component arrangement vector and its corresponding standard structural vector, enabling data-driven refinement of the final recognition outcomes.

Two types of datasets were constructed: one for ship target recognition and one for component architecture recognition. In addition, recognition performance was validated using SAR inshore wide-swath port scene images for comparison. The recognition accuracy of all categories was improved, and the false alarm rate was suppressed down to 10%, confirming the effectiveness of the proposed method. This fine-grained recognition and correction approach improves target recognition accuracy, demonstrating strong engineering applicability. It is particularly suited for complex scene target recognition tasks, providing more accurate and efficient technical support for maritime monitoring and safety surveillance.

Currently, this method still presents some false alarms, misrecognition instances, and suboptimal efficiency stemming from its cascade architecture. Future research will concentrate on developing a component-aware recognition network

structure—shifting from postcorrection adjustments to proactive component guidance—while concurrently expanding discernible ship target categories. This dual approach, enriched by domain-specific recognition expertise, seeks to establish general methodologies applicable across increasingly complex maritime operational environments.

## REFERENCES

- [1] Y. Cheng and L. Yuan, “Current situation of the development of space-borne SAR,” *Electron. Test*, no. 8, pp. 135–136, 2016. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=edscoj&AN=edscoj.dzcs201608079&lang=zh-cn&site=eds-live>
- [2] L. Chunsheng et al., “Current situation and development trend of space-borne SAR technology,” *J. Electron. Inf. Technol.*, vol. 38, no. 1, pp. 229–240, Jan. 2016.
- [3] S. Song, B. Xu, and J. Yang, “SAR target recognition via supervised discriminative dictionary learning and sparse representation of the SAR-hog feature,” *Remote Sens.*, vol. 8, no. 8, 2016, Art. no. 683. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84983786211&doi=10.3390/rs2016080683&partnerID=40&md5=195757048b17e6c5f5cfe8b1aca95551>
- [4] Y. Baozhong, L. Wenjing, B. Yan, and W. Lei, “Target recognition technology based on improved sift algorithm,” *Appl. Sci. Technol.*, vol. 45, no. 5, pp. 76–81, 2018. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=edscoj&AN=edscoj.yykj201805015&lang=en&site=eds-live>
- [5] W. Lu et al., “A method of SAR target recognition based on Gabor filter and local texture feature extraction,” *J. Radars*, vol. 4, no. 6, pp. 658–665, 2015. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=edscoj&AN=edscoj.lctxb201506006&lang=en&site=eds-live>
- [6] Y. Zhang, Z. Lei, H. Yu, and L. Zhuang, “Imbalanced high-resolution SAR ship recognition method based on a lightweight CNN,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4014405.
- [7] M. Ma, J. Chen, W. Liu, and W. Yang, “Ship classification and detection based on CNN using GF-3 SAR images,” *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 2043. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85058903243&doi=10.3390/rs10122043&partnerID=40&md5=f54c77e862611abfd7e7e748419f5147>
- [8] T.-D. Wu, Y. Yen, J. H. Wang, R. J. Huang, H.-W. Lee, and H.-F. Wang, “Automatic target recognition in SAR images based on a combination of CNN and SVM,” in *Proc. Int. Workshop Electromagn.: Appl. Student Innov. Competition*, 2020, pp. 1–2.
- [9] Y. Li, M. Cheng, X. Peng, B. Zhuo, and F. Li, “Ship detection and recognition combining one-dimensional range profile with SAR image,” *J. Eng.*, vol. 2019, no. 19, pp. 6252–6254, 2019. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/joe.2019.0191>
- [10] P. Selvam et al., “YOLO-SAIL: Attention-enhanced YOLOv5 with optimized BI-FPN for ship target detection in SAR images,” *IEEE Access*, vol. 13, pp. 29523–29540, 2025.
- [11] Q. Wang, C. Tong, X. Li, Y. Wang, Z. Wang, and T. Wang, “Composite electromagnetic scattering and high-resolution SAR imaging of multiple targets above rough surface,” *Remote Sens.*, vol. 14, no. 12, 2022, Art. no. 2910. [Online]. Available: <https://www.mdpi.com/2072-4292/14/12/2910>
- [12] Z. Hou, Z. Cui, Z. Cao, and N. Liu, “An integrated method of ship detection and recognition in SAR images based on deep learning,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1225–1228.
- [13] F. M. Khan, E. Baccour, A. Erbad, and M. Hamdi, “Adaptive resnet architecture for distributed inference in resource-constrained IoT systems,” in *Proc. Int. Wireless Commun. Mobile Comput.*, 2023, pp. 1543–1549.
- [14] Z. Zhang, Y. Zhang, H. Li, and Z. Liu, “Resnet optimization model for high-resolution range profile target recognition,” in *Proc. 4th Int. Conf. Electron. Inf. Eng. Comput. Sci.*, 2024, pp. 389–392.
- [15] C. Zhang et al., “A fine-grained car recognition method based on a lightweight attention network and regularized fine-tuning,” *Electronics*, vol. 14, no. 1, 2025, Art. no. 211. [Online]. Available: <https://www.proquest.com/scholarly-journals/fine-grained-car-recognition-method-based-on/docview/3153798638/se-2>
- [16] T. Miao et al., “An improved lightweight retinanet for ship detection in SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4667–4679, 2022.

