

中图分类号: TP391.41; TP18 文献标识码: A 文章编号: 1006-8961(2026)05-1569-14

论文引用格式: Wang Y J, Yang X P, Zhou L, Lu H X, Zhao W Y and Zhang W D. 2026. Cross-scale adaptive frequency domain enhancement for maritime vessel detection. Journal of Image and Graphics, 31(5):1569-1582(王应军, 杨晓鹏, 周玲, 路皓翔, 赵文义, 张卫东. 2026. 跨尺度自适应频域增强的海上船舶检测. 中国图象图形学报, 31(5):1569-1582)[DOI:10.11834/jig.250548]

## 跨尺度自适应频域增强的海上船舶检测

王应军<sup>1</sup>, 杨晓鹏<sup>1</sup>, 周玲<sup>1,2</sup>, 路皓翔<sup>3</sup>, 赵文义<sup>4</sup>, 张卫东<sup>1,5\*</sup>

1. 河南科技学院信息工程学院, 新乡 453003; 2. 上海大学机电工程与自动化学院, 上海 200444;
3. 桂林电子科技大学计算机与信息安全学院, 桂林 541000; 4. 北京邮电大学人工智能学院, 北京 100876;
5. 郑州大学电气与信息工程学院, 郑州 450000

**摘要:** 目的 海上船舶目标检测对海域管理和交通安全至关重要, 但受复杂环境影响, 常出现遮挡、模糊和细节丢失等问题, 现有方法检测精度不足、易误检漏检, 难以满足船舶精确识别需求。基于此, 提出一种跨尺度自适应频域增强的海上船舶检测方法。**方法** 以YOLO11(you only look once)为基线模型进行针对性改进, 首先, 设计了一个自适应频域特征增强模块(adaptive frequency-domain feature enhancement module, AFEM)用于海上船舶细节特征的增强。该模块针对不同尺度的特征信息, 采用傅里叶变换将特征信息转换到频域, 通过门控单元对全局和局部信息进行自适应增强, 全面增强网络对海上退化特征的提取能力。其次, 在颈部引入一个多尺度特征感知模块(multi-scale feature perception module, MFP)。使用不同的卷积核捕获多尺度特征, 高效挖掘并利用海上船舶图像的上下文特征信息, 引导网络精准聚焦船舶目标特征, 有效抑制复杂背景与遮挡带来的干扰, 缓解小目标船舶的特征丢失现象, 显著降低海上船舶检测的错检与漏检率。**结果** 在MVDD(marine vessel detection dataset)和RTTS(real-world task-driven testing set)数据集上的平均精确度(mean average precision at 50% IOU, mAP50)分别达到95.18%和74.79%, 对13类船舶的检测表现优异, 尤其在小目标、遮挡船舶检测中优势显著。同时, 参数量仅有6.29 M, 推理速度达到227 帧/s。通过与先进的16种不同类型方法的比较, 本文方法检测性能更优, 在检测精度和模型复杂度之间实现了更好的平衡。**结论** 本文方法不仅在海上表现出色, 对于陆地的恶劣天气条件也有较强的适应能力, 展现出较好的鲁棒性和泛化性, 同时具备较高的可部署性和实际应用价值。

**关键词:** 频域特征增强; 傅里叶变换; 复杂背景干扰; 目标检测; 退化特征; 多尺度特征感知

## Cross-scale adaptive frequency domain enhancement for maritime vessel detection

Wang Yingjun<sup>1</sup>, Yang Xiaopeng<sup>1</sup>, Zhou Ling<sup>1,2</sup>, Lu Haoxiang<sup>3</sup>, Zhao Wenyi<sup>4</sup>, Zhang Weidong<sup>1,5\*</sup>

1. School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China;
2. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;

收稿日期: 2025-11-05; 修回日期: 2026-01-29; 预印本日期: 2026-02-04

\* 通信作者: 张卫东 zwd\_wd@163.com

**基金项目:** 中国博士后科学基金项目(2024M750747); 河南省科技研发计划联合基金青年科学家项目(235200810066); 河南省科技攻关项目(252102211003); 河南省高等学校青年骨干教师培养计划项目(2025GGJS099); 河南省高等教育教学改革研究与实践项目(研究生教育类)(2025SJKLX208Y)

**Supported by:** China Postdoctoral Science Foundation Project(2024M750747); Henan Provincial Science and Technology Research and Development Joint Foundation Project(235200810066); Science and Technology Research Project of Henan Province(252102211003); The Project of Training Young Backbone Teachers in Higher Education Institutions of Henan Province(2025GGJS099); Practice Project of the Higher Education Teaching Reform in Henan Province(Postgraduate Education)(2025SJKLX208Y)

3. School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541000, China;
4. School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China;
5. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450000, China

**Abstract: Objective** Ship detection plays a crucial role in modern maritime governance, traffic safety assurance, and the precise classification of civilian and military vessels. With the rapid growth of global maritime trade, particularly China's shipping industry, ensuring maritime situational awareness and vessel tracking has become increasingly important. However, ship detection in open-sea environments is a challenging task due to the highly dynamic and complex nature of maritime scenes. Real-world conditions often include frequent inter-ship occlusions, severe weather disturbances, low visibility, and image degradation due to fog, motion blur, or sunlight glare. Additionally, cluttered backgrounds comprising waves, clouds, and reflections further complicate detection. These environmental complexities introduce visual ambiguities, including unclear object boundaries, distorted color information, and the loss of fine structural features, especially for small targets such as fishing boats, patrol craft, or distant vessels. Traditional object detection algorithms, including two-stage models such as Faster R-CNN and early one-stage models such as YOLOv3 and YOLOv4, tend to exhibit low boundary precision, high false-positive rates, and frequent missed detections under small or occluded ships. These limitations hinder accurate maritime monitoring and fail to meet the high standards required in civilian scenarios and military applications. **Method** Aiming to address the aforementioned challenges, using YOLO11 as the baseline model with targeted improvements, this paper introduces a ship detection method designed for complex maritime environments by integrating cross-scale perception with adaptive frequency-domain feature enhancement. While maintaining computational efficiency, the proposed method introduces two lightweight but effective modules to enhance robustness, localization accuracy, and overall detection performance. The first core component is the adaptive frequency-domain feature enhancement module (AFEM), which is designed to alleviate feature degradation in complex maritime scenes. AFEM enables effective separation of global structural information and local texture details by applying a Fourier transform to map spatial features into the frequency domain, thereby enhancing high-frequency edge information that is often weakened by conventional spatial convolutions. A dual-branch gated fusion mechanism is employed to adaptively model global and local frequency components: the global branch captures long-range semantic dependencies to support large-scale ship recognition under low-visibility conditions, while the local branch preserves high-frequency details to improve detection of small and low-saliency targets. Additionally, a linear gated convolutional unit dynamically regulates the contributions of different frequency components, enhancing robustness to noise, fog, motion blur, and uneven illumination. The second key component is the multiscale feature perception (MFP) module, designed to address substantial scale variation and frequent occlusions in maritime scenes. Aiming to accommodate ships of diverse sizes and limited visibility, the MFP module employs convolutional kernels with multiple receptive fields, including standard, asymmetric elongated, and dilated convolutions, enabling joint modeling of fine-grained local features and broader contextual information. These multiscale features are then fused through a channel-wise attention mechanism, which highlights ship-related regions while suppressing background clutter such as sea foam, reflections, and distant coastlines, thereby maintaining discriminative capability across scales. Furthermore, AFEM and MFP are designed with practical deployment in mind. Through parameter sharing and lightweight convolutional structures, the proposed method achieves performance improvements without substantially increasing model complexity, increasing its suitability for maritime surveillance and intelligent perception systems that require high accuracy and real-time efficiency. **Result** Extensive experiments on the MVDD (maritime vessel detection dataset) and RTTS (real-world task-driven testing set) datasets demonstrate the effectiveness of the proposed method. Experimental results show the excellent performance of the method in detecting 13 types of ships, revealing particularly remarkable advantages in small-target and occluded ship detection. On the MVDD dataset, the model achieves 95.18% mAP50, outperforming state-of-the-art models such as YOLOv5 and D-FINE, and achieving a 1.16% higher recall compared to YOLOv12s. On the RTTS dataset, the method attains 74.79% mAP50, surpassing YOLOv10s and performing comparably to specialized methods such as AMSP-UOD. The network is highly efficient, containing only 6.29 million parameters—notably fewer than Faster R-CNN (206.68 M) while achieving 227 frame/s, thereby exceeding YOLOv12s and D-FINE. Ablation studies confirm the contributions of

AFEM and MFP, revealing performance drops of 1.29% and 0.90% upon removal, respectively. **Conclusion** The proposed method balances detection accuracy, robustness, and computational efficiency. This method addresses critical limitations in existing ship detectors, particularly under adverse maritime conditions, by effectively enhancing degraded features and capturing multiscale contextual information. Beyond ocean-based applications, the model demonstrates strong adaptability to terrestrial scenarios with similar visual challenges, such as fog, snow, and poor illumination, revealing its extensive application potential in traffic surveillance, disaster response, and environmental monitoring. Owing to its lightweight design and high-speed inference, the system can be seamlessly deployed in real-time platforms, including unmanned surface vehicles, smart port surveillance, maritime radar systems, and ecological observation devices. Future work will focus on extending the framework to multi-modal ship detection through the integration of infrared and radar data, thereby improving detection reliability in extreme conditions such as nighttime, storms, and heavy fog.

