

基于多尺度特征融合的在轨目标检测算法^{*}

霍东升 王 栋 刘云贺

(长光卫星技术股份有限公司 长春 130102)

摘 要: 为提高复杂场景下遥感图像在轨目标检测的准确率,提出了一种基于多尺度特征融合的检测方法。首先,设计多尺度特征融合模块,扩展模型感受野,采用并行化多通道融合处理的方法获取目标多维特征;其次,结合注意力机制构建空间通道特征交互模块,增强模型提取显著区域细节特征的能力;最后,引入归一化 Wasserstein 距离改进损失函数,优化重叠锚框间位置偏移的度量方法,提高模型对密集重叠目标的分辨能力。此外,通过数据增强等处理,在“吉林一号”卫星影像上标注了一套船舶数据集 JS25k。模型最终检测精度为 98.2%,比最新的 YOLOv13 高 0.6%,性能优于 RT-DETR 系列模型,在边缘端嵌入式部署后检测速率为 92.3 fps。实验结果表明,模型能够有效提高复杂场景下的检测精度,同时满足在轨时效性需求。

关键词: 遥感图像;在轨目标检测;多尺度特征融合;注意力机制;损失函数;数据增强

中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2060

On-orbit object detection algorithm based on multi-scale feature fusion

Huo Dongsheng Wang Dong Liu Yunhe

(Chang Guang Satellite Technology Co., Ltd., Changchun 130102, China)

Abstract: To improve the accuracy of on-orbit object detection in remote-sensing images under complex scenarios, this paper proposes a detection method based on multi-scale feature fusion. First, a multi-scale feature fusion module is designed to expand the model's receptive field, where parallel multi-channel fusion is employed to extract multidimensional target features. Second, a spatial-channel feature interaction module is constructed by integrating an attention mechanism to enhance the model's ability to capture detailed features in salient regions. Third, the normalized Wasserstein distance is introduced to improve the loss function, optimizing the measurement of positional offsets between overlapping anchor boxes and enhancing discrimination of densely overlapped targets. In addition, a ship dataset named JS25k is annotated on "Jilin-1" satellite imagery with data augmentation and other preprocessing steps. The proposed model achieves a final detection accuracy of 98.2%, outperforming the latest YOLOv13 by 0.6% and exhibiting better performance than the RT-DETR series. When deployed on an edge-side embedded platform, the detection speed reaches 92.3 fps. Experimental results demonstrate that the model effectively improves detection accuracy in complex scenes while meeting real-time requirements for on-orbit applications.

Keywords: remote-sensing images; on-orbit object detection; multi-scale feature fusion; attention mechanism; loss function; data augmentation

0 引 言

随着遥感技术的进步和卫星数量的增加,遥感影像的分辨率不断提高,数据规模也越来越庞大^[1]。传统目标检测需要经过星上拍摄、压缩、数传、地面解密、解压、编目等处理,流程复杂耗时,难以满足灾害应急和军事应用等高效业务。在轨目标检测通过在卫星载荷上部署嵌入式模

型,能够直接处理原始影像,同时上传关键信息和目标切片数据,快速完成飞机、船舶识别解译。但在实际应用中,仍存在复杂背景、辐射畸变等造成的误检,边缘算力资源不足等问题^[2],因此需要控制推理开销,在模型规模受限的情况下,优化网络来降低误检率。

