IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2020, accepted June 25, 2020, date of publication June 29, 2020, date of current version July 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005861

HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation

SHUNJUN WEI, (Member, IEEE), XIANGFENG ZENG, QIZHE QU, (Graduate Student Member, IEEE), MOU WANG, (Student Member, IEEE), HAO SU, (Graduate Student Member, IEEE), AND JUN SHI, (Member, IEEE)

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Xiangfeng Zeng (zxf@std.uestc.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017-YFB0502700, in part by the National Natural Science Foundation of China under Grant 61501098, in part by the China Postdoctoral Science Foundation Funded Project under Grant 2015M570778, and in part by the High-Resolution Earth Observation Youth Foundation under Grant GFZX04061502.

ABSTRACT With the development of satellite technology, up to date imaging mode of synthetic aperture radar (SAR) satellite can provide higher resolution SAR imageries, which benefits ship detection and instance segmentation. Meanwhile, object detectors based on convolutional neural network (CNN) show high performance on SAR ship detection even without land-ocean segmentation; but with respective shortcomings, such as the relatively small size of SAR images for ship detection, limited SAR training samples, and inappropriate annotations, in existing SAR ship datasets, related research is hampered. To promote the development of CNN based ship detection and instance segmentation, we have constructed a High-Resolution SAR Images Dataset (HRSID). In addition to object detection, instance segmentation can also be implemented on HRSID. As for dataset construction, under the overlapped ratio of 25%, 136 panoramic SAR imageries with ranging resolution from 1m to 5m are cropped to 800 x 800 pixels SAR images. To reduce wrong annotation and missing annotation, optical remote sensing imageries are applied to reduce the interferes from harbor constructions. There are 5604 cropped SAR images and 16951 ships in HRSID, and we have divided HRSID into a training set (65% SAR images) and test set (35% SAR images) with the format of Microsoft Common Objects in Context (MS COCO). 8 state-of-the-art detectors are experimented on HRSID to build the baseline; MS COCO evaluation metrics are applicated for comprehensive evaluation. Experimental results reveal that ship detection and instance segmentation can be well implemented on HRSID.

INDEX TERMS High-resolution SAR images dataset, ship detection, instance segmentation, deep learning, convolutional neural network.

I. INTRODUCTION

Satellite-mounted synthetic aperture radar (SAR) can eliminate the effects of complex weather, working time limit and flight altitude in earth observation. As the high-resolution and vast extent characteristics of SAR imagery, ship detection with SAR imagery has a unique advantage on marine traffic safety monitoring and marine resources development compared to other remote sensing methods [1]–[4]. In recent years, with the launch of SAR satellites, such as Sentinel-1 [5], TerraSAR-X [6] and Chinese Gaofen-3, increasing amounts of high-resolution SAR imageries are available for scientific research, which tremendously promotes the development of automatic SAR ship detection [4], [7]–[10].

Traditional ship detection algorithms for SAR imageries are mainly composed of Spectral Residual (SR) [11], constant false alarm rates (CFAR) [12] and the improved algorithms derived from them. For specific needs, CFAR detection has been incorporated with diverse modules to improve detection precision [13]–[17]. But the flaws, such as manually defined feature of SAR imagery and strong dependence on the statistical distribution of sea clutters, reduce the robustness of CFAR when detecting the ships [2], [3]. Besides, without land-ocean segmentation, CFAR has the even worse performance to the panoramic SAR imagery which contains inland canal or port [18].

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

With the capability to automatically extract the deep representations of the image, convolutional neural network (CNN) based algorithms show high robustness and efficiency. Researchers attempt to promote SR and CFAR with the excellent trait of CNN. Kang *et al.* creatively take the region proposals generated by Faster R-CNN [19] as guard windows of CFAR, which combines Faster R-CNN with CFAR to detect small-sized ships [20]. Liu *et al.* combine the Land-Ocean Segmentation-based CNN (SLS-CNN) detector with corn features and the heat map of SR saliency for accurate ship detection [21]. These experiments confirm the feasibility of CNN in SAR ship detection. In computer vision, the emerging object detectors based on CNN are roughly divided into two-stage detection algorithms, multiple-stage algorithms, and one-stage detection algorithms. Two-stage detection algorithms are classification-based and combined with the network to generate region proposals. Representative two-stage algorithms are R-CNN [22], Fast R-CNN [23], Faster R-CNN, etc. They send the feature maps generated by backbone networks, such as Residual Network (ResNet) [24], [25] and Visual Geometry Group Network (VGG) [26], to the additional network (e.g. region proposed network (RPN) [19]) for preliminarily predicting the location of the object. Multiple-stage detection algorithms use the cascaded network and stepwise Intersection over Union (IoU) to improve the detection precision. Representative algorithm is Cascade R-CNN [27]. Combined with the additional network, the detection speed of two-stage detection algorithms and multiple-stage algorithms are slightly reduced compared to one-stage detection algorithms, but they perform well in precision. So, researchers have improved them for high precision SAR ship detection. To adequately utilize spatial information of SAR images, Zhao *et al.* have proposed the cascade coupled CNN-guided (3C2N-guided) visual attention method [28]; ship proposals generated by the cascaded structure combine the spatial information to improve SAR ship detection precision. Fan *et al.* have modified the Faster R-CNN to adapt PolSAR ship detection [29]; ship proposals are generated by multi-level features to detect multi-scale ships. Wei *et al.* have modified the Cascade R-CNN to realize precise and robust ship detection in high-resolution SAR imageries [30]; the proposed HRFPN structure connects high-to-low resolution subnetworks in parallel to realize high-resolution SAR ship detection. Different from two-stage and multiple-stage detection algorithms, one-stage detection algorithms squint towards regression-based detection methods and omit the network for generating region proposals. Class probability and position coordinate value of the object are generated directly to improve the detection speed, but the precision is reduced in general. Representative one-stage algorithms are You Only Look Once (YOLO v1-v3) [31]–[33], RetinaNet [34] and Single Shot MultiBox Detector (SSD) [35], etc. As the detection speed extraordinarily significant in real-time maritime disaster relief and emergency military decisions, researchers have modified the one-stage detection algorithms to SAR ship detection. Zhang *et al.* have referenced the idea of YOLO series algorithm and proposed the grid convolutional neural network (G-CNN) for real-time SAR ship detection [36]. Wang *et al.* have adjusted the hyperparameters of RetinaNet for SAR ship detection [37]. Zhang *et al.* have combined the multi-scale detection mechanism, concatenation mechanism, and anchor box mechanism into the depthwise separable convolution neural network (DS-CNN) to realize high-speed SAR ship detection [38].

Semantic segmentation divides each pixel of the input image into a semantically interpretable category, and the segmented results are highlighted by the same color for the instances within the same category. Instance segmentation combines semantic segmentation with object detection, and the predicted mask can depict the contour of the object. The bounding boxes in instance segmentation are generated by pixel-to-pixel masks; so, they are capable to locate the edge of the instances. Each instance within the same category is highlighted by a different color for determining the semantical attributes of objects. The first attempt of instance segmentation applied on CNN is Mask R-CNN [39] proposed by He, K. Based on the structure of Faster R-CNN, Mask R-CNN supplements a segmentation branch to generate the pixel-to-pixel mask. Region of interest (RoI) pooling in Faster R-CNN is replaced by RoI Align in Mask R-CNN; RoI Align can determine the value of each sampling point from the adjacent grid point through bilinear interpolation, which enables pixel-to-pixel level mapping on the feature map. Some instance segmentation detectors proposed after Mask R-CNN, such as Cascade Mask R-CNN [40] follow the idea to extend object detectors for instance segmentation. The length and contour of ships can't be measured by SAR ship detection, but these parameters can provide information about the type of ships. For example, the particular shape of the aircraft carrier can be segmented for military strikes. While there is no existing dataset that can support instance segmentation in SAR imageries, and related research is hampered.

As for existing SAR ship datasets, they have their limitations when applied to CNN-based ship detectors. OpenSARship [41] has 10 categories. But the samples are extremely imbalanced between the categories, and it's hard to train the high-performance classification model with this dataset [42]. Ship chips are designed as small size image for ship classification. Similar to OpenSARship, ship chips in the SAR-Ship-Dataset [43] have a size of 256 $\times$ 256 pixels. The small size ship chips are beneficial to ship classification [44], but they contain fewer scatterings from the land. The model trained by the ship chips may have trouble locating the ships near the highly reflective objects [37]. In the SAR ship detection dataset (SSDD) [45], the SAR images have larger size but they need to be augmented before training and testing due to the limited data, and ship detection precision tested by the test set of SSDD are generally too high [30], [37]. Besides, inappropriate annotations and less challenging detection scenes exist in

these datasets [43], [45]. Compared to OpenSARship and SAR-Ship-Dataset which fit ship classification, researchers tend to use SSDD when developing CNN-based ship detectors [28]–[30], [36], [38].

To promote the development of CNN based detectors for ship detection and instance segmentation and exclude the deficiencies in applicable SAR ship datasets for CNN, we have constructed a High-Resolution SAR Images Dataset (HRSID). Compared to the low-resolution SAR images, high-resolution SAR images have detailed and accurately represented feature of ships, and ships are more than just the bright spot. Instance segmentation in high-resolution SAR images can authentically and effectively depict the shape of ships pixel-by-pixel than low-resolution SAR images. Besides, high-resolution SAR images are beneficial to delicate tasks such as maritime transport safety and fishery enforcement. So, 136 panoramic high-resolution SAR imageries with ranging resolution from 1m to 5m are cropped to 5604 SAR images with 800 × 800 pixels. The SAR imageries have various polarization, imaging mode, imaging condition, etc., and there are 16951 ships in HRSID. To reduce wrong annotation and missing annotation of ships, optical remote sensing imagery on Google Earth [46] which has similar imaging data to SAR imagery is selected to exclude the potentially disturbing surroundings of ships. 8 state-of-the-art detectors and Microsoft Common Objects in Context (MS COCO) [47] evaluation metrics are applied for comprehensive evaluation on HRSID. HRSID is available on our website now [48]; annotations for inshore and offshore images are supplemented at the moment. We hope it can benefit the development of ship detection and instance segmentation for the community. A concise summary of our contributions are as follows:

1) A complete process of constructing the high-resolution SAR dataset for ship detection and instance segmentation is applied. HRSID is designed for CNN based detectors, which has excluded the deficiencies in the existing SAR ship dataset when constructing.
2) As the first SAR ship dataset which supports instance segmentation, the effects of instance segmentation are examined on SAR images. For ship detection, large size SAR imagery is used to examine the migration ability of the model trained on our dataset.
3) MS COCO evaluation metrics are applied for comprehensive evaluation on ship detection and instance segmentation, which include average precision (AP) for IoU threshold and small, medium, large objects. Statistical results of 8 state-of-the-art detectors are regarded as the baseline of HRSID.

This paper is organized as follows. Section II presents the process to construct the dataset. Section III describes the detectors to experiment on the dataset. Section IV presents the experimental results. Section V and VI is the conclusions and discussions, respectively.

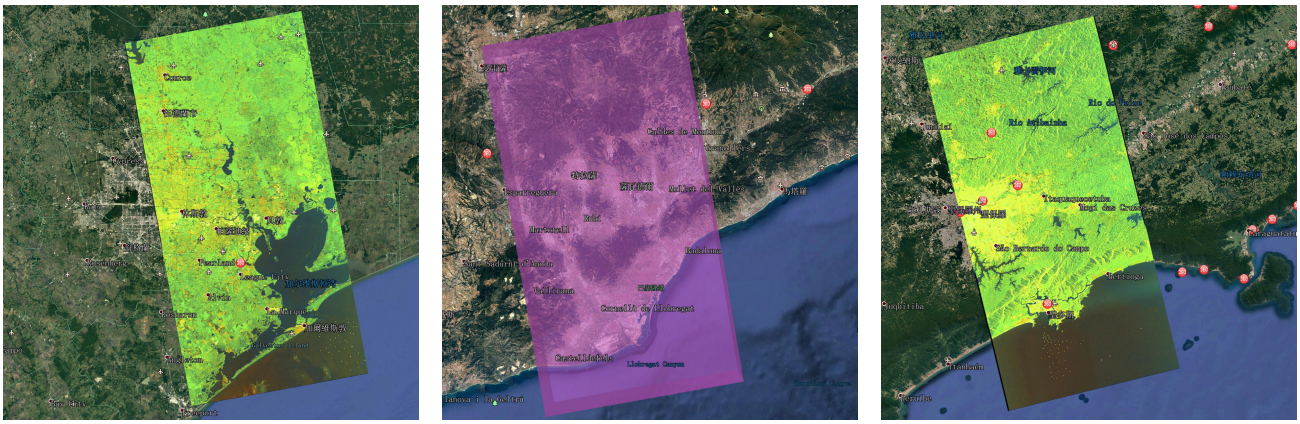
II. DATASET CONSTRUCTION AND COMPONENT ANALYSIS

A. SAR IMAGERIES FOR DATASET CONSTRUCTION

The original SAR imageries for constructing HRSID are 99 Sentinel-1B imageries, 36 TerraSAR-X and 1 TanDEM-X [49] imageries; the resolution of SAR imageries is under 3m to keep detailed and accurately represented feature of ships. Under different imaging modes of radar sensors, ships appear in different forms. For example, TerraSAR-X has several imaging modes: Staring SpotLight (ST), High Resolution SpotLight (HS), SpotLight, StripMap (SM), ScanSAR (SC), Wide ScanSAR (WSC); under the Wide ScanSAR imaging mode, the scan scope can up to 270km in azimuth and 800km in range, which can meet the broad demand of large area coverage monitoring such as marine traffic, sea ice monitoring and regular detection of oil films, but the detailed feature of ships is unclear compared to high-resolution SAR imageries.

