

文章编号 1004-924X(2023)06-0892-13

## 面向航天光学遥感复杂场景图像的舰船检测

刘忻伟<sup>1,2,3</sup>, 朴永杰<sup>1,3\*</sup>, 郑亮亮<sup>1,3</sup>, 徐 伟<sup>1,3</sup>, 籍浩林<sup>1,2,3</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100039;

3. 中国科学院 天基动态快速光学成像技术重点实验室, 吉林 长春 130033)

**摘要:** 基于深度学习的目标检测算法直接应用于航天光学遥感(Space Optical Remote Sensing, SORS)复杂场景图像中会出现舰船目标检测效果不佳的问题。针对该问题, 本文以近海复杂背景的密集排布舰船和远海多干扰中小目标舰船为检测对象, 提出一种改进的YOLOX-s(Improved You Only Look Once-s, IM-YOLO-s)算法。在特征提取阶段, 引入CA位置注意力模块, 分别从高度与宽度两个方向对目标信息的位置进行权重分配, 提高了模型的检测精度; 在特征融合阶段, 将BiFPN加权特征融合算法应用到IM-YOLO-s的颈部结构, 进一步提升了小目标船只检测精度; 在模型优化训练阶段, 以CIoU损失替代IoU损失、以变焦损失替代置信度损失、调整类别损失权重, 增大了正样本分布密集区域的训练权重, 减少了密集分布船只的漏检率。另外, 在HRSC2016数据集的基础上增加额外的离岸中小舰船图像, 自建了HRSC2016-Gg数据集, HRSC2016-Gg数据集增强了海上船只及中小像素船只检测时的鲁棒性。通过数据集HRSC2016-Gg评测算法性能, 实验结果表明: IM-YOLO-s对于SORS场景舰船检测的召回率为97.18%, AP@0.5为96.77%, F1值为0.95, 较原YOLOX-s算法分别提高了2.23%, 2.40%和0.01。这充分表明该算法可以对SORS复杂背景图像进行高质量舰船检测。

**关键词:** 舰船检测; 深度学习; CA注意力模块; 加权特征融合; 损失函数优化

中图分类号: TP394.1; TH691.9 文献标识码: A doi: 10.37188/OPE.20233106.0892

## Ship detection for complex scene images of space optical remote sensing

LIU Xinwei<sup>1,2,3</sup>, PIAO Yongjie<sup>1,3\*</sup>, ZHENG Liangliang<sup>1,3</sup>, XU Wei<sup>1,3</sup>, JI Haolin<sup>1,2,3</sup>

(1. Changchun Institute of Optics, Fine Mechanics and Physice, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100039, China;

3. Key Laboratory of Space-Based Dynamic & Rapid Optical Imaging Technology, Chinese Academy of Sciences, Changchun 130033, China)

\* Corresponding author, Email: pyj0314@163.com

**Abstract:** When deep-learning-based target detection algorithms are directly applied to the complex scene images generated by space optical remote sensing (SORS), the ship target detection effect is often poor. To address this problem, this paper proposes an improved YOLOX-S (IM-YOLO-s) algorithm, which

收稿日期: 2022-06-08; 修订日期: 2022-07-08.

基金项目: 钱学森空间技术实验室创新工作站开发基金资助项目(No. GZZKFJJ2020003)

uses densely arranged offshore ships with complex backgrounds and ships with multi-interference and small targets in the open sea as detection objects. In the feature extraction stage, the CA location attention module is introduced to distribute the weight of the target information along the height and width directions, and this improves the detection accuracy of the model. In the feature fusion stage, the BiFPN weighted feature fusion algorithm is applied to the neck structure of IM-YOLO-s, which further improves the detection accuracy of small target ships. In the training stage of model optimization, the CIoU loss is used to replace the IoU loss, zoom loss is used to replace the confidence loss, and weight of the category loss is adjusted, which increases the training weight in the densely distributed areas of positive samples and reduces the missed detection rate of densely distributed ships. In addition, based on the HRSC2016 dataset, additional images of small and medium-sized offshore ships are added, and the HRSC2016-Gg dataset is constructed. The HRSC2016-Gg dataset enhances the robustness of marine ship and small and medium-sized pixel ship detection. The performance of the algorithm is evaluated based on the dataset HRSC2016-Gg. The experimental results indicate that the recall rate of IM-YOLO-s for ship detection in the SORS scene is 97.18%, AP@0.5 is 96.77%, and the F1 value is 0.95. These values are 2.23%, 2.40%, and 0.01 higher than those of the original YOLOX-s algorithm, respectively. This indicates that the algorithm can achieve high quality ship detection from SORS complex background images.

**Key words:** ship detection; deep learning; coordinate attention; weighted feature fusion; loss function optimization

## 1 引言

随着我国海上强国战略的提出以及遥感技术的发展,提升对海洋环境的监测能力成为我国迈入海洋强国的关键。其中航天光学遥感(Space Optical Remote Sensing, SORS)场景舰船检测是遥感领域的研究热点。舰船在海洋环境中扮演着重要的角色,不仅在渔业开发、交通运输、海洋开发等民用领域发挥着重要的作用,而且在战斗保障、技术保障、后勤保障等军事领域也起着重要的作用。因此,智能化地对舰船目标进行监测,可以提高舰船工作效率,提升港口的管理能力。而在实际检测中,由于拍摄距离远、港口背景复杂、光照不均匀以及海杂波云团等的干扰,很容易造成漏检。因此,航天光学遥感复杂背景图像下的舰船检测已成为当前研究的重点<sup>[1-5]</sup>。

传统方法在对舰船进行检测时主要是通过感兴趣区域海域与陆地进行海陆分割,然后采用对比框算法、半监督分级分类<sup>[6]</sup>等算法获取候选对象的疑似区域,在疑似区域完成舰船目标特征的精细化提取,根据提取的特征完成舰船的识别与定位,最后剔除掉冗余特征的目标,整个过

程遵从由粗及细的检测流程<sup>[7]</sup>。但传统算法抵抗干扰的能力有限,在复杂背景下的检测精度不高。

近些年来随着深度学习的飞速发展,基于深度学习的图像目标检测取得了重大发展。目前,关于深度学习在目标检测方面的算法根据其完成步骤可以分为基于候选区域的双阶段检测和基于回归的单阶段检测。

早期的基于候选区域的深度学习目标检测算法在区域选择时通常采用滑窗的模式,这类方法由于采用穷举的方法,其在物体框提取上无指导,计算量巨大,效率低下不利于推广。

双阶段检测算法解决了早期算法在锚框利用上的缺陷,相比于早期滑窗穷举式搜索的方法,双阶段检测在第一阶段用候选区域方法创建目标检测的感兴趣区域,高效提升了检测的效率。在检测的第二阶段,通过卷积神经网络,对候选区域目标进行分类与目标预测框的选取。双阶段检测算法的典型代表主要有R-CNN(Region Convolutional Neural Network)<sup>[8]</sup>, SPP-net(Spatial Pyramid Pooling Network), Fast R-CNN<sup>[9]</sup>, R-FCN(R-Fully Convolutional Networks), Mask R-CNN<sup>[10]</sup>等。尽管双阶段检测算

