

引用:黄媛,曹明华,吕志胜,等 融合全局注意力机制的多尺度船舶检测算法[J]. 聊城大学学报(自然科学版), 2025,38(6): 853-863.

Citation:HUANG Yuan, CAO Minghua, LÜ Zhisheng, ZHANG Lingyun, FENG Bin. Multiscale ship detection algorithm incorporating global attention mechanism[J]. Journal of Liaocheng University (Natural Science Edition), 2025, 38(6):853-863.

文章编号 1672-6634(2025)06-0853-11

DOI 10.19728/j.issn1672-6634.2025020001

融合全局注意力机制的多尺度船舶检测算法

黄媛¹,曹明华¹,吕志胜²,张凌云¹,封斌²

(1. 兰州理工大学 计算机与通信学院,甘肃 兰州 730050;2. 广东航海学院 信息与通信工程学院,广东 广州 510725)

摘要 针对复杂海域航行过程中,仍存在小目标漏检和多目标分类性能不佳的问题,提出一种全局注意力引导的多尺度船舶检测算法,命名为 GAG-YOLO。在 YOLOv5 目标检测网络中引入全局注意力机制(GAM),以捕捉全局上下文信息,使模型能够更好地辨别船舶目标与复杂背景之间的差异。此外,优化锚框参数以更好地匹配船舶的形状,更准确地定位和识别不同尺度船舶目标。最后,改进 YOLOv5 网络的损失函数,利用 Focal-EIOU 聚焦高质量样本,通过分配不同的权重进行更精确的边界框回归。实验结果表明,GAG-YOLO 在检测准确性、鲁棒性方面均优于列出的 YOLO 系列算法,其中平均精度均值 $m_{AP0.5:0.95}$ 提高了 4.6%。该检测算法在提升整体检测速度的同时,有效降低了漏检与误检率,为复杂海域中的船舶监测提供了更快速且准确的解决方案。

关键词 YOLOv5s; 全局注意力机制; Focal-EIOU; 锚框优化; 多尺度船舶检测

中图分类号 U675.7;TP391.41

文献标志码 A

开放科学(资源服务)标识码(OSID)



Multiscale ship detection algorithm incorporating global attention mechanism

HUANG Yuan¹, CAO Minghua¹, LÜ Zhisheng², ZHANG Lingyun¹, FENG Bin²

(1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China; 2. School of Information and Communication Engineering, Guangzhou Maritime University, Guangzhou 510725, China)

Abstract The aim of this research is to address the remaining issues with small target leakage and poor multi-target classification performance that are still present during navigation in complex sea areas. To this end, a global attention-guided multi-scale ship detection algorithm, named GAG-YOLO, has been proposed. The global attention mechanism (GAM) has been introduced into the YOLOv5 target detection network in order to capture the global contextual information, thereby enabling the model to better discriminate the ship target from the complex contextual differences. Furthermore, the anchor frame parameters are optimized to better match the shape of the ship and to more accurately locate and identify ship targets at different scales. Finally, the loss function of the YOLOv5 network is improved to focus on high-quality samples using Focal-EIOU for more accurate bounding box regression by assigning different

收稿日期:2025-02-05

基金项目:国家自然科学基金项目(62265010);广东省企业科技特派员项目(GDKTP2021041300)资助

通信作者:封斌,男,汉族,博士,高级工程师,研究方向:船联网、船岸协同感知、航道感知,E-mail: fengbin@gzmtu.edu.cn。

weights. The experimental results demonstrate that GAG-YOLO outperforms the listed YOLO series algorithms in terms of detection accuracy and robustness. Specifically, the mean average precision (mAP) at 0.5:0.95 is improved by 4.6%. The detection algorithm improves the overall detection speed and effectively reduces the leakage and false detection rate, providing a faster and more accurate solution for ship monitoring in complex sea areas.

Keywords YOLOv5s; global attention mechanism; Focal-EIOU; anchor box; multi-scale ship detection

0 引言

近年来,船舶探测在海洋监视、海事安全、港口管理和环境保护等各个应用领域的重要性与日俱增,准确检测和跟踪船舶对于有效管理海洋资源和保障海洋安全至关重要。传统的船舶目标检测方法主要包括三个步骤:区域选择、特征提取和分类。在区域选择阶段,算法通过特定规则或启发式方法选择可能包含船舶的图像区域。在特征提取阶段,使用人工设计的特征,如边缘、纹理和颜色等。然而,这些特征通常是浅层特征,对船舶的表征能力有限。在分类阶段,基于提取的特征进行目标分类。由于传统方法依赖人工设计的浅层特征,在面对多变的海上环境时表现出明显的局限性,难以满足实际应用的需求。复杂的海洋环境中,天气、光照和波浪等因素的变化,对船舶检测提出了更高的要求。传统方法在这些情况下往往难以保持高精度和稳定性。

因此,迫切需要采用更加先进的技术来提升船舶检测的准确性和鲁棒性。深度学习技术的发展为船舶检测提供了新的解决方案,能够自动提取深层特征,显著提升检测效果和适应能力。目前,基于深度学习的方法可分为两类:一阶段框架和两阶段框架。两阶段框架包括基于区域的卷积神经网络(Region based Convolutional Neural Network, R-CNN)^[1]、Fast R-CNN^[2]和 Faster R-CNN^[3]。R-CNN 由 Ross B. Girshick 于 2014 年提出,它将区域建议网络(RPN)与深度卷积神经网络相结合,对图像中的物体进行定位和分类,适用于各种物体检测和识别任务。Fast R-CNN 实现了 R-CNN 与 RPN 之间的卷积层共享,通过深度卷积获得更高质量的候选区域,提高检测精度。上述框架充分利用了预测帧的特征信息,从而提高了检测性能。但是,需要生成候选区域,导致推理时间较长,实时性较差。

