

基于改进YOLOv5的船舶多尺度SAR图像检测算法

李生辉, 李晓飞, 宋璋晗, 王必祥

(南京邮电大学宽带无线通信技术教育部工程研究中心, 南京 210003)

摘要: 针对复杂场景下合成孔径雷达(Synthetic aperture radar, SAR)图像船舶目标像素尺度差异大和船舶密集排列造成目标漏检的问题, 提出一种基于改进YOLOv5的船舶多尺度SAR图像检测算法。对于YOLOv5的颈部网络, 采用双向特征金字塔结构(Bi-directional feature pyramid network, BiFPN)提升网络多尺度特征融合能力, 并在其自下而上的特征融合支路中, 基于深度可分离卷积(Depthwise separable convolution, DSC)和通道MLP构建EC-MLP(Enhanced channel-MLP)模块, 从而丰富语义信息, 提供更充分的船舶目标上下文特征; 引入全局注意力机制(Global attention mechanism, GAM), 使网络对输入特征进行针对性提取并运算, 减少网络的信息丢失; 此外, 使用SIoU损失函数进一步提高网络的训练收敛速度和检测精度。在SSDD和HRSID数据集上与其他8种方法(Faster R-CNN、Libra R-CNN、FCOS、YOLOv5s、PP-YOLOv2、YOLOX-s、PP-YOLOE-s和YOLOv7-tiny)进行对比实验。实验结果表明: 改进后算法在SSDD数据集上的AP50达到了96.7%, 在HRSID数据集上AP50达到了95.6%, 优于对比方法。

关键词: 合成孔径雷达; 船舶目标检测; 双向特征金字塔网络; 深度可分离卷积; 全局注意力机制
中图分类号: TP753 **文献标志码:** A

Multi-scale SAR Image Detection Algorithm for Ships Based on Improved YOLOv5

LI Shenghui, LI Xiaofei, SONG Zhanghan, WANG Bixiang

(Engineering Research Center of the Ministry of Education for Broadband Wireless Communication Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: An multi-scale synthetic aperture radar (SAR) image detection algorithm for ships based on improved YOLOv5 is proposed to address the large pixel scale difference of ship targets in complex scenes and missed detection caused by dense array of ships. For the neck network of YOLOv5, a bi-directional feature pyramid network (BiFPN) is adopted to enhance the multi-scale feature fusion ability of the network, and an enhanced channel-MLP (EC-MLP) module is constructed based on depthwise separable convolution (DSC) and channel MLP in its bottom-up feature fusion branch to enrich semantic information and provide more sufficient ship target context features. The global attention mechanism (GAM) is introduced to enable the network to extract input features selectively and reduce information reduction. In addition, the SIoU loss function is used to further improve the training convergence speed and detection accuracy of the network. Comparative experiments with eight other methods (Faster R-CNN, Libra R-CNN, FCOS, YOLOv5s, PP-YOLOv2, YOLOX-s, PP-YOLOE-s and YOLOv7-tiny) are

conducted on SSDD and HRSID datasets. The experimental results show that the AP50 of the improved algorithm reaches 96.7% on SSDD and 95.6% on HRSID, which is superior to the comparison methods.

Key words: synthetic aperture radar (SAR); ship target detection; bi-directional feature pyramid network (BiFPN); depthwise separable convolution (DSC); global attention mechanism (GAM)

引 言

船舶目标检测在港口管理、渔业监管、海上救援等民用和军事领域有着广泛应用。合成孔径雷达(Synthetic aperture radar, SAR)是一种高分辨率、宽测绘带的有源微波传感器,成像过程受环境因素影响较小,具有全天候、全天时、工作周期持久等特点,更适合瞬息万变的海洋环境,并且随着 TerraSAR-X、Sentinel-1 和 Gaofen-3 等 SAR 卫星的成功发射, SAR 图像中船舶目标检测已成为世界范围内的研究热点^[1]。

传统的 SAR 图像船舶目标检测方法大多采用图像分割、手工特征提取等,主要依赖于经验,并且检测速度低,如小波变换的检测方法、基于高斯模型的双参数 CFAR、恒虚警率等。近年来,卷积神经网络(Convolutional neural networks, CNN)因其强大的特征提取能力,在计算机视觉领域得到了迅速发展。目前,基于 CNN 的检测方法可以分为两类:单阶段目标检测算法和双阶段目标检测算法。双阶段目标检测算法如 Faster R-CNN^[2]、R-FCN^[3]和 Libra R-CNN^[4]等通过提取感兴趣区域来获得更高的检测精度,但其庞大的计算量难以满足实时检测的需求。单阶段目标检测算法基于回归的方法,直接通过在图像上进行密集采样来对目标进行检测,没有提取感兴趣区域的过程,代表性算法为 YOLO(You only look once)^[5]、FCOS(Fully convolutional one-stage object detection)^[6]等。其中,以 YOLO 系列为代表的单阶段目标检测算法,因检测速度快、效率高、边缘移植简单,得到了广泛应用^[7]。

