

DOI:10.19650/j.cnki.cjsi.J2413287

海面无人艇对舰船目标的细粒度检测方法^{*}

左震, 郭润泽, 孙备, 苏绍璟, 孙晓永
(国防科技大学智能科学学院 长沙 410072)

摘要:真实海面场景中舰船目标外观相似、边缘信息模糊, 现有算法无法满足海上精细、实时的细粒度检测需求。故提出基于多尺度坐标注意力和多网络自监督学习的舰船目标细粒度检测方法。首先, 设计多尺度坐标注意力和多网络自监督学习模块, 在原有特征金字塔和路径聚合网络的基础上, 进行特征增强, 提高海面场景下舰船目标的细粒度检测精度; 其次, 构建了基于光电吊舱、电子罗盘的无人艇视觉感知平台, 制备了包含渔船、快艇、商船等不同类别的舰船目标数据集; 最后, 在公开数据集和自制数据集上对本文算法进行了测试和集成验证。结果表明, 算法对舰船目标具有较高的检测精度, 真实海面场景下平均精度均值 (mAP) mAP@0.5 达到 94.6%, 相较于改进前提升了 1.1%, 运行速度 27 fps, 满足了海面无人艇鲁棒、实时的检测需求。

关键词: 舰船目标; 细粒度检测; 无人艇感知平台; 多尺度坐标注意力; 多网络自监督学习

中图分类号: TP391.4 TH865 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Fine-grained detection of ship objects by unmanned surface vehicles

Zuo Zhen, Guo Runze, Sun Bei, Su Shaojing, Sun Xiaoyong

(National University of Defense Technology, College of Intelligent Science, Changsha 410072, China)

Abstract: In real sea scenes, the appearance of ship targets is similar and the edge information is blurred. The existing algorithms cannot meet the demands for fine-grained and real-time detection at sea. Therefore, a fine-grained detection method is proposed for ship objects based on multi-scale coordinate attention and multi-network self-supervised learning. First, a multi-scale coordinate attention and multi-network self-supervised learning module is designed. Feature enhancement is carried out on the basis of the original feature pyramid network and path aggregation network to improve fine-grained detection accuracy. Secondly, an unmanned surface vehicle (USV) sensing platform based on pods and electronic compass is constructed, and a dataset containing different ship objects such as fishing boats, speedboats, and merchant ships is prepared. Finally, the algorithm is tested and integrated into public and self-made datasets. The results show that the proposed algorithm has high detection accuracy for ship targets. The mAP50 reaches 94.6% in the real sea scene, which is 1.1% higher than that before the improvement. The operation speed is 27 fps, which verifies the robust and real-time fine-grained detection capability of USVs.

Keywords: ship objects; fine-grained detection; sensing platform; multi-scale coordinate attention; multi-network self-supervised learning

0 引言

随着内陆经济加速转型, 推进海洋经济发展、维护海洋生态环境以及助力海洋强国建设已然成为我国现代化进程中的关键目标。怎样构建现代海洋监管体系, 精准识别海面舰船目标已成为众多机构聚焦的研究关键所在。

目前学术界主要借助卫星、无人机等遥感手段探舰船目标^[1]。宋志娜等^[2]提出一种极化增强检测算法, 借助多尺度特征融合达成在复杂背景下对舰船目标的检测; 张涛等^[3]针对检测算法推理能力欠佳的状况, 推出基于关键点的检测策略, 提升了模型推理的速率与精准度; 李登峰等^[4]设计轻量级注意力模块来平衡网络的精度和

速度;范云升等^[5]针对遥感场景舰船目标检测问题,改良YOLOv5 检测器以消除累积误差,有力抑制了海背景的干扰。但是,上述遥感方法存在数据处理工作量大、时间分辨率欠缺,难以持续观测,无法契合实时感知需求等问题。相较而言,无人艇部署灵活,能够自主航行与巡逻,其采集的图像能更清晰地展现目标的外观与结构,并且在处理原始影像时无需考虑几何和辐射误差,在舰船目标识别领域具备其他数据源难以企及的长处。特别是随着光学成像载荷的迅猛发展,无人艇搭载的可见光图像的清晰度与细节分辨率持续攀升,为舰船目标识别筑牢了坚实的数据根基^[6]。

然而,伴随任务环境与任务内容愈发复杂多样,现有的基于深度学习的一阶段或两阶段目标检测算法^[7-13]仅能给出目标的粗略分类结果——船舶目标,无法进一步判定其属于军用还是民用、有无威胁,难以满足更为精细的识别需求。因此,本文着重研究细粒度检测技术,捕捉并学习目标局部区域的细微差别,以给出更为精准的分类结果^[14]。

海面无人艇针对舰船目标的细粒度检测是一项富有创新性、挑战性且极具研究价值的工作。本文设计了一种基于多尺度坐标注意力与多网络自监督学习的细粒度检测方法,凭借注意力的横向衔接与局部区域信息强化,提升了真实场景下舰船的目标检测精度;搭建无人艇感知平台,借助 HEOS-20U 型光电吊舱达成海面数据的实时采集;制备涵盖渔船、快艇、商船等不同类型的舰船目标数据集,并在公开数据集与自制数据集上展开对比与全流程验证。初步证实了海面场景下无人艇对舰船目标稳健且实时的检测能力。

1 舰船目标细粒度检测相关工作

海面场景中渔船、快艇等舰船目标因其尺寸和姿态多变,艇载算法的性能面临更高的要求。

1.1 舰船目标特性分析

在复杂的海洋环境里,各种不可控和非线性的船舶特质造成舰船目标类间区别小、类内差别大,大幅提升了细粒度检测的难度。无人艇视角下海面场景中的舰船目标图像,涵盖了渔船、快艇、商船等在不同海况下的外观差异如图 1 所示。由图 1 可见,受诸多因素影响,海面场景可见光图像存在背景复杂、干扰多且目标纹理可变等难点,原因如下:1)太阳光反射、海洋表面反射、复杂气候状况等因素致使图像分辨率降低、目标色彩减弱且轮廓边缘模糊,进而难以提取有效特征;2)由于拍摄角度的不断改变,即便是相同类型的目标,在各异的视角下亦会呈现出较大差异。

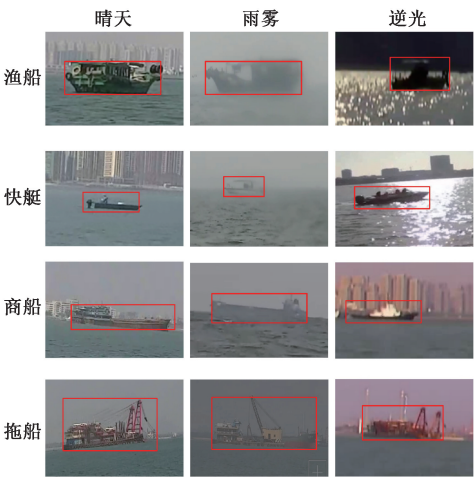


图 1 不同场景下的舰船目标
Fig. 1 Ship object in different scenarios

鉴于拍摄距离也存在差异,相同目标的尺度也发生较大变化。真实场景中典型舰船的尺寸和目标数量分布情况如图 2 所示。

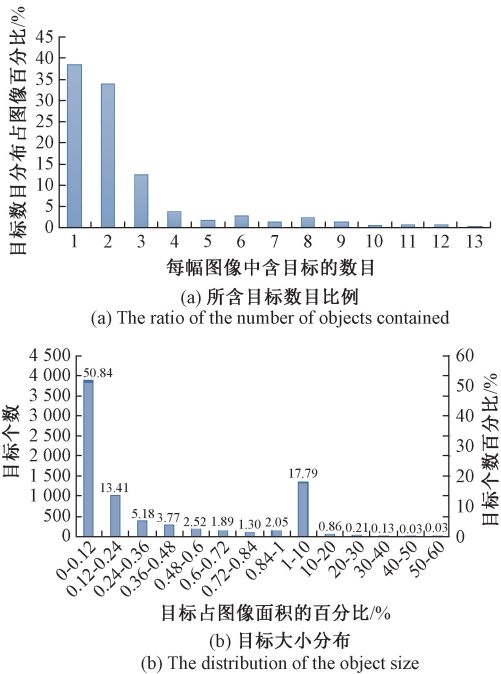


图 2 舰船目标数据集的特性统计
Fig. 2 Characteristic statistics of the datasets of ship objects

