



Article

AF-OSD: An Anchor-Free Oriented Ship Detector Based on Multi-Scale Dense-Point Rotation Gaussian Heatmap

Zizheng Hua ¹, Gaofeng Pan ², Kun Gao ^{1,*}, Hengchao Li ³ and Su Chen ¹¹ School of Optics and Photonic, Beijing Institute of Technology, Beijing 100811, China² School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100811, China³ School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610032, China

* Correspondence: gaokun@bit.edu.cn; Tel.: +86-139-1101-2489

Abstract: Due to the complexity of airborne remote sensing scenes, strong background and noise interference, positive and negative sample imbalance, and multiple ship scales, ship detection is a critical and challenging task in remote sensing. This work proposes an end-to-end anchor-free oriented ship detector (AF-OSD) framework based on a multi-scale dense-point rotation Gaussian heatmap (MDP-RGH) to tackle these aforementioned challenges. First, to solve the sample imbalance problem and suppress the interference of negative samples such as background and noise, the oriented ship is modeled via the proposed MDP-RGH according to its shape and direction to generate ship labels with more accurate information, while the imbalance between positive and negative samples is adaptively learned for the ships with different scales. Then, the AF-OSD based on MDP-RGH is further devised to detect the multi-scale oriented ship, which is the accurate identification and information extraction for multi-scale vessels. Finally, a multi-task object size adaptive loss function is designed to guide the training process, improving its detection quality and performance for multi-scale oriented ships. Simulation results show that extensive experiments on HRSC2016 and DOTA ship datasets reveal that the proposed method achieves significantly outperforms the compared state-of-the-art methods.

Keywords: anchor-free; arbitrarily oriented ship detection; deep learning; Gaussian heatmap; multi-scale



Citation: Hua, Z.; Pan, G.; Gao, K.; Li, H.; Chen, S. AF-OSD: An Anchor-Free Oriented Ship Detector Based on Multi-Scale Dense-Point Rotation Gaussian Heatmap. *Remote Sens.* **2023**, *15*, 1120. <https://doi.org/10.3390/rs15041120>

Academic Editors: Moulay A. Akhloufi and Mozhddeh Shahbazi

Received: 23 December 2022

Revised: 15 February 2023

Accepted: 16 February 2023

Published: 18 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ship detection is a highly regarded topic in remote sensing. The precise identification of ships through remote sensing images is a crucial task in the field of target recognition. Typically, ships are located offshore or near coasts, leading to a lot of similarities in the imagery. Despite this, accurately determining the size, position, and orientation of ships remains a daunting challenge due to the intricate nature of remote sensing scenarios and the varying sizes of ships. With the development of imaging hardware, remote sensing images have higher resolution, so the method of ship detection based on remote sensing images has been widely studied in various fields of marine supervision [1–3], port traffic flow [4,5], ship reconnaissance and statistics [6,7], etc.

For traditional ship detection methods [1,8], the common problems are poor system robustness, high leakage rate in complex scenarios, and low detection accuracy. With the development of deep learning in various applications, deep learning-based methods [9–11] can provide automated, high-precision, and high-accuracy results for target detection from remote sensing images. Due to the specificity of remote sensing, there are still some open problems: (1) the bounding box may contain much background in the selected area, and cannot accurately represent the position and direction of ships [12]; (2) false detection because of the small and densely distributed ships [6]; (3) difficulty of recognizing multi-scale ships [13]. Aiming at these problems, some anchor-based methods have been developed [2,13,14]. However, to ensure satisfactory detection accuracy, these methods

usually require the manual design of pre-selected boxes according to the actual scene, hindering their applications in practice because of a large number of hyper-parameters and high computational complexity.

To solve the problems of anchor-based detection methods, anchor-free methods have been further proposed [6,15–17], which do not require the anchor parameters and can directly predict the class and location information of objects. Thus, these methods can avoid hyper-parameters and reduce computational processes, gradually attracting the extensive attention of researchers. The CenterNet models [6,9,16] have been built based on the center point of the object as a positive sample point to represent an oriented object, as shown in Figure 1a. However, if a target is represented by only one point, or if the shape characteristics of the oriented target are not considered. These methods fail to balance the number of positive and negative samples well, which also causes some positive locations to be misjudged as negative samples.

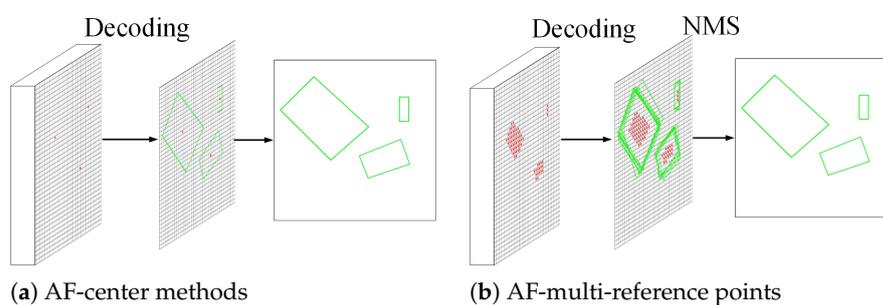


Figure 1. Comparison of target detection algorithms.

Aiming at the misjudgment problem of balancing the number of positive and negative samples, dense points-based methods are further proposed to increase the number of positive samples to achieve better detection performance. Zinelli et al. [18] proposed a dense point method to represent and predict the oriented objects, thus optimizing the imbalance of positive and negative samples to some extent. Nevertheless, the number of dense points is different with different scales of objects, as shown in Figure 1b. Specifically, the small objects usually contain fewer dense points than the large ones, and the contribution of the small objects to the loss function is small and easily ignored by the optimizer during training. Moreover, different locations of positive sample points exhibit different effects on the prediction results of the same object. The richer the object features extracted from the sample points close to the object center, the greater the impact on the prediction. However, these methods can not suppress background and noise interference well, which may result in the same confidence level for the feature points at the target centroid and edges, thus generating worse prediction results after non-maximum suppression (NMS), as shown in Figure 2. Additionally, these methods do not take the shape of the target into account, which may cause some negative locations to be misallocated as positive samples, such as the two ends of a ship.

To tackle the misallocation problem, some methods shrink the object bounding box to obtain the core region and take the core region as a positive location [18,19], while the other regions of the object bounding box are the transition from the positive location to the negative location. In this way, the problem of mislabeling can be alleviated to a certain extent. However, it fails to reflect the shape and direction characteristics of the oriented ships. Gaussian rotation heatmaps [6,20] have been used as the supervisory information to distinguish the positive and negative positions of oriented ships. Ref. [20] proposed a target detection method by adopting a Gaussian rotation heatmap as prior information and an adaptive weight-adjustment mechanism (OWAM) algorithm to weight the positive and negative samples at different positions. While these methods address the issue of mislabeling and incorporate the shape and direction of oriented ships, they only assign positive and negative labels through a continuous two-dimensional function represented by a Gaussian heatmap. This approach has a crucial limitation, as the Gaussian heatmap

cannot be utilized as the confidence output in the prediction stage. Additionally, the contribution of positive samples from different locations to the object is neglected, leading to the possibility of positive samples at the edge of the target location being overlooked and increasing the impact of noise on accurate target detection.

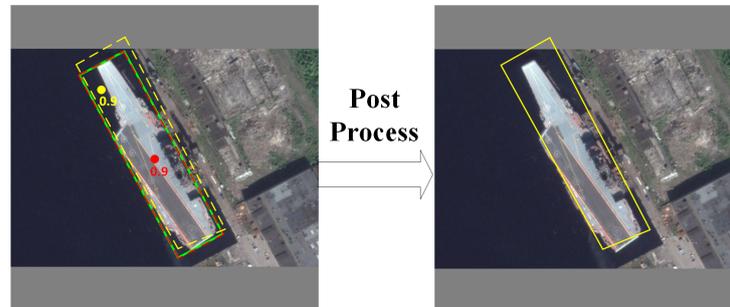


Figure 2. Diagram of prediction results of different reference points of the same rotation ship.

We aim to solve the imbalance problem caused by the number of positive and negative samples, better suppress background and noise interference, and simultaneously improve the robustness for remote sensing multi-scale ships and the training model's capability. Therefore, the multi-scale dense-point rotation Gaussian heatmap (MDP-RGH) method is proposed to overcome background and noise interference while enhancing the robustness of multi-scale ship detection and the training model's capabilities. The MDP-RGH is a discrete two-dimensional function based on a Gaussian heatmap that operates on dense points, allowing for the modeling of multi-scale oriented ships based on their shape and direction. The positive samples are weighted using the MDP-RGH to ensure rotational Gaussian distribution. The Gaussian heatmap confidence is used to predict whether a target is a positive or negative sample, thereby improving detection and reducing network computation. After that, a new anchor-free oriented ship detector network (AF-OSD) is constructed using the MDP-RGH method to detect multi-scale oriented ships. Additionally, a multi-task object size adaptive loss (OSALoss) function is designed to address the training imbalance caused by varying ship sizes. The weight of this function is determined by both the object area and the density of dense points, leading to improved ship detection accuracy. The contributions of this work can be summarized as follows.

1. An oriented ship model based on MDP-RGH is proposed, which can balance the number of positive and negative samples, suppress the interference of negative samples such as background and noise in the image, and improve the training accuracy.
2. An AF-OSD based on MDP-RGH is designed to achieve a better prediction for oriented ships with multi-scale attributes.
3. A multi-task OSALoss function is constructed to further overcome the training imbalance problem caused by different ship sizes to improve the detection quality and performance of the whole model for multi-scale ships.

The rest of the paper is organized as follows. In Section 2, the ship model based on MDP-RGH is elaborated in detail. Section 3 describes the AF-OSD based on the MDP-RGH ship model, which shows the detail of the network. Section 4 reports the hyper-parameter settings. Section 5 shows the experimental results and analysis. Section 6 discusses the experiment. Finally, Section 7 is the conclusion.

The variables themselves indicate a certain class of meaning, e.g., (x, y) for coordinate variables, **G** for ground truth, **F** for output convolutional features, and **M** for masks. The subscript of a variable indicates the qualification of the variable, indicating that the variable belongs to the range indicated by the subscript, and the next-level subscript is the qualification of the previous-level subscript.

2. The Oriented Ship Model Based on MDP-RGH

To solve the problem of an unbalanced number of positive and negative samples, predict the target using Gaussian heatmap confidence, and account for the contribution of positive samples at different locations, a new model of the ship, MDP-RGH, is proposed to suppress the background and noise interference, and effectively describe the characteristics of the shape and direction of the multi-scale ship. Specifically, there are three steps to complete the model of an oriented ship: (1) dividing the image region; (2) obtaining multi-scale dense points by down-sampling the image; (3) weighting the dense points with a rotation Gaussian Heatmap.

To address the issue of the imbalanced number of positive and negative samples, a new ship model called MDP-RGH is proposed. This model incorporates Gaussian heatmap confidence in target prediction and considers the contribution of positive samples from different locations. MDP-RGH aims to reduce background and noise interference and effectively describe the shape and direction of multi-scale ships, as shown in Figure 3. The process of creating the oriented ship model involves three steps: (1) dividing the image region; (2) obtaining multi-scale dense points through down-sampling the image; (3) weighting the dense points with a rotation Gaussian heatmap.

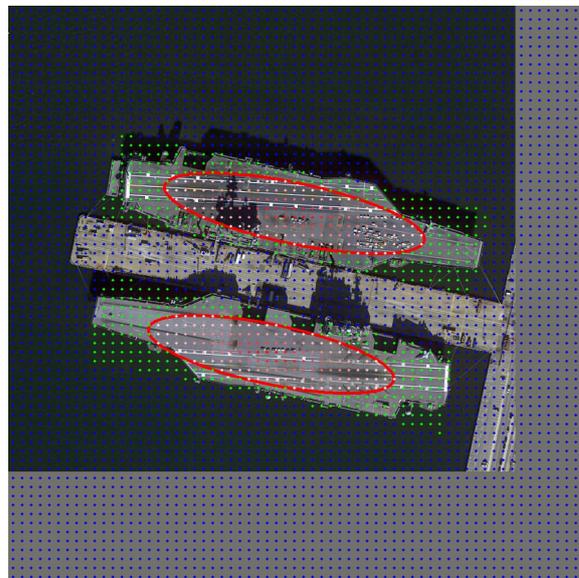


Figure 3. Ship representation format based on a dense-point Gaussian heatmap. The target in the red circle are delineated as the object regions.

