LMO-YOLO: A Ship Detection Model for Low-Resolution Optical Satellite Imagery

Qizhi Xu, *Member, IEEE*, Yuan Li, and Zhenwei Shi, *Member, IEEE*

***Abstract*—It has been observed that the existing convolutional neural network (CNN)-based ship detection models often result in high false detection rate in low-resolution optical satellite images. This problem arises from the following factors: 1) the current 8-b rescaling schemes make the images lose some important information about ships in low-resolution imagery; 2) the effective features of ships at low resolution are far fewer than those of ships at high resolution; and 3) the detection of low-resolution ships is more sensitive to object-background contrast variation. To solve these problems, a low-resolution marine object (LMO) detection YOLO model, called LMO-YOLO, is proposed in this article. First, a multiple linear rescaling scheme is developed to quantize the original satellite images into 8-b images; second, dilated convolutions are included in a YOLO network to extract object features and object-background features; finally, an adaptive weighting scheme is designed to balance the loss between easy-to-detect ships and hard-to-detect ships. The proposed method was validated by level 1 product images captured by the wide-field-of-view sensor on the GaoFen-1 satellite. The experimental results demonstrated that our method accurately detected ships from low-resolution images and outperformed state-of-the-art methods.**

***Index Terms*—Contrast sensitive loss, dilated convolution, low-resolution imagery, ships detection.**

I. Introduction

MARINE ship detection technology is of great importance to civil and military applications. With the development of remote sensing technology, the volume of remote sensing data has tremendously increased. Compared with other remote sensing image data, optical remote sensing images are rich in content and easy to understand. Therefore, many studies have been carried out on optical remote sensing images, especially in the field of ship detection [1]–[3].

Notably, to obtain more detailed information, the original remote sensing image products obtained from sensor are generally between 10 and 14 b. By way of illustration, the optical images provided by WorldView-2 [4], WorldView-3 [5], and QuickBird [6] are 11-b data. And data produced by moderate resolution imaging spectroradiometer instruments is 12 b. According to the clouds and the Earth's radiant energy system record (2000 to the present), most GEO imagers have implemented 10-b quantization [7].

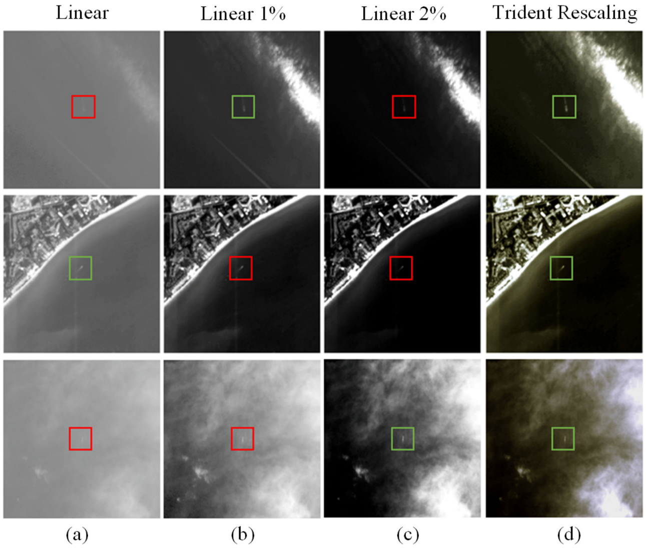

Fig. 1. Columns (a)–(d) show the comparison results of linear, 1% linear, 2% linear, and the trident rescaling (our approach) methods, respectively. The green boxes in columns (a)–(c) indicate that the target obtained by this rescaling method is the most salient object produced by the three schemes. The images in column (d) show that trident recaling achieves the best quantization in each scene.

However, in existing convolutional neural network (CNN)-based ship detection methods, the training samples are generally 8-b images. Because model training requires hand-labeled target ground-truth for supervision, the images must be visualized. Furthermore, the computational processing of the original high bit-depth data will greatly increase the memory consumption, which is not supported in practical engineering applications. Therefore, the publicly available datasets used in popular CNN-based target detection methods, such as the HRSC2016 [8], DOTA [9], and DIOR [10] datasets, are all 8-b data.

Generally, the 8-b datasets are obtained by a single quantization method. Nevertheless, a single 8-b image cannot clearly show all the useful information for low-resolution ships. As shown in Fig. 1, no one single quantization method works well for all images in different scenes. Some 8-b images rescaled by only one single quantization method will lose some details, which is not conducive to the detection of small or weak targets. Therefore, it is necessary to develop a good quantization method for deep network model to keep as much information as possible.

Manuscript received January 25, 2022; revised February 11, 2022 and April 13, 2022; accepted May 11, 2022. Date of publication May 20, 2022; date of current version June 2, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61972021 and Grant 61672076. *(Corresponding author: Yuan Li.)*

Qizhi Xu and Yuan Li are with the School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: qizhi@bit.edu.cn; liyuansme@bit.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3176141

Meanwhile, the effective features of low-resolution marine object (LMO) are fewer than that of high-resolution marine object. They are easily weakened in a deep network layer. Most of the existing methods aim to extract more target features while ignoring the importance of the background. The tiny object detection is sensitive to object-background contrast variation. For these reasons, it is difficult to obtain accurate detection results. Traditional object detection was dominated by works that make use of sample features. It includes color range, texture feature, local binary pattern (LBP), and scale-invariant feature transform, etc. In [11], structure-LBP feature descriptor is added to ship detection structure. The LBP feature can be combined with spatial information to achieve a more discriminative ship description. In [12], the histograms of oriented gradients and local binary patterns is extracted from images as effective discriminative features.

Currently, deep learning methods have shown good performance in this field and are widely studied [13], [14]. They generally can be classified to unsupervised learning and supervised learning. Unsupervised learning is a method of learning from unlabeled data and finding the patterns or high-level semantics of the data. Liao *et al.* [15] also proposed an unsupervised cluster-guided object detection method to address the dense detection problem in some scenes. For supervised object detection methods, CNN has become an good choice for many fields of image processing. For an instance, Sharifzadeh *et al.* [16] proposed an hybrid algorithm of CNN and multilayer perceptron for ship detection. It can effectively extract the images internal features and obtain good detection results. Moreover, the performance of the popular YOLO models [17]–[20] is even more surprising. They have the powerful ability to learn regions of interest and context at the same time. In particular, YOLOv4 [20] has become an efficient and accurate model with a high mean average precision. Although it is not specifically designed for small target detection, it can still be used as an excellent benchmark method [21].

Based on the above analysis, a novel LMO-YOLO network for low-resolution optical satellite imagery is proposed in this article. The main objectives of this study are as follows:

1) To develop a rescaling method that can retain more detailed information.
2) To improve the performance of LMO detection by considering more object-background features.

Only by extracting more object-background information can the network more easily distinguish the positive and negative samples, such as ship and ship-shaped scattered clouds. Overall, this study makes three contributions: first, a multiple linear rescaling scheme is designed to effectively alleviate the information loss problem of a single quantization method. Second, to capture more object-background information, a multiscale dilated convolution (MDC) module is constructed on the backbone of YOLOv4. Finally, because hard detection samples have a greater impact on model performance, a contrast-sensitive loss is employed to balance the weight between hard and easy samples.

The rest of this article is organized as follows. Section II investigates the related works and Section III describes the details of the proposed ship detection method. Experimental results and detailed comparisons are shown in Section IV to verify the superiority of our method. Finally, Section V concludes this article.

II. Related Works

A. Eight-Bit Rescaling Schemes

In the past years, a multitude of rescaling algorithms for remote sensing images have been put forward by researchers. Many software have one-click rescaling tools. Taking ENVI software as an example, there are several simple image scaling approaches that can be used, including linear [22], linear 2% [23], Gaussian [24], and square root stretching [25]. Most of these methods require few steps and are easy to complete. However, it needs to be set manually, which is not suitable for automatic operation. The results of different quantization methods also vary widely. Additionally, The quantization effect of different scenes is greatly affected by its background. For complex background images, it is difficult for a single quantization method to achieve acceptable results for all scene image blocks. As shown in Fig. 1, a single linear quantization method usually cannot obtain a satisfactory result in practice.

Currently, many studies focus on quantizing images and improving image contrast. In [26], a general framework based on histogram equalization for image contrast enhancement is presented. Conventional histogram equalization is optimized by introducing specifically designed penalty terms. In [27], a subband decomposition multiscale retinal method containing a hybrid intensity transfer function is introduced to enhance optical remote sensing images. In 2012, Celik *et al.* [28] proposed a 2-D histogram-based method. Contextual information is utilized to enhance the contrast in the input image. In addition, a guided image contrast enhancement method is proposed in [29]. This method improved the context-sensitive and context-free contrast by solving a multicriteria optimization problem and efficiently created visually pleasing enhanced images. To meet the requirements of automation and efficiency in applications, Liu *et al.* [30] presented a novel self-adaptive histogram compacting transform-based contrast enhancement method for remote sensing images. All these methods can improve the image quality of not only remote sensing images but also natural images. However, they mainly address 8-b images and cannot avoid losing some detail information of the original 16-b image.

Based on the above analysis, a good quantization method is necessary. Because a single quantization method cannot perform well for all targets, a trident linear rescaling scheme is proposed in this article. Here, "trident" denotes that three rescaling approaches are simultaneously utilized to quantize the original image data and superimpose their results as three-channel training images. In this study, linear, linear 1%, and linear 2% stretching methods are selected. Fig. 1 gives the example results, and the green boxes show that the corresponding method has a better enhancement effect on that ship. Although the linear stretching methods are commonly used, it is worth considering that how to better apply them to rescale images with different background. The trident rescaling method is simple, yet very effective. It not only improves the contrast of the image but also maximizes the retention of information. Moreover, this quantization scheme has good adaptability to various scenes.