一阶段框架在图像上等距采样,并直接以锚框和目标框为中心进行目标分类和坐标校正。这简化了预测帧的处理和图像重采样过程,使检测速度更快,但精度略有下降。具有代表性的算法包括 YOLO 和 SSD^[4]。SSD 算法结合了锚框机制和多尺度特征图,对小目标的检测性能更好。YOLO^[5]则采用 CNN 对分类概率和边界框进行回归,具有实时检测的优势。YOLOv5^[6]在输入端采用了多种增强技术,如自适应图像缩放、马赛克和自适应锚框。此外,该网络还加入了空间金字塔池化(SPPF)和路径聚合网络^[7](PANet),以增强其特征融合能力。

海上实时目标检测面临诸多挑战,包括目标集群小、船舶类型密集复杂、海面雾气导致目标模糊等。这些因素对 YOLOv5 检测和分类小型目标的能力提出了更高的要求。文献[8]提出的 ECA(Efficient Channel Attention)注意力机制被用于增强特征提取能力,但这种机制主要聚焦于通道注意力,对空间信息的提取较弱。文献[9]提出 MSAM(multiscale attention module)注意力模块,通过注意力融合,提高检测准确性,但在复杂背景环境下不同船舶类别上的检测性能存在差异。文献[10]提出多路径聚合网络结构(MPA-Net),在自底向上的特征传递过程中融合多层次特征信息,增强多尺度船舶定位能力,但该模型的性能验证依赖自定义数据集,限制了模型在实际应用中的泛化能力。文献[11]提出的多分支注意力模块使模型的检测精度有所提高,但对于模型的检测速度和实时性方面的讨论不够充分。

通道和空间注意力机制^[12]在优化特征提取方面发挥了至关重要的作用,全局注意力机制(GAM)能够更好地捕捉船舶与复杂背景之间的差异,从而提高模型在多目标和小目标场景中的识别能力。另一方面,通过使用更精确的损失函数,YOLOv5 可以精确测量锚框与实际目标框之间的重叠,并预测出更准确的目标框。此外,通过调整锚框参数,模型能够更好地处理长宽比变化较大的船舶,适应真实数据中目标形状和大小的变化,从而减少误检和漏检。受上述进展的启发,本文将 YOLOv5 算法与全局注意力机制相结合,兼顾

检测精度与速度的平衡,对损失函数及锚框参数进行进一步优化,提高检测精度的同时,加速模型的收敛,另一方面,选择 YOLOv5 作为基线,能够集中资源进行深入研究,挖掘其潜力,以解决复杂场景中的检测不足,为未来基于 YOLO 系列的进一步研究奠定基础。

本文主要介绍改进的 YOLOv5 网络结构,通过与其他基准方法进行对比的实验结果,进一步验证了本文改进方法的有效性。

1 融合全局注意力机制的多尺度船舶检测算法

1.1 算法框架

算法框架如图 1 所示,其中黄色部分为本文的改进部分。针对算法提供的初始锚框参数,采用 K-means++ 算法^[1]将 9 个不同尺度的锚框重新聚类,并将它们分别分配给小、中和大尺度特征图。经过图像预处理后,输入图像被调整为统一的 640×640 尺寸。随后,图像经过 CBS 和 C3 模块的卷积处理,得到深层次特征。在特征提取阶段,将 GAM 模块置于 SPPF 之前,通过全局注意力机制在多尺度特征融合之前获取更充分的全局上下文信息。后续实验证明这一改进既提高了特征提取的质量和丰富度,还增强了卷积核的全局感知能力。