2.1. Dividing the Image Region

As the shape of the ship is similar to a shuttle [6], if all the areas in the object bounding box are divided into object regions, part of the background pixels will be included in the object region.

To fit the shape and direction characteristics of the ship and coordinate with the subsequent rotation Gaussian heatmap, an image region division method is designed, as shown in Figure 3. Specifically, the shrink rotation ellipse regions are delineated as the object regions. Other regions in the object-bounding box are regarded as ignorable regions, while the regions not in any object-bounding box are regarded as the background regions. In this case, we only need to determine the object regions and the ignorable regions. More specifically, here are three steps to divide the image region.

2.1.1. Transformation the Coordinate for the Oriented Ship

To facilitate the subsequent calculation, we transform the rotating ship into an upright ship. Therefore, as shown in Figure 4, coordinate transformation is performed. Specifically,

we transform the points $\begin{bmatrix} x_{\text{pixel}} \\ y_{\text{pixel}} \end{bmatrix}$ in the original pixel coordinate system XOY to the new coordinate system $\text{X}'\text{O}'\text{Y}'$ with the center point of the oriented ship being set as the origin, the long axis of the ship (the line between the center points of the bow and stern) is set as the Y' axis, and the short axis is set as the X' axis. The coordinates of pixel points in the coordinate system $\text{X}'\text{O}'\text{Y}'$ can be expressed as

$$\begin{bmatrix} x'_{\text{pixel}} \\ y'_{\text{pixel}} \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \left(\begin{bmatrix} x_{\text{pixel}} \\ y_{\text{pixel}} \end{bmatrix} - \begin{bmatrix} x_c \\ y_c \end{bmatrix} \right), \tag{1}$$

where $\begin{bmatrix} x_c \\ y_c \end{bmatrix}$ means the coordinates of the center point of the oriented ship in the coordinate system XOY . α represents the counterclockwise angle between the positive half-axis of the oriented ship (the direction from the center point of the ship to the center point of the bow is positive) and the positive Y axis of the pixel coordinate system XOY .

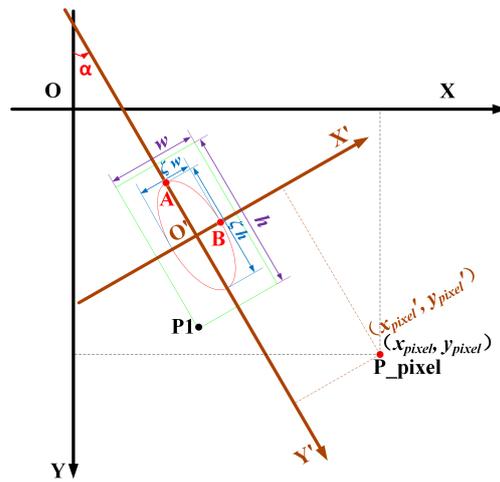


Figure 4. Coordinate transformation diagram. The green rectangular box is the boundary box of the ship. The red ellipse is a rotation ellipse with ζh as its long axis and ζw as its short axis after scaling the boundary frame of a ship according to the scale factor $\zeta \in (0, 1]$, where h and w are the width and height of the boundary frame of a ship. A and B are the two vertices of the scaled rotation ellipse.

2.1.2. Creating a Shrink Rotation Ellipse Equation for the Oriented Ship

We create a shrink rotate ellipse equation for the oriented ship to determine which region the pixels belong to. The ellipse of the shrink rotation ship bounding box in the coordinate system $\text{X}'\text{O}'\text{Y}'$ is a standard ellipse. Specifically, the ellipse equation is written as

$$f_{\text{ellipse}} \left(\begin{bmatrix} x' \\ y' \end{bmatrix} \right) = \frac{(x')^2}{\left(\frac{\zeta w}{2}\right)^2} + \frac{(y')^2}{\left(\frac{\zeta h}{2}\right)^2} = 1. \tag{2}$$

2.1.3. Identifying the Region to Which the Pixels in the Image Belong

After the above two steps, we can determine whether the pixel point $\begin{bmatrix} x_{\text{pixel}} \\ y_{\text{pixel}} \end{bmatrix}$ belongs to the object region $\text{area}_{\text{ship}}$ or the ignorable region $\text{area}_{\text{ignore}}$ as

$$\begin{cases} \begin{bmatrix} x_{\text{pixel}} \\ y_{\text{pixel}} \end{bmatrix} \in \text{area}_{\text{ship}}, & f_{\text{ellipse}} \left(\begin{bmatrix} x'_{\text{pixel}} \\ y'_{\text{pixel}} \end{bmatrix} \right) \leq 1 \text{ and } |x'_{\text{pixel}}| \leq \frac{w}{2} \\ \begin{bmatrix} x_{\text{pixel}} \\ y_{\text{pixel}} \end{bmatrix} \in \text{area}_{\text{ignore}}, & f_{\text{ellipse}} \left(\begin{bmatrix} x'_{\text{pixel}} \\ y'_{\text{pixel}} \end{bmatrix} \right) > 1 \text{ and } |y'_{\text{pixel}}| \leq \frac{h}{2}, |x'_{\text{pixel}}| \leq \frac{w}{2}. \end{cases} \quad (3)$$

Particularly, a pixel point belongs to the background region when it does not belong to any ship's object or ignorable region.

2.2. Multi-Scale Dense Points by Down-Sampling

Considering the multi-scale attributes of different ships, as shown in Figure 5, a multi-scale dense-point sampling method is further proposed to balance the number of positive samples of ships of different sizes. In particular, dense points can also reduce the amount of model computation since not all points need to be involved.

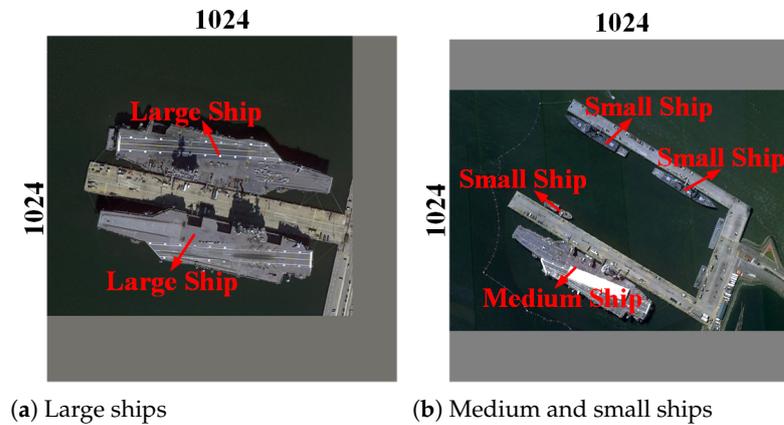


Figure 5. Some examples in the DOTA dataset.

Specifically, according to the different scales of the ships, the image is down-sampled at different multiples to obtain different low-resolution images. We conduct s ($s = 4, 8, 16$) down-samplings of the image with three scales. Then, the low-resolution image is mapped to the original image to obtain a three-scale dense point matrix. The coordinate calculation formula of dense points can be described as

$$\begin{cases} x_{\text{denps}_{ij}} = (i-1)(s) + \frac{s}{2} \\ y_{\text{denps}_{ij}} = (j-1)(s) + \frac{s}{2}, \end{cases} \quad (4)$$

where $x_{\text{denps}_{ij}}$ and $y_{\text{denps}_{ij}}$ denote the horizontal and vertical coordinates of the j -th row and i -th column sample point denps_{ij} in the original image in the dense point matrix after s samplings, respectively.

Thus, small-scale dense points are used to represent large ships and large-scale dense points are used to represent small ships, which could balance the number of positive samples of the large, medium, and small ships to a certain extent. Among them, large-scale, medium-scale, and small-scale ships can be clustered with the k -means clustering algorithm. To better delineate the effective area of the image, dense points in the background region are defined as negative sample points, dense points in the ignorable region are ignorable dense points, and dense points in the object region are positive sample points, as shown in Figure 6a–c, respectively. Particularly, the ignorable dense points do not participate in calculating the loss function when training the model.

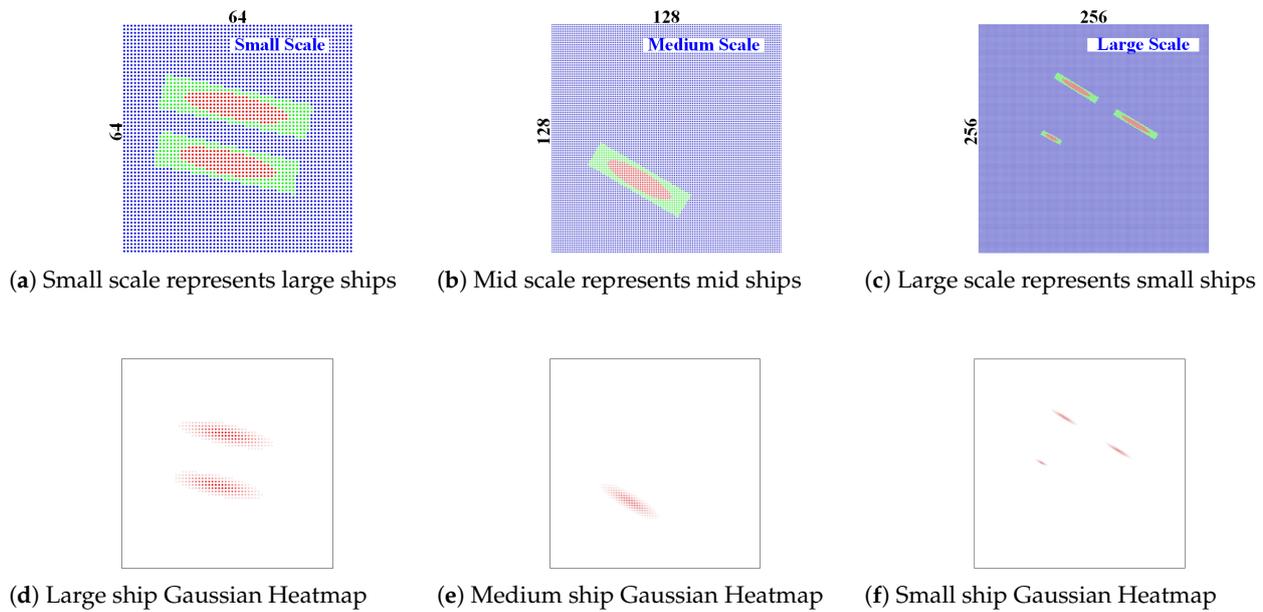


Figure 6. Schematic diagram of multi-scale dense-point sampling, in which blue points represent negative sample points, green points represent ignorable sample points, and red points represent positive sample points. Corresponding to (a–c) are two-dimensional rotation Gaussian heatmaps of large, medium, and small-sized ships, where the initial Gaussian heatmap value $g_{init} = 0.3$ and the scaling factor $\zeta = 0.7$.

2.3. Weighting the Dense Points with Rotation Gaussian Heatmap

It has been proved that the closer the sample point is to the center of the ship, the richer the features of the ship are extracted during model inference, obtaining more accurate detection results. Therefore, the Gaussian weighting on the dense points with the object region is performed, where the dense points with different positions mean the various degrees of importance of the ship. Furthermore, the multi-scale dense-point rotation Gaussian heatmap is obtained, as shown in Figure 6.

Specifically, the calculation method of the rotation Gaussian heatmap value of the dense points in the ship region at each scale can be represented as

$$g\left(x_{denp_{s_{ij}}}, y_{denp_{s_{ij}}}\right) = g'\left(x'_{denp_{s_{ij}}}, y'_{denp_{s_{ij}}}\right) = \text{Exp}\left(-\left(\frac{\left(x'_{denp_{s_{ij}}}\right)^2}{2\sigma_w^2} + \frac{\left(y'_{denp_{s_{ij}}}\right)^2}{2\sigma_h^2}\right)\right), \quad (5)$$

where $g(\cdot)$ means the rotation Gaussian heatmap function; $g'(\cdot)$ means the general Gaussian heatmap function; $x'_{denp_{s_{ij}}}$ and $y'_{denp_{s_{ij}}}$ denote the horizontal and vertical coordinates of the dense point in the coordinate system $\mathbf{X}'\mathbf{O}'\mathbf{Y}'$, respectively; $\text{Exp}(\cdot)$ is the exponential function; σ_w and σ_h are the parameters related to the width and height of the ship, respectively.