**Key words:** frequency domain feature enhancement; Fourier transform; complex background interference; object detection; degradation characteristics; multi-scale feature perception

## 0 引言

随着神经网络和人工智能的不断发展,计算机视觉逐渐得到人们的重视。目标检测作为计算机视觉的主要领域之一,也得到了空前的发展。目前已经得到广泛应用,其应用领域包括自动驾驶(Gong等,2025)、遥感影像分析(Zhang等,2024;Zhou等,2023a)以及安全监控(Phan等,2025)等。在众多应用场景中,目标检测技术凭借对复杂环境下多类目标的精准识别能力,逐步构建起成熟的技术体系。然而,当视线投向广袤无垠的海洋领域时,特殊的场景特性为这一技术带来了全新的挑战与机遇。

在海洋环境中,船舶作为重要的水上移动载体,其动态监测对于保障航海安全、打击非法越界以及应对海上救援等具有不可替代的作用。与陆地场景不同,海上船舶检测需面对多重核心挑战(见图1):除海浪扰动导致的目标姿态剧烈变化、复杂光照引发的特征模糊、远距离船舶成像尺度过小外,船舶间的相互遮挡和雨雾等恶劣天气造成的图像退化也是制约检测精度的关键瓶颈。传统图像处理方法依赖滑动窗口识别目标,在复杂海洋背景下易导致同一物体在不同窗口的特征差异显著,鲁棒性不足,难以应对遮挡和退化问题,漏检、误检率较高,无法满足实用化需求。随着深度学习的发展,目标检测算法逐步应用于海洋领域,但现有技术仍存在明显局限:双阶段检测算法(Girshick,2015;Ren等,2017;Lin等,2020;Zhao等,2025)通过生成候选区域再分类,对小目标有一定效果,但计算量大、推理效率低,难以满足实时检测的需求,且对遮挡场景中目标边缘

特征的区分能力较弱,当船舶部分被遮挡时,候选区域易包含冗余背景或丢失关键特征,导致分类偏差。

单阶段检测算法,如SSD(single shot multi-box detector)(Liu等,2016)及YOLO(you only look once)等凭借实时性优势在海洋领域快速应用。随后,其相关变体(Wang等,2021;Huang等,2023)利用迁移学习、注意力机制、物理模型(张卫东等,2025)等方法对YOLO进行改进,在一定程度上提升了小目标物体的检测能力,但仍缺乏对遮挡目标间空间关系的建模。当船舶相互遮挡时,算法难以区分重叠区域的归属,易将多船误判为单船或漏检被遮挡部分。同时,其在空域直接处理退化图像时难以分离雨雾噪声与船舶有效特征,导致退化区域的目标特征被弱化。

Transformer模型的引入为目标关系建模提供了新思路,DETR(end-to-end object detection with Transformers)(Carion等,2020)和S-DETR(Xing等,2023)通过全局注意力捕捉目标间关联,一定程度上缓解了遮挡问题,但全局注意力易被背景干扰,对遮挡区域的局部细节关注不足,且未针对图像退化的本质机理设计解决方案。退化图像在空间域多表现为对比度下降、细节模糊,而现有方法仅在空间域进行特征增强,无法精准定位并抑制退化成分。此外,Transformer模型收敛速度慢的问题也限制了其在实时海洋监测中的应用。针对上述核心挑战,现有方法尚未形成对目标遮挡以及特征退化等复合挑战的有效解决方案。遮挡问题的本质是目标特征的空间重叠与归属模糊,需同时强化目标间关系建模与局部遮挡边界的精细感知,而图像退化的核心是噪声与有效特征的混合,需从频域特性等信号本质入手实现精准分离。

基于此,本文提出端到端的跨尺度频域增强网络。通过多尺度特征感知模块捕捉目标空间关系与遮挡边界,结合自适应频域增强模块从频域分离并抑制退化成分,最终实现复杂海洋环境下船舶的高效精准检测。具体而言,本文贡献如下:1)针对海面环境复杂多样,精度与效率难以平衡的问题,提出一个端到端的海上船舶目标检测网络,在保持较高检测精度的同时兼顾轻量化,不仅对海面环境有更好的适应能力,而且在陆地上特征退化场景中也表现出色。2)针对海面图像退化造成的特征丢失问题,提出一个自适应频域特征增强模块。对于海面船舶的退化信息,使用傅里叶变换将退化特征转换到频域,根据噪声和有效信息在频域的不同分布规律,通过门控单元调整不同频段权重,实现退化特征信息的自适应增强。3)针对海面小目标漏检、遮挡目标误检问题,引入一个多尺度特征感知模块。对于海面不同尺寸的目标信息,采用差异化卷积核实现小目标纹理、大目标上下文特征信息的感知,并通过注意力抑制背景噪声凸显有效信息,增强对小目标信息的捕获能力,减少误检、漏检。



图1 海上目标检测面临的挑战

Fig. 1 Challenges in maritime target detection

## 1 相关工作

本节对基于卷积神经网络(convolutional neural network, CNN)和Transformer的海上船舶目标检测方法进行综述,对比分析了其技术特点、性能瓶颈以及前沿发展动态。

### 1.1 基于卷积神经网络的方法

基于卷积神经网络的目标检测方法按照检测流程与候选区域生成方式可以分为单阶段和双阶段。

#### 1.1.1 双阶段检测方法

许多研究人员通过对经典双阶段检测算法进行各种改进以弥补在海上复杂环境的精度和鲁棒性不

足,将其变体应用到海上船舶检测。为了更好地检测海上小目标物体,Lin等人(2019)通过融入压缩激励机制,进一步优化了用于船舶检测的Faster R-CNN(faster region-based convolutional neural network)模型,使其在训练阶段能够将注意力集中于关键特征上。Yu等人(2022)提出一种基于区域的船舶识别算法,该算法借助通道注意力机制和正则化权重改进特征融合过程,并对非极大值抑制(non-maximum suppression, NMS)方法进行优化,以增强小目标检测效果。Si等人(2023)利用K-means++生成适配多尺度船舶的候选框,通过改进的特征融合模块和注意力机制实现海上船舶的精细回归与分类。杨镇宇和石刘(2024)基于Cascade R-CNN更换主干网络,采用Soft-NMS(soft-non maximum suppression)非极大值抑制方法提高候选框的过滤精度,有效增强船舶检测的精度。Jia等人(2024)利用传统图像处理方法先对候选区域预筛选,再通过假警报过滤网络对检测结果进一步过滤,显著提升了检测准确度。双阶段检测算法虽然明显提高了检测精度、缓解了背景干扰。然而,复杂的检测流程也带来了较大的计算量,导致检测速度慢,工程部署难度高。