- [17] C. Yu, Y. Zhai, H. Huang, Q. Wang, and W. Zhou, "Capsule broad learning system network for robust synthetic aperture radar automatic target recognition with small samples," *Remote Sens.*, vol. 16, no. 9, 2024, Art. no. 1526.
- [18] C. Peng, Y. Song, and L. Yang, "Detection of weak ship radiation characteristic signal based on wavelet and chaos array," in *Proc. IEEE 6th Inf. Technol. Mechatron. Eng. Conf.*, vol. 6, 2022, pp. 139–150.
- [19] D. Xiang, W. Li, X. Sun, H. Wang, and Y. Su, "Sidelobe suppression for high-resolution SAR imagery based on spectral reshaping and feature statistical difference," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5211314.
- [20] Z. Luo, Z. Fang, and S. Ding, "Ship target classification technology fused with domain knowledge," in *Proc. Int. Conf. Intell. Comput., Automat. Appl.*, 2021, pp. 492–495.
- [21] N. Li, D. Xiang, X. Sun, C. Hu, and Y. Su, "Multiscale adaptive polar image superpixel generation based on local iterative clustering and polarimetric scattering features," *ISPRS J. Photogrammetry Remote Sens.*, vol. 220, pp. 307–322, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271624004817>
- [22] Z. Tian, W. Wang, K. Zhou, X. Song, Y. Shen, and S. Liu, "Weighted pseudo-labels and bounding boxes for semisupervised SAR target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5193–5203, 2024.
- [23] J. Deng, W. Wang, H. Zhang, T. Zhang, and J. Zhang, "PolSAR ship detection based on superpixel-level contrast enhancement," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 4008805.
- [24] Z. Huang, M. Datcu, Z. Pan, and B. Lei, "Deep SAR-net: Learning objects from signals," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 179–193, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271620300162>
- [25] L. J. Moore, B. D. Rigling, and R. P. Penno, "Characterization of phase information of synthetic aperture radar imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 676–688, Apr. 2019.
- [26] P. Ni, G. Xu, Z. Zhong, J. Chen, and W. Hong, "SAR target recognition using complex manifold multiscale feature fusion network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2024, pp. 3532–3535.
- [27] K. Li, M. Zhang, M. Xu, R. Tang, L. Wang, and H. Wang, "Ship detection in SAR images based on feature enhancement swin transformer and adjacent feature fusion," *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 3186. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85133725797&doi=10.3390/rs14133186&partnerID=40&md5=b1ef6e84f83a8f5a1cbe27b96101fece>
- [28] C. Wang et al., "SAR ship target recognition via multiscale feature attention and adaptive-weighted classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4003905.
- [29] D. Zhang, J. Liu, W. Heng, K. Ren, and J. Song, "Transfer learning with convolutional neural networks for SAR ship recognition," in *Proc. IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 322, 2018, Art. no. 07200103.
- [30] K. Balavani, D. Sriram, M. B. Shankar, and D. S. Charan, "An optimized plant disease classification system based on resnet-50 architecture and transfer learning," in *Proc. 4th Int. Conf. Emerg. Technol.*, 2023, pp. 1–5.
- [31] E. D. and N. P. G. Bhavani, "An effective DNN based resnet approach for satellite image classification," in *Proc. 4th Int. Conf. Smart Electron. Commun.*, 2023, pp. 1055–1062.
- [32] Y. Shang et al., "SA<sup>2</sup> net: Ship augmented attention network for ship recognition in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10036–10049, 2023.
- [33] Z. Xing, G. Wei, D. Bo, Z. Hong, and Z. Zhu, "Ship HRRP target recognition method based on transformer and convolutional attention network," in *Proc. 15th Int. Conf. Signal Process. Syst.*, vol. 13091, 2024, pp. 558–567.
- [34] H. Zeng et al., "An inception-textCNN model for ship target detection in SAR range-compressed domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 3501305.
- [35] Y. Li, J. Chen, M. Ke, L. Li, Z. Ding, and Y. Wang, "Small targets recognition in SAR ship image based on improved SSD," in *IEEE Int. Conf. Signal, Inf. Data Process.*, Chongqing, China, 2019, pp. 1–6, doi: [10.1109/ICSIDP47821.2019.9173411](https://doi.org/10.1109/ICSIDP47821.2019.9173411).
- [36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [37] M. Yaseen, "What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector," 2024, [arxiv:2409.07813](https://arxiv.org/abs/2409.07813).