早期的传统方法主要依靠专家知识提取图像浅层特征,输入到分类器实现目标检测,对专业知识要求高,设计

难度大,算法鲁棒性差。当前主流的方法基于深度学习开展,主要可以概括为两类^[3]:1)双阶段算法:代表性的 R-CNN 系列模型,主要利用区域候选网络确定目标位置,检测准确率较高,但参数量较大,推理速度慢,多应用于地面设备^[4];2)单阶段算法:端到端的进行特征提取和识别,符合在轨目标检测轻量化和时效性需求^[5],目前主流的检测框架为基于 CNN 架构的 YOLO 系列模型和基于 Transformer 架构的 RT-DETR 系列模型^[6]。因为工程部署的优势,大多数边缘设备普遍采用 YOLO 及其改进模型^[7]。早期 Li 等^[8]基于 YOLOv3 和通道注意力机制构建 SeMo-YOLO,提高多尺度遥感目标检测精度;Yin 等^[9]通过加入多尺度融合模块,并引入旋转框检测模块,滤除复杂背景干扰,提高模型性能;Hu 等^[10]通过数据增强,优化主干网络,设计了 DA-YOLO 模型,应用扩展可分离卷积来提高模型对小尺度目标轮廓细节特征的提取能力;王迎龙等^[11]通过引入 Biformer 注意力机制精准捕捉图像远程依赖,提高模型检测精度;李泽胤等^[12]在 YOLOv8n 基础上结合扩散机制增强特征交互,并设计轻量化的任务动态调整检测头,以特征共享方式提高模型定位分类性能。总体上,YOLO 系列模型针对特定工程问题进行优化的思路,主要可以归纳为改进主干网络结构提高模型特征提取能力,依靠特征金字塔等结构,对网络深层特征和浅层语义信息进行融合^[13];设计新颖的检测头筛选目标区域^[14];使用

改进的损失函数等引导模型解决特定的误检漏检问题^[15]。考虑到模型的轻量化、部署环境的完备性、业务适配的便利以及动态架构的优化,YOLOv5 和 YOLOv8 等仍然是该领域的主流基准框架^[16]。

为了降低复杂场景下遥感图像在轨船舶检测的误检率,进一步提升船舶检测的性能,本文在 YOLOv5l 模型的基础上进行改进,提出了一种基于多尺度特征融合的目标检测模型 MS-YOLO。1)设计了基于空洞卷积的多尺度特征融合模块 (multi-scale feature fusion module, MSFFM),扩展多层次感受野,增强主干网络特征提取能力;2)修改连接层结构,引入 SimAM 注意力机制^[17],构建空间通道特征交互模块 (spatial-channel feature attention module, SCFAM),在不增加模型参数量的情况下,提高模型对显著区域细节特征的表达能力;3)提出联合损失函数 (efficient wasserstein distance, EWD),通过修改归一化 Wasserstein 度量距离^[18],并优化锚框交并比误差的度量方式,引导模型区分重叠目标,改善对目标的定位精度;4)构建了船舶检测数据集 JS25k,利用旋转拼接、仿射变换等方式进行数据增强,丰富数据集场景规模。