#### 1.1.2 单阶段检测方法

为了提高检测速度,达到工业应用场景的实时性要求,单阶段检测算法逐渐成为研究的主流趋势。张炳焱等人(2024)通过引入轻量化骨干网络、注意力机制及动态边界框回归损失,显著提升了复杂场景下船舶检测的精度与速度。肖振久等人(2024)设计融合多重机制的合成孔径雷达舰船检测框架,在YOLOv7中集成U-Net降噪模块,并进一步改进检测头,有效抑制合成孔径雷达图像噪声,同时增强多尺度特征表达。Gao等人(2025)在YOLOv5检测头中新增角度分类分支,并提出概率与分布损失函数,有效减轻遥感图像中因船舶目标小、分布密集引发的检测难题。Zhang等人(2023)基于YOLOv5设计频率注意力模块,自适应挖掘和利用频域信息,有效抑制海波对检测结果的影响,增强网络的鲁棒性和检测精度。Zhang和Er(2024)提出了一种不对称聚合特征金字塔网络,通过自顶向下的语义聚合与自底向上的细节增强,进行不同层级之间特征图的双向信息传递,在保证实时性的同时兼顾模型的鲁棒性。Wen等人(2023)基于SSD框架,通过多尺度特征融合和上下文感知模块,提升复杂背景下多尺度船舶

的检测精度。单阶段检测算法凭借端到端检测流程大幅提升了推理速度,但在复杂海洋环境下,对退化特征、小目标及遮挡目标的检测能力仍有待提升。值得注意的是,其轻量化的设计值得借鉴。

## 1.2 基于Transformer的方法

基于Transformer的目标检测方法凭借自注意力机制强大的全局上下文建模与长距离依赖捕捉能力,为海上船舶检测提供了新的思路。Dai等人(2021)通过随机补丁检测的无监督预训练任务,解决了DETR对大规模数据和长训练时间的依赖问题,对于海上船舶检测能够达到更快收敛和更高精度。李可欣等人(2024)将Transformer的自注意力机制与长短期记忆(long short-term memory, LSTM)的时序建模能力结合,实现对船舶轨迹长序列特征的有效提取。Qin等人(2025)引入了对比去噪训练机制,通过利用噪声query的重构过程有效减少了小目标与复杂背景之间的混淆。同时,采用LTTM(look twice toward maximum)算法对目标框进行迭代精细化处理,从而显著提升了对小型船舶的检测性能。Li等人(2024)融合CNN的局部特征提取能力与Transformer的全局建模优势,构建全局一局部混合架构,有效提升合成孔径雷达图像中小型船舶的检测性能。马枫等人(2024)将Swin Transformer与CSPDarknet53融合构建特征提取网络,通过自注意力模块强化船舶与航道、岸线的关联性,有效缓解远距离船舶特征模糊与背景噪声干扰问题。Yuan等人(2024)提出空间一通道交叉Transformer块,并结合互补前馈网络,增强红外图像中船舶与背景的语义区分能力,提高小尺寸船舶检测精度。基于Transformer的船舶检测方法依赖大规模标注数据,导致模型收敛效率偏低,且对局部细节的建模能力不足,难以同时兼顾小目标检测性能与复杂背景抗干扰能力。

总的来说,基于CNN的双阶段、单阶段方法,以及基于Transformer的方法,在复杂海洋船舶检测任务中均存在三大核心瓶颈:1)对雨雾、低光照导致的图像退化缺乏针对性方案,现有频域增强策略缺乏动态适配性,未能充分利用频域特性精准分离噪声与有效特征;2)小目标与遮挡目标检测能力不足,现有多尺度融合策略多采用简单特征拼接方式,易造成语义信息损失,Transformer方法易受背景干扰,二者均难以兼顾全局上下文与局部细节感知;3)精度

与实时性难以平衡,双阶段CNN和Transformer模型计算量大、收敛慢,单阶段CNN模型牺牲复杂场景鲁棒性,部分频域与多尺度策略易引入冗余计算,制约实时部署。

针对上述核心瓶颈,本文提出一种跨尺度自适应频域增强的海上船舶检测方法,通过自适应频域特征增强与多尺度特征感知模块的协同设计,实现复杂海洋场景下船舶的高效精准检测。

## 2 本文方法

在本节中,主要介绍本文方法的整体架构,以及提出的自适应频域特征增强模块和多尺度特征感知模块的工作原理和设计流程。

### 2.1 网络整体架构

本文提出的跨尺度自适应频域增强网络以YOLO11(Jocher和Qiu, 2024)为基线进行针对性改进,以解决海上船舶检测“低能见度导致特征模糊、复杂背景下目标尺度多变、模型难部署”三大核心问题为目标,整体架构如图2所示。主要包含骨干网络、颈部网络与检测头3部分。

区别于现有频域模块需先经通道降维再进行频域处理易丢失细节信息、采用单一尺度频域加权、以及依赖预设参数缺乏动态适配能力等局限,本文在骨干网络不同尺度特征层嵌入自适应频域特征增强模块(adaptive frequency-domain feature enhancement module, AFEM),无需降维操作,直接通过傅里叶变换将空域特征映射至频域,再借助关键频段动态选择机制精准定位退化噪声集中频段与船舶有效特征频段,结合线性门控卷积实现有效特征自适应强化,最后通过残差融合保留空域原始细节,显著提升对低对比度船舶的感知能力。

同时,针对传统多尺度模块固定层级特征融合难以适配船舶尺度跨度的不足、缺乏针对遮挡场景的差异化采样设计,本文在颈部网络引入的多尺度特征感知模块(multi-scale feature perception module, MFP),如图3所示,以轻量结构为核心,采用3种差异化卷积核分支( $3 \times 3$ 捕捉小目标纹理、 $1 \times 3$ 与 $3 \times 1$ 分别提取水平、垂直方向的中尺度遮挡边界)全面覆盖不同尺寸船舶特征需求,并结合通道注意力抑制海浪、泡沫等冗余背景干扰,强化船舶特征表达。此外,架构设计兼顾部署实用性,可满足港口监控、近

岸巡逻等地面可见光场景的实时部署需求。

## 2.2 自适应频域特征增强模块

针对海上复杂雾天环境中船舶目标特征模糊、边缘不清的问题,本文提出了一种自适应频域特征增强模块(AFEM),旨在提升模型对低对比度目标的感知能力。该模块从频域的角度出发,针对不同尺度的特征

信息进行自适应增强。同时,为了加强网络的可部署性,该模块设计结构简洁、计算高效。如图2所示,其核心包括3部分:频域特征解码(frequency-domain feature decoding, FFD)、频域感知增强(frequency-perception domain enhancement, FPDE)以及线性门控卷积单元(linear gated convolutional unit, LGCU)。

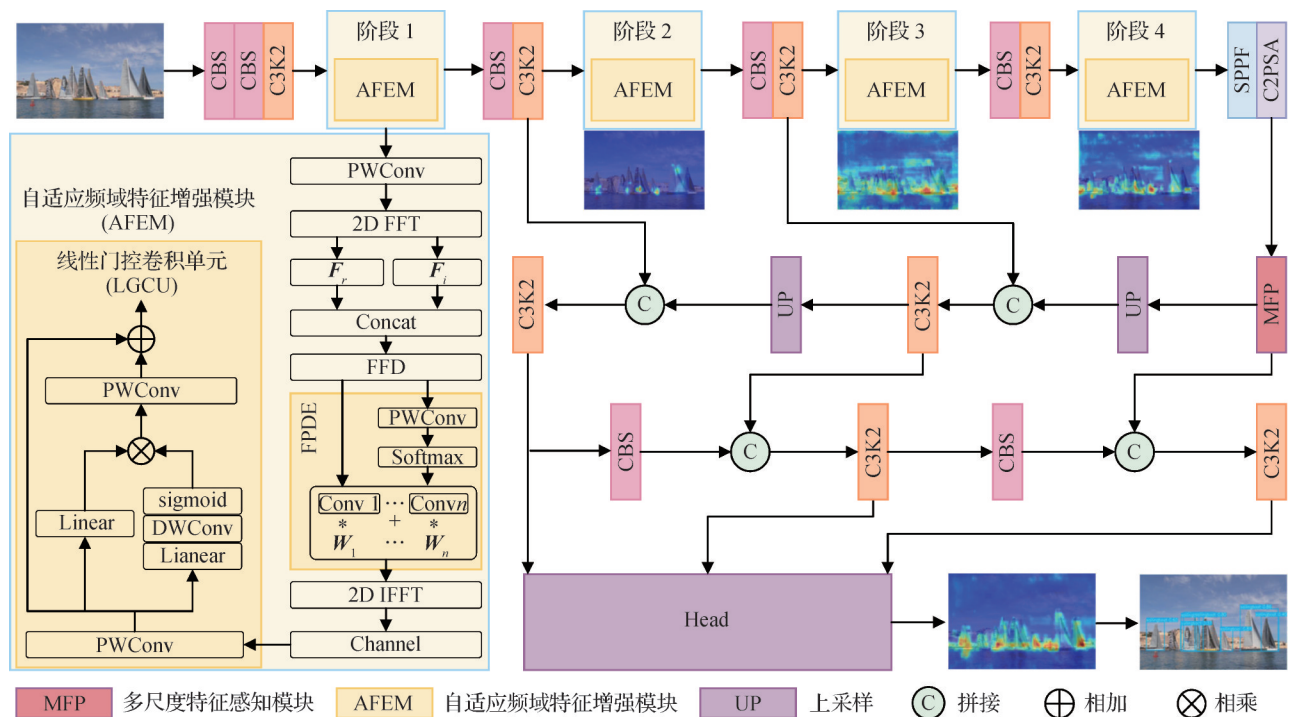


图2 跨尺度自适应频域特征增强网络整体架构

Fig. 2 The overall architecture of cross-scale adaptive frequency-domain feature enhancement network