可以发现,约 85% 的图像含 1~4 个目标,排列不密集。然而,舰船在尺寸方面呈现出较大的差异性,目标在图像中的占比从 0.01%~58% 均有分布,其中约有 50.8% 的目标仅占图像大小的 0.12%,这种情况进一步提升了目标检测和分类的难度。所以,本文基础网络要

具备多尺度识别能力,增强对弱纹理目标的特征提取能力,实现对舰船目标的鲁棒检测的本文的关键。

1.2 细粒度检测研究进展

细粒度检测是在常规检测的基础之上,针对目标进行更为详尽的类别划分^[15],其目的在于明确并辨识不同子类目标在外观上的细微差别^[16]。当下,学术界常采用强监督或者弱监督的形式来定位判别性区域,接着再开展细粒度识别工作^[17-18]。强监督方式借助边框与关键点等人工标注信息,运用区域定位、对准等方式提升识别的精准度。不过,强监督方式对人工注释信息有所依赖,所需成本高昂,并不适用于数据相对稀缺的舰船目标识别。而弱监督方式不依靠注释信息,借助网络结构的优化和注意力机制的引入,学习精细的局部特征,极大提高了细粒度识别精度。然而,不管是弱监督方法还是强监督方法,都只适用于包含单个目标的图像,难以满足海面场景下密集的舰船目标检测需求。所以,本文通过两个阶段实现细粒度检测任务。首先,利用基于多尺度坐标注意力的检测算法实时获取目标的位置信息,裁剪获得仅仅包含目标的图像;其次,设计多网络自监督学习模块,得到舰船目标的细粒度检测结果。

2 无人艇视觉感知平台搭建

为使无人艇能够在极为多变且不可预估的未知环境中自主航行,平台必须具备可靠、鲁棒的环境感知能力^[19]。为达成真实海面场景中舰船目标的精细化检测,本文设计并构建了一套无人艇视觉感知系统,可实时采集和处理舰船数据。实验平台如图 3 所示,各传感器在无人艇上的安装位点和实物景象,包括光电吊舱、航海雷达、差分 GPS 和通信天线等。

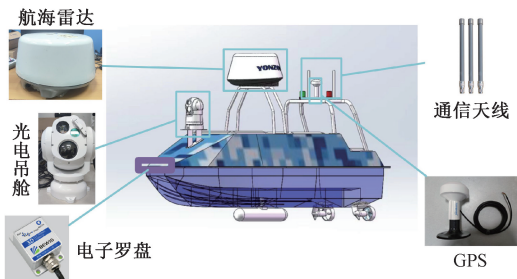


图 3 实验平台实物

Fig. 3 The physical image of experimental platform

系统中,光电吊舱采用 HEOS-20U 型,包含前端摄像头和综合控制箱两部分,能够同时采集可见光、红外两路。航海雷达可在能见度较低或者与目标距离较远的情况下,对光电吊舱的探测工作予以辅助。差分 GPS 可

以获取无人艇的实时经纬度和艏向角,而通信天线主要用于艇间或岸艇间的通联通信。

3 舰船目标数据集制作

本文制备的数据集包含两个部分,舰船目标检测数据集和细粒度分类数据集。目标检测数据集由搭载光电吊舱的无人艇视觉平台采集得到。在采集时,持续调整拍摄角度和距离,由此得到涵盖快艇、渔船、商船、军舰、桥墩等多类目标的海面场景视频。随后,截取图像帧,人工筛选掉相似度较高的图像,进而形成目标检测数据集。目标检测数据集如图 4 所示,总计有 2 229 张图像,数据场景包含晴天、雨雾天、逆光场景。



图 4 舰船目标检测数据集

Fig. 4 The dataset of ship object detection

采用人工标注的方式,得到待测目标类别、位置等信息的标注文件“.xml”。之后依据位置信息对图像进行裁剪与分割,进而获得只有舰船目标的细粒度分类数据集如图 5 所示。从图 5 可以看到,数据集中囊括了快艇、渔船、商船、浮标等多种舰船目标,共记 5 386 张图像。值得注意的事,其中快艇和渔船在特征方面记为相似,给二者的区分带来了较大困难。



图 5 细粒度分类数据集

Fig. 5 The datasets of fine-grained classification

不同类型舰船目标的数量分布如图 6 所示。从图 6 可以看出,不同子类的样本数量呈现出较大的不平衡性。例如,桥墩的数量占比最高,达 2 000。商船、水鼓、浮标

的数量不足 500。相差数倍的样本分布会加剧网络的不均衡,训练网络极易出现过拟合。因此,本文在设计网络和损失函数时,充分考虑了样本分布的不均衡性,引入了额外的超参数。

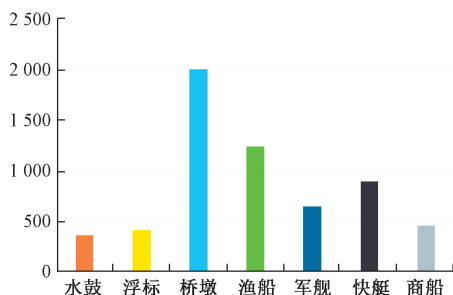


图6 舰船目标数量分布

Fig. 6 Quantity of ship objects

4 基于多尺度坐标注意力和多网络自监督学习的细粒度检测方法

为满足典型海面场景下舰船目标的实时检测需求,解决细粒度识别方法难以精确定位、目标检测方法精细化分类精度不高等难点,本文选取兼具检测精度和速度优势的 YOLOv8 为基础检测网络,并提出一种基于多尺度坐标注意力和多网络自监督学习的细粒度检测方案。首先,设计多尺度坐标注意力模块,以此减轻因尺度变动以及目标纹理特征微弱而引发的漏检状况,从而精准锁定舰船目标;其次,通过目标裁剪与分割,滤除不相干的背景信息,获取仅包含舰船目标的局部图像;最终,图像输入至自监督学习框架中,得到细粒度检测结果。

4.1 基于多尺度坐标注意力的目标区域提取

YOLOv8 属于单相检测算法,它的处理速度相当迅速,可以满足实时检测的需求。不过,海面场景具有高度的复杂性与多变性,雨雾、光照等因素会极大地干扰成像的清晰度,并且尺度的变化也极易导致目标出现漏检。基于此,本文以 YOLOv8 检测网络为依托,设计了基于多尺度坐标注意力的目标区域提取方法。通过对目标不同尺度特征的增强,实现了稳定且精准的特征提取和区域定位。

1) 多尺度坐标注意力模块

无人艇视角下舰船目标存在色彩和轮廓特征不够突出,尺度变化大等特点,难以有效定位。在确保算法实时性的前提下,为捕捉舰船目标的位置感知信息,聚焦辨别性特征,本文在轻量化坐标注意力的基础上进一步改进^[20],设计多尺度坐标注意力模块,在全面挖掘图像跨通道和多尺度信息的过程中,缓解神经网络的池化操作带来的位置信息缺失。

为防止原始信息过多丢失,本文方法引入残差模块以“跳跃连接”的方式进行级联。图 7(a)、(b)所示分别为使用残差模块前后本文方法的结构,残差模块的融入让深层的注意力模块可以与浅层特征实现级联,确保了梯度传递过程中信息的完整性。由图 7 可知,本文方法通过坐标信息嵌入和坐标信息生成两个步骤高效地将空间坐标信息融入生成的注意图中,让网络能够精准地锁定关键区域,降低由于目标模糊而引发的漏检概率,实现对舰船目标的有效聚焦。