To determine the values of σ_w and σ_h , a hyper-parameter $g_{init} \in (0, 1)$ is introduced, which represents the value of the rotation Gaussian heatmap when the dense point is located at the boundary of the shrink rotation ellipse. Take two points on the scaled rotating ellipse boundary, and then the values of σ_w and σ_h can be obtained by combining the initial Gaussian heatmap value g_{init} , scale factor ζ , (1), (2), and (5). For example, we take the two vertices of this scaled rotation ellipse (e.g., the two points A and B in Figure 4) to determine the values of σ_w and σ_h . At this point, the horizontal and vertical coordinates of these two

points in the coordinate system $X'O'Y'$ can be calculated by (1) and (2). Bringing g_{init} , A , and B into (5) yields

$$\begin{cases} g_{init} = \text{Exp}\left(-\frac{(\xi_h/2)^2}{2\sigma_h^2}\right) \\ g_{init} = \text{Exp}\left(-\frac{(\xi_w/2)^2}{2\sigma_w^2}\right), \end{cases} \quad (6)$$

then the values of σ_w and σ_h can be obtained.

In particular, the rotation Gaussian heatmap value of the dense points located in the ignorable region or background region is 0. After that, the Gaussian rotation heatmap of the ships with three scales can be generated in Figure 6d–f, respectively.

Based on the above designs, an oriented ship model based on MDP-RGH can be obtained, which can better balance the influence of positive and negative samples of the training dataset, and reduce the influence of noise and background on ship recognition. The positive samples near the edge of the ship are made to conform to the rotated Gaussian distribution by avoiding noise interference. The closer the positive sample is to the edge, the lower its value, so our positive sample is soft. A higher positive sample score means that the positive sample is more representative of the ship. Thus, the MDP-RGH improves the robustness for different size targets and the calculation speed of the proposed model. Next, we will introduce the oriented ship detection algorithm base on MDP-RGH.

3. Oriented Ship Detection Algorithm Based on MDP-RGH

Based on the proposed MDP-RGH method for training dataset improvement, an end-to-end AF-OSD is devised to detect the multi-scale oriented ship, which is accurate identification and information extraction of multi-scale ships, as shown in Figure 7. Precisely, the AF-OSD mainly consists of three parts: (1) label assignment based on MDP-RGH; (2) the AF-OSD; (3) multitask OSALoss.

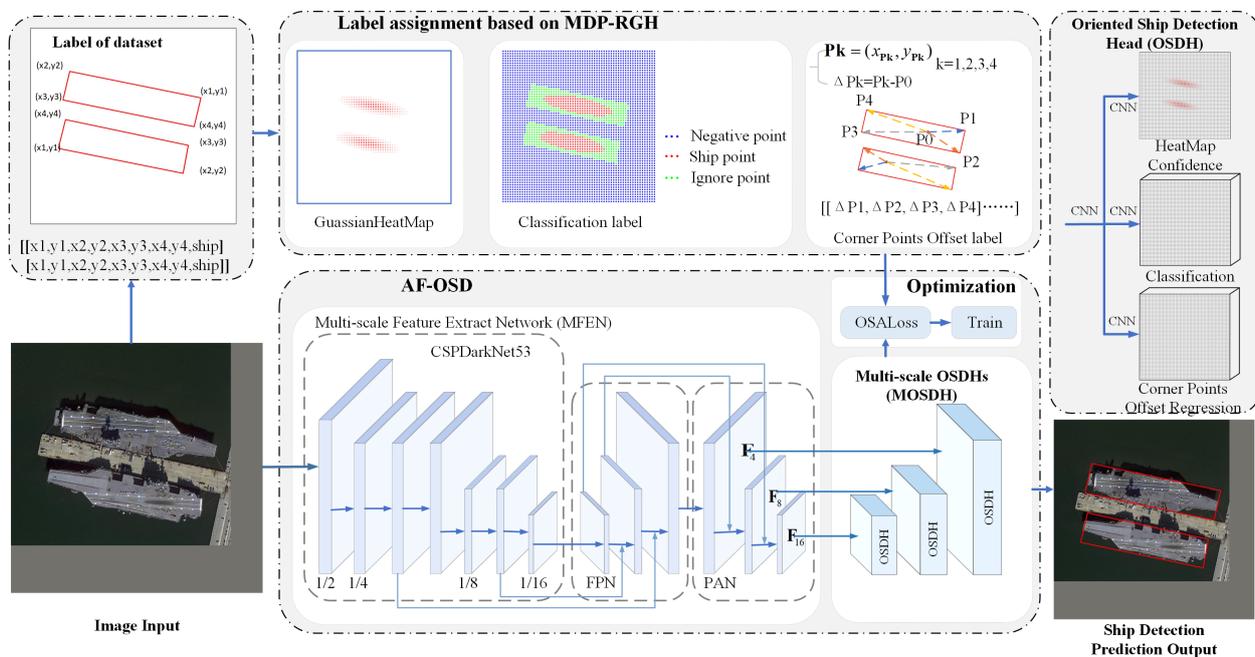


Figure 7. Overview diagram of AF-OSD based on MDP-RGH. (1) Label assignment based on MDP-RGH; (2) the AF-OSD; (3) OSALoss.

3.1. Label Assignment Based on MDP-RGH

According to Section 2, based on the MDP-RGH model, the label of supervising the training process of the AF-OSD can be obtained. The label allocation strategy processes the original dataset labels into three parts: (1) multi-scale dense-point rotation Gaussian heatmap confidence label, which is the weight of dense points; (2) multi-scale dense point category label, namely, regions divided by the image; (3) multi-scale dense point ship bounding box corner offsets label, namely, the offsets of the four corner points of the ship bounding box relative to the dense points in the corresponding object region.

3.1.1. Positive Sample Confidence Labels for Dense Points Based on MDP-RGH

The confidence labels of dense points based on MDP-RGH indicate the probability that the dense points belong to a ship, which can be expressed as $\mathbf{G}_{\text{conf}_s} \in \mathbf{R}^{(\frac{H}{s}) \times (\frac{W}{s}) \times 1}$, where $s = 4, 8, 16$ represents the sampling scales with three values over the image and each element is a scalar.

Specifically, for the dense point $\text{denp}_{s_{ij}}$ falling into the object region, its Gaussian heatmap confidence label value can be calculated by (5), while its value is 0 for the dense point outside the object region, which can be represented as

$$\mathbf{G}_{\text{conf}_{s_{ij}}} = \begin{cases} g\left(x_{\text{denp}_{s_{ij}}}, y_{\text{denp}_{s_{ij}}}\right), & \text{denp}_{s_{ij}} \in \text{area}_{\text{ship}} \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $\mathbf{G}_{\text{conf}_{s_{ij}}}$ means the label value of the Gaussian heatmap of the dense point $\text{denp}_{s_{ij}}$.

3.1.2. The Classification Label for Dense Points

The classification label $\mathbf{G}_{\text{cls}_s} \in \mathbf{R}^{(\frac{H}{s}) \times (\frac{W}{s}) \times N_c}$ of dense points is a three-dimensional matrix, which can be written as

$$\mathbf{G}_{\text{cls}_s} = \left[\mathbf{G}_{\text{cls}_{s_{ij}}}^{(1)}, \dots, \mathbf{G}_{\text{cls}_{s_{ij}}}^{(c)}, \dots, \mathbf{G}_{\text{cls}_{s_{ij}}}^{(N_c)} \right], \quad (8)$$

where each element is a vector with size N_c (here N_c represents the number of categories, including the background category). $\mathbf{G}_{\text{cls}_{s_{ij}}}^{(c)}$ denotes the score of dense point $\text{denp}_{s_{ij}}$ belonging to category c , which can be determined by

$$\mathbf{G}_{\text{cls}_{s_{ij}}}^{(c)} = \begin{cases} 1, & \text{denp}_{s_{ij}} \in c\text{-th} \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

where the dense point belongs to the c -th category. The c -th component of the above vector is set to 1, and the others are set to 0.

3.1.3. The Corner Points offsets Label of the Ship Bounding Box of the Dense Points

We built corner point offset labels of the ship bounding box for regression when training. The label $\mathbf{G}_{\text{Poffsets}_s} \in \mathbf{R}^{(\frac{H}{s}) \times (\frac{W}{s}) \times 8}$ of dense points is a two-dimensional matrix, where each element is a vector with size 8. The label represents the horizontal and vertical coordinates of the four corner points offsets of the ship bounding box of dense point $\text{denp}_{s_{ij}}$, which can be written as

$$\mathbf{G}_{\text{Poffsets}_{s_{ij}}} = \left[\mathbf{G}_{\text{Poffsets}_{s_{ij}}}^{(1)}, \dots, \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^{(8)} \right]. \quad (10)$$

If $\text{denp}_{s_{ij}}$ does not belong to the object region of any oriented ship, all components of $\mathbf{G}_{\text{Poffsets}_{s_{ij}}}$ are 0. If $\text{denp}_{s_{ij}}$ belongs to the object region of an oriented ship, the component of $\mathbf{G}_{\text{Poffsets}_{s_{ij}}}$ can be calculated as

$$\begin{cases} \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^{(2k-1)} = \left(x_{\mathbf{pk}} - x_{\text{denp}_{s_{ij}}} \right) / \text{scale}_s \\ \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^{(2k)} = \left(y_{\mathbf{pk}} - y_{\text{denp}_{s_{ij}}} \right) / \text{scale}_s, \end{cases} \quad (11)$$

where k ($k = 1, 2, 3, 4$) represents the k -th corner point of the ship bounding box, $x_{\mathbf{pk}}$ and $y_{\mathbf{pk}}$ are the horizontal and vertical coordinates of the k -th corner point of the oriented ship bounding box, respectively, and scale_s is the normalized parameter of different scales.

3.2. AF-OSD

Normally, the AF-OSD consists of two parts: (1) a multi-scale feature extraction network (MFEN) and (2) a multi-scale oriented ship detection head (MOSDH). The MFEN extracts the features of the oriented ship in the input image, while the MOSDH applies the extracted features to identify and detect the oriented ship.

3.2.1. MFEN

PANet [21] is selected for the feature extraction structure of the MFEN for the fact that the size of the ship object has a large scale range. To reduce the problem of information loss arising from long transmission distances, the bottom-up path augmentation of the FPN [22] structure is applied, which provides a shorter path for high-level information transmission and effectively reduces information loss, as shown in Figure 8.

For the backbone of MFEN, ship detection is based on remote sensing data, visual range, and complex environments to fully extract the attribute information of ships. Hence, a relatively large feature extraction backbone network is needed to design. Because the cross-stage partial network (CSPNet) [23] solves the problem of repeated gradient information in the optimization process of large CNN frameworks, CSPDarknet53 is chosen as the backbone network, which divides the feature map of the base layer into two parts and then merges them through a cross-stage hierarchy.

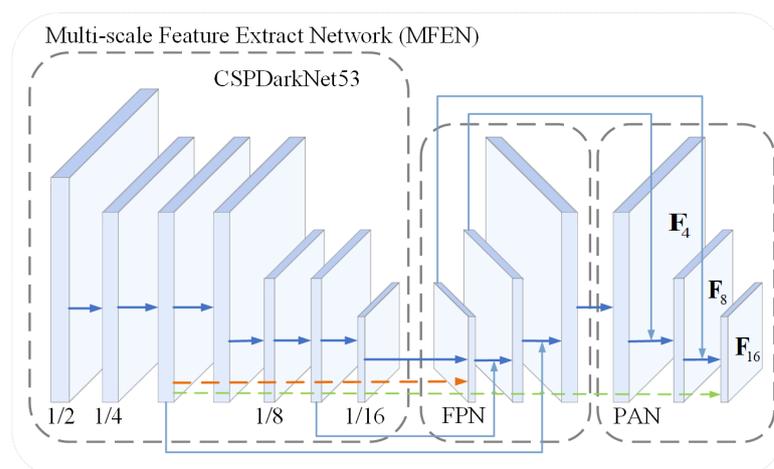


Figure 8. Multi-scale feature extraction network. The red dotted line represents the transmission path from the FPN low-level feature layer to the high-level feature layer. The dotted green line indicates the path from the PAN low-level feature layer to the high-level feature layer.