首先,输入特征经过点卷积压缩通道维度后,输入至二维快速傅里叶变换(two-dimensional fast Fourier transform, 2D FFT)中,获得实部 $F_r$ 和虚部 $F_i$ ,二者拼接后形成频域复特征表示。为了充分捕获频域中的全局语义信息,本文引入频域特征解码(FFD)机制,通过组合式通道映射与特征重建方式,将频域特征解码为更具判别力的响应图谱,频域特征解码可以表示为

$$\mathbf{F}_{\text{FFD}} = \text{PWConv}(\text{Concat}(\mathbf{F}_r, \mathbf{F}_i)) \quad (1)$$

式中,  $\text{Concat}(\cdot)$ 表示沿通道维度连接实部与虚部,  $\text{PWConv}(\cdot)$ 为逐点卷积,用于对通道间频率特征进行融合。然后,进一步输入至频域感知增强模块(FPDE)中,通过 softmax 权重控制多个可学习卷积核的组合贡献。具体而言,先基于输入特征全局平均池化生成压缩向量 $\mathbf{z}$ ,接着经映射后得到每个频域卷积核的权重分布,融合过程可以表示为

$$\alpha_i = \frac{\exp(\mathbf{w}_i^T \mathbf{z})}{\sum_{j=1}^n \exp(\mathbf{w}_j^T \mathbf{z})} \quad (2)$$

$$\mathbf{F}_{\text{FPDE}} = \sum_{i=1}^n \alpha_i \cdot (\mathbf{K}_i * \mathbf{F}_{\text{FFD}}) \quad (3)$$

式中,  $\mathbf{w}_i$ 是与卷积核 $\mathbf{K}_i$ 对应的权重向量,\*表示频域卷积操作。该机制允许模型自适应选择最合适的频率卷积响应结构,聚焦于关键频段的特征增强。随后,进行逆傅里叶变换(two-dimensional inverse fast Fourier transform, 2D IFFT)回到空间域,并通过线性门控卷积单元(LGCU)进行非线性增强。LGCU由两条路径组成:主路径为标准的点卷积分支,用于保留主干特征;门控路径中,频域特征先经过线性映射、深度卷积和激活函数生成通道间注意力响应,作为门控信号与主分支结果逐元素相乘,实现自适应增强,其增强过程可表示为

$$\mathbf{F}_{\text{LGCU}} = \text{Conv}(\mathbf{X}) \cdot \sigma(\text{DWConv}(\text{Linear}(\mathbf{X}))) \quad (4)$$

式中,  $DWConv(\cdot)$  是深度卷积, 该结构可灵活调节不同频率通道的响应强度, 从而实现全局结构建模与局部特征选择性的统一。最终, 为进一步提升整体特征的一致性与稳定性, 动态引导融合的输出  $F_{FPDE}$  通过残差链接的方式与门控增强特征  $F_{LCCU}$  进行融合, 增强上下文表示, 该过程可表示为

$$F_{out} = F_{LCCU} + IFFT2D(F_{FPDE}) \quad (5)$$

该设计使模型能够在频域层面对船舶目标的特征进行全局依赖建模与局部细节增强, 显著改善雾天等低能见度海面场景中检测的准确性与鲁棒性。

### 2.3 多尺度特征感知模块

为了更有效地提升模型在海上复杂背景下对不同尺度船舶目标的识别能力, 本文提出了一种多尺度特征感知模块 (multi-scale feature perception module, MFP)。该模块旨在融合多尺度语义信息、强化关键特征区域响应, 同时保持模型结构的高效性与灵活性, 整体结构如图3所示, 主要包含以下4个步骤: 多分支卷积特征提取、特征拼接融合、通道注意力加权、残差连接输出。

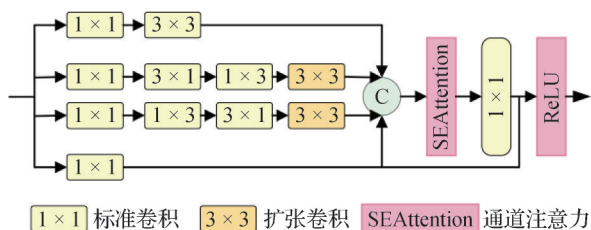


图3 多尺度特征感知模块流程图

Fig. 3 Flowchart of the multi-scale feature perception module

首先, 输入特征图  $X \in \mathbf{R}^{C_m \times H \times W}$  经过4个结构不同的并行卷积分支进行特征提取。每个分支均以点卷积进行通道压缩, 以降低后续计算开销, 并分别采用标准卷积、方向卷积、空洞卷积等策略构建多尺度感受野。这种设计可同时捕获局部细节信息与全局上下文依赖, 增强对目标结构与边缘的感知能力。各分支的输出在通道维度上进行拼接, 形成融合后的多尺度特征图, 计算为

$$F_{cat} = Concat(f_1(X), f_2(X), f_3(X), f_4(X)) \quad (6)$$

式中,  $f_i(X)$  表示第  $i$  个分支提取的特征映射,  $Concat(\cdot)$  表示沿通道维度拼接操作。通过该方式, 可显著提高特征图的多样性和鲁棒性, 适应海上场景中目标尺度多变与密集分布的特点。

然而, 简单的特征拼接可能引入冗余信息或噪

声通道, 不利于网络学习关键目标区域。为此, 本文引入基于SE(squeeze-and-excitation)的注意力机制, 对拼接后的通道进行加权重标定。首先通过全局平均池化对通道进行压缩, 然后通过两层全连接网络提取通道间依赖关系, 最后生成通道注意力权重并应用于原始特征图, 其计算过程可表示为

$$F_{att} = F_{cat} \cdot \sigma(W_2 \cdot \delta(W_1 \cdot AvgPool(F_{cat}))) \quad (7)$$

式中,  $AvgPool(\cdot)$  表示全局平均池化操作,  $W_1$  和  $W_2$  分别为注意力模块中的两层全连接权重矩阵,  $\delta(\cdot)$  和  $\sigma(\cdot)$  分别为ReLU(rectified linear unit)与sigmoid激活函数。通过该模块, 网络能够自适应地增强与目标检测任务高度相关的特征通道, 同时抑制冗余背景干扰, 提高特征表示的判别性。在获得加权特征图  $F_{att}$  后, 本文采用  $1 \times 1$  卷积进一步压缩通道数以匹配输出维度, 并与输入特征  $X$  进行残差融合, 从而保留输入信息并促进梯度反向传播。若输入输出通道数不一致, 则通过一个  $1 \times 1$  卷积层进行维度匹配, 否则直接采用恒等映射。最终输出结果表示为

$$F_{out} = ReLU(Conv(F_{att}) + R) \quad (8)$$

式中,  $Conv(\cdot)$  使用  $1 \times 1$  卷积核,  $R$  表示输入特征的残差映射。该设计不仅增强了特征的非线性表达能力, 还有效缓解了深层网络中的梯度消失问题。综上, MFP通过多分支结构感知不同尺度上下文信息, 结合通道注意力机制强化语义特征表达, 并借助残差连接保持信息流通畅, 为后续定位和回归提供判别性特征输入。该模块不仅在结构上轻量高效, 同时在实验中也展现出对复杂海洋环境下不同尺度船舶目标检测的显著性能提升。

