IceRegionShip: Optical Remote Sensing Dataset for Ship Detection in Ice-Infested Waters

Peilin Wang, Bingxin Liu, *Member, IEEE*, Ying Li, Peng Chen, and Peng Liu

***Abstract*—As shipping routes and resource exploration move toward high-latitude oceans, sea ice becomes a major threat to the safety of ship navigation, posing significant challenges to the shipping industry and offshore resource development. Continuous development of satellite remote sensing and deep learning has made large-scale and wide-ranging ship detection (SD) possible, which is of great significance for ship safety. However, existing ship datasets used for deep learning only include ship images in open waters (OW), such as ports and inland rivers. Currently, remote sensing datasets suitable for SD in ice-infested waters (IIW) are lacking. SD in IIW is more difficult than SD in OW because of complex background interference from sea ice. Thus, it is infeasible to directly use the features of ships in OW for SD in IIW. Herein, we propose a remote sensing SD dataset called IceRegionShip, which includes subdatasets IceRegionShip–red, green and blue (RGB) and IceRegionShip–ice region ship index (IRSI). IceRegionShip–IRSI consists of low-resolution images processed with IRSI. IceRegionShip–RGB and IceRegionShip–IRSI contain 11 436 and 9073 ship instances, respectively. IRSI was proposed to address false alarms caused by ice interference. To the best of our knowledge, this is the first dataset designed specifically for SD in IIW. In addition, the dataset was evaluated using several advanced detection algorithms, providing a benchmark for SD in IIW and demonstrating the effectiveness of IRSI for SD in low-resolution optical remote sensing images.**

***Index Terms*—Deep learning, ice-infested waters (IIW), remote sensing dataset, ship detection (SD).**

I. Introduction

MARITIME transportation and the utilization of ocean resources are integral to human economic activity. Thus, rational ship regulation holds significant value for ensuring stable economic development. Advances in remote sensing technology have enabled obtaining abundant ship image data, thereby becoming an indispensable ship data source. Ship detection (SD) techniques based on remote sensing imagery have broad applications, particularly in port and inland waterway scenarios.

Manuscript received 14 July 2023; revised 23 September 2023; accepted 14 November 2023. Date of publication 22 November 2023; date of current version 6 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 52271359, in part by Zhejiang Provincial Basic Public Welfare Research Project under Grant LGF21E090006, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132023018. *(Corresponding authors: Bingxin Liu.)*

The authors are with the Navigation College and the Key Laboratory of Transport Industry of Intelligent Water Transportation, Dalian Maritime University of China, Dalian 116026, China (e-mail: wpl18237976678@dlmu.edu.cn; gisbingxin@dlmu.edu.cn; yldmu@dlmu.edu.cn; chenpeng@dlmu.edu.cn; liupeng@dlmu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3335294

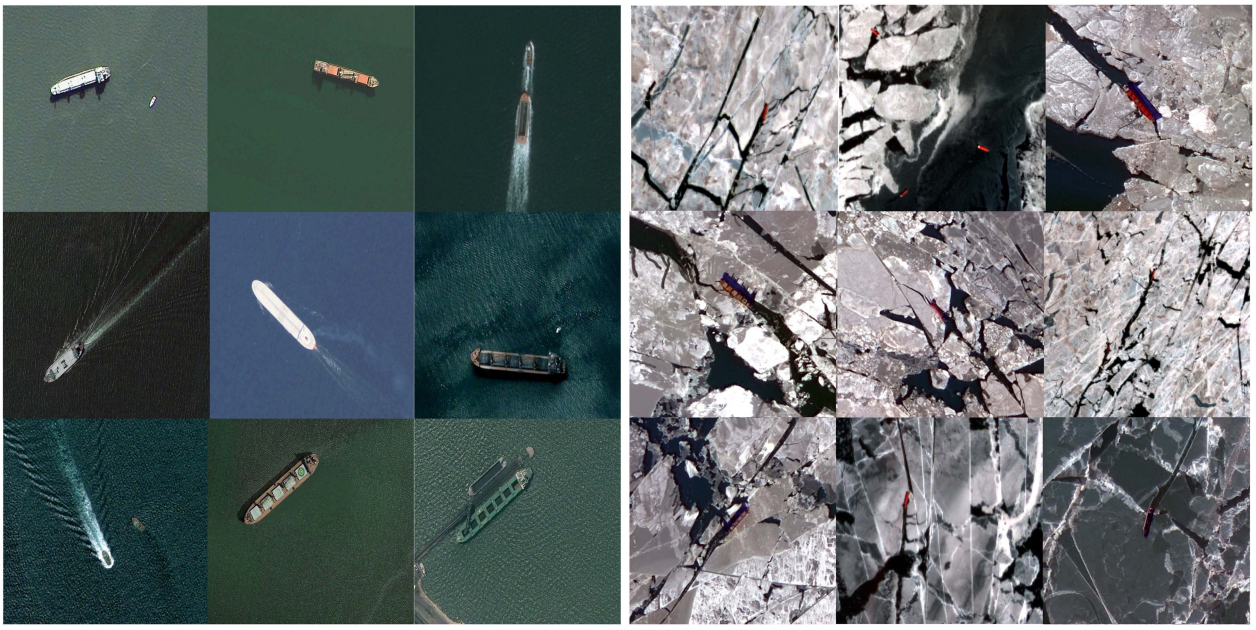
However, with global warming, maritime routes and oil and gas exploration are shifting toward high latitudes. Sea ice presents significant challenges for the development of ocean resources. For instance, the besetting of a ship in sea ice considerably threatens ship safety [1]. The infrastructure along high-latitude routes needs to be improved, necessitating the provision of navigational assistance services. Thus, SD in ice-infested waters (IIW) based on remote sensing imagery must be investigated to provide information in ice-prone environments, which is vital for maritime search and rescue operations, ship safety, high-latitude route development, and the protection of marine ecological environments.

Existing SD methods can be primarily categorized into the following types [2]:

1) threshold-based methods,
2) salient-based method,
3) shape and texture–feature-based methods,
4) transform-domain methods,
5) anomaly detection methods, and
6) deep-learning methods.

Over the past decade, deep-learning-based approaches have witnessed rapid development of computer vision. The use of convolutional neural networks [3] or visual transformer models [4] for the adaptive extraction of visual features instead of handcrafted feature design has improved the performance and generalizability of the methods. However, it is essential to note that all these new methods require large-scale data for training. Although certain progress has been attained in semisupervised [5] and zero-shot methods [6], creating large and high-quality datasets specific to the corresponding domain remains essential for deep-learning models to achieve higher accuracy and adaptability in real-world applications.

In recent years, numerous studies have released datasets to address object detection challenges in remote sensing. DOTAv2.0 [7], an extension of DOTAv1.0 [8], expands the dataset size to facilitate the training of more state-of-the-art baselines. Similarly, FAIR1M [9] not only increases the dataset size but also supplements it with additional categories. In the field of SD, a considerable number of datasets have also emerged. Liu et al. [10] introduced the first optical remote sensing SD dataset, HRSC2016, which has become a commonly used benchmark. FSGD [11] and DOSR [12] are extensions of HRSC2016, and ShipRSImageNet [13] is currently the largest optical remote sensing SD dataset, containing 17 573 ship instances. Through investigation, it was observed that the ship targets in existing optical remote sensing datasets are predominantly concentrated



(a) (b)

Fig. 1. Ship instance examples. (a) HRSC 2016. (b) IceRegionShip–RGB.

in open waters (OW), such as ports, inland rivers, and a small number of open sea areas.

In contrast to these scenes, IIW images exhibit greater complexity. Fragmented ice formed by ice fractures interferes with ship object extraction in terms of texture, shape, and other features, as illustrated in Fig. 1(b). When considering the use of deep-learning methods for SD in IIW, the existing datasets suffer from the following limitations.

1) Ship targets are primarily located in ice-free areas, resulting in a scarcity of ship instances operating in IIW. For instance, in the HRSC dataset, 1680 images capture OW but lack areas containing ice.
2) The existing datasets for fine-grained ship recognition, such as FGSD, mostly consist of images with high spatial resolutions ranging from 0.12 to 1.93 m. High-resolution remote sensing images are expensive and limited in coverage, making it difficult to conduct extensive ship searches over large areas. This limitation hinders the efficient identification of ship targets in ice-covered environments.
3) A substantial portion of the images in datasets, such as HRSC2016, FGSD, and DOSR, is derived from other datasets or Google Earth, leading to notable disparities from real-world applications.

In response to the limitations of existing datasets, herein, a novel optical remote sensing dataset called IceRegionShip is proposed specifically for SD in IIW. It comprises multiresolution optical remote sensing images obtained from various satellite platforms, including GF-1, Landsat-8, and Sentinel-2. To the best of our knowledge, this dataset is the first large-scale optical remote sensing dataset for SD in IIW, indicating its significance for relevant research.

