Multi-Scale Ship Detection from SAR and Optical Imagery via A More Accurate YOLOv3

Zhonghua Hong, *Member, IEEE,* Ting Yang, Xiaohua Tong*, *Senior Member, IEEE,* Yun Zhang*, Shenlu Jiang, Ruyan Zhou, Yanling Han, Jing Wang, Shuhu Yang and Sichong Liu, *Member, IEEE*

Abstract—Deep learning detection methods use in ship detection remains a challenge, owing to the small scale of the objects and interference from complex sea surfaces. In addition, existing ship detection methods rarely verify the robustness of their algorithms on multi-sensor images. Thus, we propose a new improvement on the 'You Only Look Once' version 3 (YOLOv3) framework for ship detection in marine surveillance, based on synthetic aperture radar (SAR) and optical imagery. First, improved choices are obtained for the anchor boxes by using linear scaling based on the k-means++ algorithm. This addresses the difficulty in reflecting the advantages of YOLOv3's multi-scale detection, as the anchor boxes of a single detection target type between different detection scales have small differences. Second, we add uncertainty estimators for the positioning of the bounding boxes by introducing a Gaussian parameter for ship detection into the YOLOv3 framework. Finally, four anchor boxes are allocated to each detection scale in the Gaussian-YOLO layer instead of three as in the default YOLOv3 settings, as there are wide disparities in an object's size and direction in remote sensing images with different resolutions. Applying the proposed strategy to 'YOLOv3-spp' and 'YOLOv3-tiny', the results are enhanced by 2-3%. Compared with other models, the improved-YOLOv3 has the highest average precision (AP) on both the optical (93.56%) and SAR (95.52%) datasets. (2) The improved-YOLOv3 is robust, even in the context of a mixed dataset of SAR and optical images comprising images from different satellites and with different scales.

***Index Terms*—Ship detection, deep learning-based object detection, YOLOv3, SAR and optical imagery.**

Manuscript received August 25, 2020. This work was supported in part by the National Key R&D Program of China under Grant 2018YFB0505400 and National Natural Science Foundation of China under Grant 41871325. (Corresponding author: Xiaohua Tong and Yun Zhang)

Z. Hong, T. Yang, Y. Zhang, R. Zhou, Y. Han, J. Wang, S. Yang, are with the College of Information Technology, Shanghai Ocean University, Shanghai 201306, China and also with the Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai Ocean University, Shanghai 201306, China (e-mail: zhhong@shou.edu.cn; m180701065@st.shou.edu.cn, y-zhang@shou.edu.cn, ryzhou@ shou.edu.cn, ylhan@shou.edu.cn, wangjing@shou.edu.cn, shyang@shou.edu.cn)

S. Jiang is with the Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China (e-mail: jestshen@hotmail.com).

X. Tong, S. Liu are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: xhtong@tongji.edu.cn, sicong.liu@tongji.edu.cn)

I. Introduction

Ship detection, in the context of maritime monitoring services, has a large number of potential applications, including those in ocean environment monitoring, illegal fishing prevention, rescue and disaster relief, irregular migration detection, and military reconnaissance. Ship detection can be narrowly defined as determining ship bounding boxes in images, and then locating the ships. The proliferation of remote sensing technology [1] has provided rapid data support for ship detection. Synthetic aperture radar (SAR) imagery is common in the maritime monitoring literature. This is because ships can be easily detected, independent of the weather and time [2]. However, SAR ship detection has several limitations: (1) it is susceptible to noise interference; (2) it is vulnerable to strong winds and waves; (3) small objects are difficult to locate and identify, and (4) different ship types are difficult to differentiate [3]. Optical images are worthy further investigation for ship detection, as they provide higher resolution and more visualised content. Moreover, many of them are free, and provide copious amounts of satellite data. Nonetheless, in optical images, weather conditions and sunlight can hinder the ship detection performance. An effective detection algorithm should not only perform well in various scenarios (such as clouds and fog, complex sea conditions, ports, and inshore ships), but should also support various forms of satellite sensor data. Therefore, optical and SAR data should be consolidated to verify the robustness of the algorithm.

Among existing ship detection methods, the constant false alarm rate (CFAR) method is based primarily on the intensity differences between ships and sea clutter [4]– [6]. The CFAR process depends on the statistical distribution of the cluttered backgrounds, and is implemented using a sliding window technique. Because of this, it is difficult for CFAR methods to efficiently process large collections of remote sensing imagery [7], [8]. More importantly, hidden deeper information, such as spectra, textures, and geometry, are not fully incorporated. Thus, detection effects of conventional ship detection approaches are often insufficient for the tasks at hand.

Object detection based on deep learning methods has attracted extensive research in recent years, and has gradually supplanted methods based on handcrafted features and machine learning. Deep learning uses a set of machine learning algorithms to learn deeper features via a deep architecture [9]. Thus, it is possible to extract deeper features and richer semantic information for ship detection from complex remote sensing images by constructing a deep learning model. Object detection is a computer vision task of combining object localisation and recognition. Currently, there are two major

branches in object detection based on deep learning. The first branch is based on an object region proposal (usually two-stage detector) approach. This includes two stages. First, candidate object proposals are generated using a regional generation algorithm. Then, features are extracted from the candidate object proposals via a CNN. Such as region-based CNN (R-CNN)[10], Fast R-CNN [11], Faster R-CNN[12], Mask R-CNN[13], Cascade R-CNN[14], Libra R-CNN[15] algorithms. The second branch is a regression-based detection (usually an one-stage detector) framework, which converts object detection to regression processing to directly predict target's coordinate and category. Such as 'You Only Look Once' (YOLO)[16], single-shot detection (SSD)[17], Deconvolutional Single Shot Detector(DSSD)[18],YOLOv2[19],YOLOv3[20],YOLOv4[21],CornetNet[22], CenterNet[23]. The development process of object detection based on deep learning is shown in Fig. 1.

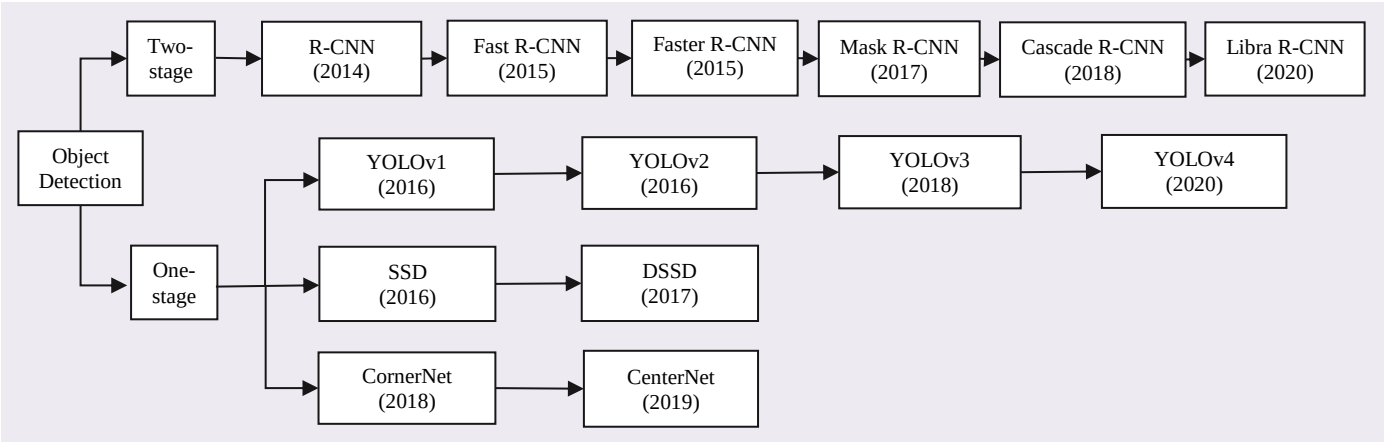


Fig. 1. Development process of two main lines (two-stage detector and one-stage detector) for classic object detection based on deep learning (SSD: Single-shot detection; R-CNN: Region-based convolutional neural network; YOLO: You only look once).

In recent years, deep learning has proved to be more effective than traditional detection methods. However, deep learning based on object detection tends to underperform when applied to remote sensing image-based ship detection. This is the fact that it is difficult to determine a detection model for small targets (e.g., ships) in remote sensing images with multiple scenes and complex backgrounds. A remote sensing image has a larger field of view and contains more numerous and complex targets than a natural scene image. This is therefore difficult to distinguish between small and clustered targets. Furthermore, ship detection from SAR and optical images is susceptible to noise (such as clouds, waves, and shadows) and ship-like objects (such as bridges, docks, and ports). These pose challenges to ship location.

Li et al. [24] improved the accuracy of the Faster R-CNN in ship detection by using feature fusion, transfer learning, and hard negative mining for SAR imagery. Similarly, Cui et al. [25] presented an improved faster R-CNN based on a dense attention pyramid network for SAR imagery. Kang et al. [26] detected ships by applying a traditional CFAR approach to a faster R-CNN for SAR imagery. Wang et al. [27] addressed SAR ship detection by applying transfer learning from SSD by considering the detection accuracy and speed. Chang et al. [28] detected ships by utilizing a YOLOv2 model for two SAR ship datasets. Cui et al. [29] realized large-scale SAR ship detection by adding spatial shuffle-group enhance attention to CenterNet[30]. In general, there are fewer reports of ship detection based on optical images than there are those based on SAR images. Bi et al. [31] introduced an optical ship detection method based on visual attention deep salient object detection and SSD methods. Li et al. [32] detected multi-scale ships by expediting the R-CNN model for optical imagery. Wu et al. [33] detected inshore ships by applying CNNs to optical satellite images. These methods did not consider whether the models supported data from multiple satellite sensors, and were only tested on SAR or optical images. An excellent detection algorithm should not only have high detection precision and recall, but should also consider the detection efficiency. The robustness of the model needs to be considered. In other words, a detection algorithm should provide fast detection and support various types of satellite sensors. It should also serve as suitable for the remote sensing of multi-scene, multi-scale, and multi-resolution images.

In this study, an improved-YOLOv3 model was utilised for ship detection. First, to determine the characteristics of the ships (e.g., narrow shapes and small sizes), an improved k-means++ [34] algorithm was used to obtain accurate anchor boxes for the ships. Second, a Gaussian model was introduced to predict the uncertainty of the bounding boxes. Third, four anchors are assigned a detection scale in the Gaussian-YOLO detection layer to improve the robustness of the model. In addition, the algorithm was adjusted to the ship object detection task through a reasonable parameter adjustment. We trained and evaluated the state-of-the-art detection methods Faster R-CNN, SSD, and YOLOv3, and the improved-YOLOv3 on ship detection benchmark datasets including multi-scene, multi-resolution, and multi-size optical and SAR images. In summary, the contributions of this study are as following.

(1) Anchor boxes generated by a traditional k-means clustering method cannot reflect the advantages of YOLOv3's multi-scale output. This study addresses the problem regarding the centralised distribution of anchor

boxes for detecting a single target type in the dataset. The improved k-means++ clustering algorithm linearly scales the anchor boxes to count the numbers and shapes of the anchor boxes from the training dataset, according to the characteristics of the ships.

(2) The original YOLOv3 model only outputs the detection object's location information (x, y, w, h); this is a poor reflection of the bounding boxes' reliability. Therefore, the Gaussian model is introduced to output the uncertainty of each prediction bounding box, and to improve YOLOv3's detection accuracy.

(3) Four anchors are assigned to a detection scale in the Gaussian-YOLO detection layer to improve the robustness of the model for multi-scale object detection.

(4) Two ship datasets, an SAR dataset and an optical dataset, are used to verify the effectiveness of the proposed method. In addition, we propose mixing two ship datasets (including multi-scale, multi-resolution, and multi-scene images) to train and verify the model, so as to verify its robustness.

The remainder of this paper is organised as following. Section 2 reviews the experimental dataset. Details of the ship detection methodology and its implementation are described in Section 3. A series of comparative experimental results are presented and discussed in Section 4. The conclusions are outlined in Section 5.

II. EXPERIMENT DATASET

The ship detection algorithm must be valid for images of various sources, sizes, and resolutions. Thus, the experiment used two ship datasets: an optical ship dataset, and a SAR ship dataset. Two datasets were used to verify the effectiveness of the proposed method. Optical and SAR imagery could be used individually, or as a source for ship detection datasets. The implementation of the datasets used in the experiments is detailed in the following sections.

A. Synthetic aperture radar (SAR) ship dataset

The public SAR dataset was supplied by Wang et al. [35], in a format similar PASCALAL visual object class (VOC) dataset. This dataset includes ships in multiple scenarios and resolutions. Accordingly, this was adequate experimental data for ship detection.

TABLE I
SYNTHETIC APERTURE RADAR (SAR) SHIP IMAGERY DETAILS

Sensor	Image Mode	Resolution (m)	Swath (km)	Incident Angle (°)	Images
GF-3	UFS	3 × 3	30	20–50	12
GF-3	FSI	5 × 5	50	19–50	20
GF-3	QPSI	8 × 8	30	20–41	20
GF-3	FSII	10 × 10	100	19–50	30
GF-3	APSII	25 × 25	40	20–38	20
Sentinel-1	SM	1.7 to 4.9	80	20–45	98
Sentinel-1	IW	20 × 22	250	29–46	10

The data were labelled by SAR experts, and were collected from 102 Chinese Gaofen-3 images and 108 Sentinel-1 images. They dataset comprised 43,819 ship chips with 256 pixels each (for both the range and azimuth). The data included SAR imagery at resolutions of 3 m, 5 m, 8 m, and 10 m. Table I provide additional details on the SAR ship dataset.