B. CNN-Based Ship Detection Methods

In recent years, encouraged by the great success of deep learning methods, many CNN-based studies have been proposed to detect marine ship [31]–[33]. They can automatically extract complex features from different levels in raw images and achieve superior detection performance. The detection performance of these methods has surpassed that of feature engineering methods [34]. This work also explores CNNs to study LMO detection. Thus far, popular remote sensing image target detectors have generally been developed for high-resolution images (large targets) and low-resolution images (small targets), respectively.

For most existing detection methods for high-resolution remote sensing images, the accurate localization of detected objects is an issue of concern. Based on this point of view, Long *et al.* [35] proposed a new method for solving the problem of automatic accurate localization of detected objects. An unsupervised score-based bounding box regression algorithm, combined with a nonmaximum suppression algorithm, was developed to optimize the bounding boxes of regions. In [36], a unified and effective method for simultaneously detecting objects was proposed. It achieved more accurate detection by adding a redesigned inception module and an accurate object detection module. Because the existing complex object detectors are not satisfactory [37], a unified part-based CNN is specifically designed. In [38], a sparse anchoring guided high-resolution capsule network (SAHR-CapsNet) was designed for geospatial object detection based on high-resolution remote sensing images. These methods have good positioning and detection performance for large objects at high resolution.

Moreover, many researchers have also been working toward high-resolution target detection using different approaches. Shi *et al.* [39] detected ships in high-resolution optical imagery in a “coarse-to-fine” manner. They transformed the panchromatic image to a “fake” hyperspectral form to amplify the separability between ships and background. After that, SVDNet [40] was proposed to solve the problems of background interference and high computational expenses. The experimental results demonstrate the superiority of SVDNet. Furthermore, a new paradigm formulated [41] from a Bayesian perspective was designed to detect targets in high-resolution aerial remote sensing images, and it outperformed many state-of-the-art methods. Different from other approaches that usually work well only at one scale, HSF-Net [42] was proposed to efficiently detect ships at various scales. This was achieved by using an added hierarchical selective filtering layer. These methods can obtain good detection results for high-resolution remote sensing image. However, the performance of these methods for small object detection in low-resolution images needs to be further improved.

Other researchers have also done work in low-resolution small object detection. On the one hand, small object detection often faces the problem of insufficient feature information in the deep layers of CNNs. To tackle the challenges brought by low resolution and noisy representation, a single perceptual generative adversarial network is developed in this article [43]. In 2020, a novel network called image pyramid guidance network (IPG-Net) [44] is proposed to ensure both the extracted spatial and semantic information are abundant. Thus, the typical problem in CNNs of information imbalance can be solved. In [45], a rotatable region-based residual network (R^3-Net) is proposed. It can detect small and dense objects by generating rotatable rectangular target boxes in a half coordinate system. In [46], Deng *et al.* proposed an extended feature pyramid network (EFPN) to reduce the damage done by feature coupling at various scales to the detection performance of small objects. These approaches have played a certain role in promoting the research of small target detection.

On the other hand, small object detection is also affected by the problem of sample imbalance. The existing sample balancing methods focus on two aspects: 1) the imbalance of positive and negative samples; and 2) the imbalance of hard and easy samples. To overcome these problems, OHEM [47] ranks all negative samples according to their loss values. The negative samples with the largest loss value are selected for targeted training to improve the detection accuracy. In [48], the author proposed a RetinaNet with focal loss to reduce the weight of many simple negative samples in training. In [49], a gradient harmonizing mechanism was proposed to solve the problem of outlier points in the sample and make the training of the model more reasonable. Libra R-CNN [50] achieved a further breakthrough in public datasets with a novel sampling method that proposed a more balanced loss function. In general, these approaches can yield competitive performance. However, they are also limited by the object-background contrast variation. Although some of these methods perform well on natural images, the LMO detection results on remote sensing images with complex backgrounds are not good enough.

III. Methodology

The proposed method, LMO-YOLO, is composed of a trident rescaling module and an object detection module. The overall framework is shown in Fig. 2. The trident rescaling module is applied to obtain the training samples. Three different quantization methods are utilized to rescale the raw satellite data. Next, the obtained three 8-b images are superimposed as an RGB three-channel image to retain all useful information. Then, the training samples after preprocessing are fed into an optimized YOLO network. In particular, the MDC layer added in the backbone network can extract more object and object-background features under different receptive fields, so the weak targets can be accurately detected by adding auxiliary information. Additionally, a contrast-sensitive loss is designed to reweight the hard samples for detection. The lower the contrast between an target anchor and its surrounding area, the harder it is to be detected. A more detailed description is given in the following subsections.

A. Trident Rescaling Approach

To gain effective training samples, the original 10 to 14-b remote sensing images were quantized to 8 b and then cut

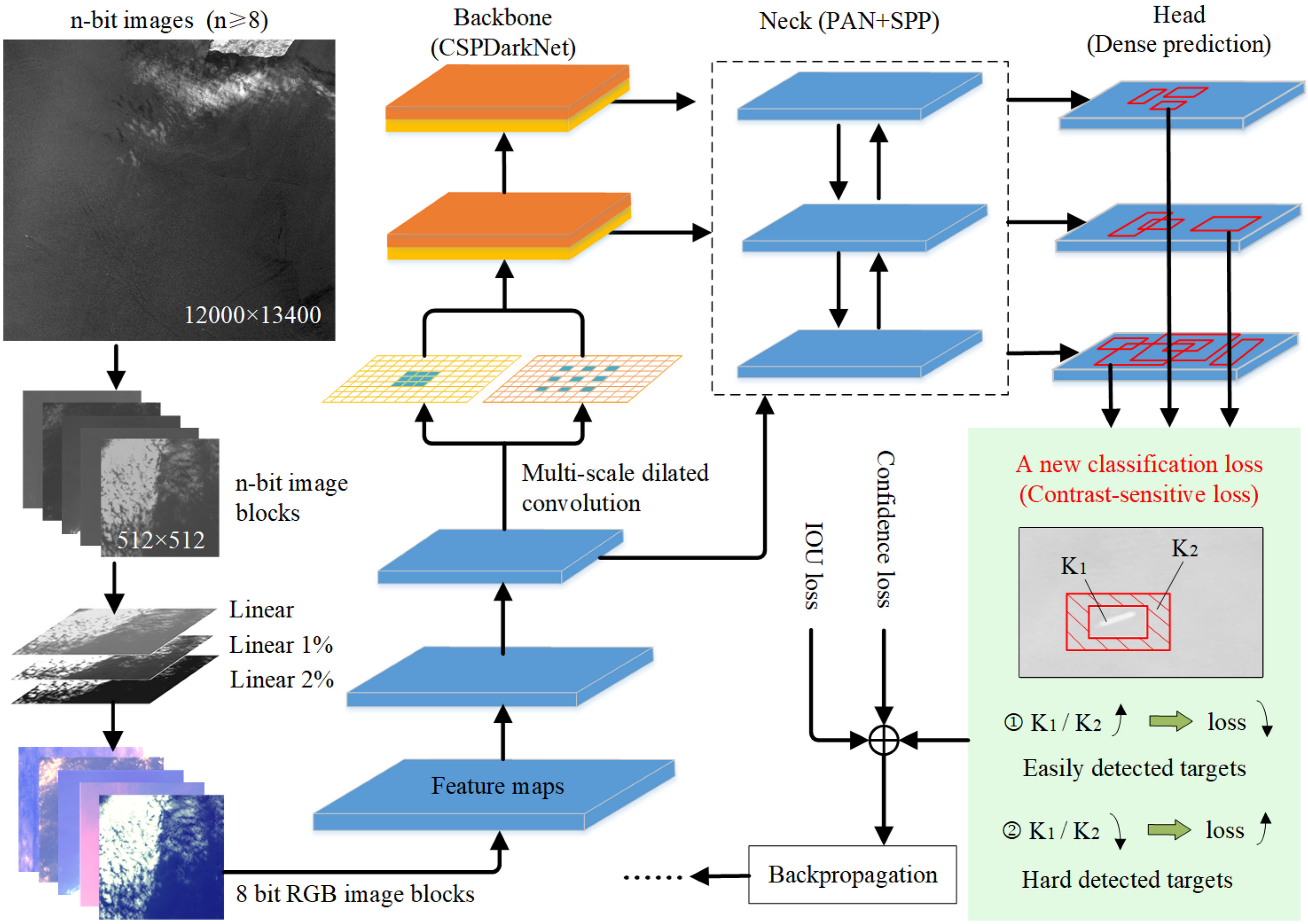

Fig. 2. Overview of the proposed ship detection method LMO-YOLO. The method consists of three stages overall: image preprocessing, feature extraction, and prediction. The details of the injection points are described in the three subsections of Section II.

into slices. A single rescaling method may cause information loss for some small and weak ships. As shown in Fig. 1(a)–(c), each linear quantization method has its own strengths and weaknesses. Specifically, compared with other quantization methods, linear quantization (that is, quantization of all gray values to [0, 255]) can preserve relatively more information. It will produce images with uniform gray values, but the ship contour is blurred. Percentage linear quantization contributes to improving target and background contrast but suffers from impaired detail. The higher the percentage value, the greater the contrast. Therefore, we chose three stretching methods, linear, linear 1$\%$, and linear 2$\%$, to simultaneously quantize the original images. Then these 8-b single-channel images were superimposed into a false-color image (RGB), as shown in Fig. 1(d). In the end, a better-quality sample set was obtained.