To ensure high imaging quality, we have chosen the high-resolution imaging mode of the satellite when constructing the dataset. As for Sentinel-1B satellite, the imaging mode of S3 StripMap is selected, which has the resolution from 1.7m × 4.3m to 3.6m × 4.9m in range and azimuth; corresponding swath width is 80km. As for TerraSAR-X, the selected imaging mode are ST, HS, and SM, corresponding resolution of imaging mode is up to 25cm, up to 1m, up to 3m, respectively; corresponding swath size is $4 \times 3.7\text{km}^2$, up to $10 \times 5\text{km}^2$, $30 \times 50\ \text{km}^2$, respectively. Several imaging areas of SAR imageries in constructing HRSID are highlighted by the rectangular box in Figure 1. The imageries are provided by Google Earth.

The imaging region is selected at the port with tremendous cargo handling capacity or the crisscrossed busy canals throughout the trading cities. These areas can simultaneously present specific scenes in need with the limited swath in high-resolution SAR imageries. For instance, the offshore areas which are covered with a wide variety of ships and the anchorage areas where ships are difficult to distinguish from the clutter interfered background can coexist in the same SAR imagery. Consequently, the limited amount of SAR imageries can be fully utilized to generate more cropped SAR images when constructing the dataset. In addition to the imaging region, the backscatter coefficients value influenced by polarization and incident angle of radar sensors will affect the imaging condition of SAR imageries. In terms of SAR imageries pre-processed by the supplier, the interferes such as foreshortening, layover, and shadowing of ships are influenced by the incident angle of radar sensors. We have chosen the incident angle which has minimized interferes. Sentinel-1B SM has 6 elevation beams and the incident angle varies from 18.3° ∼ 46.8°. S3 beam corresponds to the incident angles of 27.6° ∼ 34.8°, and ships under this elevation beam have less interferes compared to other elevation beams. Existing interferences are disposed of in subsequent annotation procedure. As for polarization, the radar remote sensing system has four fundamental polarization methods: HH, VV, HV, VH [50]. In general,



(a) (b) (c)

FIGURE 1. Several SAR imagery acquisition areas of our dataset. (a) and (c) are generated by the imaging mode of Sentinel-1B S3 SM; (b) is generated by TerraSAR-X SM. The rectangular coverage area is the imaging range.

TABLE 1. Detailed information of the SAR imageries for constructing HRSID.

Satellite	Imaging Mode	Swath (km)	Incident Angle (°)	Polarization	Resolution (m)	Position	Image (num)
Sentinel-1B	S3-SM	80	27.6~34.8	HH	3	Port Houston	40
Sentinel-1B	S3-SM	80	27.6~34.8	HV	3	Port Sao Paulo	20
Sentinel-1B	S3-SM	80	27.6~34.8	HH	3	Port Sao Paulo	21
TerraSAR-X	SM	30	20~45	VV	3	Port Barcelona	23
Sentinel-1B	S3-SM	80	27.6~34.8	VV	3	Port Chittagong	18
TerraSAR-X	ST	4	20~60	HH	0.5	Aswan Dam	2
TerraSAR-X	ST	4	20~60	HH	0.5	Shanghai	2
TanDEM	HS	10	20~55	HH	1	Panama Canal	1
TerraSAR-X	HS	10	20~55	VV	1	Port Visakhapatnam	1
TerraSAR-X	SM	30	20~45	HH	3	Port Singapore	4
TerraSAR-X	SM	30	20~45	HH	3	Strait Gibraltar	2
TerraSAR-X	SM	30	20~45	VV	3	Port Sulphur	1
TerraSAR-X	SM	30	20~45	VV	3	Bay Plenty	1

co-polarization has higher backscatter coefficients value of ships and sea clutters than cross-polarization [51], [52]. Ships and sea clutters in cross-polarized SAR imageries are brighter to the background than co-polarized SAR imageries [52]. While the calm sea in co-polarized SAR imageries is relatively darker to the background than cross-polarized SAR imageries due to specular reflection of the sea. So, when constructing HRSID, we have selected 116 co-polarized SAR imageries with a clear distinction between ships and background, and 20 cross-polarized SAR imageries are added for the supplement. Detailed descriptions of these SAR imageries are shown in Table 1. Due to the inconsistent scattering caused by the angular difference in the wide area, we mainly perform the correction and compensation according to the distance.

To make the model trained by our dataset can distinguish and segment the ships from complex backgrounds, we have analyzed the reasonable ratio of each type of detection scene in HRSID beforehand. When training the detectors, large amounts of ships with detailed and accurately represented features should be prepared for training. So, the offshore scenes with ships distributed in the sea are the main component of HRSID. Ship detection in inshore scenes is influenced by man-made facilities or buildings. The inshore scenes are regarded as the interferential scene to maintain a certain amount in HRSID. While the challenging scenes, such as the adjacent ships, cluster-distributed small ships in the canal and large size ships defined by MS COCO evaluation metrics [47], are added to HRSID as supplementary. Adjacent ships challenge the non-maximum suppression (NMS) algorithm used in CNN based ship detectors to generate precise bounding box for location; cluster-distributed small ships in the canal are dense in the space, and they challenge the location and instance segmentation; large size ships are scarce in the training samples and detectors tend to detect the component of their features as small ships.

The downloaded SAR imageries are pre-processed by the supplier beforehand. It still needs to be processed to display as a grayscale imagery, and we use the clipping function with linear transformation for implementation. The clipping function is defined in formula 1 as follows:

$$y = \begin{cases} kx, & 0 < x \leq \beta \times max(x) \\ \beta \times max(x), & x > \beta \times max(x) \end{cases} \quad (1)$$

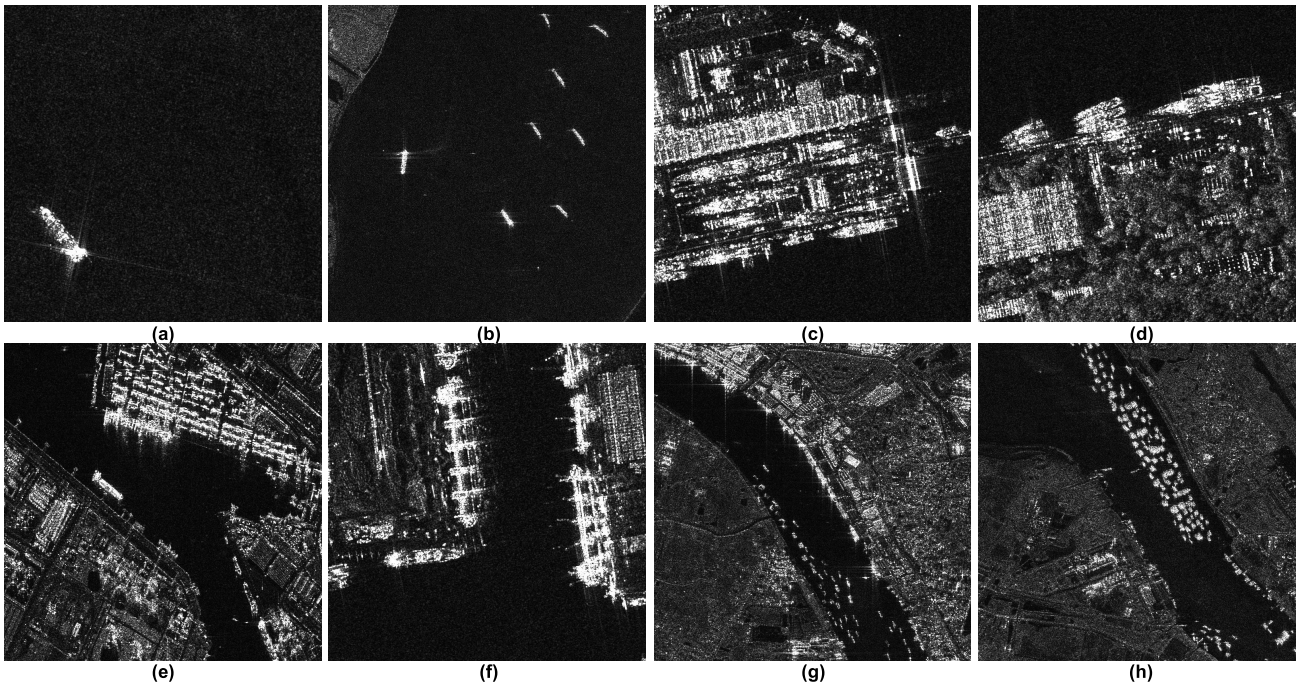


FIGURE 2. Some representative high-resolution SAR images in our dataset with 800 × 800 pixels.

x represents the pixel of the downloaded SAR image. When the value of the pixel is less than $\beta \times max(x)$, use the penalty factor k for linear transformation; we set k =1 here. When the value of the pixel is higher than $\beta \times max(x)$, reduce it to $\beta \times max(x)$. y represents the output grayscale image. Based on the default setting in the reference [50], we tune the threshold β for different imaging scenes to present distinguishable feature of ships.

Through the side-scan mechanism of SAR sensors contraposing specific regions, panoramic SAR imagery has a large image size. It is supposed to be cropped to the matched input size of CNN. The process is divided into four steps in our research. First of all, in order to avoid reduplicative cropping when constructing dataset, the port and offshore areas with relatively dense distribution of ships are separated from the panoramic SAR imagery for the subsequent sliding window procedure; besides, the sporadically distributed ships on the sea are individually separated from the SAR imagery with 800 × 800 pixels window, and threshold β is individually set for the image. Secondly, a sliding window with 800 × 800 pixels is adopted to satisfy the demand of scaling transformation in CNN based ship detectors and reserve the scene which contains ships and man-made facilities to measure the ship detection competence of detectors. Thirdly, the sliding window is shifting over the SAR imagery with a stride of 600 pixels in length and width, and the overlapped ratio of successively cropped images is set at 20% to ensure all the ships appeared in panoramic SAR imageries have complete features when cropped by sliding window. Fourthly, we have filtered out 400 cropped images with pure background. When testing the robustness of the trained model, the full negative sample can provide information of land or sea clutter.

Cropped SAR images with 800 × 800 pixels are the main components of our dataset. Ships in the high- resolution SAR images have detailed and accurately represented features. Some SAR image samples are shown in Figure 2. (a) and (b) represent offshore single and multiple ships; (c) and (d) are the adjacent ships; (e) and (f) show ships berthing at the port and large size ships, respectively. (g) and (h) display the cluster-distributed small ships in the canal.

B. ANNOTATION STRATEGIES FOR DATASET CONSTRUCTION

The bounding box is well performed in locating the objects [53]. In ship detection, the location of the ships is determined by four vertex coordinate values of the bounding box. But with the sharp shape, the annotated bounding box areas coexist ships and background features; the predicted bounding boxes only provide the four vertex coordinate values but not the shape of ships. For instance segmentation, polygons are applicable in annotation and they fit the contour of the ships well. Polygon annotated mask is also applied to generate the bounding box for object detection, and the bounding boxes are precise enough to locate the edge of ships. So, we use the polygons to annotate the ships when constructing the dataset. As for optical remote sensing images, the annotation strategy for object detection and instance segmentation can refer to this work [54]. But as SAR imagery is grayscale imagery, the corresponding annotation strategy should add additional procedure. In the inshore areas, the man-made facilities and buildings have similar features to ship, which interferes with annotation. We have designed some auxiliary means to deal with it as is shown in Figure 3.

FIGURE 3. The auxiliary means to distinguish ships in offshore areas.