B. Optical ship dataset

Optical ship dataset was provided by Kaggle for the 'Airbus Ship Detection Challenge' [36]. The dataset included tankers of various shapes and sizes, and shipping and fishing vessels located in open seas, docks, and marinas. Optical ship dataset comprised 150,000 ship chips with 768 × 768 pixels each, as extracted from SPOT satellite imagery at a resolution of 15 m. Many of the images contained no ships. Code was used to transform the data into common objects in context (COCO) format-data annotations according to a supplemental comma-separated values file. This approach provided oriented bounding boxes around the ships (in a run-length encoding format). The four primary steps were: (1) deleting images without ships, (2) searching for bad comments, (3) transforming the run-length encoding data into COCO format-data annotations, and (4) converting the COCO format-data annotations into a format identical to that of the PASCAL VOC format-data annotations.

Multi-scale, multi-resolution, and multi-sensor SAR and optical images served as the experimental data. These datasets contain various scenarios, such as cloudy and rainy weather, and complex backgrounds. For the experiment, the SAR and optical ship datasets were divided into training and test images at a ratio of 7:3. A random selection of images from the training images at a ratio of 0.2 served as the validation images. Therefore, there were 24,538 training images, 6,134 validation images, and 13,148 test images in the SAR ship dataset. For the optical ship dataset, 29,070 ship clips were collected after processing; 20,349 images were used for training, and 8,721 images were used for testing.

In addition, this study considered a mixed dataset of SAR and optical images from different satellites, scales, and channel modes (three-channel RGB optical images and one-channel SAR images). The experiments were performed on this mixed dataset, along with the other experiments, provided strong evidence for verifying the robustness of the proposed method. There were 40,817 training images, 10,205 validation images, and 21,867 test images in the mixed (optical + SAR) ship dataset.

III. SHIP DETECTION METHOD

In this section, principles of object detection based on deep learning, current mainstream methods, and the improved-YOLOv3 approach are discussed.

A. Object detection network structure

Deep learning-based object detection usually involves related concepts, such as anchor boxes, feature extraction, non-maximum suppression (NMS), and multi-scale detection. To obtain an object in an image, an algorithm must be employed to generate anchor boxes with fixed widths and heights, according to the characteristics of the target shape in the dataset. This improves the detection efficiency, owing to the uncertain position and shape of the target object. Feature extraction is a process that (traditionally) uses neural networks to propagate

candidate regions forward, combines data labels and loss functions to obtain position and category information of a detection target, and then iteratively updates model parameters through back propagation. NMS algorithm filters were overlapping boxes with low confidence, aiming to solve the problem of duplicate target detection. In a process of feature extraction, low-level feature maps usually have higher resolution and finer-grained features. This is appropriate for detecting small objects. As the receptive field increases, the resolution of higher-level feature maps decreases. This allows for additional semantic information and coarse representations, which are suitable for detecting large objects [37], [38]. Therefore, most current multi-scale object detection methods either independently detect multiple feature maps extracted from different layers of a network (such as in SSD), or fuse multiple feature maps extracted from different layers of a network (such as YOLOv3) [39].

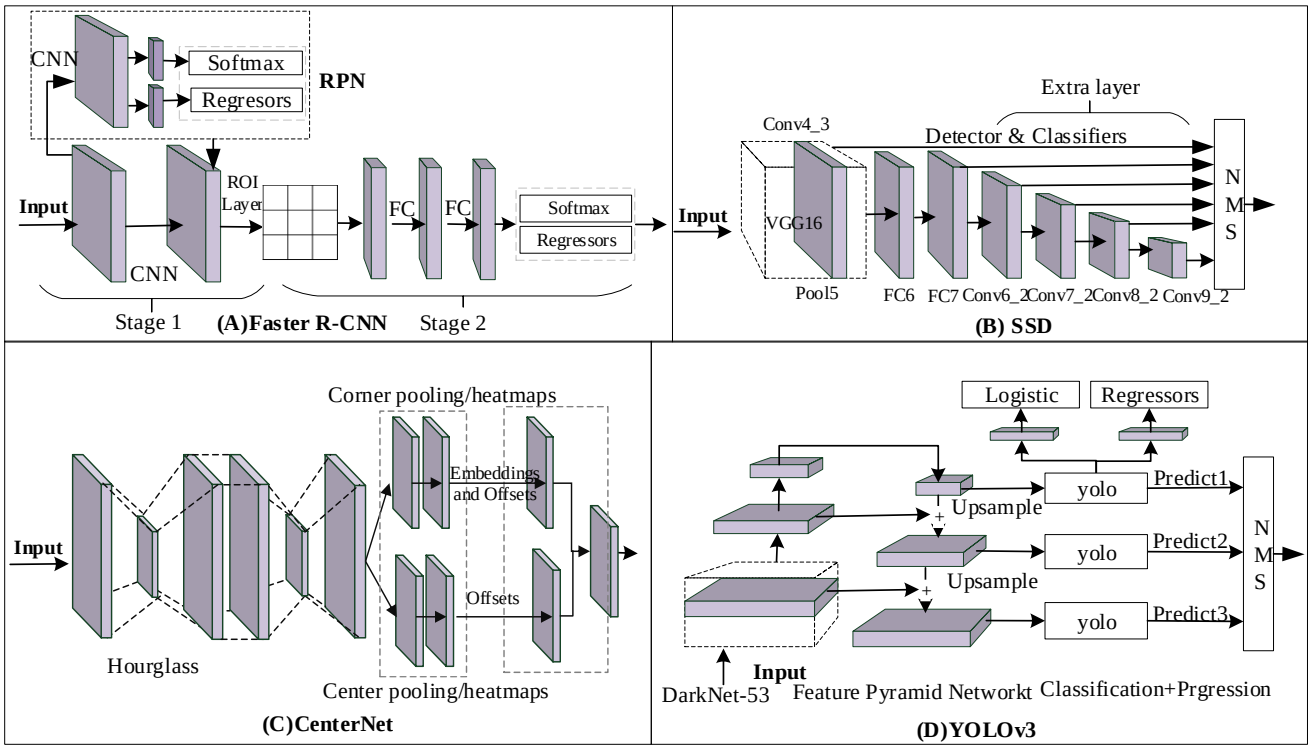


Fig. 2. A comparison of the Faster R-CNN, SSD, CenterNet, and YOLOv3 network frameworks.

The current mainstream frameworks in object detection based on deep learning incorporate the faster R-CNN, SSD, CenterNet, and YOLOv3 algorithms. In this study, the working principles of these four algorithms and possible corresponding problems in ship detection were analysed and compared to determine the most suitable framework for ship detection. Network structures of Faster R-CNN, SSD, CenterNet, and YOLOv3 are shown in Fig.2. The four algorithms are discussed in detail in the following sections.

1) Faster region-based convolutional neural network (R-CNN)

The network structure of the faster R-CNN, a two-stage object detector model, is shown in Fig. 2(A). First, the feature maps are obtained in accordance with the CNN; then, the RPN is employed to generate regional proposals. Next, the detection subnetwork uses these proposals to refine the detection results. In the second stage, the ROI pooling adjusts the object proposals into shapes of uniform sizes, and enables convolutional feature extraction for each proposal in each feature map. Finally, these converted object proposals are entered into the fully connected layer to output the object category and position of the bounding boxes. However, the improvement in detection accuracy comes at the cost of slower detection. Therefore, obtaining the results from the two-stage detector will be a slow process, and it will be difficult to meet real-time ship detection requirements.

2) Single-shot detection (SSD)

SSD is an one-stage object detector model. In the initial configuration, VGG16[40] was used as the basic network, and an anchor box was introduced. Several new convolutions were added after VGG16, based on combining low-level features to high-level features to enable multi-scale detection. As demonstrating in Fig. 2(B), SoftMax is currently used for background and object classification processing on each feature layer behind conv4_3, and the target is located using border regression. After the NMS, the object' s position is output. Regression simplifies the computational complexity of the network. However, SSD may miss small targets, such as ships. As the low-level features used for detecting small objects in the network have only one Conv4_3 layer, the feature expression ability and detail provided may be insufficient.

3) CenterNet

CenterNet is an one-stage keypoint-based object detector model that detects an object as a triplet, i.e. the top-left corner, bottom-right corner, and centre keypoint of a bounding box. The stacked hourglass [41] and HRNet-W64 [42] networks are used as backbones for comparison experiments. Centre pooling and cascade-corner pooling are introduced into CenterNet to enrich the centre and corner information. As demonstrating in Fig. 2(C), the backbone network uses cascaded corner pooling and central pooling to output two corner and centre keypoint heatmaps, and predicts the offsets. Then, pairs of corner detection are used to identify potential bounding boxes, and finally, detected central point is used to determine the final bounding box. However, this approach requires more complicated post-processing to group keypoints belonging to the same object (such as calculating distances).

4) You only look once (YOLO)v3

YOLOv3 is an one-stage object–detector model. The feature maps are resized to a uniform size, and are divided into S × S cell grids. Each cell within a target centre is responsible for the object category and position of the bounding boxes. Additionally, each grid cell predicts three bounding boxes based on three anchors. In YOLOv3, DarkNet-53[20] with residual skip connection [43] serves as the backbone network, as it can solve the vanishing gradient problem. As shown on the left side of Fig. 2(D), FPN is introduced to improve the multi-scale detection accuracy. The large feature map and up-sampled feature map are connected through an up-sampling operation. The FPN then forecasts objects from three different scales via a top-down pathway and lateral connection structure. Therefore, it efficiently constructs a multi-scale feature pyramid for obtaining global features from several convolutional layers. As shown in Fig. 2(D), YOLOv3 outputs the target coordinates and probabilities of classes on three different scales. The three detection layers' predicted information is combined and processed, and NMS is performed to eliminate redundant detection boxes and perform a local maximum search. Then, the final detection result is output.

The unremitting efforts of scholars have produced a steady stream of deep-learning-based object detection models. Therefore, it is particularly important to identify a model suitable for ship detection. As ships are small and narrow, the advantages and disadvantages of the current mainstream object detection model based on deep learning can be analysed to determine the ideal ship detection model. The detection efficiency of the CenterNet and faster R-CNN approaches cannot meet ship detection requirements. Although SSD can meet the requirements for detection efficiency, it may miss small targets, such as ships, owing to the design of its structure. Thus, in terms of the trade-off between accuracy and efficiency, YOLOv3 is better for ship detection than the other detection models.

B. Improved YOLOv3 for ship detection

YOLOv3 has produced satisfactory detection results on standard dataset detection tasks, such as PASCAL VOC and COCO. However, YOLOv3 still requires improvements for detecting ships from remote sensing images. There are three reasons why this is the case. First, images of natural scenes are different from those of remote sensing images. Natural images tend to have higher resolution, cleaner backgrounds, and a larger proportion of detection targets. Therefore, training with YOLOv3's original (default) parameters will have effects on the detection time and performance. As most ships appear as small, narrow targets in remote sensing images, the accuracy of the default anchor boxes assigned to the model should be improved. Therefore, an improved k-means++ clustering algorithm is proposed to address problems regarding the centralised distribution of anchor boxes for detecting a single ship type. Second, unlike the case with natural images, the object sizes and directions in remote sensing images vary at different resolutions. Thus, finer anchors must be assigned to the grid cells for the detection layer. Third, each prediction box in the original YOLOv3 model only contains the bounding box coordinates (i.e., t_x,t_y,t_w,t_h), and cannot reflect the reliability of the bounding box. Therefore, we introduce a Gaussian model to compensate for this drawback. In this study, we improve the YOLOv3-based prior's anchor, anchor assignment, and Gaussian model strategy to solve the problem of multiple scenes, ship targets of varying size, and detection accuracy in remote sensing imagery ship detection. This makes it more suitable for ship detection tasks.

1) Improved k-means++ clustering for priori anchors

The purpose of the anchor box in YOLOv3 is to detect multiple objects concentrated in a grid cell. This adds another dimension to the output label. The numbers and sizes of the anchor boxes also affect the detection speed and accuracy. The anchor box can be defined as the most likely width and height of an object, and can be counted from an object's benchmark dataset through a clustering algorithm. Considering this, we redesigned the anchor boxes to reduce the matching errors occurring during training.

The k-means++ clustering method was applied to the ground-truth boxes of the detected object in the training dataset automatically determine the bounding box prior. The object's width and height were calculated based on the box size of the target from the label file. The number of clusters, k, as the number of anchor boxes, was adjustable. More anchors would seem to simplify the prediction task; however, they require higher computation costs. The more similar the sizes of the rectangular boxes between the clustering centroids and ground truth, the shorter the distance. The distance matrix was constructed based on the intersection over union (IoU) between the clustering centroids and ground-truth bounding boxes of the training image. The IoU was the crossover rate between the clustering of centroid box and ground truth bounding box. The IoU can be expressed as follows:

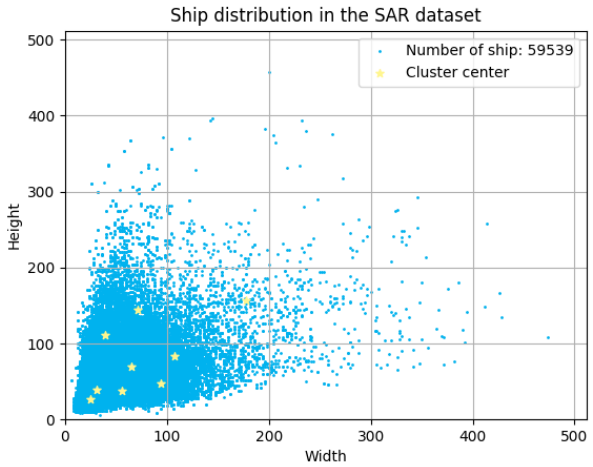
$$IoU = \frac{Area\big(Box(\text{centroids}) \cap Box(truth)\big)}{Area\big(Box(\text{centroids}) \cup Box(truth)\big)} \quad (1)$$

The distance metric should satisfy the following relationship.

