ShipRSImageNet: A Large-Scale Fine-Grained Dataset for Ship Detection in High-Resolution Optical Remote Sensing Images

Zhengning Zhang, Lin Zhang, *Member, IEEE*, Yue Wang, Pengming Feng, *Member, IEEE*, and Ran He

***Abstract*—Ship detection in optical remote sensing images has potential applications in national maritime security, fishing, and defense. Many detectors, including computer vision and geoscience-based methods, have been proposed in the past decade. Recently, deep-learning-based algorithms have also achieved great success in the field of ship detection. However, most of the existing detectors face difficulties in complex environments, small ship detection, and fine-grained ship classification. One reason is that existing datasets have shortcomings in terms of the inadequate number of images, few ship categories, image diversity, and insufficient variations. This article publishes a public ship detection dataset, namely ShipRSImageNet, which contributes an accurately labeled dataset in different scenes with variant categories and image sources. The proposed ShipRSImageNet contains over 3435 images with 17 573 ship instances in 50 categories, elaborately annotated with both horizontal and orientated bounding boxes by experts. From our knowledge, up to now, the proposed ShipRSImageNet is the largest remote sensing dataset for ship detection. Moreover, several state-of-the-art detection algorithms are evaluated on our proposed ShipRSImageNet dataset to give a benchmark for deep-learning-based ship detection methods, which is valuable for assessing algorithm improvement.[1]**

***Index Terms*—Deep learning, fine-grained image classification, image dataset, remote sensing images, ship detection.**

I. Introduction

MEASURING and monitoring human activities at sea have become increasingly important along with global economic integration. Fast-growing shipping traffic causes dramatically increasing infractions, such as environmentally devastating ship accidents, piracy, illegal fishing, drug trafficking, and illicit cargo. The International Maritime Organization defines maritime domain awareness as an effective understanding of any activity associated with the maritime domain that could impact security, safety, economy, and environment [1]. Up to the end of the year 2019, there are about 92 867 merchant ships and 5500 warships worldwide. With such high-variational categories and sizes of ships distributed on the vast sea, it is essential and technically challenging to discover and identify them quickly and accurately.

In recent years, due to the so-called orbital revolution processing, the spectral and radiometric resolution, revisit time, and spatial resolution of remote sensing satellites have been improved dramatically.

High-resolution optical remote sensing images with different swath and resolutions are often combined and analyzed for faster and more accurate ship detection. More information on ship features, e.g., structure and texture, are employed. This information has laid the foundation for fine-grained ship classification and even identifying a specific ship.

Over the past 20 years, many automatic ship detection methods have been developed and have achieved significant results. Kanjir *et al.* [1] categorized these methods into eight groups including threshold-based methods, salient-based methods, shape and texture features based methods, transform-domain methods, anomaly detection methods, computer vision methods, and deep learning methods. Deep learning methods are becoming increasingly popular because of their impressive classification performance. Object detection in optical remote sensing images with deep learning methods has achieved the state-of-the-art performance [2]–[4]. Although deep learning methods can extract features automatically, datasets with large-scale, high-quality images are essential for training models.

Significant efforts have been made to build generic image datasets, such as ImageNet [5], MS COCO [6], and PASCAL VOC [7]. However, remote sensing images from satellites are quite different from natural scene images. Satellite-based remote sensing images generally capture roof information from the bird-eye view, whereas ground-based imaging usually capture profile information, as shown in Fig. 1. It is challenging to transfer object detectors trained by ground-based imaging datasets to remote sensing datasets, especially for the ship detection task.

Manuscript received January 11, 2021; revised March 16, 2021, May 11, 2021, June 15, 2021, and July 20, 2021; accepted August 3, 2021. Date of publication August 12, 2021; date of current version September 7, 2021. This work was jointly supported in part by the National Key R&D Program of China “Integration and demonstration of maritime target recognition and surveillance system” under Grant 2017YFC1404900 and Tianjin Key Laboratory of Intelligent Information Processing in Remote Sensing. *(Corresponding author: Zhengning Zhang.)*

Zhengning Zhang and Yue Wang are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: 23880666@qq.com; wangyue@tsinghua.edu.cn).

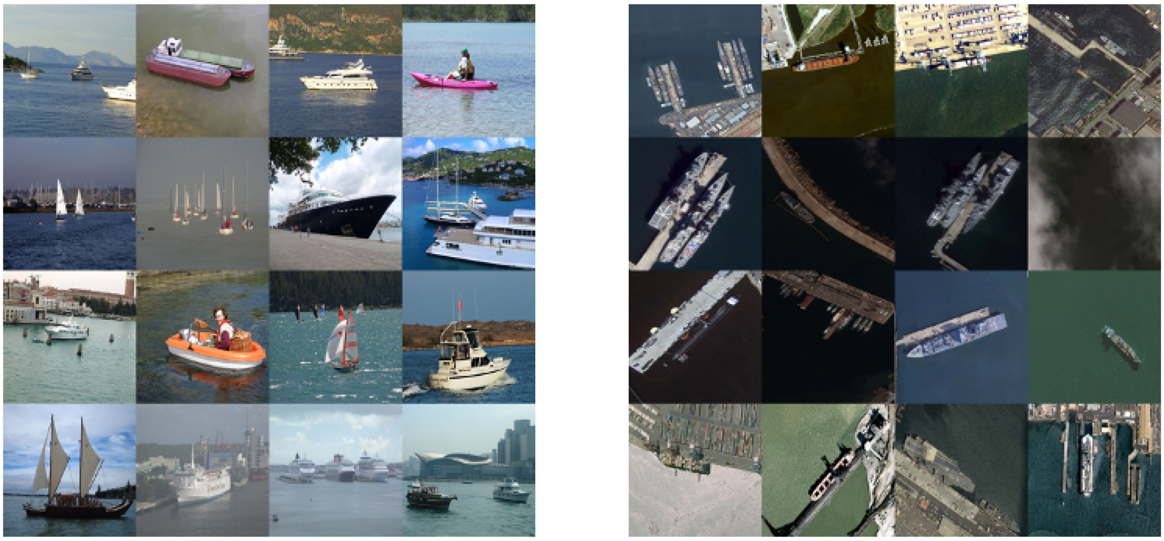
Lin Zhang is with the Tsinghua Shenzhen International Graduate School, Shenzhen 518000, China (e-mail: linzhang@sz.tsinghua.edu.cn).

Pengming Feng is with the State Key Laboratory of Space-Ground Integrated Information Technology, China Academy of Space Technology, Beijing 100095, China (e-mail: p.feng.cn@outlook.com).

Ran He is with the Aerospace ShenZhou Smart System Technology Company, Ltd., China Academy of Space Technology, Beijing 100095, China (e-mail: heran@spacesystech.com).

Digital Object Identifier 10.1109/JSTARS.2021.3104230

[1] The dataset has been released at https://github.com/zzndream/ShipRSImageNet.



(a) (b)

Fig. 1. Some ship examples, taken from (a) the PASCAL VOC dataset and (b) the proposed ShipRSImageNet dataset.

In addition to different imaging viewpoints, the task of ship detection in satellite-based remote sensing images is also different from that in ground-based images.

1) Ground-based collections in busy ports often have occluded vessels.
2) Orientation of ship instances can be extracted from the bird-eye view.
3) Environmental effects cause different problems, such as occlusions by clouds and sun glint in satellite data.

Several popular large-scale remote sensing datasets, such as DOTA [8], NWPUVHR-10 [9], and DIOR [10], are proposed in earth observation community for general object detection.

Many existing remote sensing datasets, e.g., DOTA [8], are still far from satisfying ship detection tasks since these datasets group all ships into one category. Hence, they can hardly support the detection method to distinguish different class of ship.

General object detection algorithms can distinguish categories with large interclass diversity, such as cats and dogs or airplanes and cars. A more challenging task is fine-grained detection, where subdividing objects belong to the same parent category, such as distinguishing different kinds of birds, different types of ships, and various types of cars. A large-scale dataset with fine-grained category information is necessary for fine-grained classification. Liu *et al.* [11] have compiled high-resolution ship collection 2016 (HRSC2016) to address fine-grained ship classification in remote sensing images, where a set of remote sensing images for ship detection from publicly available high-resolution imagery is given. Afterward, Chen *et al.* [12] released the fine-grained ship detection (FGSD) dataset based on HRSC2016. Nevertheless, the existing publicly available optical remote sensing datasets still have shortcomings for ship detection task in different aspects, which are as follows.

1) The number and diversity of images in the existing datasets are insufficient to support the training of deep learning methods for fine-grained ship categorization. For example, the HRSC2016 dataset contains only 1070 images for 25 categories.
2) Lack of detailed annotations. For example, in the Airbus ship dataset, all ship targets are labeled "ship."
3) The number of ship categories is small, which restricts their applicability for fine-grained classification. For example, there are only four categories in BCCT series datasets [13].
4) Most of the images in existing datasets are collected from the Google Earth database (especially in HRSC and FGSD), which have been preprocessed and manually selected, making the diversity of the images different from that in practical applications.

This work aims at facilitating ship detection research and introduces a large-scale public dataset, namely the ShipRSImageNet. Our proposed dataset contains 3435 images from various sensors, satellite platforms, locations, and seasons. Each image is around 930 × 930 pixels and contains ships with different scales, orientations, and aspect ratios (ARs). The images are annotated by experts in satellite image interpretation, categorized into 50 object categories. The fully annotated ShipRSImageNet contains 17 573 ship instances. There are five critical contributions of the proposed ShipRSImageNet dataset compared with other existing remote sensing image datasets, which are as follows.

1) Images are collected from various remote sensors covering multiple ports worldwide and have large variations in size, spatial resolution, image quality, orientation, and environment.
2) Ships are hierarchically classified into four levels and 50 ship types.
3) The number of images, ship instances, and ship types is larger than that in other publicly available ship datasets. Besides, the number is still increasing.