(a) confirms the bright pixels as a ship with partially enlarged details in optical remote sensing images; (b) shows the same imaging region in the gray image and optical image to determine the possible interference from the facilities, buildings or cranes in the port. The red oval marked objects are ships, and the orange rectangle framed objects are cranes that may interfere with annotating ships.

Ports generally have huge cargo volumes so that ships will not call at a certain berthage for long, but the facilities and buildings in the port are almost the same for a long time, and the easy moving objects in the port such as cranes have a distinct feature which is distinguishable to the ships. In optical remote sensing imageries, the location of the facilities, buildings, and cranes is easy to determine; the possible berthage of ships is consequently confirmed. So, the optical remote sensing imagery which has adjacent imaging date to the SAR imageries are taken from Google Earth to help the SAR experts with auxiliary judgment.

In the SAR images with 800 $\times$ 800 pixels, the feature of ships still has mixed interference caused by incident angle, polarization, etc. These situations, such as moving targets, ships surrounded by high-reflective objects and large antenna elevation, can twist the shape of ships. To minimize the deviation of ship detectors without causing controversy, we classify the adjacent interference emanated from ships as a composition of ships. While the highlighted pixels, for example, spindly sidelobe caused by the offset of swift navigation, are reserved according to the similarity to the principle structure of ships.

Apart from the annotating methods mentioned above, we tactfully adjust the order of ship annotating and SAR imagery cropping to avoid reduplicative annotation. If the SAR imageries are cropped before annotation, ships in the overlapped areas may lead to reduplicative and inconsistent annotations. So, a more reasonable annotation scheme is formulated as is shown in Figure 4. Firstly, we annotate the ships on panoramic SAR imagery all at once. Secondly, the annotated SAR imageries are processed to generate the corresponding imageries for semantic segmentation and instance segmentation. Thirdly, a sliding window with 800 $\times$ 800 pixels acts on the imageries to generate the SAR images with instance segmentation and semantic segmentation images. Fourthly, the annotations are regenerated from the instance segmentation and semantic segmentation images to the format MS COCO dataset [47]. The strategy can still generate annotations for the ships when boundaries of sliding window fall on it, and it has reduced the workload of annotation.

To examine the consequence of annotation, we have visualized the annotated ships in Figure 5. In the format of the MS COCO dataset, the polygons annotated by experts are transformed to mask for segmentation, and the bounding boxes for object detection are generated by the mask. The transformed mask can locate the ships with its contour and the bounding box generated by the mask is capable to locate the edge of ships. When annotating, the polygons are generated by the software named Labelme [55], which can support the annotation formats of the polygon, rectangle, circle, etc. As for dataset constructing, the annotations of each SAR image constitute a JavaScript Object Notation (JSON) file in MS COCO dataset format, facilitating the reading and transmission of information. MS COCO dataset format enables each ship instance annotation contains the category id, bounding box, and segmentation mask. Thus, guaranteeing HRSID can satisfy the demand for ship detection and instance segmentation.

C. STATISTICS ANALYSIS ON HRSID

Existing large optical datasets (e.g., MS COCO, ImageNet [56], PASCAL VOC [57]) have a large variety of categories for large-scale visual identity; correspondingly, they contain large amounts of images. Distinguished from the multiple colors in optical images, SAR imagery appears as grayscale images; but the imaging effect of SAR imageries is influenced by various factors such as clutters and incident angle of the satellite. Object detection in SAR imageries is still complicated. However, CNN based ship detectors which are trained by existing SAR datasets tend to reach the bottleneck of precision [30], [37]. So, some challenging detection scenes are supplemented in HRSID to add complexity. Besides, these high-resolution scenes can provide similar SAR detection scenes to optical scenes to add the complexity of ship detection.

As the limited operational capability in consumer-oriented graphic cards, the small datasets, for example, NWPU VHR-10 [58] constructed by optical remote sensing images and SSDD for SAR ship detection, are extensively applicated in object detection [30], [37], [59], [60]. So, HRSID is designed to have 5604 high-resolution SAR images for wide usage.

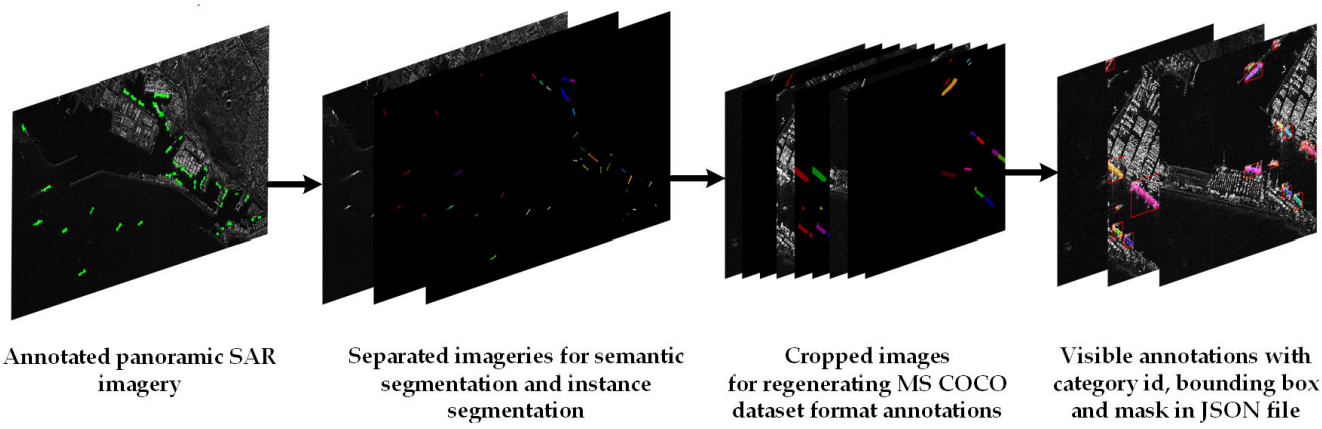


FIGURE 4. The strategy of annotation.

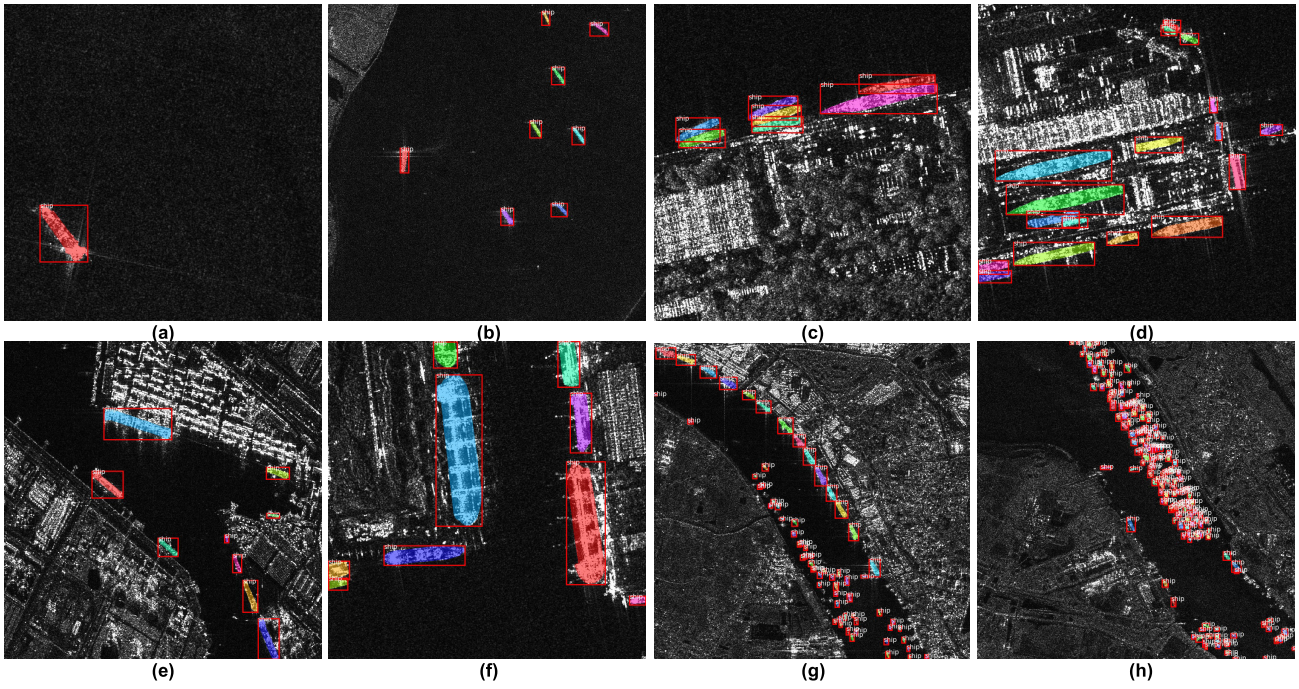


FIGURE 5. The annotated ships with category id, bounding box and mask in HRSID.

As it's designed for CNN based ship detection and instance segmentation, it contains one category and provides annotations for the ships, and other categories appeared in SAR images aren't annotated.

We refer to the MS COCO evaluation metrics [47] to analyze our dataset. HRSID is divided into the training set with the amount of 65% images, and the test set with 35% images. Statistics of HRSID, the training set, and the test set are shown as a histogram in Figure 6, the area of the bounding box and aspect ratio of the bounding box are taken into consideration. The aspect ratio of the bounding box corresponds to the shape of the bounding box, and it's essential for the CNN based detectors which adopt anchor to generate bounding box [19], [27]. The area of the bounding box is the criterion to measure the scale of ships in the MS COCO dataset. According to the scale division for object detection in MS COCO, area of the bounding box below 32×32 pixels corresponds to the small object, area of the bounding box from 32×32 pixels to 96×96 pixels correspond to the medium object and area of the bounding box above 96×96 pixels correspond to the large object. Statistically, the number of annotated ships is 16951, and each SAR image is distributed with 3 ships on average. The number of small ships, medium ships, and large ships takes up 54.5%, 43.5% and 2% of all ships, respectively. The area of the bounding box for small ships, medium ships, and large ships takes up 0∼0.16%, 0.16%∼1.5% and above 1.5% of the SAR image, respectively. So, HRSID has the characteristics of small objects but large detection scenes; ships are sparsely distributed in SAR images. As the very high-resolution (VHR) SAR imageries with specific contexts are scarce, which is the source of large ships, HRSID has a relatively low ratio of large ships. In the training set and test

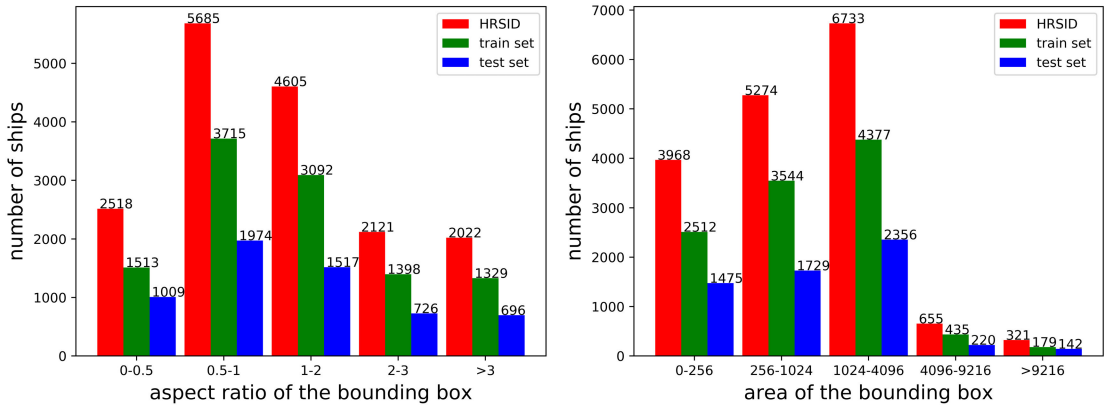


FIGURE 6. Statistics of HRSID, training set and test set, including the aspect ratio of the bounding box and area of the bounding box.

TABLE 2. Different groups of SAR images with the same imaging parameter.

Image	Satellite	Imaging mode	Incident Angle (°)	Resolution (m)	Polarization	Number of the ships
P0001~P0040, P0061~P0081	Sentinel-1B	S3-SM	27.6~34.8	3	HH	6798
P0124~P0125, P0130~P0131	TerraSAR-X	ST	20~60	0.5	HH	190
P0126~P0127, P0132~P0135	TerraSAR-X	SM	20~45	3	HH	813
P0041~P0060	Sentinel-1B	S3-SM	27.6~34.8	3	HV	1688
P0082~P0104	TerraSAR-X	SM	20~45	3	VV	2090
P0105~P0122	Sentinel-1B	S3-SM	27.6~34.8	3	VV	4985
P0129, P0136	TerraSAR-X	SM	20~45	3	VV	21
P0123	TanDEM	HS	20~55	1	HH	86
P0128	TerraSAR-X	HS	20~55	1	VV	280

set, the properties are similar to HRSID. To sum up, HRSID emphases on examining the ability of detectors in detecting small and medium ships.