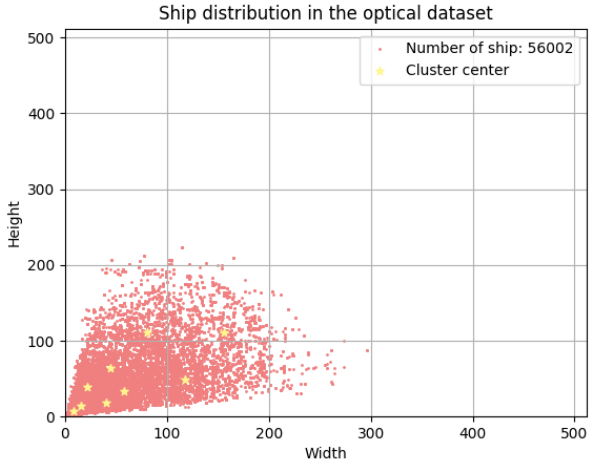
$$dis_{centroid}^{box} = 1 - IoU_{centroid}^{box} \quad (2)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3087555, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

To design anchors of similar size for the ships, a clustering method was used to count the size distribution of the bounding boxes in the SAR and optical ship datasets. SAR images with ship chips of 256 pixels and optical images with ship chips of 768 pixels used in this study contained ships of various sizes, as showed in Fig. 3. However, the anchor box distribution concentration was obtained using the k-means++ clustering algorithm, and many ships were larger than the obtained anchor boxes. This is because the ship datasets had a single category and similar size ratios. In addition, the k-means++ clustering algorithm calculates the cluster centres of each category, which narrows the real range between anchor boxes. This will result in small differences between the detection scales. Thus, the directly obtained anchor box from a single-detection target type would not emphasise the advantages of the model's multi-scale detection.



(a)



(b)

Fig. 3. Ship label bounding box size distribution relative to the image: (a) is in the synthetic aperture radar (SAR) ship dataset and (b) is in the optical ship dataset. The yellow star is the anchor box information obtained before the improvement of the k-means++ algorithm, when the number of clusters is set to nine.

Therefore, we used a linear scaling method to stretch the anchor box to both sides. The calculation formulae are as follows:

$$w_1^{'} = w_1 \times \alpha \tag{3}$$

$$h_1^{'} = h_1 \times \alpha \tag{4}$$

$$w_n^{'} = w_n \times \beta \tag{5}$$

$$h_n^{'} = h_n \times \beta \tag{6}$$

$$w_i^{'} = \left(\frac{w_i - w_1}{w_n - w_1}\right) \times (w_n^{'} - w_1^{'}) + w_1^{'} \tag{7}$$

$$h_i^{'} = w_i^{'} \times \frac{h_i}{w_i} \tag{8}$$

In the above, n is the number of anchors, w and h are the width and height of the anchor obtained from the k-means++ algorithm, respectively. α and β denote the scaling ratios of the minimum and maximum anchors, respectively. In this study, the scaling ratios of the maximum and minimum anchors were determined based on comparisons between the anchors generated by the k-means++ clustering algorithm and the actual ship's width and height distribution. As showed in Table II, the box would exceed the size range if the stretching ratio was too large, and the effect would not be evident if the stretching ratio was too small. Therefore, after comparison in this experiment, it was roughly determined that α = 0.6 and β = 1.4. w_1, h_1, w_n, h_n, $w_1^{'}$, $h_1^{'}$, $w_n^{'}$, and $h_n^{'}$ are the widths and heights of the first and last anchor box from the k-means++ and improved k-means++ algorithms, respectively; $w_i^{'}$ and $h_i^{'}$ are the width and height of the improved anchor box, and can be calculated using Equations (3), (4), (5), (6), (7), and (8).

TABLE II
ANCHOR BOX RESULTS OBTAINED BY THE IMPROVED K-MEANS ++ APPROACH FOR DIFFERENT SCALING RATIOS IN THE OPTICAL DATASET

α, β	Anchor boxes
0.9, 1.1	(7,6)(14,13)(42,19)(22,39)(62,35)(46,68)(128,53)(87,120)(171,122)
0.8, 1.2	(6,5)(14,13)(45,20)(23,40)(67,38)(50,72)(140,58)(95,130)(187,133)
0.7, 1.3	(5,4)(14,13)(47,21)(23,41)(71,40)(52,76)(151,62)(102,140)(202,144)
0.6, 1.4	(4,4)(14,13)(50,22)(24,42)(76,43)(56,81)(163,67)(109,150)(218,155)
0.5, 1.5	(4,3)(14,13)(53,24)(25,45)(81,46)(59,87)(174,72)(117,160)(234,166)
0.4, 1.6	(3,2)(14,13)(56,25)(26,46)(86,48)(62,91)(185,77)(124,170)(249,177)
0.3, 1.7	(2,2)(14,13)(58,26)(26,47)(90,51)(65,95)(197,82)(131,180)(265,188)
0.2, 1.8	(1,1)(14,13)(61,27)(27,48)(95,54)(68,100)(208,86)(138,189)(280,199)
0.1, 1.9	(0,0)(14,13)(64,28)(28,49)(100,56)(72,104)(220,91)(146,200)(296,210)

Figure 4 shows a schematic diagram of the improved k-means++ anchor box calculation. Fig. 5 presents an example process of selecting nine clusters from the SAR, optical, and mixed datasets.

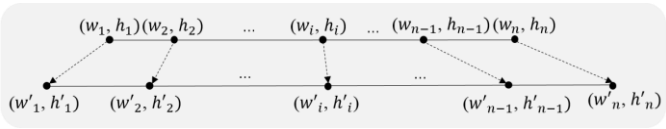


Fig. 4. Schematic diagram of improved k-means++ anchor box calculation. w_i, h_i represent the width and height of the i-th anchor box before improvement; $w_i^{'}$ and $h_i^{'}$ represent the width and height of the i-th anchor box after improvement.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3087555, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

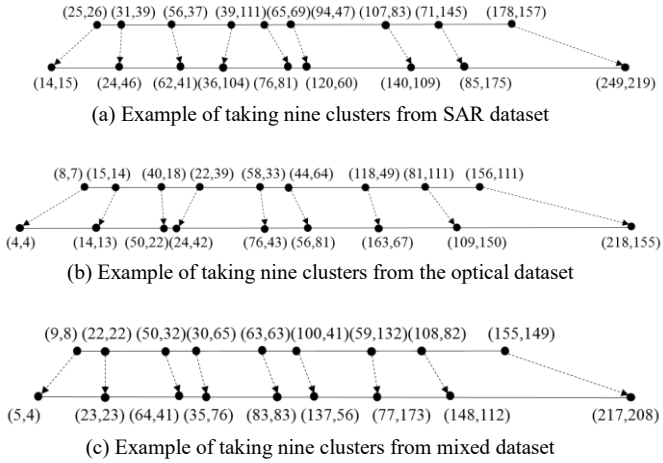


Fig. 5. Example of using the improved k-mean++ method taking nine clusters from the SAR, optical, and mixed datasets. The top and bottom are the clustering results of k-means++ before and after the improvement.

2) Assigning finer anchors to the detector layer

Unlike natural images, there can be a wide disparity in the sizes and directions of an object in remote sensing images of different resolutions. Owing to the various methods of acquiring images, objects can appear at any scale in the image, and images at the same scale may have different sizes [44]. Thus, finer anchors must be assigned to the detection layer.

On the one hand, design the different detection layer structures. As shown in figure 6, Structure 1 is the structure used by YOLOv3. Structure 2 and Structure 3 add the first and second feature scales to the third feature scale without changing the original number of 9 anchor boxes. The difference between them is that the number of anchor boxes allocated to the second detection layer. On the other hand, by increasing the total number of anchor boxes to evenly distribute to each detection layer. Structure (d) is to add a new detection scale to YOLOv3 and do not change the preset number of anchors 3 for each prediction layer. The structure of (e) is to add an anchor box to each detection scale.

In this study, 12 clusters were selected, and an average of four anchor boxes corresponded to the detection layer in the proposed method for the SAR, optical, and mixed datasets. We sorted these anchors, and distributed clusters evenly among the scales. The original nine clusters obtained on the PASCAL VOC dataset for YOLOv3 were as follows: (10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 98), (156, 98), and (373, 326). As showed in Table III, the clusters were more concentrated than the original clusters, and the widths and heights were smaller than in the original clusters.

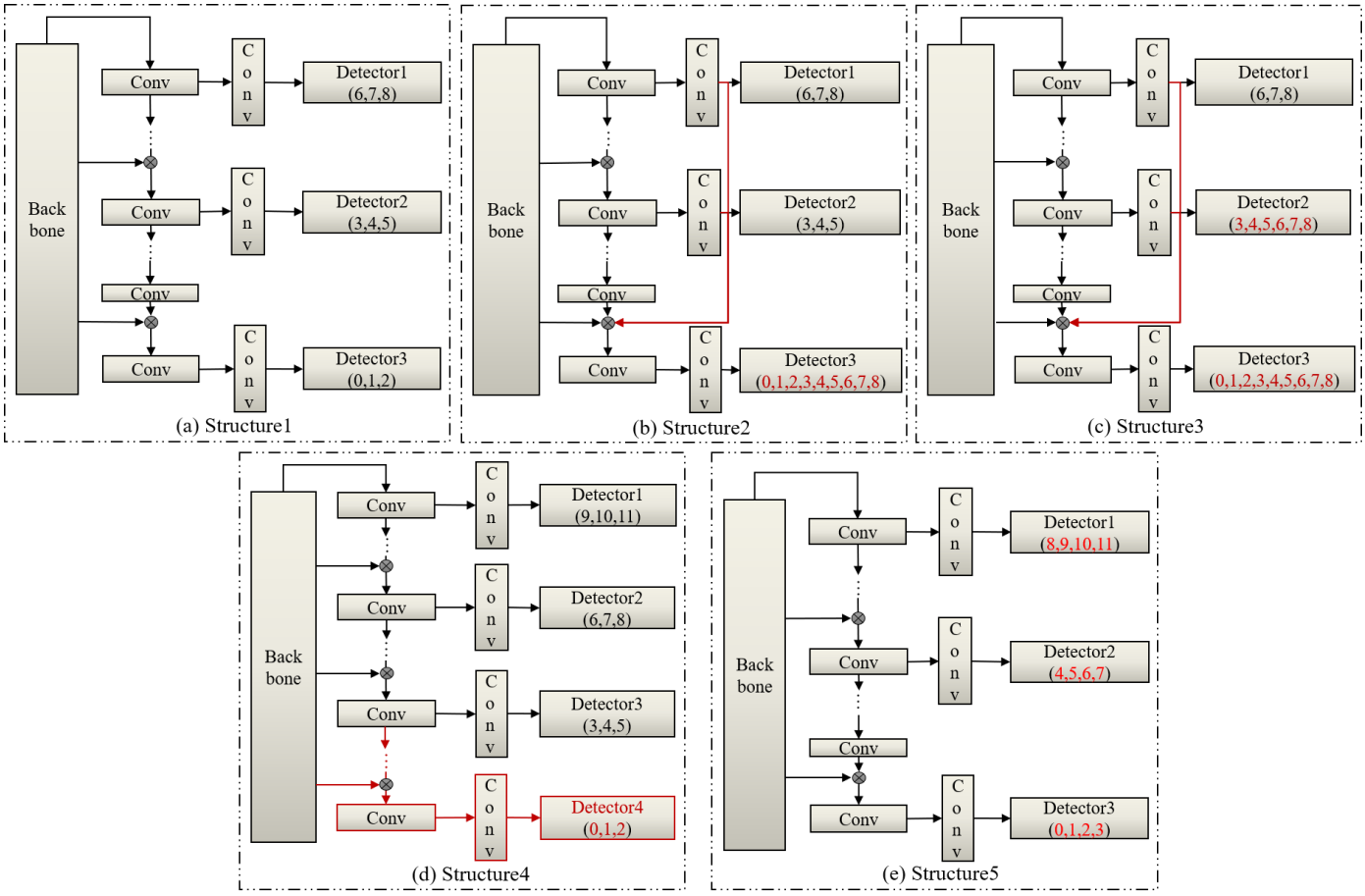


Fig. 6. Different detection layer structures. (a) is the structure employed by YOLOv3. (b) indicates add the first and second detection features to the third detection feature, and uses all anchor boxes to the third detection layer without changing the total number of preset anchor boxes. The structure of (c) is the same as (b), except that the number of anchor boxes in the second detection layer is increased. Structure (d) is to add a new detection scale to YOLOv3 and keep the number of anchors of each detection scale is 3. Structure (e) is to allocate one more anchor box to each detection scale more finely. Modified part is shown by the red line and font.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3087555, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

TABLE III
ANCHOR BOX RESULTS OBTAINED BY THE IMPROVED K-MEANS++ APPROACH FROM THE TRAINING DATASETS

Data Source	SAR training set		Optical training set		Mixed training set	
Cluster	9	12	9	12	9	12
Third detector layer	(14, 15) (24, 26) (62, 41)	(12, 13) (46, 36) (22, 46) (99, 46)	(4, 4) (14, 13) (50, 22)	(4,3) (12,11) (35, 18) (22, 35)	(5, 4) (23, 23) (64, 41)	(5, 4) (20, 18) (29, 51) (65, 33)
Second detector layer	(36, 104) (76, 81) (120, 60)	(32, 89) (63, 67) (127, 76) (93, 106)	(24, 42) (76, 43) (56, 81)	(70, 29) (32,78) (60, 50) (155, 48)	(35, 76) (83, 83) (137, 56)	(68, 64) (41, 100) (132, 55) (101, 100)
First detector layer	(85, 175) (140, 109) (249, 219)	(57,164) (176, 118) (115, 190) (282, 242)	(163, 67) (109, 150) (218, 155)	(86, 83) (101, 158) (188, 96) (207, 184)	(77, 173) (148, 112) (217, 208)	(64, 178) (179, 102) (124, 179) (249, 203)

3) Gaussian model

In YOLOv3, an input image is divided into an S × S grid cell, and the grid cells are responsible for detecting objects whose centre falls into them. Every grid cell predicts the object category and position of the bounding boxes. As shown in Fig. 7, the YOLOv3 prediction map has three prediction boxes per grid. Each grid outputs $3 \times ((t_x, t_y, t_w, t_h) + obj_{score} + class_{score})$ information, where $class_{score}$ refers to the probability of the object category, and obj_{score} refers to whether an object is present in the bounding box. However, the bounding box only provides coordinates; the results do not adequately reflect the reliability of the box (i.e., $t_x, t_y, t_w,$ and t_h indicate the centre coordinates, width, and height of the predicted bounding box, respectively).