Conventional optical remote sensing systems typically have sensors that capture multiple spectra, each offering specific information. The use of multiband remote sensing data for band calculations allows for emphasizing the desired features and properties of the target objects, thereby enhancing the accuracy and reliability of target extraction. In this article, the ice region ship index (IRSI) was proposed for SD in IIW based on the analysis of the spectral characteristics of ship decks, sea ice, and seawater. Incorporating IRSI into the image preprocessing stage yielded a subdataset named IceRegionShip–IRSI. Experimental results demonstrate that the preprocessing images with IRSI aids the discrimination between ships and sea ice, particularly in low-spatial-resolution images. This approach improves the detection performance of models and effectively addresses the issue of false alarms caused by gaps in sea ice.

Now, we summarize the main contributions of this article as follows.

1) The construction of the first optical remote sensing dataset, IceRegionShip, designed specifically for SD in IIW. It includes 11 436 ship instances, addressing the limitations of existing datasets by providing samples from water areas with ice coverage.
2) The proposal of the IRSI for SD in IIW. The IceRegionShip–IRSI subdataset was created for detecting ships in low-resolution images. The IRSI reduces the effect of sea ice on SD, enhances ship outlines, and effectively reduces the number of false alarm targets.
3) Testing of the IceRegionShip dataset using multiple benchmark models. The results validate the advantages of IceRegionShip–IRSI for low resolutions.

TABLE I
COMPARISON OF SHIP INSTANCES IN PUBLIC OPTICAL REMOTE SENSING OBJECT DATASETS

Dataset	Annotation	Images total	Ship categories	Ship instances	Year
NWPU VHR-10 [14]	HBB	800	1	302	2014
DOTA-v1.0 [8]	OBB	2806	1	52 516	2017
DIOR [15]	HBB	23 463	1	64000	2019
HRRSD [16]	HBB	21 761	1	3886	2019
XView [17]	HBB	1413	9	5672	2018
DOTA-v2.0 [7]	OBB	11 268	1	251 883	2021
FAIR1M [9]	OBB	15 266	9	≈115 000	2021
SODA-A [18]	OBB	2510	1	61 916	2022

Now, we summarize the main contributions of this article as follows.

II. RELATED WORK

A. Optical Remote Sensing Datasets

We surveyed popular optical remote sensing datasets containing ship targets, such as NWPU VHR-10 [14], DOTAv1.0 [8], and DIOR [15]. Table I presents information on the annotation method, number of images, number of ship instance categories, number of ship instances, and year of release for each dataset. By investigating the spatial distribution of ships in existing datasets, we found that most ship instances are concentrated in OW, such as inland rivers, ports, and open sea areas; moreover, sea surface images and ship instances in IIW are lacking in these datasets.

B. Optical Remote Sensing Ship Datasets

The existing optical remote sensing ship datasets are limited in quantity—the representative ones include HESC2016, FGSD, DOSR, and ShipRSImageNet. However, these datasets are mainly intended for research on fine-grained ship recognition and thus lack ship instances in complex environments, particularly in IIW.

1) HRSC2016: The HRSC2016 dataset comprises 1070 images and 2976 ship instances obtained from Google Earth [10]. It is the earliest large-scale and multiclass optical remote sensing dataset designed explicitly for SD. It has become one of the commonly used benchmarks in the field of SD [13]. HRSC2016 initially employed a three-stage hierarchical classification system, dividing ships into 25 categories, which greatly influenced subsequent datasets.

2) FGSD: In 2020, Chen et al. [11] collected images of ships from multiple large ports using Google Earth; they built upon the HRSC2016 dataset and introduced a new dataset called FGSD. FGSD comprises 2612 images and 5634 ship instances, with spatial resolutions ranging from 0.12 to 1.93 m. Each ship instance is annotated with both horizontal bounding boxes (HBBs) and oriented bounding boxes. FGSD not only subdivides ships into 43 categories but also adds a ship dock category, with an average of 131 instances per category [13].

3) DOSR: To address the limitations of HRSC2016, such as the need for annotations for small ships and incorrect ship category labels, Han et al. [12] constructed a new publicly available dataset called DOSR. Similar to the datasets mentioned above, DOSR consists of images from Google Earth and includes 1066 optical remote sensing images and 6127 ship instances. The ship types in DOSR are categorized into 20 classes. Additionally, DOSR classifies the image scene into 832 coastal scenes (e.g., densely populated ports) and 234 offshore scenes (consisting solely of the sea surface).

4) ShipRSImageNet: Zhang et al. [13] released a large-scale optical remote sensing image dataset, which included 3425 images with 17 573 ship instances containing ship targets from multiple sources, such as XView [17], FGSD, and GF-2. Influenced by HRSC2016, ShipRSImageNet employs a hierarchical classification system, which divides ship types into four levels: class, category, subcategory, and type, resulting in 49 ship categories. All instances have annotations for bounding boxes, oriented bounding boxes, and polygons. ShipRSImageNet also considers weather conditions, such as cloudiness, sunlight reflected from the ocean surface, and low lighting to make the dataset more applicable to real-world scenarios. Compared to HRSC2016 and FGSD, ShipRSImageNet is larger and offers a more diverse set of ship categories, further promoting the advancement of fine-grained recognition of ship targets and related applications.

III. CONSTRUCTION OF ICE REGION SHIP DETECTION ICEREGIONSHIP

Our dataset, named IceRegionShip, consists of two subdatasets: IceRegionShip–red, green and blue (RGB), obtained by merging RGB images from three spectral bands, and IceRegionShip–IRSI, preprocessed using the IRSI. This section elucidates the construction of the IceRegionShip dataset. Fig. 2 exhibits the entire production workflow.

A. Images Acquisition

Different satellites have varying revisit times and imaging characteristics. Compared to a single satellite data source, combining multiple satellite data sources enhances the diversity of

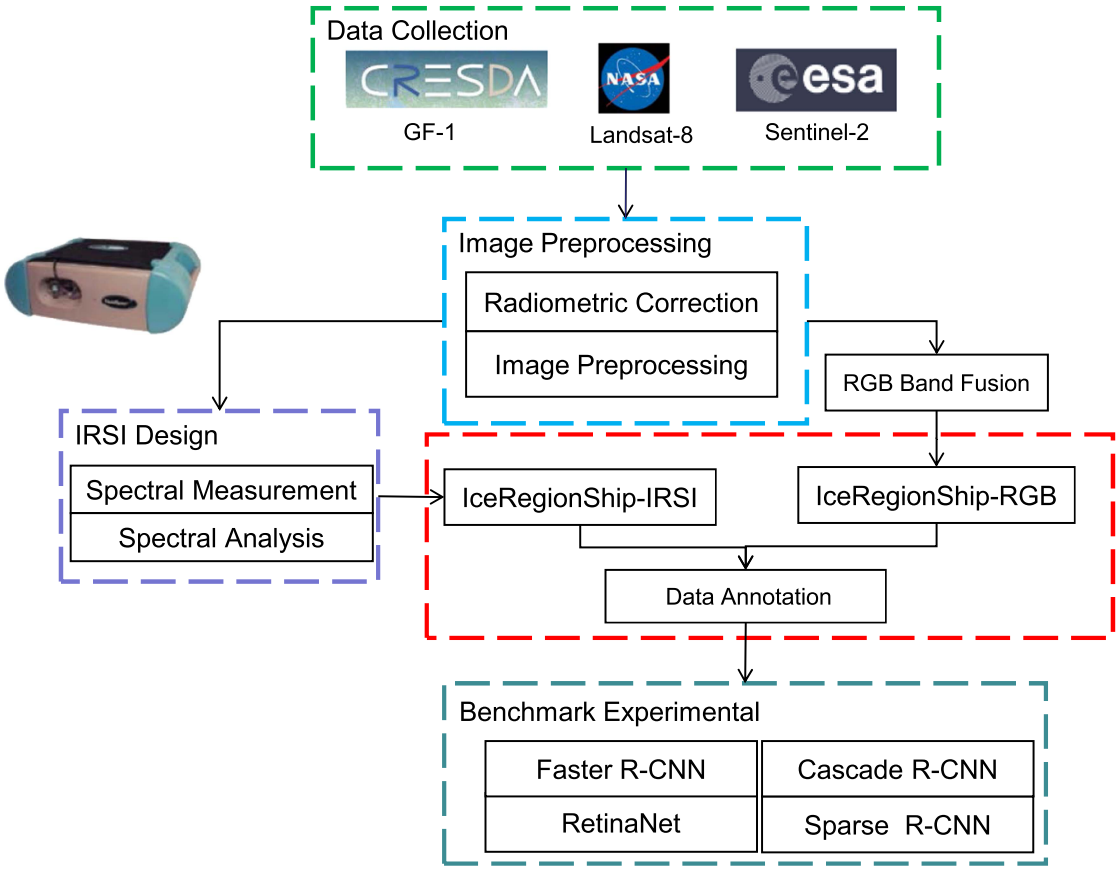


Fig. 2. Construction process of IceRegionShip.