Considering the imaging parameters such as incident angle, polarization, resolution, etc., are different, we divide the dataset into different groups according to the serial number. Each group has the same imaging parameters and the number of ships is counted as is shown in Table 2. Meanwhile, statistics of the SSDD and the SAR-Ship-Dataset are summarized in Table 3. Parameters, including the size of ships, size of images, number of images, annotations and resolution are used for analysis. Quantitatively, all the dataset emphasis on detecting small and medium ships. As for the SAR-Ship-Dataset, the small size ship chips are beneficial to ship classification; but they are incompetent to delicate ship detection tasks. As for the SSDD, it has wide range usage for CNN based ship detection relative balanced size of ships and multiple sizes of SAR images; but with the fewer number of SAR images, SSDD needs to be augmented before training and testing. In terms of HRSID, the resolution of SAR images varies from 0.5m~3m. Ships are annotated by polygons, and the annotations contain masks and bounding boxes for ship detection and instance segmentation, respectively. The model trained by high-resolution SAR images fit the delicate tasks such as maritime transport safety and fishery enforcement. Besides, taking the difference in the stride while cropping and size of cropped images into account, the capacity of HRSID is equivalent to the SAR-Ship-Dataset.

III. STATE-OF-THE-ART ALGORITHMS FOR BUILDING THE BASELINE

A. BACKBONE NETWORK

In order to compare the performance of the detectors on our dataset under the same conditions, we use the ResNet-FPN [61] architecture as the backbone network of the detectors. With the residual module, ResNet is deeper but stable [25], [26]. Feature Pyramid Networks (FPN) [62] construct the top-down feature pyramid structure, which is based on fusing the inherent multi-scale feature map. As for ResNet, conv2, conv3, conv4, and conv5 are recorded as $\{C_2, C_3, C_4, C_5\}$ with corresponding strides of {4, 8, 16, 32} relative to the pixel of the input image. By upsampling the C_5 layer with a top-down pathway to match the size of convolutional layers, reducing the dimension of the channels

TABLE 3. Statistics of SSDD, SAR-Ship-Dataset and HRSID.

Datasets	Size of ships (num)			Size of images (pixels)		Images (num)	Annotations	Resolution (m)
	Small	Medium	Large	height	width			
SSDD	1529	935	76	190~526	214~668	1160	Bounding box	1~10
SAR-Ship-Dataset	35695	23660	180	256	256	43819	Bounding box	3~25
HRSID	9242	7388	321	800	800	5604	Polygon	0.5~3

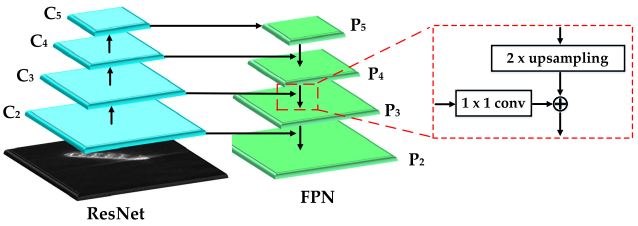


FIGURE 7. The architecture of ResNet-FPN.

with a 1 × 1 convolutional layer and merging the maps with a 3 × 3 convolutional network to eliminate the effects of upsampling. The output feature maps are {P_2, P_3, P_4, P_5} which corresponds to {C_2, C_3, C_4, C_5}. The composite structure of ResNet and FPN makes the detectors adaptable to detect small objects. It's beneficial to ship detection because ships often appear as small objects in SAR images [61]. The architecture of ResNet-FPN is shown in Figure 7.

B. STATE-OF-THE-ART DETECTORS

Improved backbone networks, optimized loss functions (e.g., Focal Loss [34] applied in RetinaNet) and functional subnet are beneficial to develop object detectors. So, building a baseline with state-of-the-art detectors that contain common structure for reference is essential for HRSID to be implemented in further study. We have selected 8 state-of-the-art detectors to build the baseline of HRSID.

1) FASTER R-CNN

Faster R-CNN consists of three modules: ResNet, Region Proposed Network (RPN) and Fast R-CNN. ResNet-FPN extracts the feature map of the SAR image. RPN can generate the ship proposals for preliminarily predicting the location of ships. The RoI Pooling layer can transform the scale of the region proposed feature map to fit the input size of fully connected layers. Fast R-CNN finishes the binary classification and bounding box regression.

2) CASCADE R-CNN

As for detectors using the anchor mechanism, the IoU threshold value of the bounding box is used to distinguish the positive and negative samples. The precision of the predicted bounding box under the IoU threshold value will be classified into negative samples and filtered. Cascade R-CNN replaces Fast R-CNN in Faster R-CNN with cascaded Fast R-CNN. The cascaded Fast R-CNN assigns increasing IoU threshold value in sequence for each Fast R-CNN. Besides, the IoU threshold value is assigned at an incremental interval to avoid mismatching.

3) RETINANET

RetinaNet consists of two components: ResNet-FPN, Fully Convolutional Networks (FCN). Two independent FCN branches perform the classification and location tasks separately. FCN can adapt to the flexible size of feature maps, and it's more robust than full connected layers. Focal Loss lowers the weight of negative samples and strengthens the effects of positive samples. It has solved the category imbalance in one-stage detection algorithms.

4) MASK R-CNN

Based on the structure of Faster R-CNN, Mask R-CNN adds a mask branch to predict the segmentation mask for each Region of Interest (RoI), paralleling to the classification and bounding box regression branch in Fast R-CNN. The mask branch utilizes FCN to predict the segmentation mask in a pixel-to-pixel manner. Besides, RoI Pooling in Faster R-CNN is replaced by RoI Align in Mask R-CNN. RoI Align determines the value of each sampling point from the adjacent grid point on the feature map through bilinear interpolation so that it can finish one-to-one correspondence between input pixels and output pixels. Without quantization in the coordinates, Mask R-CNN can generate a pixel-to-pixel mask for instance segmentation. The loss function of the mask is calculated separately to ship detection.

5) MASK SCORING R-CNN

Among the instance segmentation tasks in Mask R-CNN, the quality of the segmented mask is determined by the classification confidence of object detection branches. But there are no strong correlations between the two. When segmenting the instances, the mask quality of overlapped congeneric objects tends to be poor. Mask Scoring R-CNN [63] adds MaskIoU Head to improve the mask quality, and the mask score is defined by the product of classification score and MaskIoU score. MaskIoU Head transforms the output scale of the mask branch with the MaxPooling layer and concatenates it with

the RoI feature map as its input, through 4 convolution layers and 3 fully connected layers to get the output.

6) CASCADE MASK R-CNN

As the name implies, Cascade Mask R-CNN is the hybrid of Mask R-CNN and Cascade R-CNN. It combines the excellent characteristics of the two detectors, so that the Cascade R-CNN, which performs well in object detection, can finish the instance segmentation tasks. Each cascade structure adds a mask branch to finish the instance segmentation task and generate the pixel level mask.

7) HYBRID TASK CASCADE

In the structure of Cascade Mask R-CNN, each stage contains the bounding box branch and mask branch. But there are no association between the parallel structure. To improve the detection precision, Hybrid Task Cascade [40] interweaves bounding box and mask branches for joint multi-stage processing, and uses semantic segmentation branches to provide spatial context.

8) HRSDNET

Apart from the above standard ship detection methods, we have added the dedicated ship detection method for SAR images. HRSDNet adopts the high-resolution feature pyramid network (HRFPN) as the backbone network. HRFPN connects the high-to-low resolution subnetworks in parallel for obtaining accurate spatial precision. The Soft Non-Maximum Suppression (Soft-NMS) is used to detect the cluster distributed ships.

IV. SHIP DETECTION AND INSTANCE SEGMENTATION PERFORMANCE ON HRSID

In this section, we will evaluate the experimental results on HRSID generated by state-of-the-art detectors mentioned. Not only the measured AP be regarded as the baseline of our dataset, but we also visualize the detection and segmentation results of the detectors.

A. EVALUATION METRICS

For quantitatively and comprehensively evaluation of the performance of object detectors, the evaluation metrics such as IoU, precision, recall, and mAP are the normative means [57]. In supervised learning, the coordinates of the object's location are annotated by the experts, which is called ground truth in object detection and instance segmentation. The overlap rate of predicted result and ground truth is the measurement of the correlation between the two; a higher degree of overlap indicates a better correlation and more precise prediction. As is shown in formula 2, bounding box IoU is defined by the overlap rate of the predicted bounding box and ground truth bounding box:

$$IoU_{bbox} = \frac{Bbox_{pd} \cap Bbox_{gt}}{Bbox_{pd} \cup Bbox_{gt}} \tag{2}$$

Analogously, mask IoU for instance segmentation is defined by the overlap rate of predicted mask and ground mask to measure the segmentation precision, as is shown in formula 3:

$$IoU_{mask} = \frac{Mask_{pd} \cap Mask_{gt}}{Mask_{pd} \cup Mask_{gt}} \tag{3}$$

During classification, algorithms may misjudge the background and objects. There are four classification results: True Positives (TP), True Negatives (TN), False Negatives (FN) and False Positives (FP). TP denotes the amount of correctly classified positive samples; TN shows the amount of correctly classified negative samples; FN represents the amount of missed positive samples; FP indicates the number of false alarms in the background. The precision and recall are defined by these criteria, as is shown in formula 4 and formula 5.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Based on the quantities of precision and recall, AP is defined. In the Cartesian coordinate system, if the horizontal coordinate is recall value and the vertical coordinate is precision value, the area under the recall-precision curve is AP value, as is shown in formula 6:

$$AP = \int_0^1 P(r)dr \tag{6}$$

where P represents precision and r represents recall. If there are multiple categories in the dataset, the numerical average of all categories is defined as mean AP (mAP).

Common dataset evaluation formats are Pascal Visual Object Classes (Pascal VOC) and Microsoft Common Objects in Context (MS COCO). The calculation criterion of mAP for the Pascal VOC dataset is based on an IoU threshold of 0.5, while the evaluation metrics in MS COCO are abundant and comprehensive. In the evaluation metrics of MS COCO, objects with multiple sizes in an identical category are assessed individually due to their wide disparity in AP; except for the same AP_{50} in evaluation metrics of Pascal VOC, MS COCO has the strict metric of IoU thresholds such as AP_{75} and AP. AP_{75} represents the calculation under the IoU threshold of 0.75, and AP is the primary challenge metric with the calculation of average IoU, which has ten IoU thresholds distributed from 0.5 to 0.95 with the step of 0.05. In terms of the capabilities in multi-scale object detection, there are AP_S, $AP_{M,}$ and AP_L for evaluation. Specifically, the three indicators denote the objects with small (area $< 32^2$ pixels), medium (32^2 $<$ area $< 64^2$ pixels) and large (area $> 64^2$ pixels) size. We have taken AP, AP_{50}, AP_{75}, AP_S, $AP_{M,}$ and AP_L to characterize the performance of the detectors on our test set. Except for the IoU computation which is respectively performed on bounding boxes and masks, the evaluation metrics above are in all respects for object detection with bounding boxes and instance segmentation with masks.

TABLE 4. Ship detection statistics generated by bounding box AP on test set of HRSID.