TABLE I
COMPARISON BETWEEN EIGHT PUBLICLY AVAILABLE OPTICAL REMOTE SENSING OBJECTS DATASETS

Datasets	Image Source	#Images	#Instances /Categories	Image Size	#Ship instances/ #Ship Categories	Spatial resolution	Year
NWPU VHR-10 ([10])	Google Earth	800	3,775/10	1k×1k	302/1	0.5∼2m	2014
WHU-RS19[12]	Google Earth	1,005	-/19	600×600	50/1	0.5m	2012
RSC11 ([17])	Google Earth	,232	-/11	512×512	100/1	0.2m	2016
DOTA ([8])	Google Earth /GF-2/JL-1	2,806	188,282/15	800×800∼4k×4k	2,702/1	max 0.5m	2017
NWPU-RESISC45 ([18])	Google Earth	31,500	31,500/45	256×256	700/1	0.2∼30m	2017
HRRSD ([19])	Google Earth Beidu Map	21,761	55,740/13	-	3,886/1	0.15∼1.2m	2018
DIOR ([10])	Google Earth	23,463	190,288/20	800×800	64,000/1	0.5∼30m	2018
xView ([14])	DigitalGlobe's WorldView 3	1,413	800,636/60	1.5k×1.2k	5,672/9	max 0.3m	2019

TABLE II
COMPARISON OF EIGHT OPTICAL REMOTE SENSING SHIP DETECTION DATASETS

Dataset	**Image source**	**#Images**	**Image Size**	**#Ship Categories**	**#Ship instances**	**Spatial resolution**	**Year**
BCCT200 ([13])	RAPIER Ship Detection System	800	Various size	4	-	<5m	2011
BCCT200-resize ([13])		800	300×150	4	-	<5m	2011
BCCT200-Synth ([16])		200k	300×150	4	-	<5m	2011
HRSC2016 ([11])	Google Earth	1,070	300×300 ∼1.5k×900	25	2,976	0.4-2m	2016
Airbus ship ([20])	SPOT6/7, Pléiades	192,556	768×768	1	231,723	0.3-6m	2019
MASATI ([15])	Microsoft Bing maps	6,212	512×512	1	3,313	-	2019
FGSD ([12])	Google Earth	2,612	930×930	43	5,634	0.12-1.93m	2020
Proposed ShipRSImageNet	Multi sources	3,435	930×930 ∼1.4k×1k	50	17,573	0.12-6m	2020

4) We simultaneously use both horizontal and oriented bounding boxes, and polygons to annotate images, providing detailed information about direction, background, sea environment, and location of targets.
5) We have benchmarked several state-of-the-art object detection algorithms on ShipRSImageNet, which can be used as a baseline for future ship detection methods.

II. REVIEW ON REMOTE SENSING IMAGE DATASETS FOR SHIP DETECTION

We investigate eight publicly available optical remote sensing object datasets containing ship targets, including DOTA, DIOR, and xView [14], etc. As shown in Table I, these datasets contain different ship instances; however, they are often labeled as the same ship-parent category. Unlike other datasets, xView is the only generic optical remote sensing image dataset with various ship categories.

We also investigate seven publicly available ship detection datasets in optical remote sensing images, including BCCT200 [13], HRSC2016, FGSD, etc., for detection tasks and MASATI [15] (see Fig. 2) for the classification tasks, as shown in Table II. Except the Airbus ship dataset that employs images from satellite platforms directly (see Fig. 3), all other datasets collect images from public image resource platforms, such as Google Earth and Microsoft Bing Maps, which causes insufficient diversity and variation in terms of resolution and environment. Many of these images were selected to minimize cloud coverage, to have good illumination conditions, and

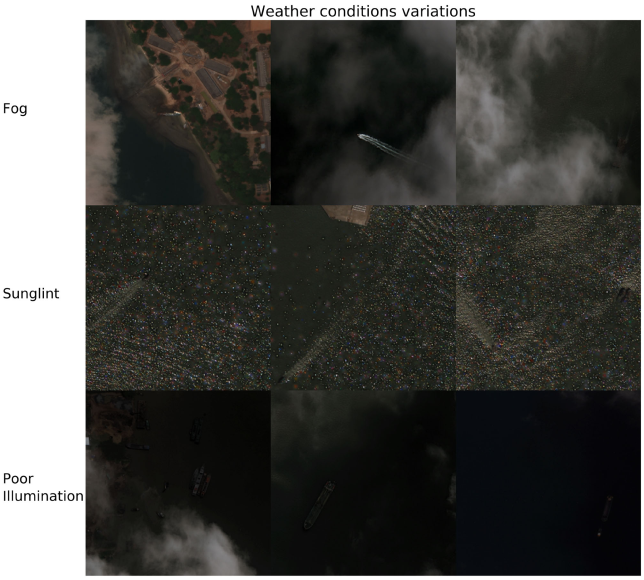


Fig. 2. Image examples of different classes from MASATI dataset.

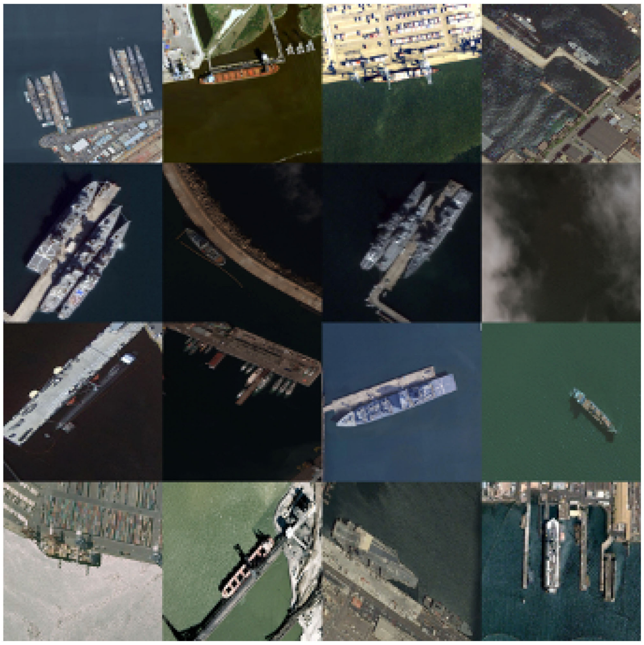


Fig. 3. Samples of Airbus ship detection dataset.

sea surface. This selection would negatively impact training detection techniques that are intended to operate in real-life environmental conditions.

Another problem that needs to be addressed is that weather conditions and scenes are not labeled in existing ship detection datasets. Therefore, it is impossible to evaluate performance of ship detection in different environmental conditions and scenes.

A. BCCT200 Series Datasets

In 2011, Space and Naval Warfare Systems Command proposed a vessel objects dataset using optical remote sensing images, namely, BCCT200. As far as we know, BCCT200 is the earliest public ship detection dataset, which contains four different classes of ships with various image sizes, including barge, cargo, container, and tanker, with 200 images per image category. Based on the BCCT200 dataset, BCCT200-resize dataset was created by rotating, resizing, and aligning BCCT200 images. Besides, another BCCT200 series dataset, BCCT-Synth, consists of 200 000 labeled images of the same ship classes as BCCT200, captured from 15 virtual overhead sensors with different illumination conditions, weather, sea states, and clouds [16].

B. HRSC2016 Dataset

Liu *et al.* [11] released the HRSC2016 dataset in 2016, a milestone in the ship detection research community. It is the first large-scale, detailed classified high spatial resolution optical remote sensing image dataset for ship detection. The HRSC2016 dataset contains 1070 images from Google Earth, and a total number of 2976 ship instances of 25 categories. Before HRSC2016, the instances are only categorized into background and ship or few categories in most ship-detection-related works. Liu *et al.* organized the ship images into a tree structure, which consists of three levels: ship class, ship category, and ship type. The image size changes from 300 × 150 to 1500 × 900, as shown in Fig. 4.

However, there is much room for the HRSC2016 dataset to improve because of the following.

1) Diversity of images collected from Google Earth is not enough.
2) Each category contains 144 instances on average, but the number of instances in most types is less than 50.
3) Only limited ports are covered in the dataset.
4) Ships are grouped into general categories without further finer categorization.
5) Variations of rotation, position, shape, and sea surface are not enough to support practical applications.

C. FGSD Dataset

The FGSD is a new large-scale ship detection dataset released by Chen *et al.* [12]. It is expanded based on HRSC2016 and contains 2612 images from Google Earth and 17 different ports. It consists of a total number of 5634 instances, covers 43 classes. All the instances in the FGSD are annotated manually with both horizontal bounding box (HBB) and oriented bounding box. Compared with HRSC2016, besides ship category, a new class, dock, is annotated in the FGSD. The 42 classes of ships (except for dock) were divided into a three-level category structure, similar to the HRSC2016 dataset. Although the FGSD dataset almost doubles the number of images, it also doubles the number of classes. The average number of instances per class changed only from 119 to 131.

Therefore, for most deep-learning-based methods, the number of images in these datasets to represent varying scale and appearance of ship of different categories is still not enough. Moreover, HRSC2016 and FGSD mainly focus on military ship targets, but there are many categories of merchant ships.

III. Proposed ShipRSImageNet Dataset

A. Images Collection

The ShipRSImageNet dataset collects images from a variety of sensor platforms and datasets, in particular, the following are given.

1) Images of the xView dataset are collected from WorldView-3 satellites with 0.3-m ground resolution. Images in xView are pulled from a wide range of geographic locations. We only extract images with ship targets from them. Since the image in xView is huge for training, we slice them into 930 × 930 pixels with 150 pixels overlap to produce 532 images and relabeled them with both HBB and oriented bounding box.
2) We also collect 1057 images from HRSC2016 and 1846 images from FGSD datasets, corrected the mislabeled and relabeled missed small ship targets.
3) 21 images from the Airbus Ship Detection Challenge.
4) 17 images from Chinese satellites, such as GaoFen-2 and JiLin-1.