法在检测效率方面较早期的滑窗穷举的方式有较大提升,但仍旧受限于提取的候选区域目标框的数量较多,所以速度较慢。

基于回归的单阶段检测器不需要提前筛选感兴趣区域,它通过对特征图进行网络化划分,直接对特征图进行目标的分类与定位,典型代表是 YOLOv3~v4<sup>[11-12]</sup> 系列算法和 SSD (Single Shot Detection)<sup>[13]</sup> 算法以及 RetinaNet<sup>[14]</sup> 等。

上述基于深度学习的目标检测算法主要是基于锚框(Anchor-based)的,这类模型通常在航天光学遥感图像中通用性较差、正负样本容易失衡,近些年来,有不少科学家提出了基于无锚框(Anchor-free)的目标检测,其不依赖于预定义框,避免了与锚框相关的复杂计算。基于无锚框检测模型可以大致分为基于关键点的方法和基于中心区域的方法。基于关键点的方法首先确定几个可学习的目标热力图的代表点,然后依据目标框的纵横比确定预测框的预测值,代表算法有 CornerNet<sup>[15]</sup>, CenterNet<sup>[16]</sup> 等。基于中心区域的方法是先确定目标的正样本区域,然后回归它到目标边界框的四个边的距离以进行检测,代表算法有 FCOS(Fully Convolutional One-Stage)<sup>[17]</sup>, FoveaBox<sup>[18]</sup> 等。

以上算法大多已被广泛应用于遥感图像舰船目标检测中。Wang<sup>[19]</sup> 等在 YOLOv3-Tiny 算法的基础上保留了小尺度特征图,增强了小目标舰船的检测精度,引入残差网络结构,增强了特征提取阶段的目标提取能力,借助  $1 \times 1$  卷积和  $3 \times 3$  卷积特性,融合了高层特征与底层特征信息,从而有效提取舰船特征,提升了检测的准确性;Shi<sup>[20]</sup> 等对 SSD 算法进行了改进,利用特征融合卷积模块加强了各层网络之间的关联,利用反卷积模块减少了特征提取阶段信息的丢失,从而减少了小目标检测的漏检问题以及舰船重复检测等问题。同时改进了网络的损失函数优化训练阶段,改善了训练过程中正负样本不平衡的局面;Jiao<sup>[21]</sup> 等基于 RetinaNet 网络进行了改进,将原有矩形框预测改为旋转框预测,可以更加精确地匹配真实目标,同时引入了 CBAM 模块,增强了特征提取网络对舰船目标的聚焦,改善了近岸舰船检测效果。

但上述算法在场景复杂、尺度多变的舰船目

标图像中大多检测精度较低。对舰船目标进行检测时,由于近海岸舰船易受港口复杂环境的干扰且大多停靠紧密易发生非极大值抑制,造成检测时的误检漏检率骤升;远海岸图像中小目标船只由于其特征较少,易受到海杂波、岛屿等的干扰,检测时也易发生漏检。

基于以上分析,本文提出 IM-YOLO-s 海上舰船检测算法,该算法可以很好地应对尺度变化,同时在复杂场景船只检测时也能降低漏检率。在 YOLOX-s<sup>[22]</sup> 的基础上,首先引入了 CA (Coordinate Attention)<sup>[23]</sup> 位置注意力模块,然后对将要融合的特征图进行权重分配,最后将 CIoU<sup>[24]</sup> 损失和变焦损失引入到模型优化训练阶段,并重新分配了类别损失的权重。

为了验证本文算法的有效性,在数据集 HRSC2016-Gg 上进行了训练与评估,结果表明该方法可以有效解决复杂场景下密集型舰船检测漏检问题,有效提高多尺度舰船图像的目标召回率。

## 2 SORS 复杂场景下舰船检测算法

本文提出的 IM-YOLO-s 算法可以用于 SORS 复杂场景下的舰船检测,IM-YOLO-s 算法总体组成如图 1 所示,网络由三部分组成,分别是 Backbone, Neck, Head。2.1 节是模型的整体概述,2.2 节和 2.3 节模型中两个关键组件的详细实现,2.4 节介绍了模型训练阶段损失函数的改进。

### 2.1 IM-YOLO-s 算法概述

YOLOX 通过控制参数 `depth_multiple` 和 `width_multiple` 调整网络的深度和宽度。本文采用 YOLOX-s 结构,该网络结构宽度深度适中,能够兼顾准确性与快速性要求。

算法在输入端进行了基于马赛克(Mosaic)算法数据增强,将 4 幅经过不同数据增强操作的图片合成一幅,扩大数据规模的同时丰富了检测背景,增强了复杂背景下检测的鲁棒性。增强训练后的数据接入 Focus 网络,其操作过程如图 2 所示,图中不同颜色的方块代表不同的像素,以  $4 \times 4 \times 3$  图像输入为例,经过 Focus 操作后,图片中每隔一个像素取一个值,尺寸变为  $2 \times 2$ ,3 通道变为 12 通道。