Model	Backbone	Model Size (Mb)	Test Speed	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN	ResNet-50+FPN	330.2	0.074s	63.5	86.7	73.3	64.4	65.1	16.4
	ResNet-101+FPN	482.4	0.096s	63.9	86.7	73.6	64.8	66.2	24.2
Cascade R-CNN	ResNet-50+FPN	552.6	0.086s	66.6	87.7	76.4	67.5	67.7	28.8
	ResNet-101+FPN	704.8	0.109s	66.8	87.9	76.6	67.5	68.8	27.7
RetinaNet	ResNet-50+FPN	290.0	0.068s	60.0	84.7	67.2	60.9	60.9	26.8
	ResNet-101+FPN	442.3	0.091s	59.8	84.8	67.2	60.4	62.7	26.5
Mask R-CNN	ResNet-50+FPN	351.2	0.083s	65.0	88.0	75.2	66.1	66.1	17.3
	ResNet-101+FPN	503.4	0.105s	65.4	88.1	75.7	66.3	68.0	23.2
Mask Scoring R-CNN	ResNet-50+FPN	481.1	0.079s	64.1	87.6	75.0	65.3	65.8	22.2
	ResNet-101+FPN	633.1	0.103s	64.9	88.6	75.4	66.2	67.3	19.6
Cascade Mask R-CNN	ResNet-50+FPN	615.6	0.097s	67.5	88.5	77.4	68.6	67.4	22.6
	ResNet-101+FPN	767.8	0.119s	67.6	88.8	77.4	68.4	69.9	23.9
Hybrid Task Cascade	ResNet-50+FPN	639.3	0.132s	68.2	87.7	78.8	69.0	71.2	38.1
	ResNet-101+FPN	791.6	0.156s	68.4	87.7	78.8	69.2	72.0	31.9
HRSDNet	HRFPN-W32	598.1	0.127s	68.6	88.4	79.0	69.6	70.0	25.2
	HRFPN-W40	728.2	0.154s	69.4	89.3	79.8	70.3	71.1	28.9

B. EXPERIMENTAL DETAILS

All the experiments on our dataset are supported by the personal computer (PC) with the 64-bits Ubuntu 18.04 operating system. The software configuration consists of python programming language, PyTorch 1.3.0, CUDA 10.1 and cuDNN 7.6.1. The hardware capabilities include NVIDIA RTX-2080 GPU (8GB memory), Intel®i7-8700 CPU @3.20GHz and 32 GB RAM. To maintain the same hyperparameters of the detectors, we choose mmdetection (a flexible toolkit for reimplementing existing methods) [64] for training and testing. To make more accurate location and segmentation, the SAR images are proportionally resized to 1000 × 1000 pixels in the process of training and testing [65]–[68]. All the detectors are trained with GPU and finished in 12th epochs; the momentum and weight decay are set to 0.9 and 0.0001, respectively. IoU threshold is set to 0.7 when training and testing for rigorous filtering to the bounding boxes with low precision. The IoU thresholds in Cascade R-CNN are set to {0.5, 0.6, 0.7}. We choose SGD with the initial learning rate of 0.0025 as the optimizer, the other hyperparameters are set to the default values in mmdetection.

C. SHIP DETECTION RESULTS ON STATE-OF-THE-ART DETECTORS

In Table 4, we have shown the ship detection statistics generated by bounding box AP on the test set of HRSID. Each detector adopts ResNet50-FPN and ResNet101-FPN as the backbone network for contrast. Considering the feasibility in practical application, we have added the model size after training and the test speed per SAR image for each detector. Generally, with more functional structure and deeper network, the model size and AP will increase, but the detection speed is reduced in return. To build the baseline of our dataset, the settings of hyperparameters are consistent. The precision-recall curve (PR curve) of each detector is shown in Figure 8 and Figure 9.

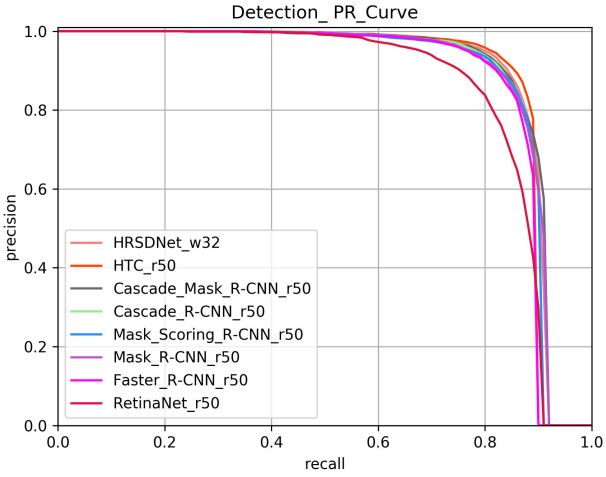


FIGURE 8. The detection PR curve with the backbone of ResNet-50.

Through the comparison of the statistics, detectors with ResNet101-FPN perform better in bounding box AP than detectors with ResNet50-FPN as a backbone in general, but the deeper network adds the size of the model after training and lower the detection speed on SAR images. With the same backbone network, RetinaNet has the minimum model size of 290.0Mb and outperforms other two-stage and multiple-stage detection algorithms in detection speed, but it's inferior in bounding box AP; HRSDNet receives the highest bounding box AP of 69.4% and the model size of 728.2Mb with the cascaded networks, while it takes 0.154s to detect the ships. Compared to Cascade Mask R-CNN, the bounding box AP of

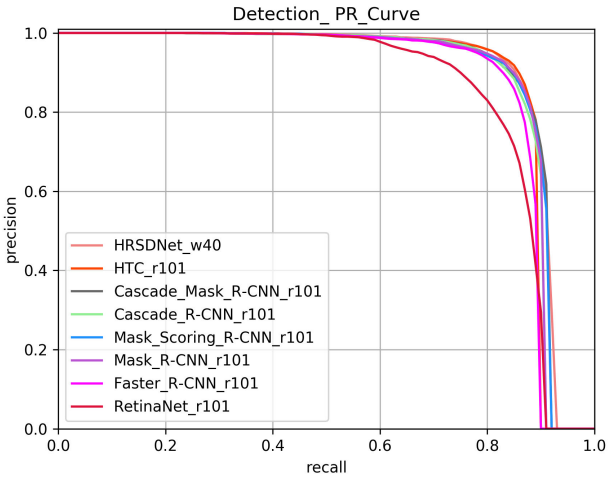


FIGURE 9. The detection PR curve with the backbone of ResNet-101.

Hybrid Task Cascade (HTC) is improved by 0.7% and 0.8% with the backbone of ResNet50-FPN and ResNet101-FPN, respectively. Mask R-CNN performs better in bounding box AP than Faster R-CNN with its mask branch and RoI Align. Using the stepwise increasing IoU threshold in each cascaded detection structure, Cascade R-CNN receives a prominent improvement of 3.1% and 2.9% than Faster R-CNN in bounding box AP with the backbone of ResNet50-FPN and ResNet101-FPN, respectively.

Under the bounding box IoU threshold of 0.5, the bounding box AP of state-of-the-art detectors is above 84.7%. While under the relatively strict bounding box IoU threshold of 0.75, bounding box AP is above 67.2%. As for multi-scale ship detection, when detecting the small and medium ships in our dataset, the bounding box AP of detectors is above 60.4% and 60.9%, respectively. Benefit from the ResNet-FPN backbone, detectors are applicable to detect small ships. But they have abrupt decreasing in bounding box AP when detecting large ships on account of the relatively strict definition for large SAR ships in MS COCO evaluation metrics and the fewer training samples.

To examine the ship detection ability of detectors to complex detection scenes, we have selected 4 representative scenes in the test set of HRSID for ship detection. Visible results are shown in Figure 10. Green bounding boxes denote ground truth and red bounding boxes denote predicted results. The bounding box IoU threshold for testing is set to 0.7 to avoid excessive amounts of the false alarms; the predicted bounding box under the confidence coefficient of 0.7 is filtered. Column 1 shows ships which have a similar feature to the objects in the port, Column 2 denotes the cluster-distributed small ships in the canal, Column 3 exhibits the adjacent ships, Column 4 is the large size ships mooring in the port. Row 1 to Row 8 represents the ship detection results of Faster R-CNN, Cascade R-CNN, RetinaNet, Mask R-CNN, Mask Scoring R-CNN, Cascade Mask R-CNN, Hybrid Task Cascade, and HRSDNet, respectively.

As is shown in Column 1 and Column 2, all the two-stage and multiple-stage detection algorithms have high performance in detecting ships near man-made facilities in the port but missed detection and false alarm appear when the scene switches to the canal with cluster-distributed small ships; RetinaNet has difficulty in detecting ships in the above scenes. In Column 3, it appears as if the performance of detectors with the NMS algorithm has reached the bottleneck in detecting adjacent ships. In Column 4, large size ships are hard to detect due to insufficient amounts of corresponding samples in the training set, which accords with statistical AP_L in Table 4.

TABLE 5. Ship detection in the inshore and offshore scenes of HRSID.

Model	Scenes	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN	Inshore	51.4	78.3	58.1	50.4	64.0	24.1
	Offshore	80.7	98.0	94.5	82.0	78.2	31.3
Cascade R-CNN	Inshore	55.9	79.6	63.6	54.5	69.6	32.7
	Offshore	83.6	98.0	95.5	84.9	81.1	65.4
RetinaNet	Inshore	41.3	69.0	42.5	39.4	57.9	28.4
	Offshore	79.6	98.6	93.2	81.2	75.1	57.4
Mask R-CNN	Inshore	53.1	79.0	60.7	52.5	63.6	20.0
	Offshore	81.0	98.8	94.6	82.3	79.0	44.9
Mask Scoring R-CNN	Inshore	52.8	78.6	60.8	52.3	64.7	24.9
	Offshore	80.2	98.0	94.6	81.6	77.9	43.4
Cascade Mask R-CNN	Inshore	56.3	80.0	64.9	55.5	67.6	24.8
	Offshore	84.1	98.9	95.6	85.2	81.9	59.0
Hybrid Task Cascade	Inshore	61.8	82.7	71.0	59.9	77.4	48.8
	Offshore	87.1	99.0	96.8	88.5	86.9	65.5
HRSDNet	Inshore	58.9	81.3	68.3	57.7	72.3	30.1
	Offshore	84.7	98.6	96.0	86.1	82.3	68.2

D. SHIP DETECTION IN INSHORE AND OFFSHORE SCENES

Ship detection in the pure sea background is less challenging to the CNN-based detectors as there are no interferential objects in these scenes. So, we have divided HRSID into inshore and offshore scenes to measure the capability of state-of-the-art detectors in detecting ships with interferences. Statistically, inshore scenes occuy the proportion of 18.4%, and offshore scenes occupy the proportion of 81.6%. In Table 5, we have shown the ship detection results in inshore and offshore scenes. As for detecting the offshore scenes, AP, AP_{50}, and AP_{75} of state-of-the-art detectors is above 79.6%, 98%, 93.2%, respectively. When detecting the small, medium and large ships, the highest bounding box AP among the detectors is still maintained at 88.5%, 86.9%, 68.2%, respectively. While detecting the inshore scenes, the detection precision has dropped significantly; the bounding box AP of all detectors ranges from 41.3% to 61.8%. Compared with the offshore scenes, AP_{50} and AP_{75} for inshore scenes has reduced by 20% and 22% respectively. As for detecting small, medium and large ships, the highest bounding box AP is 59.9%, 77.4% and 48.8% respectively. In summary, detectors can precisely detect the ships in offshore scenes but the inshore scenes in HRSID are still challenging to the state-of-the-art detectors.

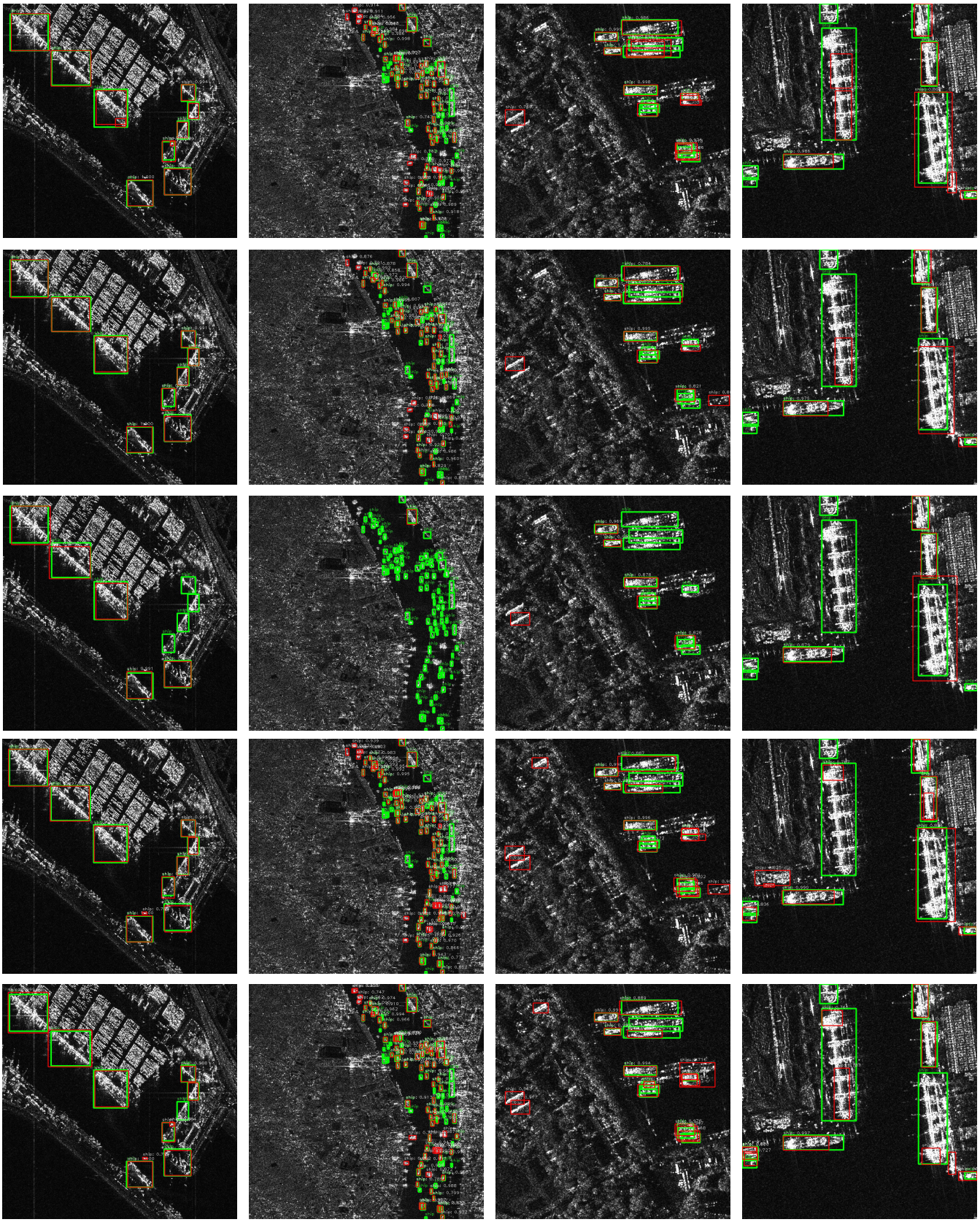


FIGURE 10. Visible ship detection results of state-of-the-art detectors with ResNet50-FPN backbone on complex detection scenes from the test set of HRSID. Green bounding box denotes ground truth and red bounding box denotes predicted results.