- [38] Y. Gong, Z. Chen, W. Deng, J. Tan, and Y. Li, "Real-Time Long-Distance Ship Detection Architecture Based on YOLOv8," *IEEE Access*, vol. 12, pp. 116086–116104, 2024, doi: [10.1109/ACCESS.2024.3445154](https://doi.org/10.1109/ACCESS.2024.3445154).
- [39] X. Tang, J. Sa, and J. Zhou, "BP-YOLOv8: Rotating target detection in synthetic aperture radar ship images based on improved YOLOv8," *J. Appl. Remote Sens.*, vol. 18, no. 3, Jul. 2024, Art. no. 034509.
- [40] B. Xiong, Z. Sun, J. Wang, X. Leng, and K. Ji, "A lightweight model for ship detection and recognition in complex-scene SAR images," *Remote Sens.*, vol. 14, no. 23, 2022, Art. no. 6053.
- [41] C. Y. Pui, S. Ghio, B. Ng, E. Giusti, L. Rosenberg, and M. Martorella, "Robust 3D ISAR ship classification," in *Proc. IEEE Radar Conf.*, vol. 23, 2023, pp. 1–6.
- [42] X. Yang, F. Bi, Y. Yu, and L. Chen, "An effective false-alarm removal method based on OC-SVM for SAR ship detection," in *Proc. IET Int. Radar Conf.*, 2015, pp. 1–4.
- [43] R. Yang, G. Wang, Z. Pan, H. Lu, H. Zhang, and X. Jia, "A novel false alarm suppression method for CNN-based SAR ship detector," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1401–1405, Apr. 2021.
- [44] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, Sep. 2021, Art. no. 3690.
- [45] X. Zheng, Y. Feng, H. Shi, B. Zhang, and L. Chen, "Lightweight convolutional neural network for false alarm elimination in SAR ship detection," in *Proc. IET Int. Radar Conf.*, vol. 2020, 2020, pp. 287–291.
- [46] X. Xiong, X. Wang, J. Zhang, B. Huang, and R. Du, "TCUNet: A lightweight dual-branch parallel network for sea-land segmentation in remote sensing images," *Remote Sens.*, vol. 15, no. 18, 2023, Art. no. 4413. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=172418748&lang=zh-cn&site=eds-live>
- [47] G. Wei, J. Xu, Q. Chong, and J. Huang, "FMPNet: A fuzzy-embedded multi-scale prototype network for sea-land segmentation of remote sensing images," *Eur. J. Remote Sens.*, vol. 57, no. 1, Dec. 2024, Art. no. 2343531.
- [48] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," 2024, [arxiv:2305.09972](https://arxiv.org/abs/2305.09972).
- [49] S. Cheng, Z. Wang, S. Liu, Y. Han, P. Sun, and J. Li, "Attention-based lightweight YOLOv8 underwater target recognition algorithm," *Sensors*, vol. 24, no. 23, Dec. 2024, Art. no. 7640.
- [50] S. Khokhar and D. Kedia, "Integrating YOLOv8 and CSPbottleneck based CNN for enhanced license plate character recognition," *J. Real-Time Image Process.*, vol. 21, no. 5, 2024. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=edsj&AN=edsj.21ACBBF2&lang=zh-cn&site=eds-live>
- [51] T. Cheng, J. Li, J. Luo, and Z. Li, "Improved YOLOv8 for complex environmental fish detection," in *Proc. IEEE 7th Adv. Inf. Technol., Electron. Automat. Control Conf.*, 2024, pp. 797–801.
- [52] M. Biswas and R. Hoque, "Construction site risk reduction via YOLOv8: Detection of PPE, masks, and heavy vehicles," in *Proc. IEEE Int. Conf. Comput., Appl. Syst.*, 2024, pp. 1–6.
- [53] R. Min, H. Lan, Z. Cao, and Z. Cui, "A gradually distilled CNN for SAR target recognition," *IEEE Access*, vol. 7, pp. 42190–42200, 2019.
- [54] Y. Pan et al., "Fine-grained recognition framework for ship targets in SAR images based on component architecture," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2024, pp. 1–5.
- [55] R. Jin, J. Cheng, W. Wang, D. Xiang, C. Jiang, and J. Zhang, "A method of SAR target data augmentation based on component-level perturbation," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2024, pp. 1–5.
- [56] Y. Li, R. Min, J. Li, and Y. Pi, "Terahertz circular SAR imaging algorithm based on the extraction of scattering characteristics of target structures," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 3504905.
- [57] Y. Li and L. Du, "Design of the physically interpretable SAR target recognition network combined with electromagnetic scattering characteristics," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 4988–4991.
- [58] L. Wang, Z. Liu, and Z. Zhang, "Scattering and optical cross-modal attention distillation framework for SAR target recognition," *IEEE Sensors J.*, vol. 25, no. 2, pp. 3126–3137, Jan. 2025.
- [59] S. Zhao, X. Li, J. Ren, L. Wang, and Z. Huang, "Interpretable SAR target recognition based on white box neural network," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2024, pp. 1–5.