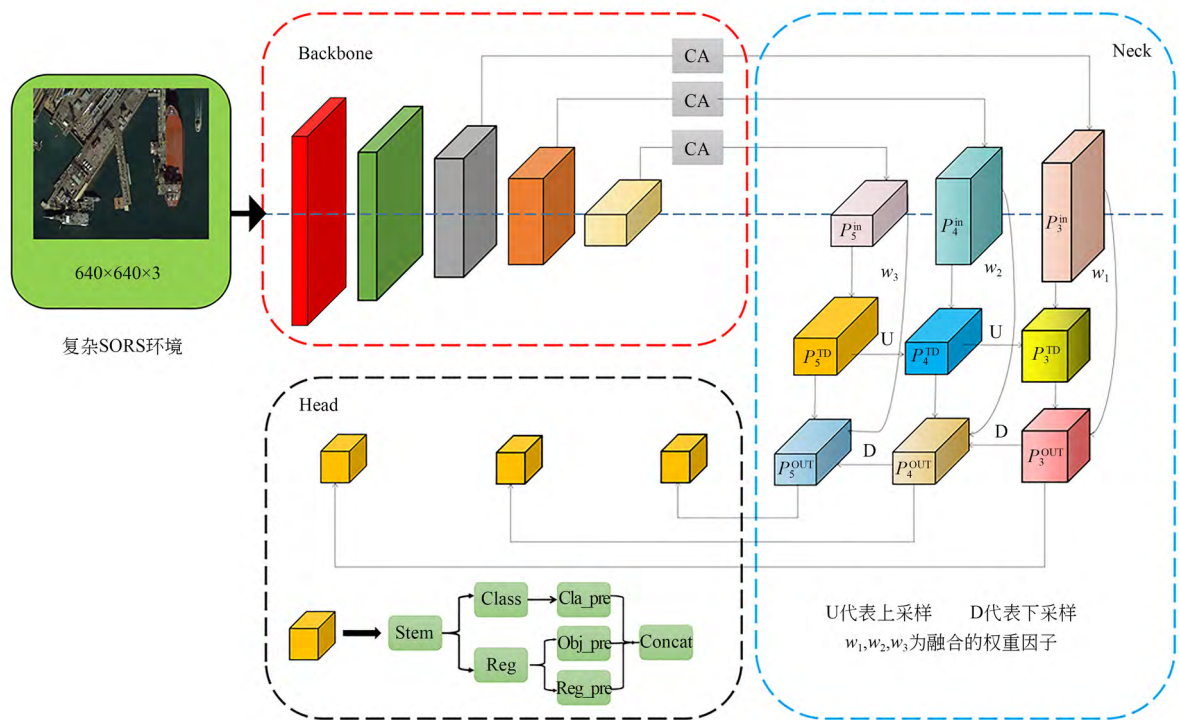


图 1 IM-YOLO-s 网络结构示意图  
Fig. 1 Structure chart of IM-YOLO-s

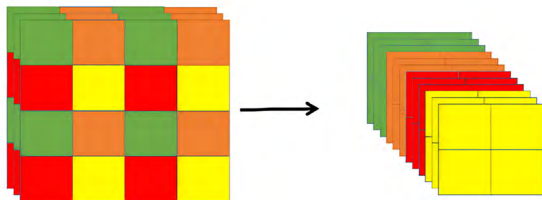


图 2 Focus 操作过程示意图  
Fig. 2 Schematic diagram of Focus operation process

输入到网络中的图片首先经过 Resize 操作统一以 640×640×3 的图像输入 Focus 结构,变成 320×320×12 的特征图,再经过 1×1 卷积操作调整通道数,最后以 320×320×32 的特征图输入到 CSPLayer 结构中。

特征提取阶段,提取出 3 个不同层的特征,低层特征  $P_3$ 、中层特征  $P_4$ 、高层特征  $P_5$ ,其尺度分别为 80×80,40×40,20×20,低层特征更多拥有舰船目标的位置信息,高层特征更多拥有遥感图像的语义信息。

特征融合阶段,特征提取阶段的 3 个不同层特征图输入到 Neck 部分分别为  $P_3^{\text{in}}, P_4^{\text{in}}, P_5^{\text{in}}$ ,不同层不同级加权融合后输出  $P_3^{\text{out}}, P_4^{\text{out}}, P_5^{\text{out}}$ 。

特征预测阶段,沿用原 YOLOX-s 算法:(1)解耦头改变了 YOLOv3 将分类与回归在单一卷积中实现的方式,而是将其分为两部分分别实现,最终将预测结果整合;(2)摒弃了自 YOLOv3 以来的锚框设计,而是采用 anchor free 的方式,避免使用复杂的锚框设计,同时也减少了后续预测框与真实框之间大量的拟合优化计算;(3)通过三种不同的下采样率得到不同尺度的特征图,每种特征图对应不同大小的锚框,借助锚框充当“桥梁”,完成标签分配,挑选出正样本预测框。该过程用了一种简化版的 SimOTA<sup>[25]</sup>方法,该方法是由旷视科技在 2021 年初提出的,极大地提升了正样本匹配的效率。最终,通过计算目标框与预测框之间的损失值,不断优化为目标框确定最终的预测框。

## 2.2 融入 CA 注意力模块

考虑到光学遥感图像中弱小舰船目标较多,背景干扰较复杂,易发生漏检误检,同时本文实验的主要目的在于能够精准定位舰船的位置信息,因此引入一种新型 CA 坐标注意力机制,可以有效提高算法的检测精度。CA 注意力机制不同于 SE (Squeeze and Excitation)<sup>[26]</sup>注意力机制和

CBAM(Convolution Block Attention Module)<sup>[27]</sup>注意力机制,CA可获取图像大范围的目标信息。SE注意力机制是通道域注意力机制的一种,它通过关注图像内部的通道信息来调整图像中目标权重,但不能获取图像的空间信息;CBAM是通道注意力与空间注意力的结合体,对特征图的每个位置的多个通道进行最大池化与平均池化操作,但只能得到图像的局部信息,无法获取全局信息。

CA注意力机制的结构如图3所示,输入端与输出端在经过CA操作之后张量尺寸没有发生变化,可以嵌入模型结构的任意位置。

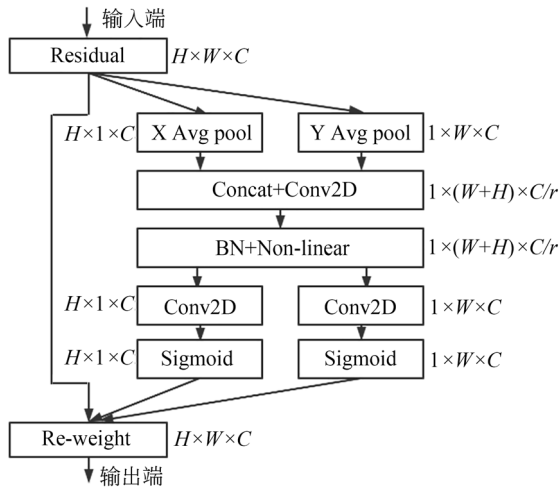


图3 CA注意力机制结构图

Fig. 3 Structure chart of CoordinateAttention

CA模块实现过程大致可以分为五步:

(1) 输入特征张量  $x = [x_1, x_2, \dots, x_c] \in R^{H \times W \times C}$ , 其中  $H, W, C$  分别表示特征图的高度、宽度和维度;

(2) 将输入之后的特征图分为  $X, Y$  两个方向进行平均池化, 获得特征图两个方向的信息;

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_c(h, i), \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w). \quad (2)$$

(3) 将两个方向的特征图进行拼接, 拼接之后输入到  $1 \times 1$  卷积核中。  $1 \times 1$  卷积不改变图像的宽度与高度, 但可以改变图像的维度。借助这个特性, 将拼接之后的特征图维度降为原来的  $C/r$ , 得到  $F1$ ;

(4) 而后进行批归一化处理以及非线性优化来编码  $X$  方向和  $Y$  方向的空间信息。经过处理后的特征图记为  $f$ :

$$f = \delta(F_1([z^h, z^w])); \quad (3)$$

(5) 将特征图再次沿  $X, Y$  两方向经过  $1 \times 1$  卷积处理, 然后对两个方向的特征图经 Sigmoid 处理得到特征图在宽度、高度两个方向的注意力权重  $g^h$  与  $g^w$ ;

(6) 结合原始特征图, 得到最终在宽度和高度方向上带有注意力权重的特征图。

$$y_c(i, j) = x_c(i, j) \times g^h(i) \times g^w(j). \quad (4)$$

本文将CA模块融入在3个特征提取层之后, 可以增强模型的特征提取能力, 前景目标会得到更多的权重关注。

### 2.3 Neck 结构改进

YOLOX-s 在 Neck 部分采用的是 FPN+PAN 结构, FPN 即在特征融合阶段先引入一条侧向连接层, 然后将高层特征不断上采样达到中层和底层特征图相同的分辨率进行融合, 传递高层特征舰船语义信息。PAN 是在 FPN 基础上增加了另外一条侧向连接层, 然后将低层特征不断下采样达到 FPN 融合后的特征图分辨率, 最终将不同特征以相同的权重进行相加。但实际上不同分辨率的特征图对融合输出后检测的影响是不同的, 相同权重的融合会影响最终的融合效果<sup>[28]</sup>。