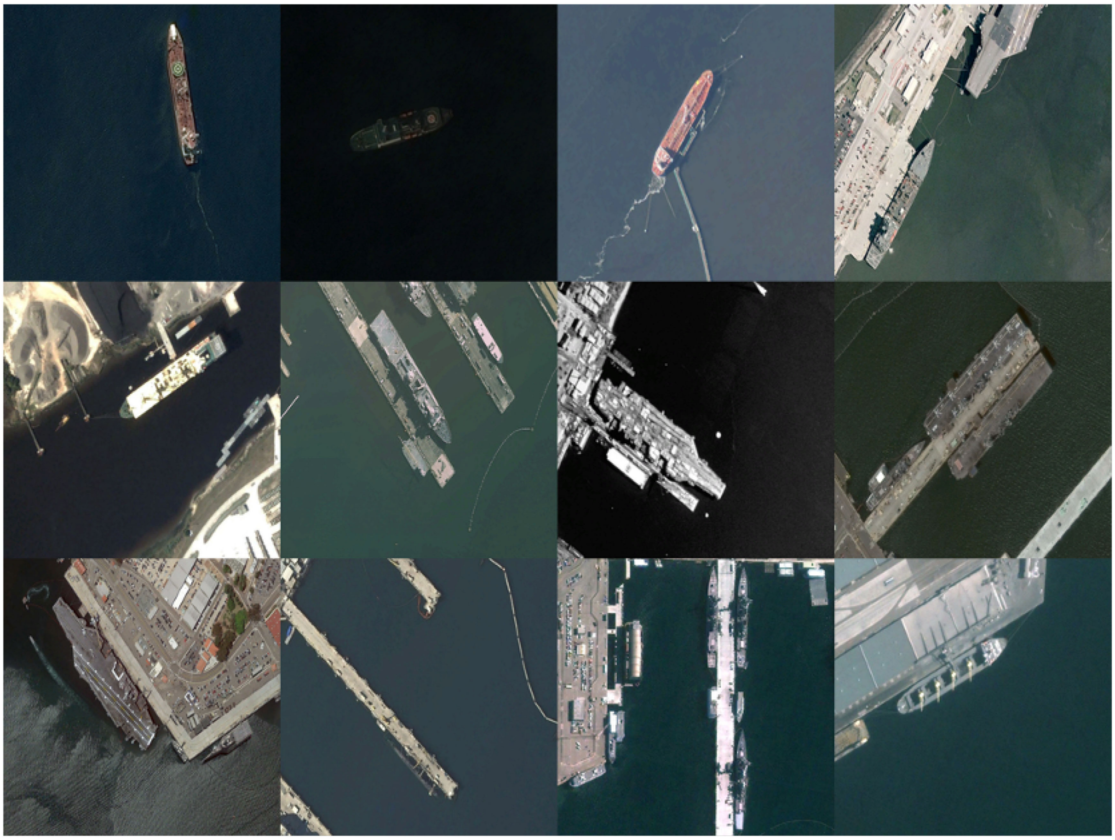


Fig. 4. HRSC2016 dataset samples.

B. Ship Category Selection

Determining which type of ship should be included in our dataset is the first step for constructing ShipRSImageNet, and significantly impacts the performance of the ship detection algorithm. For the existing optical remote sensing datasets, e.g., DOTA, DIOR, the categories are selected based on the value for applications. The interclass distances between these categories are often significant to be distinguished. For example, the DOTA dataset contains 16 categories, including plane, ship, large vehicle helicopter, swimming pool, etc., which are sharply different from each other.

However, for fine-grained categorization, detailed features between different types of ships are small. Fine-grained ship detection is necessary for practical applications. Detecting a ship as an *Arleigh Burke Destroyer* gives more valuable information than detecting it as a *Destroyer*. However, it is difficult to distinguish between different types of ships, such as *Destroyer* and *Frigate*, or *Cargo* and *Container* through optical remote sensing satellite images without expert information.

We inherit the classification system of the HRSC2016 and the FGSD datasets but increase the number of levels in ship categorization hierarchy from 3 to 4 and supplement the subcategory division. The ship classification tree of proposed ShipRSImageNet is shown in Fig. 5. Level 0 distinguish whether the object is a ship, namely *Class*. Level 1 further classifies the ship object category, named as *Category*. Level 2 further subdivides the categories based on level 1. For example, *Warship* is subdivided into *Submarine, Aircraft Carrier, Destroyer, Frigate*, etc., namely the Subcategory. Level 3 is the specific type of ship, named as *Type*. For example, the *Aircraft Carrier* subcategory includes *Nimitz, Enterprise*, etc. However, it needs to be pointed out that for *Merchant ship*, the number of specific types is huge and have limit value to be classified in practice, so they are not further divided at level 3.

At level 0, two categories are labeled, *Ship* and *Dock*, which means all positive targets are classified as *Ship*. *Dock* does not belong to any ship category but is labeled because the rectangular shape of docks influences near-shore ship detection performance.

At level 1, ships are labeled into three categories: *Warship, Merchant ship*, and *Other ship*. *Other ship* means an object is a ship, but we cannot identify whether it is a *Warship* or *Merchant ship*. This standard also applies to level 2 and level 3.

At level 2, we subdivide *Warship* into ten subcategories. There is no unified standard for the classification of military ships globally for *Warship*, especially for modern warships. It is challenging to distinguish modern warships in the optical remote sensing image from bird-eye view. For example, modern *Frigate* and *Destroyer* are very similar, with only little differences in size; sometimes, even experts cannot easily distinguish between them. We added the *Auxiliary Ship* category at level 2 as a new subcategory. The *Medical Ship, Test Ship*, and *Training Ship* in FGSD are all subcategories of *Auxiliary Ship*. We also carefully

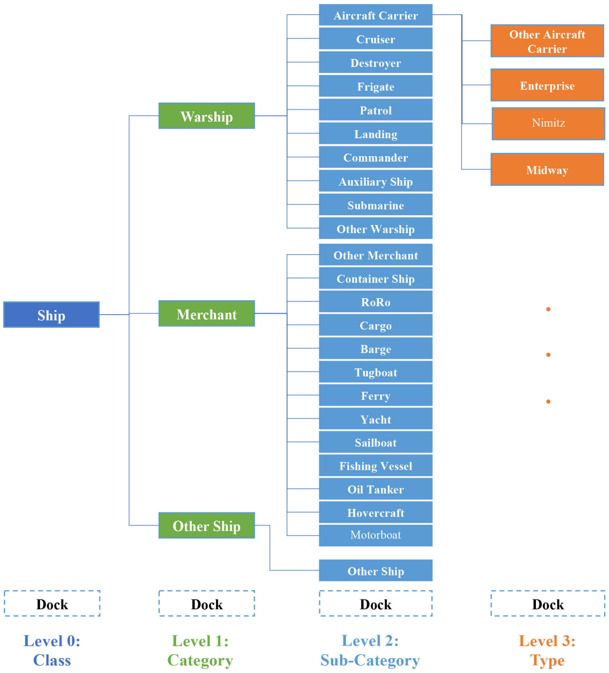


Fig. 5. ShipRSImageNet class hierarchy.

choose 13 subcategories for *Merchant ship*, which are more accurate than those given in the FGSD and HRSC2016 2016 datasets.

Finally, at level 3, ship objects are divided into 50 types. For brevity, we use the following abbreviations: *DD* for *Destroyer*, *FF* for *Frigate*, *LL* for *Landing*, *AS* for *Auxiliary Ship*, *LSD* for *Landing Ship Dock*, *LHA* for *Landing Helicopter Assault Ship*, *AOE* for *Fast Combat Support Ship*, *EPF* for *Expeditionary Fast Transport Ship*, and *RoRo* for *Roll-on Roll-off Ship*. These 50 object classes are *Other Ship, Other Warship, Submarine, Other Aircraft Carrier, Enterprise, Nimitz, Midway, Ticonderoga, Other Destroyer, Atago DD, Arleigh Burke DD, Hatsuyuki DD, Hyuga DD, Asagiri DD, Other Frigate, Perry FF, Patrol, Other Landing, YuTing LL, YuDeng LL, YuDao LL, YuZhao LL, Austin LL, Osumi LL, Wasp LL, LSD 41 LL, LHA LL, Commander, Other Auxiliary Ship, Medical Ship, Test Ship, Training Ship, AOE, Masyuu AS, Sanantonio AS, EPF, Other Merchant, Container Ship, RoRo, Cargo, Barge, Tugboat, Ferry, Yacht, Sailboat, Fishing Vessel, Oil Tanker, Hovercraft, Motorboat*, and *Dock*.

We assigned a specific code with four fields for each of the 50 categories, each with two digits. The first field describes the category of level 0, the second field describes the category of level 1, the third field describes the subcategory of level 2, and the fourth field describes the type of level 3, respectively. As an example, *Dock* would have the code *02-04-25-50*. The codes and abbreviations of ship categories are shown in Table III.

It should be pointed out that the number of types can be expanded easily in level 3, whereas the subcategories in level 2 should be kept the same to maintain consistency of ship categorization in the dataset.

C. Dataset Scale and Splits

ShipRSImageNet dataset contains 3435 optical remote sensing images with 17 573 ship instances, including 49 ship types and *Dock*. The size of most original images in the dataset is 930 × 930, the maximum width is 1238, and the maximum height is 930.

For evaluation, we partitioned the ShipRSImageNet dataset into train and test splits. We randomly select 64% of the original images as the training set, 16% as a validation set, and 20% as the testing set. This resulted in 2198 images for training, 550 images for validating, and 687 images for testing, respectively. For each ship category and subset, the number of instances is given in Table III.

D. Annotation Pipeline

Motivated by existing datasets, such as MS-COCO, we design the annotation pipeline as the following three steps. The first step performs a two-class classification. If the image contains at least one ship instance, it will be passed to the next step. In the second step, each image is labeled as containing particular ship categories using a hierarchical labeling approach. In the third step, the individual ship instances are labeled with the location information and weather characteristics.

Our annotation team has three groups: annotators, inspectors, and examiners. Images were divided into subsets and were annotated by the annotators, then checked by the inspectors. Finally, the examiners review the quality of annotations. The image annotation job is rejected if any error is found in any of the steps.