1 MS-YOLO 架构

基于多尺度特征融合改进后的模型 MS-YOLO 整体结构如图 1 所示。

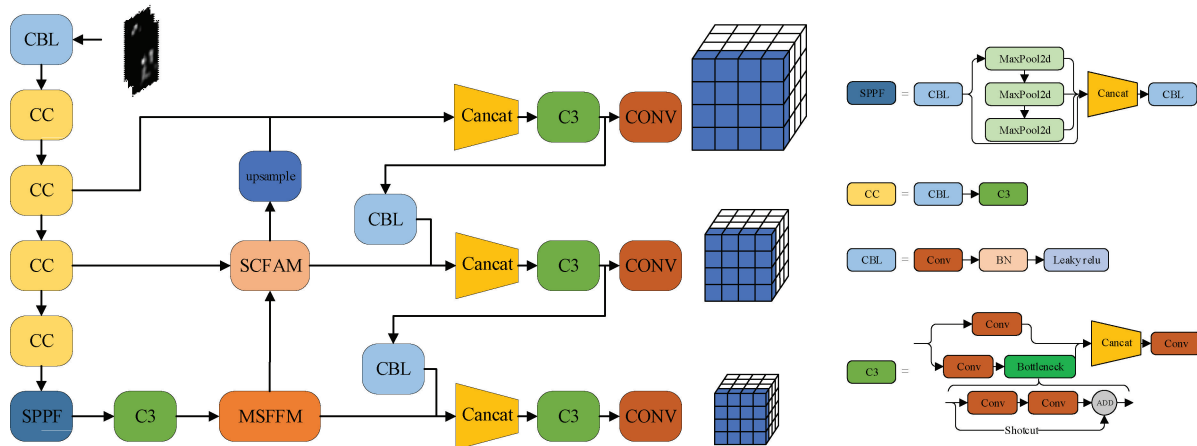


图 1 MS-YOLO 模型主体架构

Fig. 1 The overall network structure of the MS-YOLO

首先通过数据增强等预处理手段,针对背景噪声、雨雪遮挡、大视角拍摄等情况,对来自“吉林一号”星座真实标注的船舶检测数据进行了扩充,确保样本规模尽可能覆盖各类畸变工况。考虑到检测场景目标区域占图像区域的比值动态范围较大,单一的锚框无法兼顾关键区域和细节信息,因此在特征提取阶段,使用多尺度空洞卷积,扩大模型感受野,同时融合多区域、多层次特征,在特征金字塔尾端构建了 MSFFM,获取目标的多维度信息;检测过程中,由于云层遮挡、地物干扰,会造成对集装箱船、油槽船以及散货船等

目标的误检,因此在主干连接处构建特征交互模块,融合空间和通道维度特征,增强模型对显著区域的捕捉能力;在检测头位置,通过结合二维高斯分布空间下的距离度量方式,设计了损失函数 EWD,提高预测边界框对密集重叠目标的区分能力。下面对模型各部分进行详细介绍。