TABLE II
DATA SOURCE OF ICEREGIONSHIP

Satellite	Sensor	Revisit time(d)	Swath (km)	Resolution (m)	Image
Landsat-8	OLI	16	185	15	28
Sentinel-2	MSI	5	290	10	15
GF-1	WFV	4	800	16	41
GF-1	PMS	4	60	2	5
				Total	89

the data and compensates for data missing on specific dates in a single source. The proposed IceRegionShip dataset primarily consists of 41 wide field view (WFV) images from GF-1, 28 images from the operational land imager (OLI) sensor of Landsat-8, and 15 images from the multispectral instrument sensor of Sentinel-2 (Table II). Official websites of these satellites provide publicly available low-resolution satellite images. Landsat-8 OLI images were fused to achieve a spatial resolution of 15 m and swath size of 185 km. Sentinel-2 images have a spatial resolution of 10 m and a swath size of 290 km. GF-1 WFV images have a spatial resolution of 16 m and swath size of 800 km. These images meet the requirements for large-scale coverage monitoring and are cost-effective, even though detailed features of the ship may not be as prominent as in high-resolution images. Additionally, we supplemented the dataset with higher resolution (spatial resolution: 2 m) images from the GF-1 PMS camera to provide additional ship instance samples. During the data acquisition process, we strive to select data with observation angles as close as possible to eliminate radiation differences caused by observation geometry. Subsequently, sensor noise and atmospheric effects are eliminated through radiometric calibration and atmospheric correction.

Located in the northeastern Bohai Sea, Liaodong Bay is a water area with the most prolonged ice period and the most extensive range in China. Most of the area is covered with drifting sea ice, with only a small portion near the coastline covered by fixed ice. Liaodong Bay is one of China's busiest maritime transportation regions, encompassing ports, such as Huludao, Jinzhou, and Yingkou, with dozens of shipping routes [19]. Ships frequently commute between these ports, increasing

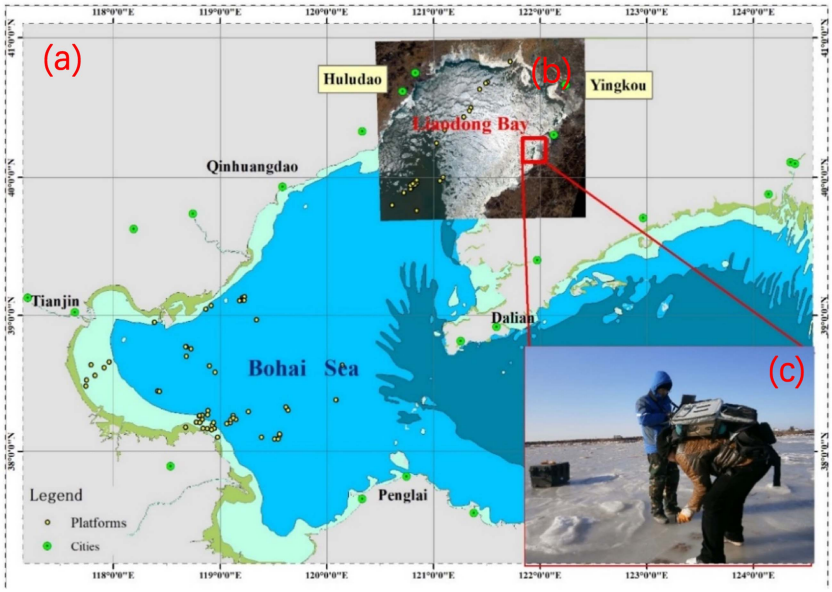


Fig. 3. Area for collecting image source data. (a) Location of Liaodong Bay. (b) Remote sensing image of sea ice in Liaodong Bay. (c) Onsite sampling of sea ice samples.

TABLE III
DIFFERENT THICKNESS OF SEA ICE SAMPLES USED IN THE EXPERIMENT

Sample no.	1	2	3	4	5	6	7	8	9	10	11
Thickness / cm	5.2	9.5	28.2	30	34.1	35.3	36.2	37.8	39.4	41.1	42.5

the complexity of maritime traffic. Ice class indicates the ship's ability to navigate in ice. Ships with high ice class usually have an additional level of strengthening for navigation and sustenance in ice. However, relatively few ships in this region have Ice Class B or higher. Thus, the presence of abundant sea ice poses a significant threat to maritime structures and ship navigation in the Liaodong Bay area, highlighting the critical importance of remote imaging of this region. Therefore, we selected the remote sensing images of the Liaodong Bay region as the source images. These images contain a large number of ships operating in IIW. We also conducted onsite sampling of sea ice in this region, as shown in Fig. 3.

B. Ice Region Ship Index

Although the selected image data satisfies the practical needs of large-scale detection, low-resolution images of IIW still pose challenges to SD. In RGB images, ships tend to blend with the sea ice, making it difficult for human eyes to discern them. Moreover, the elongated shapes formed by gaps between ice layers confuse the model and generate false detections. Therefore, we used additional bands and constructed spectral indices to assist SD. To capture the spectral differences between ships, seawater, and sea ice in the visible and near-infrared wavelength ranges, we conducted spectral measurements in the waters of Liaodong Bay. Subsequently, the required spectral bands were identified through the analysis of reflection spectra, and spectral indices for ship extraction in the ice zone were developed.

1) Spectral Measurement Experiment: In this experiment, 11 sea—ice samples with different thicknesses were used. The thicknesses of the employed sea–ice samples are listed in Table III. The ship deck data were collected from several typical decks (made of steel and painted in white, gray, iron red, and medium green colors) of ships sailing in the area. The ASD FieldSpec3 spectrometer was used as the measurement instrument. During the measurement process, the field of view of the spectrometer was set to 25°, zenith angle was set to 0°, and probe was positioned vertically on the ice surface (∼45 cm above the observed object). Reflectance data were obtained for different deck colors, different thicknesses of sea ice, and seawater samples. Each reflectance spectral curve consisted of a total of 1563 data points, and after removing the outliers, the average reflectance spectral curve was obtained as the reflectance spectrum for the sample.

2) Spectral Analysis Results: Through spectral measurements, we obtained the reflectance spectra of different-colored decks, different thicknesses of sea ice, and seawater, as shown in Fig. 4. To better differentiate between sea ice and ships, we compared the reflectance of different thicknesses of sea ice and ships (Fig. 5). The contrast between different-colored decks and sea ice is the most prominent in the shortwave infrared band, and the difference in reflectance tends to remain constant at different

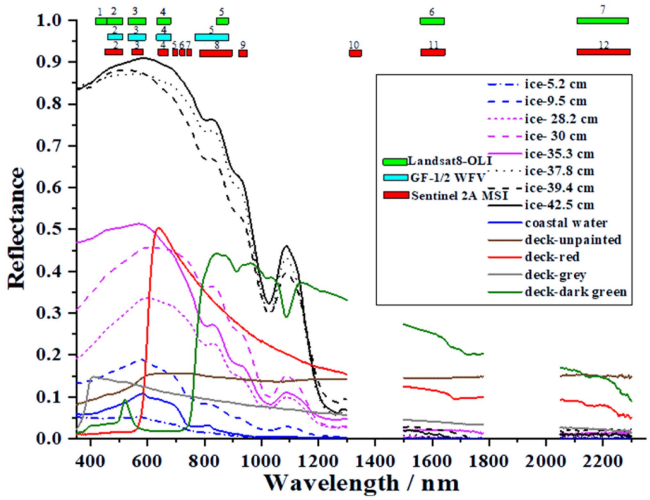


Fig. 4. Measured reflectance of water, sea ice, and ship decks. The band ranges for Landsat-8 OLI (green), GF-1/2 WFV (blue), and Sentinel-2A MSI (red) are indicated on the graph.

ice thicknesses, consistently exhibiting a significant difference. The contrast between decks and sea ice is also noticeable in the near-infrared band. Moreover, the contrast between decks and sea ice is slightly lower in the visible light range. In the case of thin sea ice, the contrast between decks, except for white decks, and sea ice is relatively low. As the thickness of sea ice increases, the reflectance of white decks becomes comparable to that of sea ice, hindering the extraction of ship targets in IIW.

3) Construction of Ice Region Ship Index (IRSI): The experiment shows that sea ice, seawater, and ships exhibit different reflectance values in different spectral bands. Therefore, ship targets can be effectively extracted by combining spectral reflectance in different bands. As GF-1 data lacks the infrared band, the near-infrared band was utilized in this article. The normalized difference water index (NDWI) highlights water bodies in sea ice using the near-infrared and green bands. However, our objective is to differentiate between sea ice and ship targets. Inspired by NDWI and considering the spectral properties of ship decks and sea ice, we designed the IRSI specifically for SD in IIW

$$\mathrm{IRSI} = \frac{R_{\mathrm{Green}} + R_{\mathrm{NIR}}}{R_{\mathrm{Green}} - R_{\mathrm{NIR}}} \tag{1}$$

where R_{Green}represents the green spectral band and R_{NIR}represents the near-infrared spectral band. IRSI is considered 0 when $R_{\mathrm{Green}} = R_{\mathrm{NIR}}$. At this time, the pixels appear black and will not have a major impact on the ship target. We applied NDWI and IRSI for ship target processing and compared the results. Ship targets processed with NDWI appear predominantly black, making it difficult to differentiate them from the surrounding ice layers, as shown in Fig. 6(b). In contrast, in the IRSI-processed image, the outer contours of ships are highlighted, distinguishing them from ice.