The percentage rate of quantization is aligned with the different grayscale distribution of the dataset. The image I' after percentage quantization can be obtained with the following equations:

$$I' = \begin{cases} 255\,, & I_{x,y} \geq I_a \\ \dfrac{I_{x,y} - I_b}{I_a - I_b} \times 255\,, & I_b < I_{x,y} < I_a \\ 0\,, & I_{x,y} \leq I_b \end{cases} \tag{1}$$

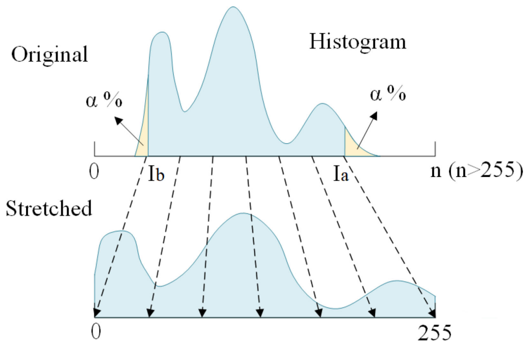

Fig. 3. Schematic diagram of percent linear stretches.

where I_a and I_b are calculated as percentages, as shown in Fig. 3. I_a corresponds to the gray value obtained by subtracting the first $\alpha\%$ pixels from the maximum gray value in the gray histogram of the original data. Similarly, I_b corresponds to the gray value obtained by adding the last $\alpha\%$ pixels from the minimum gray value. $I_{x,y}$ denotes the gray value at the point (x, y), and α is predefined according to the grayscale distribution of the dataset. When $\alpha = 0$, the percentage quantization is converted to linear quantization. In this article, we set α to 0, 0.01, and 0.02.

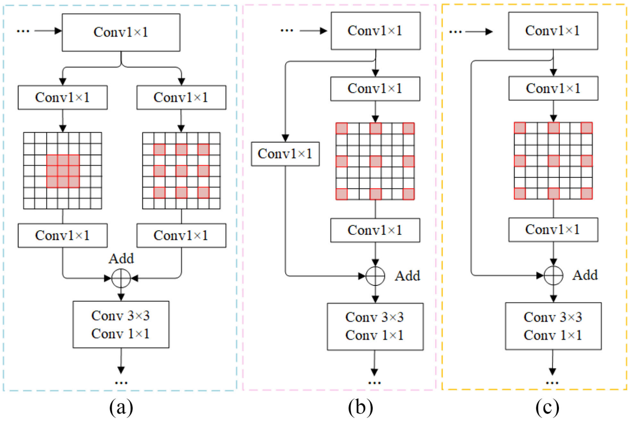

Fig. 4. Structure of proposed MDC blocks with different receptive fields. The dilate rate is 1 in (a), and 2 in (b) and (c). In the backbone, MDC 1, 2, and 3 are connected in series to extract multiscale features.

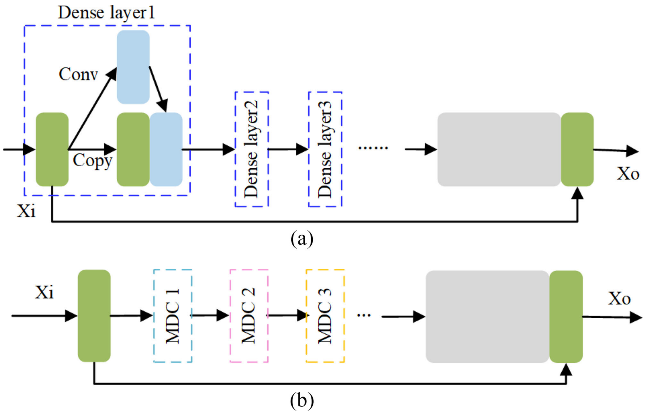

Fig. 5. Flowchart of backbone modules. (a) Original cross stage partial module in CSPDarkNet. (b) The dilated module developed in this article. MDC: Multiscale dilated convolution blocks.

B. Dilated Convolution Improved Backbone

It can be known from many experiments that detection accuracy has a close relationship with the amount of information captured. However, information from small objects is easily weakened as the network deepens and the spatial resolution decreases. Thus, this is one of the major drawback of small object detection. In addition, background variations also have a large interference on LMO detection and most existing methods ignores the background information. Based on this consideration, we optimized the original YOLO network and introduced an MDC module in the backbone to extract more object and object-background difference features at multiple scales. When the network is more accurate at discriminating negative samples, the detection rate of positive samples will also increase. The standard dilated convolution operation is expressed as follows:

$$(F * g)(\mathbf{r}) = \sum_{\mathbf{m}+\mathbf{n}=\mathbf{r}} F(\mathbf{m}) g(\mathbf{n}) \tag{2}$$

where F denotes a discrete function and g is a discrete filter. Then, the dilation factor can be generalized as

$$(F *_l g)(\mathbf{r}) = \sum_{\mathbf{m}+l\mathbf{n}=\mathbf{r}} F(\mathbf{m}) g(\mathbf{n}). \tag{3}$$

The $*_l$ is defined as a dilated convolution or an s-dilated convolution. The standard convolution $*$ is simply the 1-dilated convolution. The dilated convolutions can expand receptive fields without losing resolution or coverage. In the dilated convolution module designed in this study, there are three different dilated convolution blocks. All dilated convolution kernels are discrete 3×3 filters.

$$F_{i+1} = F_i *_{2^i} g_i \text{ for } i = 0, 1, 2. \tag{4}$$

The specific structure of each block is illustrated in Fig 4. MDC 1 includes a standard convolution and a dilated convolution kernel with a dilated rate of 1. MDC 2 and 3 mainly contain a dilated convolution kernel with a dilated rate of 2. Benefiting from these dilated convolution kernels, the receptive fields of the network can be expanded without decreasing the resolution, and more effective information can be captured.

In addition, the whole detection framework of our approach is constructed based on YOLOv4. It has three main components: 1) backbone, 2) neck, and 3) detection heads. First, the backbone is a CSPDarkNet module [20] that is optimized for small target detection tasks and that combines many tricks, such as CSPX, the Mish activation function, and Dropblock. It follows the filter size and overall structure of the ResNe(X)t network, adding a cross stage partial structure to each group of residual blocks. It also reduces the parameters to make it easier to train. There are five cross stage partial modules in the backbone. The first three modules keep the original CSP structure, the latter two modules use the developed dilated module, which is given in Fig. 5. This dilated convolution module is constructed by connecting three MDC blocks in series. Meanwhile, we fix the spatial resolution after stage 3. Then, the high spatial resolution of the feature maps can be maintained, and a large receptive field can also be kept. Therefore, the added MDC module can capture more information about objects and object-background features. Second, the neck part includes spatial pyramid pooling (SPP) module, feature pyramid network (FPN), and path aggregation network (PAN) structures. The SPP component contains a multiscale pooling structure, which can complete multiscale feature fusion. Third, the prediction part includes three multiscale detection heads. Both target prediction and loss calculation are realized in this step. We fixed these modules with the same structure as YOLOv4 to verify the effectiveness of our method. More details can be seen in [20].

C. Contrast-Based Loss

According to the analysis above, we believe a satisfactory loss function for LMO detection should focus more on indistinguishable samples and reduce the effect of numerous simple samples. The contrast-based (or contrast-sensitive) loss is designed to improve the comprehensive detection performance of LMO-YOLO. Common classification loss will lead to insufficient learning of the hard detection object area. On the one hand, because the contrast between some small objects and the background is very small, it is difficult for the network to detect.

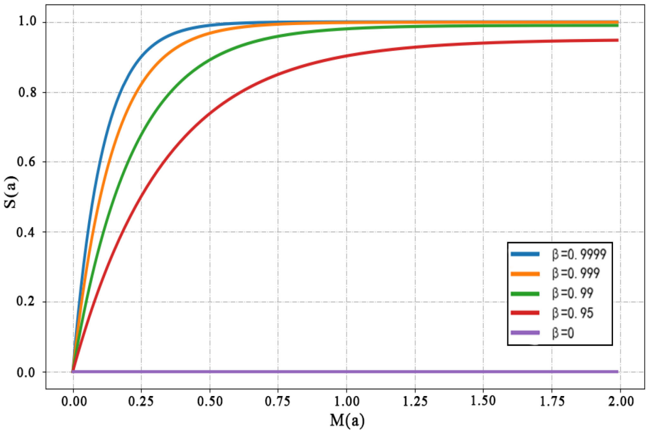

Fig. 6. Visualization of the proposed contrast-based term $S(a)$.

On the other hand, these hard samples only make up a fractional part of whole samples. The contrast-based loss function improves the feature quality by computing the contrast of each anchor with its surrounding area. In this loss function, the weight of each object sample is inversely proportional to the contrast.

The contrast of the two regions can also be conversely referred to as the similarity of the two regions. There are many methods that can be utilized, namely, Euclidean distance, Manhattan distance, Minkowski distance, and Mahalanobis distance. They all have their own advantages and disadvantages. In this study, Mahalanobis distance is more suitable for regional contrast calculation. It has higher stability and can eliminate the interference of correlation between variables. Here, regional contrast based on Mahalanobis distance is calculated by

$$M^2(a) = (\mu_a - \mu_b)^T C_b^{-1} (\mu_a - \mu_b) \tag{5}$$

where $\mu_a = \frac{1}{M}\sum_{i=1}^{M} x_i$ is the mean value of the inner area of an anchor, which is shown as K1 in Fig. 2. $\mu_b = \frac{1}{N}\sum_{j=1}^{N} x_j$ is the mean of the surrounding window, which is shown as K2 in Fig. 2. The size of K2 is designed to be twice as large as k1. That is, keep the center point unchanged and multiply the original length and width of each anchor by 2. C_b is the covariance matrix of the surrounding window

$$C_b = \frac{1}{N}\sum_{j=1}^{N} (x_j - \mu_b)(x_j - \mu_b)^T. \tag{6}$$

Hence, the larger the $M(a)$, the greater the distance between the anchor and its surrounding area, or the greater the contrast.

For a sample x with label y in a ship detection task, when it is a positive sample (that is, y=1), the larger the $M(a)$, the easier it is to detect. On the contrary, when y=0, the smaller the $M(a)$, the easier it is to accurately distinguish. To eliminate the adverse effects of singular sample data, we normalized $M(a)$ to [0,1], as shown in Fig. 6. The contrast-based weighting factor $S(a)$ is set as follows:

$$S(a) = \beta \times \left(1 - (1-\beta)^{M(a)}\right). \tag{7}$$

According to the distribution of training sample data, we set belta value to 0.99. In this way, the samples with different contrast against background can be spread out evenly in a large range. The detection results is better with this setting value. The $S(a)$ notation is independent of the model and the loss.