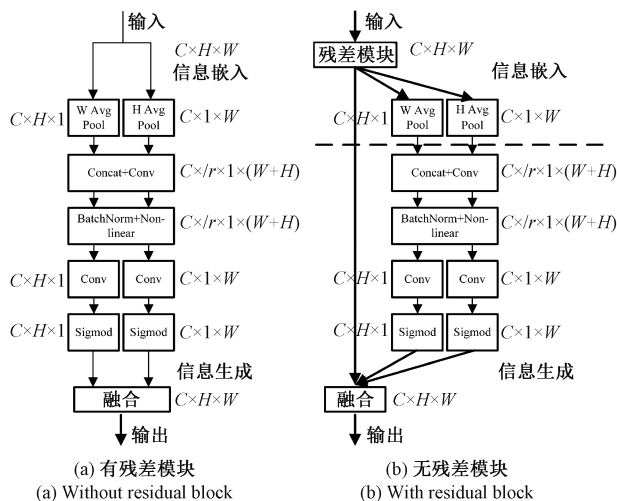


图7 残差坐标注意力结构

Fig. 7 Structure of residual coordinate attention

注意力结构中,池化操作通常用于全局信息的编码。然而,此过程中,局部位置信息难以保留。因此,在坐标信息嵌入阶段,分别使用特定大小的卷积核对输入特征 X 沿垂直和水平方向进行编码,以此输出一组不同方向的特征图。

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

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

式中: $z_c^h(h)$ 表示高度为 h 通道为 c 的特征图输出; $z_c^w(w)$ 表示宽度为 w 通道为 c 的特征图输出。为学习不同通道间的空间依赖关系,在坐标信息生成阶段,模块使用 1×1 的共享卷积函数 C_1 将水平和垂直方向的空间编码信息并联,生成特征图 $c \in \mathbf{R}^{C/r \times (H+W)}$,如式(3)所示。其中 r 是缩减注意力模型的比例,用来 $c^h \in \mathbf{R}^{C \times H}$ 减少计算开销。随后,将 c 分解为两个独立的张量和 $c^w \in \mathbf{R}^{C \times W}$,并利用卷积函数 C_h, C_w 转换为与输入 X 通道数相同的张量 g_c^h 和 g_c^w 。最后,张量与初始特征相乘,得到坐标注意力的输出特征,如式(6)所示。

$$c = C_1(z_c^h, z_c^w) \quad (3)$$

$$\mathbf{g}_c^h = C_h(\mathbf{c}^h) \quad (4)$$

$$\mathbf{g}_c^w = C_w(\mathbf{c}^w) \quad (5)$$

$$C(i, j) = X \times \mathbf{g}_c^h(i) \times \mathbf{g}_c^w(j) \quad (6)$$

虽然,上述方法的引入可以使网络聚焦于舰船目标的位置,但是仍无法解决因目标尺寸变化大而导致的漏检难题,因此,本文针对网络的颈部进行改进,设计并联的多尺度坐标注意力模块,如图 8 所示。

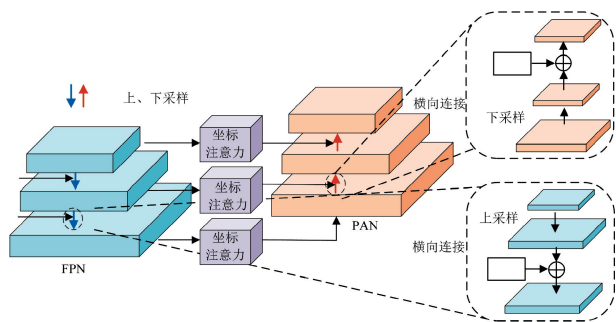


图 8 多尺度坐标注意力结构

Fig. 8 Structure of multiscale coordinate attention

本文在不同特征层级上构建特征金字塔网络 (feature pyramid networks, FPN) 和路径聚合网络 (path aggregation network, PAN), 首先,借助主干网络不同尺度的输出特征图 P3、P4、P5 与 FPN 低层特征经最近邻上采样 (2 倍) 的结果进行横向连接,自上而下传递语义特征;其次,在保证轻量化的同时,FPN 不同尺度的输出经由坐标注意力模块增强局部空间特征;最后,通过来自不同尺度的坐标注意力特征图进和 PAN 低层特征下采样之间的横向连接,自下而上传递定位特征,以实现多尺度信息的增强。FPN 和 PAN 中的横向连接方式为相应特征图通过相加求和的方式实现融合。

2) 损失函数

网络的损失由两部分组成,置信度损失和定位损失:

$$L = L_{\text{conf}} + L_{\text{locate}} \quad (7)$$

(1) 置信度损失

网络中,置信度损失采用 Focal 损失^[21],主要解决目标定位阶段不同类型样本数量不均衡的问题,其定义如下:

$$L_{\text{conf}} = -\alpha_i (1 - p_i)^{\tau} \log(p_i) \quad (8)$$

式中: p_i 是不同类别的概率; α_i 用于解决正负样本的不平衡,例如在目标检测数据集中,较多的桥墩目标和较少的水鼓目标导致的样本不均衡问题; τ 主要是解决难易样本的不平衡,例如数据集中互相重叠或尺度变化导致的置信度较低等问题,两个参数分别被设置为 0.25、2。

(2) 定位损失

3 种重叠方式如图 9 所示, A、B、C 分别表示真实框、预测框和最小外接矩形框面积,而深色区域对应的是重叠面积。交并比 (intersection over union, IoU) 用于计算

边框预测值和真值之间的差值,可以反映算法的定位效果。然而,用 IoU 衡量定位损失仍存在一些不足之处: 1) 若预测值和真值间不存在重合区域,则 IoU 为 0,无法度量样本间的距离,网络也就无法收敛和优化; 2) IoU 无法精确的预测框和真实框两者的拟合度,如图 9(a) ~ (c) 所示, 3 种情况的重叠区域面积一样, IoU 相等,说明仅使用 IoU 无法评价其定位效果。因此,本文在 IoU 之外再添加一个距离计算项,计算预测值和真实值之间的最小外接矩形,进而计算预测框和真实框偏移过大或没有交集时的距离值,其定义如下:

$$L_{\text{locate}} = 1 - \text{IoU}(B, A) + \frac{|C - (BUA)|}{|C|} \quad (9)$$

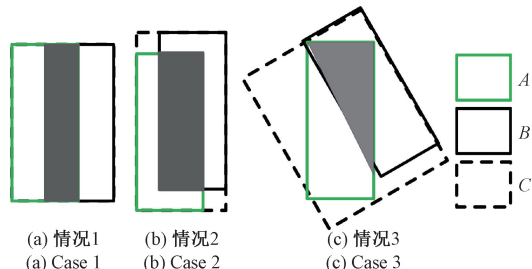


图 9 3 种重叠方式

Fig. 9 Three overlapping methods

鉴于单阶段目标检测算法细粒度分类精度方面存在一定的局限性,基于多尺度坐标注意力的 YOLOv8 算法主要用于目标位置信息的获取。准确定位后,根据位置信息,利用 OpenCV 的裁剪函数对输入图像进行分割,得到滤除背景后的目标图像。

4.2 多网络自监督学习框架

舰船目标相似程度高,识别难度大,仅靠检测算法无法对舰船目标精细分类。因此,本文采用无需标注信息的多网络自监督学习框架,对裁剪后图像 (仅包含目标及其类别标签) 进行辨别,得到精细化的分类结果,本文算法整体架构如图 10 所示。