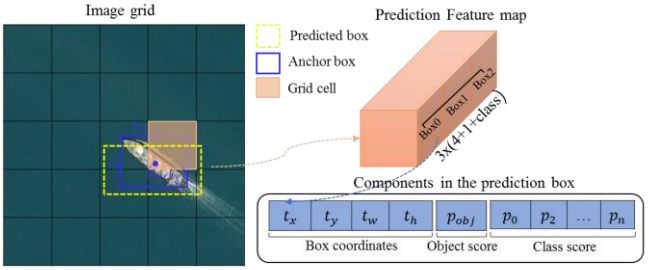


Fig. 7. Attributes of the prediction map in the YOLOv3 architecture.

Based on this limitation, Choi et al. [45] proposed using Gaussian distribution characteristics to evaluate the reliability of each bounding box's coordinate information, to thereby improve the accuracy of the network. The output of a bounding box after introducing the Gaussian model contains $\hat{\mu}t_x$, $\hat{\Sigma}t_x$, $\hat{\mu}t_y$, $\hat{\Sigma}t_y$, $\hat{\mu}t_w$, $\hat{\Sigma}t_w$, $\hat{\mu}t_h$, and $\hat{\Sigma}t_h$, where $\mu(x)$ and $\sum(x)$ are the mean and variance functions, respectively. The data must be transformed as follows:

$$\sigma(x) = \frac{1}{(1 + \exp(-x))} \tag{9}$$

$$\begin{cases} \Sigma t_x = \sigma\left(\widehat{\sum t_x}\right) \\ \Sigma t_y = \sigma\left(\widehat{\sum t_y}\right) \\ \Sigma t_w = \sigma\left(\widehat{\sum t_w}\right) \\ \Sigma t_h = \sigma\left(\widehat{\sum t_h}\right) \end{cases} \tag{10}$$

$$\begin{cases} \mu t_x = \sigma(\hat{\mu}t_x) \\ \mu t_y = \sigma(\hat{\mu}t_y) \\ \mu t_w = \sigma(\hat{\mu}t_w) \\ \mu t_h = \sigma(\hat{\mu}t_h) \end{cases} \tag{11}$$

The maximum likelihood estimations μt_x, μt_y, μt_w, and μt_h obtained from Equations (9), (10), and (11) can be used to calculate the coordinates of the bounding box regressions. After assigning four anchors to the Gaussian-YOLO detection layer, the output information for each grid cell is as follows:

$$4 \times ((\mu t_x, \Sigma t_x, \mu t_y, \Sigma t_y, \mu t_w, \Sigma t_w, \mu t_h, \Sigma t_h) + obj_{score} + class_{score})$$

Now, each bounding box coordinate satisfies a Gaussian distribution with a mean μ and variance σ, and a loss function. Taking the coordinate x of the centre point of the bounding box as an example, the modified bounding box x' coordinate prediction error calculation formula also must be modified accordingly.

$$L_x = -\sum_{n}^{W}\sum_{m}^{H}\sum_{l}^{K} \gamma_{nml} \times \log\left(N\left(x_{nml}^{G} \middle| \mu_{tx}(x_{nml}), \sum x_{tx}(x_{nml})\right) + \varepsilon\right) \tag{12}$$

Here, W and H correspond to each grid's width and height; K is the number of anchors; $\mu_{tx}(x_{nml})$ is the mean value of t_x of the k-th anchor in the (n, m) grid of the output layer; $\sum x_{tx}(x_{nml})$ is the uncertainty of the corresponding t_x value; x_{nml}^{G} is the true value of t_x; γ_{nml} is a weight parameter; and ε is a constant, with a value of 10^{-9}. Therefore, the detection formula for the bounding boxes is as follows:

$$Cr. = \sigma(Object) \times \sigma(Class_i) \times (1 - Uncertainty_{aver}) \quad (13)$$

In the above, $\sigma(Object)$ reflects the probability of a box containing an object, $\sigma(Class_i)$ indicates the probability of the i-th class, and $Uncertainty_{aver}$ represents the average reliability of the predicted bounding box coordinates.

In addition to the Guassian model, GIoU (Generalized Intersection over Union) loss [46], DIoU (Distance Intersection over Union) loss [47], and CIoU (Complete Intersection over Union) loss [47] can also improve the accuracy of bounding box regression. Among them, GIoU Loss introduces the smallest bounding rectangle between the predicted and the ground truth bounding boxes to deal with the situation that the two bounding boxes do not overlap in the IoU loss. DIoU loss takes into consideration the distance between the center points of the two bounding boxes to speed up the convergence speed of GIoU loss. CIoU loss increases the aspect ratio scale of the bounding box to improve the regression accuracy of the DIoU loss bounding box.

C. Network architecture of the improved-YOLOv3 for ship detection

In this study, an improved-YOLOv3-based convolutional neural network was proposed for improving the performance of ship detection. Fig. 8 presents the improved-YOLOv3 network architecture.

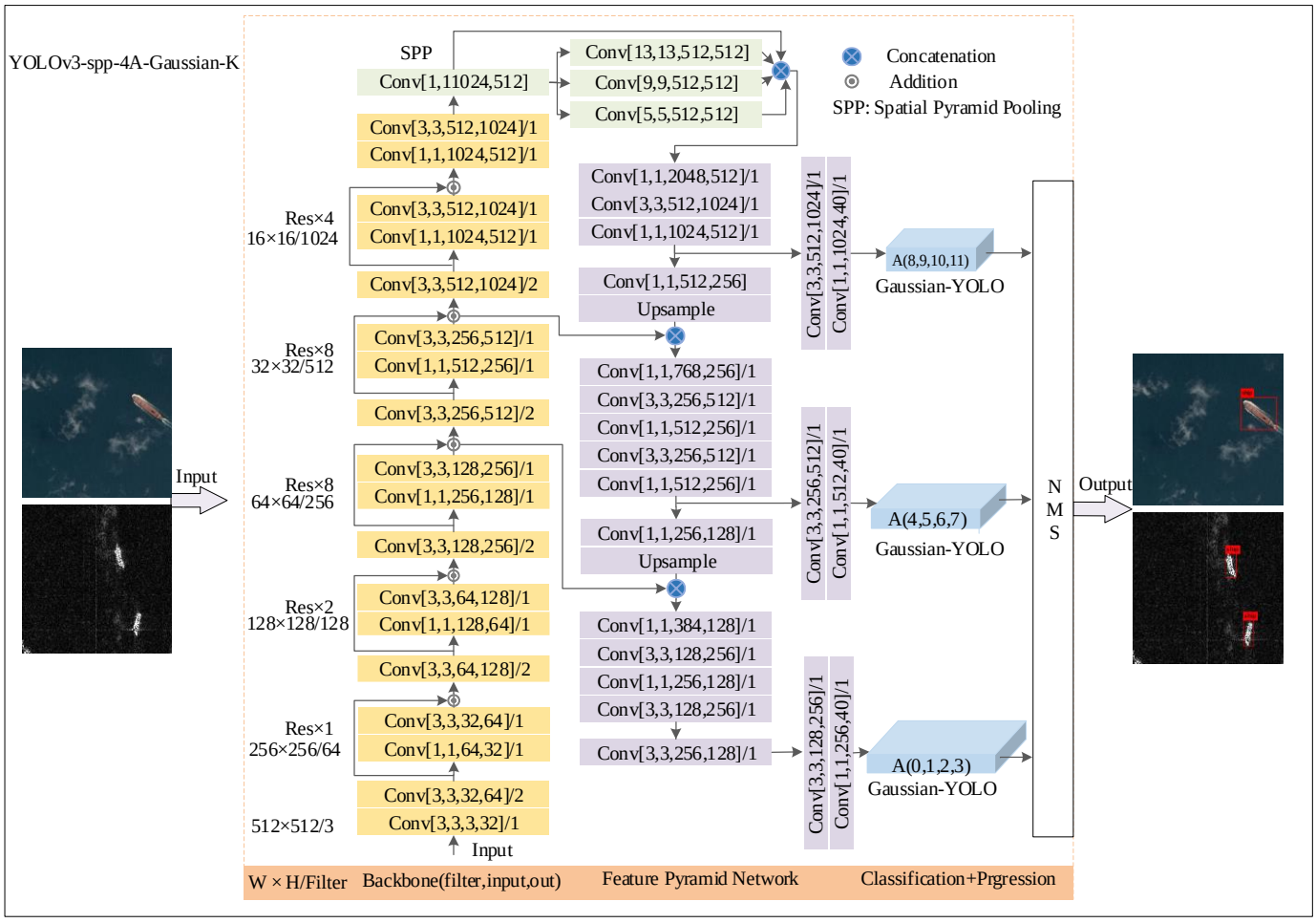


Fig. 8. Architecture of the improved-YOLOv3.The yellow module is the backbone network Darknet-53, the purple module is the feature pyramid network (FPN) multi-scale feature detection, and the blue part is the improved module

Darknet-53 served as the backbone network for feature extraction from the bottom up. It contained 23 residual units, consisting of 3 × 3 and 1 × 1 convolutional layers. Objects were detected with the FPN architecture at three different scales to improve the multi-scale prediction. In addition, the four anchors were used to determine the detection bounding box in the Gaussian-YOLO detection layer. The network structure comprised inputting the ship image, feature extraction, Gaussian-YOLO layer detection, and classification and regression based on NMS screening of the candidate boxes to

output the ship detection results. These processes are described more fully below.

(1) Image Input: 256×256 pixels SAR ship slices and the 768×768 pixels optical ship slices were resized to 512×512 pixels as the network input, and the processed images were evenly divided into S × S cell grids. Every grid cell used three anchor boxes to predict the position of the bounding box, confidence score, and class probabilities.

(2) Feature Extraction: The processed unified images were sent to DarkNet-53. Using Multi-Scale training way, feature map will be randomly changed from 320×320 to 608×608 in a step of 32 during the training process. The purpose is to make the model adaptable various input sizes to improve the robustness of detection.

(3) Gaussian-YOLO Layer Detection: The 83rd, 61st, and 36th layers outputted the feature vector through horizontal and top-down connections, and then the convolution and up-sampling operations were used on the feature vector to obtain the next feature vector. Finally, the feature vector outputs (in dimensions of 16 × 16, 32 × 32, and 64 × 64) were sent to the three Gaussian-YOLO layers for the decode operations, calculation of the loss rates, and prediction of the classification and boundary regression.

(4) NMS: NMS filters pout the redundant bounding boxes. The steps are as following. Step 1: Arrange the bounding boxes in set B according to their confidence. Step 2: Select the bounding box with the highest confidence from set B as reserved box M. Step 3: Calculate the IoU values of the remaining bounding boxes and M in set B, and delete the bounding boxes whose IoU values are higher than the NMS threshold (0.45). Step 4: Repeat steps 2 and 3 until set B is empty.

D. Evaluation metrics

In the comparison experiments, the detection speed, precision-recall curve, precision, recall, and average precision (AP) were used to evaluate the performance of the detection model. The detection speed was expressed as detection time/images (total detection time/number of detected images), i.e., the time required to detect each image. The metric measured the fraction of detections that were true positives (TP), and the metric measured the fraction of positives that were correctly identified [49]. True positive (TP) denoted the number of ships correctly detected as ships, false positive (FP) denoted the number of backgrounds incorrectly detected as ships, and false negative (FN) denoted the number of ships incorrectly detected as backgrounds.

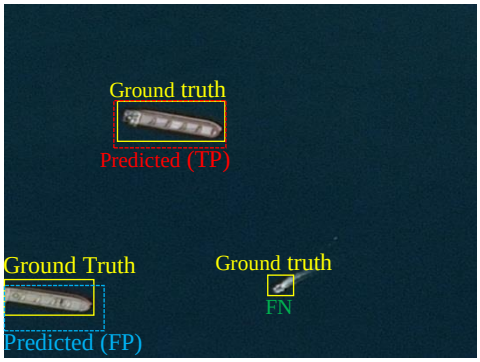


Fig. 9. Schematic diagram of calculating evaluation indicators of intersection-over-union (IoU), true positive (TP), false positive (FP), and false negative (FN).