E. Annotation Method

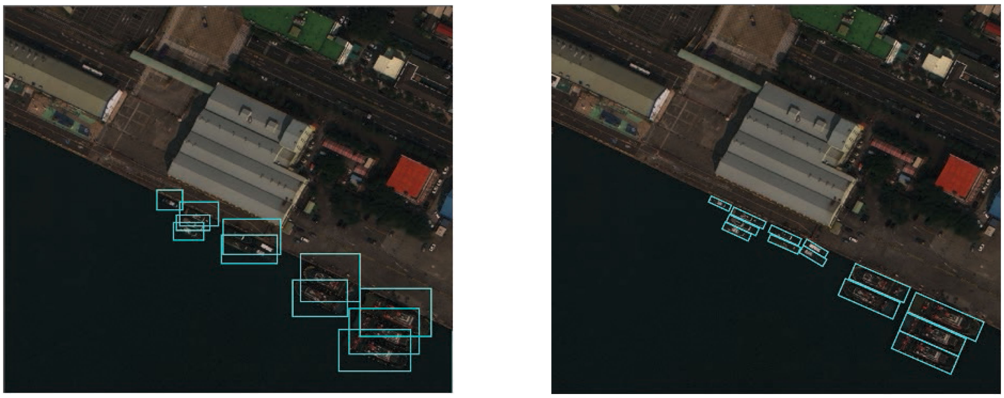
A typical HBB is presented with (x_c, y_c, w, h), where (x_c, y_c) is the central location of the target, w and h are the width and height of the surrounded bounding boxes, respectively. However, the HBB cannot accurately represent ship targets because of the overlap between the HBBs, especially when ships are densely distributed, as shown in Fig. 6.

To address this problem, Liu *et al.* have proposed to replace the HBBs with oriented bounding boxes [21]. In this way, the overlap between bounding boxes is minimized [22]. An oriented bounding box is characterized with five tuples (x_c, y_c, w, h, θ), where (x_c, y_c) is the central location, (w, h) is the width and height of the surrounded orientation bounding box, respectively, and θ denotes the angle between the oriented bounding box and the horizontal direction. Another equivalent representation method is to use the coordinates of the four corners of the oriented rectangular box, similar to the polygon labeling method, which can be denoted as $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$.

In our dataset, each ship instance is manually labeled by experts with a HBB, an oriented bounding box, as well as a polygon annotation. We used the (x_c, y_c, w, h, θ) method to represent the oriented bounding box. However, θ is defined as the angle at which the horizontal axis (*X*-axis) rotates counterclockwise to the head vector of the ship since the head position has a particularly significant value for identifying the direction in the

TABLE III
NUMBER OF INSTANCES PER SHIP CATEGORY AND PER SUBSET

Type	Code: level 0-1-2-3	Abbv.	Total	Train	Val	Test	Color Code: RGB
Other Ship	01-01-01-01	c1	1696	1050	297	349	(66, 77, 59)
Other Warship	01-02-02-02	c2	1455	962	209	284	(127, 224, 99)
Submarine	01-02-03-03	c3	1017	666	171	180	(17, 104, 104)
Other Aircraft Carrier	01-02-04-04	c4	46	30	4	12	(24, 135, 138)
Enterprise	01-02-04-05	c5	101	69	18	14	(233, 127, 10)
Nimitz	01-02-04-06	c6	115	84	19	12	(222, 214, 167)
Midway	01-02-04-07	c7	23	13	5	5	(53, 90, 30)
Ticonderoga	01-02-05-08	c8	440	258	77	105	(57, 2, 47)
Other Destroyer	01-02-06-09	c9	277	171	46	60	(66, 149, 21)
Atago DD	01-02-06-10	c10	272	167	41	64	(103, 135, 239)
Arleigh Burke DD	01-02-06-11	c11	652	396	130	126	(173, 202, 6)
Hatsuyuki DD	01-02-06-12	c12	140	88	22	30	(23, 143, 224)
Hyuga DD	01-02-06-13	c13	108	70	16	22	(113, 178, 123)
Asagiri DD	01-02-06-14	c14	76	48	14	14	(31, 29, 113)
Other Frigate	01-02-07-15	c15	215	140	28	47	(174, 15, 208)
Perry FF	01-02-07-16	c16	659	423	77	159	(85, 43, 116)
Patrol	01-02-08-17	c17	154	102	37	15	(150, 70, 208)
Other Landing	01-02-09-18	c18	108	69	18	21	(193, 139, 220)
YuTing LL	01-02-09-19	c19	101	61	19	21	(126, 222, 16)
YuDeng LL	01-02-09-20	c20	83	53	11	19	(154, 64, 177)
YuDao LL	01-02-09-21	c21	63	40	5	18	(233, 41, 152)
YuZhao LL	01-02-09-22	c22	44	31	6	7	(129, 133, 155)
Austin LL	01-02-09-23	c23	138	76	27	35	(217, 175, 146)
Osumi LL	01-02-09-24	c24	41	28	6	7	(16, 3, 163)
Wasp LL	01-02-09-25	c25	29	14	6	9	(103, 108, 66)
LSD 41 LL	01-02-09-26	c26	145	95	21	29	(160, 159, 111)
LHA LL	01-02-09-27	c27	192	126	31	35	(136, 157, 249)
Commander	01-02-10-28	c28	146	88	32	26	(42, 255, 213)
Other Auxiliary Ship	01-02-11-29	c29	93	60	18	15	(245, 48, 123)
Medical Ship	01-02-11-30	c30	33	22	5	6	(175, 156, 7)
Test Ship	01-02-11-31	c31	55	43	7	5	(38, 156, 133)
Training Ship	01-02-11-32	c32	55	31	11	13	(220, 213, 40)
AOE	01-02-11-33	c33	62	37	11	14	(169, 22, 232)
Masyuu AS	01-02-11-34	c34	50	28	8	14	(246, 74, 50)
Sanantonio AS	01-02-11-35	c35	73	48	13	12	(25, 194, 118)
EPF	01-02-11-36	c36	62	42	10	10	(98, 151, 99)
Other Merchant	01-03-12-37	c37	252	150	50	52	(49, 199, 150)
Container Ship	01-03-13-38	c38	376	232	72	72	(152, 115, 248)
RoRo	01-03-14-39	c39	170	107	20	43	(151, 41, 140)
Cargo	01-03-15-40	c40	1082	657	169	256	(12, 34, 202)
Barge	01-03-16-41	c41	239	161	22	56	(59, 91, 227)
Tugboat	01-03-17-42	c42	290	197	46	47	(249, 85, 231)
Ferry	01-03-18-43	c43	309	191	53	65	(134, 203, 83)
Yacht	01-03-19-44	c44	712	501	140	71	(56, 95, 99)
Sailboat	01-03-20-45	c45	796	325	341	130	(35, 176, 189)
Fishing Vessel	01-03-21-46	c46	606	318	99	189	(172, 18, 93)
Oil Tanker	01-03-22-47	c47	204	129	32	43	(19, 240, 187)
Hovercraft	01-03-23-48	c48	334	229	31	74	(64, 9, 100)
Motorboat	01-03-24-49	c49	2091	1190	398	503	(7, 47, 41)
Dock	02-04-25-50	c50	1093	744	154	195	(255, 25, 180)



(a) (b)

Fig. 6. HBB and oriented bounding box annotation. (a) HBB. (b) Oriented bounding box.

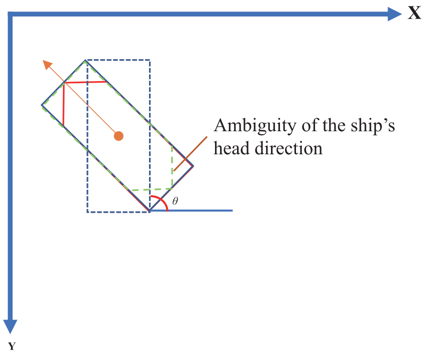


Fig. 7. Annotation method of the orientation bounding box and ship head.

densely distributed situation. When the ship's head points to the X-axis direction, the rotation angle θ is 0°, and the range of θ is $[0, 2\pi]$. Since there are no clear visual clues for some types of ships to determine the head position of the ship, such as a ship with a V-shape at both ends, we choose the top-left point of it as the starting point, which is consistent with the annotation method in DOTA [8] dataset.

Polygon annotations are also provided, where the right point is chosen carefully as the first point of the oriented bounding box, as shown in Fig. 7. Then, we can determine the other three points of an oriented bounding box clockwise. Finally, these four points are denoted as $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ in the polygon field of the ShipRSImageNet annotation file. Examples of annotated patches in ShipRSImageNet are shown in Fig. 8. If the image contains only part of the ship, then we use the HBB to annotate this ship instance and add an additional tag, namely *Truncated*. If a ship object is only partially visible in the image then set *Truncated* to 1, otherwise set *Truncated* to 0.

F. *Characteristics of ShipRSImageNet*

We believe that, up to now, the proposed ShipRSImageNet dataset is one of the largest public datasets in the ship detection and classification community, with the most diversity in spatial resolution, object size, environmental conditions, location, and AR. Compared with existing ship detection datasets, including BCCT, HRSC2016, FGSD, and Airbus Ship Detection, the proposed ShipRSImageNet dataset has differentiating characteristics as follows.

1) Large Variation: The proposed ShipRSImageNet dataset has the largest variation in terms of the number of images, number of instances, number of categories, and average number of instances in each category. Therefore, the release of ShipRSImageNet will help the earth observation community to evaluate different ship detection and recognition algorithms.

Fig. 9 shows the distribution of the number of ship objects in each image of the ShipRSImageNet dataset, where most images contain instances less than 15, and the average number of the ship in each image is 5.12, with a maximum number of 317. Moreover, the average number of instances of each type is 308, with a minimum number of 23.

TABLE IV
Comparison of Instance Pixel Size Distribution of Some Remote Sensing Datasets

Dataset	<1,024 pixel	1,024-9,216 pixel	>9,216 pixel
ShipRSImageNet	0.24	0.36	0.40

2) Variable Spatial Resolution: The spatial resolution of ShipRSImageNet ranges from 0.12 to 6 m, which is a wider variation than in FGSD and other optical remote sensing ship image datasets. In the ShipRSImageNet dataset, we provide the spatial resolution for most images, which could be used to calculate the actual size of a ship instance.

3) Object Size: The size of a ship varies in different ship categories, as is shown in Fig. 10.

Following the definition of the MS-COCO dataset, we divide all the instances in our dataset into three splits according to their area: small for the range of pixel area smaller than 1024, middle for range from 1024 to 9216, and large for range above 9216. Table IV illustrates the percentages of these three instances splits in our dataset. ShipRSImageNet dataset achieves a good balance between small, middle, and large instances, which helps deep-learning-based detectors to capture the various size of ships.

Furthermore, the size of the ship in different categories is also very different. Fig. 11 gives the distribution of pixel area in each category.