本文将 BiFPN<sup>[29]</sup>的思想应用到特征融合中。首先将输入到 Neck 部分的特征图按照原算法操作, 然后将输入的 3 个不同尺度的特征图  $P_3^{\text{in}}, P_4^{\text{in}}, P_5^{\text{in}}$  以不同的权重  $w_1, w_2, w_3$  直接引入到 FPN+PAN 融合结果中。

与之前结构对比如图4所示。

其中  $P_3^{\text{in}}, P_4^{\text{in}}, P_5^{\text{in}}$  是主干特征提取网络输出的 3 个不同层的特征图,  $P_5^{\text{TD}}$  是  $P_5^{\text{in}}$  经过  $1 \times 1$  卷积调整通道得到,  $P_4^{\text{TD}}$  由  $P_5^{\text{TD}}$  上采样后和  $P_4^{\text{in}}$  先 Concat 融合, 后经过 CSPLayer 结构得到,  $P_3^{\text{TD}}$  由  $P_4^{\text{TD}}$  上采样后先与  $P_3^{\text{in}}$  经过 Concat 融合, 后经过 CSPLayer 结构得到。

加权融合之后各为:

$$P_3^{\text{out}} = P_3^{\text{TD}'} \cdot w_1 + P_3^{\text{in}} \cdot (1 - w_1), \quad (5)$$

$$P_4^{\text{out}} = P_4^{\text{TD}'} \cdot w_2 + P_4^{\text{in}} \cdot (1 - w_2), \quad (6)$$

$$P_5^{\text{out}} = P_5^{\text{TD}'} \cdot w_3 + P_5^{\text{in}} \cdot (1 - w_3), \quad (7)$$

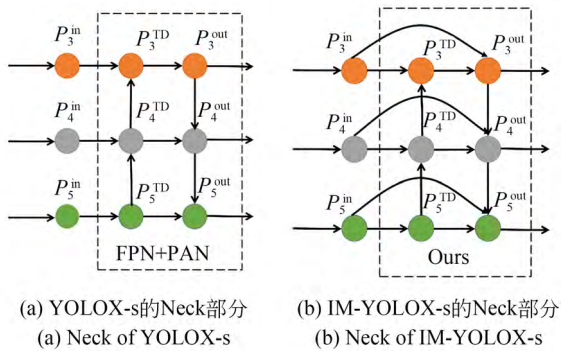


图 4 Neck 网络结构对比图

Fig. 4 Network structure comparison diagram of Neck

其中:  $P_3^{TD'}$  即  $P_3^{TD}$ ,  $P_4^{TD'}$  由  $P_3^{TD}$  下采样与  $P_4^{TD}$  先 Concat 融合, 后经过 CSPLayer 结构得到,  $P_5^{TD'}$  由  $P_4^{TD'}$  下采样与  $P_5^{TD}$  先 Concat 融合, 后经过 CSPLayer 结构得到。CSPLayer 结构如图 5 所示。

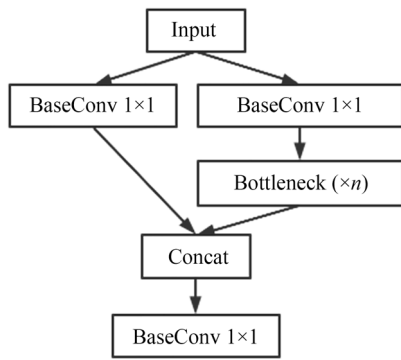


图 5 CSPLayer 结构图

Fig. 5 Network structure chart of CSPLayer

主干部分采用 1 个基本卷积块和堆叠 N 个 Bottleneck 残差块结构提取特征, 残差部分采用 1 个基本卷积块, 最后 Concat 操作合并两个分支再作一次基本卷积块。

$w_1, w_2, w_3$  设置为可被学习训练的参数并经过归一化处理, 可以消除权重的无界性, 加快模型收敛。因此, 采用权重归一化的方式进行融合输出特征图不再是简单的等权重相加, 而是采用训练权重值进行融合操作。这样, 输出后的特征图根据训练反馈自适应地融合不同层不同级特征, 增强信息融合。

### 2.4 损失函数改进

深度学习中损失值可以反映预测值与真实值之间的差距, 而损失函数在模型预测值与真实

值不断拟合的过程中发挥着重要的作用。YOLOX-s 的损失函数分为 IoU 存在损失、置信度损失、类别损失三类, 总损失是三种损失函数加权相加的结果。本文对损失函数的改进主要有以下三方面:

#### (1) IoU 损失改进

IoU 为预测框与实际框之间的交并比, 如式 (8) 所示, 其中  $A$  为预测框,  $B$  为实际框:

$$IoU = \frac{A \cap B}{A \cup B} \quad (8)$$

YOLOX-s 的 IoU 损失如式 (9) 所示:

$$IoU_{Loss} = 1 - IoU \quad (9)$$

$IoU_{Loss}$  越接近 0, 说明预测效果越好, 反之  $IoU_{Loss}$  若接近 1, 说明预测框与实际框之间几乎无交集, 但损失函数在 0~1 之间时无法准确地表达预测框与实际框之间真实的重合情况, 如图 6 (a)。同时当  $IoU_{Loss} = 1$  时, 也无法判断预测框与实际框之间真实的相对位置, 如图 6 (b)。

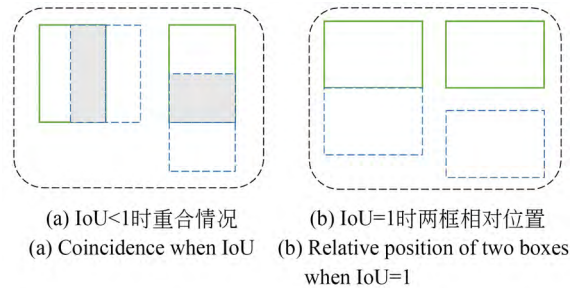


图 6 IoU 取不同值时两种框相交情况

Fig. 6 Intersection of two boxes with different values IoU

这种只借助相交面积来计算损失函数导致模型优化阶段无法给训练提供准确的反馈。针对上述问题, 本文提出将 CIoU 损失应用到 IoU 损失中, 以提高模型优化的效率。CIoU 损失如式 (10) 所示:

$$CIoU_{Loss} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (10)$$

其中:  $b, b^{gt}$  分别代表了预测框和实际框的中心点, 且  $\rho$  代表的是计算两个中心点间的欧式距离,  $c$  代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离, 如图 7 如示。通过这一项可以加快最小化预测框与真实框中心点之间的距离。  $\alpha v$  可以用来拟合预测框与实际框之间的纵横比, 可以使预测框更加贴近真实框的长宽

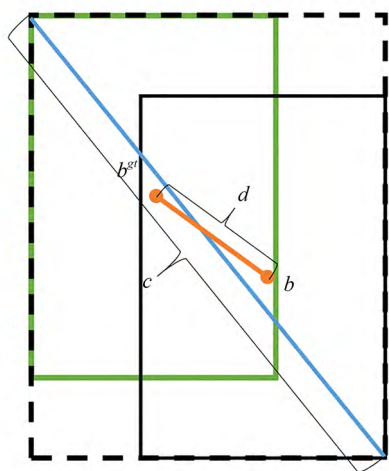


图 7 CIoU 中心点惩罚项示意图

Fig. 7 Schematic diagram of penalty item of central point of CIoU

比例。 $v$ 是用来衡量预测框与真实框纵横比的一致性,定义如式(11):