Fig. 9 shows an example calculation of the IoU, FP, and FN. In general, AP calculates the area enclosed by the precision-recall curve through numerical integration. In object detection, precision and recall are generally a pair of contradictory indicators. Therefore, AP is a common evaluation indicator for object detection.

E. Ship detection workflow

Fig. 10 presents the overall framework of the improved-YOLOv3 ship detection. There are three main components of the workflow: ship image pre-processing, ship detection, and detection model performance evaluation. The first component comprises the pre-processing of the SAR and optical ship datasets. The images are converted into a unified PASCAL VOC data format that can be fed into the detection model, and are randomly divided into training, test, and validation sets in equal proportions. The second component comprises improved-YOLOv3 training and prediction process for ship datasets. After pre-processing, the test and training images are input into the network. Then, the features are extracted from the training images using the Darknet-53 neural network. The parameters are continuously adjusted through forward and backward propagation until the model converges, and the model weight file is saved. Finally, the generated weight file is called for detection based on the test images, and the candidate boxes are filtered by NMS to output the ship detection results. The third component of the process comprises evaluating the detection model's ship detection performance according to evaluation metrics.

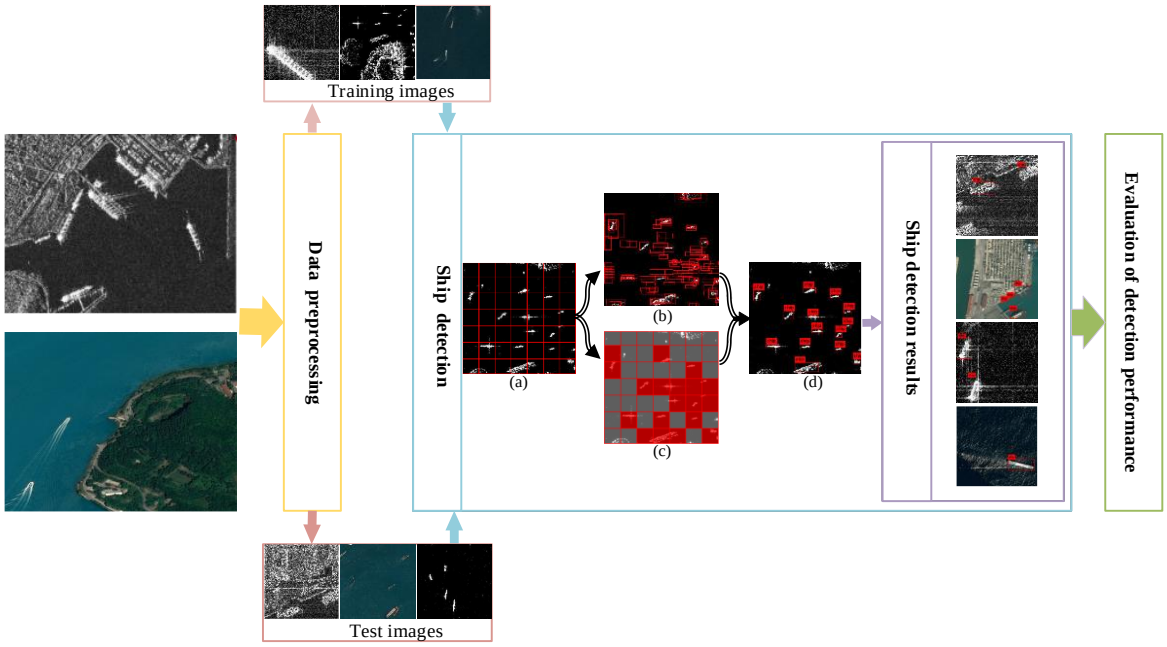


Fig. 10. General framework for ship detection using deep learning techniques: (a) division of the input image into an S × S grid(S can be one of 64, 32 and 16); (b) Each grid cell in which the centre of the object falls predicts bounding boxes and confidence intervals for those boxes; (c) Each grid cell into which the centre of the object falls predicts the class probability map; (d) final detections are obtained by non-maximum suppression (NMS) filtering.

IV. Experiment and Analysis

In this section, we discuss the ship detection performance based on deep learning algorithms for SAR and optical images. To highlight the effectiveness of the proposed method, the improved-YOLOv3, YOLOv3, SSD, CenterNet, and Faster R-CNN are compared on the same dataset and the same experimental environment.

A. Experiment environment

The configurations of the software and hardware used in the experiments are listed in Table IV.

TABLE IV
ENVIRONMENTAL PARAMETERS FOR THREE DEEP LEARNING DETECTION MODELS

	YOLOv3	SSD	CenterNet	Faster R-CNN
Processor	Intel Core™ i7-9800X CPU @ 3.80GHz × 16			
RAM	31.1GiB			
GPU	GeForce RTX 2080Ti 11GB			
Operating system	Ubuntu 16.04 LTS			
Deep learning framework	DarkNet	Caffe	Keras	Tensorflow 1.10
Programming language	C	Python 3.5, Shell	Python 3.6	Python 2.7

For improved-YOLOv3, the batch size is set to 64, and the subdivisions are set to 16 to maximise the GPU's memory utilisation and reach a quick convergence. For the network training tasks, the learning rate is set to 0.000261, the momentum is set to 0.9, and the decay is set to 0.0005. The ignore threshold is set to 0.5, and the total training is set at approximately 50,200 iterations. Faster-RCNN and SSD models iterate are each iterated 500,200 times, the epoch of CenterNet is set to 146, and the other parameters remain the same as before. Eventually, these detection models converge, and save the weight parameters.

B. Experimental results and evaluation

The model and data selection influences the ship detection performance. In this section, a quantitative representation of each comparison is provided for the Faster R-CNN, SSD, CenterNet, YOLOv3, and improved-YOLOv3 approaches, on both the SAR and optical ship detection datasets. The detection accuracy is measured based on the AP, and the detection efficiency is measured based on the detection time/image (total detection time/number of detected images). Additionally, the YOLOv3 and improved-YOLOv3 approaches are trained and tested on a mixed SAR and optical dataset, to verify the robustness of the proposed method. In addition to the quantitative detection accuracy results, we also report qualitative forecast results for the Faster R-CNN, SSD, CenterNet, YOLOv3, and improved-YOLOv3 ship detection methods on different datasets.

1) Ship detection comparison results for the improved method

YOLOv3-tiny is a lightweight YOLOv3 with only a 31-layer network. Although the detection accuracy is lower, the training time is relatively short. On the contrary, YOLOv3-spp has 114 -layers, and the detection accuracy is relatively high, but it needs a longer training time. Therefore, consider adding different strategies to YOLOv3-tiny under the same conditions to quickly observe the results of the improved method.

Table V lists the optical ship detection results of five structures used in YOLOv3-tiny. Structure 2 and Structure 3 keep the total number of anchor boxes at 9, and add the first and second feature scales to the third feature scale. Structure 4 and

Structure 5 increase the total number of anchor boxes to 12, and then evenly distribute the anchor boxes to each detection scale in two different ways. It can be concluded that among the five structures, Structure 5 that assigns finer anchor boxes to each detection scale performs best.

Table VI lists the detection number of small, medium, and large scales optical ships before and after increasing the number of anchor boxes of the corresponding scale prediction layer based on Structure 5. The purpose is to observe which scale ship objects are hard to be detected by YOLOv3. Small, medium, and large are divided according to the sum of the width and height of the extreme value of the anchor box distribution of different detection scales. For example, the anchor box distribution before improvement is (8, 7), (15, 14), (40, 18), (22, 39), (58, 33), (64, 62), (118, 49), (81,111), (156,111). Thus, width +height $\leqslant$ 58 is small-scale ship, 58< width +height $\leqslant$ 126 is medium-scale ship, width +height >126 is large-scale ship.

In Table VI, O represents the original YOLO. $4A_s$, $4A_m$, and $4A_l$ are based on the O model and only increase the number of anchors with the small, medium, and large scales detection layer, respectively. 4A is a combination of $4A_s$, $4A_m$, and $4A_l$ models, increase the number of anchors of each detection layer. It can be concluded that the original YOLOv3 has more missed detection for medium-scale ships and false alarms for small-scale ships. It can be seen from the detection results of $4A_s$, $4A_m$, and $4A_l$ that increasing the number of anchor boxes of a certain scale will improve the detection results of that scale. From the detection result of 4A, we can see that the deviation obtained by increasing the anchor box for each detection scale is the smallest. Table VII lists the SAR and optical ship detection results using Gaussian model, GIoU loss, DIoU loss, and CIoU loss methods in YOLOv3-tiny. It can be noticed that adding GIoU loss, DIoU loss, CIoU loss, and Gaussian parameters to YOLOv3 can improve the regression accuracy of SAR and optical ship bounding boxes, and the Gaussian model has the most obvious improvement effect compared to other methods.

TABLE V
DETECTION ACCURACY ON THE OPTICAL SHIP DATASET FOR FIVE STRUCTURES

Detection layer structure	AP(%)	Recall(%)	Precision(%)
Structure1	85.91	90.74	56.32
Structure2	85.94	90.83	56.38
Structure3	84.39	89.95	55.67
Structure4	85.92	90.63	**63.11**
Structure5	**85.97**	**90.99**	57.14

TABLE VI
COUNT THE NUMBER OF LARGE, MEDIUM, AND SMALL SCALES SHIPS ON OPTICAL IMAGES SHIP DETECTION

Model		AP(%)	Number of detected ships			Deviation
			Large	Medium	Small	
YOLOv3-tiny	O	85.91	4122(-48)	3753(-604)	8583(+385)	1037
	$4A_s$	85.94	4181(+11)	3782(-575)	**8299(+101)**	687
	$4A_m$	85.94	4100(-70)	**4065(-292)**	8566(+368)	730
	$4A_l$	85.96	**4201(+31)**	3755(-602)	8528(+330)	963
	4A	85.97	3942(-228)	4462(-105)	8062(-136)	**469**
YOLOv3-spp	O	92.08	4148(-22)	3808(-549)	8423(+225)	796
	$4A_s$	92.08	4177(+7)	4040(-317)	**8287(+98)**	422
	$4A_m$	92.07	4122(-48)	**4116(-241)**	8419(+221)	510
	$4A_l$	92.06	**4155(-15)**	3895(-462)	8362(+164)	641
	4A	92.07	4043(-127)	4410(+53)	7968(-230)	**410**
Number of real ships			4170	4357	8198	0

TABLE VII
DETECTION ACCURACY ON THE OPTICAL AND SAR SHIP DATASETS FOR FIVE METHODS

Method	AP(%)	
	SAR ship dataset	Optical ship dataset
YOLOv3-tiny	91.46	85.91
GIoU	91.77	85.97
DIoU	92.09	85.93
CIoU	92.15	85.96
Gaussian(IoU)	**93.45**	**85.98**

Based on the above experiments, the proposed strategies are successively added to YOLOv3-spp and YOLOv3-tiny on the optical ship detection dataset to verify the effectiveness of the proposed method. To evaluate the proposed method, six experiments are conducted in YOLOv3-spp and YOLOv3-tiny. The model relationship are as follows: (1) anchor boxes obtained by k-means ++; (2) -K: anchor boxes obtained by improved K-means++; (3)-4A: four anchor boxes assigned to a detection scale, based on Experiment 1; (4)-Gaussian: a Gaussian model is introduced, based on Experiment 1; (5) -4A-K: four anchor boxes are assigned to a detection scale, based on Experiment 2; and (6) -4A-Gaussian-K: a Gaussian model is

introduced based on Experiment 5. In Table VIII, the model 'YOLOv3-tiny-o' refers to the original YOLOv3-tiny; its anchor boxes are produced by clustering with the PASCAL VOC dataset. Training settings applied to YOLOv3-spp and YOLOv3-tiny are the same.

As showed in Table VIII, the improved-YOLOv3 has a 2% improvement in the accuracy of the optical dataset. It is a well-known fact that most of the improvements in detection accuracy come at the cost of deepening the network and requiring additional calculations. However, the addition of the proposed strategy in this study has little effect on the detection efficiency.

TABLE VIII
DETECTION ACCURACY ON THE OPTICAL SHIP DATASET FOR EACH METHOD

Model relationship	Model	(AP) (%)	Detection times/images(ms)
0	YOLOv3-tiny-o	84.98	8.7
1	YOLOv3-tiny	85.91	8.7
1-1	YOLOv3-tiny-K	86.22	8.7
1-2	YOLOv3-tiny-4A	85.97	9.2
1-3	YOLOv3-tiny-Gaussian	85.98	9.8
1-1-2	YOLOv3-tiny-4A-K	86.79	9.2
1-1-2-3	YOLOv3-tiny-4A-Gaussian-K	87.91	10.1
2	YOLOv3-spp	92.08	18.7
2-1	YOLOv3-spp-K	92.35	18.7
2-2	YOLOv3-spp-4A	92.07	18.7
2-3	YOLOv3-spp-Gaussian	93.19	20.0
2-1-2	YOLOv3-spp-4A-K	92.37	18.6
2-1-2-3	YOLOv3-spp-4A-Gaussian-K	93.56	21.3

Figs. 11 and 12 show each curve comparisons between each 'added strategy' and the method before adding (as represented by different colours). The solid and dashed lines represent YOLOv3-spp and YOLOv3-tiny, respectively. Fig. 11 shows the AP values of each method under different iteration times, and Fig. 12 shows the PR curves of each method in the optical ship detection dataset. As shown, the improved methods (indicated in red) are at the top. This indicates that the improved methods are effective.