The thin strip-shaped ships blend with the ice layers, almost merging with black cracks, making them difficult to discern even to the human eye [Fig. 6(a)]. After applying IRSI, visual differences between ship targets and ice become more pronounced, facilitating SD. In addition to making ship targets more prominent, IRSI enables the model to further differentiate between ships and polynya during the learning process. In unprocessed images, the polynyas in ice may be identified as ship targets with a high level of confidence (Fig. 7). This leads to a number of false alarms, thus reducing the performance of the detector. In the IRSI-processed image, there are no bright spots outlining ship contours around nonship targets, and the polynyas appear blurred, reducing the probability of false alarms caused by elongated ship-resembling polynyas (Fig. 8).

C. IceRegionShip–RGB and IceRegionShip–IRSI

Owing to the differences between RGB images and images processed using the IRSI for SD in IIW, two subdatasets were constructed and named IceRegionShip–RGB and IceRegionShip–IRSI. IceRegionShip focuses on detecting ships in IIW using optical remote sensing images at spatial resolutions of 2, 10, 15, and 16 m. After preprocessing, which involves radiometric calibration and atmospheric correction, the RGB bands were fused and the images were cropped, resulting in a dataset of 9789 images containing 11 436 ship instances. IceRegionShip–IRSI primarily focuses on SD in low-resolution images (spatial resolutions of 10, 15, and 16 m). After performing the same preprocessing steps, a single-band image was obtained by fusing the bands using the IRSI, followed by cropping to generate a dataset of 7990 images containing 9073 ship instances. At the same resolution, ship targets in IceRegionShip–IRSI are generally brighter and more easily detectable than those in IceRegionShip, as shown in Fig. 9.

D. Data Annotation

In object detection, targets are commonly enclosed in rectangular bounding boxes, which is called HBB annotation. The HBB annotation is typically represented by a quadruple, which represents the coordinates of the object center and the width and height of the horizontal box. Well-known datasets, such as PASCAL VOC [20] and MS COCO [21], utilize HBB annotation. Similarly, we provide corresponding XML and JSON files for our dataset for each image containing ship targets. In the case of annotation for ice-covered environments, annotators need to exercise extra caution. The complexity of the background and the presence of floating ice fragments can affect the model's judgment of ships. To ensure high-quality annotations, we divided the annotation process into three stages, with five annotators and two reviewers assigned to three groups.

In the first stage, three annotators worked on the initial annotation for the IceRegionShip–RGB dataset, while two annotators annotated the IceRegionShip–IRSI dataset. During the annotation process, we aligned the ship's bow with the shorter side of the bounding box, aiming to tightly enclose the ship target and minimize the proportion of background area within the box. In the second stage, the two groups of annotators reviewed each other's annotations from the first stage, addressing any identified

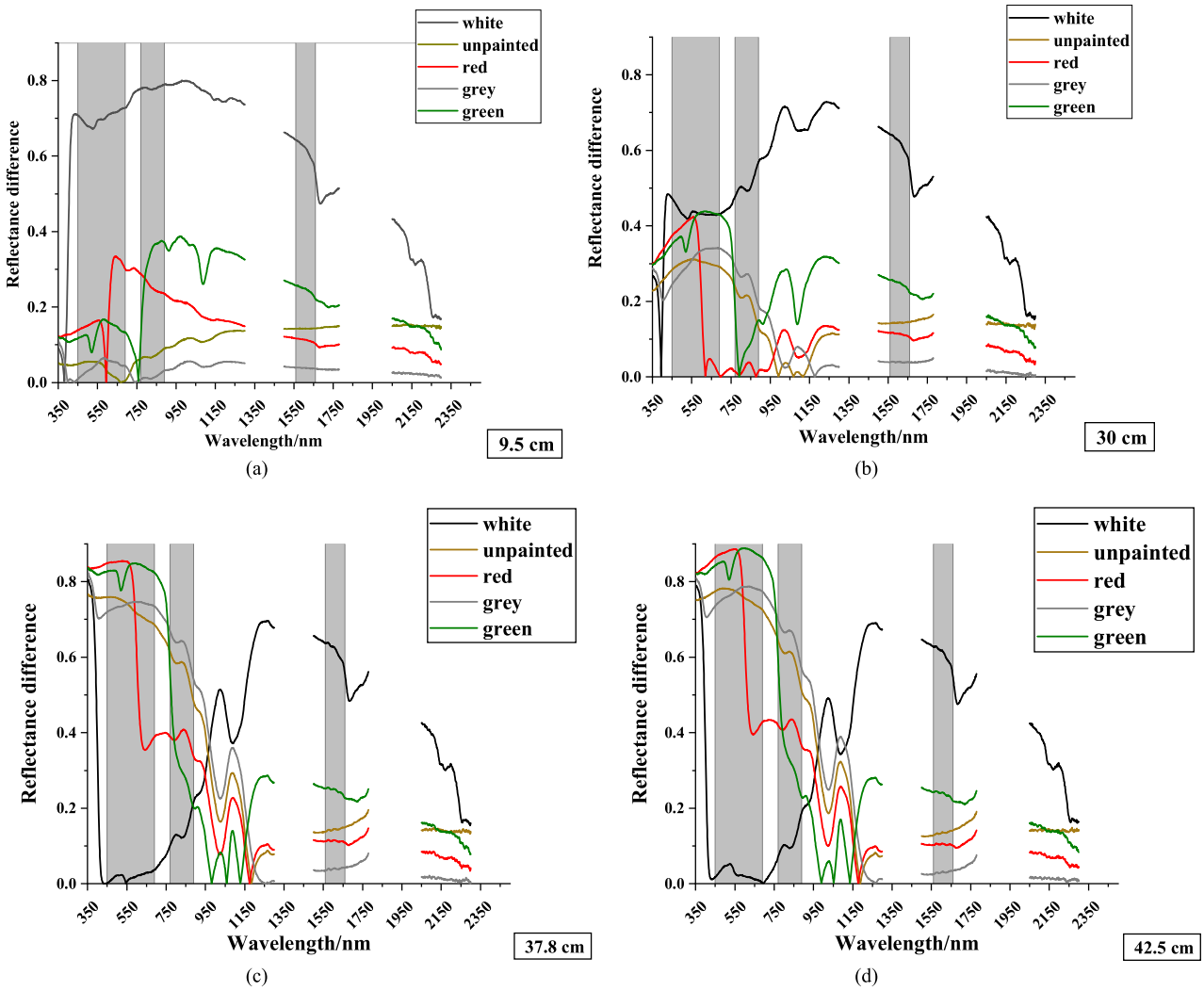


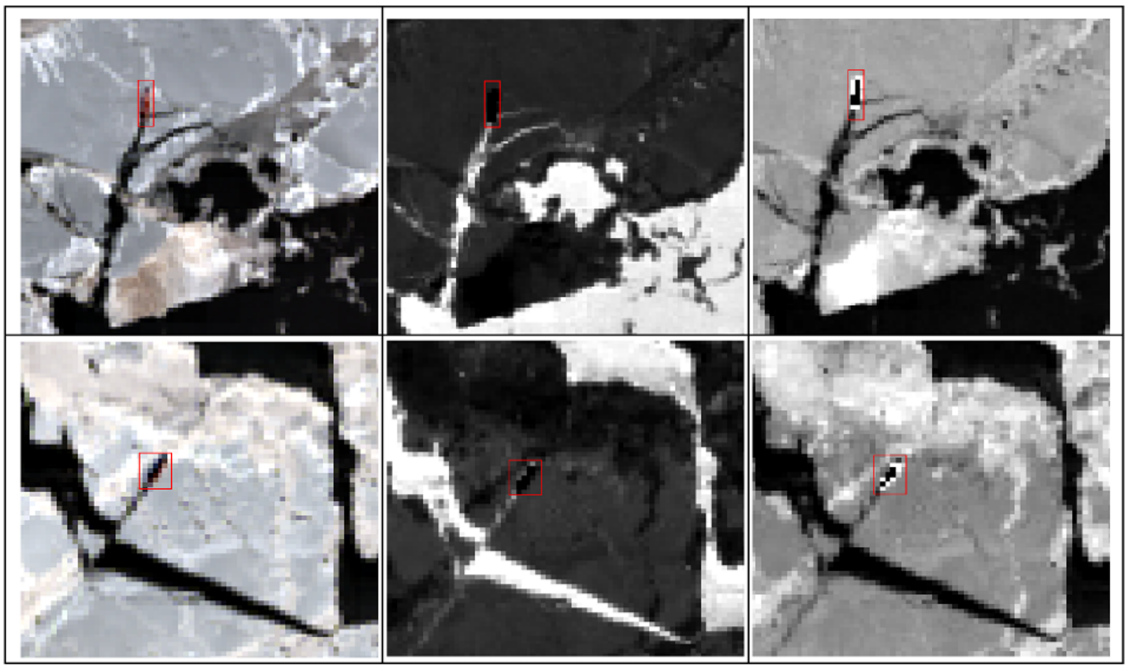
Fig. 5. Differences in reflectance between sea ice and ship deck. (a–d) show differences in reflectance between sea ice of multiple thicknesses (9.5, 30, 37.8, and 42.5 cm) and decks of different colors. Gray areas represent the region conducive to distinguishing between sea ice and the ship deck.

issues by returning to the first stage for resolution. In the third stage, two expert reviewers conducted a thorough examination. If there were any doubts or concerns regarding the annotations in an image, the two groups of annotators engaged in discussions and performed the first stage of annotation again for the image.