4) Variable Sources: We used eight different remote sensing platforms as image sources, and images were captured at different locations, seasons all around the world with different spatial resolution and image quality. Besides ships on the open sea, images are collected from 58 ports worldwide, which outperforms six ports in the HRSC2016 dataset and 17 ports in FGSD dataset.

5) Variable Weather Conditions and Scenes: As weather conditions impact the performance of ship detection, we provide information on weather, including whether there are clouds or fog, strong sea reflection, and whether the light is insufficient (see Fig. 12). Yang *et al.* show ship detection accuracy decreases drastically from a quiet sea to a cluttered sea [23]. We have added weather-related tags, namely *Ship _ env_is _ fog*, *Ship _ env_is _ glint*, and *Ship _ env_is _ dark*. Depending on the presence or lack of clouds, reflection, and light, the tags are set to 1 or 0, respectively.

Moreover, the spatial context also impacts detector performance significantly. For example, annotation is provided if the ships are on the sea, near the shore, or on the river. In this case, ShipRSImageNet annotates the location of the ship: sea, river, near shore, and land.

6) AR of Instances Variations: The AR of ship objects is an important feature, often ranges from 3 to 6 and is used as one of the main features in ship classification and recognition tasks. For deep-learning-based methods, the AR is a critical parameter for designing anchor-based models, such as Faster RCNN. Fig. 13 illustrates the ship aspect ratio distribution in ShipRSImageNet.

7) Variable Orientations of Instances: The orientation of ship objects ranges from 0 to 2π, the distribution of rotation angle illustrated in Fig. 14.

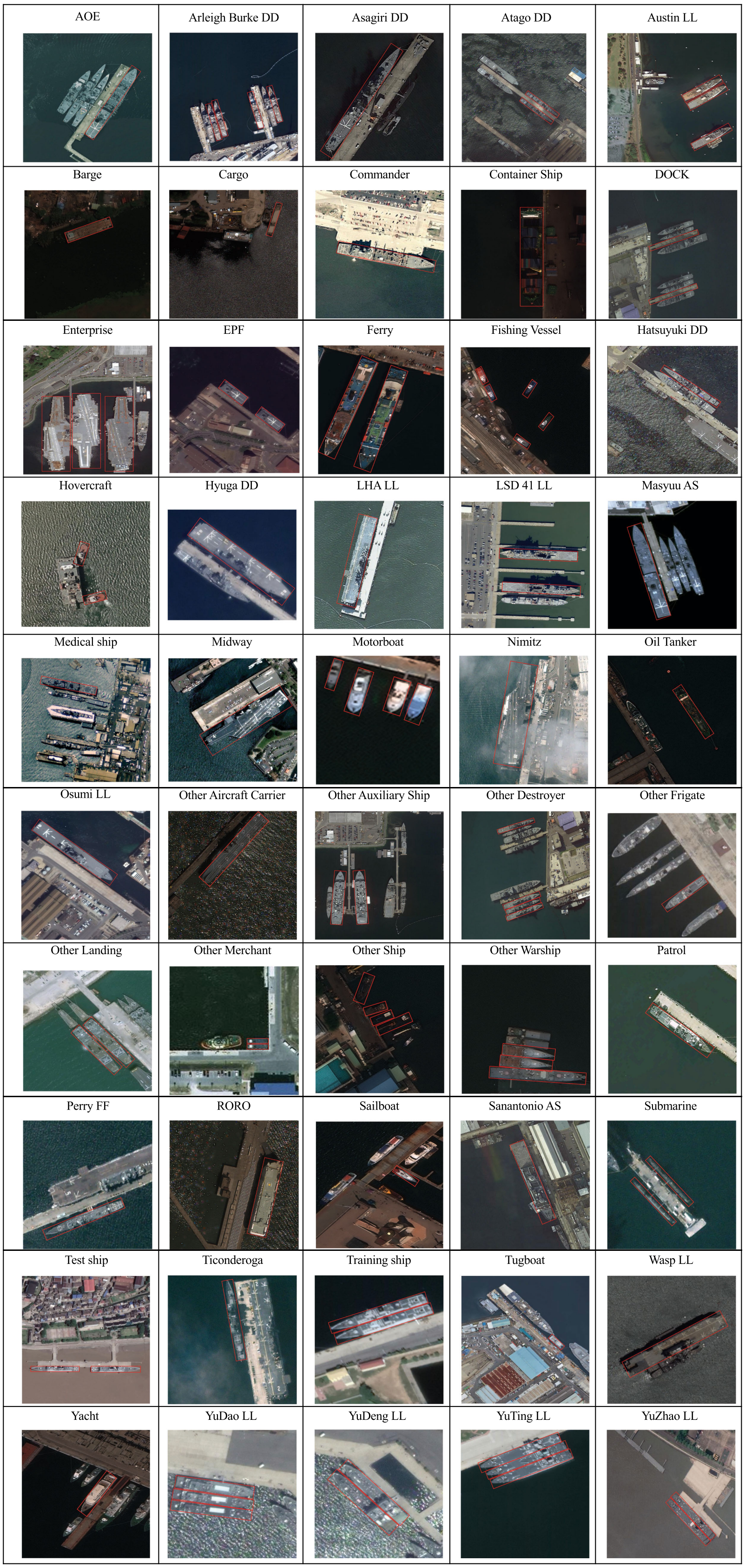


Fig. 8. Samples of annotated images in ShipRSImageNet.

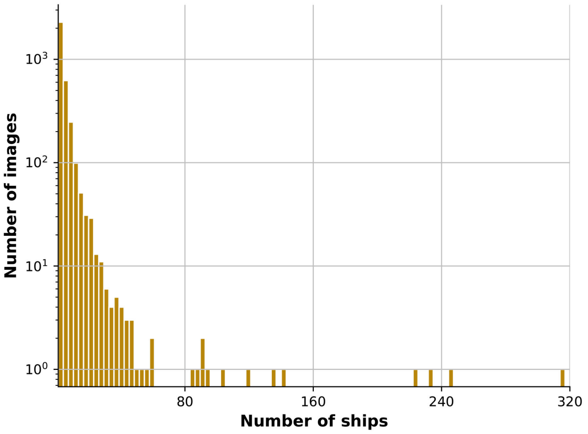


Fig. 9. Ship count distribution.

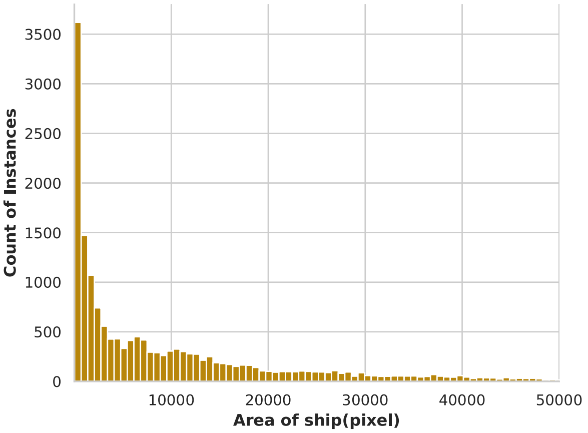


Fig. 10. Ship area distribution.

IV. Benchmark for Ship Detection

To evaluate state-of-the-art object detection methods on the ShipRSImageNet, we divided the dataset into a training set, validation set, and test set according to the ratio given in Section III-C. We conducted all experiments on the computer with a single Intel Xeon E5-2667 CPU, 32 GB of memory, and two NVIDIA Tesla V100 GPU with 24-GB display memory for deep learning acceleration.

A. *Experimental Setup*

To create an evaluation baseline, we used deep learning algorithms, such as the following:

1) three region-proposal-based approaches: Faster RCNN [24], Mask RCNN [25], and Cascade Mask RCNN [26];
2) two regression-based methods: SSD [22] and RetinaNet [27]; and
3) two anchor-free methods: FoveaBox [28] and FCOS [29].

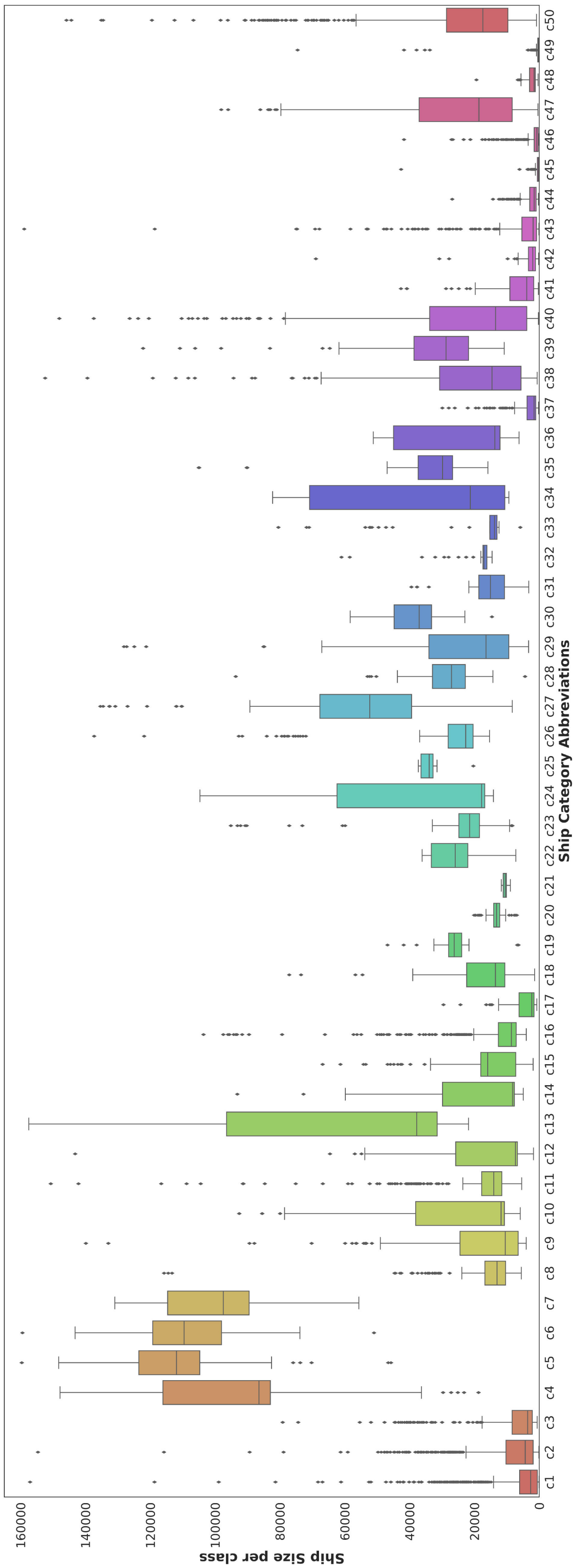


Fig. 11. Object pixel size distribution per type. The short names for types are defined in Table III.