## 3 实验结果与分析

本节介绍了实验所使用的数据集、实验细节以及评估指标, 为了全面评估本文方法的有效性, 在MVDD (marine vessel detection dataset) 数据集和RTTS (real-world task-driven testing set) 数据集上进行了实验验证, 与当前的SOTA (state-of-the-art) 方法分别从数据分析和视觉效果对比两方面展现本文方法的实际检测性能, 并通过消融实验验证各模块的有效性。

### 3.1 数据集

1) MVDD数据集。专门用于海洋船舶检测的公

开数据集,由 35 474 幅图像组成。这些图像经过精确标注,涵盖了 13 种船舶类别,包括货船、客船、游轮、散货船、油轮、帆船、拖船、渔船、钻井船、消防船、集装箱船、军舰和潜艇。

2) RTTS 数据集。这是一个公开的真实世界的雾天数据集,涵盖不同雾浓度场景,包含大巴车、自行车、汽车、行人、摩托车等 5 类检测对象,共 4 322 幅图像。本文实验所用数据集的具体数据分布如表 1 所示。

表 1 MVDD 和 RTTS 数据集的数据分布  
Table 1 Data distribution of MVDD and RTTS datasets

数据集	类别	数量/幅
MVDD	Cargo	7 640
	Passenger	5 685
	Cruise	5 307
	Bulker	5 270
	Tanker	4 589
	Sailingboat	3 830
	Tug	2 733
	Fishing	1 427
	Drill	1 336
	Firefighting	1 278
	Containership	1 074
	Warship	420
	Submarine	250
RTTS	Bicycle	534
	Bus	1 838
	Motorbike	862
	Person	7 950
	Car	18 415

### 3.2 实验细节

本文方法按照输入图像大小为  $640 \times 640$  像素进行训练,处理图像的批量大小设置为 16,学习率设置为 0.01,权重衰减因子为 0.000 5,epoch 设置 300,测试图像大小与训练输入图像大小相同。实验是在 13th Gen Intel (R) Core (TM) i5-13600KF、NVIDIA GeForce RTX 4060Ti 上进行,内存大小 32 G,采用的是 PyTorch-GPU 深度学习架构。所有方法遵循同样的实验细节,训练时两个数据集均按照目标检测常用的划分比例 9:1 分为训练集和测试

集,并以相同比例从训练集中划分出验证集。

### 3.3 评估指标

本文实验中采用的评估指标包括精确率 (precision)、召回率 (recall)、平均 AP 值 (mAP)、参数量 (parameters)、浮点计算次数 (floating point operations, FLOPs) 以及每秒帧数 (frames per second, FPS)。其中,precision 与 recall 分别用于衡量模型对目标检测结果的准确性与完整性,mAP 用于综合评估整体检测性能;parameters 与 FLOPs 反映模型的复杂度与计算开销,而 FPS 则用于衡量模型的推理速度与实时处理能力。这些评估指标分别从精度、鲁棒性、实时性 3 个方面综合评估所提方法在海上及其他复杂场景中的实际检测性能。

### 3.4 对比实验

1) 为了评估本文方法在真实海上环境检测性能,在 MVDD 数据集与 16 种不同类型的 SOTA 方法进行对比。包括双阶段检测方法 Fast R-CNN (Girshick, 2015) 和 Faster R-CNN (Ren 等, 2017); 基于 Transformer 的检测方法 DETR (Carion 等, 2020) 和 D-FINE (redefine regression task in DETRs as fine-grained distribution refinement) (Peng 等, 2024); 单阶段检测方法 YOLOv3 (Redmon 和 Farhadi, 2018)、YOLOv4 (Bochkovskiy 等, 2020)、YOLOv5 (Jocher, 2020)、YOLOv7 (Wang 等, 2023)、YOLOv8s (Jocher 等, 2023)、YOLOv10s (Wang 等, 2024b)、YOLO11s (Jocher 和 Qiu, 2024)、YOLOv12s (Tian 等, 2025)、AodeMar (attention-aware occlusion detection of vessels for maritime autonomous surface ships) (Wang 等, 2024a)、AMSP-UOD (when vortex convolution and stochastic perturbation meet underwater object detection) (Zhou 等, 2024)、IA-YOLO (image-adaptive YOLO for object detection in adverse weather conditions) (Liu 等, 2022) 和 SSDA-YOLO (semi-supervised domain adaptive YOLO for cross-domain object detection) (Zhou 等, 2023b)。

本文方法与不同类型 SOTA 方法的对比结果如表 2 所示。可以看出,所提方法在检测精度和模型复杂度之间取得了较好的平衡,以 6.29 M 的参数量在 MVDD 数据集上实现了 95.18% 的平均精确度 (mAP50)。与单阶段的 YOLO 系列经典算法 YOLOv12s 相比,所提方法的参数量比其少 2.86 M,平均精确度更高,同时,推理速度也比其高 21 帧/s。

表2 不同方法在MVDD数据集上的结果

Table 2 Results of different methods on the MVDD dataset

方法	类型	mAP50/%	mAP50-90 /%	precision /%	recall /%	parameters /M	FLOPs /G	FPS /(帧/s)
Fast R-CNN(Girshick,2015)	双阶段	80.36	51.60	85.51	81.21	33.04	134.38	21
Faster R-CNN(Ren等,2017)	双阶段	88.47	67.57	93.15	81.28	206.68	41.14	39
DETR(Carion等,2020)	Transformer	90.21	77.92	81.38	89.87	36.76	114.25	34
D-FINE(Peng等,2024)	Transformer	94.62	82.96	93.47	90.15	10.36	25.07	161
YOLOv3(Redmon和Farhadi,2018)	单阶段	89.20	66.37	93.50	84.63	61.95	66.17	26
YOLOv4(Bochkovskiy等,2020)	单阶段	92.68	66.93	94.86	83.37	64.36	60.53	21
YOLOv5(Jocher,2020)	单阶段	94.08	79.25	94.61	88.89	7.06	16.71	128
YOLOv7(Wang等,2023)	单阶段	94.05	81.50	92.52	90.70	37.62	106.47	53
YOLOv8s(Jocher等,2023)	单阶段	93.54	81.52	93.35	89.97	11.17	28.82	116
YOLOv10s(Wang等,2024b)	单阶段	94.51	82.89	93.21	89.00	7.23	21.69	192
YOLO11s(Jocher和Qiu,2024)	单阶段	94.92	83.21	94.36	89.20	9.43	21.54	172
YOLOv12s(Tian等,2025)	单阶段	94.65	82.62	94.67	89.58	9.15	19.47	206
IA-YOLO(Liu等,2022)	单阶段	92.15	80.32	78.65	85.12	58.93	65.61	28
SSDA-YOLO(Zhou等,2023b)	单阶段	91.08	75.26	73.19	83.05	20.87	27.14	51
AodeMar(Wang等,2024a)	单阶段	<b>95.43</b>	82.57	94.02	<b>91.02</b>	8.28	67.40	99
AMSP-UOD(Zhou等,2024)	单阶段	93.62	83.15	92.70	90.53	10.37	23.82	20
本文	单阶段	95.18	<b>83.38</b>	<b>94.89</b>	90.74	<b>6.29</b>	<b>13.16</b>	<b>227</b>

注:加粗字体表示各列最优结果。

与Transformer类型的检测方法相比,本文方法比D-FINE少4.07 M的参数,却实现了更高的检测精度。相较于当前流行的频域增强方法IA-YOLO和SSDA-YOLO,本文方法在精度上明显要更高,并且由于主流的频域增强方法大多采用滤波器进行图像增强,往往需要较多的参数去实现频域滤波,导致模型较大,本文方法比IA-YOLO参数量少52.64 M,推理速度高出199帧/s,在精度和速度方面都更具优势。在参数量相差不大的情况下,本文方法比D-FINE的检测精度更高,同时召回率也更高,模型收敛速度更快。

本文方法与不同SOTA方法在MVDD数据集上的视觉对比如图4所示。可以看出,对于因海面反射而造成过曝光的船舶图像和夜晚环境中的低光照船舶图像,本文方法不仅能够精准识别,而且还具有较高的置信度。在复杂背景和有船舶互相遮挡的情况下,Faster R-CNN、D-FINE和YOLO等系列的模型普遍存在错检和漏检等问题,而本文方法对于这种