随着人工智能的兴起,越来越多的学者将计算机视觉领域的目标检测算法引入到 SAR 图像船舶目标检测任务中。但与光学遥感图像中的物体不同, SAR 图像中船舶目标的纹理信息和对比度信息较弱,并且还有如成像散焦、旁瓣效应等特性,因此其检测难度更大。文献[8]发布了第一个开放的 SAR 图像船舶目标检测数据集(SSDD),并提出了一种基于改进 Faster R-CNN 的 SAR 图像船舶目标检测方法。文献[9]提出了一种由特征金字塔网络、旋转区域推荐网络以及多级旋转检测网络三部分构成的 SAR 图像船舶目标检测网络,但其网络结构复杂,检测速度较慢。文献[10]使用深度分离卷积对 YOLO 网络中的原始卷积进行替换来减少网络参数,并结合注意力机制进一步提高船舶目标的检测精度。文献[11]针对复杂近海场景,基于 Darknet-53 和特征金字塔结构提出了一种可以在近岸场景实现船舶高精度检测的算法,但仍无法满足实时检测的需求。

SAR 图像往往具有多种不同的图像分辨率,加之不同船型的尺寸也不尽相同,这就导致同一数据集中,船舶目标的最大像素面积和最小像素面积相差近千倍。因此,如何在保持速度与精度平衡的同时,有效地提高网络的多尺度目标检测能力成为 SAR 图像船舶目标检测的一个难题。文献[12]利用卷积块注意机制来突出特定尺度船舶目标的显著特征,进而提升网络的多尺度检测能力。文献[13]针对特征融合网络提出了多分支扩张模块来增强不同尺度的感受野,能在增强语义信息的同时提高小尺寸船舶的检测能力。文献[14]利用迁移学习和旋转检测框来提高小样本船舶的检测性能,并使用无锚检测框来更好地适应多尺度船舶目标。目前基于深度学习的 SAR 船舶目标检测方法虽然取得了令人瞩目的成果,但复杂场景 SAR 图像中的船舶目标往往尺度差异大,且密集排列,相比其他目标更难检测,因此检测性能仍有很大提升空间。

1 YOLOv5 算法改进

YOLOv5网络分为主干网络、颈部网络和预测网络3个部分,根据网络层数的深度和特征图宽度可以分为s、m、l和x四个版本。为了满足检测速度和准确性,本文选用YOLOv5s作为改进的基线。首先,采用双向特征金字塔结构(Bi-directional feature pyramid network, BiFPN)^[15]来提升网络多尺度特征融合能力,并在自下而上的特征融合支路中增加本文提出的基于深度可分离卷积(Depthwise separable convolution, DSC)^[16]和通道MLP^[17]构建的EC-MLP(Enhanced channel-MLP)模块来提升通道维度间的信息交互,进而丰富语义信息;其次,使用全局注意力机制(Global attention mechanism, GAM)^[18]帮助网络选取关键的输入特征进行运算,更好地结合不同尺度下的目标特征;最后,使用SIoU^[19]损失函数考虑预测框和真实框的角度不匹配问题,进一步提高网络的收敛速度和检测精度。改进网络结构如图1所示。

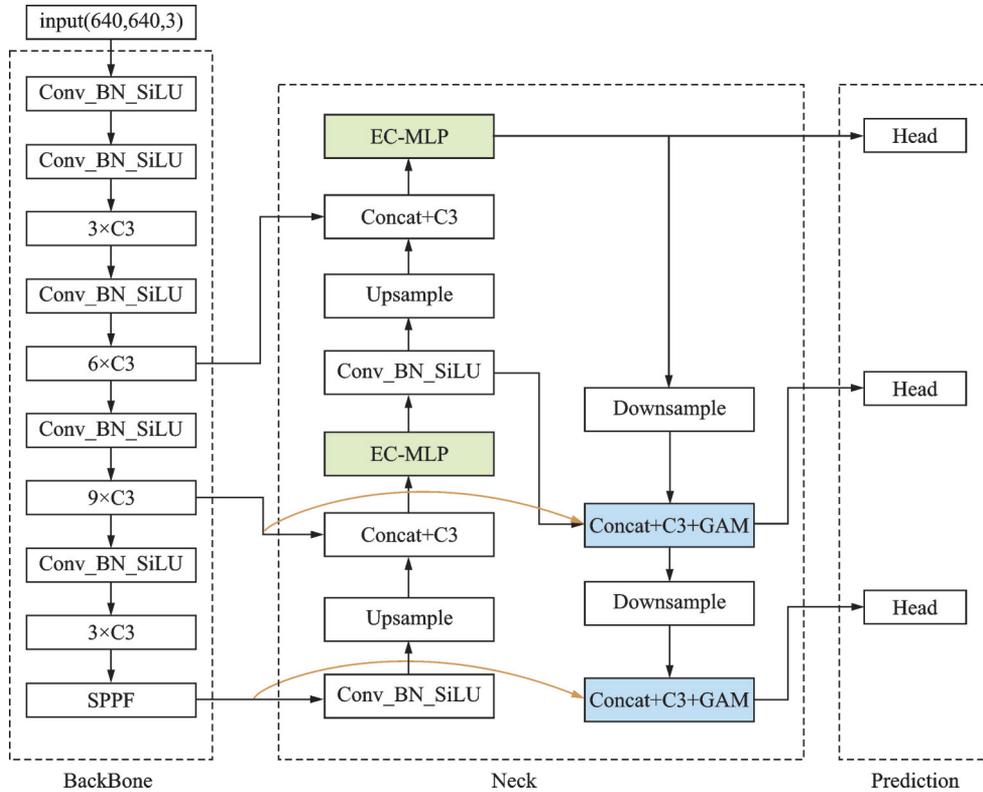


图1 改进网络模型结构

Fig.1 Architecture of improved network model

1.1 颈部网络改进

YOLOv5的颈部网络使用路径聚合网络(Path aggregation network, PANet)^[20]进行特征提取及融合,采用自下而上和自上而下的双向特征融合方法,但由于高分辨率SAR图像中船舶尺度差异大,并且有如岛礁、沿岸港口、船舶密集排列等诸多干扰因素,因此需要融合更多的特征才能提升检测精度。BiFPN是一种双向特征金字塔网络,将同一层的输入和输出节点进行跨层连接,缩短特征信息在高层和低层之间传递的路径,以实现更高层的特征融合。