2) SAR ship detection comparison results

The detection accuracies of the five SAR ship detection models are listed in Table IX. Each of these models achieves a higher detection probability and lower false alarm probability on the SAR dataset. Improved-YOLOv3 has the highest AP (95.52%), and a higher detection precision than YOLOv3.

Fig. 13 presents the SAR ship detection results for five deep learning detection models under different scenarios, in which the improved-YOLOv3 refers to YOLOv3-spp-4A-Gaussian-K. Most of the orange triangles are distributed in the SSD. This indicates that small ships are hard for SSD to detect. As showed by the position of the blue circle, CenterNet and Faster R-CNN are poor at combating interference. improved-YOLOv3 misdetection (the blue circle in the second row of Fig. 13(a)) and missed ship detection (the orange triangles in the second row of Fig. 13(b)) are improved.

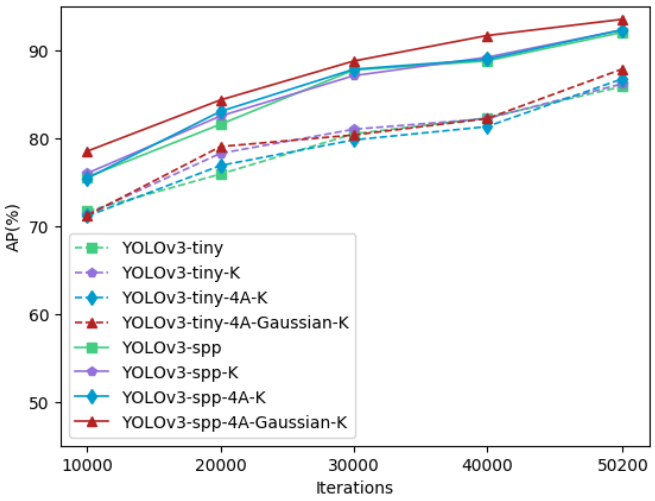


Fig. 11. Average precision (AP) values for each method based on improved-YOLOv3 at different iteration numbers on the optical ship dataset.

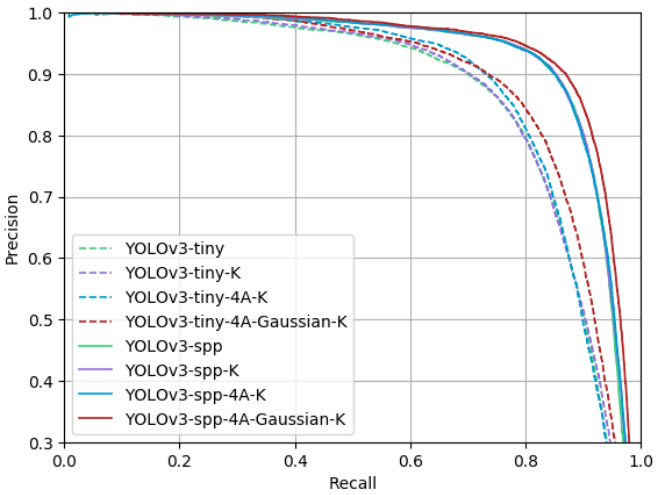


Fig. 12. Precision-recall (PR) curve of each method based on improved-YOLOv3 for the optical ship dataset.

TABLE IX
SAR SHIP DETECTION FOR THE FOUR DEEP LEARNING DETECTION MODELS

Model	AP (%)	Recall (%)	Precision (%)	Detection times/images(ms)
Faster-RCNN	90.37	92.84	78.58	51.7
SSD	89.47	93.48	76.43	10.4
CenterNet	92.89	95.98	78.09	57.8
YOLOv3-tiny	91.46	92.55	77.05	8.7
YOLOv3-tiny-4A-Gaussian-K	94.41	95.03	82.62	10.1
YOLOv3-spp	92.69	95.16	78.05	18.7
YOLOv3-spp-4A-Gaussian-K	95.52	95.88	83.70	21.3

3) Optical ship detection comparison results

Table X shows five models' comparison results on the optical ship dataset. The overall detection accuracy is lower than that of the SAR ship-detection dataset. As optical images have higher spatial resolution to facilitate the detection of additional details in the image (e.g., smaller ships and various features on large ships), a more complex analysis is required. The YOLOv3 and improved-YOLOv3 approaches still perform better.

However, the SSD, CenterNet, and Faster R-CNN approaches have poor detection results for the optical dataset. They may perform better when the parameters are adjusted, but we kept the original parameters of their model for the comparison experiments. The improved-YOLOv3 has the highest AP (93.59%), approximately 2% higher than that of YOLOv3.

TABLE X
OPTICAL SHIP DETECTION FOR FOUR DEEP LEARNING DETECTION MODELS

Model	AP (%)	Recall (%)	Precision (%)	Detection times/images(ms)
Faster-RCNN	77.43	86.15	68.80	51.1
SSD	69.09	87.07	55.21	10.2
CenterNet	70.98	87.41	62.17	57.5
YOLOv3-tiny	85.91	90.73	56.31	8.6
YOLOv3-tiny-4A-Gaussian-K	87.91	89.55	57.49	10.0
YOLOv3-spp	92.08	92.64	71.43	18.6
YOLOv3-spp-4A-Gaussian-K	93.56	94.15	67.95	21.2

Fig. 14 presents the verification of the optical ship detection results for the five deep learning detection algorithms under difficult conditions ('improved-YOLOv3' refers to YOLOv3-spp-4A-Gaussian-K). The distribution of the orange and blue triangles shows the same trend as that of the SAR image detection. The SSD misses many dense and small ships, as showed by the distribution of orange triangles. Ship-like objects (such as bridges and docks) both in ports and onshore are easily misidentified as ships, as showed in Fig. 14(c). Notably, there are fewer cases of false identification as compared with SAR image ship detection, owing to the superior optical image interpretation. In addition, as showing in the second row in Fig. 14(c), the CenterNet, SSD, and Faster R-CNN approaches cannot differentiate between crowded ships well. In contrast, our method shows good detection performance under the aforementioned circumstances.

4) Ship detection comparison results on the SAR-optical mixed dataset

Ideally, a good ship detection algorithm should be effective in representative sea conditions for different ship types, geographic areas, and sensor data. To compare the data suitability and verify the robustness of the proposed method, we compared the results from SAR and optical hybrid ship detection with YOLOv3 and the improved-YOLOv3 on a mixed dataset. The mixed dataset contained multiple resolutions, scenes, scales, sizes, data sources, and modes (three-channel RGB optical images and one-channel SAR images).

Table XI shows the ship detection accuracy of the YOLOv3 and improved-YOLOv3 approaches on the optical-SAR remote sensing image mixed dataset. The detection precision is lower than the training and testing performance based on SAR or optical data alone, as the mixed dataset is more heterogeneous. However, improved AP of YOLOv3 can still reach 90.91%, and it is still increased by 3-4% on the mixed dataset as it is on the respective dataset alone.

We separately recorded the results of ship detection on the SAR and optical of the mixed dataset. Judging from the experimental results, the optical image has greater interference in the mixed data training model. There are two possible reasons. The first one may be because the optical image has a higher resolution, requiring the model to identify more detailed details (such as small boats and large ships with different characteristics). The second point may be caused by an imbalance in the number, because there are 29,070 optical images and 43,819 SAR images in the mixed dataset. The data enhancement method can be used to solve the problem of the imbalance in the number of optical images.

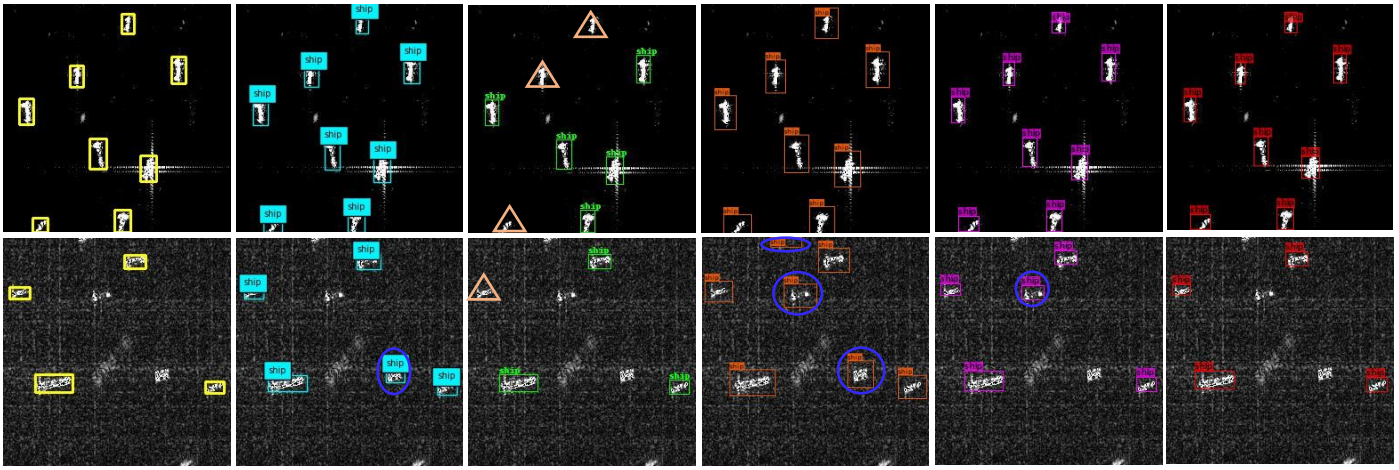
In order to verify the second conjecture, the optical ship training samples in the mixed dataset were enhanced by random combinations of flipping, translation, rotation, mirroring, brightness, and cropping. And keep testing samples unchanged to compare experiments. Finally, the number of SAR and optical ship slices in the mixed dataset is 50,807 and 43,819 respectively. The experimental results show that addressing the issue of data imbalance through data enhancement of optical data can improve the accuracy of ship detection with mixed dataset. But the optical image still has greater interference than SAR image in the mixed dataset.

TABLE XI
SHIP DETECTION RESULTS FOR THE OPTICAL-SAR MIXED DATASET

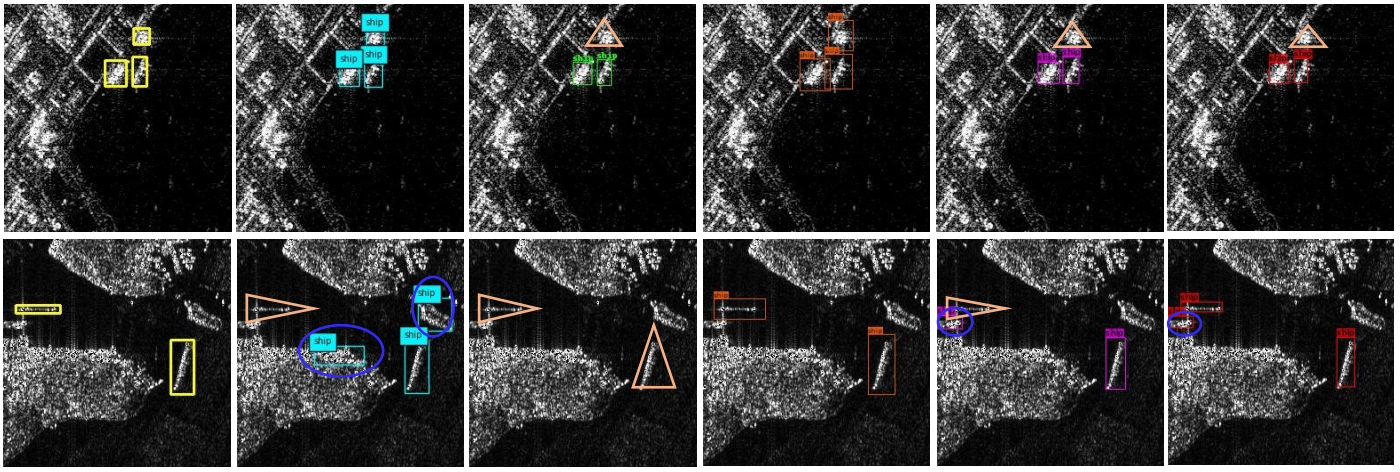
Model	AP (%) Before optical data enhancement			AP (%) After optical training data enhancement		
	Mixed	Of with SAR	Of with optical	Mixed	Of with SAR	Of with optical
YOLOv3-tiny	82.40	92.51	72.29	84.98	91.88	78.08
YOLOv3-tiny-4A-Gaussian-K	86.00	93.62	78.38	87.36	93.56	81.16
YOLOv3-spp	87.94	92.77	79.71	89.19	93.32	85.06
YOLOv3-spp-4A-Gaussian-K	90.91	96.03	85.80	91.24	95.87	86.61

To the algorithm's credit, the mixed data training model tested the ship detection results on large-scale SAR images, most of the ships can be detected as shown in Fig.15. Thus, the improved-YOLOv3 provides better ship detection than the original YOLOv3. This verifies that the improved-YOLOv3 has good robustness and generalization ability in various complex scenarios.