如图 10(b) 所示,自监督学习框架由 3 部分组成: 定位网络、监督网络以及分类网络。定位网络与常规方法中的锚框类似,图像输入后,生成一系列比例与尺寸各异的局部矩形区域,用 $\xi_1, \xi_2, \dots, \xi_i$ 表示第 i 个区域所包含的有效信息量,并记将包含信息量最多的前 M 个矩形区域 (图 10(b) 中 ξ_1, ξ_2, ξ_3) 输入监督网络; 随后,监督网络生成置信度 $\kappa_1, \kappa_2, \dots, \kappa_M$, 通过分类结果和置信度间分类损失的收敛,来优化定位网络; 随着定位网络逐渐收敛和优化,前 K (本文设置为 2) 个信息最丰富的局部区域与输入图像一同输入全连接层,预测舰船目标的细粒度类别。上述 3 种网络的自监督学习可以有效缩减目标的类内差异,取得更高的分类置信度。本文使用“通道注意力+ResNet-101”作为多网络自监督学习框架的特征提

取器。其中,定位网络采用具有横向连接和自上而下的结构在目标图像上生成一系列不同大小的局部区域特征图;监督网络在接收到前者生成的局部特征图后,借助特征提取器和 softmax 层计算其属于真实类别的置信度,并

使用交叉熵损失来优化定位网络;分类网络从本质上来说属于全连接层,将之前 K 个信息丰富的局部区域与全局图像生成的特征向量进行拼接融合,得到细粒度分类结果。

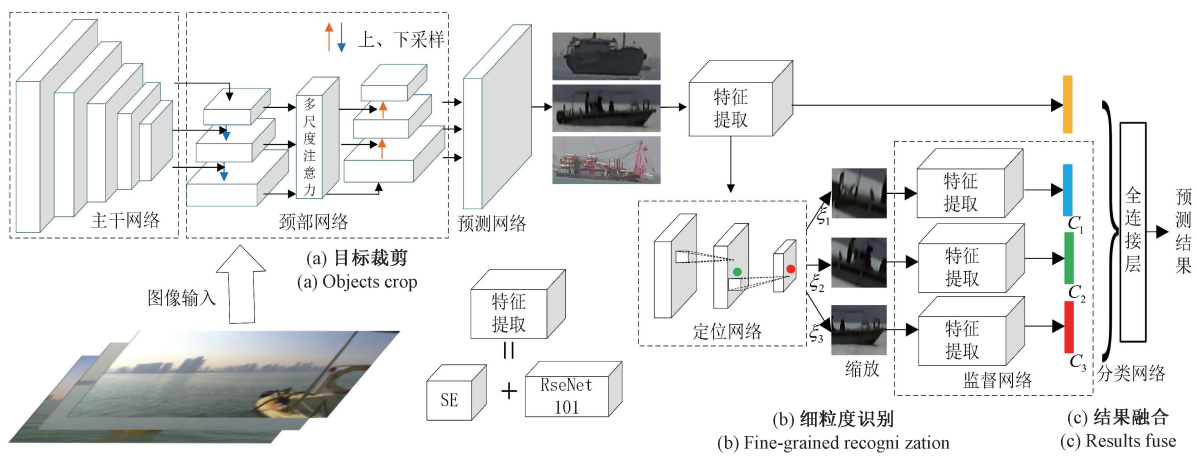


图 10 本文算法整体架构
Fig. 10 The overall architecture of the algorithm in this article

4.3 网络结构

由图 10 可知,舰船目标的细粒度检测算法通过 3 个步骤完成:目标区域提取;细粒度识别;结果融合。
目标裁剪阶段,本文使用基于多尺度坐标注意力的 YOLOv8 算法对目标精准定位,滤除无关的背景信息。因拍摄视角和距离不同,无人艇视角下舰船目标尺度多样,固定尺寸的锚框难以匹配目标的真实框,导致了定位精度的下降。因此,为缓解尺度变化带来的影响,本文采用无锚框的方法,简化了预测过程,减少了超参数数量,并提高了模型对不同宽高比目标的适应能力。细粒度识别阶段,将裁剪后的图像输入至多网络自监督学习框架,通过 3 种网络分工协作,得到辨别性区域和目标的细粒度识别结果。
结果融合阶段,将多网络自监督学习得到的细粒度类别信息与检测得到的位置信息融合,得到最终的细粒度检测结果。

5 实验结果与分析

5.1 实验设置与评价指标

本文实验依据自制数据集和公开数据集开展,自制数据集涵盖 2 229 张图像共 7 个类别,其中训练集为 1 560 张,测试集为 669 张。公开数据集包括细粒度车辆和飞行器分类数据集 Stanford 和 FGVC。训练参数设置方面,网络迭代次数为 100,批处理数量为 4,优化方法选用 SGD 优化,并对数据进行马赛克增强处理。训练与

测试的实验环境为 ubuntu 20.04 操作系统,pytorch 1.9.1 框架,配置有 NVIDIA RTX 1060Ti 显卡。
评价指标包含精度和速度两个部分,精度的评价指标包括平均精度均值 (mean average precision, mAP) mAP@ 0.5、mAP@ 0.5:0.95 等,mAP@ 0.5 是指在 0.5 阈值下多个类别对应的平均精度均值,mAP@ 0.5:0.95 是指在 0.5:0.95 阈值下多个类别对应的平均精度均值,一般而言,mAP@ 0.5:0.95 更为严苛。此外,为验证算法对不同尺度目标的检测能力,本文还引入平均精度 (average precision, AP) AP_s 、 AP_m 、 AP_l 指标,分别代表小、中、大 3 种不同尺寸目标的分类精度。如表 1 所示,参照 COCO 数据集大中小目标定义,在舰船目标检测数据集 (图片像素为 1 920×1 080 pixels) 中,本文将边框长宽乘积小于 1 867 pixels 的目标记作小目标,位于 1 867 与 1 6796 pixels 之间的目标记为中等目标,大于 1 6796 pixels 的目标记为大目标。速度的评价指标为帧率,其含义为每秒能够处理图片的帧数。

表 1 目标尺寸定义
Table 1 The definition of the sizes of targets

目标类型	COCO 数据集	舰船目标数据集/pixels (1 920×1 080)
小目标	$\leq 32 \times 32$	$w \times h \leq 1\,867$
中等目标	$32 \times 32 \sim 96 \times 96$	$1\,867 < w \times h < 16\,796$
大目标	$\geq 96 \times 96$	$w \times h \geq 16\,796$

5.2 对比实验与分析

1) 与其他检测算法性能对比

首先在自制舰船目标检测数据集上设计了本文算法和其他先进的目标检测算法的对比实验,包括两阶段算法 Faster-RCNN、Cascade-RCNN 和一阶段算法 SSD^[22]、YOLOv3、YOLOX、YOLOv5、YOLOv6、YOLOv8、YOLOv9。为确保实验结果的有效性和可参考性,各算法的训练参数设置均保持一致,图像输入尺寸统一调整至 512×512。算法的定量测试结果如表 2 所示,本文算法对小目标和大目标的分类精度最优,分别达到 45.2% 和 86.1%,

比基础网络 YOLOv8 的对应指标分别高 2.5% 和 3.6%,比 YOLOv9 算法的对应指标分别高 0.6% 和 1.9%。两阶段算法 Faster-RCNN 对中等目标的分类精度最高,达到 61.6%。同时,本文算法的 mAP@ 0.5 达到 94.6%,优于其余的两阶段和一阶段检测算法,这表明本文算法在变尺寸目标检测方面有显著的性能优势。特别是 AP_s 和 mAP@ 0.5 的显著提升,证明了其先进性和有效性。此外,检测速度方面,本文方法仅次于 YOLOv9,检测速度达到 27 fps,充分说明了本文算法在速度和精度上达到了较好的平衡。

表 2 不同算法精度对比
Table 3 The comparison of different algorithms

方法	AP _s /%	AP _M /%	AP _L /%	mAP@ 0.5/%	mAP@ 0.5:0.95/%	检测速度/fps
Faster-RCNN	39.1	61.6	71.5	92.3	58.4	4
Cascade-RCNN	12.6	29.2	37.3	64.3	31.1	7
SSD	26.3	48.6	70.4	86.0	48.1	13
YOLOv3	40.2	59.2	71.0	93.6	56.4	12
YOLOX	40.2	58.5	72.0	94.4	57.4	15
YOLOv5	40.9	57.3	79.1	93.2	57.9	18
YOLOv6	30.8	56.9	71.7	90.2	55.1	23
YOLOv8	42.7	57.9	82.5	93.5	58.5	29
YOLOv9	44.6	58.7	84.2	94.1	59.2	32
本文	45.2	60.4	86.1	94.6	60.9	27