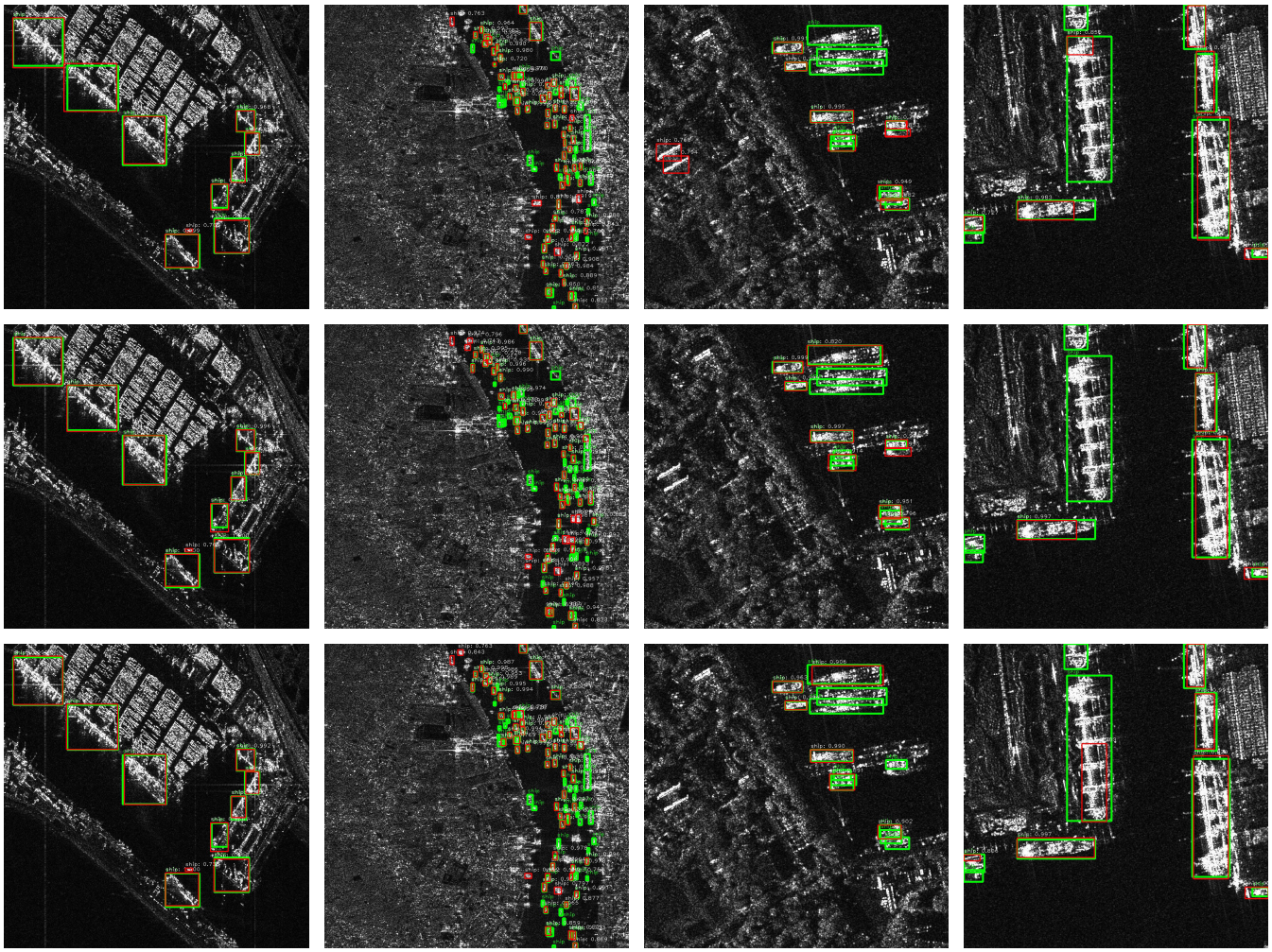


FIGURE 10. ***(Continued.)*** **Visible ship detection results of state-of-the-art detectors with ResNet50-FPN backbone on complex detection scenes from the test set of HRSID. Green bounding box denotes ground truth and red bounding box denotes predicted results.**

E. COMPARISON WITH THE BASELINE OF SSDD

Compared to the ship chips in OpenSARship and SAR-Ship-Dataset which are dedicated to the ship classification [42], SAR images in SSDD can be better implemented in ship detection [28]–[30], [36], [38]; researchers tend to augment the images in SSDD to fix its flaws when developing algorithms. To verify that HRSID is more applicable to CNN-based detectors, we have experimented on SSDD for further comparison in detection precision. All the detectors and corresponding hyperparameters are consistent with the experiments on HRSID when measuring the baseline of SSDD. The baseline of SSDD is shown in Table 6.

SSDD is randomly divided into the training set (65% SAR images) and test set (35% SAR images). The model size of the state-of-the-art detectors is the same as HRSID. Since the training set of SSDD has 4.8 times fewer SAR images than the training set of HRSID, the detection speed of the model trained by SSDD has decreased for about 0.02s. Under the bounding box IoU threshold of 0.5, the bounding box AP of the models trained by the training set of SSDD is above 90%. But with a relatively strict bounding box IoU threshold of 0.75, bounding box AP has a sharp decrease for about 25%. So that ships are easy to detect but hard to locate precisely in the SSDD. While in HRSID, AP_{50} is decreased by about 7% than SSDD; but AP_{75} has increased by about 8% than SSDD. There is no sharp decrease between AP_{50} and AP_{75} in HRSID. Compared to the AP_S in HRSID, AP_S in SSDD is reduced for about 10% under the ResNet-FPN backbone. As for the AP_L in SSDD, it receives the abnormal results. There are 76 large size ships in SSDD and 50 large ships for training, but AP_L of detectors vary from 45.4% to 61.2%. The few training samples but relatively high precision shows that the feature of large ships hasn't been clearly distinguished from small and medium ships. In HRSID, the feature of the large ship in high-resolution SAR image is more detailed and distinguishable to small and medium ships as is shown in Figure 5. With the lack of large ships, statistics of AP_L in HRSID are lower but real than SSDD. To sum up, ship detection in HRSID is more challenging and detectors can locate the ships precisely than SSDD. Ship detection statistics in high-resolution SAR images are more authentical than low-resolution SAR images.

TABLE 6. The baseline of SSDD.

Model	Backbone	Model Size (Mb)	Test Speed	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN	ResNet-50+FPN	330.2	0.057s	59.1	93.8	68.6	55.2	66.0	47.3
	ResNet-101+FPN	482.4	0.074s	58.4	94.0	65.9	54.5	64.7	51.9
Cascade R-CNN	ResNet-50+FPN	552.6	0.070s	59.7	93.1	67.6	54.8	67.1	57.8
	ResNet-101+FPN	704.8	0.088s	60.3	94.0	69.6	56.0	66.6	59.3
RetinaNet	ResNet-50+FPN	290.0	0.052s	55.5	90.2	62.3	51.2	62.6	45.4
	ResNet-101+FPN	442.3	0.070s	55.2	90.8	60.2	50.9	62.2	49.7
Mask R-CNN	ResNet-50+FPN	351.2	0.060s	58.9	93.4	66.6	55.3	64.9	49.7
	ResNet-101+FPN	503.4	0.077s	59.4	93.9	67.7	54.9	66.2	53.9
Mask Scoring R-CNN	ResNet-50+FPN	481.1	0.059s	59.4	94.7	67.8	55.6	65.3	51.2
	ResNet-101+FPN	633.1	0.077s	59.8	94.9	68.2	54.9	67.0	53.8
Cascade Mask R-CNN	ResNet-50+FPN	615.6	0.075s	59.7	93.1	68.9	55.5	65.9	53.2
	ResNet-101+FPN	767.8	0.093s	59.9	92.4	68.9	55.6	66.4	58.3
Hybrid Task Cascade	ResNet-50+FPN	639.3	0.110s	61.1	94.5	70.3	55.9	68.4	59.2
	ResNet-101+FPN	791.6	0.128s	60.7	93.6	69.5	55.7	67.6	61.2
HRSDNet	HRFPN-W32	598.1	0.101s	61.1	93.9	70.1	56.6	67.7	58.8
	HRFPN-W40	728.2	0.120s	60.9	94.4	69.7	56.2	67.8	58.9

TABLE 7. Descriptions of Alos-2 SAR imagery.

Position	Waveband	Resolution (m)	Polarization	Time	Image Size (pixels)
Tokyo Bay	L	3	HH	2014-4-10	10389 x 6487

TABLE 8. Instance segmentation statistics generated by mask AP on test set of HRSID.

Model	Backbone	Model Size (Mb)	Test Speed	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Mask R-CNN	ResNet-50+FPN	351.2	0.083s	54.0	86.0	64.4	53.5	62.0	16.4
	ResNet-101+FPN	503.4	0.105s	54.3	86.0	64.3	53.6	63.2	21.3
Mask Scoring R-CNN	ResNet-50+FPN	481.1	0.079s	53.8	84.7	64.2	53.3	61.2	20.5
	ResNet-101+FPN	633.1	0.103s	54.4	85.8	64.6	54.0	62.3	17.2
Cascade Mask R-CNN	ResNet-50+FPN	615.6	0.097s	54.6	86.6	64.7	54.1	62.4	19.1
	ResNet-101+FPN	767.8	0.119s	54.7	86.9	65.5	54.1	63.7	21.2
Hybrid Task Cascade	ResNet-50+FPN	639.3	0.132s	55.2	86.5	66.1	54.3	65.4	28.5
	ResNet-101+FPN	791.6	0.156s	55.4	86.4	66.7	54.5	65.3	24.4

F. SHIP DETECTION RESULT ON ALOS-2

To examine the migration ability of the model trained on our dataset, we have obtained a panoramic Alos-2 SAR imagery with multiple inshore and offshore ships for the experiment. Detailed descriptions are shown in Table 7.

The size of large-scale SAR imagery doesn't fit the input of CNN based detectors. So, the detection process is divided into several steps. Firstly, the SAR imagery is vertically and parallelly cropped by 800 × 800 pixels sliding window; each successively cropped image has an overlapped ratio of 20% to ensure the stitching process can be implemented. Secondly, 187 cropped SAR images are inputted into the detectors to get the detection results. Thirdly, detection results are stitched to form the detected panoramic SAR imagery. The visible ship detection result of Cascade R-CNN with ResNet50-FPN is shown in Figure 11. Green bounding boxes denote ground truth and red bounding boxes denote predicted results.

The model trained by HRSID performs well in detecting the offshore ships, and there are few false alarms on the land. But as the man-made facilities or buildings in the port have a similar feature to the ships, false alarms and missing detections increase when the model detects inshore ships. To sum up, the model trained by HRSID has the migration ability to detect large size SAR imagery and is of value in practical application.