遮挡情况适应较好,能够检测到被遮挡船只。在小目标的检测上,本文方法能够检测出更多的小目标船舶,检测性能显著优于对比的先进方法。总体而言,对于真实的海上情况,所提方法在检测性能方面优于对比的先进算法,对于各种复杂的海上环境具备更好的适应能力,进一步体现了本文方法的有效性和稳定性。

2)为了进一步验证所提方法的鲁棒性和泛化能力,与SOTA方法在RTTS数据集进行实验对比。RTTS数据集涵盖各种驾驶场景,阴天、雾天和霾天等,反映了各种能见度条件下的实际情况,能够对本文方法的泛化能力进行更全面评估。

不同方法在RTTS数据集上的对比结果如表3所示。可以看出,对比YOLO系列的检测模型,本文方法不仅在精度上比YOLO的s级模型高,而且本文方法的参数量也相对更小,检测速度更快。与基于Transformer的方法相比,本文方法检测精度和速度都更优。相比于经典的双阶段检测方法Fast-RCNN

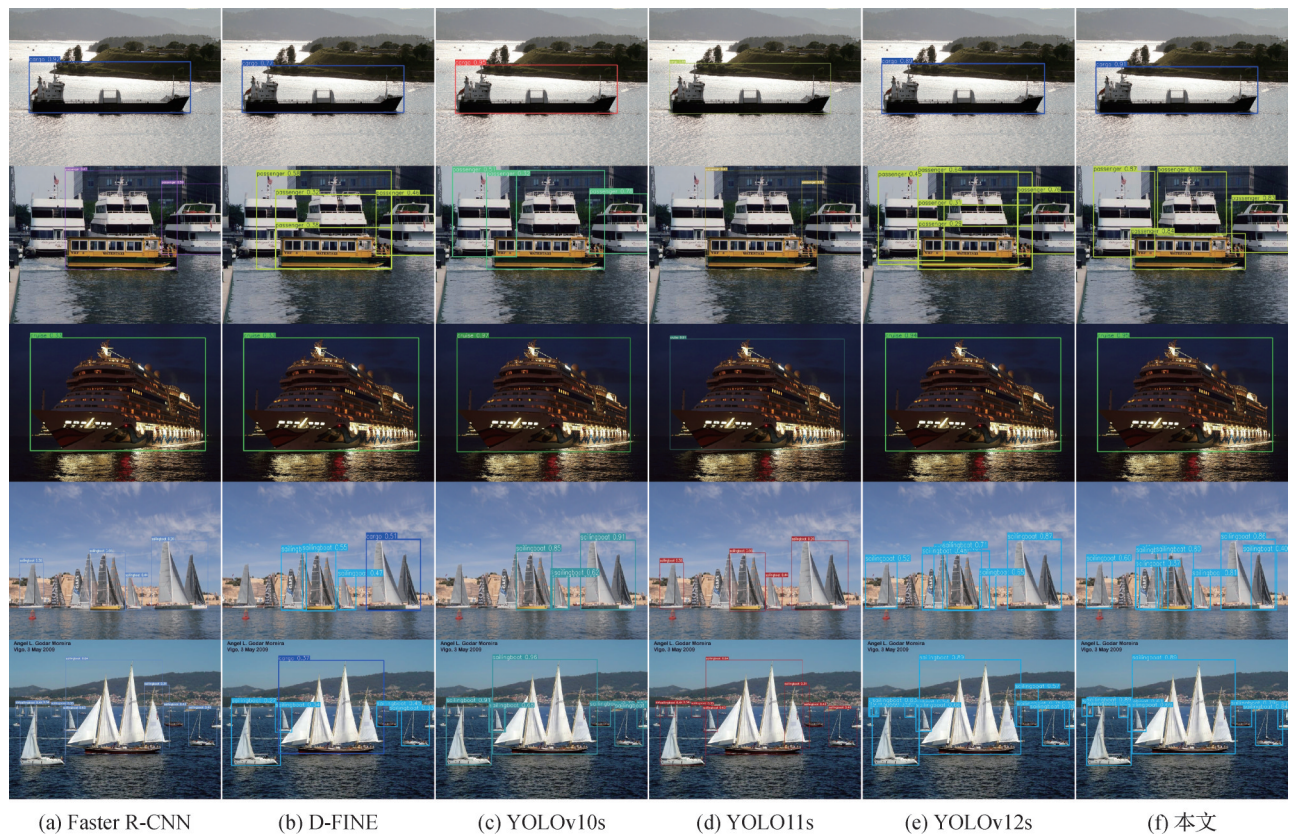


图4 不同SOTA方法与本文方法在MVDD数据集的视觉对比

Fig. 4 Visual comparison of different SOTA methods and our method on the MVDD dataset ((a) Faster R-CNN; (b) D-FINE; (c) YOLOv10s; (d) YOLO11s; (e) YOLOv12s; (f) ours)

表3 不同方法在RTTS数据集上的结果

Table 3 Results of different methods on the RTTS dataset

方法	类型	mAP50/%	mAP50-90 /%	precision /%	recall /%	parameters /M	FLOPs /G	FPS /(帧/s)
Fast R-CNN(Girshick, 2015)	双阶段	67.32	41.37	76.63	57.81	33.04	134.38	19
Faster R-CNN(Ren等, 2017)	双阶段	68.43	42.17	78.59	59.60	206.68	41.14	31
DETR(Carion等, 2020)	Transformer	73.78	48.52	77.62	65.43	36.76	114.25	21
D-FINE(Peng等, 2024)	Transformer	73.55	49.57	75.50	66.28	10.36	25.07	124
YOLOv3(Redmon和Farhadi, 2018)	单阶段	69.65	43.48	81.15	60.17	61.95	66.17	22
YOLOv4(Bochkovskiy等, 2020)	单阶段	70.87	45.92	78.62	62.59	64.36	60.53	19
YOLOv5(Jocher, 2020)	单阶段	74.00	49.37	80.29	61.30	7.06	16.71	95
YOLOv7(Wang等, 2023)	单阶段	73.49	44.19	81.32	60.85	37.62	106.47	37
YOLOv8s(Jocher等, 2023)	单阶段	74.13	49.15	80.10	65.33	11.17	28.82	105
YOLOv10s(Wang等, 2024b)	单阶段	73.78	47.52	<b>80.90</b>	60.08	7.23	21.69	181
YOLO11s(Jocher和Qiu, 2024)	单阶段	74.34	48.51	78.42	63.58	9.43	21.54	163
YOLOv12s(Tian等, 2025)	单阶段	73.95	48.30	74.58	65.72	9.15	19.47	197
IA-YOLO(Liu等, 2022)	单阶段	71.62	48.64	77.31	62.54	58.93	65.61	20
SSDA-YOLO(Zhou等, 2023b)	单阶段	70.37	45.17	75.82	63.32	20.87	27.14	37
AMSP-UOD(Zhou等, 2024)	单阶段	<b>75.21</b>	49.62	78.28	65.71	10.37	23.82	18
本文	单阶段	74.79	<b>49.93</b>	78.69	<b>66.48</b>	<b>6.29</b>	<b>13.16</b>	<b>206</b>

注:加粗字体表示各列最优结果。

和Faster-RCNN,本文方法的平均精确度更高,模型复杂度更低。虽然AMSP-UOD的检测精度比本文方法高0.42%,但是本文方法的FLOPs仅为AMSP-UOP的二分之一,推理速度更快,对于实际的应用场景,本文方法具有更大优势。

本文方法与不同SOTA方法在RTTS数据集上的视觉对比如图5所示。从图5可以看出,无论是经典的双阶段算法还是比较先进的YOLO系列算法和基于Transformer的检测算法,在浓雾中的表现都不佳,总有漏检的问题,而本文方法却能达到较好的检测结果。对于快速移动、低光照和雪天环境下的车辆,所提方法也能实现精准检测,而基线模型在面对小目标物体时,Faster R-CNN等经典的双阶段目标检测方法往往会忽略掉一些重叠的物体,从而导致漏检。D-FINE等基于Transformer的方法可能会将大巴车检测成小汽车,检测精度有待提升。总体而言,本文方法不仅能在海上达到较好的检测效果,在陆地上的恶劣环境中也有一定的优势。