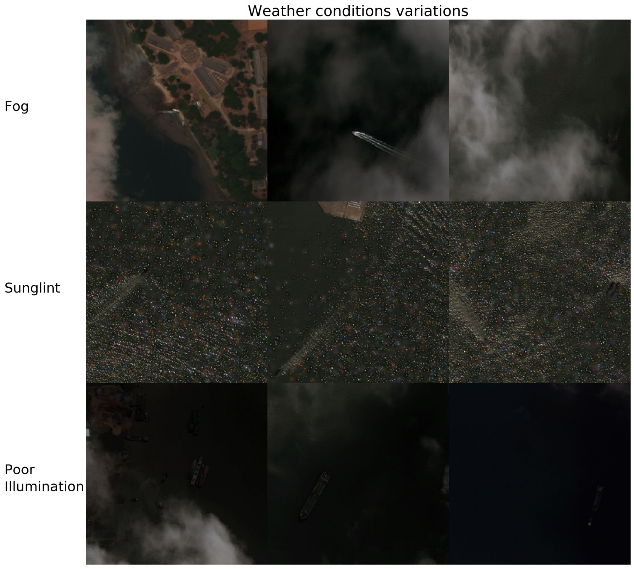


Fig. 12. Weather conditions' variations.

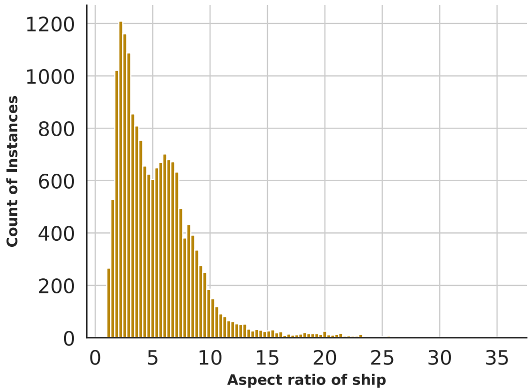


Fig. 13. AR distribution of oriented bounding box.

TABLE V
ANCHOR SETUP PARAMETERS OF FIVE REPRESENTATIVE METHODS

Methods	Anchor scales	Anchor ratios	Anchor strides
Faster RCNN	8	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
Mask RCNN	8	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
Cascade Mask RCNN	8	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
SSD	-	[[2], [2, 3], [2, 3], [2, 3], [2], [2]]	[8, 16, 32, 64, 100, 300]
RetinaNet	-	[0.5, 1.0, 2.0]	[8, 16, 32, 64, 128]

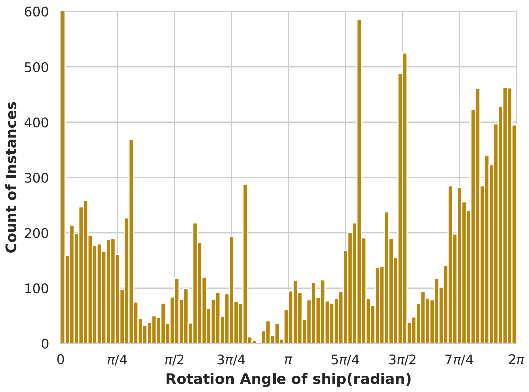


Fig. 14. Orientations distribution of instances.

TABLE VI
SEGMENTATION AVERAGE PRECISION AND RECALL OF TWO REPRESENTATIVE METHODS

Task Level	Metric	Mask RCNN with ResNet-50-FPN	Cascade Mask R-CNN-ResNet50
0	mAP	**0.440**	**0.430**
	mAP of area small	0.063	**0.065**
	mAP of area medium	**0.472**	0.467
	mAP of area large	**0.544**	0.542
	mAR	**0.507**	0.495
1	mAP	0.347	**0.365**
	mAP of area small	0.081	**0.083**
	mAP of area medium	0.393	**0.438**
	mAP of area large	0.414	**0.425**
	mAR	0.456	**0.463**
2	mAP	0.377	**0.389**
	mAP of area small	**0.145**	0.140
	mAP of area medium	0.337	**0.341**
	mAP of area large	0.393	**0.415**
	mAR	**0.495**	0.493
3	mAP	0.450	**0.483**
	mAP of area small	**0.151**	0.132
	mAP of area medium	**0.349**	**0.333**
	mAP of area large	0.463	**0.498**
	mAR	0.561	**0.577**

Note: The entries with the best APs and ARs for each level are bold-faced. The task level means ship detection task from level 0 to level 3.

ResNet-50 with feature pyramid network (FPN) or ResNet-101 with FPN are used as backbone network for feature extraction to make comparison, as shown in Table VII.

We keep all the experiment settings and hyperparameters the same, as depicted in MMDetection (v2.11.0) config files except for the number of categories and parameters. MMDetection is an open-source object detection toolbox based on PyTorch. It is a part of the OpenMMLab project developed by Multimedia Laboratory, CUHK [30].

Experiments are implemented on four different levels of the proposed ShipRSImageNET, and the weights of the backbone network pretrained on ImageNet and loaded. For more details, the input images and their annotations are rescaled to 800 × 1333 during the augmentation pipeline, which is the same in the training process and testing process; the batch size is 4. The learning rate initialized to 0.005, and it is decreased exponentially in each epoch. The number of iterations has been fixed to 100 epochs, as it was sufficient for these five networks to converge to a good solution. For Faster RCNN, Mask RCNN, and Cascade Mask RCNN network, the anchor scales are set to 8, with three different ratios: [0.5, 1.0, 2.0], and five different strides that are [4, 8, 16, 32, 64]. For SSD network, the anchor strides are set to [8, 16, 32, 64, 100, 300]. For RetinaNet, the anchor strides is set to [8, 16, 32, 64, 128], as shown in Table V for detail.

B. Evaluation Measurements

The precision–recall curve and average precision are two widely used measures in a lot of object detection works. The precision metric is defined as the number of true positives (TPs)

TABLE VII
DETECTION OF AVERAGE PRECISION AND RECALL (@IoU0.50:0.95) OF EIGHT REPRESENTATIVE METHODS

Task Level	Metric	Faster RCNN with FPN -ResNet50	Mask RCNN with FPN -ResNet50	Cascade Mask RCNN -ResNet50	SSD -VGG16	Retinanet with FPN -ResNet50	Retinanet with FPN -ResNet101	FCOS with FPN -ResNet101	FoveaBox-ResNet101
Ritinanet 0	mAP	0.550	0.440	**0.568**	0.464	0.418	0.419	0.333	0.453
	mAP of area small	0.077	0.063	**0.091**	0.069	0.067	0.460	0.044	0.071
	mAP of area medium	0.569	0.472	**0.595**	0.529	0.493	0.501	0.468	0.502
	mAP of area large	0.682	0.544	**0.702**	0.580	0.520	0.526	0.444	0.567
	mAR	0.605	0.507	**0.617**	0.543	0.543	0.548	0.480	0.535
1	mAP	0.366	0.456	**0.485**	0.397	0.368	0.359	0.351	0.389
	mAP of area small	0.074	0.115	0.131	0.111	0.228	0.086	0.089	**0.240**
	mAP of area medium	0.383	0.513	**0.561**	0.490	0.458	0.458	0.475	0.473
	mAP of area large	0.467	0.537	**0.559**	0.467	0.446	0.438	0.405	0.463
	mAR	0.460	0.557	**0.574**	0.506	0.519	0.503	0.508	0.488
2	mAP	0.455	0.377	**0.492**	0.423	0.369	0.411	0.431	0.427
	mAP of area small	0.191	0.145	**0.204**	0.173	0.190	0.194	0.195	0.186
	mAP of area medium	0.421	0.337	**0.446**	0.402	0.375	0.397	0.425	0.387
	mAP of area large	0.498	0.393	**0.531**	0.467	0.371	0.410	0.438	0.453
	mAR	0.572	0.495	0.599	0.551	0.566	**0.603**	0.595	0.556
3	mAP	0.375	0.545	**0.593**	0.483	0.326	0.483	0.498	0.459
	mAP of area small	0.196	0.191	0.184	0.220	0.108	0.172	0.164	**0.252**
	mAP of area medium	0.406	**0.467**	0.456	0.406	0.350	0.424	**0.467**	0.401
	mAP of area large	0.559	0.563	**0.612**	0.520	0.320	0.488	0.513	0.478
	mAR	0.647	0.661	**0.695**	0.618	0.561	0.689	0.674	0.622

Note: The entries with the best APs for each level are bold-faced.

divided by the sum of TPs and false positives (FPs)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{1}$$

The recall metric is defined as the number of TPs divided by the sum of TPs and false negatives (FNs)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2}$$

Intersection over Union (IoU) is defined as the area of the intersection divided by the area of the union of a predicted bounding box (B_p) and a ground-truth box (B_{gt})

$$\text{IoU} = \frac{\text{area}(B_p \bigcap B_{\text{gt}})}{\text{area}(B_p \cup B_{\text{gt}})}. \tag{3}$$

AP is the precision averaged across all unique recall levels. Average recall (AR) is the recall averaged over all IoU $\in [0.5, 1.0]$. The calculation of AP only involves one class. However, in object detection, there are usually K classes. Mean average precision (mAP) is defined as the mean of AP across all K classes

$$\text{mAP} = \frac{\sum_{i=1}^{K} \text{AP}_i}{K}. \tag{4}$$

Mean AR is defined as the mean of AR across all K classes

$$\text{mAR} = \frac{\sum_{i=1}^{K} \text{AR}_i}{K}. \tag{5}$$

Several mAP metrics are defined using different thresholds following MS-COCO, including the following:

1) mAP(@IoU=0.50:0.95), which is mAP averaged over ten IoU thresholds for all the ship categories;
2) mAP(@IoU=0.50), which is a strict metric, using IoU=0.50 as the threshold; and
3) mAP(@IoU=0.75), which is a strict metric, using IoU=0.75 as the threshold.