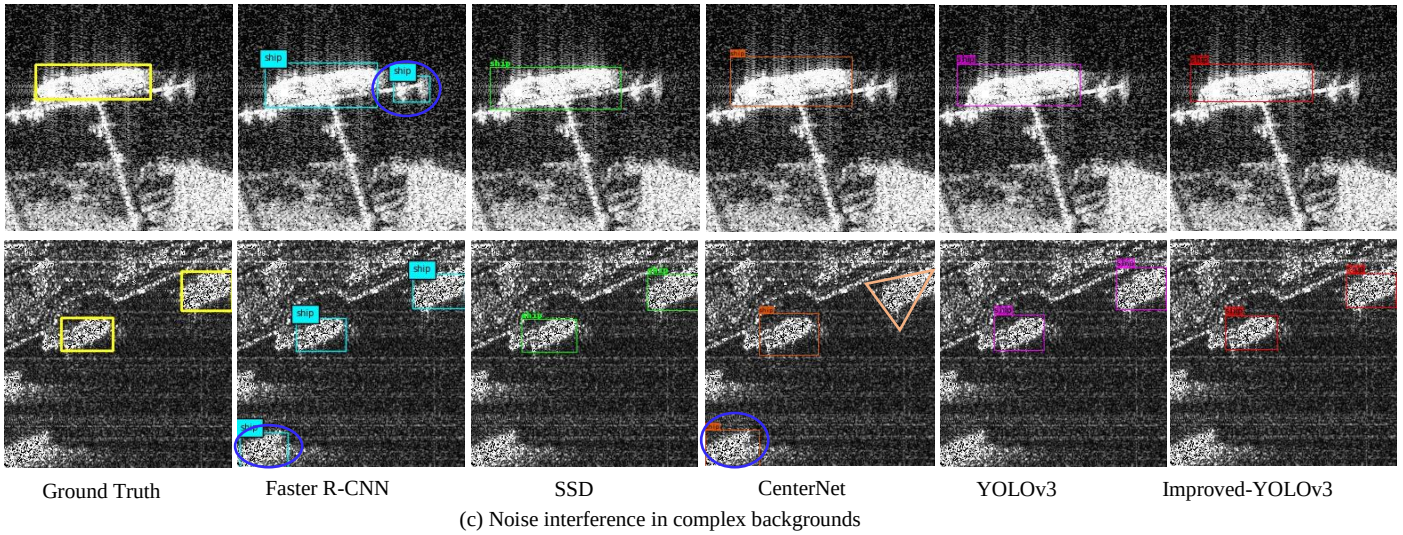
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3087555, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



(a) Multi-target in the ocean scenes



(b) Like-ships interference in the port scenes

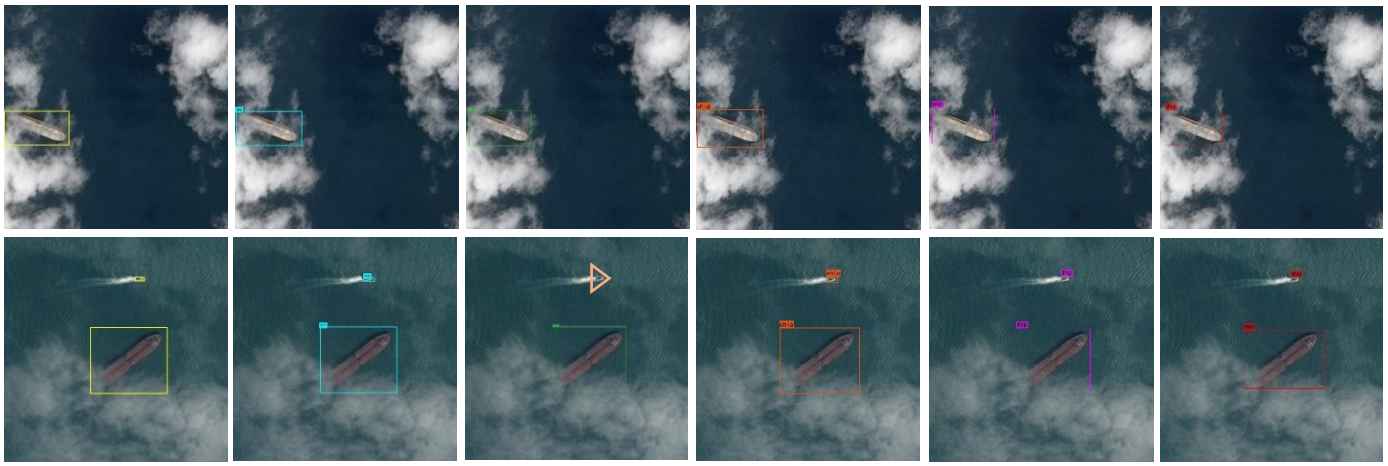


(c) Noise interference in complex backgrounds

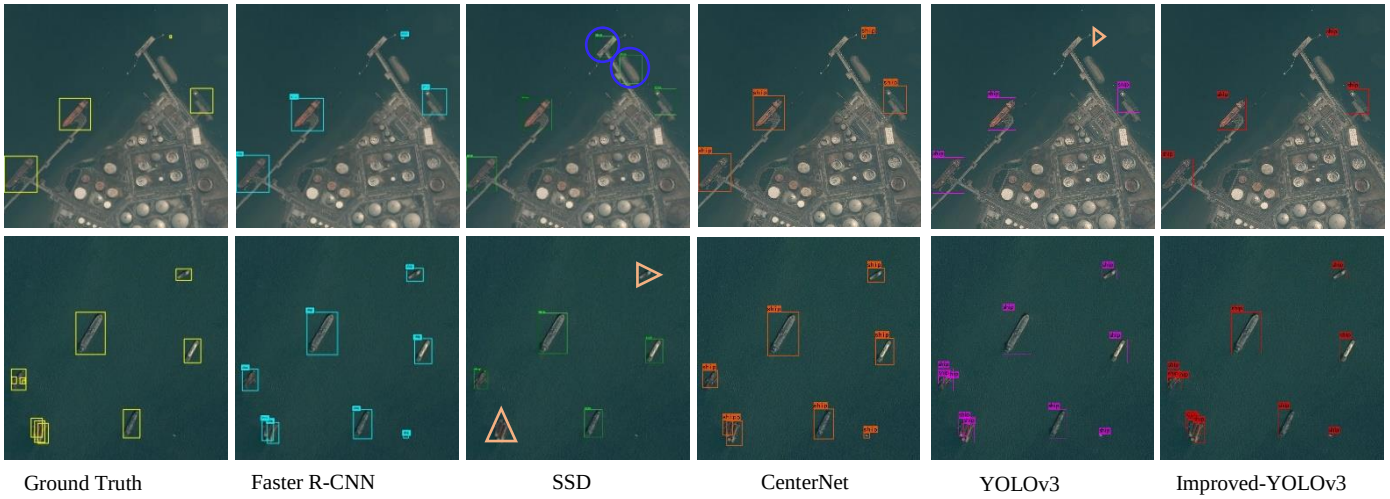
Fig. 13. Comparison of SAR image ship detection results for five deep learning detection models under different scenarios. The yellow is the real ship location in the images. The blue, green, dark orange, purple and red rectangles indicate corresponding detection results for Faster R-CNN, SSD, CenterNet, YOLOv3 and the improved-YOLOv3. Orange triangles and blue circles represent the missing ships and false alarms, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3087555, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

(a) Low contrast in complex sea conditions

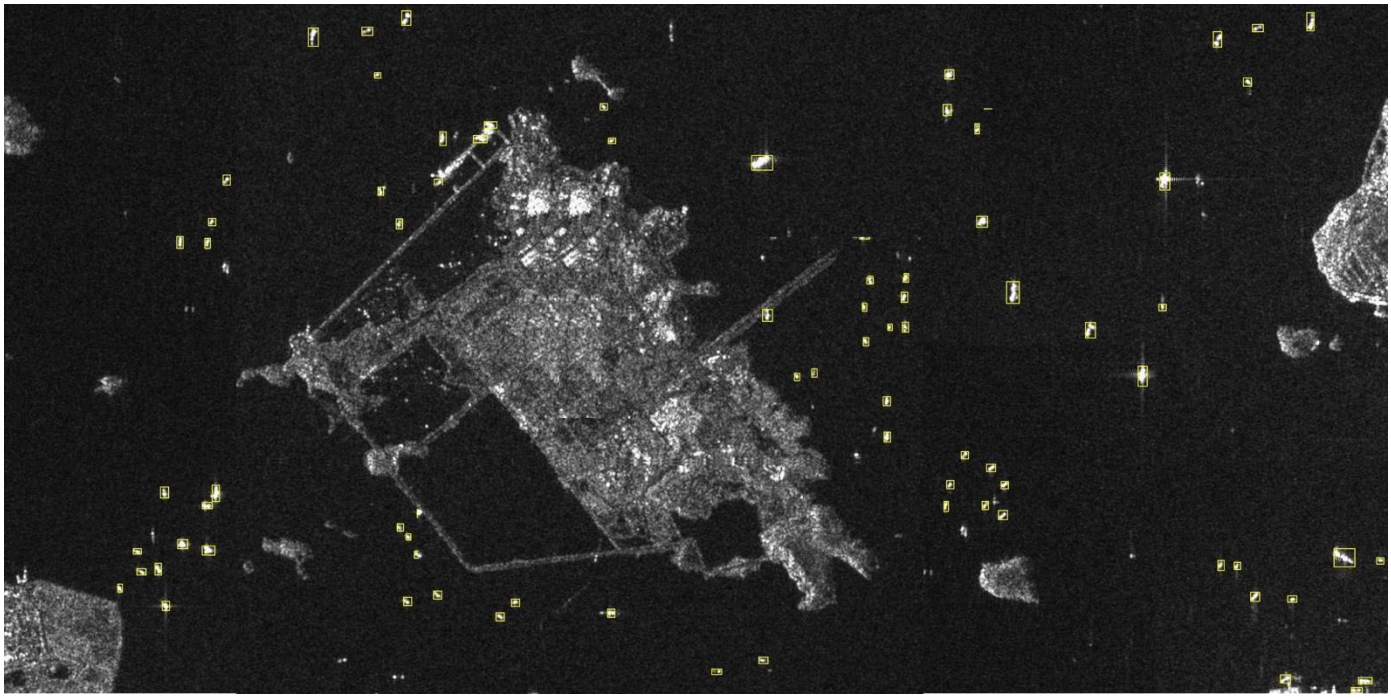


(b) Cloud interference

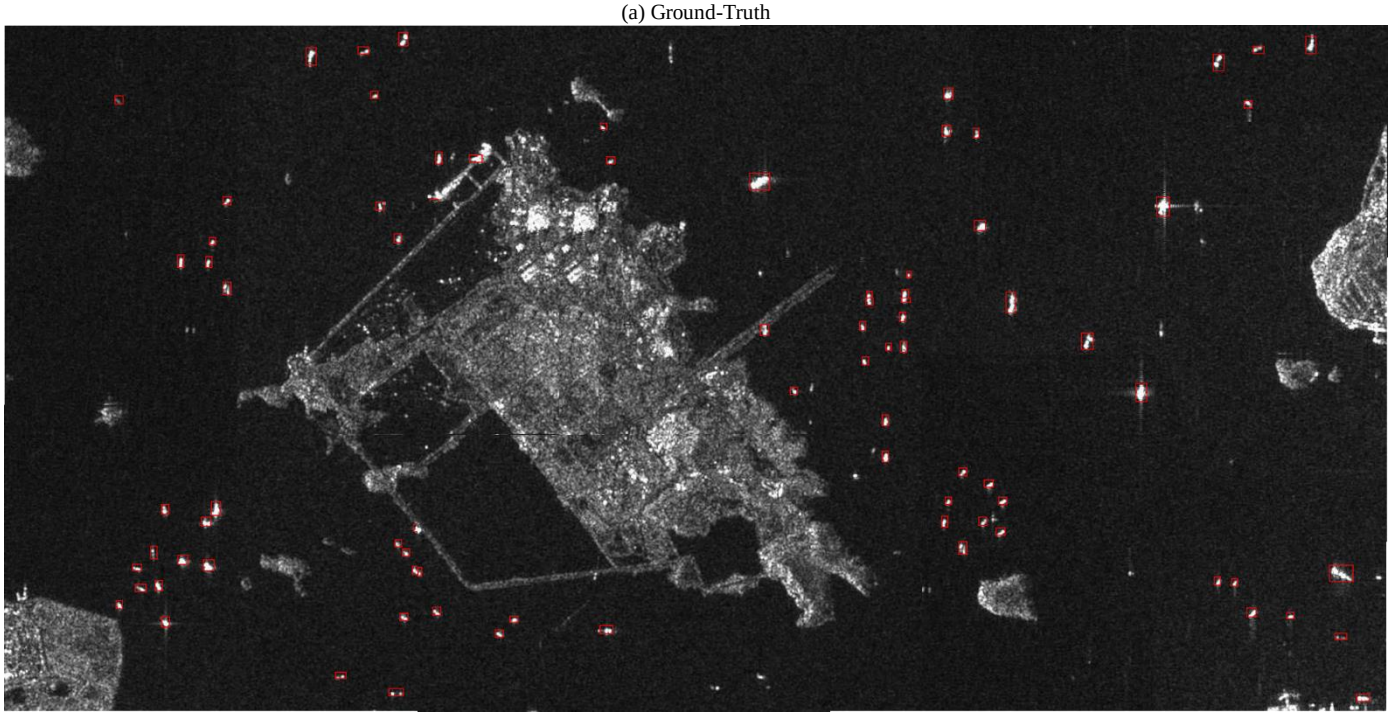


(c) Crowded ships and similar interference

Fig. 14. Comparison of optical image ship detection results for five deep learning detection models under difficult conditions. The yellow is the real location of the ship in the images. Blue, green, dark orange, purple and red rectangles indicate corresponding detection results for Faster R-CNN, SSD, CenterNet, YOLOv3 and the improved-YOLOv3. Orange triangles and blue circles represent missing ships and false alarms, respectively.