E. Characteristics of IceRegionShip

1) Specification: IceRegionShip is proposed to address the challenge of large-scale SD in IIW. To the best of our knowledge, this is the first introduced dataset in this field. IceRegionShip–RGB comprises fused images from the red, green, and blue spectral bands, faithfully representing the authentic IIW. IceRegionShip–IRSI, on the other hand, is derived through band calculations, which effectively eliminate certain portions of sea ice and enhance ship outlines. In essence, this corresponds to two feasible approaches: directly training models using the original images or training models after applying suitable sea ice preprocessing methods. We expect these two subdatasets will contribute to solving the complexities associated with SD in challenging ice-covered scenes and provide a robust pathway toward achieving reliable detection models in ice-covered regions.

2) Small Scale of Instance: With reference to the scale division for object detection in MS COCO, bounding box regions below 32×32 pixels are referred to as small objects, regions from 32×32 pixels to 96×96 pixels are categorized as medium objects, and bounding box regions larger than 96×96 pixels are considered large objects [18]. Considering the practical applications of wide-range monitoring, most images in IceRegionShip have lower resolutions, making most ship targets small objects. Thus, IceRegionShip–IRSI was primarily proposed for detecting complex ice area scenes with low spatial resolution, where ship targets are mostly small, with an average bounding box region of 221 pixels. Fig. 10 presents the distribution of ship instance sizes in IceRegionShip–RGB and IceRegionShip–IRSI. It is evident that small objects dominate the dataset, posing certain challenges for detection.



(a) (b) (c)

Fig. 6. Comparison of the effects of NDWI and IRSI on ship targets. (a) Preprocessed Image. (b) NDWI-processed Image. (c) IRSI-processed Image.

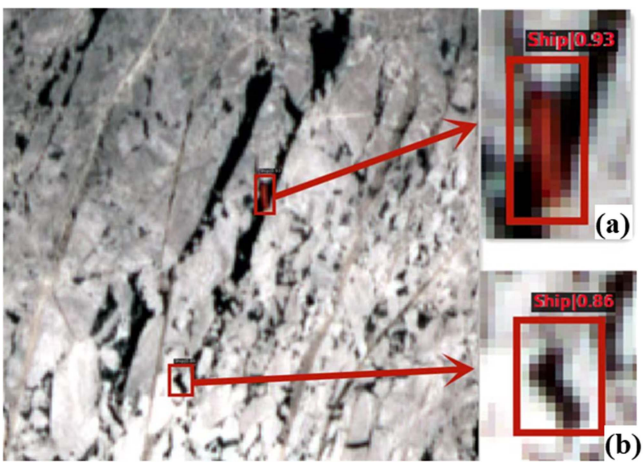


Fig. 7. Detection results in RGB images. (a) Ship. (b) False alarm, polynya was falsely detected as a ship.

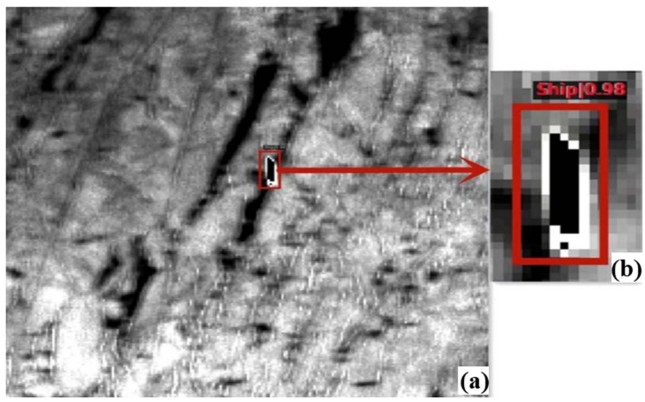
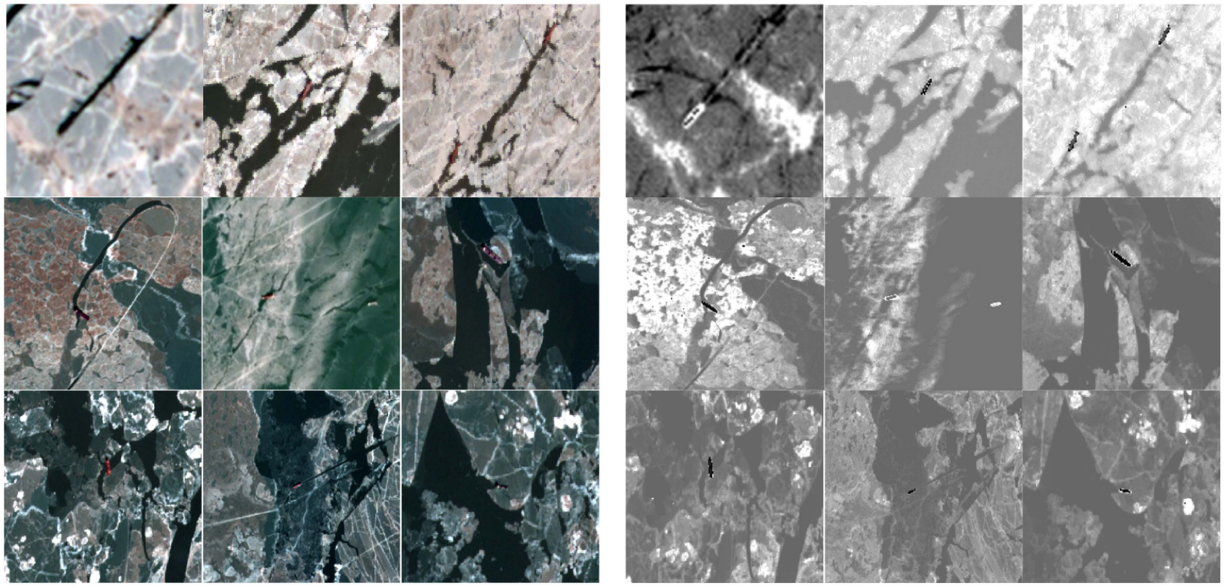


Fig. 8. (a) Detection results in IRSI images. (b) Ship target processed by the IRSI.

3) Diversity: The diversity of the dataset enables it to encompass a broad range of states that may occur in real-world scenarios, making the dataset representative of practical applications. The IceRegionShip dataset sources data from multiple publicly available optical remote sensing satellites with different sensors, containing images with different spatial resolutions and quality. The IceRegionShip dataset considers different sea–ice concentrations in IIW. As shown in Fig. 11, regions with low sea–ice concentration primarily consist of seawater with minimal sea–ice presence. Seawater separates ice in regions with medium ice concentration, forming multiple fragmented pieces with a significant proportion of broken ice. Regions with ice concentrations mainly feature large, solid ice blocks with relatively smooth sea–ice-surrounding ship targets. In addition to diverse scenes, the dataset also features various ship instances varying in deck color, size, orientation, and position, thus representing scenarios ranging from freely navigable sea surfaces with minimal sea ice to ice-congested situations where ships are surrounded by thick ice and mimicking ships beset in ice, among others.

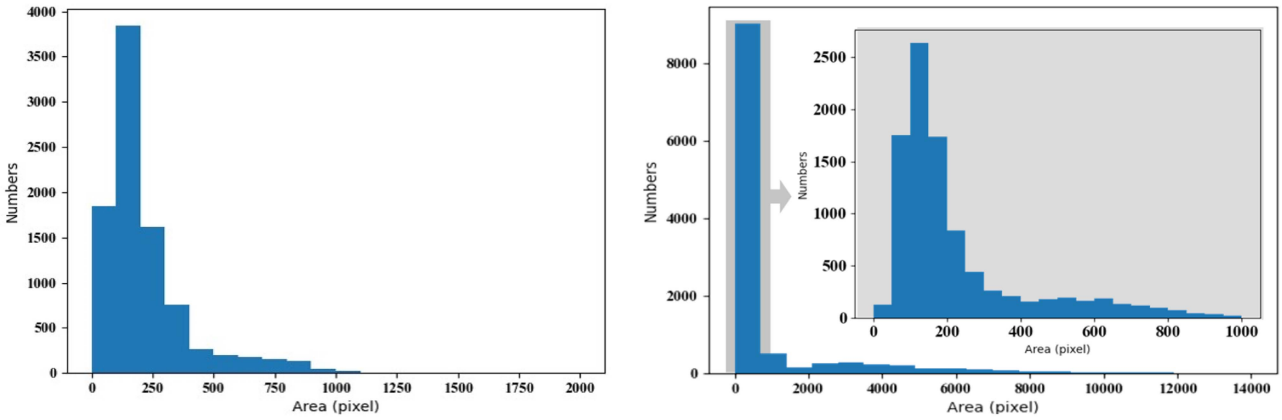
IV. Benchmark for Ship Detection of Ice Region

In this section, we evaluate the IceRegionShip–RGB and IceRegionShip–IRSI datasets using state-of-the-art detection models. We employ multiple metrics, including average precision (AP), to assess the experimental results.