### 3.5 类别级检测结果分析

为了进一步验证本文方法的有效性,详细对比了不同方法在MVDD和RTTS数据集针对小、中、大等不同尺度目标的检测精度( $AP_s$ 、 $AP_m$ 、 $AP_l$ ),结果如表4所示。可以看出,本文方法在多尺度目标检测上展现出较高的鲁棒性。具体而言,在MVDD数据集上,该方法在所有尺度上均取得了最优结果,其中, $AP_s$ 达到15.8%,优于YOLOv12s(15.3%),有效改善了小目标漏检问题。在更具挑战性的RTTS数据集上,本文方法在中目标(54.2%)和大目标(66.0%)检测上保持显著领先,同时在小目标检测上也表现出极强的竞争力(25.6%),优于大多数对比模型。这一结果充分证明了改进策略能有效增强特征提取能力,在复杂场景下实现了对不同尺度目标的高精度覆盖。

### 3.6 消融实验

为了验证本文方法各模块的有效性,在MVDD数据集上进行了消融实验。根据表5的消融实验结果可知,未引入任何模块的基准模型mAP50为



图5 不同SOTA方法与本文方法在RTTS数据集上的视觉对比

Fig. 5 Visual comparison of different SOTA methods and our method on the RTTS dataset ((a) Faster R-CNN; (b) D-FINE; (c) YOLOv10s; (d) YOLO11s; (e) YOLOv12s; (f) ours)

表4 不同方法在 MVDD 与 RTTS 数据集上的  
小、中、大目标检测精度对比  
Table 4 Comparison of detection accuracy for small,  
medium, and large targets by different methods on  
MVDD and RTTS datasets

方法	MVDD数据集			RTTS数据集		
	AP <sub>s</sub>	AP <sub>m</sub>	AP <sub>l</sub>	AP <sub>s</sub>	AP <sub>m</sub>	AP <sub>l</sub>
Fast R-CNN	5.2	12.5	55.4	16.5	40.1	51.2
Faster R-CNN	9.8	18.6	70.1	18.2	43.5	54.8
DETR	11.5	22.1	79.5	21.5	50.1	61.2
D-FINE	14.8	27.5	84.8	24.8	53.2	65.5
YOLOv3	8.5	16.4	68.5	17.5	45.2	56.8
YOLOv4	9.2	17.8	70.2	18.9	47.1	59.4
YOLOv5	12.5	24.5	81.2	23.5	51.8	63.5
YOLOv7	13.1	25.8	83.5	22.1	46.5	58.9
YOLOv8s	14.2	26.9	83.8	24.6	53.1	64.8
YOLOv10s	14.9	27.5	84.5	23.8	51.5	63.2
YOLO11s	15.1	27.8	84.9	24.5	52.6	64.3
YOLOv12s	15.3	27.9	85.0	24.4	52.4	64.1
IA-YOLO	13.5	25.2	82.1	24.1	52.0	63.8
SSDA-YOLO	10.8	21.5	77.4	20.5	47.8	59.5
AMSP-UOD	15.0	28.1	85.2	<b>26.2</b>	53.5	65.2
本文	<b>15.8</b>	<b>28.6</b>	<b>85.8</b>	25.6	<b>54.2</b>	<b>66.0</b>

注:加粗字体表示各列最优结果。

表5 对模型中不同组件的消融实验

Table 5 Abandonment experiments for different  
components in the model

AFEM	MFP	mAP50 /%	recall /%	params /M	FLOPs /G	FPS /(帧/s)
×	×	93.35	89.21	<b>2.65</b>	<b>6.34</b>	<b>308</b>
√	×	94.28	89.78	4.28	9.42	269
×	√	93.89	89.57	5.17	10.19	252
√	√	<b>95.18</b>	<b>90.74</b>	6.29	13.16	227

注:加粗字体表示各列最优结果。“√”和“×”分别表示使用和未使用对应模块。

93.35%,精度相对较低。在此基础上仅加入自适应频域特征增强模块(AFEM)后,mAP50提升至94.28%,表明该模块能够有效增强频域特征表达能力。仅加入多尺度特征感知模块(MFP)时,mAP50

达到93.89%,说明该模块能够加强跨尺度特征融合能力,提高小目标检测性能,但同样带来一定的计算开销。当同时引入AFEM与MFP时,mAP50达到95.18%,召回率提升至90.74%,均为最高值,验证了两种模块在特征增强与多尺度感知方面的互补性,能够显著提高复杂海上场景船舶目标的检测精度,且不带来参数量的过多增加。

## 4 结论

本文针对复杂海上环境中船舶目标特征模糊、尺度变化大及背景干扰强等问题,提出了一种跨尺度自适应频域增强的船舶检测方法。该方法在基准检测框架中引入自适应频域特征增强模块(AFEM),有效提升了频域特征的表达力;同时设计了多尺度特征感知模块(MFP),增强了跨尺度特征融合与小目标感知能力。通过在真实海上船舶数据集(MVDD)上的实验验证,所提方法在mAP50、mAP50-90及召回率等指标上均优于基准模型和SOTA方法,并在多组消融实验中进一步证明了两个模块的独立贡献与协同效应。尽管引入新模块在一定程度上增加了参数量与计算量,但整体性能提升显著,具有较高的应用价值。未来的工作将进一步优化模型结构与回归策略,以在保持高精度的同时降低计算开销,并探索其在无人船舶导航、海上交通监控等实际场景中的应用潜力。

## 参考文献(References)

- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2025-11-01]. <https://arxiv.org/pdf/2004.10934.pdf>
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8\_13]
- Dai X Y, Chen Y P, Yang J W, Zhang P C, Yuan L and Zhang L. 2021. Dynamic DETR: end-to-end object detection with dynamic attention//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2968-2977 [DOI: 10.1109/ICCV48922.2021.00298]
- Gao G, Wang Y J, Chen Y H, Yang G, Yao L B, Zhang X, et al. 2025. An oriented ship detection method of remote sensing image with