Based on the designed MFEN, the feature extraction of the input image can be realized. Specifically, the image $I \in \mathbb{R}^{H \times W \times 3}$ (H and W are the height and width, respectively) is

input into the MFEN E_{MFEN} . Then, the multi-scale deep feature $F_s \in \mathbb{R}^{(H/s) \times (W/s) \times C}$ can be extracted as

$$F_s = E_{MFEN}(I, s; \theta_E), \tag{12}$$

where C is the channel number, and θ_E is the weight parameters of the designed MFEN E_{MFEN} . The multi-scale deep feature F_s is fed into the subsequent MOSDH to identify and detect the ship.

3.2.2. MOSDH

The MOSDH is used to identify and detect the oriented ships through the features extracted by the MFEN. The output structure of MOSDH refers to the decoupling output mode of YoloX [24]. Corresponding to the three feature layers of PANet structure, the MOSDH contains three scales of oriented ship detection heads (OSDHs), where the small-scale detection head predicts large ships, the medium-scale detection head predicts medium ships, and the large-scale detection head predicts small ships.

Moreover, each OSDH has three decoupling output branches, as shown in Figure 9: (a) Gaussian heatmap confidence output branch of dense points: to predict the score that each dense point to be a positive sample (belonging to a ship); (b) Gaussian heatmap classification output branch of dense points: to predict the category of each dense point; (c) ship bounding box corner points offsets output branch of dense points: to predict the offsets of the four corner points of the ship (which ship the dense point belongs to) relative to the dense point.

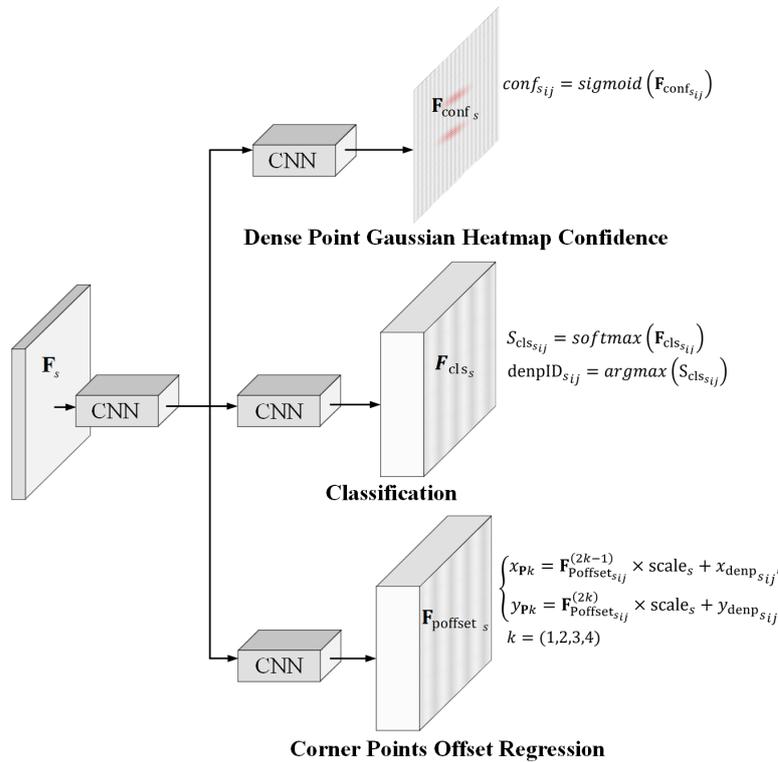


Figure 9. Structure of OSDH.

(a) The Dense Point Gaussian Heatmap Confidence Output Branch

To suppress background and noise interference, the confidence output of the Gaussian heatmap of dense points is designed. The multi-scale deep feature F_s is fed into the dense

point Gaussian heatmap confidence output branch of MOSDH $H_{\text{MOSDH-conf}_s}$, and the Gaussian heatmap confidence output feature $\mathbf{F}_{\text{conf}_s} \in R^{(H/s) \times (W/s) \times 1}$ can be obtained as

$$\mathbf{F}_{\text{conf}_s} = H_{\text{MOSDH-conf}_s}(\mathbf{F}_s, s; \theta_{\mathbf{H}}), \quad (13)$$

where $\theta_{\mathbf{H}}$ is the weight parameters of the MOSDH. Each feature point of the Gaussian heatmap confidence output feature layer predicts the score of the dense point corresponding to this feature point belonging to the ship object region ($\text{area}_{\text{ship}}$). When the score is greater than the set threshold $T_{\text{conf}} \in (0, 1)$, the dense point is judged to be a dense positive point (belonging to the ship object region). The judgment method can be expressed as

$$\begin{cases} \text{denp}_{s_{ij}} \in \text{area}_{\text{ship}}, & \text{sigmoid}(\mathbf{F}_{\text{conf}_{s_{ij}}}) \geq T_{\text{conf}} \\ \text{denp}_{s_{ij}} \notin \text{area}_{\text{ship}}, & \text{otherwise,} \end{cases} \quad (14)$$

where $\mathbf{F}_{\text{conf}_{s_{ij}}}$ means the eigenvalue of the j -th row and the i -th column of the Gaussian heatmap confidence output feature layer, the activation function of the confidence output uses the $\text{sigmoid}(\cdot)$ function, and T_{conf} represents the confidence threshold.

Suppose that the dense point is determined to be a positive dense point. In that case, the category of the oriented object to which the dense point belongs and the coordinates of the four corner points of the oriented ship can be determined. Otherwise, it is directly judged that the dense matter belongs to the background (dense negative point), and no further discrimination and calculation will be carried out.

(b) Dense Point Classification Output Branch

We design the dense point classification output branch to determine the category of dense points. When a dense point is determined as a positive sample point, the dense point classification output branch $H_{\text{MOSDH-cl}_s}$ is used to predict the category of the dense point.

Specifically, the multi-scale deep feature \mathbf{F}_s is input into the classification output branch of the MOSDH $H_{\text{MOSDH-cl}_s}$, and the classification output feature $\mathbf{F}_{\text{cl}_s} \in R^{(H/s) \times (W/s) \times N_c}$ is obtained as

$$\mathbf{F}_{\text{cl}_s} = H_{\text{MOSDH-cl}_s}(\mathbf{F}_s, s; \theta_{\mathbf{H}}). \quad (15)$$

Each feature point with size N_c of the feature layer of the category output predicts the score of each category that the corresponding dense point is, which can be written as

$$\text{denpID}_{s_{ij}} = \text{argmax}\left(\text{softmax}\left(\mathbf{F}_{\text{cl}_{s_{ij}}}\right)\right), \quad (16)$$

where $\text{denpID}_{s_{ij}}$ denotes the category number of the dense point $\text{denp}_{s_{ij}}$, and $\mathbf{F}_{\text{cl}_{s_{ij}}}$ represents the N_c dimension feature vector of the j -th row and the i -th column of the output feature layer of the category.

Therefore, the output feature vector of the category is first converted into the score of each type by the $\text{softmax}(\cdot)$ function, and then the category ID number of the highest category score is calculated by $\text{argmax}(\cdot)$ function to obtain the category of the oriented object.

(c) The Corner Point offsets Output Branch

To determine the position of the bounding box, the corner points to offset the output branch are further built. When a dense matter is judged as a dense positive point, the corner point offset output branch of the dense points predicts the offsets of the four corner points of the ship bounding box relative to the dense point.

Specifically, \mathbf{F}_s is input into the corner point offset output branch of the dense points of the MOSDH $H_{\text{MOSDH-Poffset}_s}$, and the output of offsets feature $\mathbf{F}_{\text{Poffset}_s} \in R^{(H/s) \times (W/s) \times 8}$ is calculated as

$$\mathbf{F}_{\text{Poffset}_s} = H_{\text{MOSDH-Poffset}_s}(\mathbf{F}_s, s, \theta_{\mathbf{H}}). \quad (17)$$

Each feature point with size 8 of the output feature layer predicts the offsets of the four corner points of the ship bounding box relative to the dense point corresponding to the feature point. Therefore, the coordinates of the four corner points are the offsets plus the coordinate of the dense point, which can be expressed as

$$\begin{cases} x_{Pk} = x_{\Delta Pk} + x_{denp_{s_{ij}}} = \mathbf{F}_{Poffsets_{s_{ij}}}^{(2k-1)} \times scale_s + x_{denp_{s_{ij}}} \\ y_{Pk} = y_{\Delta Pk} + y_{denp_{s_{ij}}} = \mathbf{F}_{Poffsets_{s_{ij}}}^{(2k)} \times scale_s + y_{denp_{s_{ij}}} \end{cases} \quad (18)$$

where x_{Pk} and y_{Pk} are the horizontal and vertical coordinates of the k -th corner point Pk of the ship bounding box, respectively, $x_{\Delta Pk}$ and $y_{\Delta Pk}$ represent the offsets of the k -th corner point Pk of the ship bounding box relative to the dense point $denp_{s_{ij}}$, respectively, and $\mathbf{F}_{Poffsets_{s_{ij}}}^{(2k-1)}$ is the $(2k - 1)$ -th value of the j -th row and i -th column feature vector of the output feature.

Multiple dense points may simultaneously predict the same ship object, or output layers of different scales may predict the same object. Therefore, after decoding all prediction boxes, it is necessary to use the NMS technology to filter out the redundant oriented object bounding boxes, obtaining the final detection result of the oriented ship.

3.3. Multi-Task Object Size Adaptive Loss (MOSALoss)

The multi-task OSALoss function, as the target function of the optimizer in the training stage, guides the optimization of the weight parameters of the proposed AF-OSD.

3.3.1. OSALoss Weight

The multi-scale structure of the AF-OSD network has been designed to solve the unbalanced training of different sizes of ships due to the other numbers of positive dense points. Specifically, three detection heads are designed to detect oriented ships of various sizes. However, since the object size is approximately continuous from small to large, the predicted objects of a certain output scale are large or small. The loss function model [25] focuses more on a small target with only one best anchor as a positive sample paired with it. Different from [25], we design the loss function OSALoss suitable for the MDP-RGH-based multi-scale directional ship detection in this paper.

To further solve the training imbalance caused by different object sizes, increasing the number of output characteristic layers at different scales will not only fail to completely solve this problem but also increase the amount of calculation, as shown in Figure 10. As such, an OSALoss weight (OSAWeight) is further designed, which can be adaptive to ship size.

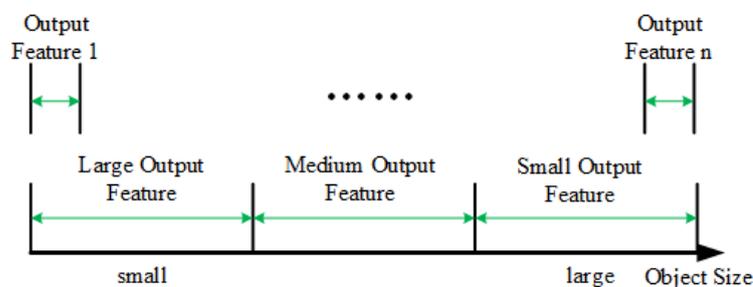


Figure 10. The relationship between the scale size of the target and the model output.

Specifically, if the dense point $denp_{s_{ij}}$ belongs to the object region of a ship $Ships_{s_n}$, the weight of the sample point $W_{adap-densp_{s_{ij}}}$ is obtained as

$$\begin{cases} W_{\text{adap-denp}_{s_{ij}}} = \frac{d_{\text{denp}_s}}{\sqrt{A_{\text{Ships}_{s_n}}}}, & \text{denp}_{s_{ij}} \in \text{area}_{\text{ship}} \\ W_{\text{adap-denp}_{s_{ij}}} = 1, & \text{otherwise,} \end{cases} \quad (19)$$

where Ships_{s_n} represents the n -th ship on the down-sampling scale of s times, $A_{\text{Ships}_{s_n}}$ is the area of the ship, and d_{denp_s} is the distance between two adjacent dense points. According to (19), the weight of this dense point is inversely proportional to the square root of the area of the ship in a dense point matrix of the same scale. When the shipping area is large, the weight is small, while the weight is significant when the shipping area is small.