本文算法改进前后 mAP@ 0.5 和 mAP@ 0.5:0.95 值的定性分析结果如图 11 所示。由图 11 可知,改进后的方法收敛更快,约在 10 个 epoch 时达到 80%,而改进前的方法在 20 个 epoch 后才达到 80%。另一方面,改进后方法的 mAP 值更高更稳定,波动也较小,说明其更加鲁棒。

使用多尺度坐标注意力模块前后算法的类激活热力图如图 12 所示。一般而言,通过热力图,可以得到对正确分类起到作用的局部区域,区域灰度越深,越有助于模型进行分类。图中可以明显看到,原算法下深色区域覆盖范围大,所用无关信息过多。使用多尺度坐标注意力

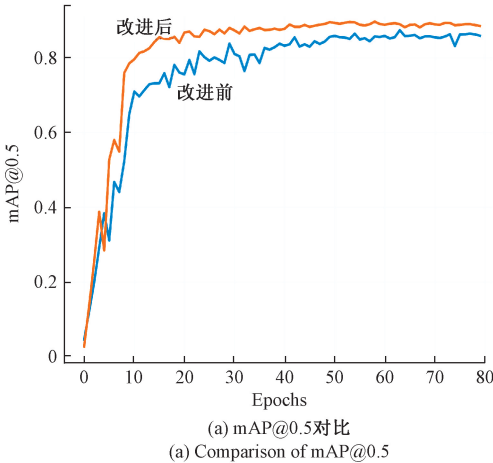
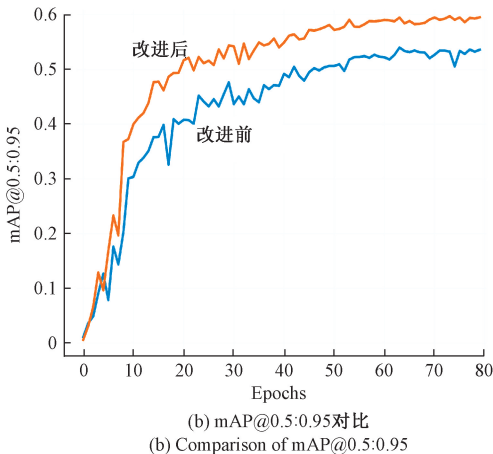


图 11 本文算法改进前后性能对比
Fig. 11 Comparison of performance before and after improvement

模块后,模型更加聚焦含有舰船目标的局部区域,在背景冗杂的海面场景中,抑制无关信息,产生更具辨别性的特征表示。

2) 与其他注意力的对比分析

为验证本文提出的多尺度坐标注意力模块的有效性,本文在自制舰船目标数据集上进行了与其他注意力模块的对比实验。对比实验以 YOLOv8 为基准模型,分

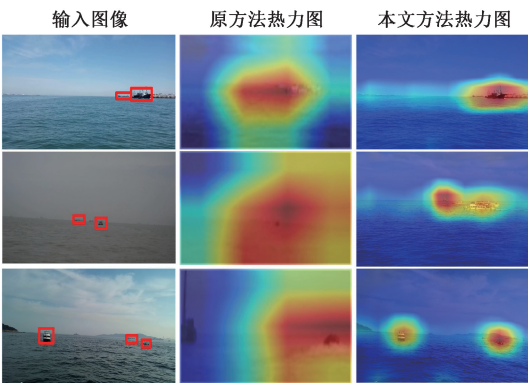


图 12 改进前后热力图对比

Fig. 12 Comparison of heat maps before and after improvement

别添加卷积块注意力模块(convolutional block attention mojule,CBAM)、选择性内核(selective kernel, SK) 注意力、挤压激励(squeeze-excitation, SE) 注意力、S2 注意力、Shuffle 注意力以及几种注意力的组合,与本文设计的多尺度坐标注意力进行对比。将这些注意力模块嵌入到模型的颈部网络之间,训练过程中所有的参数设置、实验软硬件配置均相同,保证对比实验的可参考性。此外,为定量比较添加各种注意力结构后网络的性能,记录了各个模块的参数数量。对比实验结果如表 3 所示。

表 3 不同注意力在舰船目标数据集上的对比结果
Table 3 The comparative results of different attention on datasets of ship targets

编号	方法	mAP@ 0. 5/%	mAP@ 0. 5 : 0. 95/%	参数量 /KB	检测速度 /fps
1	YOLOv8s	93. 5	58. 5	—	29
2	+CBAM	94. 0	59. 1	130. 5	28
3	+SK	94. 3	59. 7	86 368. 13	25
4	+SE	93. 9	59. 5	128	28
5	+S2	94. 3	59. 7	8 200	27
6	+Shuffle	94. 2	59. 7	0. 75	29
7	+S2+Shuffle	94. 1	59. 5	8 200. 75	27
8	+S2+SE	94. 2	59. 5	8 328	27
9	+Shuffle+SE	93. 7	58. 6	128. 75	28
10	+S2+SE+ Shuffle	93. 9	59. 2	88 328. 75	25
11	+CA	94. 2	59. 6	100. 19	28
12	本文	94. 6	60. 9	300. 6	27

从表 3 可以看出,相比于基准模型,在加入本文的多尺度坐标注意力模型后,mAP@ 0. 5、mAP@ 0. 5 : 0. 95 分别提升了 1. 1% 和 2. 4%,提升检测精度均位列第 1,达到最佳,相比于 CBAM 注意力,本文算法的

mAP@ 0. 5 : 0. 95 提升 1. 8%。相比于仅使用坐标注意力模块,本文算法的 mAP@ 0. 5 : 0. 95 提升 1. 3%。这表明,本文设计的多尺度坐标注意力可以取得更好的效果。方法 7~10 的结果表明,并不是注意力模块间的组合叠加都能够改善检测性能,部分情况下注意力模块的叠加也会降低检测性能。对比 CBAM 注意力,本文设计的模型参数量有所增加,在检测速度仅下降 2 fps 的同时,性能大大提高,说明了本文的多尺度坐标注意力模块与其他注意力模块相比的优越性。此外,原始的 YOLOv8 的检测速度较快,但检测性能相对较低,原因在于无人艇视角下舰船目标的特征弱和多尺度特性,从而降低了目标的检测性能,加入多尺度坐标注意力机制之后,网络对于不同尺度目标赋予不同的权重再混洗融合,算法运行速率达到 27 fps,基本满足无人艇自主航行的感知需求。说明本方法在保持轻量化的同时,进一步提升了对多尺度舰船目标的检测性能。

3) 细粒度分类性能的比分析

本文使用弱监督细粒度识别算法解决相似目标检测精度低的问题,不同于目标检测图像,弱监督细粒度图像仅包含目标和其类别标签。除细粒度舰船目标数据集外,本文算法还在公开细粒度分类数据集 Stanford Cars 和 FGVC Aircraft 上进行了测试。训练中,图像被统一缩放到 448×448 大小,辨别性区域数量 K 为 2。

不同的细粒度识别算法在公开和自制数据集上的识别准确率如表 4 所示。由表 4 可知,本文算法在 FGVC 数据集和舰船目标细粒度数据集的识别正确率明显优于其他算法,分别达到 93. 5% 和 83. 5%,比常见的弱监督识别算法 Bilinear-CNN 高了 2. 2% 和 3. 9%。在 Stanford 数据集上,本文算法的识别结果仅次于 MA-CNN,达到 89. 6%。实验结果表明,无论是在公开数据集还是自制数据集上,多网络自监督学习算法取得了比典型弱监督算法更好的细粒度识别结果。