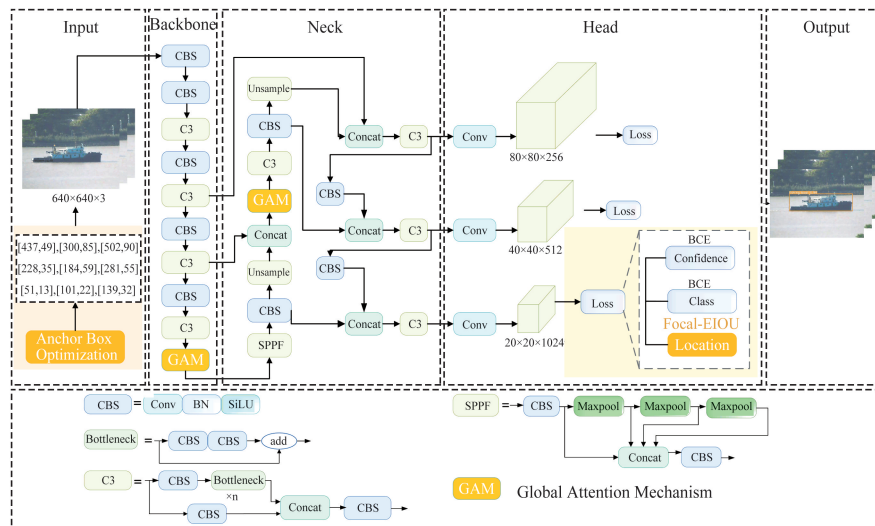


图 1 GAG-YOLO 网络结构

Fig. 1 GAG-YOLO network structure

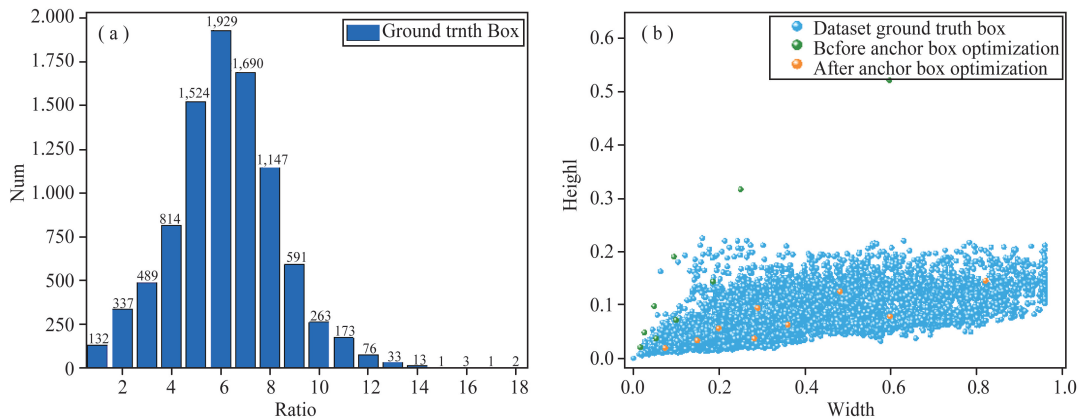
在特征融合阶段 PANet 自下而上的路径中,利用 GAM 模块增强模型对包含目标的图像区域的注意力,从而更好地融合不同尺度的特征信息。后续实验证明这一改进显著提高了模型感知不同大小目标的能力,使目标的检测和定位更加准确。

在特征预测阶段,检测头对来自三个尺度的预测信息进行处理,以得出置信度损失、分类损失和定位损失。其中 Focal-EIOU 通过^[2]调整不同样本的权重,重点关注那些对模型性能提升最为关键的高质量样本,从而加速训练过程中定位损失的收敛。后续实验证明这一改进不仅提高了模型的训练效率,更进一步提升了检测的精度和鲁棒性。

1.2 锚框参数优化

原始 YOLOv5 的锚框参数是基于 COCO^[15]数据集聚类得到的。该数据集涵盖了各种不同大小和长宽比的目标类型。然而,由于船舶仅属于其中一类,且通常具有较大的长宽比,原始锚框不适用于船舶检测。在网络训练之前,对数据集中船舶目标的几何特征进行了详细分析,例如目标的纵横比分布情况,如图 2(a) 所示,可以看出,数据集中相当一部分目标的纵横比集中在 6 左右。而原始 YOLOv5 中边界框纵横比的最大阈值默认设置为 4,因此,将该值调整为 6。随后,采用 K-means++ 聚类算法,欧氏距离度量,迭代 100 次,初始质心选择策略为距离最远优先,对数据集中的目标框重新进行聚类,得到了 9 个不同尺度的参数:

[51,13]、[101,22]、[139,32]、[228,35]、[184,59]、[281,55]、[437,49]、[300,85]和[502,90]。前三个锚框分配给小尺度特征图,中间三个分配给中尺度特征图,最后三个锚框分配给大尺度特征图。



注:(a)数据集目标框的宽高比分布;(b)优化前后数据集目标框和锚框的宽高比分布图。

图2 数据集分析

Fig.2 Dataset analysis

图2(b)中的对比图展示了优化前后的锚框分布,很明显,优化后的锚框宽度和高度值在整个数据集中的分布更加均匀,而原始锚框的尺寸则表现出更大的差异。此外,原始锚框的宽度、高度和纵横比值也与目标框的值存在偏差。YOLOv5对锚框采用了三种评估指标: B_{PR} (best possible recall,最佳召回率)、 A_{AT} (anchors above threshold,最优锚框率)和 I_{ou} (intersection over union,交并比)。 B_{PR} 和 A_{AT} 评估目标框宽高与锚框宽高之间的一致性,而 I_{ou} 则衡量真实框与锚框大小的一致性。优化后的三个评估值如表1所示: B_{PR} 提高了1.17%, A_{AT} 提高了44.97%, I_{ou} 提高了45.61%。这表明优化后的锚框在宽高比方面与船舶数据集的形状更加吻合,因此与目标框的交并比更高。

表1 锚框参数优化前后指标对比

Tab. 1 Comparison of indicators before and after optimizing anchor box parameters

Anchor Box	$B_{PR}/\%$	$A_{AT}/\%$	$I_{ou}/\%$
Pre-optimization	98.80	26.81	34.32
Post-optimization	99.97	71.78	79.93

1.3 全局注意力机制 GAM

注意力机制旨在通过引导网络在特定对象特征提取过程中关注空间或通道注意力,从而提升算法性能。然而,先前的方法^[3-4]忽视了在通道和空间方面增强跨维度交互的重要性。传统的卷积神经网络(CNN)主要依赖局部感受野来提取特征,而全局注意力机制(GAM)则通过全局池化操作捕捉整个图像的全局信息,通过减少信息分散并放大全局交互表示,提升了深度神经网络的性能。具体地,GAM能够提高模型从输入数据中捕捉全局上下文信息的能力,使其能够对整个图像范围进行全局上下文建模,而不仅仅局限于关注局部区域。

在船舶目标检测中,由于存在多艘船舶重叠、波浪、天气条件和许多其他海事环境因素,捕捉全局上下文信息对于实现准确的目标检测至关重要。因此,引入GAM来提高网络特征提取能力。GAM采用了CBAM的顺序通道-空间注意力机制,同时重新设计了子模块,以更好地处理全局信息。GAM模块的实现如图3所示。给定输入特征图 $F_1 \in \mathbb{R}^{C \times H \times W}$,中间状态 $F_2 = M_c(F_1) \otimes F_1$,输出状态 $F_3 = M_s(F_2) \otimes F_2$ 。其中, M_c 和 M_s 为通道和空间注意力子图, $\otimes$ 表示元素级乘法。