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2. \quad (11)$$

$\alpha$ 用作平衡参数,定义如式(12):

$$\alpha = \frac{v}{(1 - IoU) + v}. \quad (12)$$

(2)置信度损失改进

近海岸船只分布一般具有多而密的特点,而YOLOX-s检测此类密集船只时漏检率较高。IM-YOLO-s将VariFocalLoss<sup>[30]</sup>作为置信度损失,可以有效解决难易样本不平衡的问题,一定程度降低算法在检测密集船只时的漏检率。

(3)类别损失设置

由于本文对于舰船类别检测要求不敏感,因此总损失函数适当减少了类别损失的权重。总损失函数如式(13)所示:

$$Loss_{Total} = \lambda Loss_{ciou} + Loss_{VF} + Loss_{obj}, \quad (13)$$

其中 $\lambda$ 在本文中设置为5。

### 3 实验结果及分析

#### 3.1 舰船数据集及评价指标

目前受限于国防安全、商业隐私、拍摄成本等因素,公开的光学遥感数据集较少,本文实验需同时检测近海岸舰船与海上舰船,且要对大、中、小舰船都有较好的检测效果,因此自建了

HRSC2016-Gg数据集,空间分辨率从15 m到0.5 m不等,图片尺寸范围在 $300 \times 300 \sim 1\,500 \times 900$ 之间。其中包括Liu<sup>[31]</sup>等制作的HRSC2016舰船数据集,该数据集中舰船目标像素相对较大,可用来本文中大、中型舰船识别。并额外从谷歌地图爬虫得到了一些舰船图片,采用标注软件LabelImg进行了标注,可应用于小像素船只的识别。数据集涵盖了不同气候条件(多云、多雾)、不同光照条件以及含有各种背景干扰(海杂波、尾迹波、海上岛屿、岸上建筑物)的图片。

数据集共有1 963张图片,表1列出了训练集、验证集和测试集的样本数量。

表 1 数据集划分

Tab. 1 Data set division

| 数据集 | 数量    |
|-----|-------|
| 训练集 | 1 413 |
| 验证集 | 157   |
| 测试集 | 393   |

为了测试改进算法对不同大小船只的检测效果,根据像素值的大小,给出三类船只的定义,并统计了各自在数据集中的数量,如表2所示。

表 2 小中大目标定义

Tab. 2 Definition of small, medium and large goals

| 类型          | 像素值(area)               | 目标数量  |
|-------------|-------------------------|-------|
| 小目标(small)  | $0 < a \leq 32^2$       | 309   |
| 中目标(medium) | $32^2 \leq a \leq 96^2$ | 1 402 |
| 大目标(large)  | $96^2 \leq a$           | 2 920 |

实验主要采用以下评价指标:精确率P(Precision)、召回率R(Recall)、平均精度(Average Precision, AP)、平衡分数F1(F1-Score)。

本文将预测框与真实框之间的重合度交并比(Intersection over Union, IoU)阈值设为0.5,其计算方式如式(14)所示:

$$IoU = \frac{A \cap B}{A \cup B}, \quad (14)$$

其中: $A$ 为真实框面积, $B$ 为预测框面积。如果预测框与真实框之间的 $IoU > 0.5$ ,则将预测结果视为真正的正样本(True Positives, TP);如果预测框与真实框之间的 $IoU < 0.5$ ,则将预测结果视为

假的正样本(False Positives, FP)。如果没有检测出真实目标,称作错的负样本(False Negatives, FN)。精确率 $P$ 、召回率 $R$ 和平衡分数F1值计算方式如式(15)~式(17):

$$P = \frac{TP}{TP + FP} \times 100\%, \quad (15)$$

$$R(\text{recall}) = \frac{TP}{TP + FN} \times 100\%, \quad (16)$$

$$F1 = 2 \times \frac{P \cdot R}{P + R}. \quad (17)$$

平均精度(Average Precision, AP)是精确率-召回率曲线(Precision-Recall Curve)下的面积,其计算方式如式(18):

$$AP = \int_0^1 P(R) dR. \quad (18)$$

在本文中,主要采用AP@0.5, AP@0.5:0.95来评估模型的整体精度,采用APs@0.5:0.95, APm@0.5:0.95, API@0.5:0.95来评估小、中、大目标的检测精度。

每秒帧率(Frame Per Second, FPS)是指画面每秒传输帧数,本文是将测试图片检测100次得到各个模型的FPS值。

### 3.2 实验环境及相关参数

本文的实验环境如表3所示。

表3 实验环境

| 名称   | 型号、参数                 |
|------|-----------------------|
| 操作系统 | Windows10             |
| CPU  | i9-10900 CPU@2.80 GHz |
| RAM  | 32.0 GB               |
| 显卡   | GeForce RTX 3070      |
| DL框架 | Pytorch1.8            |

实验过程借鉴了迁移学习的思想,首先进行了冻结训练,冻结训练中骨干网络采用经过(ImageNet)预训练后的模型,其他网络部分不进行训练,冻结训练50轮,采用预训练模型可以比从零开始训练更能加快模型收敛,减少训练次数;解冻训练250轮,这一过程中,模型中所有参数都会进行训练。具体训练参数如表4所示。

训练过程采用了余弦退火学习率衰减策略,损失函数衰减如图8所示。

表4 训练参数

Tab. 4 Training parameters

| 训练方式 | 训练轮次 | 学习率    | batch_size | 优化器  | 动量   | 置信度阈值 |
|------|------|--------|------------|------|------|-------|
| 冻结训练 | 50   | 0.001  | 32         | Adam | 0.94 | 0.5   |
| 解冻训练 | 250  | 0.0001 | 8          | Adam | 0.94 | 0.5   |

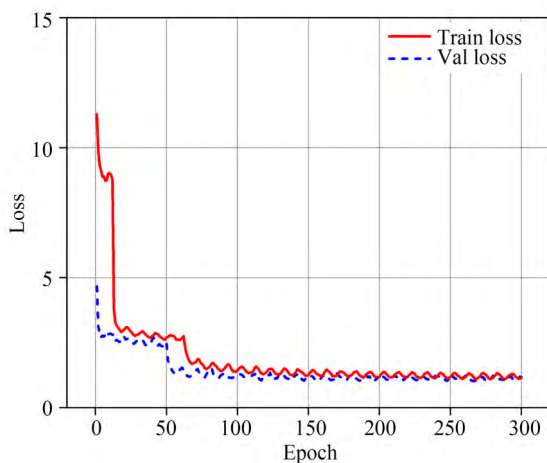


图8 训练损失函数图

Fig. 8 Training loss function diagram

### 3.3 实验结果及对比分析

为验证IM-YOLO-s算法在SORS复杂场景下检测舰船的优越性,通过HRSC2016-Gg数据集进行了评测。同时也与传统经典检测算法与当前主流的基于深度学习的目标检测算法进行了对比,表5展现了具体的实验结果。