表 4 细粒度识别性能对比
Table 4 The comparison of fine-grained recognition performance (%)

方法	Stanford Cars	FGVC Aircraft	舰船
Bilinear-CNN	84. 1	91. 3	79. 6
RA-CNN	88. 2	92. 5	92. 5
Boost-CNN	88. 5	92. 1	81. 5
MA-CNN	89. 9	92. 8	83. 2
本文	89. 6	93. 5	83. 5

与其他海面目标相比,舰船目标的共同点在于其外观呈扁平状。因此,目标宽、高尺度差异可以判断其属于舰船这一大类,但无法判断目标的精细化类别。为直观

剖析多网络自监督框架于细粒度识别任务中的效能,多网络自监督框架所预测的信息区域如图 13 所示。以各异的矩形框标识出蕴含信息量最为充裕的 4 个区域。图 13A 图像源于 CUB200-2011 数据集,且 K 值设定成 2;图 13B 图像亦取自 CUB200-2011 数据集, K 值设定成 4。图 13C 和 D 图像分别源自 Stanford 数据集以及 FGVC 数据集, K 值均设定为 4。图 13E 和 F 图像出自舰船目标检测数据集,其超参数分别设为 2 和 4。



图 13 预测区域展示
Fig. 13 Display of forecast area

由图 13 能够观察到,图 13A 运用两个区域可覆盖鸟类的部分信息,其中有图像中鸟类的颜色与背景极为相似,借助本文算法能预测出信息量最为充裕的区域。由图 13B 可以发现,鸟类头部、翅膀以及主体部分是信息量最丰富的区域,这与人类的认知相符。针对图 13C,大灯和格栅被视作汽车中信息量最充裕的区域。而图 13D,定位网络将焦点置于飞机的机翼和机头处,这些区域更具有辨别性。从图 13E 和 F 可以看出,船身和船尾属于船舶目标信息量最为丰富的区域。实验结果显示,本文所提出的多网络自监督学习算法有助于网络精准定位到信息量最丰富的区域,进而提升细粒度分类的精度。

此外,为表征算法的识别性能,舰船目标细粒度分类的混淆矩阵如图 14 所示,其中每一列代表该类别数据的预测值,每一行代表该类别数据的真实值,通过比较预测结果和实际结果来计算模型的分类准确程度。

由图 14 可知,水鼓、浮标、桥等典型海面目标识别精度较高,而渔船、快艇、拖船等舰船目标的识别精度略低,特别是渔船、快艇以及商船、拖船目标间容易被相互混淆。所有渔船目标中,被准确识别的占到 82%,被误判为

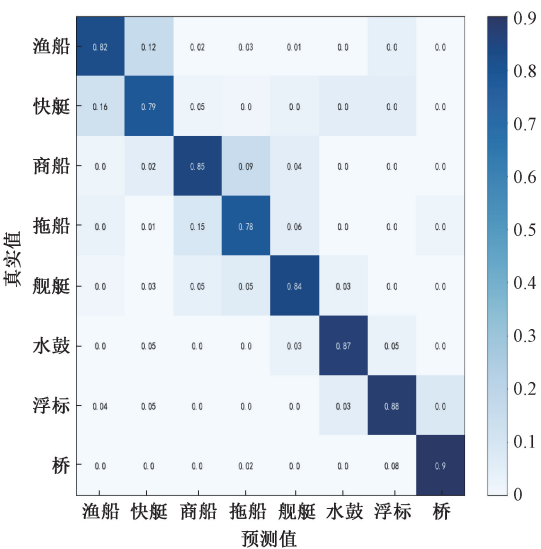


图 14 混淆矩阵
Fig. 14 The confusion matrix

快艇的占到 12%;快艇目标中,准确识别的占到 79%,被误识别成渔船占到 16%。此外,有 9%的商船目标被识别为拖船,15%的拖船目标被识别为商船,导致此类问题的原因是目标在外观上的相似性。值得注意的是,多网络自监督学习方法对渔船、快艇、商船的识别正确率均高于正常的检测算法,这说明本文的细粒度检测算法能够帮助网络定位到特征更加丰富、更有辨别性的局部区域,提高舰船目标的细粒度识别结果。

5.3 消融实验

为定量分析和比较本文算法中不同模块、不同参数对舰船目标检测结果带来的影响,本文在自制的舰船目标数据集上进行了不同的消融实验,结果如表 5 所示。其中,MC-YOLOv8 表示仅使用多尺度坐标注意力的 YOLOv8 检测算法,LSC-Net 表示本文提出的细粒度检测算法,用 LC-Net 表示去掉监督网络后的检测算法(即定位网络直接输出信息区域)。此外,实验还研究了超参数

表 5 消融实验结果
Table 5 Results of ablation studies

编号	方法	细粒度检测准确率/%
1	YOLOv8	77.8
2	MC-YOLOv8	80.9
3	ResNet-101	79.6
4	DC-Net ($K=4$)	78.2
5	DSC-Net ($K=0$)	81.5
6	DSC-Net ($K=2$)	83.2
7	DSC-Net ($K=4$)	83.3

K 对实验结果的影响,表示被用于细粒度分类的局部区域数量,超参数分别设置为 0、2、4。

实验结果表明,本文所提出的细粒度检测方法对提升舰船目标细粒度检测性能有着积极作用。相较于基于多尺度坐标注意力的 YOLOv8 算法,本文方法的检测准确率实现了 2.4% 的提升。不过,方法 4 的实验结果显示,在去掉监督网络后,识别准确率降至 78.2%,较之前的 83.3% 有所下降,且 78.2% 的精度甚至低于直接运用 ResNet-101 进行测试所得到的结果。由此可见,若缺乏监督网络的反馈修正,细粒度识别的性能有所下降。其次,当超参数从 0 提升至 2 时,识别准确率提升了 1.7%,而当从 2 提高到 4(即特征维度扩大 1 倍)时,识别准确率仅提升了 0.1%。据此可知,简单地增加特征维度虽能带来些许改善,但同时也会造成参数冗余,进而致使性能下降,所以超参数 K 的选择不宜过多,本文超参数 K 设定为 4。

5.4 典型舰船目标检测集成应用

本文在典型海面环境下构建了舰船目标检测场景。系统的总体架构以及数据传输过程如图 15 所示,该系统主要包含无人艇平台和岸上平台这两大部分。其中,无人艇平台涵盖了艇载多源传感器、感知平台以及艇上电源这几部分,而岸上平台包含岸上电源与显示终端^[23]。在集成应用时,先是通过艇外安装的光电吊舱来采集真实海况场景下的图像或视频,随后借助网络交换机把数据传输至感知平台。紧接着,感知平台会对图像开展实时细粒度检测处理,处理后的数据经艇上电台、岸上电台传输,最终传至岸上终端,以此为无人艇的决策提供感知数据方面的有力支撑。

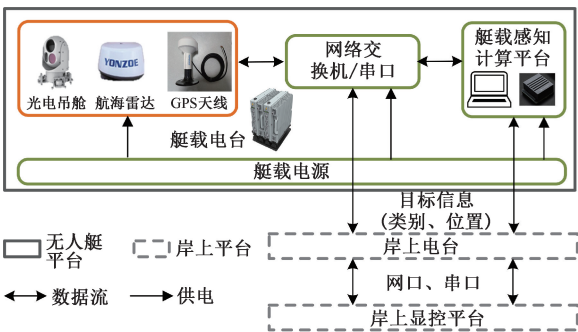


图 15 无人艇实验平台总体架构

Fig. 15 The architecture of USV platform

本文选择 Nvidia Xavier 作为无人艇载的感知平台,可将本文算法所使用的 Opencv、CUDA 学习库迁移至感知平台上,再通过自组网电台连接,使岸上显控端与无人艇载处理端直接相连。岸上显控端可直接下达指令,明确无人艇运动方向,并控制光电吊舱转动至目标方向。如图 16 所示,为无人艇在不同天气、不同时序下的算法