In addition to different IoU thresholds, COCO challenge also defines mAP calculated across different object scales, and these variants of mAP are all averaged over ten IoU thresholds(@IoU=0.50:0.95), which are as follows:

1) mAP of area small, which is mAP for small ship objects that covers area less than 1024 pixel;
2) mAP of area medium, which is mAP for medium ship objects that covers area greater than 1024 but less than 9216 pixel; and
3) mAP of area large, which is mAP for large ship objects that covers area greater than 9216 pixel.

We applied deep learning algorithms to: detection with HBBs and segmentation with oriented bounding boxes (SBB for short). HBB aims at extracting bounding boxes with the same orientation of the image, it is an object detection task. SBB aims at semantically segmenting the image, it is a semantic segmentation task. We also performed detection and classification at four levels: from level 0 to level 3, as defined in Section III-B. For evaluation metrics, we adopt the same mAP of area small/medium/large and mAP(@IoU=0.50:0.95) calculation following MS-COCO. We adopt the same mAR(max=100) metric that is used by MS COCO, which is mAR given 100 detections per image and averaged over 10 IoU thresholds (IoU=0.50:0.95).

C. Experimental Results

In Fig. 15, we show the testing results of HBB and SBB tasks on the ShipRSImageNet using Cascade Mask R-CNN. The average detection precision and an AR of different level tasks on the test set are shown in Tables VI and VII. The average detection precision (%) for each category in the level 3 ship detection task is shown in Table VIII. Three categories of the detectors are shown in Tables VII and VIII. They include: two-stage detectors

TABLE VIII
DETECTION AVERAGE PRECISION (@IOU0.50:0.95) OF EIGHT REPRESENTATIVE METHODS ON THE PROPOSED SHIPRSIMAGENET IN LEVEL 3 TASK

Methods		**Backbone**	**c1**	**c2**	**c3**	**c4**	**c5**	**c6**	**c7**	**c8**	**c9**	**c10**	
Two stage	Faster RCNN with FPN	ResNet-50	0.087	0.207	0.350	0.640	0.633	0.488	0.517	0.553	0.226	0.393	
		ResNet-101	0.209	0.332	0.601	0.647	0.728	0.820	0.671	0.779	0.333	0.511	
	Mask RCNN with FPN	ResNet-50	0.256	0.338	0.599	0.559	0.709	0.764	0.744	0.768	0.336	0.530	
		ResNet-101	0.241	0.375	**0.614**	0.627	**0.767**	0.803	0.672	0.760	0.334	0.504	
	Cascade Mask RCNN with FPN	ResNet-50	**0.271**	**0.386**	0.601	0.740	0.701	**0.861**	**0.809**	**0.807**	**0.410**	**0.613**	
One stage	SSD	VGG16	0.223	0.305	0.522	**0.742**	0.723	0.707	0.684	0.634	0.286	0.363	
	RetinaNet with FPN	ResNet-50	0.138	0.195	0.349	0.471	0.565	0.560	0.308	0.614	0.154	0.433	
	RetinaNet with FPN	ResNet-101	0.230	0.302	0.572	0.730	0.618	0.792	0.738	0.753	0.252	0.445	
Anchor-free	FoveaBox	ResNet-101	0.188	0.292	0.562	0.410	0.685	0.659	0.288	0.719	0.318	0.491	
	FCOS with FPN	ResNet-101	0.208	0.323	0.564	0.551	0.685	0.674	0.676	0.722	0.248	0.434	
Methods		**Backbone**	**c11**	**c12**	**c13**	**c14**	**c15**	**c16**	**c17**	**c18**	**c19**	**c20**	
Two stage	Faster RCNN with FPN	ResNet-50	0.625	0.327	0.866	0.392	0.284	0.608	0.098	0.095	0.192	0.189	
		ResNet-101	0.821	0.634	0.880	0.307	0.426	0.730	0.354	0.146	0.447	0.260	
	Mask RCNN with FPN	ResNet-50	0.832	0.675	0.888	0.512	**0.355**	0.733	0.256	0.120	0.448	0.386	
		ResNet-101	0.849	0.630	0.892	0.551	0.380	**0.781**	**0.459**	**0.096**	0.453	0.283	
	Cascade Mask RCNN with FPN	ResNet-50	**0.854**	**0.724**	**0.942**	**0.615**	**0.433**	0.771	0.401	**0.156**	**0.525**	0.287	
One stage	SSD	VGG16	0.698	0.510	0.855	0.516	0.397	0.662	0.358	0.088	0.374	0.266	
	RetinaNet with FPN	ResNet-50	0.646	0.229	0.772	0.205	0.065	0.576	0.127	0.060	0.119	0.169	
	RetinaNet with FPN	ResNet-101	0.767	0.579	0.896	0.308	0.403	0.670	0.320	0.051	0.377	0.337	
Anchor-free	FoveaBox	ResNet-101	0.763	0.553	0.842	0.151	0.334	0.673	0.336	0.055	0.284	0.397	
	FCOS with FPN	ResNet-101	0.783	0.629	0.865	0.380	0.331	0.690	0.346	0.072	0.507	**0.428**	
Methods		**Backbone**	**c21**	**c22**	**c23**	**c24**	**c25**	**c26**	**c27**	**c28**	**c29**	**c30**	
Two stage	Faster RCNN with FPN	ResNet-50	0.000	0.543	0.430	0.872	0.914	0.469	0.579	0.414	**0.276**	0.049	
		ResNet-101	0.534	**0.913**	0.649	0.961	0.895	**0.740**	0.730	0.804	0.363	0.570	
	Mask RCNN with FPN	ResNet-50	0.366	0.787	0.718	0.944	**0.980**	0.667	0.761	0.808	0.375	0.608	
		ResNet-101	**0.722**	0.858	0.625	**1.000**	0.938	0.618	**0.810**	0.753	0.381	0.550	
	Cascade Mask RCNN with FPN	ResNet-50	0.583	0.909	**0.735**	0.983	**0.953**	**0.759**	0.799	**0.832**	**0.420**	**0.699**	
One stage	SSD	VGG16	0.360	0.797	0.556	0.860	0.900	0.545	0.727	0.621	0.340	0.299	
	RetinaNet with FPN	ResNet-50	0.000	0.422	0.307	0.888	0.792	0.195	0.625	0.351	0.101	0.019	
	RetinaNet with FPN	ResNet-101	0.628	0.792	0.434	0.955	0.811	0.326	0.803	0.433	0.157	0.263	
Anchor-free	FoveaBox	ResNet-101	0.516	0.762	0.463	0.966	0.909	0.417	0.761	0.479	0.381	0.222	
	FCOS with FPN	ResNet-101	0.478	0.897	0.612	0.929	0.823	0.477	0.726	0.652	0.339	0.405	
Methods		**Backbone**	**c31**	**c32**	**c33**	**c34**	**c35**	**c36**	**c37**	**c38**	**c39**	**c40**	
Two stage	Faster RCNN with FPN	ResNet-50	**0.142**	0.291	0.299	0.780	0.442	0.540	0.028	0.401	0.533	0.471	
		ResNet-101	0.559	0.670	0.499	0.818	0.734	0.682	0.059	0.513	**0.749**	0.574	
	Mask RCNN with FPN	ResNet-50	**0.583**	0.702	0.557	0.815	0.752	0.687	0.042	0.495	0.654	0.564	
		ResNet-101	0.522	**0.889**	0.414	0.889	0.689	**0.770**	0.042	**0.579**	0.734	**0.614**	
	Cascade Mask RCNN with FPN	ResNet-50	0.536	0.848	**0.676**	**0.962**	**0.815**	0.747	**0.035**	0.536	**0.696**	0.610	
One stage	SSD	VGG16	0.520	0.676	0.392	0.670	0.584	0.587	0.049	0.448	0.687	0.535	
	RetinaNet with FPN	ResNet-50	0.036	0.259	0.173	0.650	0.471	0.419	0.025	0.381	0.509	0.456	
	RetinaNet with FPN	ResNet-101	0.565	0.814	0.277	0.866	0.477	0.549	0.026	0.528	0.658	0.590	
Anchor-free	FoveaBox	ResNet-101	0.552	0.748	0.442	0.677	0.589	0.535	0.013	0.511	0.655	0.501	
	FCOS with FPN	ResNet-101	**0.590**	0.564	0.409	0.627	0.671	0.663	**0.081**	0.490	0.667	0.520	
Methods		**Backbone**	**c41**	**c42**	**c43**	**c44**	**c45**	**c46**	**c47**	**c48**	**c49**	**c50**	**mAP**
Two stage	Faster RCNN with FPN	ResNet-50	**0.112**	0.276	0.230	0.453	0.002	0.070	0.420	0.466	0.048	0.437	0.375
		ResNet-101	0.202	**0.475**	0.358	0.598	0.016	**0.238**	0.598	0.402	0.055	**0.577**	0.543
	Mask RCNN with FPN	ResNet-50	**0.226**	0.450	0.370	0.593	0.019	0.189	0.634	0.447	**0.068**	0.575	0.545
		ResNet-101	**0.226**	0.462	**0.397**	**0.602**	0.014	0.216	**0.679**	**0.521**	0.053	0.569	0.564
	Cascade Mask RCNN with FPN	ResNet-50	**0.226**	0.449	0.367	**0.602**	**0.029**	0.205	0.669	0.445	**0.065**	**0.570**	**0.593**
One stage	SSD	VGG16	0.219	0.414	0.271	0.422	0.002	0.223	0.577	0.489	0.036	0.446	0.483
	RetinaNet with FPN	ResNet-50	0.099	0.314	0.277	0.426	0.004	0.118	0.376	0.456	0.032	0.357	0.326
	RetinaNet with FPN	ResNet-101	0.184	0.360	0.350	0.514	0.007	0.187	0.530	0.469	0.034	0.439	0.483
Anchor-free	FoveaBox	ResNet-101	0.133	0.400	0.274	0.533	0.012	0.123	0.505	0.466	0.029	0.379	0.459
	FCOS with FPN	ResNet-101	0.210	0.412	0.395	0.507	0.018	0.207	0.623	0.467	0.045	0.293	0.498

Note: The short names for types are defined in Table III. The entries with the best APs for each level are bold-faced.

that use either ResNet-50-FPN or ResNet-101-FPN for feature extraction, one-stage detectors, and the anchor-free detectors.