1.1 基于空洞卷积的 MSFFM

遥感图像的目标尺度分布广泛,覆盖像素面积差别较大,通常还受到云层、港岸码头等背景的遮挡干扰。为了克服复杂相似地物的干扰,区分目标边缘背景的差异,需

要小尺度的感受野提取目标区域纹理的细节特征;为了获取大目标的全局信息,提取区域轮廓的全局特征,又需要扩大感受野。传统的特征金字塔感受野受限,只能获取单

一的尺度特征,想要同时获得目标区域和轮廓背景的多维信息,需要扩展骨干网络的采样尺度范围。因此本文设计了一个 MSFFM,模块主体结构如图 2 所示。

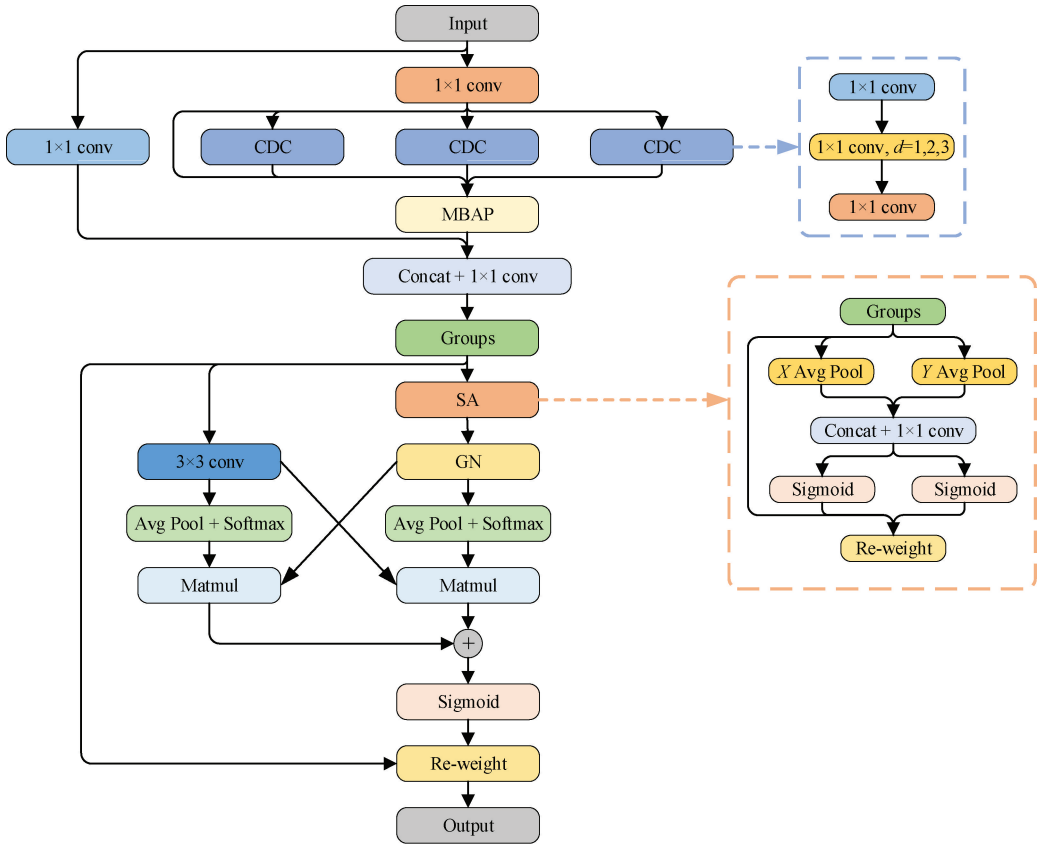


图 2 多尺度特征融合模块

Fig. 2 Multi-scale feature fusion module structure

模块的上半部分引入了一组膨胀率分别为 1~3 的不同尺度的空洞卷积来捕获不同感受野下的特征,再对多分支进行平均池化操作,合并不同尺度信息,同时使用残差连接规避梯度爆炸或梯度消失,之后连接特征送入 1×1 的卷积层,获取不同采集尺度下的融合信息。模块下半部分首先对特征进行了分组批量化拆分,通过构建并行化处理结构,提高特征提取的效率。其中支路 1 引入 SA 结构^[19],分别沿通道维度和空间维度提取特征进行拼接,将空间位置信息融入不同维度的通道,之后再做进一步的组归一化和池化处理,实现多维度的特征融合;支路 2 首先经过一个 3×3 的卷积核操作,对深度特征信息做进一步融合,之后同支路 1 进行局部跨通道交互,通过矩阵 Matmul 相乘的方式聚合多尺度特征,最后再与支路 3 残差连接,充分结合层间和区域内的上下文信息,实现多维度特征融合,整体提高模型的特征挖掘和表达能力。

1.2 基于注意力机制的 SCFAM

为了对目标与背景做进一步的区分,提高模型在目标位置的权重,需要引导模型关注图像显著区域。传统的注意力机制分为空间注意力、通道注意力和空间通道联合

3 类,通过特征压缩获取对应维度权重,之后再对相应维度进行激励赋值,以加权形式表达显著特征的重要程度,但这些方式会在不同程度上提高模型的计算复杂度,同时增加网络的参数量。因此本文在网络连接层部分嵌入了 SCFAM,其具体执行方式如图 3 所示。

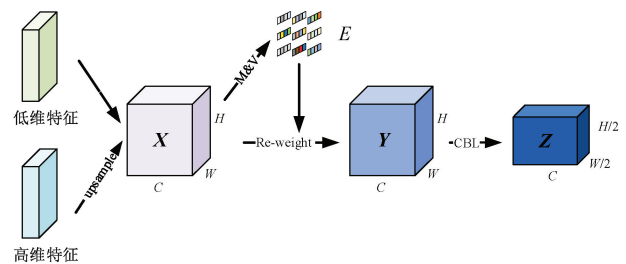


图 3 空间通道特征交互模块

Fig. 3 Spatial-channel feature attention module

模块首先对高维特征做近邻差值上采样处理,再与低维特征进行融合得到输入特征,之后分别从空间维度和通道维度为输入特征的每个神经元定义如下能量函数:

$$e_i(w_i, b_i, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \quad (1)$$

式中: t 和 x_i 分别是输入特征 $X \in \mathbf{R}^{C \times H \times W}$ 上的目标神经元和其他神经元, $\hat{t} = w_i t + b_i$ 和 $\hat{x}_i = w_i x_i + b_i$ 分别是关于 t 和 x_i 的线性变化, i 是神经元对应空间索引, $M = H \times W$ 是通道神经元数量, w_i 和 b_i 是线性变换的权重与偏置。神经元能量越低, 与周围神经元的差别越大, 神经元就越重要。用上述能量函数表示神经元权重, 最小化该函数的过程就等价于训练同一通道内, 目标神经元 t 与其他神经元 x_i 间的线性可分性。通过采用二值标签, 并为函数添加正则项, 可以获得每个通道的能量函数, 之后对能量函数进行解析, 可以求得最小能量函数:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (2)$$

式中: $\hat{\mu}$ 和 $\hat{\sigma}^2$ 分别为通道内所有神经元的平均值和方差。然后使用最小能量函数加权处理输入特征, 重新赋值对应区域权重, 可以有效增强模型对显著特征的关注程度。最后经过一次 3×3 的卷积处理、批归一化操作以及 ReLU 函数激活, 提高模块的稳定性和鲁棒性, 避免过拟合。整个处理过程无需过多的卷积神经网络结构, 仅依靠对输入特征的矩阵转化计算, 在引入注意力机制的同时有效的控制了模型参数规模, 有助于边缘端的轻量化部署。

1.3 密集重叠目标检测损失函数

为了引导模型区分密集区域的重叠目标, 提高检测精度, 本文对检测头损失函数做了改进。针对重叠目标漏检现象, 设计了一个改善密集目标检测的损失函数 EWD。传统的损失函数中, 定位损失和置信度损失均与边界框损失有关, 基准模型默认使用 CIoU 损失来计算交并比, 但其内部的纵横比损失并不能代表锚框宽高与真实标注的差异。因此本文使用改进的 EIoU 损失计算矩形框交并比, 其公式为:

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \frac{\rho^2(w, w^{\text{gt}})}{C_w^2} + \frac{\rho^2(h, h^{\text{gt}})}{C_h^2} \quad (3)$$

式中: $\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})$ 代表两框中心点间的欧氏距离, c 代表两框最小闭包区域的对角线距离, C_w 和 C_h 分别是包含预测框和真实框两框最小闭合框的宽和高。损失函数合理考虑了 IoU 损失 L_{IoU} 、距离损失 L_{dis} 和宽高损失 L_{asp} , 能够有效加快参数收敛速度, 提高预测框和真实框对比相似性。此外, 由于训练样本锚框存在质量分布不平衡的问题, 低质量样本回归误差较大, 会使得计算梯度较大, 从而影响整个训练过程, 因此使用 Focal Loss 方式, 从梯度运算的角度引导模型更加关注高质量锚框。修改后损失函数为:

$$L_{\text{Focal-EIoU}} = \text{IoU}^\gamma L_{\text{EIoU}} \quad (4)$$

式中: $\text{IoU} = |A \cap B| / |A \cup B|$, γ 为调节控制异常值程度的超参。

单一的通过交并比来计算边界框损失, 仍然会使得模型对相邻重叠目标的轻微偏差不够敏感, 产生合并误检的现象。考虑到数据集及实测影像内不同尺寸目标的分布情况, 需要设计新的度量方式来评估重叠程度, 优化损失函数对不同大小目标的分配比重。归一化 Wasserstein 距离可以改变锚框差异的评估维度, 对不同大小的目标归一化处理, 使得锚框的重叠损失过渡平滑, 且具有良好的尺度一致性。本文首先将边界框 $R = (cx, cy, w, h)$ 进行二维高斯分布建模 $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$:

$$\mathbf{m} = \begin{bmatrix} cx \\ cy \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (5)$$