(a) Ground-Truth



(b) Detection results

Fig. 15. Results in large-scale SAR images using hybrid data training models for improved-YOLOv3. The yellow is the real ship location in the images and the red rectangles indicate detection results for improved-YOLOv3.

V. Conclusion

This study presents an improved-YOLOv3 for realising automatic ship detection from SAR and optical ship datasets. Three major modifications to YOLOv3 are proposed. First, linear stretching is introduced into the anchor box generation clustering algorithm, to solve the problem of the concentrated distribution of anchor boxes, and to highlight the advantages of YOLOv3's multi-scale detection for single-type target detection. Second, by comparing GIoU loss, DIoU loss and CIoU, the Gaussian parameters of the bounding box coordinates are introduced to predict the positioning uncertainty. So as to address the unreliable bounding box coordinate information in YOLOv3. Finally, four anchors are assigned a detection scale in the Gaussian-YOLO detection layer to address the variations in the directions and sizes of each target in remote sensing images at different resolutions. This improves the robustness of the model.

To evaluate the proposed method, three experiments were conducted, based on expanded SAR and optical ship detection

datasets. The experiments were conducted as following.

(1) The improved k-means ++ algorithm, Gaussian model, and anchor assignment strategy were simultaneously added to the YOLOv3-spp and YOLOv3-tiny models to enable experimentation on the same dataset. The experimental results show that each improved strategy is effective. The final improved-YOLOv3 is 2–3% higher than the original YOLOv3.

(2) Experiments were conducted using a mainstream detection model based on deep learning. Optical and SAR ship detection datasets in different scenes and complex backgrounds were analysed, and the improved-YOLOv3 produced satisfactory results in regards to both efficiency and AP. In addition, the overall detection accuracy for SAR ship detection was higher than that for optical ship detection. This is because optical images have a higher spatial resolution, allowing for the detection of additional image details. However, there were fewer cases of misdetection in optical ship detection, owing to the superior interpretations of the optical images.

(3) A comparison experiment was conducted between the YOLOv3 and improved-YOLOv3 approaches on the SAR and optical hybrid ship detection datasets. The intent was to verify that the improved model was effective under various conditions (such as multiple resolutions, scenes, sizes, data sources, and models). The results show that the improved-YOLOv3 approach achieves an AP of 90.91%, even under a mixed SAR-optical dataset. In addition, the performance of the mixed data training model for ship detection is tested in large SAR images, which has good robustness.

REFERENCES

[1] H. Mehner, M. E. Cutler, D. Fairbairn, and G. Thompson, "Remote sensing of upland vegetation: the potential of high spatial resolution satellite sensors," *Glob. Ecol. Biogeogr.*, vol. 13, no. 4, pp. 359-369, Jul. 2004.

[2] G. Mattyus, "Near real-time automatic marine ship detection on optical satellite images," *in Proc. Int. Soc. Photogrammetry Remote Sens. Hannover Workshop,* pp. 233–237, May. 2013.

[3] H. Greidanus, "Assessing the operationality of ship detection from space," *in Proc. EURISY Symposium*, vol. 584, Feb. 2005.

[4] A. C. Frery, H. J. Müller, C. C. F. Yanasse, and S. J. S. Santanna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May. 1997.

[5] T. Liu, Z. Yang, J. Yang, and G. Gao, "CFAR Ship Detection Methods Using Compact Polarimetric SAR in a K-Wishart Distribution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no.10, pp. 3737-3745, Jul. 2019.

[6] T. Zhang, Z. Yang, H. Gan, D. Xiang, S. Zhu and J. Yang, "PolSAR Ship Detection Using the Joint Polarimetric Information," *IEEE Trans. Geosci. Remote Sens*., doi: 10.1109/TGRS.2020.2989425, May. 2020.

[7] Y. Cui, J. Yang, Y. Yamaguchi, G. Singh, S. E. Park, and H. Kobayashi, "On semiparametric clutter estimation for ship detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens*., vol. 51, no. 5, pp. 3170-3180, Nov. 2012.

[8] C. Wang, S. Jiang, H. Zhang, F. Wu and B. Zhang, "Ship Detection for High-Resolution SAR Images Based on Feature Analysis," *IEEE Geosci Remote Sens Lett.*, vol. 11, no. 1, pp. 119-123, Jan. 2014.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May. 2015.

[10] 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. (CVPR)*, pp. 580-587, Jun. 2014.

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

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, pp. 91-99, Dec. 2015.

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn", in *Proc. IEEE Conf. Comput. Vis. (ICCV)*, pp. 2961-2969, 2017

[14] Z. Cai, and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection", in *Proc. IEEE Conf. Comput. Vis. (ICCV)*, pp. 6154–6162, 2018.

[15] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards Balanced Learning for Object Detection" 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (*CVPR*), 2020.

[16] 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)*, pp. 779-788, Jun. 2016.

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 21-37, Oct. 2016.

[18] C.Y. Fu, W. Liu, A. Ranga, A. Tyagi, A. C. Berg, "Dssd: Deconvolutional single shot detector", in: arXiv preprint arXiv:1701.06659, 2017.

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

[20] J. Redmon, and A. Farhadi, "Yolov3: An incremental improvement," Unpublished paper, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767.

[21] A. Bochkovskiy, C. Y. Wang, and H. Liao, "Yolov4: optimal speed and accuracy of object detection", 2020 [Online]. Available: http://export.arxiv.org/pdf/2004.10934.

[22] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV),2018*, pp. 734–750.

[23] *K. W. Duan, S. Bai, L. X. Xie, H. G. Q, Q. M. Huang, and Q. Tian,* "*CenterNet: Keypoint Triplets for Object Detection,"* in *Proc. IEEE Conf. Comput. Vis. (ICCV),2019, pp*.6569-6578.

[24] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," *in Proc. IEEE BIGSARDATA*, pp. 1-6, Nov. 2017.

[25] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images." *IEEE Trans. Geosci. Remote Sens*., vol. 57, no. 11, pp. 8983-8997, Nov. 2019.

[26] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. IEEE IRSIP*, pp. 1-4, May. 2017.

[27] Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images," *Remote Sens Lett.*, vol. 9, no. 8, pp. 780-788, May. 2018.

[28] Y. L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C. Y. Hsiao, and W. H. Lee, "Ship Detection Based on YOLOv2 for SAR Imagery," *Remote Sens.*, vol. 11, no. 7, pp. 786, Apr. 2019.

[29] Z.Cui, X. Wang, N. Liu , Z. Cao, & J.Yang, "Ship detection in large-scale sar images via spatial shuffle-group enhance attention". *IEEE Trans. Geosci. Remote Sens*., pp. 1-13, 2020.

[30] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, arXiv:1904.07850. [Online]. Available: http://arxiv.org/abs/1904.07850

[31] F. Bi, J. Hou, L. Chen, Z. Yang, and Y. Wang, "Ship Detection for Optical Remote Sensing Images Based on Visual Attention Enhanced Network," *Sensors*, vol. 19, no.10, pp. 2271, May. 2019.

[32] Q. Li, L. Mou, Q. Liu, Y. Wang, and X. X. Zhu, "HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7147-7161, Dec. 2018.

[33] F. Wu, Z. Zhou, B. Wang, and J. Ma, "Inshore ship detection based on convolutional neural network in optical satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*., vol. 11, no. 11, pp. 4005-4015, Oct. 2018.

[34] Y. Guan, A. A. Ghorbani, and N. Belacel, "K-means+: An autonomous clustering algorithm," *Technical Report TR04-164*, University of New Brunswick, 2004.

[35] 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, pp. 765, Mar. 2019.

[36] Kaggle, accessed on Aug. 19, 2020. [Online]. Available: https://www.kaggle.com/c/airbus-ship-detection.

[37] J. Jiao, Y. Zhang, H. Sun, X. Yang, X. Gao, W. Hong, K. Fu and X. Sun, *"*A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection*," IEEE Access*, vol. 6, pp. 20881-20892, Apr. 2018.

[38] C. Zhang, C. C. Chang, and M. Jamshidi, "Concrete bridge surface damage detection using a single‐stage detector," *Comput-Aided Civil Infrastruct. Eng.*, vol. 35, no. 4, pp. 389-409, Otc. 2020.

[39] Z. Huang, J. Wang, X. Fu, T. Yu, Y. Guo, and R. Wang, "DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection," *Information Sciences*., 2020, pp. 241-258.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *in Proc. Int. Conf. Learn. Represent. (ICLR)*, pp. 1–14, Sep. 2015.

[41] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis. (ECCV),2016*, pp. 483–499.

[42] K. Sun, Y. Zhao, B. R. Jiang, T. H. Cheng, B. Xiao, D. Liu, Y. D. Mu, X. G. Wang, W. Y. Liu, and J. D. Wang, "High-resolution representations for labelling pixels and regions," CoRR, abs/1904.04514.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.770-778, Oct. 2016.

[44] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens*., vol. 55, no. 5, pp.2486-2498, Jan. 2017.

[45] J. Choi, D. Chun, H. Kim, and H. Lee, "Gaussian YOLOv3: An Accurate and Fast Object Detector Using Localization Uncertainty for Autonomous Driving," in *Proc. IEEE Conf. Comput. Vis. (ICCV)*, pp. 502-511, Otc. 2019.

[46] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression". in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*., pp. 658-666, 2019.

[47] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, Distance-IoU loss: "Faster and better learning for bounding box regression", *in Proc. IEEE Conf. AAAI Artif Intell.*, Vol. 34, No. 07, pp. 12993-13000, April. 2020.

[48] 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. 2017.

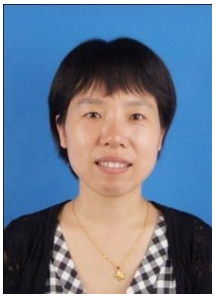
Zhonghua Hong received the Ph.D. degrees in GIS from Tongji University, Shanghai, China, in 2014. Currently, he has been an Associate Professor in the College of Information Technology, Shanghai Ocean University since 2019. His research interests include 3D damage detection, coastal mapping, photogrammetry, GNSS-R and deep learning.

Ting Yang received the B.S. degree in computer science and technology from Jiujiang University, Jiangxi, China, in 2018. She is currently pursuing the M.S. degree in software engineering with Shanghai Ocean University, Shanghai, China. Her research interests include deep learning, object detection and semantic segmentation.

Shenlu Jiang received his Ph.D. Degree in electronic and electrical engineering in the School of Electrical and Electronics Engineering at Sungkyunkwan University in 2020. From 2020.2 – 2020.9, he was a research assistant in the Institute of Space and Earth Information Science of The Chinese University of Hong Kong. Currently, he is a postdoctoral researcher in the TITANE team of the French Institute for Research in Computer Science and Automation (INRIA). His research interests include computer vision, robot vison, remote sensing and deep learning.

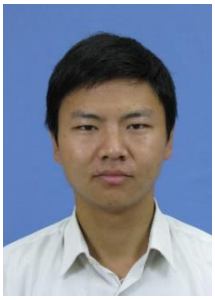
Yun Zhang received the Ph.D. degree in applied marine environmental studies from Tokyo University of Maritime Science and Technology, Tokyo, Japan, in 2008. From 2011 to present, he was a Professor with the College of Information and Technology, Shanghai Ocean University, Shanghai, China. His research interests include the study of navigation system reflection signal technique and its maritime application.

Ruyan Zhou received her Ph.D. degree in agricultural bio-environment and energy engineering in Henan Agricultural University in 2007. From 2007-2008, she worked in Zhongyuan University of Technology. Currently, she is working with Shanghai Ocean University, Shanghai, China.



Yanling Han received his B.E. degree in mechanical design and manufacturing and the M.E. degree in mechanical automation from Sichuan University, Sichuan, China, and his Ph.D. degree in engineering and control theory from Shanghai University, Shanghai, China. She is a Professor and currently working with the Shanghai Ocean University, Shanghai, China. Her research interests include the study of ocean remote sensing, flexible system modelling, and deep learning.

Jing Wang received her PhD. degree of Biomedical Engineering, in the department of biomedical engineering of Shanghai Jiaotong University in 2014. Currently, she has been a Lecturer in the College of Information Technology, Shanghai Ocean University since 2015. Her research interests include computer vision, medical image processing.



Shuhu Yang received Ph.D. degrees in physics of physics from School of Physics, Nanjing University. Since 2012.9, he has been the lecturer in the College of Information Technology, Shanghai Ocean University. His research interests include Evolution of the Antarctic ice sheet, hyperspectral remote sensing, and the use of navigational satellite reflections.

Xiaohua Tong received his Ph.D. degree from Tongji University, Shanghai, China, in 1999. From 2001 to 2003, he was a Postdoctoral Researcher at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow at Hong Kong Polytechnic University, Hong Kong, in 2006, and a visiting scholar at the University of California, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include photogrammetry and remote sensing, trust in spatial data, and image processing for high-resolution satellite images.

Sichong Liu received the B.Sc. degreein geographical information system and the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2009 and 2011, respectively, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy,

2015. He is currently an Associate Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research interests include multitemporal remote sensing data analysis, change detection, multispectral/hyperspectral remote sensing, signal processing, and pattern recognition. Dr. Liu was the winner (ranked as third place) of Paper Contest of the 2014 IEEE GRSS Data Fusion Contest. He is the Technical Co-Chair of the Tenth International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp 2019). He was the Session Chair for many international conferences such as International Geoscience and Remote Sensing Symposium. He is also a Referee for more than 30 international journals.