通道注意力子模块如图3(b)所示,首先对输入特征图 F_1 的每个通道进行全局平均池化,得到 $\text{avg_pool}(F_1) \in \mathbb{R}^{C \times 1 \times 1}$,将空间维度压缩为1,捕捉每个通道的全局信息,如船舶整体形状、颜色分布。全连接层 W_1 将通道数从 C 压缩到 C/r ,进行降维,减少计算量。ReLU激活引入非线性,增强模型表达能力。 W_2 将通道

数从 C/r 恢复到 C , 生成通道权重。通过两层全连接层学习通道间的依赖关系, 例如增强船舶金属纹理相关通道, 抑制海面波浪纹理通道, 当图像中包含船舶和波浪时, 通道注意力会增强与船舶边缘(高频特征)相关的通道, 抑制与波浪(低频纹理)相关的通道, 最终输出 $M_c(F_1) = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{avg_pool}(F_1)))$ 。

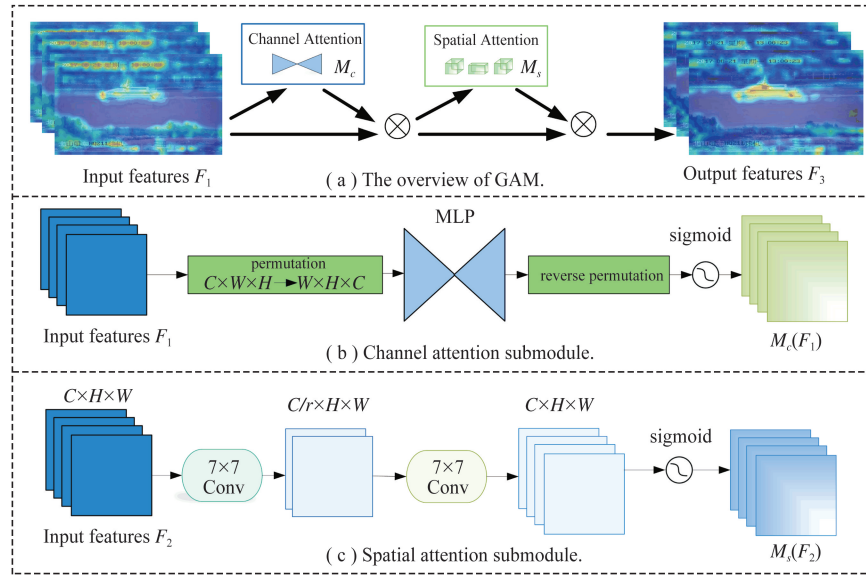


图 3 GAM 的整体结构

Fig. 3 The overall structure of GAM

空间注意力子模块如图 3(c)所示。其中采用了两层卷积层来融合空间信息并强调其重要性。其中, 第一层使用分组卷积, 每个通道独立卷积, 保留空间细节, 第二层卷积将通道数压缩为 1, 生成空间注意力图。在雾天图像中, 空间注意力会聚焦船舶轮廓区域(如桅杆、船体), 抑制均匀雾天背景的背景响应, 最终输出 $M_s(F_2) \sigma(\text{Conv}_{7 \times 7}(\text{Conv}_{7 \times 7}(F_2)))$ 。

最后, 通道注意力与空间注意力相乘, 实现 "双重聚焦", 先通过通道注意力筛选重要通道(如船舶边缘通道), 再通过空间注意力在这些通道中聚焦船舶区域, 优先增强船舶所在通道的空间区域, 将目标更好地与背景干扰区分开来。例如, 对集装箱船细长形状的检测中, 通道注意力增强水平船舶目标边缘通道, 空间注意力聚焦水平方向的长条形船舶区域。通过引入 GAM 注意力机制, YOLOv5 模型能够增强对重要特征的关注, 同时抑制背景噪声。

1.4 边界框回归损失函数优化

在船舶目标检测中, 海浪、天气条件和不同视角等因素会对目标产生不利影响, 导致目标模糊或难以识别。为增强模型处理复杂目标的能力, 提高检测的准确性和效率, 对回归损失函数进行优化。通过 Focal-EIOU 调整正负样本的权重, 有效应对复杂目标, 从而提升检测精度。Focal-EIOU 通过赋予复杂目标不同的权重, 使模型能够准确捕捉它们的位置和形状。此外, 损失函数的优化不仅加快了模型收敛速度, 还缩短了训练时间, 从而增强了模型在各种海事船舶检测任务中的适应能力, 包括在不同天气条件、摄像机视角和海洋环境下地表现。

YOLOv5 使用 CIOU 作为回归损失的度量指标。现有实验结果表明, CIOU 损失在收敛速度和检测准确性方面相比之前的损失函数有显著提高。然而, L_{CIOU} 最后一项中的变量 " v " 定义不明确, 降低了模型有效减少预测框和目标框宽高比之间真实差异的能力。与之前的 YOLO 网络类似, EIOU 损失函数由三个部分组成: I_{ou} 损失 (L_1)、距离损失 (L_{dis}) 和宽高比损失 (L_{asp})。给定一个预测框 B 和一个目标框 B^{gt} , 其中 b 和 b^{gt} 分别表示它们的中心点, 损失函数如

$$L_{\text{EIOU}} = L_1 + L_{\text{dis}} + L_{\text{asp}} = 1 - I_{\text{ou}} + \frac{\rho^2(b, b^{\text{gt}})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{\text{gt}})}{(w^c)^2} + \frac{\rho^2(h, h^{\text{gt}})}{(h^c)^2} \quad (1)$$