实测结果,目标距无人艇约 100~200 m。由结果可知,本文算法能有效检测出真实海面环境下的舰船目标,并且能区分出舰船目标的细粒度类别。此外,集成应用时,岸上端可随时控制无人艇或光电吊舱。



图 16 无人艇载算法运行结果

Fig. 16 Results of the algorithm on USV

5.5 结果讨论

本文设计了基于多尺度坐标注意力和多网络自监督学习的细粒度检测方法。多网络自监督学习由细粒度分类方法启发,而多尺度坐标注意力是由坐标注意力启发。本文最开始使用坐标注意力进行实验和测试,发现网络在训练至 25 epoch 时,出现过拟合现象,本文记录了此时训练过程中的损失情况,如图 17 所示。原因在于舰船目标尺度多样,网络如果不能提取目标的多尺度特征,小尺寸和中尺寸目标在经过下采样后特征更加微弱,无法有效完成舰船目标检测任务。

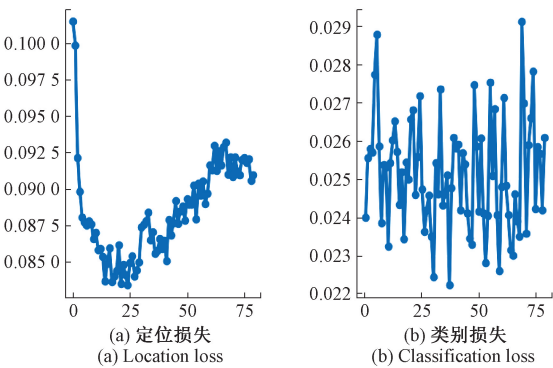


图 17 注意力模块改进前性能

Fig. 17 Performance of the attention module before improvement

因此,本文选择多尺度坐标注意力模块,将注意力模块嵌入网络小、中、大 3 种不同尺度的特征层中,改善网络对于多尺度目标的检测性能。使用多尺度坐标注意力

模块后网络训练过程的损失情况如图 18 所示。由图 18 可知,随着 epoch 的增加,模型的损失逐渐降低,约在 epoch 为 15 时,就收敛至较为稳定的值。

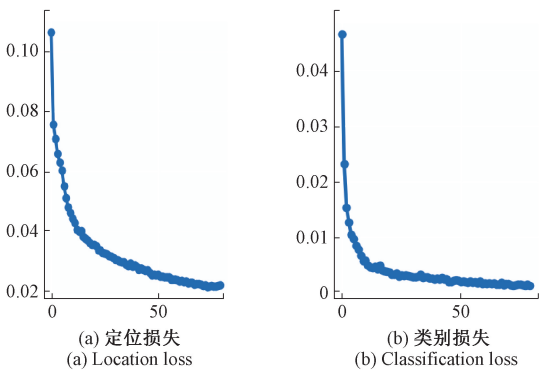


图 18 注意力模块改进后性能

Fig. 18 Performance of the attention module after improvement

此外,为直观彰显本文算法在海面场景中的优越性,不同输入图像所对应的尺度不变特征变换 (Scale invariant feathre transform, SIFT) 以及本文检测算法的可视化结果,如图 19 所示。在输入图像中,矩形框标记出水面目标的位置。SIFT 特征以圆框标注具有尺度不变属性的特征点。本文算法的可视化结果借由热力图展现,在热力图中,区域颜色越深,则表明该区域对检测、分类等任务的贡献越大。从图 19B ~ E 结果可看出, SIFT 特征会把岸边建筑物等那些颜色和轮廓与水面目标相近的物体匹配为特征点,还会致使一些小尺度目标被遗漏检测。而图 19 (a3) 和 (b3) 结果表明,本文算法剔除了海面场景里大量的背景信息,使网络聚焦于图中的深色区域,且与输入图像的水面目标所处区域逐一对应,由此说明本文的算法有助于舰船目标的检测和识别。

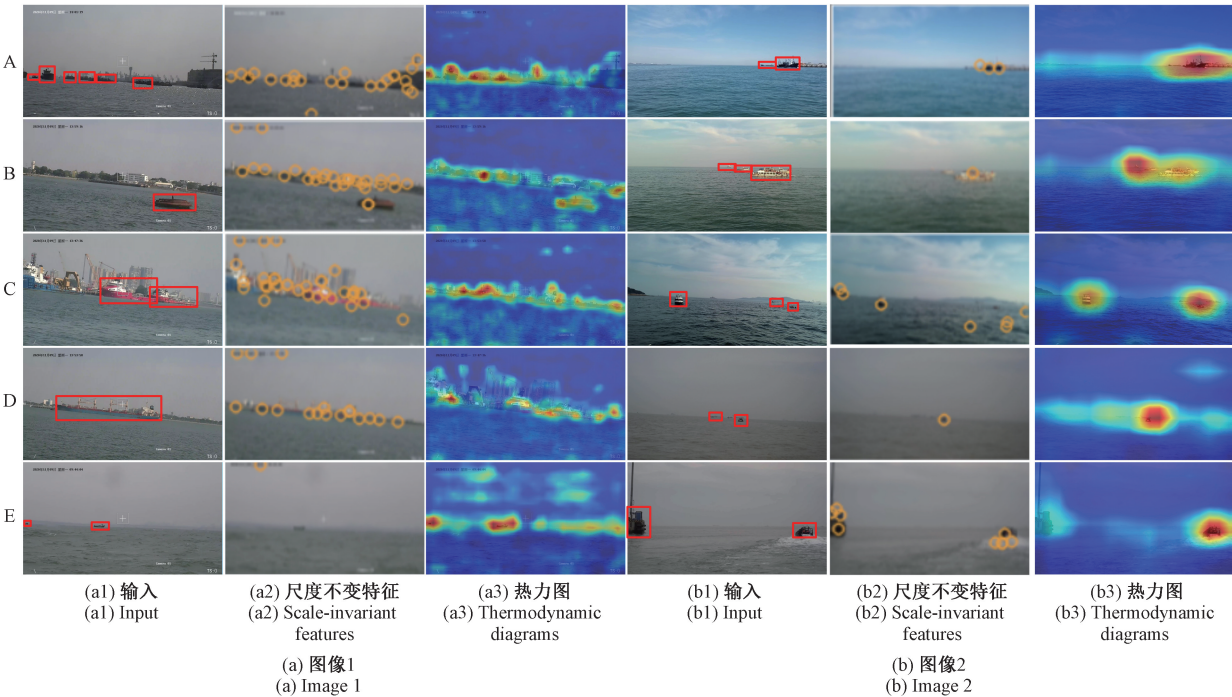


图 19 可视化结果

Fig. 19 Visualization results

6 结 论

舰船目标的细粒度检测有助于海面场景不明来源舰船目标的精细划分和意图分析,在军事和民用领域都发挥着无可替代的作用。本文针对海面场景下无人艇视角目标检测问题,首先,提出了一种多尺度坐标注意力和多网络自监督学习的细粒度检测方法,以提升

海面场景下舰船目标的检测和识别精度;其次,构建了无人艇视觉感知平台,通过光电吊舱采集、人工筛选、标注,自制了舰船目标检测数据集;最后,在自制数据集和公开数据集上进行了对比实验和集成验证。结果表明,本文方法能明显提升无人艇对舰船目标的细粒度检测性能,且具有较好的运行速率。未来的研究应聚焦于目标遮挡、重叠等条件下或多视角条件下舰船目标的识别和稳定感知。此外,无人艇视角下舰船目标检测数据集的开源工作也有待进一步完善。