(a) (b)

Fig. 9. Sample images chosen from low spatial resolutions of 10, 15, and 16 m from two datasets. (a) IceRegionShip–RGB. (b) IceRegionShip–IRSI.



(a) (b)

Fig. 10. Ship size distributions in (a) IceRegionShip–RGB and (b) IceRegionShip–IRSI, with the majority of small targets.

A. State-of-the-Art Detection Models

1) Faster R-CNN: Faster R-CNN [23] is one of the classical two-stage models composed of a backbone network [region proposal network (RPN)] and classification and detection heads. The RPN network can rapidly generate proposal boxes by abandoning the original sliding window and selective search methods, considerably improving detection efficiency and accuracy. The introduction of the feature pyramid network (FPN) has further enhanced the multiscale detection capability of faster R-CNN, making it a commonly used benchmark model.

2) Cascade R-CNN: Detectors that employ an anchoring mechanism typically utilize an intersection over union (IoU) threshold to distinguish positive and negative samples based on bounding boxes. Predicted bounding boxes with IoU below the threshold are considered negative samples. Cascade R-CNN [24] utilizes a multistage object detection framework consisting of a series of detectors with progressively increasing IoU thresholds. By gradually training the detectors, high-quality detection models can be obtained, reducing overfitting.

3) RetinaNet: RetinaNet [25] is a classical single-stage model consisting of a backbone network, FPN, and box regression and classification subnetworks. The model employs a simple and effective loss function called focal loss, which reduces the weight of negative samples to address the issue of class imbalance in single-stage detection models. Compared to two-stage models, RetinaNet improves detection accuracy while ensuring good detection quality.

4) Sparse R-CNN: Sparse R-CNN [26] is a purely sparse object detection method that utilizes a predefined number, K,

TABLE IV
DETECTION PERFORMANCE OF STATE-OF-THE-ART MODELS ON ICEREGIONSHIP–RGB

Model	Backbone	Sched.	FPS	Param (M)	AP50	AP75	mAP
Faster R-CNN	ResNet-50+FPN	1×	68.0	42.12	79.50	24.40	35.95
	ResNet-101+FPN		55.3	60.11	79.20	21.90	35.06
Cascade R-CNN	ResNet-50+FPN	1×	48.7	68.93	86.00	23.70	38.59
	ResNet-101+FPN		38.5	87.92	86.30	**28.40**	**40.42**
RetinaNet	ResNet-50+FPN	1×	82.1	36.1	83.80	25.20	37.71
	ResNet-101+FPN		59.0	55.1	83.80	26.50	38.24
Sparse R-CNN	ResNet-50+FPN	1×	45.2	105.94	87.70	23.50	37.57
	ResNet-101+FPN		34.9	124.94	**89.10**	23.60	39.67
ATSS	ResNet-50+FPN	1×	74.3	31.89	88.10	26.90	39.67
	ResNet-101+FPN		56.8	50.88	86.90	23.50	37.36

The bold font indicates the best result in an evaluation metric.

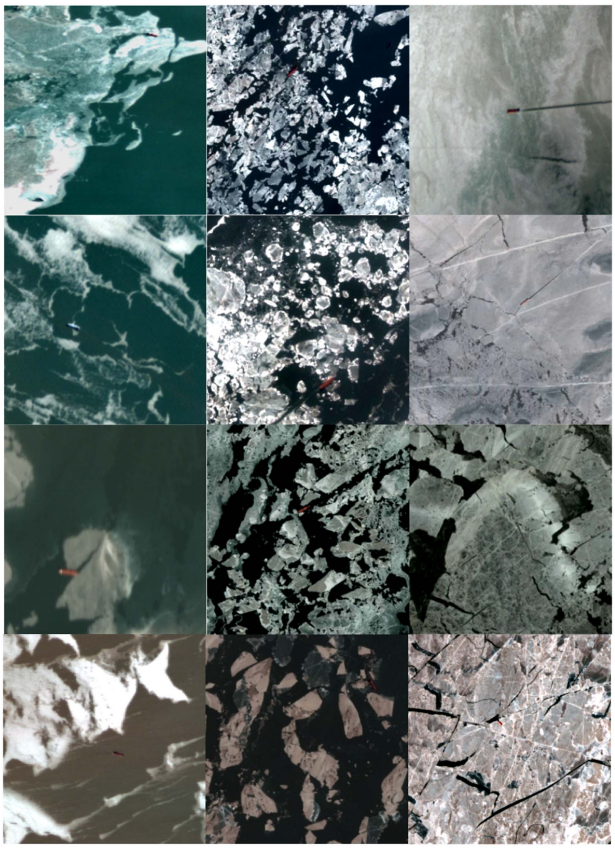


Fig. 11. Images of waters with different sea–ice concentrations: (a) low, (b) medium, and (c) high.

of candidate boxes (e.g., 100) instead of relying on a dense set of hundreds of thousands of candidate boxes. It primarily consists of a backbone network, learnable proposals, and multiple dynamic instance heads. The learnable proposal module learns the initial distribution of targets and generates the initial K of candidate boxes. The dynamic instance interaction heads take the ROI features extracted from each box and input them into their dedicated heads for target localization and classification. The model is trained using one-to-one label assignment and set prediction loss, avoiding the postprocessing nonmaximum suppression step in the results.

5) ATSS: In recent years, the anchor-free methods have rapidly developed due to their reduced reliance on hyperparameters and enhanced generalization capabilities. Adaptive training sample selection (ATSS) is one of the more novel approaches in this regard. It introduces the ATSS method, demonstrating the crucial significance of sample selection for detectors. Automatically selecting positive and negative samples based on target statistical features [29] significantly improves the detector's performance.

B. Experimental Setup

All experiments were conducted on a personal computer with an Intel Core i7-12700KF CPU, 32 GB of RAM, and NVIDIA RTX3090 GPU (24 GB VRAM). The software used included Python 3.8, CUDA 11.7, and PyTorch 1.13.0.

In many previous benchmark tests, different models were implemented using different codebases [18], making fair comparisons among algorithms on the same benchmark difficult. To address this issue and unify the implementation differences across models, all models in this section were trained and tested using MMDetection [27]. All detectors were optimized using the AdamW [28] algorithm. The initial learning rate is set to 0.0001. Each batch consisted of 4 images, and a weight decay 0.05 was applied. A learning rate warm-up was employed for the first 500 iterations.

C. Evaluation Metrics

In this article, we employ authoritative and standard detection evaluation metrics, such as IoU, precision, recall, and AP, to quantitatively evaluate the detection performance of the detectors mentioned above on IceRegionShip. From a localization perspective, IoU is introduced as a measure of the overlap between the predicted and ground truth regions. It represents the ratio of the overlapping area between the two regions to their total area. A higher IoU indicates a closer match between the prediction and the ground truth, indicating more accurate localization. The calculation of IoU is given by the following formula:

$$\text{IoU} = \frac{\text{area}(\text{Bbox}_{\text{pb}} \cap \text{Bbox}_{\text{gt}})}{\text{area}(\text{Bbox}_{\text{pb}} \cap \text{Bbox}_{\text{gt}})} \tag{2}$$

where Bbox_{pb} represents the predicted bounding box and Bbox_{gt}represents the ground truth box.

During ship classification, the model can make misclassifications, confusing targets with the background. Each target can have four outcomes: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP represents the number of ship targets correctly classified; FP represents the number of nonship targets incorrectly classified as ships; FN represents the number of ship targets incorrectly classified as background; and TN represents the number of nonship targets correctly classified. Precision reflects the model's accuracy in predicting positive samples, while recall measures the proportion of correctly detected positive samples among all positive samples. The specific calculation equations are given as follows:

$$\text{Precison} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

AP represents the AP of the detector at different recall levels. It is calculated as the mean precision values at all recall levels. The formula for AP is as follows:

$$\text{AP} = \int_0^1 P(r)dr \tag{5}$$

where P represents precision and $r$represents recall. MS COCO has established a comprehensive and rigorous metric system. Similarly, we adopt three metrics, namely AP50, AP75, and mAP, to comprehensively evaluate the performance of the detector. AP50 uses an IoU threshold of 0.50, considering predicted boxes with an overlap more significant than this threshold as correct predictions for calculating AP. AP75 uses an IoU threshold of 0.75. mAP is calculated by setting the IoU threshold between 0.5 and 0.95 with a step size of 0.05, computing AP at each threshold, and averaging the 10 obtained AP values.