式中 w^c 和 h^c 表示覆盖两个框的最小外接矩形的宽和高。EIOU 损失通过直接最小化目标框和锚框之间的宽度和高度差异,加速收敛并增强定位效果,其计算原理如图 4 所示。

在边界框回归(BBR)任务中,训练样本通常存在固有的不平衡问题,即高质量样本(回归误差较小的锚框)在数量上远少于低质量样本。这种不平衡现象主要源于图像中目标物体相对稀疏的分布。因此,在网络训练过程中,有必要通过优化梯度使用,赋予高质量样本更大的权重,以提高模型的学习效果。为了应对这一挑战,并进一步提升 EIOU 损失函数的性能,本文引入了 Focal-EIOU 损失。该损失函数通过专注于高质量样本,使模型在训练时更加

关注那些定位精度较高的锚框,从而在整体上提升了检测精度和边界框回归的性能。Focal-EIOU 损失的最终表达式为 $L_{\text{Focal-EIOU}} = -\alpha(1-p_t)\gamma \cdot L_{\text{EIOU}}$,其中 α 为类别平衡因子,通过调整正负样本权重,解决船舶类别不平衡问题(如渔船样本少则 α 设为 1.67); γ 为难例挖掘指,增大 γ 可提升难例(如密集、小目标船舶)的权重,抑制易例(如清晰的大型船舶)的权重; p_t 为预测概率,通过 $(1-p_t)^\gamma$ 降低易分类样本的权重,聚焦难例,确保预测框在位置、大小和形状上更接近真实框,尤其适合船舶细长形状的回归。

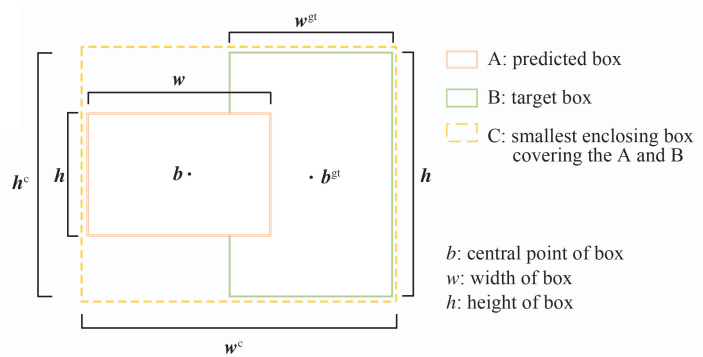


图 4 EIOU 原理

Fig. 4 Principle of EIOU

2 实验与结果分析

2.1 数据集与评价指标

为验证所提方法的性能,本文选用 SeaShips^[5]数据集进行对比实验。该数据集包含不同背景、光照条件、角度、天气条件、时间点以及不同程度船舶重叠的图像。涵盖六种常见的船舶类型:矿砂船(ore carrier)、散装货船(bulk cargo carrier)、杂货船(general cargo ship)、集装箱船(container ship)、渔船(fishing boat)和客船(passenger ship)。图像分辨率为 $1\,920 \times 1\,080$ 。整个数据集共 7 000 张图像,按 8 : 2 的比例分为训练集(5 600 张)和测试集(1 400 张)。

本文采用 $m_{\text{AP}@_{0.5, 0.95}}$ 评估所提出网络模型在船舶检测中的性能。P(Precision,精度)用于评估船舶预测的准确性,反映正确预测的正样本占有所有预测正样本的比例。R(Recall,召回率)用于评估测试数据集中所有船舶是否被准确预测,表示模型正确预测的正样本占总正样本的比例。AP 值反映了模型的准确性,值越高,识别能力越强,表示 P-R 曲线的积分。通过不同的 I_{OU} 阈值确定两个逐渐严格的 AP 指标: $\text{AP}@_{0.5}$ 和 $\text{AP}@_{0.5, 0.95}$ 。这些指标分别对应 I_{OU} 阈值为 0.5 和从 0.5 到 0.95(间隔为 0.05)的平均 AP 值。AP 值反映了特定类别的检测准确性,而 m_{AP} 通过平均各个类别的 AP 值反映了所有类别的检测准确性。 $m_{\text{AP}@_{0.5}}$ 和 $m_{\text{AP}@_{0.5, 0.95}}$ 等综合指标用于有效衡量船舶检测任务中的精度和鲁棒性。除了这些评估指标之外,还通过每秒帧数(FPS)评估模型的实时性能。

2.2 训练环境及参数配置

本文实验在 NVIDIA GeForce RTX 3060(64 GB)的设备上进行,使用 Pytorch 框架。在训练过程中,对所有模型进行了 300 个 epoch 训练, batchsize 设置为 16。考虑数据集的特性,采用权重衰减(0.000 5)和动量(0.937)的随机梯度下降法(SGD)。此外,通过马赛克数据增强将四个训练样本合并为一个样本,有效地减少了对大量小批量样本的需求。

2.3 损失函数对比试验

本文采用 Focal-EIOU 作为边界框回归损失函数,其结合 Focal 损失和 EIOU(Efficient Intersection over Union)损失的特点,其中 EIOU 损失主要用于衡量预测框与真实框之间的重叠程度,同时考虑框的位置