本文利用BiFPN强大的多尺度特征融合能力,去除其中的加权特征融合,以实现颈部网络轻量化。

在此基础上,在颈部网络自下而上的特征融合支路中增加两组EC-MLP模块(见图1),丰富船舶目标的上下文特征。EC-MLP模块的结构如图2所示。

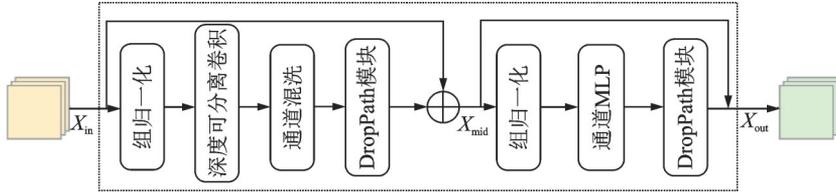


图2 EC-MLP模块结构图

Fig.2 Structure of the proposed EC-MLP module

EC-MLP主要由深度可分离卷积DSC和通道MLP两个模块构成,具体执行流程为:首先对输入特征 X_{in} 进行组归一化(Group normalization)^[21]处理,将其馈送到DSC模块中进一步丰富语义信息,然后对其进行通道混洗(Channel shuffle)^[22]来减少特征通道信息的分离,最后进行DropPath^[23]操作并和残差边进行逐元素相加。上述过程可以表述为

$$X_{mid} = CSDP(DSC(GN(X_{in}))) + X_{in} \tag{1}$$

式中:GN(•)为组归一化,DSC(•)为卷积核大小为 1×1 的深度可分离卷积,CSDP(•)为通道混洗和DropPath操作。

接下来,对输入特征 X_{mid} 进行组归一化处理,然后将输出特征馈送到通道MLP模块,使不同通道维度间的信息相互感知。之后,对输出特征进行DropPath操作并和 X_{mid} 的残差边进行逐元素相加。上述过程可以表示为

$$X_{out} = DP(CMPL(GN(X_{mid}))) + X_{mid} \tag{2}$$

式中:CMPL(•)表示通道MLP,DP(•)表示DropPath操作。

DSC模块由深度卷积和逐点卷积构成,相较于普通卷积,DSC能够在提取特征更深层次语义信息的同时减少参数数量和计算量,但DSC会在计算过程中使输入特征的通道信息分离。为了解决这个问题,本文通过对DSC的输出特征进行通道混洗来进行改进,如图3(a)所示。通道MLP由两个全连接层和GELU激活函数组成,能够增强不同通道特征之间的依赖性,如图3(b)所示。

DSC和通道MLP两个模块的输入特征均会经过组归一化操作,输出特征会经过DropPath操作。组归一化将特征图按照通道进行分组,并在每一个分组内计算各自的均值和方差,分组进行归一化。与批量归一化(Batch normalization)^[24]相比,组归一化不会受每次处理图片的批量大小影响,可以在不同批量大小下保持精度的稳定。

DropPath操作是一种正则化方式,通过对网络的子路径随机删除,可以有效防止过拟合。

1.2 全局注意力机制

注意力机制的合理应用有助于网络选取关键的输入特征进行运算,提高目标检测网络的检测效率,进而提升SAR图像船舶目标检测任务的性能。SAR图像船舶目标的分布较为离散且周围存

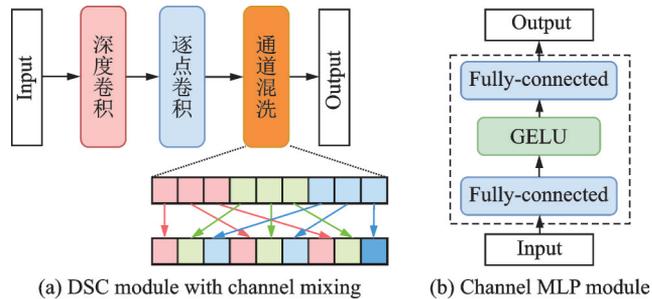


图3 添加通道混洗的DSC模块和通道MLP模块结构图

Fig.3 Structure of the DSC module with channel mixing and channel MLP module

在陆地等干扰因素,因此在颈部网络末端引入集成全局注意力机制 GAM 模块,可以使模型更好地结合不同尺度下的目标特征,帮助网络更好地关注到全局感兴趣区域。

GAM 模块通过强化网络全局性通道和空间的跨维度特征关系,帮助网络选取关键输入特征进行运算,整体结构如图 4 所示。GAM 在 CBAM^[25]的顺序通道-空间注意力机制结构的基础上,重新设计了通道注意力子模块和空间注意力子模块。CBAM 依次执行通道和空间注意力子模块,两个子模块中均采用最大池化和平均池化操作来提取特征图最显著特征及综合考虑所有特征,但忽略了特征图通道和空间方面跨维度交互的重要性,并且最大池化操作会导致特征图的信息缩减。GAM 在通道注意力子模块中使用三维通道置换来保留特征图的空间信息,并通过双层 MLP 强化通道和空间的跨维度特征关系;在空间注意力子模块中使用了两个卷积层代替 CBAM 中的平均池化和最大池化操作,解决最大池化操作带来的特征图信息缩减问题,提取空间关键特征信息。GAM 的执行过程如下