- contextual global attention mechanism and lightweight task-specific context decoupling. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #4200918 [DOI: 10.1109/TGRS.2024.3520658]
- Girshick R. 2015. Fast R-CNN//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Gong Y, Zhang X Y, Lu J L, Jiang X M, Wang Z C, Liu H, et al. 2025. Steering angle-guided multimodal fusion lane detection for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 26 (2) : 1470-1481 [DOI: 10.1109/TITS.2024.3507536]
- Huang Z X, Jiang X N, Wu F L, Fu Y, Zhang Y, Fu T J, et al. 2023. An improved method for ship target detection based on YOLOv4. *Applied Sciences*, 13(3) : #1302 [DOI: 10.3390/app13031302]
- Jia H C, Pu X Y, Liu Q Y, Wang H P and Xu F. 2024. A fast progressive ship detection method for very large full-scene SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5206615 [DOI: 10.1109/TGRS.2024.3369637]
- Jocher G. 2020. Ultralytics YOLOv5 [EB/OL]. [2025-11-01]. <https://github.com/ultralytics/yolov5>
- Jocher G, Chaurasia A and Qiu J. 2023. Ultralytics YOLOv8 [EB/OL]. [2025-11-01]. <https://github.com/ultralytics/ultralytics>
- Jocher G and Qiu J. 2024. Ultralytics YOLO11 [EB/OL]. [2025-11-01]. <https://github.com/ultralytics/ultralytics>
- Li C, Hei Y Q, Xi L H, Li W T and Xiao Z. 2024. GL-DETR: global-to-local transformers for small ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 21: #4016805 [DOI: 10.1109/LGRS.2024.3461212]
- Li K X, Guo J, Li R C, Wang Y J, Li Z M and Liao K. 2024. Ship trajectory anomaly detection method based on encoder-decoder architecture composed of Transformer\_LSTM modules. *Chinese Journal of Ship Research*, 19(2) : 223-232 (李可欣, 郭健, 李冉冲, 王宇君, 李宗明, 缪坤. 2024. 基于Transformer\_LSTM编解码器模型的船舶轨迹异常检测方法. *中国舰船研究*, 19(2): 223-232) [DOI: 10.19693/j.issn.1673-3185.03291]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2020. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42 (2) : 318-327 [DOI: 10.1109/TPAMI.2018.2858826]
- Lin Z, Ji K F, Leng X G and Kuang G Y. 2019. Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 16 (5) : 751-755 [DOI: 10.1109/LGRS.2018.2882551]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. 2016. SSD: single shot MultiBox detector//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0\_2]
- Liu W Y, Ren G F, Yu R S, Guo S, Zhu J K and Zhang L. 2022. Image-adaptive YOLO for object detection in adverse weather conditions//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*. Virtually: AAAI: 1792-1800 [DOI: 10.1609/aaai.v36i2.20072]
- Ma F, Shi Z H, Sun J, Chen C, Mao X B and Yan X P. 2024. Lightweight and robust ship detection method driven by self-attention mechanism. *Chinese Journal of Ship Research*, 19(5) : 188-199 (马枫, 石子慧, 孙杰, 陈晨, 毛显斌, 严新平. 2024. 自注意力机制驱动的轻量化高鲁棒船舶目标检测方法. *中国舰船研究*, 19(5): 188-199) [DOI: 10.19693/j.issn.1673-3185.03389]
- Peng Y S, Li H B, Wu P X, Zhang Y Y, Sun X Y and Wu F. 2024. D-FINE: redefine regression task in DETRs as fine-grained distribution refinement [EB/OL]. [2025-11-01]. <https://arxiv.org/pdf/2410.13842.pdf>
- Phan D T, Doan V H M, Choi J, Lee B and Oh J. 2025. AADC-Net: a multimodal deep learning framework for automatic anomaly detection in real-time surveillance. *IEEE Transactions on Instrumentation and Measurement*, 74: #5025713 [DOI: 10.1109/TIM.2025.3551832]
- Qin C, Zhang L P, Wang X Q, Li G, He Y and Liu Y H. 2025. RDB-DINO: an improved end-to-end transformer with refined de-noising and boxes for small-scale ship detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #5200517 [DOI: 10.1109/TGRS.2024.3515150]
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2025-11-01]. <https://arxiv.org/pdf/1804.02767.pdf>
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6) : 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Si J H, Song B B, Wu J X, Lin W, Huang W and Chen S Y. 2023. Maritime ship detection method for satellite images based on multi-scale feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 6642-6655 [DOI: 10.1109/JSTARS.2023.3296898]
- Tian Y J, Ye Q X and Doermann D. 2025. YOLOv12: attention-centric real-time object detectors [EB/OL]. [2025-11-01]. <https://arxiv.org/pdf/2502.12524.pdf>
- Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]
- Wang N, Wang Y Y, Feng Y and Wei Y. 2024a. AodeMar: attention-aware occlusion detection of vessels for maritime autonomous surface ships. *IEEE Transactions on Intelligent Transportation Systems*, 25(10) : 13584-13597 [DOI: 10.1109/TITS.2024.3398733]
- Wang A, Chen H, Liu L H, Chen K, Lin Z J, Han J G, et al. 2024b. YOLOv10: real-time end-to-end object detection [EB/OL]. [2025-11-01]. <https://arxiv.org/pdf/2405.14458.pdf>
- Wang Z H, Zhang X Y, Li J and Luan K F. 2021. A YOLO-based target

- detection model for offshore unmanned aerial vehicle data. Sustainability, 13(23): #12980 [DOI: 10.3390/su132312980]
- Wen G Q, Cao P, Wang H N, Chen H L, Liu X L, Xu J H, et al. 2023. MS-SSD: multi-scale single shot detector for ship detection in remote sensing images. Applied Intelligence, 53(2): 1586-1604 [DOI: 10.1007/s10489-022-03549-6]
- Xiao Z J, Lin B H and Qu H C. 2024. SAR ship detection with multi-mechanism fusion. Journal of Image and Graphics, 29(2): 545-558 (肖振久, 林渤翰, 曲海成. 2024. 融合多重机制的SAR舰船检测. 中国图象图形学报, 29(2): 545-558) [DOI: 10.11834/jig.230166]
- Xing Z J, Ren J, Fan X Z and Zhang Y. 2023. S-DETR: a transformer model for real-time detection of marine ships. Journal of Marine Science and Engineering, 11(4): #696 [DOI: 10.3390/jmse11040696]
- Yang Z Y and Shi L. 2024. Boat object detection method based on improved Cascade R-CNN algorithm. Ship Science and Technology, 46(6): 144-149 (杨镇宇, 石刘. 2024. 基于改进Cascade R-CNN算法的船舶目标检测方法. 舰船科学技术, 46(6): 144-149) [DOI: 10.3404/j.issn.1672-7649.2024.06.025]
- Yu M, Han S J, Wang T F and Wang H Y. 2022. An approach to accurate ship image recognition in a complex maritime transportation environment. Journal of Marine Science and Engineering, 10(12): #1903 [DOI: 10.3390/jmse10121903]
- Yuan S, Qin H L, Yan X, Akhtar N and Mian A. 2024. SCTransNet: spatial-channel cross transformer network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 62: #5002615 [DOI: 10.1109/TGRS.2024.3383649]
- Zhang L P, Liu Y, Zhao W D, Wang X Q, Li G and He Y. 2023. Frequency-adaptive learning for SAR ship detection in clutter scenes. IEEE Transactions on Geoscience and Remote Sensing, 61: #5215514 [DOI: 10.1109/TGRS.2023.3249349]
- Zhang W D, Li Z X, Li G H, Zhuang P X, Hou G J, Zhang Q, et al. 2024. GACNet: generate adversarial-driven cross-aware network for hyperspectral wheat variety identification. IEEE Transactions on Geoscience and Remote Sensing, 62: #5503314 [DOI: 10.1109/TGRS.2023.3347745]
- Zhang B Y, Zhang C, Shi Z N and Liu S T. 2024b. Lightweight ship detection method based on YOLO-FNC model. Chinese Journal of Ship Research, 19(5): 180-187 (张炳焱, 张闯, 石振男, 刘松涛. 2024. 基于YOLO-FNC模型的轻量化船舶检测方法, 中国舰船研究, 19(5): 180-187) [DOI: 10.19693/j.issn.1673-3185.03487]
- Zhang Y N and Er M J. 2024. Asymmetric aggregation network for accurate ship detection in optical imagery. IEEE Transactions on Geoscience and Remote Sensing, 62: #4211914 [DOI: 10.1109/TGRS.2024.3481370]
- Zhang W D, Wang Z, Hao Z B and Gao X Y. 2025. Marine species image enhancement method based on box-cox transform. Journal of Naval Aviation University, 40(4): 600-606, 640 (张卫东, 王中, 郝志斌, 高兴运. 2025. 基于Box-Cox变换的海洋物种图像增强方法. 海军航空大学学报, 40(4): 600-606, 640) [DOI: 10.7682/j.issn.2097-1427.2025.04.010]
- Zhao L, Li B, Jiang J X and Wei X X. 2025. Classification committee for active deep object detection. IEEE Transactions on Multimedia, 27: 1277-1288 [DOI: 10.1109/TMM.2024.3521778]
- Zhou X F, Shen K Y, Weng L, Cong R M, Zheng B L, Zhang J Y, et al. 2023a. Edge-guided recurrent positioning network for salient object detection in optical remote sensing images. IEEE Transactions on Cybernetics, 53(1): 539-552 [DOI: 10.1109/TCYB.2022.3163152]
- Zhou H Y, Jiang F and Lu H T. 2023b. SSDA-YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection. Computer Vision and Image Understanding, 229: #103649 [DOI: 10.1016/j.cviu.2023.103649].
- Zhou J C, He Z X, Lam K M, Wang Y D, Zhang W S, Guo C L, et al. 2024. AMSP-UOD: when vortex convolution and stochastic perturbation meet underwater object detection//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 7659-7667 [DOI: 10.1609/aaai.v38i7.28599]

## 作者简介

王应军,男,副教授,硕士生导师,主要研究方向为计算机视觉。E-mail:wangyingjun@hist.edu.cn

张卫东,通信作者,男,副教授,硕士生导师,主要研究方向为图像增强和深度学习。E-mail:zwd\_wd@163.com

杨晓鹏,男,硕士研究生,主要研究方向为计算机视觉与深度学习。E-mail:pxy\_0123@163.com

周玲,女,讲师,主要研究方向为视觉图像处理。

E-mail: 1026160568@qq.com

路皓翔,男,讲师,硕士生导师,主要研究方向为深度学习和图像处理。E-mail:hclu1005@163.com

赵文义,男,讲师,硕士生导师,主要研究方向为人工智能和计算机视觉。E-mail:zwy@bupt.edu.cn