和大小差异。特别地,Focal 损失通过引入一个调节因子,在训练初期动态调整样本权重,将困难样本(如小目标或重叠目标)在损失计算中的贡献放大,使模型在早期阶段集中精力优化最困难的样本,从而加快了整体训练过程的收敛速度。如图 5 所示,GAG-YOLO 通过引入 Focal-EIOU,其损失曲线展现出更快的收敛速度,并达到了更低的最终收敛值。这表明 Focal-EIOU 不仅加速了模型的训练过程,还提高了模型对船舶目标定位的精度和在复杂海域中的稳定性。相比于原始 YOLOv5 模型中的损失函数,由于其无法针对性地调整难样本的权重,通常需要更多的训练轮次才能达到同样的收敛效果。

3 实验结果与分析

为验证所提出的模块和新的损失函数的有效性,将本文方法与一些其他改进方法在 SeaShips 数据集上进行了一系列消融实验。改进方法在原骨干网络及颈部网络部分融入了 SE^[16],CBAM^[12]和 CA^[18]三种不同的注意力机制,上述的 3 种改进方法均在加入锚框参数优化和 Focal-EIOU 损失函数的结构中实现,最后与本文改进算法进行比对。将锚框优化命名为 ABO,实验结果如表 2 所示。表中中间几行表示各种消融实验。在表 2 中,“√”表示使用了当前模块,“—”则表示未使用。实验结果表明,以上 8 种改进均能提高检测算法的性能。就 m_{AP} 值而言,在损失函数相同及锚框参数优化的情况下,注意力机制 GAM 相对 SE、CBAM 能够为原算法带来更多提升。故选择全局注意力机制 GAM。直接采用 YOLOv5 时,得到的 m_{AP} 值为 82.23%。在这种情况下,引入 GAM 模块使 m_{AP} 值提高了 2.7%。进一步通过优化的损失函数使网络有效收敛, m_{AP} 值更是提高了 3.6%。此外,观察 F_{PS} 指标:引入 GAM 模块略微降低了计算速度(F_{PS} 从 90 降至 74),但优化的损失函数和 ABO 模块均提高了计算效率(F_{PS} 从 90 增加到 93)。

对比结果表明,本文在三个方面的综合改进最大限度地提升了原始算法的目标检测性能。在 SeaShips 数据集上, m_{AP} 从初始的 82.23%增加到 86.83%,显著提升 4.6%。总的来说,所提出的模型在不增加额外计算负担的情况下,实现了更高的检测准确率。

表 2 消融实验

Tab. 2 Ablation experiments

Methods	SE	CBAM	CA	ABO	GAM	Focal-EIOU	$P/\%$	$R/\%$	$m_{AP@0.5}$ / $\%$	$m_{AP@0.5:0.95}$ / $\%$	F_{PS}
YOLOv5s	—	—	—	√	—	—	97.72	97.92	99.14	82.23	90
Improvement 1	—	—	—	√	—	—	98.58	97.77	98.72	85.51	93
Improvement 2	—	—	—	—	√	—	97.81	97.43	98.37	84.49	74
Improvement 3	—	—	—	—	—	√	98.81	97.33	98.28	85.19	93
Improvement 4	—	—	—	—	√	—	98.95	98.70	99.04	86.18	73
Improvement 5	—	—	—	—	√	√	99.05	97.43	98.26	85.83	74
Improvement 6	√	—	—	√	—	√	98.97	97.72	97.82	85.77	88
Improvement 7	—	√	—	√	—	√	99.08	98.93	98.75	85.85	90
Improvement 8	—	—	√	√	—	√	99.03	98.82	98.98	85.92	91
GAG-YOLO	—	—	—	√	√	√	99.15	99.33	99.34	86.83	87

表 3 列出了 6 种类型船舶的检测结果。与最初的 YOLOv5 相比,GAG-YOLO 在检测矿砂船方面表现出色, $m_{AP@0.5}$ 、 $m_{AP@0.5:0.95}$ 、 P 和 R 四个指标上都有显著提高,其中 $m_{AP@0.5:0.95}$ 指标显著提高了 6.7%。对于

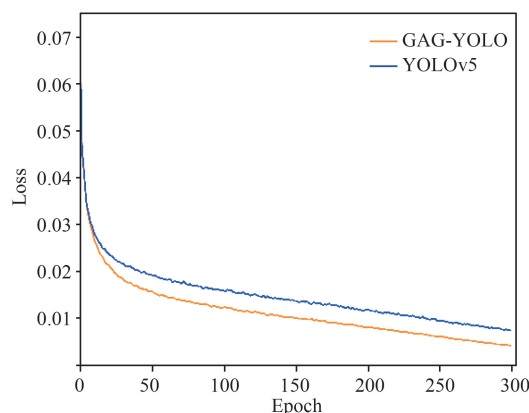


图 5 边界框回归损失曲线

Fig. 5 Curve of training loss function

普通货船和集装箱船,GAG-YOLO 的 P 和 R 指标都有大幅提高,特别是集装箱船的 P 和 R 值均达到 1。总体而言,GAG-YOLO 在散货船、集装箱船、矿砂船、客船和普通货船上的表现都非常出色。

表 3 YOLOv5 与 GAG-YOLO 对六种类型船舶的检测结果
Tab. 3 Detection results of each model for different type of ship