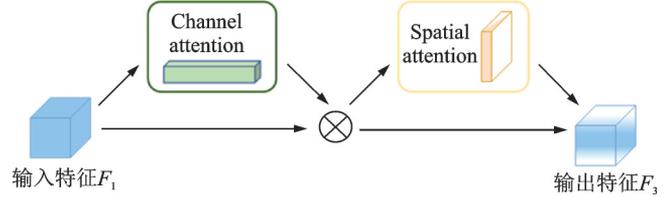


图 4 GAM 模块结构

Fig.4 Structure of the GAM module

$$F_2 = M_c(F_1) \otimes F_1 \quad (3)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (4)$$

式中: F_1 表示输入特征图, F_3 表示输出特征图, M_c 表示通道注意力机制, M_s 表示空间注意力机制, \otimes 表示元素乘法。

GAM 的两个注意力机制子模块具体执行步骤为:(1)通道注意力模块如图 5 所示。首先将维度为 $C \times W \times H$ 的输入通过三维通道置换方式置换为 $W \times H \times C$,然后通过双层 MLP 增强不同维度的通道空间之间的依赖性,最后通过 Sigmoid 激活函数获得权重系数 M_c 。(2)空间注意力模块如图 6 所示。首先将输入特征 F_2 经过一个缩减率为 r 的 7×7 卷积层对通道进行缩减,提取空间共性关键特征信息,然后利用一个 7×7 卷积层恢复特征通道数,最后通过 Sigmoid 激活函数获得权重系数 M_s 。

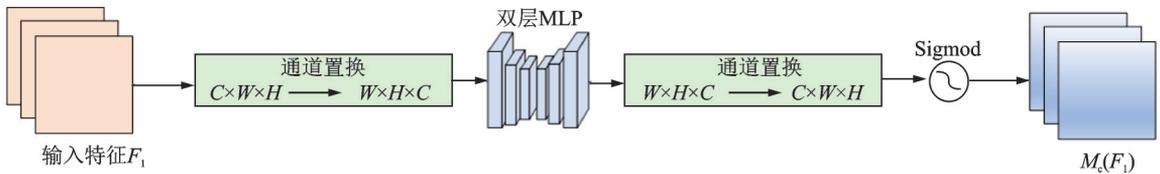


图 5 通道注意力机制

Fig.5 Channel attention mechanism

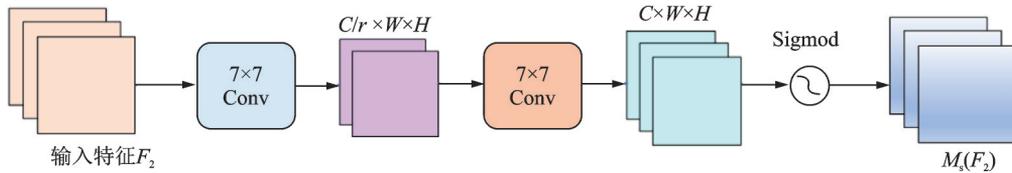


图 6 空间注意力机制

Fig.6 Spatial attention mechanism

1.3 损失函数改进

YOLOv5 采用 CIoU^[26]作为预测框的损失函数,在考虑真实框和预测框的重叠面积及中心坐标的

欧氏距离的基础上增加了对二者长宽比的惩罚因子 v ,使得预测框的回归变得更加稳定,定义如下

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{d^2} + \tau v \quad (5)$$

$$v = \frac{4}{\pi} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (6)$$

$$\tau = \frac{v}{(1 - \text{IoU}) + v} \quad (7)$$

式中: IoU 为真实框和预测框面积的交并比, b 和 b^{gt} 分别表示预测框和真实框的中心坐标, $\rho^2(b, b^{\text{gt}})$ 表示 b 和 b^{gt} 之间的欧氏距离, $w, h, w^{\text{gt}}, h^{\text{gt}}$ 分别表示预测框和真实框的宽和高, τ 为权重参数。

CIoU虽然同时考虑了真实框和预测框的长宽比例以及二者的中心距离,但是没有考虑预测框和真实框的角度不匹配问题,这导致CIoU在训练速度和预测准确性方面不如SIoU。SIoU包含4个部分:IoU损失、距离损失、形状损失以及角度损失,定义如下

$$L_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Omega + \Delta}{2} \quad (8)$$

式中: Ω 表示形状损失, Δ 表示加入了角度损失平衡因子的距离损失,分别定义如下

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (9)$$

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad (10)$$

式中: $\omega_w = \frac{|w - w^{\text{gt}}|}{\max(w, w^{\text{gt}})}$, $\omega_h = \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})}$; θ 表示损失函数对形状损失的关注程度; $\rho_x = \left(\frac{b_{c_x}^{\text{gt}} - b_{c_x}}{c_w} \right)^2$,