根据表中,HOG+SVM的检测效果较差。和一些现有的深度学习方法相比,该方法并不适用于大规模数据集下的舰船检测。一方面该方法的速度较慢,主要是由于候选区域提取采用效率低下的滑窗法。这种直接在全局图像上滑动特定大小的窗口获得候选区域的方法,会大大增加计算的负担。对于一些没有目标的区域,滑窗法没有实际意义。另一方面该方法的检测性能



表 5 模型结果  
Tab. 5 Model results

| 模型         | Backbone      | Recall/%     | Precision/%  | AP@0.5/%     | F1          | Params/MB    | FLOPs/G       | Model Size/kB | FPS       |
|------------|---------------|--------------|--------------|--------------|-------------|--------------|---------------|---------------|-----------|
| HOG+SVM    | /             | 40.97        | 69.99        | 48.36        | 0.52        | /            | /             | /             | 2         |
| FasterRenn | Resnet50      | 91.54        | 70.18        | 89.36        | 0.79        | 28.275       | 454.292       | 110 773       | 17        |
| Yolov3     | Darknet53     | 91.66        | 94.66        | 91.23        | 0.93        | 61.524       | 32.759        | 240 693       | 65        |
| Yolov4     | Darknet53+SPP | 93.65        | 92.89        | 92.97        | 0.93        | 69.338       | 29.879        | 250 265       | 48        |
| CenterNet  | Resnet50      | 83.63        | <b>97.81</b> | 83.63        | 0.90        | 32.664       | 22.125        | 127 932       | <b>73</b> |
| RetinaNet  | Resnet50      | 82.84        | 91.56        | 82.26        | 0.87        | 36.351       | 74.042        | 142 251       | 37        |
| Yolox-s    | CSPDarknet    | 94.95        | 93.09        | 94.37        | 0.94        | <b>8.938</b> | <b>13.317</b> | <b>35 177</b> | 67        |
| IM-YOLO-s  | IM-CSPDarknet | <b>97.18</b> | 92.61        | <b>96.77</b> | <b>0.95</b> | 8.973        | 13.319        | 35 341        | 60        |

较差,主要是由于HOG特征只能描述整体的梯度方向直方图信息,不能很好地捕捉图像内部的上下文语义信息。

对比其他深度学习算法可以得出,IM-YOLOX-s算法具有最高的平均精度、召回率和F1值。与YOLOX-s算法相比,检测准确率略有下降,模型参数量、计算复杂度、模型权重都略有增加,导致FPS值略有降低,但都在可接受范围内并且算法保持着良好的实时性,召回率提升了2.23%,AP@0.5值提升了2.40%,F1值提高了0.01,精确率和召回率达到了最佳的平衡。这表明,IM-YOLO-s算法可以有效解决漏检问题,进行高质量的舰船检测。与CenterNet算法相比,尽管该算法在检测准确率与FPS上达到了上述

所有算法的最优结果,但在本文的主要评价指标AP@0.5与F1值方面均逊色于本文所提算法,在模型参数量、计算复杂度和模型权重方面也远远超过了本文算法。综上,所提IM-YOLOX-s算法可以有效提高舰船检测的召回率,进行高质量的舰船检测。