This balance term is applied to focal Loss. The recently proposed focal loss [20] adds a modulation factor to the sigmoid cross-entropy loss to reduce the relative loss of well-classified samples and to focus on difficult samples. The focal loss can be expressed as

$$F_{fl} = \begin{cases} -\alpha(1-y')^{\gamma}\log(y') & y=1 \\ -(1-\alpha)(y')^{\gamma}\log(1-y') & y=0 \end{cases} \tag{8}$$

where the factor $(1-y')^{\gamma}$ to the standard cross entropy criterion. The balance factor α is added to balance the uneven ratio of positive and negative samples. However, for the approach in this study, the contrast-sensitive focal loss can be written as

$$C_{fl} = \begin{cases} -(1-S)(1-y')^{\gamma}\log(y') & y=1 \\ -Sy'^{\gamma}\log(1-y') & y=0. \end{cases} \tag{9}$$

Consequently, the larger the contrast between an anchor and its surrounding area, the larger the S. If it corresponds to a positive label ($y=1$), the loss will be small. On the contrary, the loss will be large. In addition, intersection over union loss (L_{IoU}) and confidence loss (L_{conf}) are also weighted with this optimized classification loss to form total loss in this method. IoU loss is the CIoU loss including the aspect ratio factor; confidence loss is binary cross entropy loss (BCEloss), see [20] for details. Then, the total loss is defined as follows, and λ_1, λ_2, and λ_3 are the set parameters

$$F_{\text{total}} = \lambda_1 \cdot L_{\text{conf}} + \lambda_2 \cdot L_{\text{IoU}} + \lambda_3 \cdot C_{fl}. \tag{10}$$

IV. Experiments

In this section, we show the efficacy of the proposed LMO-YOLO and compare it with state-of-the-art methods on the Gaofen (GF-1) satellite dataset. For some satellites, only low-resolution single-band image can be obtained. This article is proposed for such datasets, but they cannot be made public due to confidentiality. Therefore, all experiments in this study are performed using GF-1 single-channel data. The proposed method was implemented using the PyTorch deep learning framework and was trained on a workstation with an NVIDIA GeForce RTX 3090 GPU with 16-GB memory. In the rest of the manuscript, the datasets are introduced first. Next, we describe the experimental setup. After that, we show the experimental results and analysis.

A. Datasets

To evaluate the performance of the proposed method, we conducted experiments on wide-field-viewing (WFV) data and the panchromatic and multispectral sensor (PMS) data of the GF-1 satellite. The spatial resolutions of these two images are 16 and 8 m, respectively. Their spectral ranges are 0.45–0.52 and 0.77-0.89 μm. Specific parameters of the two datasets are shown in Table I. Most of the ships in these images are distributed in the South China Sea, USA (e.g., Norfolk Harbor, Pearl Harbor, and Naval Base San Diego), Russia (e.g., Murmansk), France (e.g., Toulon), and Italy (e.g., Taranto). Due to the large size

TABLE I
DATASET OVERVIEW

Satellite	Sensor	Resolution	Spectral information	Image size	Sample size	Total samples
GaoFen-1	WFV	16m	0.45-0.52μm	12000×13400	512×512	3800
	PMS	8m	0.77-0.89μm	4548×4500	512×512	3800

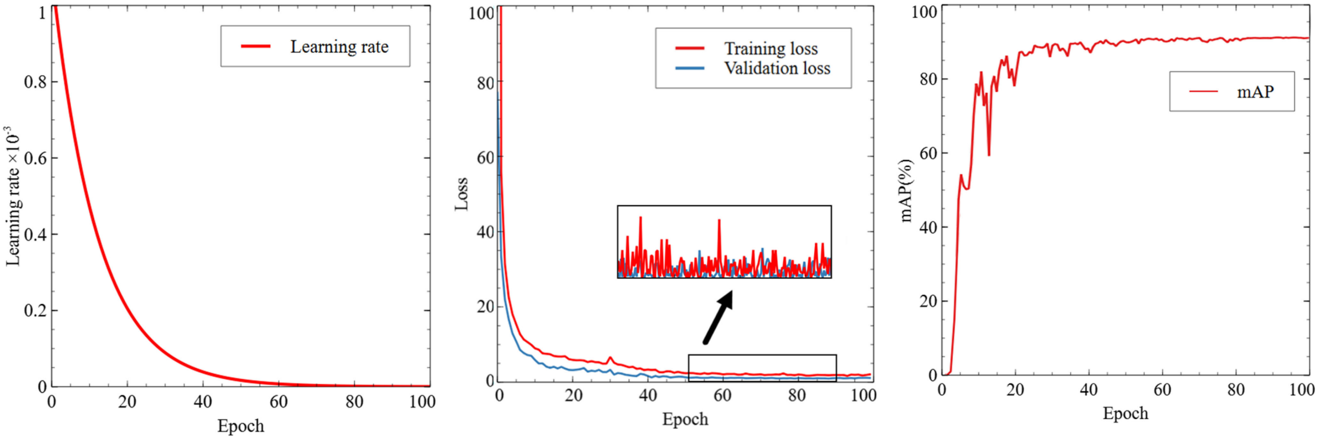

Fig. 7. Training information. (a) Learning rate versus epochs; (b) Training loss and validation loss; (c) mAP versus epochs.

of the original remote sensing image, we cut the images to 512 × 512 pixels, with 15% overlap. There is a total of 7600 slices, which constitute the entire dataset. The ratio of positive to negative sample images was 1:2. Furthermore, there are many types of ships in the training samples, and the size of the ship target varies from 8 pixels to 35 pixels.

Moreover, many image blocks contain few target samples due to the large amount of cloud and fog occlusion. Therefore, we used the simulation method for small object augmentation [51] to simulate samples. Ship samples of different types and sizes were cut out and randomly added to the background image. Then, the boundary was blurred by multisize Gaussian filters so that the target and the background were better blended. After that, we rescreened the obtained simulation samples and removed the images with poorly fused images to obtain the final dataset.

B. Implementation Details

To compare the proposed LMO-YOLO with state-of-the-art methods more fairly, the training hyperparameters were set to be the same as or similar to those of the comparison methods. Because the proposed method is a one-stage detector, the YOLOv4 is selected as the baseline of the comparison experiments. Furthermore, because the datasets were newly constructed, YOLOv4 was reimplemented based on our datasets and many improvements were added to the baseline. As a result, the final detection precision on the two datasets were 92.32% and 93.07%, respectively. The number of training epochs is 100. The IoU threshold was set to 0.2 to obtain better results because the object size was small. The confidence threshold was set to 0.3, and the NMS threshold was 0.5.

Additionally, the original YOLOv4 pretraining weights were utilized as the initial parameters. High learning rates lead to undesirable nonconvergence and smaller ones slow down the convergence speed. Therefore, we choose a dynamic change strategy for the learning rate [see Fig. 7(a)], as follows:

$$lr = lr_{(\text{last\_epoch})} * \lambda^n, n = \text{epoch}/\text{step\_size} \qquad (11)$$

where we set the original learning rate (lr) to 0.001, and λ was set to 0.92.

Fig. 7(b) gives the training and validation loss curves. On the one hand, we can observe that this network converges rapidly when training. On the other hand, the difference between training loss and verification loss is small. This shows that there is no overfitting phenomenon in the proposed model. Fig. 7(c) shows the mean average precision (mAP) versus epochs on the validation set. We can see that the mAP increases dramatically when the number of epochs is less than 20, and then slowly grows and stabilizes. Eventually, it reaches the best fit the dataset.

C. Evaluation Metrics

To quantitatively evaluate the ship detection performance of these methods, we chose accuracy evaluation indexes from remote sensing community ($\mathrm{P_d}$, $\mathrm{P_m}$, and $\mathrm{P_f}$) and deep learning community (precision, recall, and AP). However, to compute these indicators, the true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs) in the detection results need to be found first. Further, Intersection over Union (IoU) is required, which represents the overlap ratio between the prediction box S_p and ground truth box S_gt. It can be defined as

$$\text{IoU} = \left(S_p \cap S_{gt}\right) / \left(S_p \cup S_{gt}\right). \qquad (12)$$