$\rho_y = \left(\frac{b_{c_y}^{\text{gt}} - b_{c_y}}{c_h} \right)^2$; γ 用于平衡距离损失和角度损失,定义如下

$$\gamma = 1 + 2 \times \sin^2 \left(\alpha - \frac{\pi}{4} \right) \quad (11)$$

式中: $\alpha = \arcsin \left(\frac{|b_{c_y}^{\text{gt}} - b_{c_y}|}{\sqrt{(b_{c_x}^{\text{gt}} - b_{c_x})^2 + (b_{c_y}^{\text{gt}} - b_{c_y})^2}} \right)$ 表示真实框和预测

框中心坐标之间的角度,如图7所示。可以看出,当 $\alpha \rightarrow 0$ 时,角度损失的占比会上升,距离损失对损失函数的贡献会降低。当 $\alpha \rightarrow \frac{\pi}{4}$ 时,角度损失的占比会减少,损失函数对其的关注度也就越少。

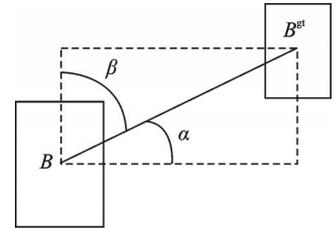


图7 角度损失

Fig.7 Angle cost

2 实验结果与分析

2.1 实验环境

网络结构基于Pytorch1.9.0构建,编程语言为Python3.8,使用的操作系统为Ubuntu 18.04 LTS,使用CUDA10.2和CUDNN7.6.5加速训练,显卡为显存24 GB的GeForce RTX 2080Ti。在消融实验和对比实验中,将整个数据集按照7:2:1的比例随机划分。在网络训练过程中使用SGD优化器对网络参数进行更新,动量参数设为0.9。图片的批处理量batchsize设为32,训练迭代次数Epoch设为400。设置初始学习率为0.01,最终学习率为0.002。训练和测试环境的输入图片大小均为 640×640 。

2.2 数据集简介

选择了HRSID^[27]和SSDD^[8]两个SAR船舶数据集来验证改进YOLOv5s算法的有效性,两个数据集均包括含有船舶目标的近海、远海、海港和海岛等丰富场景。数据集的具体参数如表1所示。

表1 SSDD和HRSID数据集参数信息

Table 1 Parameters about SSDD and HRSID datasets

数据集		HRSID	SSDD
极化		HH, HV, VV	HH, HV, VV, VH
图像数量		5 604	1 160
船舶目标数量		16 956	2 551
图像大小		800×800	500×500等多种尺寸
分辨率/m		0.5, 1, 3	1~15
不同尺寸船舶目标的数量	Small	9 242	1 529
	Medium	7 388	935
	Large	321	76

SSDD数据集是第一个用于SAR船舶目标检测的开放数据集,它的SAR图像来自RadarSat-2卫星、TerraSAR-X卫星以及Sentinel-1卫星,SSDD数据集已经广泛应用于SAR船舶目标检测的研究中。SSDD数据集共有1 160幅图像(包括2 551个船舶目标实例),SAR图像分辨率范围为1~15 m,图像尺寸主要为500×500像素。

HRSID数据集中的图像来自Sentinel-1卫星和TerraSAR-X卫星,它的SAR图像分辨率分别为0.5、1和3 m。HRSID数据集共有5 604幅图像(包括16 965个船舶目标实例),图片的像素大小为800×800。与SSDD数据集相比,HRSID数据集具有更高的分辨率和更丰富的船舶目标实例。

2.3 评价指标

实验的评价指标包括精度(Precision)、召回率(Recall)、平均精度(Average precision, AP)、参数量(Parameters)、浮点运算数(Giga floating point of operations, GFLOPs)以及每秒帧数(Frames per second, FPS)。TP(True positives)表示被正确检测的船舶目标数目, FN(False negatives)表示漏检的船舶目标数目, FP(False positives)表示非船舶目标被错检为船舶目标的数目。精度表示在所有被检测为船舶目标的检测结果中正确检测到的船舶目标比例,召回率表示在所有船舶实例中被正确检测到的船舶目标比例,定义如下

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

AP表示 $P(R)$ 曲线的积分面积。AP50表示IoU阈值为0.5(即,当预测框与真实框的重叠区域超过50%时,就认为该检测框是正确的)时计算的平均精度,AP表示IoU从0.5到0.95的0.05间隔内获得的平均精度的平均值,定义如下

$$\text{AP} = \int_0^1 P(R) dR \quad (14)$$

式中 $P(R)$ 为精度和召回率曲线。

2.4 实验结果分析

在本节中,首先分别在SSDD和HRSID数据集上进行不同改进组合的消融实验,验证本文对YOLOv5s算法改进的有效性,然后将改进算法的效果与Faster R-CNN^[2]、Libra R-CNN^[4]、FCOS^[6]、YOLOv5s、PP-YOLOv2^[28]、YOLOX-s^[29]、PP-YOLOE-s^[30]以及YOLOv7-tiny^[31]进行实验对比。