为了更直观地感受本文算法的改进效果,在测试集图片上进行了逐张检测,图9显示了不同算法在同一测试集图片上的检测结果,其中9(a)是真实框标注图像。

图10展示了真实目标框标注结果,图11~12分别展示了不同场景下YOLOX-s算法与本文的IM-YOLOX-s算法的检测效果。

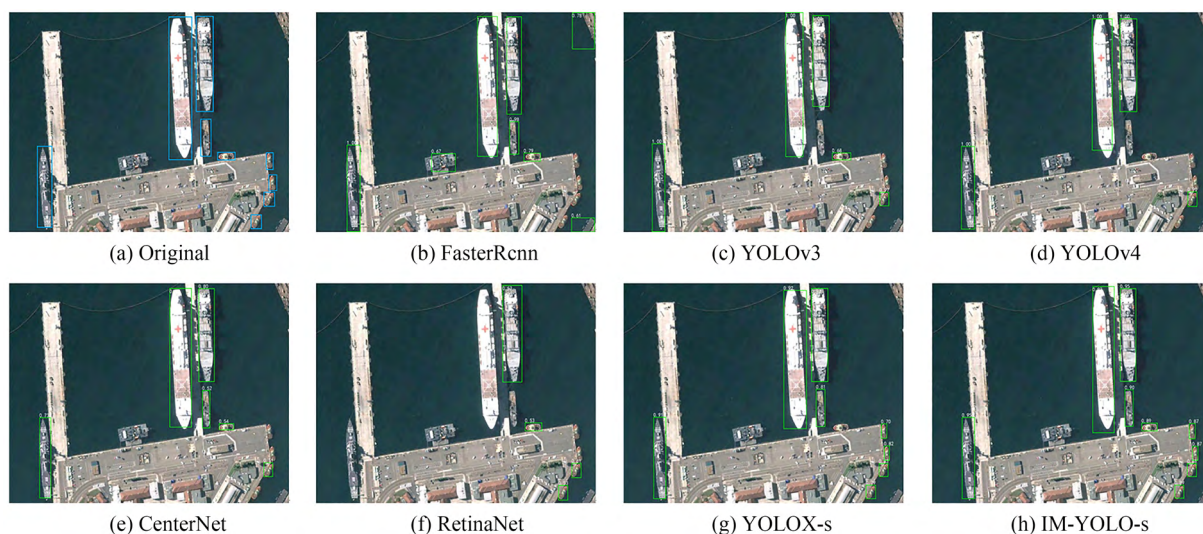


图 9 不同方法实际检测效果对比图

Fig. 9 Comparison chart of actual detection effect of different algorithm

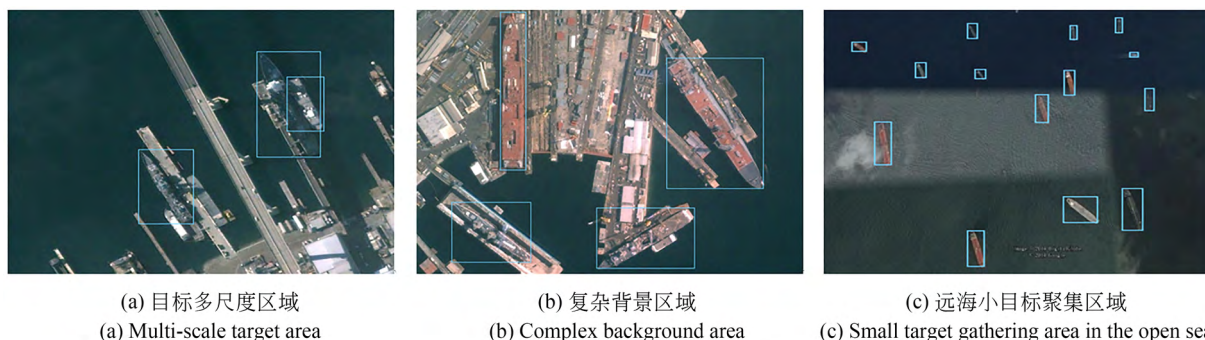


图 10 真实目标框标注效果图

Fig. 10 Real target box annotation renderings

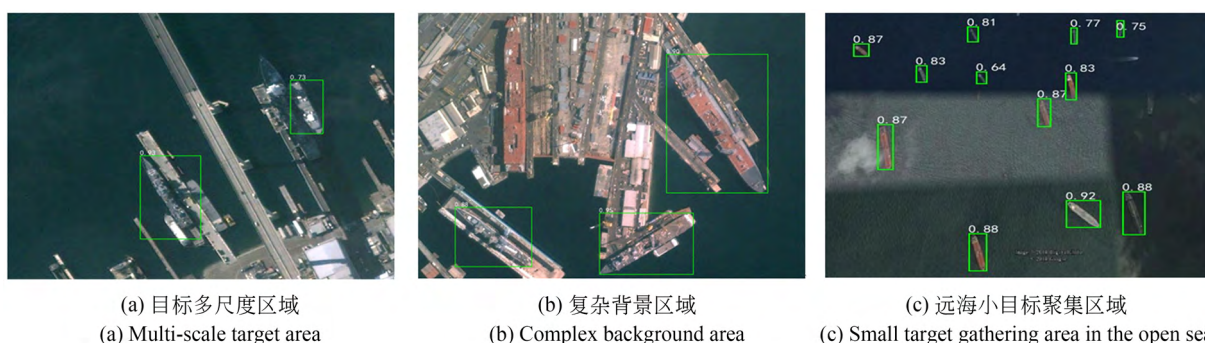


图 11 YOLOX-s 检测结果

Fig. 11 Test results of YOLOX-s

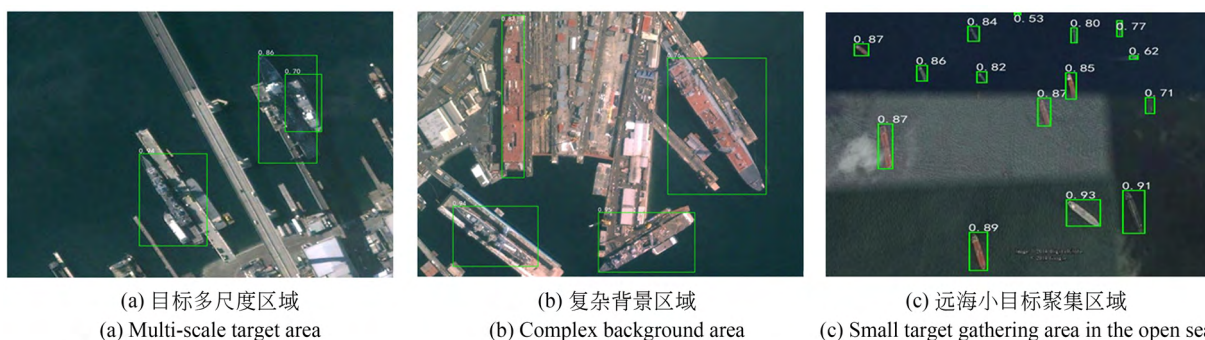


图 12 IM-YOLO-s 检测结果

Fig. 12 Test results of IM-YOLO-s

### 3.4 消融实验

为了定量分析本文方法中融入CA注意力模块、Neck改进、损失函数改进对舰船目标检测精度以及对大中小舰船检测精度的影响,设计了消融实验,将上述改进分别加入到YOLOX-s算法中与原算法进行对比,实验结果如表6所示。

由表6可知,在YOLOX-s中只融入CA注意力模块,AP@0.5提升了1.11%,R提升了1.05%,P基本保持不变,但是FPS下降明显,可

以看出在融入CA模块之后,会导致模型的推断时间提高;只进行Neck改进,AP@0.5提升了0.67%,R提升了0.58%,FPS基本没有下降,并且对小目标船只检测效果提升明显,APs@0.5:0.95提升了0.015,表明改进模型的特征融合方式可以有效提高小目标的检测精度;对于只进行损失函数改进来说,AP@0.5提升了0.89%,R提升了1.05%,其他指标基本保持不变。

在对YOLOX-s同时进行neck改进和融入

表 6 消融实验结果

Tab. 6 Ablation experimental results

| YOLOX-s | CA | Neck<br>改进 | Loss<br>改进 | AP@          | R/%          | P/%          | AP@          | APs@         | APm@         | API@         | FPS/<br>(frame·<br>s <sup>-1</sup> ) |
|---------|----|------------|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------------------------------|
|         |    |            |            | 0.5/%        |              |              | 0.5:         | 0.5:         | 0.5:         | 0.5:         |                                      |
| √       |    |            |            | 94.37        | 94.95        | 93.09        | 0.775        | 0.438        | 0.688        | 0.857        | 67                                   |
| √       | √  |            |            | 95.48        | 96.00        | 93.16        | 0.778        | 0.453        | 0.694        | 0.859        | 61                                   |
| √       |    | √          |            | 95.03        | 95.53        | 93.23        | 0.770        | 0.473        | 0.686        | 0.849        | 66                                   |
| √       |    |            | √          | 95.26        | 96.00        | 91.49        | 0.762        | 0.443        | 0.682        | 0.838        | <b>67</b>                            |
| √       | √  | √          |            | 95.71        | 96.00        | <b>93.48</b> | 0.781        | 0.453        | <b>0.700</b> | 0.861        | 60                                   |
| √       | √  |            | √          | 95.75        | 96.36        | 92.97        | 0.780        | <b>0.479</b> | 0.684        | 0.861        | 61                                   |
| √       |    | √          | √          | 95.27        | 95.65        | 93.14        | <b>0.787</b> | 0.429        | 0.691        | <b>0.876</b> | 66                                   |
| √       | √  | √          | √          | <b>96.77</b> | <b>97.18</b> | 92.61        | 0.786        | 0.442        | 0.692        | 0.869        | 60                                   |

CA 模块之后,改进算法的检测准确度达到了所有消融实验模型的最佳结果,并且中目标检测结果也达到了最好,其他指标均有不同程度的提升,但 FPS 下降到 60;在对 YOLOX-s 同时进行损失函数改进和融入 CA 模块之后,对于小目标船只的检测结果达到了最高;在对 YOLOX-s 进行损失函数改进和融入 CA 模块之后,大目标船只的检测精度达到了最高。

对于 IM-YOLOX-s, AP@0.5 和 P 值达到了最高,表明可以有效解决漏检问题,同时对于不同大小舰船的检测结果也都有提升,综合指标达到了最佳。

综上,可以看出影响 IM-YOLOX-s 推断时间下降的原因主要是 CA 模块的融入,后续将会研究更加轻量级的注意力机制在模型改进方面的作用,同时也观察到不同方法的改进措施会相互制约,从而影响检测效果,后续也会对这一问题

展开探究。

## 4 结 论

本文针对目前大多基于深度学习的舰船检测算法在航天光学遥感复杂场景条件下目标易漏检的问题,提出 IM-YOLO-s 算法,介绍了 IM-YOLO-s 的网络结构和工作原理,研究了 CA 注意力模块在检测中的作用和损失函数的优化算法,并改进了模型的 Neck 结构。给出了融合 CA 注意力模块、加权特征融合的 IM-YOLO-s 算法。实验结果证明:在 HRSC2016-Gg 数据集下平均精度达到了 96.77%,召回率达到了 97.18%,F1 值达到了 0.95,FPS 达到了 60,基本满足了复杂场景下舰船检测的需求,一定程度上降低了模型的漏检率。在未来的工作中,将基于 IM-YOLO-s 算法在物理平台上完成算法的植入,对算法的落地部署进行研究。

## 参考文献:

- [1] 姜鑫,陈武雄,聂海涛,等. 航空遥感影像的实时舰船目标检测[J]. 光学精密工程, 2020, 28(10): 2360-2369.
- JIANG X, CHEN W X, NIE H T, *et al.* Real-time ship target detection based on aerial remote sensing images[J]. *Opt. Precision Eng.*, 2020, 28(10): 2360-2369. (in Chinese)
- [2] 王慧利,朱明,蔺春波,等. 光学遥感图像中复杂海背景下的舰船检测[J]. 光学精密工程, 2018,

26(3): 723-732.