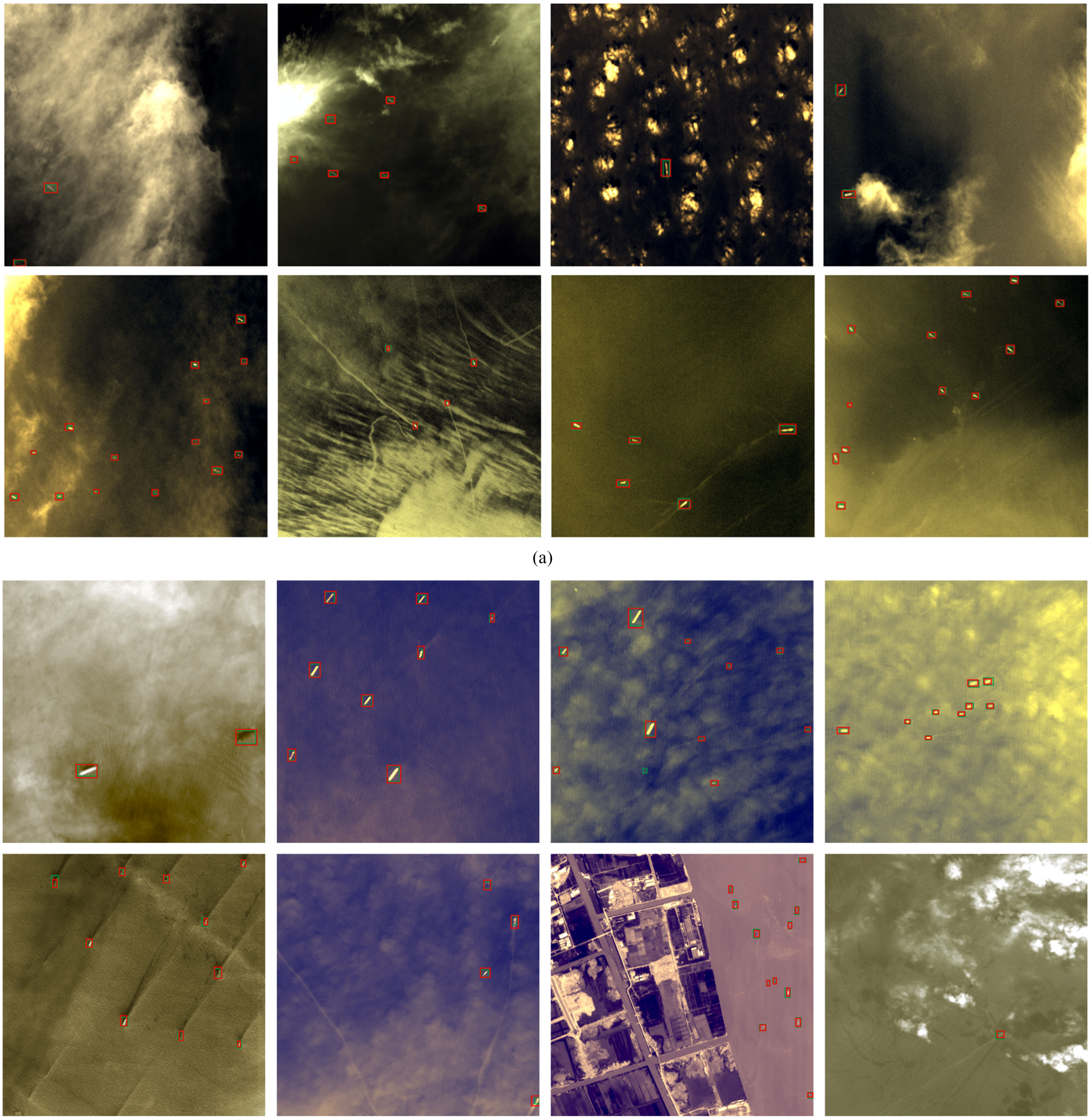

Fig. 8. Visualization Results of the proposed LMO-YOLO on the constructed two GF-1 datasets. The green boxes are groundtruth and red boxes are predicted results.

If IoU $>$ the setting threshold value, this predicted box is considered as true positive, otherwise it is considered as false positive. If no predicted box covers the target area, it is treated as a false negative. Otherwise, the region is a true negative.

Consequently, the detection probability ($\mathrm{P_d}$), missed-detection probability $\mathrm{P_m}$, and false alarm probability $\mathrm{P_f}$ are defined as

$$\mathrm{P_d} = TP\,/\,GT \tag{13}$$

$$\mathrm{P_m} = FN\,/\,GT \tag{14}$$

$$\mathrm{P_f} = FP\,/\,(TP + FP). \tag{15}$$

The precision and recall can be calculated as follows:

$$\mathrm{Precision} = TP\,/\,(TP + FP) \tag{16}$$

$$\mathrm{Recall} = TP\,/\,(TP + FN) \tag{17}$$

where, the GT is the number of true objects. It is not sufficient to evaluate the performance of the model only using above indexes. Another comprehensive indicator, average precision (AP) score,

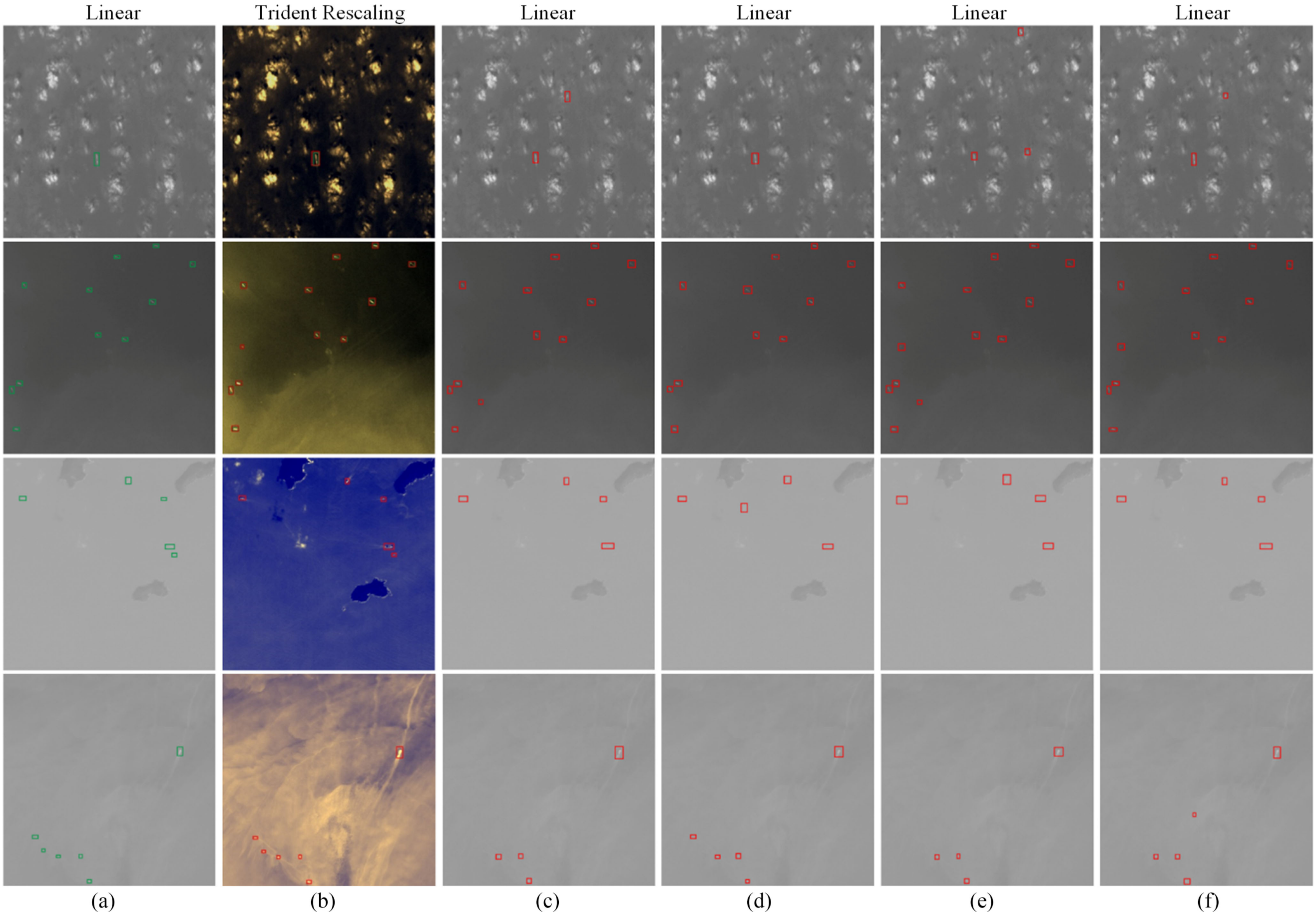

Fig. 9. Comparison results between the proposed method with trident rescaling and the compared methods with linear rescaling. The first two rows are images from the WFV dataset, and the rest are images from the PMS dataset.

is shown as follows:

$$\text{AP} = \int_0^1 P(R)\,dR \tag{18}$$

where $P(R)$ is the precision–recall (P–R) curve. It is the average value of precision for each object category when the recall varies from 0 to 1. Compared with other indexes, mAP reflects the performance of the detection model more accurately and intuitively. The model detection speed can be quantitatively evaluated using time (t) and frames per second (FPS)

$$\text{FPS} = 1\,/\,t. \tag{19}$$

D. Analysis of Different Scheme Settings

1) Analysis of the Trident Rescaling Scheme: This section focuses on verifying the superiority of the proposed quantization method. To illustrate the performance of the proposed quantization method, we conducted validation experiments. We compared the proposed quantization method with five other methods: linear, linear 1%, linear 2%, histogram equalization [52], and linear postcontrast enhancement [53], [54]. We processed the data using different quantization methods and fed them into the same baseline network (YOLOv4) to detect the ships. The detection networks shared the same parameter settings for a fair comparison, and the training samples were labelled identically. The precision, recall and AP were used as evaluation metrics. The results of the quantitative method validation are shown in Table III. The comparative results of the detection performance show the superiority of the method in this article. The proposed quantization method effectively preserves the texture detail features of the vessel while improving the contrast between the ships and the background. Moreover, the method in this article can effectively improve the recall rate and reduce the number of missed ships.