D. Experimental Analysis

According to the experimental results, we got the following observations.

1) FPN [21] improves the ship detection accuracy due to its capacity of feature extraction for different sizes of the target. Faster RCNN, Mask RCNN, and Cascade Mask RCNN with FPN methods show better advances for detecting ship objects in comparison with one-stage methods, such as SSD and RetinaNet. These results are consistent with detection results in the DIOR dataset using the same methods.
2) A deeper backbone network performs better than a shallow backbone network because it has a stronger representation capability. The performance of ResNet101 outperforms ResNet50.
3) Cascade RCNN significantly improves high-quality detection on large objects. The Cascade RCNN method has the highest average accuracy in detection and segmentation at the level 3 task. The Performance of small ships is far from satisfactory due to their small size and usually a dense packed groups of small ships in optical satellite images. On the contrary, for medium and large ship objects, the detection accuracy is higher.
4) Compared with the RetinaNet and ResNet101 backbone network, the FCOS method improved the AP by 1.5%, as shown in Table VII. These results prove that anchor-free

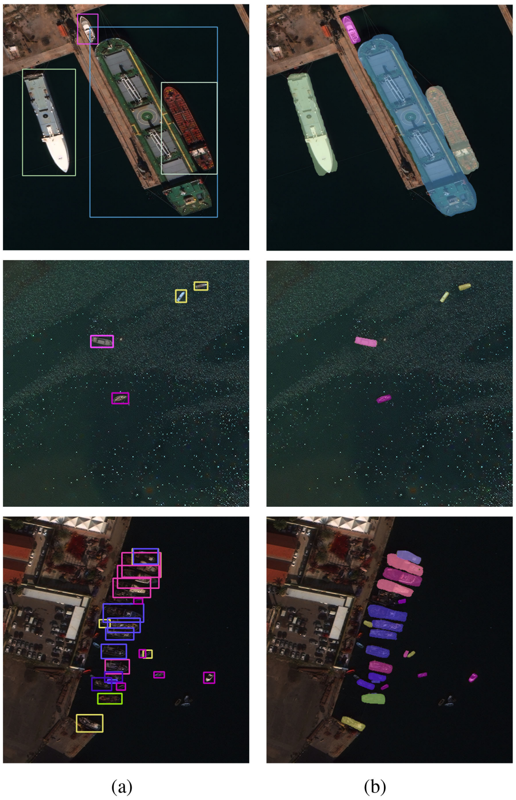


Fig. 15. Visualization results of testing on ShipRSImageNet using Cascade Mask R-CNN. Left and right, respectively, illustrate the results for HBB and SBB tasks in cases of large ships near shore, small ships on the sea, and densely distributed ships near shore. Different colors correspond to different ship types. (a) HBB. (b) SBB.

methods could significantly improve ship detection performance in striking contrast with the one-stage detector and avoid extra computation and hyperparameters related to predefined anchors.

5) The evaluation results for HBB ship detection in level 3 are significantly lower than the results produced by the Faster-RCNN methods on the FGSD and HRSC2016 dataset. Our experiments illustrate that ShipRSImageNet dataset is challenging.

V. Conclusion

In this article, we proposed ShipRSImageNet, an overhead ship detection dataset in high-resolution optical remote sensing images. To the best of our knowledge, the proposed dataset is the largest ship detection dataset in the computer vision and earth observation communities. We extensively benchmarked the dataset on ship instance segmentation and object detection tasks in remote sensing images using various state-of-the-art deep-learning-based approaches. The experimental results can be used as a useful performance baseline.

We will keep extending ShipRSImageNet. In particular, we will evaluate few shot learning [31] and fine-grained recognition [32] algorithms and add ship images taken under different environmental conditions and spatial context.

We hope the release of ShipRSImageNet will facilitate development and validation of new deep learning techniques for ship detection in remote sensing imagery.

Acknowledgment

The authors would like to thank Dr. M. Wu and Dr. C. Zhang from the University of Posts and Telecommunications for providing the FGSD training dataset.

References

[1] 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.

[2] K. Li, G. Cheng, S. Bu, and X. You, “Rotation-insensitive and context-augmented object detection in remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2337–2348, Apr. 2018.

[3] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, “When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[4] X. Yang *et al.*, “Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks,” *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 132.

[5] O. Russakovsky, J. Deng, and H. Su, “ImageNet large scale visual recognition challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[6] T.-Y. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[7] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The PASCAL Visual Object Classes challenge: A retrospective,” *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2015.

[8] G.-S. Xia *et al.*, “DOTA: A large-scale dataset for object detection in aerial images,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.

[9] G. Cheng, J. Han, P. Zhou, and D. Xu, “Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection,” *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.

[10] 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.

[11] 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, vol. 2, pp. 324–331.

[12] K. Chen, M. Wu, J. Liu, and C. Zhang, “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

[13] K. Rainey and J. Stastny, “Object recognition in ocean imagery using feature selection and compressive sensing,” in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop*, 2011, pp. 1–6.

[14] D. Lam *et al.*, “xView: Objects in context in overhead imagery,” *CoRR*, vol. abs/1802.07856, 2018. [Online]. Available: http://arxiv.org/abs/1802.07856

[15] A.-J. Gallego, A. Pertusa, and P. Gil, “Automatic ship classification from optical aerial images with convolutional neural networks,” *Remote Sens.*, vol. 10, no. 4, Mar. 2018, Art. no. 511.

[16] C. M. Ward, J. Harguess, and C. Hilton, “Ship classification from overhead imagery using synthetic data and domain adaptation,” *CoRR*, vol. abs/1905.03894, 2019. [Online]. Available: http://arxiv.org/abs/ 1905.03894

[17] L. Zhao, P. Tang, and L. Huo, “Feature significance-based multibag-of-visual-words model for remote sensing image scene classification,” *J. Appl. Remote Sens.*, vol. 10, no. 3, 2016, Art. no. 035004.

[18] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[19] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, “Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.

[20] Airbus, "Airbus Ship Detection Challenge," 2018. Accessed: Dec. 1, 2020. [Online]. Available: https://www.kaggle.com/c/airbus-ship-detection/data

[21] L. Liu, Z. Pan, and B. Lei, "Learning a rotation invariant detector with rotatable bounding box," *CoRR*, vol. abs/1711.09405, 2017. [Online]. Available: http://arxiv.org/abs/1711.09405

[22] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[23] F. Yang, Q. Xu, and B. Li, "Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 602–606, May 2017.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[25] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[26] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High quality object detection and instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1483–1498, May 2021.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[28] T. Kong, F. Sun, H. Liu, Y. Jiang, and J. Shi, "FoveaBox: Beyond anchor-based object detector," *IEEE Trans. Image Process.*, vol. 29, pp. 7389–7398, 2020.

[29] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9627–9636.

[30] 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

[31] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, 2020.

[32] Y. Chen, Y. Bai, W. Zhang, and T. Mei, "Destruction and construction learning for fine-grained image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5157–5166.

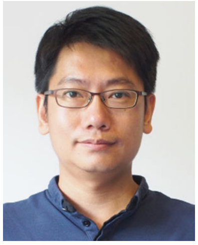
Zhengning Zhang received the M.Eng. degree in communication engineering from the China Academy of Space Technology, Beijing, China, in 2009. He is currently working toward the Ph.D. degree in information and communication engineering with the Department of electronic engineering, Tsinghua University, Beijing, China.

His research interests include remote sensing image processing, object detection, and machine learning.

Lin Zhang (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1998, 2001, and 2006, respectively.

He was a Visiting Professor with the University of California at Berkeley, from 2011 to 2013. He has been teaching the courses in selected topics in communication networks and information theory to senior undergraduate and graduate students at Tsinghua University. He is currently a Professor with Tsinghua Shenzhen International Graduate School, Tsinghua University. Since 2006, he has been implementing wireless sensor networks in a wide range of application scenarios, including underground mine security, precision agriculture, industrial monitoring, and also the 2008 Beijing Olympic Stadium (the Bird's Nest) structural security surveillance project and a metropolitan area sensing and operating network in Shenzhen. His research interests include efficient protocols for sensor networks, statistical learning, and data mining algorithms for sensory data processing, and information theory.

Dr. Zhang was the recipient of the IEEE/ACM SenSys 2010 Best Demo Awards, IEEE/ACM IPSN 2014 best demo awards, IEEE CASE 2013 best paper awards, and the excellent teacher awards from Tsinghua University, in 2004 and 2010.

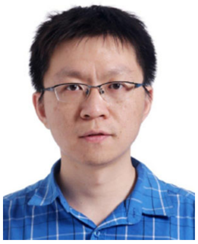


Yue Wang received the B.Sc. and Ph.D. degrees in electronic engineering from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1999 and 2005, respectively.

He is currently an Associate Professor with Tsinghua University. His research interests include computer networks, data fusion, and complex networks.

Pengming Feng (Member, IEEE) was born in Jilin, China. He received the B.Sc. degree in automatic control from the Beijing University of Chemical Technology, Beijing, China, in 2012, the M.Sc. degree in digital communication systems from Loughborough University, Loughborough, U.K., in 2013, and the Ph.D. degree in intelligent signal processing from Newcastle University, Newcastle Upon Tyne, U.K., in 2016.

During his Ph.D. degree, he was with the University Defence Research Collaboration sponsored by the U.K. Defence Science and Technology Laboratory and the Engineering and Physical Science Research Council, on the project Signal Processing in Networked Battlespace. He joined China Aerospace Science and Technology Corporation, Beijing, China, in 2017, as a Doctoral Engineer with the State Key Laboratory of Space-Ground Integrated Information Technology, where he is currently a Senior Doctoral Engineer. His research interests include remote sensing, multiple target tracking, target detection, machine learning, and sparse representation.



Ran He received the M.Eng. degree in computer science from the China Academy of Space Technology, Beijing, China, in 2009.

His research interests include industrial Internet, system integration, and machine learning.