The difference in object size directly affects the difference in the number of positive sample points, which affects the loss related to positive sample points, leading to the problem of training imbalance. Therefore, the OSAWeight will be introduced into the loss function related to positive sample points to eliminate the influence caused by different object sizes.

3.3.2. Multi-Task OSALoss

In order to solve the situation of easy sample imbalance during training, the multi-task OSALoss $Loss_{\text{adap-multi}}$ consists of the Gaussian heatmap confidence loss $\mathbf{GHCLoss}_s = Loss(\mathbf{F}_{\text{conf}_s}, \mathbf{G}_{\text{conf}_s})$ of dense points, the category prediction loss $\mathbf{CPLoss}_s = Loss(\mathbf{F}_{\text{cls}_s}, \mathbf{G}_{\text{cls}_s})$ of dense points, and the corner points offsets prediction loss $\mathbf{OPLoss}_s = Loss(\mathbf{F}_{\text{Poffsets}_s}, \mathbf{G}_{\text{Poffsets}_s})$ of the ship bounding box. Then, the proposed multi-task OSALoss can be written as

$$Loss_{\text{adap-multi}} = \frac{1}{3} \sum_{s \in (4,8,16)} (\mathbf{GHCLoss}_s + \gamma \mathbf{CPLoss}_s + \lambda \mathbf{OPLoss}_s), \quad (20)$$

where $\mathbf{G}_{\text{conf}_s}$, $\mathbf{G}_{\text{cls}_s}$, and $\mathbf{G}_{\text{Poffsets}_s}$ represent the confidence ground truth of Gaussian heatmap, the classification ground truth of dense points and the corner offsets ground truth of ship bounding box, respectively. γ and λ are the weight parameters, which are adaptive with the training.

Next, we will introduce the Gaussian heatmap confidence loss, the category prediction loss, and the corner points offset prediction loss in detail.

(a) *The Gaussian Heatmap Confidence Loss:*

Adaptive wing loss [26] is used for the $\mathbf{GHCLoss}_s$, which can be represented as

$$\mathbf{GHCLoss}_s = \mathbf{AWingLoss}(\mathbf{F}_{\text{conf}_s}, \mathbf{G}_{\text{conf}_s}). \quad (21)$$

Since the $\mathbf{GHCLoss}_s$ adopts the relatively mature adaptive wing loss, we do not introduce the OSAWeight into this loss function.

(b) *The Category Prediction Loss:*

The cross-entropy function is used for the \mathbf{CPLoss}_s . Particularly, the \mathbf{CPLoss}_s is composed of positive $\mathbf{CPLoss}_s^{\text{Pos}}$ and negative $\mathbf{CPLoss}_s^{\text{Neg}}$, which can be written as

$$\mathbf{CPLoss}_s = \mathbf{CPLoss}_s^{\text{Pos}} + \mathbf{CPLoss}_s^{\text{Neg}}, \quad (22)$$

where the positive category prediction loss $\mathbf{CPLoss}_s^{\text{Pos}}$ can be given as

$$\begin{aligned} \mathbf{CPLoss}_s^{\text{Pos}} = & \sum_{j=1}^{H/sW/s} \sum_{i=1}^{H/sW/s} W_{\text{adap-denp}_{s_{ij}}} \times \mathbf{M}_{\text{obj}_{s_{ij}}} \times \sum_{c=1}^{N_c} \left(\mathbf{G}_{\text{cls}_{s_{ij}}}^c \log \left(\mathbf{S}_{\text{cls}_{s_{ij}}}^c \right) \right. \\ & \left. + \left(1 - \mathbf{G}_{\text{cls}_{s_{ij}}}^c \right) \log \left(1 - \mathbf{S}_{\text{cls}_{s_{ij}}}^c \right) \right), \end{aligned} \quad (23)$$

where $\mathbf{M}_{\text{obj}_{s_{ij}}}$ is the positive dense point mask. $\mathbf{S}_{\text{cls}_{s_{ij}}}^c$ is the score when the dense point $\text{denp}_{s_{ij}}$ is the c -th category. $\mathbf{G}_{\text{cls}_{s_{ij}}}^c$ is the ground truth of whether the dense point $\text{denp}_{s_{ij}}$ is the c -th category.

$\mathbf{M}_{\text{obj}_{s_{ij}}}$ can be written as

$$\begin{cases} \mathbf{M}_{\text{obj}_{s_{ij}}} = 1, & \text{denp}_{s_{ij}} \in \text{area}_{\text{ship}} \\ \mathbf{M}_{\text{obj}_{s_{ij}}} = 0, & \text{otherwise} \end{cases} \quad (24)$$

and $\mathbf{S}_{\text{cls}_{s_{ij}}}^c$ can be expressed as

$$\mathbf{S}_{\text{cls}_{s_{ij}}} = \text{softmax}(\mathbf{F}_{\text{cls}_{s_{ij}}}). \quad (25)$$

For negative dense point category prediction loss, to further solve the imbalance between the number of positive and negative samples, the hard negative example mining technique is applied. Specifically, the negative dense category prediction loss for each dense point can be written as

$$\text{Loss}_{\text{neg}}(\mathbf{F}_{\text{cls}_{s_{ij}}}, \mathbf{G}_{\text{cls}_{s_{ij}}}) = \sum_{c=1}^{N_c} (\mathbf{G}_{\text{cls}_{s_{ij}}}^c \log(\mathbf{S}_{\text{cls}_{s_{ij}}}^c) + (1 - \mathbf{G}_{\text{cls}_{s_{ij}}}^c) \log(1 - \mathbf{S}_{\text{cls}_{s_{ij}}}^c)) \times \mathbf{M}_{\text{bg}_{s_{ij}}}, \quad (26)$$

where $\mathbf{M}_{\text{bg}_{s_{ij}}}$ is a negative sample mask, which can be written as

$$\begin{cases} \mathbf{M}_{\text{bg}_{s_{ij}}} = 0, & \text{denp}_{s_{ij}} \in \text{area}_{\text{ship}} \\ \mathbf{M}_{\text{bg}_{s_{ij}}} = 1, & \text{otherwise.} \end{cases} \quad (27)$$

Then, the negative sample category prediction loss of dense points is sorted from the largest to the smallest, and the ranked negative sample category prediction loss $\text{Loss}_{\text{neg}_{\text{sorted}}}(\mathbf{F}_{\text{cls}_{s_l}}, \mathbf{G}_{\text{cls}_{s_l}})$ can be obtained, where l denotes the serial number of the l -th dense point after the negative dense category prediction loss ranking of dense points. Finally, only the sum of the negative sample category prediction loss of the top L is taken as the negative sample category prediction loss. The category prediction loss $\text{CPLoss}_s^{\text{Neg}}$ can be represented as

$$\text{CPLoss}_s^{\text{Neg}} = \sum_{l=1}^L \text{Loss}_{\text{neg}_{\text{sorted}}}(\mathbf{F}_{\text{cls}_{s_l}}, \mathbf{G}_{\text{cls}_{s_l}}). \quad (28)$$

In particular, if the number of negative samples is greater than twice the number of positive samples, L takes two times the number of positive samples. Otherwise, L is equal to the number of positive samples.

(c) *The Corner Points Offsets Prediction Loss:*

The ship bounding box corner point offset prediction loss uses the object size adaptive weighted Smooth-L1 error loss. Specifically, we first use the Smooth-L1 loss to predict each dense point of the ship bounding box corner point offsets as

$$\text{Loss}(\mathbf{F}_{\text{cls}_{s_{ij}}}, \mathbf{G}_{\text{cls}_{s_{ij}}}) = \begin{cases} \sum_{m=1}^8 0.5 \times (\mathbf{F}_{\text{Poffsets}_{s_{ij}}}^m - \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^m)^2, & \text{if } \left| \mathbf{F}_{\text{Poffsets}_{s_{ij}}}^m - \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^m \right| < 1 \\ \sum_{m=1}^8 \left(\left| \mathbf{F}_{\text{Poffsets}_{s_{ij}}}^m - \mathbf{G}_{\text{Poffsets}_{s_{ij}}}^m \right| - 0.5 \right), & \text{otherwise.} \end{cases} \quad (29)$$

Then, the target-scale adaptive weighted summation of the predicted losses for all positive dense points ship target frame corner point offsets is performed. The total loss \mathbf{OPLoss}_s can be written as

$$\mathbf{OPLoss}_s = \sum_{j=1}^{H/sW/s} \sum_{i=1}^{W/s} W_{\text{adap-dens}_{s_{ij}}} \times \mathbf{M}_{\text{obj}_{s_{ij}}} \times \text{Loss}(\mathbf{F}_{\text{cls}_{s_{ij}}}, \mathbf{G}_{\text{cls}_{s_{ij}}}). \quad (30)$$

Using the above label assignment strategy, network design, and loss function, we can train the deep learning network of the AF-OSD through the Adam optimization method [27], as summarized in Algorithm 1.

Algorithm 1: Training AF-OSD based on Adam optimization

INPUT: Image data set $\mathbb{I} = \{\mathbf{I}^{(1)}, \mathbf{I}^{(2)}, \dots\}$, and the corresponding labels

$\mathbb{Y} = \{\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots\}$

OUTPUT: The trained AF-OSD with fixed neural network parameters $\theta_{\mathbf{E}}$ and $\theta_{\mathbf{H}}$

Construct AF-OSD

Initialize $\theta_{\mathbf{E}}, \theta_{\mathbf{H}}$ randomly

for k to iterations do

Select randomly minibatch $\mathcal{I}_k \subset \mathbb{I}, \mathcal{Y}_k \subset \mathbb{Y}$

Label assign as Section 3.1: $\mathcal{G}_k \leftarrow \mathcal{Y}_k$

$\mathcal{F}_k = \text{AF-OSD}(\mathcal{I}_k; \theta_{\mathbf{E}k}, \theta_{\mathbf{H}k})$

$\mathcal{L}_k = \text{Loss}_{\text{adap-multi}}(\mathcal{F}_k, \mathcal{G}_k)$

$\theta_{\mathbf{E}k+1}, \theta_{\mathbf{H}k+1} \leftarrow \text{Optimizer}_{\text{Adam}}(\mathcal{L}_k; \theta_{\mathbf{E}k}, \theta_{\mathbf{H}k})$

end for

Freeze $\theta_{\mathbf{E}}, \theta_{\mathbf{H}}$

4. Experimental Conditions

In this section, experiments on two public-oriented ship datasets are conducted to quantitatively and qualitatively evaluate the proposed AF-OSD. A series of experiments are conducted using the DOTA ship dataset [28] and HRSC2016 dataset [29]. Next, we will introduce the experimental conditions, including experimental platforms, datasets, evaluation metrics, and implementation details.

4.1. Experimental Platforms

All the experiments are implemented on a desktop computer with an Intel Core i7-9700 CPU, 32 GB of memory, and a single NVIDIA GeForce RTX 3090 with 24 GB GPU memory.

4.2. Dataset

4.2.1. DOTA Ship Dataset

DOTA is a large-scale aerial image dataset containing 2806 images, with 15 categories labeled with oriented boxes and image sizes ranging from 800×800 to 1600×1600 . We select 434 images containing ships (including 37,028 ships), where 90% of them are randomly selected as the training set, and the other 10% are as the validation set. The size distribution of ships is not uniform, and there are few large-scale ships. So, we enhance and expand the large ships in the dataset (rotation and flip enhancement), and the size distribution of the original dataset and the enhanced dataset is shown in Figure 11a, and the rotation angle distribution is shown in Figure 11b.