对各模块进行消融实验评估的结果如表2所示。表中“√”表示添加相应模块,YOLOv5s_A、YOLOv5s_B和YOLOv5s_C分别表示利用BiFPN和EC-MLP模块改进的颈部网络、添加GAM注意力机制和使用Siou损失函数。从表2的实验结果可以看出,颈部网络改进对SAR船舶目标检测效果有较为显著的影响(AP50在SSDD上提高了1.2%,在HRSID上提高了1.1%),原因是颈部网络利用BiFPN和本文提出的EC-MLP模块,提升了网络的全局信息感知能力和多尺度特征融合能力,更有利于不同尺度船舶的检测,有效地解决了诸如陆地干扰、船舶目标密集排列造成目标漏检等问题。应用GAM模块能帮助网络有效选择关键输入特征进行运算,提高网络的检测效率。Siou损失函数在IoU损失、距离损失、形状损失的基础上,引入了预测框和真实框的角度损失,考虑的因素更加综合全面,对训练中损失函数的收敛更加有利。

表2 SSDD和HRSID数据集上的消融实验结果

Table 2 Results of ablation experiments on SSDD and HRSID datasets

模型	改进颈部网络	GAM	Siou loss	SSDD		HRSID	
				AP50	AP	AP50	AP
YOLOv5s				94.6	67.7	93.2	67.9
YOLOv5s_A	√			95.8	68.4	94.3	68.4
YOLOv5s_B	√	√		96.4	68.7	94.8	68.5
YOLOv5s_C	√	√	√	96.7	68.9	95.6	68.7

图8(a)和图8(b)分别为在SSDD和HRSID数据集上训练时预测框Loss的下降曲线,横坐标为训练次数Epoch,纵坐标为预测边框的Loss损失值,可以看出本文方法在使用Siou损失函数后收敛性更好且预测框损失函数的最终值更小。图9(a)和图9(b)分别为在SSDD和HRSID数据集上训练时AP指标的曲线图,其中AP50表示IoU阈值为0.5时对应的AP,可以看出改进算法的AP50值更高。综上所述,与YOLOv5s方法相比,改进后算法损失函数的收敛性和检测准确率均有更优的表现。

与其他8种算法的定量对比结果如表3所示,其中加粗字体表示各列最优结果。从整体上看,本文

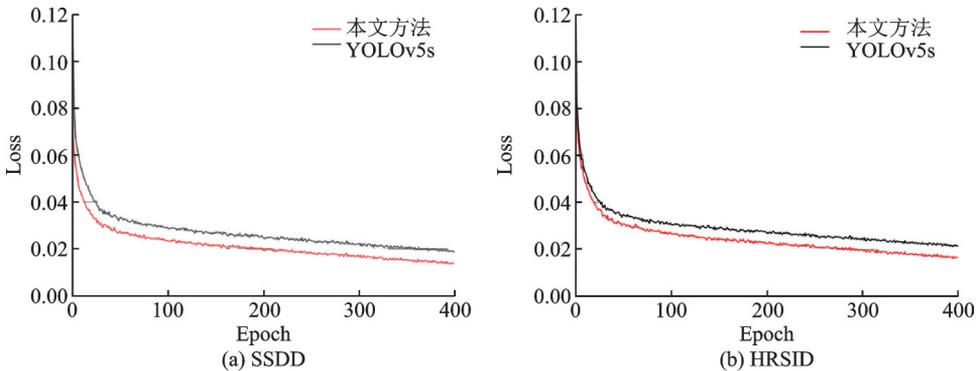


图8 SSDD和HRSID数据集上预测框Loss下降曲线

Fig.8 Loss decline curves of prediction box on SSDD and HRSID datasets

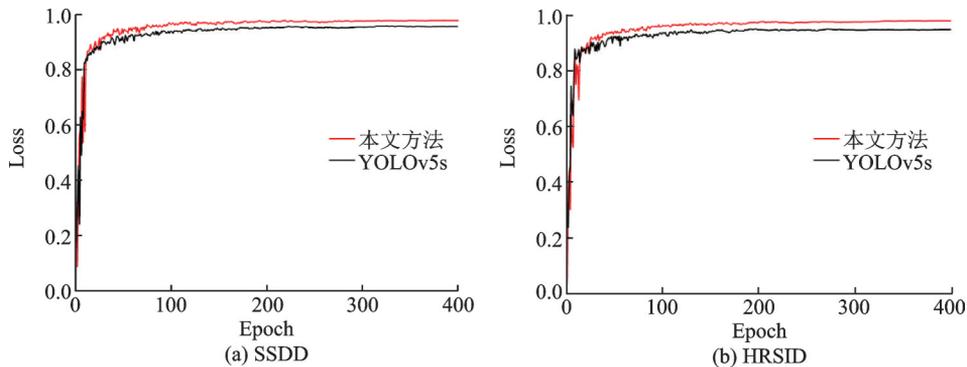


图9 SSDD和HRSID数据集上AP50训练曲线

Fig.9 AP50 training curves on SSDD and HRSID datasets

表3 SSDD和HRSID数据集上对比实验结果

Table 3 Comparative experimental results on SSDD and HRSID datasets

模型	Params/MB	GFLOPs/GB	FPS	SSDD				HRSID			
				P	R	AP50	AP	P	R	AP50	AP
Faster R-CNN	28.27	272.46	12.6	82.4	85.6	92.6	64.9	85.6	87.2	89.2	67.4
Libra R-CNN	34.34	293.58	11.3	90.8	89.7	93.4	65.8	87.2	89.6	90.6	68.1
FCOS	41.12	126.38	32.6	91.2	90.3	92.8	65.2	89.6	88.4	89.3	65.6
PP-YOLOv2	52.67	61.26	69.4	92.6	90.1	95.6	66.9	93.1	87.5	94.8	67.1
YOLOv5s	7.23	16.42	89.3	93.2	88.7	94.6	66.3	92.9	87.2	93.2	67.9
YOLOX-s	9.12	26.72	92.6	94.8	89.3	96.1	68.6	93.9	89.4	95.3	68.4
PP-YOLOE-s	7.93	18.56	83.2	94.6	90.5	95.6	68.7	93.8	89.3	95.1	68.6
YOLOv7-tiny	6.41	14.23	103.4	94.3	90.8	95.8	68.3	94.2	89.1	94.9	68.1
本文方法	9.87	21.65	76.8	95.1	91.2	96.7	68.9	94.4	90.2	95.6	68.7