D. Experimental Results and Analysis

Table IV presents the detection results for several advanced detectors on the IceRegionShip–RGB dataset, which includes images with spatial resolutions of 2, 10, 15, and 16 m. The notation "1×" represents training for 12 epochs. To address the multiscale variations in the dataset, the FPN structure was employed in all models. The model parameter sizes are also listed as essential references for application. From the results, the cascade R-CNN with ResNet101 backbone achieves the best performance with an mAP of 40.42 and the highest AP75. ATSS demonstrates promising results, reaching high levels of AP75 and mAP with the smallest Param size. Sparse R-CNN, as a sparse object detector, also performs well, achieving the highest AP50 value of 89.10, which can be attributed to the sparsity of ships on the ice-covered sea surface. Fig. 12 visually shows the detection effect of the model.

To assess the detection performance on low-resolution data processed using IRSI, we conducted a comparative experiment between IceRegionShip–RGB and IceRegionShip–IRSI. To obtain fairer results, we extracted the low-resolution parts of 10, 15, and 16 m from IceRegionShip–RGB into a new dataset, IceRegionShip–RGB∗, for comparison with IceRegionShip–IRSI. The results are presented in Table V. Owing to the overall lower resolution, the mAP for IceRegionShip–RGB∗ was considerably lower than that for IceRegionShip However, it is evident that all benchmark models tested on IceRegionShip–IRSI exhibit notable improvements. In the context of the IceRegionShip–IRSI dataset, ATSS achieved a notable mAP value of 42.51, which is significantly higher than its mAP value of 38.07 on the IceRegionShip–RGB dataset. Similarly, Cascade R-CNN also increased from 33.85 to 40.11. Other models also have very significant growth. Particularly, for the more stringent high-precision metric, AP75, faster R-CNN demonstrated a 6.00 improvement and cascade R-CNN exhibited a 7.3 improvement.

V. Discussion

In this article, we have built a novel ship dataset tailored explicitly for SD in IIW, IceRegionShip, which fills a gap in the existing ship datasets. We introduced the IRSI index and constructed IceRegionShip–IRSI. Our experimental results have shown that applying the IRSI index enhances the contrast between ships and the background sea ice, resulting in a substantial improvement in the performance of all detection models. Compared to the IceRegionShip–RGB∗ dataset, the models exhibited the highest detection performance improvement of 8.72 (cascade R-CNN) on the IceRegionShip–IRSI dataset. On average, all models showed an improvement of 4.60. Visualizations of the detection results further confirm it, as depicted in Fig. 13.

The performance of different models is also worth noting. First, one-stage models exhibit high FPS and low Param, rendering them more suitable for applications. Among these, ATSS demonstrated outstanding results, emphasizing the significance of positive and negative sample selection for IceRegionShip [29]. Among the two-stage models, cascade R-CNN outperformed faster R-CNN, illustrating the advantages of the cascading structure and different IOU thresholds [22].

We tested the dataset using end-to-end detection networks based on transformers (DETR, deformable-DETR) and the Swin Transformer-Tiny (Swin-T) backbone network. Regrettably, due to its slow convergence, end-to-end detection networks based

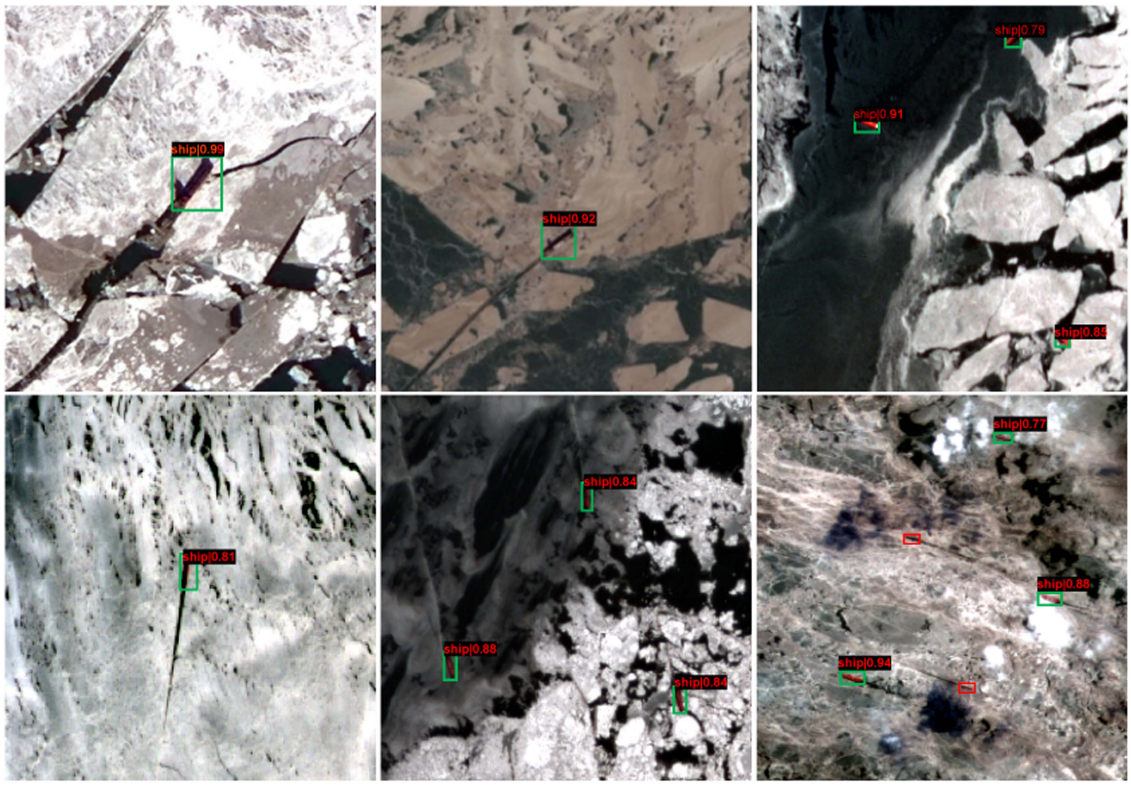


Fig. 12. Visualization of the detection results of cascade R-CNN on IceRegionShip–RGB.

TABLE V
DETECTION PERFORMANCE OF STATE-OF-THE-ART MODELS ON ICEREGIONSHIP–RGB∗ AND ICEREGIONSHIP–IRSI

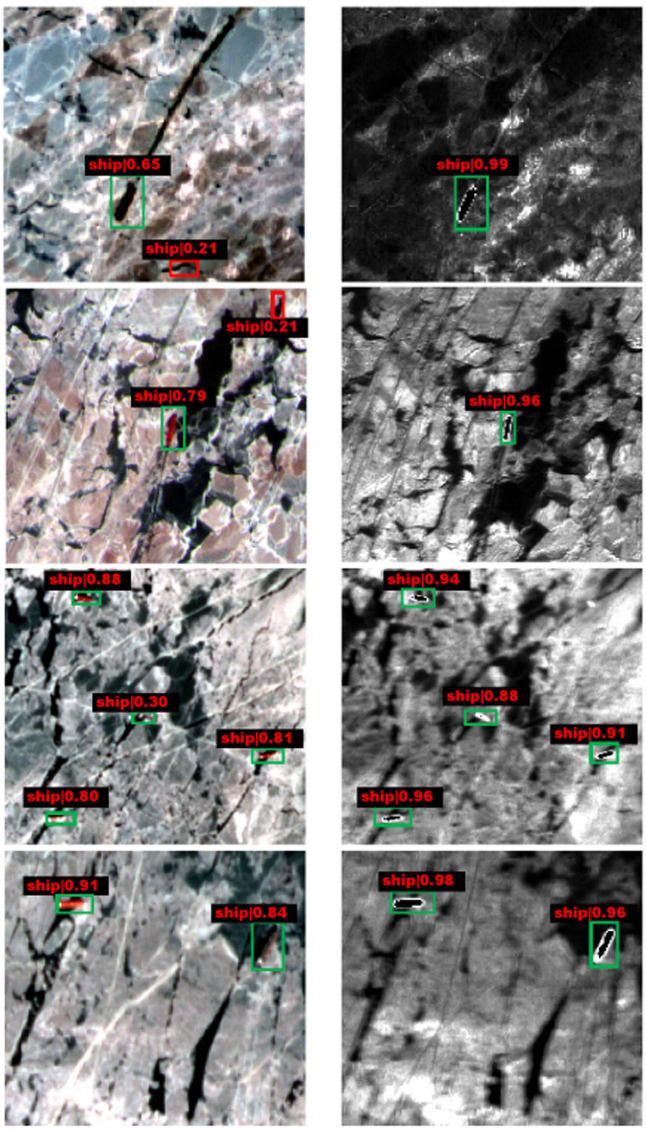
Model	Backbone	FPS	Dataset	Sched.	AP50	AP75	mAP
Faster R-CNN	ResNet-50+FPN	68.2	IceRegionShip–RGB*	1×	78.40	11.90	30.55
		67.4	IceRegionShip–IRSI		80.00	17.90	34.26
	ResNet-101+FPN	55.5	IceRegionShip–RGB*		78.00	15.10	30.50
		50.4	IceRegionShip–IRSI		86.80	17.40	33.72
Cascade R-CNN	ResNet-50+FPN	46.8	IceRegionShip–RGB*	1×	77.20	15.90	28.95
		46.0	IceRegionShip–IRSI		87.30	23.20	37.67
	ResNet-101+FPN	38.6	IceRegionShip–RGB*		83.20	19.90	33.85
		37.9	IceRegionShip–IRSI		87.50	26.10	40.11
RetinaNet	ResNet-50+FPN	80.6	IceRegionShip–RGB*	1×	78.90	11.20	29.65
		78.8	IceRegionShip–IRSI		80.90	12.70	31.32
	ResNet-101+FPN	58.8	IceRegionShip–RGB*		81.60	19.10	32.80
		61.1	IceRegionShip–IRSI		83.30	19.70	33.04
Sparse R-CNN	ResNet-50+FPN	46.9	IceRegionShip–RGB*	1×	77.50	8.20	26.65
		46.1	IceRegionShip–IRSI		87.30	16.30	33.80
	ResNet-101+FPN	38.6	IceRegionShip–RGB*		87.80	11.90	31.89
		38.4	IceRegionShip–IRSI		88.50	20.70	37.29
ATSS	ResNet-50+FPN	72.3	IceRegionShip–RGB*	1×	87.50	21.90	38.07
		72.1	IceRegionShip–IRSI		**89.90**	**27.40**	**42.51**
	ResNet-101+FPN	55.7	IceRegionShip–RGB*		86.50	16.80	33.72
		55.5	IceRegionShip–IRSI		88.20	23.50	38.99