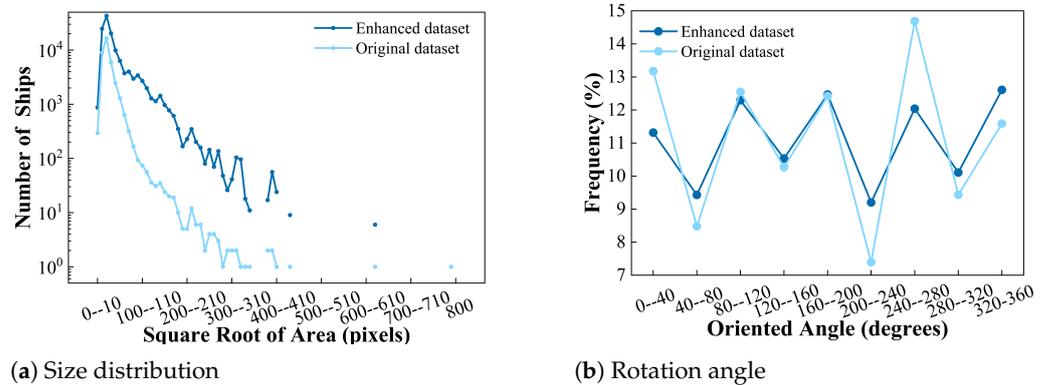


Figure 11. Size and rotation angle of ship targets in the DOTA dataset.

4.2.2. HRSC2016 Dataset

HRSC2016 is a remote sensing image dataset containing ship targets labeled as arbitrary orientations. It consists of 1061 images (436 images are the training set, 181 images are the validation set, and 444 images are the test set), whose spatial sizes range from 300×300 to 1500×900 . We use the training set to train the model and the validation set to test the model. The size and rotation angle distributions of ships are illustrated in Figure 12, and the ship sizes and angles are more uniformly distributed. Therefore, we do not target special scale ships when we perform enhancement expansion (flip enhancement).

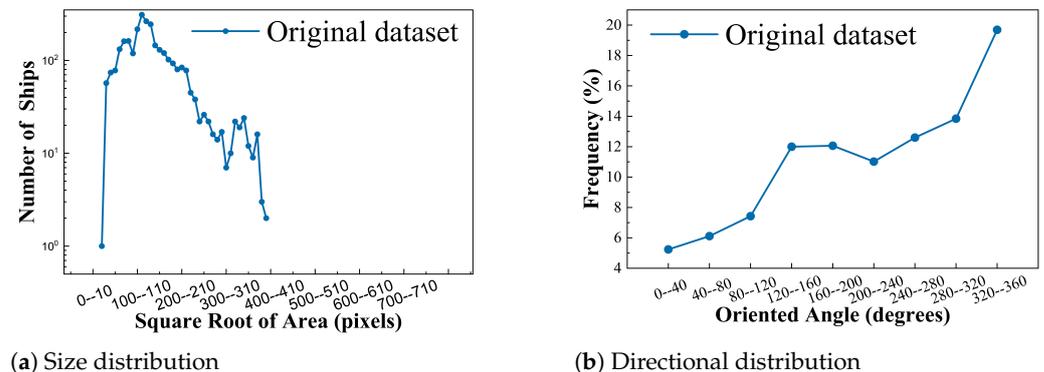


Figure 12. Size and directional distributions of ship targets in the HRSC2016 dataset.

4.3. Evaluation Metrics

In this article, the widely used metrics in object detection are adopted to measure the detection performance, i.e., precision, recall, and average precision (AP) (IOU threshold set to 0.5). Moreover, we use the number of parameters (Params) and average running time to evaluate the complexity and speed of the model, respectively.

4.4. Implementation Details

Due to the large size of the original input image, we cut it before network training. Specifically, the input resolution is set to 800×800 . As such, the original image and the corresponding labels are first preprocessed, where the cut image is uniform 800×800 in size (if the image is smaller than 800×800 , the padding operation is applied). To avoid the situation that an object is cut into two halves and disappears, an overlapping area of 200 pixels is set in the experiments. Of course, there will be some cases where only a part of the object is in the cropped image.

If this happens, we will judge the ratio of the area of the target on the cropped image to the area of the target itself. If this ratio is greater than the set threshold (taken as 0.6), the

label information of the object will be kept. The judgment method of whether to keep or discard the target on the cropped picture is represented as

$$\begin{cases} \text{Keep,} & \frac{A(\text{box}_{\text{Object}} \cap \text{box}_{\text{Crop_image}})}{A(\text{box}_{\text{Object}})} > \text{Threshold} \\ \text{Discard,} & \text{otherwise,} \end{cases} \quad (31)$$

where $A(\cdot)$ denotes the function of the area and $\text{box}_{\text{Object}} \cap \text{box}_{\text{Crop_image}}$ means the intersection of the target box and the cut picture box to be calculated. If it is determined that the target needs to be retained, its coordinates on the cut image are calculated as

$$p^{O*} = p_i^{O*} = p_i^O - p_1^{\text{Crop}}, i \in 1, 2, 3, 4 \quad (32)$$

where p^{O*} is the new coordinate of the target frame in the cropped picture, p_i^O denotes the i -th coordinate point of the target frame in the original picture, and p_1^{Crop} denotes the first coordinate point of the cropped picture frame in the original pictures. The image cutting schematic is shown in Figure 13.

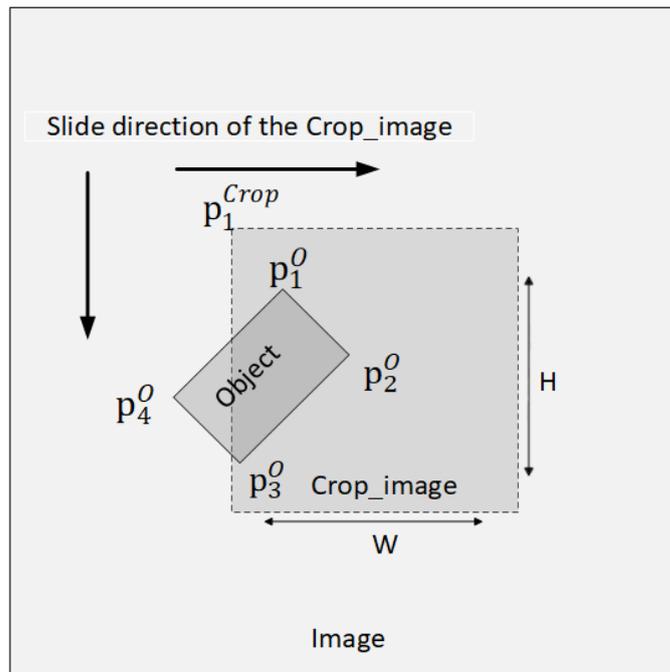


Figure 13. Image cutting schematic.

During the training process, the Adam optimizer is used to optimize the weight parameters of our AF-OSD, and the initial learning rate is set to 0.001 with an exponential decay of 5%. The batch size is set to 10, and all networks are trained for 200 epochs.

During the testing process, since the image size of HRSC2016 is not very large, before input to the model, we resize the image and then pad its size to 800×800 . For the test images of DOTA, we used the same processing method as the training data. The confidence threshold $score_t$ is set to 0.1 and the NMS threshold is set to 0.6.

5. Experimental Results

To verify the rationality and advantages of the proposed AF-OSD, some comparison experiments are implemented to study the performance differences between the proposed oriented ship detector and other state-of-the-art (SOTA) oriented-object detection methods. Moreover, some ablation experiments are conducted to analyze the effects of each part on the performance of the proposed oriented ship detector.

5.1. Ablation Experiments

To testify to the necessity of the oriented ship detector proposed, three sets of ablation comparison experiments are designed: (1) comparing the performance of a multi-scale network structure and a single-scale network structure for detecting oriented ships; (2) comparing the performance of a network without MDP-RGH confidence output branch and a network with MDP-RGH confidence output branch for detecting oriented ships; (3) comparing the performance of the AF-OSD trained with the traditional multi-task loss function and the AF-OSD trained with multi-task OSALoss function.

All the ablation comparisons of DNN networks are trained using the HRSC2016 training set, parameter settings, and training algorithms mentioned in the previous section. They are then tested in the HRSC2016 validation set for measuring their detection performance.

5.1.1. Ablation Experiment on Different Scales

To verify the effect of multi-scale detection head on model enhancement, three single-scale models and multi-scale models were compared in terms of detection accuracy on the HRSC2106 dataset. Single-scale oriented ship detection network contains different detection scales, including 50×50 , 100×100 , and 200×200 , and the multi-scale oriented ship detection network has three detection scales output.

As shown in Table 1, the comparison results show that the detection accuracy of the multi-scale oriented ship detection network is better than that of any single-scale detection network. Since it is difficult for a single-scale detection head to cover the remote sensing oriented ship with a large scale range, the multi-scale detection head can solve this problem well. So, compared with the AF-OSD network that does not use the multi-scale detection head structure, the performance of the whole model can be improved by 1.31–2.69% by introducing this structure.

Table 1. Ablation study of the multi-scale structure of the AF-OSD. The ✓ means network with the structure. The ✗ means network without the structure. The bold results mean the best performance.

Single Head	Multi Head	Recall	Precision	AP
50×50	✗	0.9020	0.8761	0.8789
100×100	✗	0.9020	0.7948	0.8700
200×200	✗	0.9039	0.8180	0.8833
✗	✓	0.9068	0.8963	0.8969

5.1.2. Ablation Experiment on MDP-RGH

To ensure the MDP-RGH confidence is favorable for the system, the comparison of the detection accuracy of the oriented ship detection network without the MDP-RGH confidence output branch and the oriented ship detection network with this branch is designed.

As shown in Table 2, the comparison results show that the detection accuracy of the oriented ship detection network with the MDP-RGH confidence branch is better. Compared with the AF-OSD without the MDP-RGH structure, the performance of the whole framework can have achieved 4.75% AP improvement by introducing this structure on HRSC2016.

Table 2. Ablation study of the MDP-RGH confidence output branch of the AF-OSD.

FPN	PAN	MDP-RGH	Recall	Precision	AP
✓	✓	✗	0.8854	0.8615	0.8494
✓	✓	✓	0.9068	0.8963	0.8969

5.1.3. Ablation Experiment on Multi-Task OSALoss

To demonstrate the effectiveness of the OSALoss designed in this paper, the detection accuracy between the AF-OSD trained with the traditional multi-task loss (TMLoss)

function and the network trained with the multi-task OSALoss (MOSALoss) function on HRSC2016 is compared.

As shown in Table 3, comparison results show that our AF-OSD trained with the MOSALoss function has better detection accuracy. The OSALoss weight in the MOSALoss function is proportional to the distance between the dense points and inversely proportional to the square root of the ship area. Therefore, the MOSALoss function can be adjusted adaptively with the different ship sizes. So, compared with the AF-OSD network trained with TMLoss, the performance of the model which trained through MOSALoss can be improved by 4.56%. Therefore, the AF-OSD network with the MOSALoss function can effectively solve the training imbalance caused by the different scales of oriented ships.

Table 3. Ablation study of the multi-task OSALoss function.

TMLoss	MOSALOSS	Recall	Precision	AP
✓	✗	0.8595	0.9082	0.8513
✗	✓	0.9068	0.8963	0.8969

Based on the ablation experiments, the three parts of the AF-OSD are fully proven effective in remote sensing ship detection.

5.2. Comparisons with SOTA

To demonstrate the effectiveness of the method in this paper, the AF-RPN [18] is used as a baseline method, and the anchor-based methods RRPN [30], ROI-Transformer [31], ref. [20], and R³Det [32] are compared with our method. The anchor-free methods IENet [33], GGHL [20], and GRS-DET [6] are also used for the comparison. The baseline method shrinks the object bounding box to obtain the core region and takes all the core region points as positive. The proposed method utilized the MDP-RGH to model a ship compared to the baseline method. We can compare the benefits of MDP-RGH through experiments.

Repeatability experiments on the DOTA dataset by the proposed method in this paper yield better results on the DOTA test dataset, as shown in Figure 14. The first row in the figure is the ground truth of the dataset. From Figure 14a,b, we can find that the proposed method has a high recognition rate for small targets in remote sensing images in the DOTA dataset. From Figure 14a,c, it can be seen that our method can achieve a good result both for large and dense small-oriented ships at the same time. Moreover, AF-OSD can accurately identify the small targets in the ground truth that are not completely labeled in the ground truth, as shown in Figure 14a. However, the missed detection phenomenon of the method proposed is significantly less than the ground truth. Meanwhile, it can be seen that the bounding box accuracy of the detection algorithm in this paper is high, as shown in Figure 15, in which the green bounding box is our method and the red bounding box is the ground truth.