**Yongchen Pan** (Graduate Student Member, IEEE) was born in 2000. She received the B.S. degree in electronics and information engineering as well as a second major in English in 2021 from Beihang University, Beijing, China, where she is currently working toward the Ph.D. degree in signal and information processing with School of Electronics and Information Engineering.

Her research interests include SAR target recognition and scene information mining on the application of artificial intelligence.



**Hongcheng Zeng** (Member, IEEE) was born in 1989. He received the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2016.

From 2016 to 2019, he held a postdoctoral position with the School of Electronics and Information Engineering, Beihang University, where he has been an Assistant Professor since August 2019. From 2017 to 2018, he was a Visiting Researcher with the School of Mathematics and Statistics, The University of Sheffield, Sheffield, U.K. Since 2022, he has been an

Associate Professor with the School of Electronics and Information Engineering, Beihang University. He has authored or coauthored more than 40 journal and conference publications. His research interests include SAR target detection, high-resolution spaceborne SAR image formation, and passive radar signal processing.



**Yingjie Zhou** was born in 2003. She received the B.S. degree in communication engineering in 2025 from Beihang University, Beijing, China, where she is currently working toward the M.S. degree in signal and information processing with the School of Electronics and Information Engineering.

Her research interests include developing artificial intelligence methodologies for spaceborne SAR target recognition processing.



**Wei Yang** (Member, IEEE) was born in 1983. He received the M.S. and Ph.D. degrees in signal and information processing from Beihang University, Beijing, China, in 2008 and 2011, respectively.

From 2011 to 2013, he held a postdoctoral position with the School of Electronics and Information Engineering, Beihang University, where he has been a Lecturer with the School of Electronics and Information Engineering since 2013. From 2016 to 2017, he was a Visiting Researcher with the Department of Electronic and Electrical Engineering, The University

of Sheffield, Sheffield, U.K. He has been an Associate Professor with the School of Electronics and Information Engineering, BUAA, since 2018, and a Professor since 2023. He has authored or coauthored more than 60 journals and conference publications. His research interests include moving target detection, high-resolution spaceborne synthetic aperture radar (SAR) image formation, SAR image quality improvement, and 3-D imaging.