方法检测效果稳定,并可以达到实时检测的要求。在SSDD和HRSID数据集上的实验结果表明,相较于其他8种算法,本文改进YOLOv5s算法达到了最佳检测精度。与基线YOLOv5s相比,改进算法的参数量和GFLOPs分别增加了2.67 MB和6.23 GB,但精度和AP50均有显著的提升。具体而言,精度在SSDD上从93.2%提升到95.1%,在HRSID上从92.9%提升到94.4%;AP50在SSDD数据集上达到了96.7%,相较于原算法提升了2.1%,在HRSID上达到了95.6%,相较于原算法提升了2.4%。在实验过程中,YOLOv5s的检测帧数为89.6 FPS,本文方法的检测帧数(76.8 FPS)虽然较之减慢了12.5 FPS,但在增加网络复杂性为代价获得更高精度的同时,依然能达到实时监测的要求。本文方法在参数量、GFLOPs上高于PP-YOLOE-s、YOLOv7-tiny等算法,但AP50在SSDD和HRSID数据集上相较于先进的PP-YOLOE-s和YOLOv7-tiny算法分别提升了1.1%、0.9%和0.5%、0.7%,且精度、召回率、AP等指标也有提升,并且本文方法的检测帧率可以满足实时检测的需求。因此,本文方法在准确性和速度的平衡方面优于对比算法,进一步证明改进方法的有效性。

图10(a~d)分别为本文方法对离岸船舶、近岸船舶、大型船舶和密集分布的小型船舶的检测结果,其中绿线框为真实船舶目标,红线框为本文方法的检测结果。图片中共有357个船舶目标,其中346个船舶目标被正确检测到,11个船舶目标被漏检,10个船舶目标被误检。准确率为97.2%,召回率为

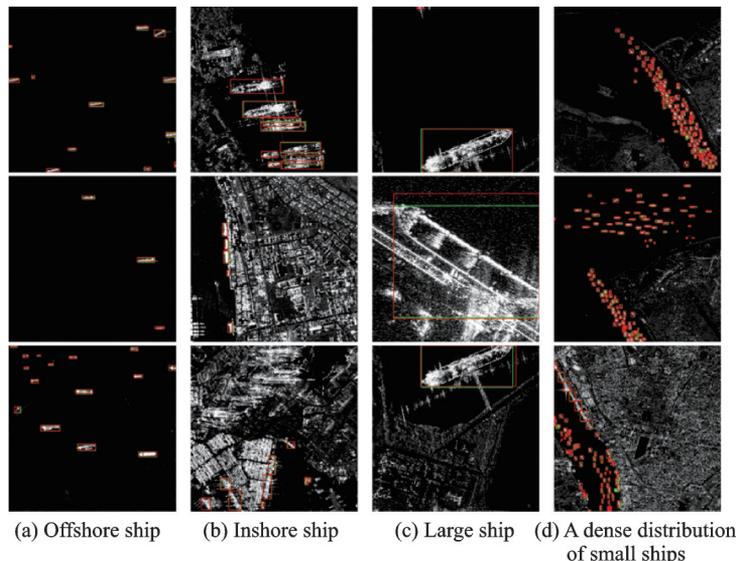


图10 改进方法检测效果

Fig.10 Detection results of improved method

96.9%。可以看出,本文方法对离岸船舶和近岸船舶的检测精度均较高,这表明改进后的网络可以有效区分如陆地、港口等复杂背景。同时,对大型船舶和密集分布的小型船舶的检测结果也相对稳定,这表明改进后的网络具有更强大的特征提取和特征融合能力,可以降低复杂背景的噪声干扰并降低船舶小目标密集排列漏检和误检的可能性。但在船舶目标密集排列以及周围具有复杂背景的情况下,仍会导致漏检和误检。船舶目标周围存在与其形状、大小相似的海岛等背景目标时容易发生误检;多个船舶目标密集排列时,目标检测网络易将多个目标识别为一个,导致漏检和误检数的增加。

3 结束语

综上所述,针对SAR图像船舶目标像素尺度差异大和船舶密集排列造成目标漏检的问题,本文提出一种基于改进YOLOv5的多尺度SAR图像船舶目标检测算法,在满足实时检测的条件下,获得了更高的检测精度。本文方法以YOLOv5s作为改进基线,采用BiFPN提升网络多尺度特征融合能力,并在自下而上的特征融合支路中,增加基于深度可分离卷积DSC和通道MLP构建的EC-MLP模块,从而丰富语义信息,提供更充分的船舶目标上下文特征;在网络模型末端引入GAM模块,使网络对输入特征进行针对性提取并运算,减少网络的信息丢失;利用SIoU损失函数考虑预测框和真实框的角度不匹配问题来提高网络训练的收敛速度和检测精度。在SSDD和HRSID数据集的实验结果表明,相较原YOLOv5s模型,本文算法对离岸船舶、近岸船舶、大型船舶和密集分布的小型船舶的多尺度船舶目标,均能保证更高的检测精度,能够适应未来要求更高的船舶目标检测任务。