G. INSTANCE SEGMENTATION RESULTS ON STATE-OF-THE-ART DETECTORS

In Table 8, we have shown the instance segmentation statistics generated by mask AP on the test set of HRSID, which

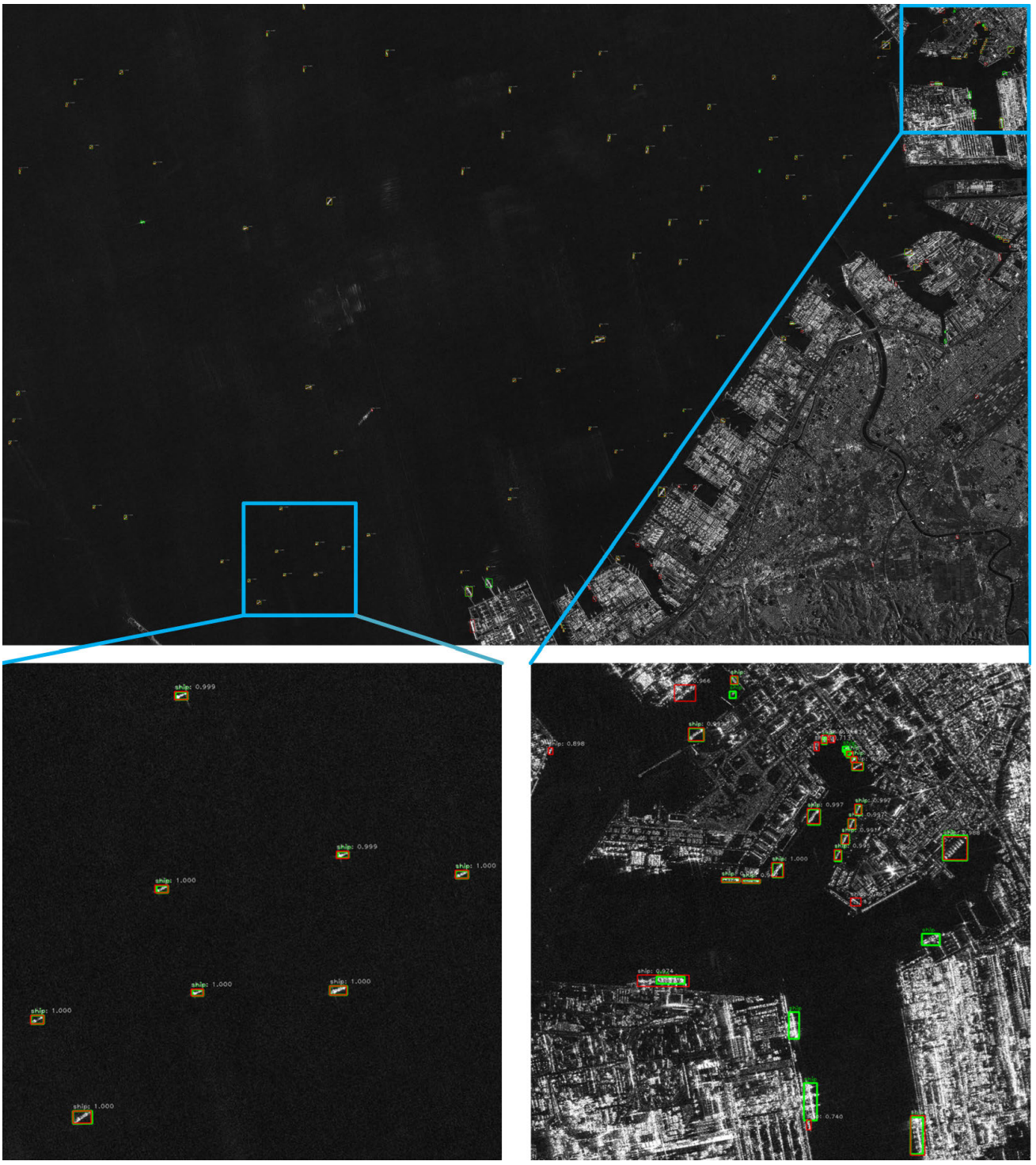


FIGURE 11. Visible ship detection result on Alos-2 SAR imagery of Cascade R-CNN with ResNet50-FPN. Green bounding boxes denote ground truth and red bounding boxes denote predicted results.

are the results of Mask R-CNN, Mask Scoring R-CNN, Cascade Mask R-CNN, and Hybrid Task Cascade. Similar to the bounding box AP in object detection, mask AP is generated by the IoU of the predicted mask and ground truth mask. Mask prediction is more complicated than bounding box prediction due to the alterable shape of the mask. So, the mask AP is slightly reduced compared to bounding box AP. In three-dimensional space, objects may lose some partial features due to occlusion from the same category; but ships are generally distributed on the surface of the water, occlusion of ships is rare. Mask IoU head has limited effects to improve the mask quality when segmenting the ships,

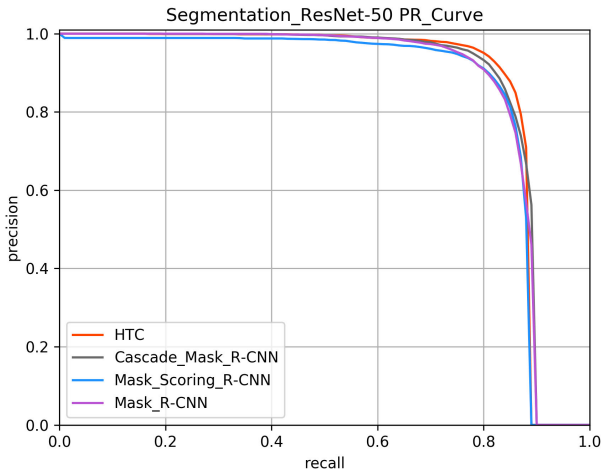


FIGURE 12. The segmentation PR curve with the backbone of ResNet-50.

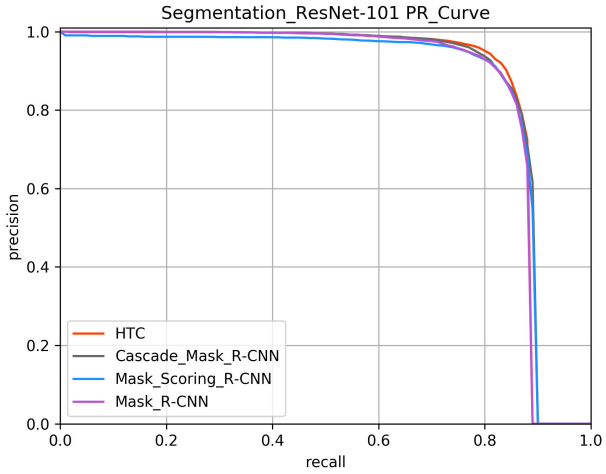


FIGURE 13. The segmentation PR curve with the backbone of ResNet-101.

and mask AP in Mask Scoring R-CNN is almost the same compared to Mask R-CNN. With the stepwise increasing IoU threshold in each cascade of instance segmentation structure, Cascade Mask R-CNN performs better in mask AP than Mask R-CNN. Benefit from the interactive bounding box and mask branch, the mask AP of HTC has improved by 0.6% and 0.7%, with the backbone of ResNet50-FPN and ResNet101-FPN respectively. Specifically, the mask AP has increased by 3% when detecting the medium ships.

On account of sharing the same model, model size and test speed per image of instance segmentation are the same as ship detection statistics. All the detectors have the mask AP for about 54%. Under the mask IoU threshold of 0.5, mask AP of Mask R-CNN, Cascade Mask R-CNN, and Hybrid Task Cascade is above 86%, 86.6%, and 86.4%, respectively. While with the relatively strict mask IoU threshold of 0.75, mask AP is above 64.3% for the detectors. As for multi-scale instance segmentation, when segmenting the small and medium ships in HRSID, mask AP of detectors is above 53.3% and 61.2%, respectively. When segmenting the large ships, mask AP is still low with a limited amount of large size ships. The statistics indicate that detectors can generate a more precise mask for medium ships than small ships. The precision-recall curve (PR curve) of each detector is shown in Figure 12 and Figure 13.

We have selected 4 representative scenes in the test set of HRSID for instance segmentation. Visible results are shown in Figure 14. Row 1 is the ground truth of bounding boxes and masks; the bounding boxes appear as green for distinguishing from the predicted red bounding boxes below. Row 2 and Row 3 show the instance segmentation results of Mask R-CNN with ResNet50-FPN backbone and ResNet101-FPN backbone, respectively. Row 4 and Row 5 represent Cascade Mask R-CNN with ResNet50-FPN and ResNet101-FPN, respectively.

Different from visible results in ship detection, the predicted mask in instance segmentation can depict the ships with concrete shape, which is beneficial to determine the type of ships. In the visible results of instance segmentation, detectors can segment offshore ships, berthed ships in the canal and berthed ship near man-made facilities well; but it appears overlapped masks in segmenting adjacent ships. Compared with the ResNet50-FPN backbone, detectors with ResNet101-FPN backbone can generate the more accurate mask.

V. DISCUSSION

With a different incident angle of the radar signal, environmental factor, polarization methods, etc., the pre-processed SAR imageries exist clutter noise which interferes with the feature of ships then the ship detection and instance segmentation with CNN. So, distinguished from constructing an optical remote sensing dataset for object detection and instance segmentation [54], ships should be accurately and completely annotated when constructing the SAR dataset for ship detection and instance segmentation. Existing SAR ship datasets prepared for CNN have respective defects, and detectors tend to reach too high AP_{50} when testing on these datasets [30], [37]. For example, the HR-SDNet, which is dedicated to CNN based ship detection, has reached 98.8% of AP_{50} when experimenting on SSDD [30].

In this paper, we have designed a complete and efficient process to construct a high-resolution SAR dataset for CNN based ship detection and instance segmentation. To avoid wrong annotation and missing annotation caused by man-made facilities which are similar to ships [37], we have developed the auxiliary method for annotation in Section II.B. As for ships in low-resolution SAR images are presented as highlighted spots, the effects of instance segmentation on low-resolution SAR images may be limited. So, high-resolution SAR imageries are applied to construct the dataset, and they are cropped to 800 $\times$ 800 pixels SAR images for better implementation on the functions such as multi-scale training. To comprehensively evaluate the performance of detectors, we have employed MS COCO dataset evaluation metrics for comprehensive evaluation on HRSID with

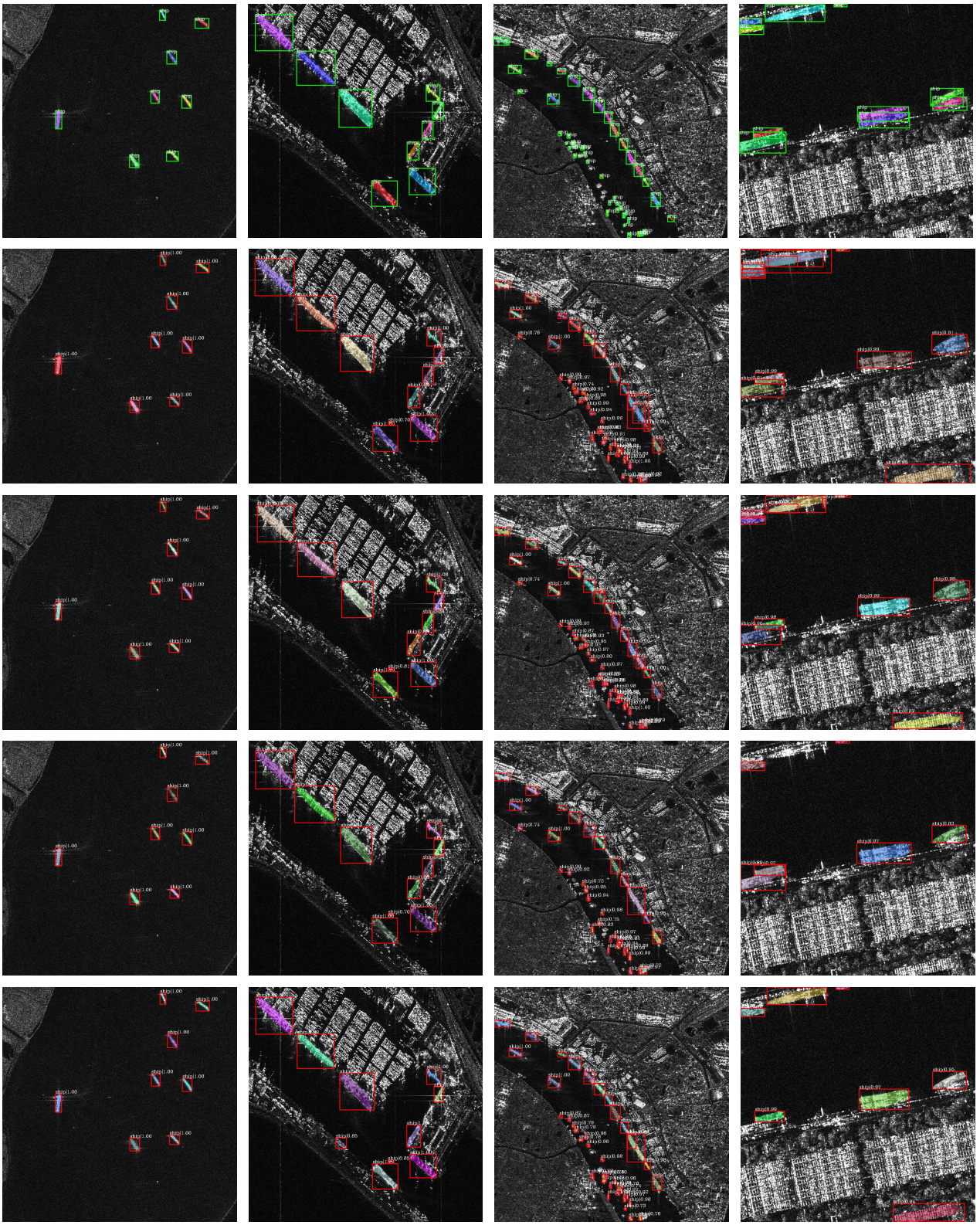


FIGURE 14. Visible instance segmentation results of Mask R-CNN and Cascade Mask R-CNN with ResNet50-FPN and ResNet101-FPN backbone on test set of HRSID. Row 1 is the ground truth. Row 2 to Row 5 are the predicted results.