Note: * indicates that IceRegionShip–RGB utilized data from the low spatial resolution portion for a fair comparison with IceRegionShip–IRSI.
The bold font indicates the best result in an evaluation metric.

TABLE VI
DETECTION PERFORMANCE OF SWIN-T BACKBONE MODELS ON ICEREGIONSHIP–RGB∗ AND ICEREGIONSHIP–IRSI

Model	Backbone	FPS	Dataset	Sched.	AP50	AP75	mAP
Cascade R-CNN	Swin-T+FPN	41.4	IceRegionShip–RGB*	1×	86.70	26.30	37.80
		40.8	IceRegionShip–IRSI		86.80	20.60	38.05
Sparse R-CNN	Swin-T+FPN	41.8	IceRegionShip–RGB*	1×	87.90	15.60	33.24
		38.1	IceRegionShip–IRSI		85.70	16.00	34.09
ATSS	Swin-T+FPN	59.3	IceRegionShip–RGB*	1×	86.10	19.20	34.88
		61.1	IceRegionShip–IRSI		**89.20**	**25.40**	**42.01**

The bold font indicates the best result in an evaluation metric.



(a) (b)

Fig. 13. Visualization of the detection results of cascade R-CNN on (a) IceRegionShip–RGB* and (b) IceRegionShip–IRSI. Green bounding boxes are ships that are correctly detected. Red bounding boxes are false alarms.

on transformers (DETR, deformable-DETR) need an extremely long training schedule on large-scale datasets like COCO [30]. The number of instances in our dataset challenges meeting their convergence requirements. Table VI presents the detection results using Swin-T as the backbone network. Clearly, this outcome also underscores the role of IRSI, as all three models exhibited improved detection performance to some degree.

In the next phase of this article, we will explore the integration of multimodal RGB and IRSI data to enhance target detection accuracy. RGB data retain radiometric information from three spectral bands of the target, allowing for differentiation from the surrounding environment based on color variations. IRSI data enhance distinctions between targets and the environment in terms of spatial features, such as texture and geometry. In the future, this fusion can be achieved by combining RGB and IRSI data at the shallow layers (or input layers) of deep-learning models. By jointly considering the differences in color features and spatial features between targets and the environment, we aim to improve the precision of target detection.

VI. CONCLUSION

This article proposes the IceRegionShip dataset for SD in IIW. This dataset includes subdatasets IceRegionShip–RGB and IceRegionShip–IRSI. To the best of our knowledge, this is the first optical remote sensing dataset designed specifically for SD in IIW. During the dataset processing, we constructed IRSI through spectral measurement experiments. The IRSI was used to reduce the influence of sea ice on SD and enhance ship outlines. Comparative experiments demonstrate that the use of IRSI facilitates SD in IIW. We extensively benchmarked various state-of-the-art deep-learning-based methods on this dataset. Thus, the obtained results of the experiment can be used as a useful baseline and contribute to the advancement of SD in IIW. In the future, we aim to continue to conduct research in related areas, further enlarge the dataset, and develop detection models that are more suitable for IIW.

REFERENCES

[1] J. Vanhatalo, J. Huuhtanen, and M. Bergström,, "Probability of a ship becoming beset in ice along the northern sea route—A Bayesian analysis of real-life data," *Cold Regions Sci. Technol.*, vol. 184, Apr. 2021, Art. no. 103238.

[2] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sens. Environ.*, vol. 207, pp. 1–26, Mar. 2018.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[4] M. Bi, M. Wang, and Z. Li, "Vision transformer with contrastive learning for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 738–749, Dec. 2022.
[5] B. Chen, P. Li, X. Chen, and B. Wang, "Dense learning based semi-supervised object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4805–4814.
[6] S. Chen et al., "MSDN: Mutually semantic distillation network for zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 7602–7611.
[7] J. Ding et al., "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7778–7796, Nov. 2022,.
[8] G. S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[9] X. Sun et al., "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 116–130, Feb. 2022.
[10] Z. Liu, L. Yuan, L. Weng, and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proc. Int. Conf. Pattern Recognit. Appl. Methods*, 2017, pp. 324–331.
[11] K. Chen et al., "FGSD: A dataset for fine-grained ship detection in high resolution satellite images," *CoRR*, vol. abs/2003.06832, 2020. [Online]. Available: https://arxiv.org/abs/2003.06832
[12] Y. Han, X. Yang, T. Pu, and Z. Peng, "Fine-grained recognition for oriented ship against complex scenes in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5612318.
[13] Z. Zhang, L. Zhang, Y. Wang, P. Feng, and R. He, "ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8458–8472, Sep. 2021.
[14] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014.
[15] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020.
[16] Y. Zhang, Y. Yuan, Y. Feng, and X. Lu, "Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5535–5548, Aug. 2019,.
[17] D. Lam R. Kuzma and K. McGee, "View: Objects in context in overhead imagery," *CoRR*, Feb. 2018. [Online]. Available: http://arxiv.org/abs/1802.07856
[18] G. Cheng et al., "Towards large-scale small object detection: Survey and benchmarks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13467–13488, Nov. 2023.
[19] Y. Ding et al., "Origins and features of oil slicks in the Bohai sea detected from satellite SAR images," *Mar. Pollut. Bull.*, vol. 106, no. 1–2, pp. 149–154, May 2016.
[20] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[21] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[22] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[24] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.
[25] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[26] P. Sun et al., "Sparse R-CNN: End-to-end object detection with learnable proposals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14449–14458.
[27] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," *CoRR*, vol. abs/1906.07155, 2019. [Online]. Available: http://arxiv.org/abs/1906.07155
[28] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *CoRR*, vol. abs/1711.05101, Nov. 2017. [Online]. Available: http://arxiv.org/abs/1711.05101
[29] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9756–9765.
[30] Z. Dai, B. Cai, Y. Lin, and J. Chen, "UP-DETR: Unsupervised pre-training for object detection with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1601–1610.

Peilin Wang received the B.S. degree in information management and information system from Dalian Maritime University, Dalian, China, in 2022. He is currently working toward the M.S. degree in traffic engineering with Navigation College, Dalian Maritime University, Dalian, China.

His research interests include deep learning and vessels object detection.

Bingxin Liu (Member, IEEE) was born in Shandong Province, China, in 1984. He received the B.S. degree in geographical information science from Northeast Normal University, Changchun, China, in 2008, and the Ph.D. degree in environmental science from Dalian Maritime University, Dalian, China, in 2013.

He is currently a Professor with the Navigation College of Dalian Maritime University. His research interests include surveillance of vessels using remote sensing, remote sensing mechanisms, and deep learning.

Dr. Liu was the recipient of the second prize of National Technological Invention Award and the first prize of Liaoning Provincial Science and Technology Progress Award.

Ying Li was born in Liaoning, China, in 1968. She received the Ph.D. degree in geosciences from Tohoku University, Sendai, Japan, in 1996.

She is currently a Professor with the Navigation College, Dalian Maritime University, Dalian, China. She has authored more than 100 refereed journals and conference papers. Her research interests include geographic information science, marine environment information technology, and ship pollution detection.

Peng Chen was born in Dalian, Liaoning, China, in 1982. He received the B.S. degree in geographical information science from the China University of Geosciences, Wuhan, China, in 2005, and the Ph.D. degree in environmental science from Dalian Maritime University, Dalian, China, in 2012.

He is currently an Associate Professor with the Dalian Maritime University. His research interests include machine learning and remote sensing image processing.

Peng Liu received the Ph.D. degree in marine science from Kobe University, Kobe, Japan, in 2015.

He is currently an Associate Professor with Navigation College, Dalian Maritime University, Dalian, China. His research interests include image processing, data mining, and pattern recognition.