在实验过程中发现,数据集部分图像的分割不当会导致图像边缘出现误检和漏检的情况。在未来的工作中会尝试对图像分割方法进行优化,同时也将使用高分辨率更高的SAR图像数据集对网络进行训练,进一步优化网络对多尺度目标和密集排列小目标的检测能力。

参考文献:

[1] 朱岱寅,耿哲,俞翔,等. 地面目标多角度 SAR 数据集构建与目标识别方法[J]. 南京航空航天大学学报, 2022, 54(5): 985-994.
 ZHU Daiyin, GENG Zhe, YU Xiang, et al. SAR database construction for ground targets at multiple angles and target recogni-

- tion method[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2022, 54(5): 985-994.
- [2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *Advances in Neural Information Processing Systems*, 2016, 39(6): 1137-1149.
- [3] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[J]. *Advances in Neural Information Processing Systems*, 2016(10): 379-387.
- [4] PANG J, CHEN K, SHI J, et al. Libra R-CNN: Towards balanced learning for object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2019: 821-830.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2016: 779-788.
- [6] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.]: IEEE, 2019: 9627-9636.
- [7] 张睿, 王尧尧, 段雅琦, 等. 面向飞行机械臂的实时目标检测与定位算法[J]. *南京航空航天大学学报*, 2022, 54(1): 27-33.
ZHANG Rui, WANG Yaoyao, DUAN Yaqi, et al. Real-time object detection and location algorithm for aerial manipulator[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2022, 54(1): 27-33.
- [8] LI J, QU C, SHAO J. Ship detection in SAR images based on an improved faster R-CNN[C]//*Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*. [S.l.]: IEEE, 2017: 1-6.
- [9] PAN Z, YANG R, ZHANG Z. MSR2N: Multi-stage rotational region based network for arbitrary-oriented ship detection in SAR images[J]. *Sensors*, 2020, 20(8): 2340.
- [10] 张晓玲, 张天文, 师君, 等. 基于深度分离卷积神经网络的高速高精度 SAR 舰船检测[J]. *雷达学报*, 2019, 8(6): 841-851.
ZHANG Xiaoling, ZHANG Tianwen, SHI Jun, et al. High speed and high precision SAR ship detection based on deep separation convolutional neural network[J]. *Journal of Radar Science*, 2019, 8(6): 841-851.
- [11] CHEN Y, DUAN T, WANG C, et al. End-to-end ship detection in SAR images for complex scenes based on deep CNNs[J]. *Journal of Sensors*, 2021(3): 1-19.
- [12] CUI Z, LI Q, CAO Z, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 8983-8997.
- [13] GUO W, LI W, GONG W, et al. Extended feature pyramid network with adaptive scale training strategy and anchors for object detection in aerial images[J]. *Remote Sensing*, 2020, 12(5): 784.
- [14] AN Q, PAN Z, YOU H, et al. Transitive transfer learning-based anchor free rotatable detector for SAR target detection with few samples[J]. *IEEE Access*, 2021, 9: 24011-24025.
- [15] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2020: 10781-10790.
- [16] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. *arXiv preprint arXiv:1704.04861*, 2017.
- [17] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: An all-MLP architecture for vision[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 24261-24272.
- [18] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. *arXiv preprint arXiv:2112.05561*, 2021.
- [19] GEVORGYAN Z. SIOU Loss: More powerful learning for bounding box regression[J]. *arXiv preprint arXiv:2205.12740*, 2022.
- [20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018: 8759-8768.
- [21] WU Y, HE K. Group normalization[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.]: ECCV, 2018: 3-19.
- [22] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018: 6848-6856.
- [23] LARSSON G, MAIRE M, SHAKHNAROVICH G. Fractalnet: Ultra-deep neural networks without residuals[J]. *arXiv preprint arXiv:1605.07648*, 2016.

- [24] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. [S.l.]: PMLR, 2015: 448-456.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCV, 2018: 3-19.
- [26] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52: 8574-8586.
- [27] WEI S, ZENG X, QU Q, et al. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation [J]. Ieee Access, 2020, 8: 120234-120254.
- [28] HUANG X, WANG X, LV W, et al. PP-YOLOv2: A practical object detector[J]. arXiv preprint arXiv:2104.10419, 2021.
- [29] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.
- [30] XU S, WANG X, LV W, et al. PP-YOLOE: An evolved version of YOLO[J]. arXiv preprint arXiv:2203.16250, 2022.
- [31] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprint arXiv:2207.02696, 2022.

作者简介:



李生辉(1998-),男,硕士研究生,研究方向:计算机视觉、目标检测, E-mail: 1020010508@njupt.edu.cn。



李晓飞(1964-),通信作者,男,教授,研究方向:广域目标跟踪、视频分析, E-mail: lixf@njupt.edu.cn。



宋璋晗(1998-),女,硕士研究生,研究方向:图像处理、图像超分辨率重建, E-mail: 1020010519@njupt.edu.cn。



王必祥(1999-),男,硕士研究生,研究方向:计算机视觉、姿态识别, E-mail: 1220013630@njupt.edu.cn。

(编辑:夏道家)