Methods	Targets	Bulk cargo carrier	Container ship	Fishing boat	General cargo ship	Ore carrier	Passenger ship
YOLOv5s	$mAP@_{0.5} / \%$	98.6	99.4	98.4	99.1	93.6	99.4
	$mAP@_{0.5:0.95} / \%$	86.2	87.4	80.3	87.1	79.8	84.7
	$P / \%$	96.7	99.5	98.1	98.6	97.7	99.1
	$R / \%$	97.9	98.1	97.2	98.3	89.1	99.1
GAG-YOLO	$mAP@_{0.5} / \%$	99.2	99.5	99.3	99.5	99.3	99.2
	$mAP@_{0.5:0.95} / \%$	88.2	90.4	81.3	88.6	85.2	87.2
	$P / \%$	98.2	100	98.5	100	98.6	99.1
	$R / \%$	98.7	100	97.6	98.4	97.9	100

为了展示 GAG-YOLO 模型在不同环境条件下的通用性和鲁棒性,从测试集中选取了雨天、雾天和傍晚条件下的检测结果,如图 6 所示。检测结果表明,GAG-YOLO 在迷雾环境中能够以较高的置信度检测到多个重叠的船舶目标,显著改善了漏检和误检的情况。通过引入 GAM,模型能够有效捕捉全局上下文信息,使其在复杂背景下更好地辨别船舶目标与干扰物之间的差异。在光线条件不足的情况下,GAG-YOLO 也展现出优越的定位能力,能够准确识别部分被遮挡的船舶目标。这得益于优化后的锚框参数和 Focal-EIOU 损失函数,使模型在处理高质量样本时更加专注,减少了对容易预测样本的依赖。然而,尽管 GAG-YOLO 在复杂场景下表现出色,但在极端天气条件(如强光反射、严重遮挡)下,模型的检测精度仍有提升空间,未来的研究可进一步探索这些情况下的优化策略。



图 6 GAG-YOLO 检测结果

Fig. 6 GAG-YOLO detection results

然而,尽管 GAG-YOLO 在复杂场景下表现出色,但在极端天气条件(如强光反射、严重遮挡)下,模型的检测精度仍有提升空间,未来的研究可进一步探索这些情况下的优化策略。

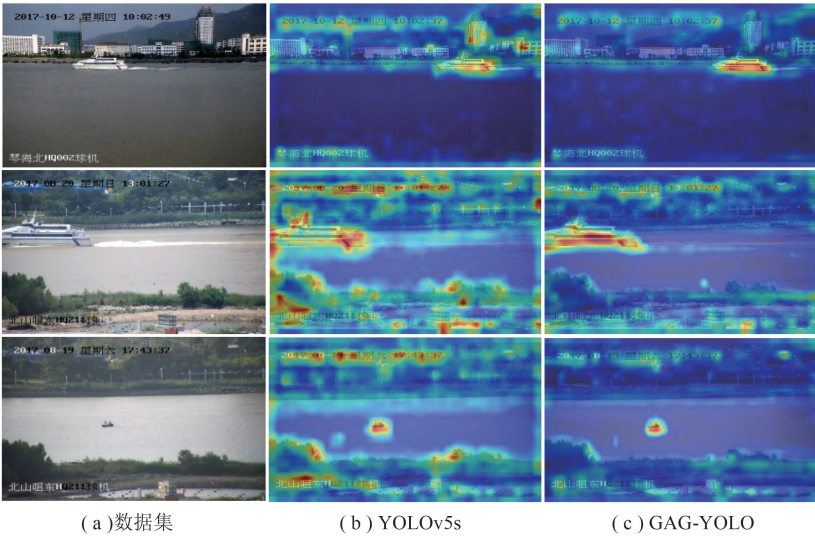


图 7 热力图检测结果

Fig. 7 Heatmap test results

图7的热力图检测结果可以看出,GAG-YOLO在关注船舶目标的区域上更加突出,无论是远处小目标还是与背景重叠的大目标,都在热力图上表现为更加清晰集中的高响应区域,而非目标区域(如海面、天空、岸边植被与建筑)的响应强度明显降低。这意味着GAM的全局建模能力使船舶目标的边缘和细节部分被更好的捕捉,并有效抑制了背景噪声的影响。总体而言,GAG-YOLO在复杂场景下也能高质量地完成船舶检测任务。

为验证所提模型的先进性和整体性能,将GAG-YOLO与YOLO系列中的其他先进方法进行对比实验,包括YOLOv7、YOLOv8n、YOLOv8s及YOLOv9s,并将GAG-YOLO算法与Transformer^[19]及Transformer模型变体DETR^[20]进行比较。为全面展示GAG-YOLO在不同实际应用场景中的适用性,图8展示了多种典型场景下的检测结果,涵盖了多种复杂背景元素和照明条件,如港口设施、雨天、雾天以及傍晚黄昏等复杂环境。由检测结果可见,GAG-YOLO在检测精度方面显著优于其他对比方法。具体而言,在小尺度目标(第1列)、大尺度目标(第2列)和多目标场景(第3列)中,可以明显观察到,YOLOv8n、YOLOv8s和YOLOv9模型存在不同程度的误检或漏检现象,而YOLOv7在多目标检测中也存在漏检问题。相比之下,GAG-YOLO在这些场景中实现了更精确和可靠的检测结果。在第四列展示的雨雾条件下,尽管所有模型均能成功检测到目标,YOLOv8n、YOLOv8s和YOLOv9仍存在一定程度的误检现象。第五列中的傍晚黄昏条件下,YOLOv8n、YOLOv8s同样表现出误检问题。GAG-YOLO在这些复杂环境下表现出更优越的检测效果,其标注的边界框更为紧贴目标物,准确反映了目标的真实位置和大小,显著提升了检测的精度和可靠性。综合来看,GAG-YOLO在多种典型场景下均表现出优越的检测性能,尤其是在复杂环境和多目标检测中表现突出,相较于其他最新YOLO版本方法,进一步提高了目标检测的准确性和鲁棒性,但在多目标重叠场景中的检测置信度还有待进一步提高。