8 state-of-the-art detectors. Quantitatively, detectors with improved structure have received proper improvement on bounding box AP when detecting ships; the maximum bounding box AP_{50} of the results is 89.3%. The AP_L of the detectors varies from 16.4% to 38.1% on account of insufficient amounts of large ships in HRSID, and this phenomenon

also appears in the existing SAR ship dataset prepared for CNN as is discussed in Section II.C. We are looking forward to the subsequent supplement from the community to perfect HRSID. The statistics of model size, test speed, and COCO evaluation metrics reasonably vary from detectors. To challenge the detectors and examine the ability of detectors to detect the ship in complex scenes, some complex scenes are added in HRSID and the correspondingly visible results of detectors are tested. The results show that the complex scenes such as cluster-distributed small ships and adjacent ships are still challenging to the detectors. As for the visible detection results in instance segmentation, the generated mask can authentically depict the distribution of ships with its concrete shape pixel-by-pixel, establishing a preliminary basis for further research on instance segmentation.

The statistical detection results are regarded as the baseline of state-of-the-art detectors, including ship detection and instance segmentation statistics. With reasonable AP to state-of-the-art detectors and challenging detection scenes, HRSID is worth further research to promote the development of ship detection and instance segmentation. The novel structure and algorithms can also be tested on HRSID. We hope HRSID can promote the development of ship detection and instance segmentation in SAR images just like MS COCO in optical images.

Future work will be conducted on ship detection and instance segmentation with HRSID. Existing problems in our experiments, such as poor instance segmentation performance in cluster-distributed small ships and adjacent ships are the major indicator of our further research.

VI. CONCLUSION

In this research, we have constructed a high-resolution SAR dataset for CNN based ship detection and instance segmentation. 136 SAR imageries with resolution under 5m are cropped to 5604 SAR images with 800 x 800 pixels. When building the baseline of our dataset, we have applicated 8 state-of-the-art detectors to our dataset for ship detection and instance segmentation. The large size SAR imagery is used for examining the migration ability of the model trained on our dataset. Besides, we have measured the baseline of SSDD to verify the novelty of HRSID. The experimental results reveal (1) the process we have designed for constructing HRSID is effective as the statistical results of detectors are reasonable; (2) ship detection and instance segmentation can be implemented on HRSID, and the predicted pixel-by-pixel mask can depict the shape of ships which is beneficial to determine the type of ships; (3) the baseline of HRSID shows its superiority compared to SSDD; (4) The model trained by HRSID can detect ships in large size SAR imagery and is of value in practical application. We hope HRSID can promote the development of ship detection and instance segmentation.

REFERENCES

[1] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 3–22, Nov. 2018.

[2] K. El-Darymli, E. W. Gill, P. Mcguire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, Sep. 2016.

[3] K. El-Darymli, P. McGuire, D. Power, and C. R. Moloney, "Target detection in synthetic aperture radar imagery: A state-of-the-art survey," *J. Appl. Remote Sens.*, vol. 7, no. 1, pp. 7–35, Mar. 2013.

[4] Z. Zhao, K. Ji, X. Xing, H. Zou, and S. Zhou, "Ship surveillance by integration of space-borne SAR and AIS—Review of current research," *J. Navigat.*, vol. 67, no. 1, pp. 177–189, Jan. 2014.

[5] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, and I. N. Traver, "GMES sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.

[6] W. Pitz and D. Miller, "The TerraSAR-X satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 615–622, Feb. 2010.

[7] A. Marino, M. Sanjuan-Ferrer, I. Hajnsek, and K. Ouchi, "Ship detection with spectral analysis of synthetic aperture radar: A comparison of new and well-known algorithms," *Remote Sens.*, vol. 7, no. 5, pp. 5416–5439, Apr. 2015.

[8] X. Huang, W. Yang, H. Zhang, and G.-S. Xia, "Automatic ship detection in SAR images using multi-scale heterogeneities and an a contrario decision," *Remote Sens.*, vol. 7, no. 6, pp. 7695–7711, Jun. 2015.

[9] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.

[10] S.-Q. Huang, D.-Z. Liu, G.-Q. Gao, and X.-J. Guo, "A novel method for speckle noise reduction and ship target detection in SAR images," *Pattern Recognit.*, vol. 42, no. 7, pp. 1533–1542, Jul. 2009.

[11] H. Wang, F. Xu, and S. Chen, "Saliency detector for SAR images based on pattern recurrence," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2891–2900, Jul. 2016.

[12] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[13] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[14] X. Qin, S. Zhou, H. Zou, and G. Gao, "A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 806–810, Jul. 2013.

[15] G. Gao and G. Shi, "CFAR ship detection in nonhomogeneous sea clutter using polarimetric SAR data based on the notch filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4811–4824, Aug. 2017.

[16] G. Gao, L. Liu, L. Zhao, G. Shi, and G. Kuang, "An adaptive and fast CFAR algorithm based on automatic censoring for target detection in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1685–1697, Jun. 2009.

[17] X. Wang, C. Chen, Z. Pan, and Z. Pan, "Fast and automatic ship detection for SAR imagery based on multiscale contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1834–1838, Dec. 2019.

[18] L. Zhai, Y. Li, and Y. Su, "Inshore ship detection via saliency and context information in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1870–1874, Dec. 2016.

[19] 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.

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

[21] Y. Liu, M.-H. Zhang, P. Xu, and Z.-W. Guo, "SAR ship detection using sea-land segmentation-based convolutional neural network," in *Proc. Int. Workshop Remote Sens. Intell. Process. (RSIP)*, May 2017, pp. 1–4.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[23] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

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

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual network," in *Proc. Europ. Conf. Comp. Visi (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 630–645.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[27] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

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

[29] W. Fan, F. Zhou, X. Bai, M. Tao, and T. Tian, "Ship detection using deep convolutional neural networks for PolSAR images," *Remote Sens.*, vol. 11, no. 23, p. 2862, Dec. 2019.

[30] S. Wei, H. Su, J. Ming, C. Wang, M. Yan, D. Kumar, J. Shi, and X. Zhang, "Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet," *Remote Sens.*, vol. 12, no. 1, p. 167, Jan. 2020.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[32] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[33] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

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

[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Europ. Conf. Comp. Visi (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[36] T. Zhang and X. Zhang, "High-speed ship detection in SAR images based on a grid convolutional neural network," *Remote Sens.*, vol. 11, no. 10, p. 1206, May 2019.

[37] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, p. 531, Mar. 2019.

[38] Zhang, Zhang, Shi, and Wei, "Depthwise separable convolution neural network for high-speed SAR ship detection," *Remote Sens.*, vol. 11, no. 21, p. 2483, Oct. 2019.

[39] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Venice, Italy, Oct. 2017, pp. 2961–2969.

[40] K. Chen, W. Ouyang, C. C. Loy, D. Lin, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, and J. Shi, "Hybrid task cascade for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4974–4983.

[41] L. Huang, B. Liu, B. Li, W. Guo, W. Yu, Z. Zhang, and W. Yu, "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.

[42] J. Li, C. Qu, and S. Peng, "Ship classification for unbalanced SAR dataset based on convolutional neural network," *J. Appl. Remote Sens.*, vol. 12, no. 3, Aug. 2018, Art. no. 035010.

[43] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019.

[44] J. Shao, C. Qu, J. Li, and S. Peng, "A lightweight convolutional neural network based on visual attention for SAR image target classification," *Sensors*, vol. 18, no. 9, p. 3039, Sep. 2018.

[45] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era: Models, Methods Appl. (BIGSARDATA)*, Nov. 2017, pp. 1–6.

[46] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017.

[47] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Proc. Europ. Conf. Comp. Visi (ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 740–755.

[48] (2020). *High-Resolution SAR Images Dataset*. [Online]. Available: https://github.com/chaozhong2010/HRSID

[49] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.

[50] J. C. Curlander, and R. N. McDonough, *Synthetic Aperture Radar-Systems and Signal Processing*. New York, NY, USA: Wiley, 1991.

[51] A. Ludwig, "The definition of cross polarization," *IEEE Trans. Antennas Propag.*, vol. AP-21, no. 1, pp. 116–119, Jan. 1973.

[52] C. Oliver, and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Chennai, India: SciTech, 2004.

[53] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. Europ. Conf. Comp. Visi (ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 391–405.

[54] S. W. Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. S. Khan, and X. Bai, "iSAID: A large-scale dataset for instance segmentation in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops. (CVPRW)*, Long Beach, CA, USA, Jun. 2019, pp. 28–37.

[55] K. Wada. (2016). *Labelme: Image Polygonal Annotation With Python*. [Online]. Available: https://github.com/wkentaro/labelme

[56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[57] 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, Sep. 2009.

[58] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[59] H. Su, S. Wei, M. Yan, C. Wang, J. Shi, and X. Zhang, "Object detection and instance segmentation in remote sensing imagery based on precise mask R-CNN," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 1454–1457.

[60] C. Wang, X. Bai, S. Wang, J. Zhou, and P. Ren, "Multiscale visual attention networks for object detection in VHR remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 310–314, Feb. 2019.

[61] C. Wang, J. Shi, X. Yang, Y. Zhou, S. Wei, L. Li, and X. Zhang, "Geospatial object detection via deconvolutional region proposal network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3014–3027, Aug. 2019.

[62] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[63] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6409–6418.

[64] K. Chen *et al.*, "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*. [Online]. Available: http://arxiv.org/abs/1906.07155

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

[66] Z. Cui, M. Zhang, Z. Cao, and C. Cao, "Image data augmentation for SAR sensor via generative adversarial nets," *IEEE Access*, vol. 7, pp. 42255–42268, 2019.

[67] M. Zhang, Z. Cui, X. Wang, and Z. Cao, "Data augmentation method of SAR image dataset," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 5292–5295.

[68] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

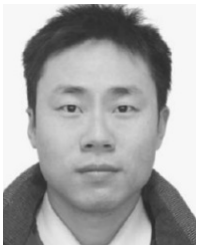
SHUNJUN WEI (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China, in 2002, 2005, and 2009, respectively. He is currently an Associate Professor. His research interests include radar signal processing, machine learning, and synthetic aperture radar systems.

XIANGFENG ZENG received the B.S. degree in information and communication engineering from the University of Electronic Science and Technology of China, in 2019, where he is currently pursuing the M.S. degree with the School of Information and Communication Engineering. His current research interests include SAR ship detection and machine learning.

QIZHE QU (Graduate Student Member, IEEE) received the B.S. degree in information countermeasure technology from the North University of China, in 2018. He is currently pursuing the M.S. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His current research interests include radar signal recognition and machine learning.

MOU WANG (Student Member, IEEE) received the B.S. degree in communication engineering from the Chongqing University of Posts and Telecommunications. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His current research interests include compressed sensing, SAR imaging, and machine learning.

HAO SU (Graduate Student Member, IEEE) is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include SAR ship detection and machine learning.



JUN SHI (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China, in 2002, 2005, and 2009, respectively. He is currently an Associate Professor. His research interests include radar signal processing and synthetic aperture radar systems.