5.2.1. Results on the DOTA Ships Dataset

Table 4 further shows the quantitative comparison results between the proposed AF-OSD and the SOTA-oriented ship detection algorithms on the DOTA ship dataset. Among these methods, the R³Det method has the highest accuracy among anchor-based methods. Compared with the highest accuracy of the anchor-based method R³Det, our method has improved by 0.32%. Moreover, among anchor-free methods, the experimental results show that the GRS-DET achieves about 2.36% AP improvement compared to the baseline method because of the Gaussian mask and selective concatenation module (SCM) structure. Compared to the GRS-DET, the proposed MDP-RGH can better suppress the interference of negative samples such as background and noise in the image, and the OSALoss function can further overcome the training imbalance problem caused by the different ship sizes, so our method outperforms the GRS-DET with 3.39% AP improvement on the DOTA ship dataset. The positive and negative sample distribution of GGHL is based on Gaussian

prior and is adjusted during the training process, making it applicable to remote sensing targets of any shape, but it will have some impact on regular symmetrical targets (the distribution of positive samples in symmetrical rules will change to some extent after adjustment), resulting in its AP value being 2% lower than the proposed method on the DOTA ship targets.

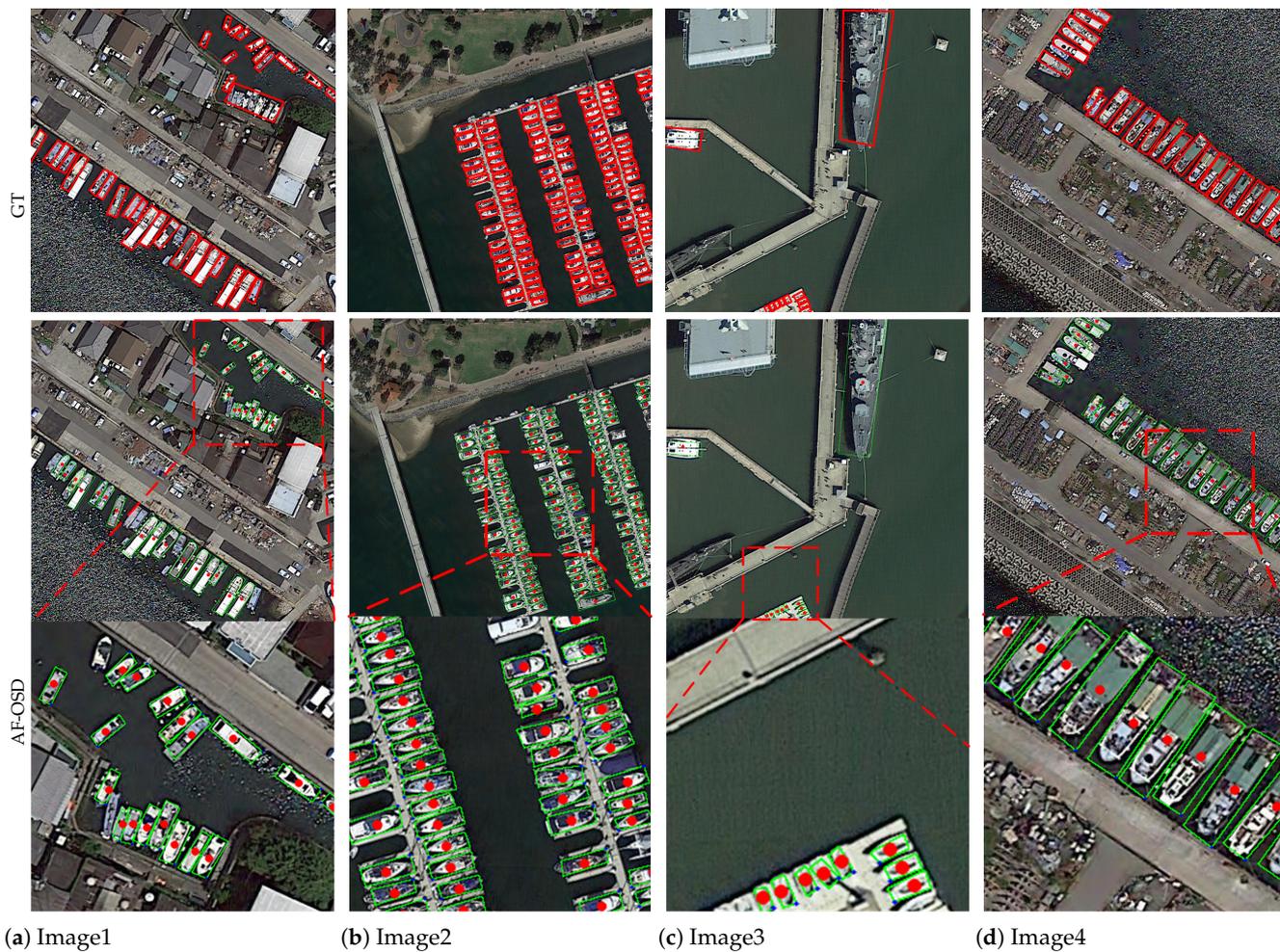


Figure 14. Detection comparisons of different test images on the DOTA ships dataset (red box: ground truth bounding box; red point: ship central points of proposed method; green box: proposed method bounding box).

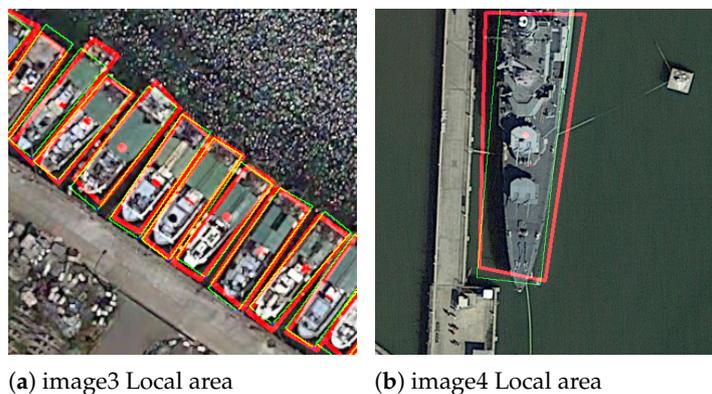


Figure 15. Comparison of AF-OSD bounding box on DOTA dataset.

Table 4. Comparison results on the DOTA ship dataset.

Methods	Anchor-Free	Backbone	AP
RRPN	×	ResNet-101	0.4454
ROI-Transformer	×	ResNet-101	0.8359
R ³ Det	×	ResNet-152	0.8784
IENet	✓	ResNet-101	0.7161
AF-RPN	✓	CSPDarkNet53	0.8241
GRS-DET	✓	ResNet-101	0.8477
GGHL	✓	DarkNet53	0.8616
AF-OSD (Ours)	✓	CSPDarkNet53	0.8816

5.2.2. Results on the HRSC2016 Dataset

For the HRSC2016 test dataset, we directly resize and pad the image to 800×800 pixels, and then input it into the model for processing. Based on the qualitative comparison of the detection effectiveness of the proposed AF-OSD and the baseline method (AF-RPN) under the HRSC2016 dataset, it can be seen that our method tends to use the point near the center of the ship to predict the ship's bounding box. The point marked red means it had the highest score to predict the ship. This indicates that the proposed MDP-RGH can overcome the interference of negative samples such as background and noise in the image by using the center point to detect the ship. Thus, a better result can be realized by our method, as shown in Figure 16. From Figure 16b, it can be obtained that the proposed method effectively reduces the false detection rate of the ships. Moreover, AF-OSD can accurately identify the small targets in ground truth that are not completely labeled in the ground truth, as shown in Figure 16a,d.

Table 5 further shows the quantitative comparison results between the proposed AF-OSD and the SOTA-oriented ship detection algorithm on the HRSC2016 dataset. Among the anchor-based methods, the R³Det method has the highest accuracy. Compared with R³Det, the accuracy can be improved by 0.43% by our method. Moreover, among anchor-free methods, the experimental results show that the GRS-DET [6] achieves about 4.63% AP improvement compared to the baseline method. Due to the scale of ships in the HRSC2016 dataset being more uniform, our method, with the help of the MDP-RGH and the OSALoss function, outperforms the GRS-DET with 0.12% AP improvement on the HRSC2016 dataset.

Table 5. Comparison results on the HRSC2016 dataset.

Methods	RRPN	ROI-Transformer	R ³ Det	IENet	AF-RPN	GRS-DET	AF-OSD (Ours)
AP	0.7905	0.8620	0.8926	0.7501	0.8494	0.8957	0.8969

The results proved that the proposed AF-OSD based on MDP-RGH can achieve SOTA through qualitative and quantitative comparative experiments.

5.3. Network Complexity Analysis

Based on the experimental process analysis, considering the network model's complexity and the computational effort, the proposed AF-OSD based on MDP-RGH in this paper is compared with the anchor-based and anchor-free models, as shown in Table 6. The comparison of results is based on the amount of computation and computing time within the corresponding methods' articles.

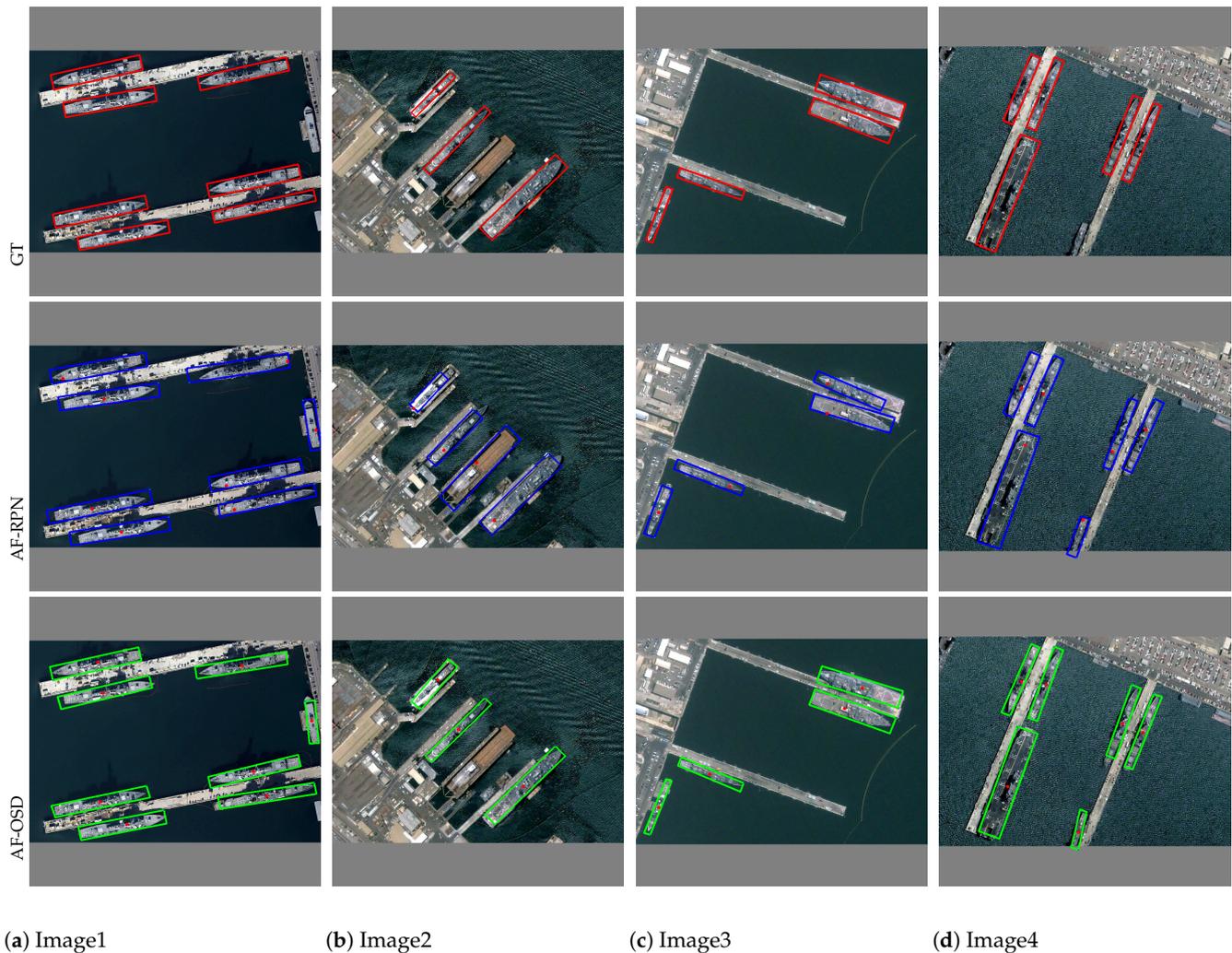


Figure 16. Detection comparisons of different test images on HRSC2016 (red box: ground truth bounding box; blue box: baseline method bounding box; green box: proposed method bounding box).