参考文献

- [1] 赵其昌, 吴一全, 苑玉彬. 光学遥感图像舰船目标检测与识别方法研究进展[J]. 航空学报, 2024, 45(8): 51-84.
ZHAO Q CH, WU Y Q, YUAN Y B. Progress of ship target detection and recognition methods in optical remote sensing images[J]. Journal of Aeronautics, 2024, 45(8): 51-84.
- [2] 宋志娜, 李莎, 杨建明, 等. 基于特征与区域定位增强的遥感舰船目标检测[J]. 计算机工程, 2023, 49(8): 257-264.
SONG ZH N, LI SH, YANG J M, et al. Remote sensing ship target detection based on feature and area localization enhancement[J]. Computer Engineering, 2023, 49(8): 257-264.
- [3] 张涛, 杨小冈, 卢瑞涛, 等. 基于关键点的遥感图像舰船目标检测[J]. 系统工程与电子技术, 2022, 44(8): 2437-2447.
ZHANG T, YANG X G, LU R T, et al. Remote sensing image ship target detection based on keypoints[J]. Systems Engineering and Electronics, 2022, 44(8): 2437-2447.
- [4] 李登峰, 高明, 叶文韬. 结合轻量级特征提取网络的舰船目标检测算法[J]. 计算机工程与应用, 2023, 59(23): 211-218.
LI D F, GAO M, YE W T. Ship target detection algorithm combined with lightweight feature extraction network[J]. Computer Engineering and Applications, 2023, 59(23): 211-218.
- [5] 范云生, 张凯, 牛龙辉, 等. 海上无人机对运动船舶的长期检测跟踪算法[J]. 仪器仪表学报, 2024, 45(3): 326-335.
FAN Y SH, ZHANG K, NIU L H, et al. Long-term detection and tracking algorithm for moving ships by maritime UAV[J]. Chinese Journal of Scientific Instrument, 2024, 45(3): 326-335.
- [6] 车思文, 汪宇玲. 一种改进的 YOLOv7 光学遥感图像舰船目标检测算法[J]. 电光与控制, 2024, 31(5): 34-39, 65.
CHEN S W, WANG Y L. An improved YOLOv7 optical remote sensing image ship target detection algorithm[J]. Electronics Optics and Control, 2024, 31(5): 34-39, 65.
- [7] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] 王宇, 魏宇, 孙传猛, 等. 复杂恶劣环境下水位智能检测方法研究[J]. 电子测量与仪器学报, 2023, 37(11): 119-131.
WANG Y, WEI Y, SUN CH M, et al. Research on intelligent water level detection method in complex and adverse environments[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(11): 119-131.
- [9] 刘玉洁, 补冲. 基于多特征融合的舰船目标检测方法研究[J]. 舰船科学技术, 2024, 46(9): 164-167.
LIU Y J, BU CH. Research on ship target detection method based on multi-feature fusion[J]. Ship Science and Technology, 2024, 46(9): 164-167.
- [10] REN D W, MENG L W, WANG R. Borehole depth recognition based on improved YOLOx detection[J]. The Computer Journal, 2024, 67(7): 2408-2420.
- [11] 王振, 方海峰, 曹晋, 等. 基于 YOLOv5s 的轻量化可回收饮料瓶颜色识别[J]. 国外电子测量技术, 2023, 42(3): 160-166.
WANG ZH, FANG H F, CAO J. Color recognition of lightweight recyclable beverage bottles based on YOLOv5s[J]. Foreign Electronic Measurement Technology, 2023, 42(3): 160-166.
- [12] LI Y W, ZHANG X X. Object detection for UAV images based on improved YOLOv6[J]. International Journal of Computer Science, 2023, 50(2): 759-768.
- [13] 戚玲珑, 高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程, 2023, 49(1): 41-48.
QI L L, GAO J L. Small object detection based on improved YOLOv7[J]. Computer Engineering, 2023, 49(1): 41-48.
- [14] 赵春晖, 肖舒匀, 宿南. 基于注意力机制特征增强的舰船目标识别[J]. 黑龙江大学学报, 2022, 13(1): 43-51.
ZHAO CH H, XIAO SH Y, SU N. Ship recognition based on feature enhancement using attention mechanism[J]. Journal of Engineering of Heilongjiang University, 2022, 13(1): 43-51.
- [15] 杨晓强, 黄加诚. 基于动态定位和特征融合的多分支细粒度识别方法[J]. 计算机工程与科学, 2024, 46(2): 253-263.
YANG X Q, HUANG J CH. Multi branch fine-grained recognition method based on dynamic positioning and feature fusion[J]. Computer Engineering and Science, 2024, 46(2): 253-263.
- [16] 张拯宁, 张林, 王钺, 等. 基于层间分类一致性准则的舰船目标细粒度识别方法[J]. 中国空间科学技术, 2023, 43(3): 93-104.
ZHANG ZH N, ZHANG L, WANG Y, et al. A fine-

- grained recognition method for ship targets based on inter layer classification consistency criteria[J]. China Space Science and Technology, 2023, 43(3): 93-104.
- [17] 王彬州,肖志勇. 面向细粒度图像识别的通道注意力多分支网络[J]. 激光与光电子学进展, 2021, 58(22): 172-180.
- WANG B ZH, XIAO ZH Y. Channel attention multi-branch network for fine-grained image recognition[J]. Laser and Optoelectronics Progress, 2021, 58(22): 172-180.
- [18] 姜孟超,范灵毓,李硕豪. 基于注意力双线性池化的细粒度舰船识别[J]. 计算机技术与发展, 2022, 32(8): 66-70.
- JIANG M CH, FAN L Y, LI SH H. Fine-grained ship recognition based on attention bilinear pooling[J]. Computer Technology and Development, 2022, 32(8): 66-70.
- [19] 孙备,党昭洋,吴鹏,等. 多尺度互交叉注意力改进的单无人机对地伪装目标检测定位方法[J]. 仪器仪表学报, 2023, 44(6): 54-65.
- SUN B, DANG ZH Y, WU P, et al. Multi-scale intersecting attention improved single UAV ground camouflage target detection and localization method[J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 54-65.
- [20] 闫祎巧,王宏生,赵怀慈,等. 融合多尺度卷积和注意力机制的场景提取方法[J]. 电子测量技术, 2023, 46(16): 172-178.
- YAN Y Q, WANG H SH, ZHAO H C. Scene extraction method integrating multi-scale convolution and attention mechanism[J]. Electronic Measurement Technology, 2023, 46(16): 172-178.
- [21] 童小钟,魏俊宇,苏绍璟,等. 融合注意力和多尺度特征的典型水面小目标检测[J]. 仪器仪表学报, 2023, 44(1): 212-222.
- TONG X ZH, WEI J Y, SU SH J, et al. Fusion of attention and multi-scale features for detection of typical small targets on water surface[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 212-222.
- [22] 陈欣佳,刘艳霞,洪晓斌,等. 基于 SSD-CF 的无人艇目标检测跟踪方法[J]. 中国测试, 2019, 45(2): 145-150.
- CHEN X J, LIU Y X, HONG X B, et al. Unmanned surface vehicle target detection and tracking method based on SSD-CF[J]. China Testing, 2019, 45(2): 145-150.
- [23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

作者简介



左震, 分别在 2002 年、2004 年和 2012 年于国防科技大学获得学士学位、硕士学位和博士学位, 现为国防科技大学副研究员, 主要研究方向为智能侦测与对抗。

E-mail: z. zuo@nudt.edu.cn

Zuo Zhen received his B. Sc. degree, M. Sc. degree and Ph. D. degree all from National University of Defense Technology in 2002, 2004, and 2012, respectively. He is currently an associate researcher at the University of National Defense Technology. His main research interests include intelligent detection and countermeasures.



郭润泽(通信作者), 分别在 2018 年、2021 年于国防科技大学获得学士学位、硕士学位, 现为国防科技大学博士研究生, 主要研究方向为智能侦测与对抗。

E-mail: guorunze14@nudt.edu.cn

Guo Runze (Corresponding author) received his B. Sc. degree and M. Sc. degree both from National University of Defense Technology in 2018 and 2021, respectively. He is currently a Ph. D. candidate at National University of Defense Technology. His main research interests include intelligent detection and countermeasures.