2) Analysis of Innovations: Except for the rescaling method, the other schemes proposed in this article have also been evaluated. As presented in Table II, the trident rescaling scheme (TRS), MDC, and contrast-sensitive loss function (CSL) have been verified for their effectiveness. "Selected modules" means that the method adds different schemes based on the baseline network (YOLOv4). For a method *n*, if there is a checkmark under the corresponding schemes in the next cell, it means that this scheme is included in this method. We can see that each proposed schemes can indeed improve the model detection performance. Although the most helpful operation for model performance improvement is the quantization (about 3% in AP), the MDC module and contrast-based loss strategy can also produce positive effects. Finally, when all three schemes

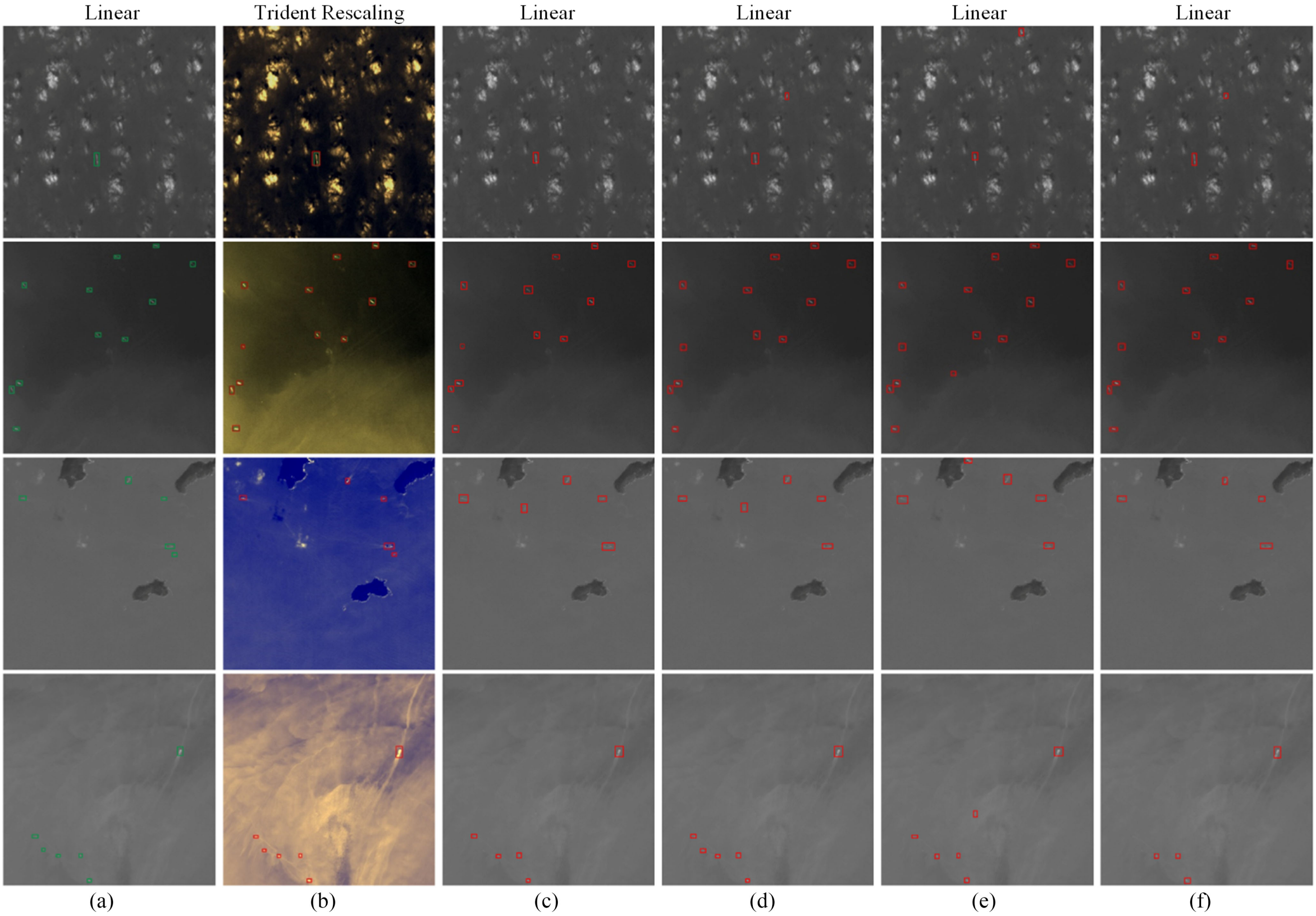

Fig. 10. Comparison results between the proposed method with trident rescaling and the compared methods with linear 1% rescaling. The first two rows are images from the WFV dataset, and the rest are images from the PMS dataset.

TABLE II
QUANTITATIVE EVALUATION OF INNOVATION IN THIS ARTICLE

Methods	Baseline	TRS	MDC	CSL	Recall	Precision	AP	Inference time
1	✓				86.22	92.77	87.69	13.10
2	✓	✓			89.12	94.06	91.14	13.10
3	✓		✓		86.76	93.58	89.27	13.65
4	✓			✓	87.03	92.95	89.41	13.37
5	✓	✓	✓	✓	90.53	94.93	92.32	14.69

Note: TRS: Trident rescaling scheme. MDC: Multiscale dilated convolution CSL: Contrast-sensitive loss. The units of the above accuracy indexes are all percentages (%), and the time is given in milliseconds (ms / image).

proposed in this study are included, the AP can be increased by 4.63% and the recall rate is improved by 4.31%. Although there is an increase in inference time, this is acceptable.

E. Detection Results and Comparison

To illustrate the detection performance of LMO-YOLO, we conducted comparative experiments on the constructed GF-1 datasets to compare the proposed method with state-of-the-art object detection methods. These comparison methods include not only one-stage methods, such as YOLOv4 [20], and SSD [55], and two-stage methods, such as RetinaNet [48] and Faster-RCNN [56] but also a detection model R^3-Net [45] for small objects. Moreover, FMSSD [57] and CSFF [58], which specifically designed for remote sensing object detetion, are also used for comparison. The comparative experiments are fair and extensive. For methods with open-source code, we use them directly for testing; and for popular detectors like faster R-CNN, the MMDetection project is utilized.

1) Experiments for LMO-YOLO: The visualization results of LMO-YOLO for the WFV and PMS datasets are given in Fig. 8, which includes image slices of different scenes containing many ships. The test experiment is completed based on the images after trident rescaling. However, due to the different gray scale

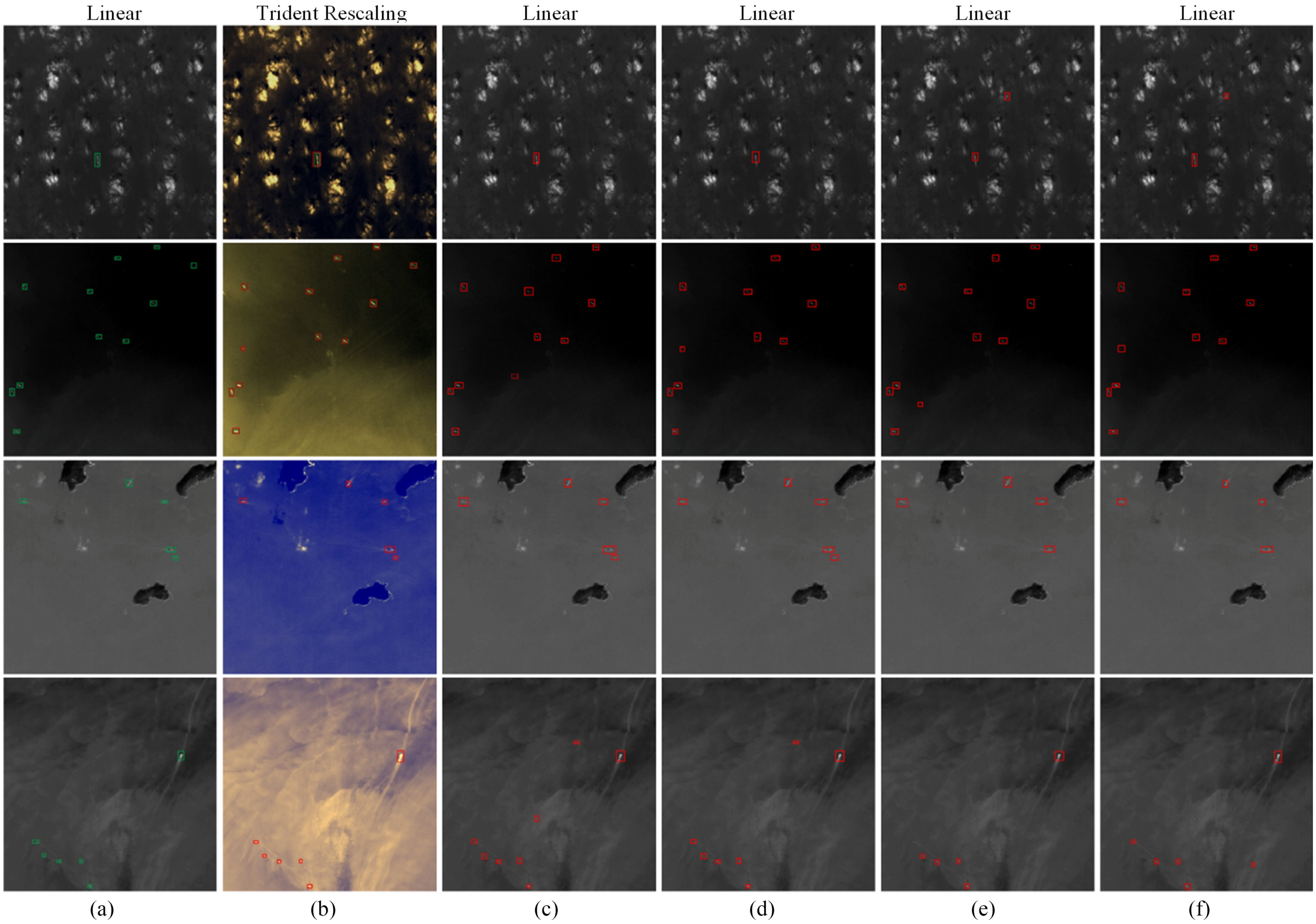

Fig. 11. Comparison results between the proposed method with trident rescaling and the compared methods with linear 2% rescaling. The first two rows are images from the WFV dataset, and the rest are images from the PMS dataset.

TABLE III
QUANTITATIVE EVALUATION OF QUANTIZATION METHOD

No.	Preprocess Method	Recall	Precision	AP
1	Linear	86.22	92.77	87.69
2	Linear 1%	87.10	92.94	87.85
3	Linear 2%	86.42	92.06	87.41
4	Histogram equalization	87.35	91.16	88.54
5	Linear + CE	86.43	93.02	88.75
6	Ours	89.16	93.97	91.28

Note: The units of the above accuracy indexes are all percentages (%).

ranges of the original image blocks, the false color images after trident rescaling presents a different color. No matter what the color is, the definition in the image blocks is higher, and the targets are more salient than in the original images. We can see that the size of the ship in our low-resolution WFV images is small, while the size of the ship in the PMS images varies greatly. Nonetheless, our approach can obtain good detection performance. In Tables IV and V, specific numbers also show that LMO-YOLO has the highest AP of the state-of-the-art detectors.