Table 6. Comparison of different networks on HRSC2016.

Methods	AP	Params	GPU	Speed
RRPN	0.7905	181 MB	GTX 1080TI	0.2857 s
ROI-Transformer	0.8620	273 MB	RTX 3090	0.1282 s
R ³ Det	0.8926	227 MB	RTX 3090	0.0950 s
IENet	0.7501	212 MB	GTX 1080TI	0.0592 s
AF-RPN	0.8494	189 MB	RTX 3090	0.0109 s
GRS-DET	0.8957	200 MB	GTX 1080TI	0.0729 s
AF-OSD (Ours)	0.8969	192 MB	RTX 3090	0.0162 s

The network parameters of our proposed AF-OSD are significantly less, and its parameters are 35 MB less than those in the R³Det method. Moreover, compared to the anchor-free method, the parameters of our method are 3 MB more than the baseline, 20 MB less than IENet, and 8 MB less than GRS-DET. Based on the analysis of the toxic network model, the computational complexity of this paper is 136 G FLOPS.

6. Discussion

In this paper, in order to solve the sample imbalance problem and suppress the interference of negative samples such as background and noise, the oriented ship is modeled

via the proposed MDP-RGH according to its shape and direction to generate ship labels with more accurate information, which can predict the target for the contribution of positive samples at different positions to judge the target. Additionally, we designed an end-to-end anchor-free oriented ship detector (AF-OSD) net based on MDP-RG and validate its detection performance.

For the necessity study of network modules, the ablation experiment includes three parts: multi-scale structure, MDP-RGH-based label confidence, and MOSA loss. It has been proved that the framework with a multi-scale structure can extract the features of the oriented ship with different scales better. Multi-scale-based detection head setup enhances the robustness of the model. Therefore, the design of the multi-scale structure of the oriented ship detection network is reasonable. We use MDP-RGH to weight the positive samples so that the samples are rotationally Gaussian distributed and have values closer to the edge. The positive samples of our method are soft. As the result of the ablation experiment, the output can use the Gaussian heatmap confidence to predict whether the target is a negative sample or a positive one. A higher positive sample score means that the sample is more representative of the ship. Therefore, the MDP-RGH confidence can better suppress the interference of negative samples such as background and noise in the image, so the detection performance has improved. The improved loss function comparison experiments can prove that the weight can better solve the loss imbalance caused by the large and small ships. The larger the target, the higher the number of positive sample densities included; the smaller the target, the lower the number of positive sample densities included. The ablation experiment shows that the weight can adaptively solve the situation of sample imbalance that easily occurs during training.

Based on the test results on the DOTA ship dataset and HRSC2016 dataset, one can reach that the proposed AF-OSD has the best target recognition effect in remote sensing images with multi-scale ships, complex scenes, and positive and negative sample imbalance. The AF-OSD has features of high accuracy, few parameters of the network model, and high robustness. Due to the complexity of airborne remote sensing scenes, strong background and noise interference, positive and negative sample imbalance, and multiple ship scales, ship detection is a key and challenging task in remote sensing. The proposed method in this paper can better solve the above challenges. In summary, the proposed AF-OSD has a low number of parameters and higher model accuracy, detection accuracy, and computational speed.

As our main research object is ship targets. Most ships in the HRSC2016 and DOTA ship datasets have the characteristic of being narrow at both ends and wide in the middle, and their external contours are similar to rectangles. Therefore, the proposed method in this paper can be extended to the detection of remote sensing targets with similar external contours as rectangles (such as vehicles and stadium-like remote sensing targets) in any orientation, and even be possibly applied to symmetrical targets. However, AF-OSD is not an optimal solution for irregular and asymmetrical remote sensing targets (such as port-like targets). The application of the method in this paper to the detection of irregular targets is somewhat affected by the poor matching of the Gaussian heatmap. Therefore, the detection of full targets will be developed in the subsequent work. Moreover, to improve recognition speed and increase detection accuracy by decoupling the scale and task, the channel and spatial attention mechanism need to be considered in the design. However, the results showed that the inclusion of the attention mechanism actually led to a decrease in the detection accuracy of the network, which means that aspect requires further research and study in future work.

7. Conclusions

This paper presents the AF-OSD (anchor-free oriented ship detector) based on the MDP-RGH (multi-scale dense-point rotation Gaussian heatmap), which is designed for detecting arbitrarily oriented ships. The AF-OSD adopts a PANet structure for multi-scale feature extraction and a decoupled detection head for predicting the rotate Gaussian

heatmap confidence, class, and location information of the oriented ship. The AF-OSD is capable of handling ships of different sizes, whose bounding boxes can be of any shape that fits the target. Experiments on two commonly used oriented ship datasets demonstrate the superiority of the AF-OSD over other state-of-the-art methods. These results showcase the high robustness of the AF-OSD in remote sensing ship identification applications and its ability to balance the number of positive and negative samples of various scale targets unsupervised while reducing the impact of noise and background on target identification. As a result, the proposed model has achieved state-of-the-art performance.

Author Contributions: Conceptualization, Z.H.; Methodology, Z.H.; Software, Z.H.; Formal analysis, G.P., H.L. and S.C.; Investigation, H.L.; Writing—original draft, Z.H.; Writing—review & editing, Z.H., G.P., K.G., H.L. and S.C.; Visualization, S.C.; Supervision, G.P., K.G., H.L. and S.C.; Project administration, G.P. and K.G.; Funding acquisition, K.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number U2241275 and 61827814; Beijing Natural Science Foundation, grant number Z190018; China High-resolution Earth Observation System Project, grant number 52-L10D01-0613-20/22.

Data Availability Statement: Data available in a publicly accessible repository. The data presented in this study are openly available in DOTA dataset at [<https://doi.org/10.1109/CVPR.2018.00418>], reference number [28], and HRSC dataset at [<https://doi.org/10.5220/0006120603240331>], reference number [29].

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Yang, G.; Li, B.; Ji, S.; Gao, F.; Xu, Q. Ship Detection From Optical Satellite Images Based on Sea Surface Analysis. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 641–645. [[CrossRef](#)]
2. Zou, Z.; Shi, Z. Ship Detection in Spaceborne Optical Image With SVD Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5832–5845. [[CrossRef](#)]
3. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Wang, J. Improving Performance and Adaptivity of Anchor-Based Detector Using Differentiable Anchoring With Efficient Target Generation. *IEEE Trans. Image Process* **2021**, *30*, 712–724. [[CrossRef](#)]
4. Hu, Z.; Gao, K.; Zhang, X.; Wang, J.; Wang, H.; Han, J. Probability Differential-Based Class Label Noise Purification for Object Detection in Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6509705. [[CrossRef](#)]
5. Yang, Y.; Tang, X.; Cheung, Y.M.; Zhang, X.; Liu, F.; Ma, J.; Jiao, L. AR2Det: An Accurate and Real-Time Rotational One-Stage Ship Detector in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5605414. [[CrossRef](#)]
6. Zhang, X.; Wang, G.; Zhu, P.; Zhang, T.; Li, C.; Jiao, L. GRS-Det: An Anchor-Free Rotation Ship Detector Based on Gaussian-Mask in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3518–3531. [[CrossRef](#)]
7. Ren, Z.; Tang, Y.; He, Z.; Tian, L.; Yang, Y.; Zhang, W. Ship Detection in High-Resolution Optical Remote Sensing Images Aided by Saliency Information. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5623616. [[CrossRef](#)]
8. Liu, G.; Zhang, Y.; Zheng, X.; Sun, X.; Fu, K.; Wang, H. A New Method on Inshore Ship Detection in High-Resolution Satellite Images Using Shape and Context Information. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 617–621. [[CrossRef](#)]
9. Guo, W.; Chen, H.; Zhang, Z.; Zhang, Y.; Yu, W. Direct Oriented Ship Localization Regression in Remote Sensing Imagery with Curriculum Learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2584–2587. [[CrossRef](#)]
10. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
11. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
12. Liu, Z.; Wang, H.; Weng, L.; Yang, Y. Ship Rotated Bounding Box Space for Ship Extraction From High-Resolution Optical Satellite Images With Complex Backgrounds. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1074–1078. [[CrossRef](#)]
13. Yang, X.; Sun, H.; Fu, K.; Yang, J.; Sun, X.; Yan, M.; Guo, Z. Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. *Remote Sens.* **2018**, *10*, 132. [[CrossRef](#)]
14. Lin, H.; Shi, Z.; Zou, Z. Fully Convolutional Network With Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1665–1669. [[CrossRef](#)]

15. Pan, X.; Ren, Y.; Sheng, K.; Dong, W.; Yuan, H.; Guo, X.; Ma, C.; Xu, C. Dynamic Refinement Network for Oriented and Densely Packed Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11204–11213. [[CrossRef](#)]
16. Yi, J.; Wu, P.; Liu, B.; Huang, Q.; Qu, H.; Metaxas, D. Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 2149–2158. [[CrossRef](#)]
17. Yang, F.; Xu, Q.; Li, B. Ship Detection From Optical Satellite Images Based on Saliency Segmentation and Structure-LBP Feature. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 602–606. [[CrossRef](#)]
18. Zinelli, A.; Musto, L.; Pizzati, F. A Deep-Learning Approach for Parking Slot Detection on Surround-View Images. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 683–688. [[CrossRef](#)]
19. Zhong, Z.; Sun, L.; Huo, Q. An Anchor-Free Region Proposal Network for Faster R-CNN-based Text Detection Approaches. *Int. J. Doc. Anal. Recognit.* **2019**, *22*, 315–327. [[CrossRef](#)]
20. Huang, Z.; Li, W.; Xia, X.G.; Tao, R. A General Gaussian Heatmap Label Assignment for Arbitrary-Oriented Object Detection. *IEEE Trans. Image Process* **2022**, *31*, 1895–1910. [[CrossRef](#)]
21. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768. [[CrossRef](#)]
22. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [[CrossRef](#)]
23. Wang, C.Y.; Mark Liao, H.Y.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580. [[CrossRef](#)]
24. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
25. Wang, P.; Sun, X.; Diao, W.; Fu, K. FMSSD: Feature-Merged Single-Shot Detection for Multiscale Objects in Large-Scale Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3377–3390. [[CrossRef](#)]
26. Wang, X.; Bo, L.; Fuxin, L. Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October 2019–2 November 2019; pp. 6970–6980. [[CrossRef](#)]
27. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
28. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983. [[CrossRef](#)]
29. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines. In Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, Porto, Portugal, 20 October 2017; pp. 324–331. [[CrossRef](#)]
30. Ma, J.; Shao, W.; Ye, H.; Wang, L.; Wang, H.; Zheng, Y.; Xue, X. Arbitrary-Oriented Scene Text Detection via Rotation Proposals. *IEEE Trans. Multimed.* **2018**, *20*, 3111–3122. [[CrossRef](#)]
31. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2844–2853. [[CrossRef](#)]
32. Yang, X.; Yan, J.; Feng, Z.; He, T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3163–3171. [[CrossRef](#)]
33. Lin, Y.; Feng, P.; Guan, J.; Wang, W.; Chambers, J. IENet: Interacting Embranchment One Stage Anchor Free Detector for Orientation Aerial Object Detection. *arXiv* **2019**, arXiv:1912.00969.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.