**Yamin Wang** (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from Beihang University, Beijing, China, in 2016 and 2022, respectively.

From 2022 to 2024, she was a Postdoctoral Fellow with the School of Electronics and Information Engineering, Beihang University, where she has been a Lecturer from 2024. Her research interest includes synthetic aperture radar (SAR), video SAR, ground moving target indication (GMTI), and SAR signal processing.



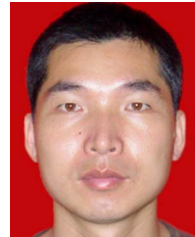
**Xinwei An** (Graduate Student Member, IEEE) was born in 2001. He received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2022. He is currently working toward the Ph.D. degree in signal and information processing with the School of Electronic Information Engineering, Beihang University, Beijing, China.

His research interests include spaceborne synthetic aperture radar image quality improvement and intelligent interpretation.



**Can Su** (Graduate Student Member, IEEE) was born in 1998. He received the undergraduate degree in 2020. He is currently working toward the doctoral degree in signal and information processing with the School of Electronic Information Engineering, Beihang University, Beijing, China.

His research interests include remote sensing intelligent interpretation algorithms based on AI technology, specifically in intelligent target detection, target recognition, and target behavior prediction analysis of spaceborne SAR data.



**Wei Xiong** received the B.S., M.S., and Ph.D. degrees from Naval Aviation University, Yantai, China, in 1998, 2001, and 2005, respectively.

From 2007 to 2009, he was a Postdoctoral Researcher with the Department of Electronic Information Engineering, Tsinghua University, Tsinghua, Beijing. He is currently a Full Professor with the Naval Aviation University, Yantai, China, where he teaches random signal processing and information fusion. He is one of the Founders and the Directors with the Research Institute of Information Fusion,

Naval Aviation University. His research interests include pattern recognition, remote sensing, and multisensor information fusion.

Dr. Xiong is the Member and the Director General with Information Fusion Branch, Chinese Society of Aeronautics and Astronautics.



**Wei Liu** (Senior Member, IEEE) received the B.Sc. degree in space physics, and the LL.B. degree in intellectual property law from Peking University, Beijing, China, in 1996 and 1997, respectively, the M.Phil. degree from the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2001, and the Ph.D. degree from the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., in 2003.

He was a Postdoctoral first with the University of Southampton and later with Imperial College London, London, U.K. From 2005 to 2023, he was a Lecturer and a Senior Lecturer with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K. From 2023 to 2024, he was a Reader with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London. Since 2024, he has been a Professor with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong. He has authored or coauthored more than 430 journal articles and conference papers, six book chapters, and two research monographs titled "Wideband Beamforming: Concepts and Techniques" (Wiley, 2010) and "Low-Cost Smart Antennas" (Wiley, 2019), respectively. His research interests include a wide range of topics in signal processing, with a focus on sensor array signal processing and its various applications, such as robotics and autonomous systems, human-computer interface, radar, sonar, and wireless communications.

Dr. Liu is a Member of the Applied Signal Processing Systems Technical Committee of the IEEE Signal Processing Society (SPS) for the term 2023–2025. He was the Chair of the Sensor Array and Multichannel Signal Processing Technical Committee of the IEEE SPS from 2021 to 2022, the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society from 2022 to 2024, and is currently the Chair of the IEEE SPS Education Board of its Educational Conference Program Committee for the term 2024–2026. He was a Member of the Sensor Array and Multichannel Signal Processing Technical Committee of the IEEE SPS from 2021 to 2022, the IEEE SPS Technical Directions Board from 2021 to 2022, and the IEEE SPS Conference Board and its Executive Subcommittee from 2022 to 2023. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE ACCESS, and *Journal of the Franklin Institute*. He is currently an Associate Editor of IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS and an Executive Associate Editor-in-Chief of the *Frontiers of Information Technology and Electronic Engineering*. He was an IEEE Distinguished Lecturer for the Aerospace and Electronic Systems Society, from 2023 to 2024.