2) *Comparative Experiments on the WFV Dataset:* The results on the WFV dataset compared with other state-of-the art methods are shown in Table IV. The optimal results are shown in bold. The evaluation indexes demonstrate that the AP of LMO-YOLO reaches 92.32%. This AP is significantly higher than that of the other detectors investigated. Furthermore, the recall and precision of LMO-YOLO are also better. Our approach also performs well on remote sensing evaluation indicators. Compared with early classic detection algorithms, such as FasterRCNN and SSD, our approach produces an approximately 13% improvement in detection probability ($\mathrm{P_d}$). LMO-YOLO obtains the lower false alarm probability ($\mathrm{P_f}$) than the methods specifically designed for remote sensing object detection, like FMSSD. In comparison with the model ($\mathrm{R^3}$-Net) utilized detect small objects, the proposed method also outperforms it. Moreover, although its inference speed is not as fast as that of SSD, LMO-YOLO still performed well in the studies we investigated.

We applied the TRS to all methods and compared it with different single quantization methods (e.g., linear, linear 1%, and linear 2%) to verify its effectiveness. Table IV presents the detailed quantitative results. Among all the methods evaluated, the linear 1% rescaling scheme has a relatively better detection effect than other single rescaling methods. However, there is no significant difference between them. Nevertheless, the AP of the TRS is 2%–3% higher than any single rescaling method.

TABLE IV
EVALUATION INDEXES OF DIFFERENT METHODS BASED ON WFV DATASETS

Method	Rescaling scheme	Accuracy (Remote Sensing)			Accuracy (Deep learning)			Speed	
		P_d	P_m	P_f	Recall	Precision	AP	Time	FPS
FasterRCNN	Linear	76.87	23.13	15.22	76.87	84.78	81.31	43.49	22.99
	Linear 1%	77.30	22.70	15.05	77.30	84.95	81.62		
	Linear 2%	76.93	23.07	15.39	76.93	84.61	81.10		
	Trident Rescaling	78.79	21.21	13.93	78.79	86.07	83.42		
RetinaNet	Linear	79.44	20.56	14.19	79.44	85.81	84.05	47.17	21.20
	Linear 1%	79.70	20.30	13.95	79.70	86.05	84.13		
	Linear 2%	79.52	20.48	14.31	79.52	85.69	83.88		
	Trident Rescaling	81.88	18.12	12.09	81.88	87.91	86.01		
SSD	Linear	78.16	21.84	14.86	78.16	85.14	81.92	**24.42**	**40.95**
	Linear 1%	78.65	21.35	14.67	78.65	85.33	82.33		
	Linear 2%	78.42	21.58	14.92	78.42	85.08	81.63		
	Trident Rescaling	80.86	19.14	13.37	80.86	86.63	84.17		
FMSSD	Linear	82.84	17.16	11.66	82.84	88.34	86.02	33.17	30.15
	Linear 1%	83.24	16.76	11.48	83.24	88.52	86.12		
	Linear 2%	83.08	16.92	11.60	83.08	88.40	85.64		
	Trident Rescaling	84.77	15.23	10.03	84.77	89.97	87.98		
CSFF	Linear	84.10	15.90	9.67	84.10	90.33	86.81	49.55	20.18
	Linear 1%	84.84	15.16	9.34	84.84	90.66	87.00		
	Linear 2%	84.22	15.78	9.75	84.22	90.25	86.68		
	Trident Rescaling	86.38	13.62	8.13	86.38	91.87	88.64		
YOLOv4	Linear	86.22	13.78	7.23	86.22	92.77	87.69	26.39	37.89
	Linear 1%	87.10	12.90	7.06	87.10	92.94	87.85		
	Linear 2%	86.42	13.58	7.94	86.42	92.06	87.41		
	Trident Rescaling	89.16	10.84	6.03	89.16	94.97	91.28		
R^3-Net	Linear	84.55	15.45	8.32	84.55	91.68	86.74	22.20	45.05
	Linear 1%	85.61	14.39	8.94	85.61	91.06	87.11		
	Linear 2%	85.27	14.73	9.19	85.27	90.81	86.35		
	Trident Rescaling	87.56	12.44	7.56	87.56	92.44	89.06		
LMO-YOLO	Linear	87.33	12.67	6.43	87.33	93.57	88.92	28.61	34.95
	Linear 1%	87.82	12.18	5.98	87.82	94.02	89.57		
	Linear 2%	87.47	12.53	6.74	87.47	93.26	88.76		
	Trident Rescaling	**90.53**	**9.47**	**5.07**	**90.53**	**94.93**	**92.32**		

Note: Except for AP, which has no units, the units of the above accuracy indexes are all percentages (%), and the time is given in milliseconds (ms / image).

Figs. 9–11 also illustrate the comparison results between the LMO-YOLO with trident rescaling and other compared methods with linear *n%* rescaling. The first two rows in each figure are the images from WFV data. It is observed that the targets in trident rescaling images is clearer, and the image contrast is higher.

3) Comparative Experiments on the PMS Dataset: To further verify the effectiveness and versatility of LMO-YOLO for different dataset, comparative experiments on the PMS dataset with 8-m spatial resolution were also conducted. As shown in Table V, our approach has the highest AP of 93.07% and its P_d reaches 90.75%, which is higher than that of the other state-of-the-art detectors. Although the inference speed of LMO-YOLO is not the fastest, it is much faster than other two-stage methods and is acceptable. The analysis of the quantization method produces the same conclusion as the above experiment. The advantage of the TRS is quite obvious. Figs. 9–11 show the visual detection maps to demonstrate the results more intuitively. The last two rows in each figure show images from PMS dataset. Except for LMO-YOLO, the quantization scheme of the methods is the single linear *n%*. From these figures, we can see that the missed detection rate as well as false alarm rate of our approach is lower than those of the other methods. This also confirms that the quantization scheme proposed in this article is indeed important. We can found that our model does well in both datasets, which indicates that it has a high generalization ability.

F. Discussion

Thanks to the trident rescaling module, the 8-b images quantized from the original remote sensing data have higher quality. The constructed MDC module and contrast-sensitive loss can more accurately detect targets with low contrast against the background. We can also observe from Tables III–V that the approach proposed in this study can indeed improve the detection accuracy for small-sized objects in low-resolution imagery. Unfortunately, some targets that are covered by thick clouds cannot be detected, and some sea clutter or broken clouds may be recognized as false alarms. Object detection tasks will face different challenges in different scenes. In the follow-up research,

TABLE V
EVALUATION INDEXES OF DIFFERENT METHODS BASED ON PMS DATASETS

Method	Rescaling scheme	Accuracy (Remote Sensing)			Accuracy (Deep learning)			Speed	
		P_d	P_m	P_f	Recall	Precision	AP	Time	FPS
FasterRCNN	Linear	77.14	22.86	15.65	77.14	84.35	81.53	43.88	22.79
	Linear 1%	77.31	22.69	15.88	77.31	84.12	81.43		
	Linear 2%	77.72	22.28	15.38	77.72	84.62	82.06		
	Trident Rescaling	79.03	20.97	14.17	79.03	85.83	84.26		
RetinaNet	Linear	79.58	20.42	14.37	79.58	85.63	84.55	47.29	21.15
	Linear 1%	79.77	20.23	14.53	79.77	85.47	84.33		
	Linear 2%	80.03	19.97	13.98	80.03	86.02	84.81		
	Trident Rescaling	82.95	17.05	12.41	82.95	87.59	86.23		
SSD	Linear	78.04	21.96	14.74	78.04	85.26	82.92	**24.36**	**41.05**
	Linear 1%	78.54	21.46	14.85	78.54	85.15	82.48		
	Linear 2%	78.92	21.08	14.58	78.92	85.42	83.26		
	Trident Rescaling	81.45	18.55	13.41	81.45	86.59	85.25		
FMSSD	Linear	82.98	17.02	11.66	82.98	88.34	86.72	33.27	30.06
	Linear 1%	83.27	16.73	12.01	83.27	87.99	86.53		
	Linear 2%	83.48	16.52	11.52	83.48	88.48	86.94		
	Trident Rescaling	85.22	14.78	10.05	85.22	89.95	88.81		
CSFF	Linear	84.35	15.65	9.66	84.35	90.34	87.46	50.31	19.88
	Linear 1%	84.58	15.42	9.88	84.58	90.12	87.26		
	Linear 2%	85.03	14.97	9.25	85.03	90.75	87.65		
	Trident Rescaling	86.64	13.36	8.12	86.64	91.88	89.41		
YOLOv4	Linear	86.43	13.57	7.82	86.43	92.18	88.24	26.54	37.68
	Linear 1%	86.61	13.39	8.08	86.61	91.92	88.02		
	Linear 2%	87.03	12.97	7.23	87.03	92.77	88.53		
	Trident Rescaling	89.25	10.75	5.87	89.25	94.13	91.69		
R^3-Net	Linear	85.23	14.77	8.74	85.23	91.26	86.98	22.13	45.19
	Linear 1%	85.78	14.22	9.13	85.78	90.87	86.36		
	Linear 2%	86.02	13.98	7.98	86.02	92.02	87.45		
	Trident Rescaling	88.77	11.23	6.98	88.77	93.02	90.11		
LMO-YOLO	Linear	87.16	12.84	6.97	87.16	93.03	90.05	29.26	34.18
	Linear 1%	87.48	12.52	7.12	87.48	92.88	89.87		
	Linear 2%	87.81	12.19	6.68	87.81	93.32	90.32		
	Trident Rescaling	**90.75**	**9.25**	**5.07**	**90.75**	**94.93**	**93.07**		

Note: Except for AP, which has no units, the units of the above accuracy indexes are all percentages (%), and the time is given in milliseconds (ms / image).

we will conduct more in-depth exploratory research from the view of complex scenes perception. A faster and lighter detector for embedded devices is another work we intend to study.