注:(a) 原始图像;(b) YOLOv7;(c) YOLOv8n;(d) YOLOv8s;(e) YOLOv9s;(f) GAG-YOLO。

图8 GAG-YOLO和YOLO系列不同模型的检测结果

Fig. 8 Detection results of different models of GAG-YOLO and YOLO series

由表4结果可见,与其他先进方法相比,GAG-YOLO在所有指标上均优于YOLO系列模型, $m_{AP@0.5,0.95}$ 比

次优的 YOLOv8s 高 1.31%, 与 Transformer 模型相比, GAG-YOLO 的 $m_{AP@0.5:0.95}$ 分别比 DETR 和 Swin Transformer 高 8.63% 和 5.33%, 且 F_{PS} 提升 3 倍以上。对比结果可见, GAG-YOLO 在检测性能与效率的平衡方面更加出色。

为了更好地验证改进模型的泛化性能, 选取包含更多小型船舶(平均尺寸比 SeaShips 数据集小 30%), 且背景复杂(港口、岛屿等)的 SAR 数据集 SSDD^[21] 进行模型测试, 表 5 结果显示, GAG-YOLO 在未见过的数据集上仍保持 83.5% 的 $m_{AP@0.5:0.95}$, 验证了其跨数据集泛化能力。

表 4 不同先进模型的检测结果对比

Tab. 4 Detection performance of different methods

Methods	$P/\%$	$R/\%$	$m_{AP@0.5}/\%$	$m_{AP@0.5:0.95}/\%$	F_{PS}
YOLOv7	97.68	98.42	99.41	83.68	46
YOLOv8n	98.45	95.82	98.95	82.62	110
YOLOv8s	98.93	98.63	99.26	85.52	95
YOLOv9s	98.95	98.52	99.13	85.48	108
DETR	94.12	92.35	97.28	78.20	28
Swin Transformer	96.54	95.87	98.43	81.50	35
GAG-YOLO	99.15	99.33	99.34	86.83	87

表 5 在 SSDD 数据集上的测试结果

Tab. 5 Detection results of SSDD dataset

Methods	$P/\%$	$R/\%$	$m_{AP@0.5}/\%$	$m_{AP@0.5:0.95}/\%$
YOLOv5s	95.12	93.45	97.02	79.30
GAG-YOLO	98.24	97.56	98.87	83.50

4 结论

为解决复杂海上场景中存在的小目标漏检和多目标检测性能不佳的问题, 提出一种全局注意力引导的多尺度船舶检测算法。首先, 引入全局注意力机制捕捉更详细的船舶目标信息, 从而增强对目标与复杂背景的区分, 提高检测的准确性。其次, 结合高效的边界框回归损失函数, 通过分配不同的权重, 提高对高质量样本目标区域的关注度, 从而减少误检并加速算法收敛。最后, 对锚框进行优化, 增强模型对不同船只类型的适应性, 减轻背景干扰, 进一步提高检测精度。经验证, GAG-YOLO 在平均精度均值方面达到了 99.34%, 具有比其他基准方法更高的准确性和鲁棒性。

基于可见光图像的船舶检测算法应用面临着复杂场景中目标遮挡和密集分布等诸多问题。未来的研究需要进一步探索高密度多尺度船舶目标检测算法, 以及参数优化和检测速度的提升。特别地, 未来将利用公开多模态数据集或开发合成数据生成技术, 探索基于注意力机制的多模态融合方法, 以提升极端天气下的检测鲁棒性。

参 考 文 献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.
- [2] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [3] REN S, HE K, 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.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [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. 2016: 779-788.
- [6] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations [DB/OL]. Zenodo, 2022. [2024-05-20]. <https://doi.org/10.5281/zenodo.7359008>.
- [7] 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. 2018: 8759-8768.
- [8] NING Y, ZHAO L, ZHANG C, et al. STD-Yolov5: a ship-type detection model based on improved Yolov5[J]. Ships and Offshore Structures, 2024, 19(1): 66-75.
- [9] 萧箏, 王继业, 夏叶亮. 基于 MSAM-YOLOv5 的内河航道船舶识别方法 [J]. 华中科技大学学报(自然科学版), 2023, 51(5): 67-73.
- [10] 王文亮, 李延祥, 张一帆, 等. MPANet-YOLOv5: 多路径聚合网络复杂海域目标检测 [J]. 湖南大学学报(自然科学版), 2022, 49(10): 69-76.
- [11] 胡欣, 马丽军. 基于 YOLOv5 的多分支注意力 SAR 图像舰船检测[J]. 电子测量与仪器学报, 2022, 36(8): 141-149.
- [12] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [13] VASSILVITSKII S, ARTHUR D. K-means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2007: 1027-1035.
- [14] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [15] LIN TY, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, 2014: 740-755. Springer.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE Press, 2018: 7132-7141.
- [17] SHAO Z, WU W, WANG Z, et al. Seaships: A large-scale precisely annotated dataset for ship detection [J]. IEEE Transactions on Multimedia. Piscataway: IEEE, 2018, 20(10): 2593-2604.
- [18] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern recognition. Nashville, Tennessee, USA: IEEE, 2021: 13713-13722.
- [19] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [20] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 10012-10022.
- [21] ZHANG T, ZHANG X, LI J, et al. SAR ship detection dataset (SSDD): official release and comprehensive data analysis[J]. Remote Sensing, 2021, 13(18): 3690.

(责任编辑:赵海军)