WANG H L, ZHU M, LIN CH B, *et al.* Ship detection of complex sea background in optical remote sensing images[J]. *Opt. Precision Eng.*, 2018, 26(3): 723-732. (in Chinese)

- [3] 丁鹏,张叶,贾平,等. 基于多尺度多特征视觉显著性的海面舰船检测[J]. 光学精密工程, 2017, 25(9): 2461-2468.

DING P, ZHANG Y, JIA P, *et al.* Ship detection on sea surface based on multi-feature and multi-scale visual attention[J]. *Opt. Precision Eng.*, 2017, 25

- (9): 2461-2468. (in Chinese)
- [4] 兰旭婷, 郭中华, 李昌昊. 基于注意力与特征融合的光学遥感图像飞机目标检测[J]. 液晶与显示, 2021, 36(11): 1506-1515.  
LAN X T, GUO ZH H, LI CH H. Attention and feature fusion for aircraft target detection in optical remote sensing images[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1506-1515. (in Chinese)
- [5] 王浩桐, 郭中华. 改进SSD的飞机遥感图像目标检测[J]. 液晶与显示, 2022, 37(1): 116-127.  
WANG H T, GUO ZH H. Improved SSD based aircraft remote sensing image target detection [J]. *Chinese Journal of Liquid Crystals and Displays*, 2022, 37(1): 116-127. (in Chinese)
- [6] 高彦宇, 尹怡欣. 一种基于支持向量机和半监督期望最大化算法的分级图像标识方法[J]. 自动化学报, 2010, 36(7): 960-967.  
GAO Y Y, YIN Y X. A hierarchical image annotation method based on SVM and semi-supervised EM [J]. *Acta Automatica Sinica*, 2010, 36(7): 960-967. (in Chinese)
- [7] DONG C, LIU J H, XU F. Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor [J]. *Remote Sensing*, 2018, 10(3): 400.
- [8] JIANG H Z, LEARNED-MILLER E. Face Detection with the Faster R-CNN[C]. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). May 30 - June 3, 2017, Washington, DC, USA. IEEE, 2017: 650-657.
- [9] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV). December 7-13, 2015, Santiago, Chile. IEEE, 2016: 1440-1448.
- [10] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. IEEE, 2017: 2980-2988.
- [11] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. *arXiv preprint arXiv: 1804.02767*, 2018.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020: arXiv: 2004.10934. <https://arxiv.org/abs/2004.10934>
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[J]. *Computer Vision*, 2016: 21-37.
- [14] WANG Y Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery [J]. *Remote Sensing*, 2019, 11(5): 531.
- [15] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [J]. *Computer Vision*, 2018: 765-781.
- [16] DUAN K W, BAI S, XIE L X, et al. CenterNet: Keypoint Triplets for Object Detection[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27 - November 2, 2019, Seoul, Korea (South). IEEE, 2020: 6568-6577.
- [17] TIAN Z, SHEN C H, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27 - November 2, 2019, Seoul, Korea (South). IEEE, 2020: 9626-9635.
- [18] KONG T, SUN F C, LIU H P, et al. FoveaBox: beyond anchor-based object detection[J]. *IEEE Transactions on Image Processing*, 2020, 29: 7389-7398.
- [19] 王玺坤, 姜宏旭, 林珂玉. 基于改进型YOLO算法的遥感图像舰船检测[J]. 北京航空航天大学学报, 2020, 46(6): 1184-1191.  
WANG X K, JIANG H X, LIN K Y. Remote sensing image ship detection based on modified YOLO algorithm[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2020, 46(6): 1184-1191. (in Chinese)
- [20] 史文旭, 鲍佳慧, 姚宇. 基于深度学习的遥感图像目标检测与识别[J]. 计算机应用, 2020, 40(12): 3558-3562.  
SHI W X, BAO J H, YAO Y. Remote sensing image target detection and identification based on deep learning [J]. *Journal of Computer Applications*, 2020, 40(12): 3558-3562. (in Chinese)
- [21] 焦军峰, 靳国旺, 熊新, 等. 旋转矩形框与CBAM改进RetinaNet的SAR图像近岸舰船检测[J]. 测绘科学技术学报, 2020, 37(6): 603-609.  
JIAO J F, JIN G W, XIONG X, et al. SAR images nearshore ship detection based on RetinaNet

- algorithm with rotated rectangular box[J]. *Journal of Geomatics Science and Technology*, 2020, 37(6): 603-609. (in Chinese)
- [22] GE Z, LIU S, WANG F, *et al.* YOLOX: exceeding YOLO series in 2021 [EB/OL]. 2021: arXiv: 2107.08430. <https://arxiv.org/abs/2107.08430>
- [23] HOU Q B, ZHOU D Q, FENG J S. Coordinate Attention for Efficient Mobile Network Design [C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [24] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993-13000.
- [25] GE Z, LIU S T, LI Z M, *et al.* OTA: Optimal Transport Assignment for Object Detection [C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 303-312.
- [26] HU J, SHEN L, SUN G. Squeeze and Excitation Networks [C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [27] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [J]. *Computer Vision*, 2018: 3-19.
- [28] 邱晓华, 李敏, 邓光芒, 等. 多层卷积特征融合的双波段决策级船舶识别[J]. *光学精密工程*, 2021, 29(1): 183-190.
- QIU X H, LI M, DENG G M, *et al.* Multi-layer convolutional features fusion for dual-band decision-level ship recognition [J]. *Opt. Precision Eng.*, 2021, 29(1): 183-190. (in Chinese)
- [29] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 10778-10787.
- [30] ZHANG H Y, WANG Y, DAYOUB F, *et al.* VarifocalNet: an IoU-aware Dense Object Detector [C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 8510-8519.
- [31] LIU Z K, YUAN L, WENG L B, *et al.* A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines [C]. *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*. February 24-26, 2017. Porto, Portugal. SCITEPRESS-Science and Technology Publications, 2017: 324-331.

## 作者简介:



刘忻伟(1996—),男,山西忻州人,硕士研究生,2019年于内蒙古大学获得学士学位,主要从事图像处理、计算机视觉和机器学习方面的研究。E-mail: 984281928@qq.com

## 通讯作者:



朴永杰(1984—),男,吉林梅河口人,博士,研究员,2006年于吉林大学获得学士学位,2011年于中科院长春光机所获博士学位,目前主要从事星上电子学系统及图像传输处理方面的研究。E-mail: pyj0314@163.com