V. CONCLUSION

In this study, a novel ship detection model, LMO-YOLO, for low-resolution optical satellite imagery is proposed. First, a trident linear rescaling scheme was developed to quantize the original satellite images, so the obtained 8-b images can contain more detailed information. Second, we kept the spatial resolution of the last few stages of the backbone and added dilated convolution to expand the receptive field and extract more object and object-background features. Finally, to balance the easy-to-detect and hard-to-detect samples in terms of contrast, an adaptive weighting scheme was designed. The experiments demonstrated the following: 1) The rescaling module retained the 8-b images with more useful features; 2) the optimized backbone effectively prevented the weakening of small target information and learned much useful information; and 3) the contrast-sensitive loss scheme improved the robustness of the network to variation in object-background contrast. The experimental results show that the proposed LMO-YOLO outperformed other state-of-the-art methods.

REFERENCES

[1] Y. Zhuang, L. Li, and H. Chen, “Small sample set inshore ship detection from VHR optical remote sensing images based on structured sparse representation,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2145–2160, 2020.

[2] Z. Zhang, L. Zhang, Y. Wang, P. Feng, and R. He, “Shiprsimagenet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8458–8472, 2021.

[3] P. Qin, Y. Cai, J. Liu, P. Fan, and M. Sun, “Multilayer feature extraction network for military ship detection from high-resolution optical remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11 058–11 069, 2021.

[4] D. Poli, E. Angiuli, and F. Remondino, “Radiometric and geometric analysis of worldview-2 stereo scenes,” in *Proc. Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, 2010, pp. 15–18.

[5] T. Salehi and M. H. Tangestani, “Large-scale mapping of iron oxide and hydroxide minerals of zefreh porphyry copper deposit, using worldview-3 vnir data in the northeastern Isfahan, Iran,” *Int J. Appl. Earth Observation Geoinformation*, vol. 73, pp. 156–169, 2018.

[6] K. S. Krause, “Relative radiometric characterization and performance of the quickbird high-resolution commercial imaging satellite,” in *Proc SPIE*, vol. 5542, 2004, pp. 35–44.

[7] R. Bhatt, D. Doelling, C. Haney, B. Scarino, and A. Gopalan, “Consideration of radiometric quantization error in satellite sensor cross-calibration,” *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1131.

[8] Z. Liu, H. Wang, L. Weng, and Y. Yang, “Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2018.

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

[10] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, “Object detection in optical remote sensing images: A survey and a new benchmark,” *ISPRS-J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, 2020.

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

[12] D. Konstantinidis, T. Stathaki, V. Argyriou, and N. Grammalidis, “Building detection using enhanced hog–LBP features and region refinement processes,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 888–905, Mar. 2017.

[13] Y. Sun, Z. Wang, X. Sun, and K. Fu, “Span: Strong scattering point aware network for ship detection and classification in large-scale SAR imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6083–6101, 2022.

[14] Z. Hong *et al.*, “Multi-scale ship detection from SAR and optical imagery via a more accurate yolov3,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6083–6101, 2021.

[15] B. Du, Y. Wang, C. Wu, and L. Zhang, “Unsupervised scene change detection via latent dirichlet allocation and multivariate alteration detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4676–4689, 2018.

[16] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, “Ship classification in SAR images using a new hybrid CNN–MLP classifier,” *J. Indian Soc. Remote Sens.*, vol. 47, no. 4, pp. 551–562, 2019.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[18] J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.

[19] A. Farhadi and J. Redmon, “Yolov3: An incremental improvement,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1804–1804.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 2020, *arXiv:2004.10934*.

[21] Z. Zakria, J. Deng, R. Kumar, M. S. Khokhar, J. Cai, and J. Kumar, “Multi scale and direction target detecting in remote sensing images via modified yolo-v4,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1039–1048, 2022.

[22] P. Lemenkova, “A technical approach of image segmentation in ENVI GIS to identify thematic clusters for visualization of urban transformations,” in *Proc. Conf. Proc. Reality Sum Inf. Technol.*, 2015, pp. 100–104.

[23] J. Mootz and L. Mathews, “Displaying and stretching 16-bit per band digital imagery,” in *USDA FSA Aerial Photogr. Field Office*.

[24] C. Hu, X. Li, and W. G. Pichel, “Detection of oil slicks using modis and SAR imagery,” in *Handbook of Satellite Remote Sensing Image Interpretation: Applications for Marine Living Resources Conservation and Management*, IOCCG, Dartmouth, NS, Canada, 2011, pp. 21–34.

[25] A. T. Manual, “ALOS PALSAR 50-meter mosaic step-by-step processing manual on forest cover classification,” 2013.

[26] T. Arici, S. Dikbas, and Y. Altunbasak, “A histogram modification framework and its application for image contrast enhancement,” *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, 2009.

[27] J. H. Jang, S. D. Kim, and J. B. Ra, “Enhancement of optical remote sensing images by subband-decomposed multiscale retinex with hybrid intensity transfer function,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 983–987, 2011.

[28] T. Celik, “Two-dimensional histogram equalization and contrast enhancement,” *Pattern Recognit.*, vol. 45, no. 10, pp. 3810–3824, 2012.

[29] S. Wang, K. Gu, S. Ma, W. Lin, X. Liu, and W. Gao, “Guided image contrast enhancement based on retrieved images in cloud,” *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 219–232, 2015.

[30] J. Liu, C. Zhou, P. Chen, and C. Kang, “An efficient contrast enhancement method for remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1715–1719, 2017.

[31] R. Zhang, J. Yao, K. Zhang, C. Feng, and J. Zhang, “S-CNN-based ship detection from high-resolution remote sensing images,” *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 3–22, 2016.

[32] X. Li, S. Wang, B. Jiang, and X. Chan, “Inshore ship detection in remote sensing images based on deep features,” in *Proc. IEEE Int. Conf. Signal Process. Commun. Comput.*, 2017, pp. 1–5.

[33] 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, 2016.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[35] 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, 2017.

[36] 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, 2018.

[37] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, “PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery,” *ISPRS-J. Photogramm. Remote Sens.*, vol. 173, pp. 50–65, 2021.

[38] Y. Yu *et al.*, “Sparse anchoring guided high-resolution capsule network for geospatial object detection from remote sensing imagery,” *Int. J. Appl. Earth Observation Geoinf.*, vol. 104, 2021, Art. no. 102548.

[39] Z. Shi, X. Yu, Z. Jiang, and B. Li, “Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, 2013.

[40] Z. Zou and Z. Shi, “Ship detection in spaceborne optical image with SVD networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, 2016.

[41] Z. Zou and Z. Shi,, “Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1100–1111, 2017.

[42] 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, 2018.

[43] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, “Perceptual generative adversarial networks for small object detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1222–1230.

[44] Z. Liu, G. Gao, L. Sun, and L. Fang, “IPG-Net: Image pyramid guidance network for small object detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, 2020, pp. 1026–1027.

[45] Q. Li, L. Mou, Q. Xu, Y. Zhang, and X. X. Zhu, “R^{3-}net: A deep network for multioriented vehicle detection in aerial images and videos,” *IEEE Trans. Geosci. Remote Sens.*, pp. 5028–5042, 2019.

[46] C. Deng, M. Wang, L. Liu, Y. Liu, and Y. Jiang, “Extended feature pyramid network for small object detection,” *IEEE Trans. Multimedia*, vol. 24, pp. 1968–1979, 2021.

[47] A. Shrivastava, A. Gupta, and R. Girshick, “Training region-based object detectors with online hard example mining,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 761–769.

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

[49] B. Li, Y. Liu, and X. Wang, “Gradient harmonized single-stage detector,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 8577–8584.

[50] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, “Libra R-CNN: Towards balanced learning for object detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 821–830.

[51] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, “Augmentation for small object detection,” 2019, *arXiv:1902.07296*.

[52] Y.-T. Kim, “Contrast enhancement using brightness preserving bi-histogram equalization,” *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, 1997.

[53] E. Lee, S. Kim, W. Kang, D. Seo, and J. Paik, “Contrast enhancement using dominant brightness level analysis and adaptive intensity transformation for remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 62–66, 2012.

[54] Y. Wang, Q. Chen, and B. Zhang, “Image enhancement based on equal area dualistic sub-image histogram equalization method,” *IEEE Trans. Consum. Electron.*, vol. 45, no. 1, pp. 68–75, 1999.

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

[56] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[57] P. Wang, X. Sun, W. Diao, and K. Fu, "FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3377–3390, 2019.

[58] G. Cheng, Y. Si, H. Hong, X. Yao, and L. Guo, "Cross-scale feature fusion for object detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 431–435, 2020.

Qizhi Xu (Member, IEEE) received the B.S. degree from Jiangxi Normal University, Nanchang, China, in 2005, and the Ph.D. degree from Beihang University, Beijing, China, in 2012.

He was a Postdoctoral Fellow with the University of New Brunswick, Fredericton, NB, Canada. He is currently an Associate Professor with the School Of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China. His research interests include image fusion, image understanding, and big data analysis of remote sensing.

Dr. Xu was the recipient of the Technological Invention Award First Prize from the Chinese Institute of Electronics for his image fusion research, in 2017.

Yuan Li received the B.S. degree in College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, in 2018, and the M.S. degree, in 2021. She is currently working toward the Ph.D. degree with the School of Mechatronical Engineering, Beijing Institute of Technology.

Her research interests include remote sensing image process and pattern recognition.

Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005.

From 2005 to 2007, he was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing, China. He has authored or coauthored more than 100 scientific articles in related journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi was the recipient of the best reviewer awards for his service to IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, in 2017. He has been an Associate Editor for *Infrared Physics and Technology*, since 2016.