式中: c_x, c_y, w, h 分别为锚框中心点坐标、框宽和框高。之后使用 Wasserstein 距离来度量两框之间的分布相似程度, 对于转换后的二维高斯分布 $\mu_1 = \mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)$ 和 $\mu_2 = \mathcal{N}(\mathbf{m}_2, \mathbf{\Sigma}_2)$, 其 Wasserstein 距离可以简化表示为:

$$W_2^2(\mu_1, \mu_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_2^2 + \frac{\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2}{2} \quad (6)$$

式中: $\|\cdot\|_F$ 为弗罗贝尼乌斯范数。代入框 $A = (c_{xa}, c_{ya}, w_a, h_a)$ 和框 $B = (c_{xb}, c_{yb}, w_b, h_b)$ 的二维高斯分布 $\mathcal{N}_a$ 和 $\mathcal{N}_b$, 可以得到:

$$W_2^2(\mathcal{N}_a, \mathcal{N}_b) = \left\| \left(\begin{bmatrix} c_{xa}, c_{ya}, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} c_{xb}, c_{yb}, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_F^2 \quad (7)$$

最后对其进行指数归一化处理, 得到可以直接用作相似性度量的归一化 Wasserstein 距离:

$$\text{NWD}(\mathcal{N}_a, \mathcal{N}_b) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a, \mathcal{N}_b)}}{C}\right) \quad (8)$$

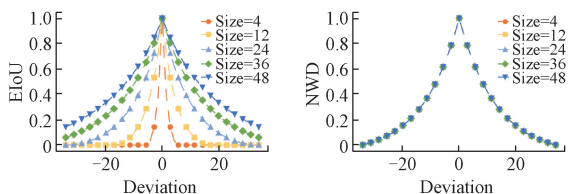
式中: C 是需要根据数据集特性进行调节的常数。综上, 本文最终设计了损失函数 EWD, 其公式为:

$$L_{\text{EWD}} = (1 - \lambda) \left(1 - \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a, \mathcal{N}_b)}}{C}\right) \right) + \lambda \left(1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \frac{\rho^2(w, w^{\text{gt}})}{C_w^2} + \frac{\rho^2(h, h^{\text{gt}})}{C_h^2} \right) \quad (9)$$

式中: λ 为调节目标损失函数计算比重的常数。

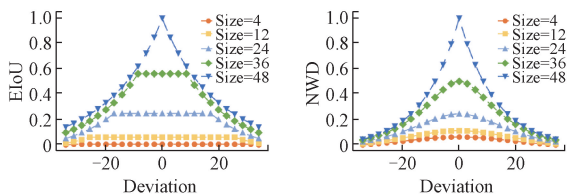
在目标框对角线方向上, 按照等差偏移步长构建预测框, 分别计算 NWD 函数与原始单一 IoU 损失在不同偏移下得到的重叠误差, 具体可视化结果如图 4 所示。通过图 4(a)可以看出当目标框与预测框尺寸相同时, NWD 相比传统的 IoU 损失具有更好的尺度一致性, 不同大小的锚框偏移相同的距离时, 造成的误差相同。通过图 4(b)可知当目标框一定, 预测框尺寸从大到小变化, NWD 的损失函数在两者相交后面积一定时, 误差损失仍能够随着中心偏移程度变化, 并不会产生水平不变的现象, 整体具有更

好的平滑度。因此模型可以更加准确的评估密集目标框间的重叠程度。



(a) 目标框与预测框尺寸一致

(a) The size of the object box is the same as that of the prediction box



(b) 目标框尺寸一定, 预测框尺寸变化

(b) The size of the object box is fixed, and the size of the prediction box is variable

图 4 偏移误差曲线

Fig. 4 Offset error curve

2 实验与分析

2.1 实验环境

实验工作站 CPU 为 48 核的 Intel® Xeon(R) Silver 4214CPU@2.20 GHz, 内存 128 G, GPU 为 2 块 NVIDIA GeForce RTX3090, 显存 48 G。操作系统为 Ubuntu18.04, 运算平台为 CUDA11.4, cuDNN8.2.4, 深度学习框架为 Pytorch1.8.1。边缘端测试部署使用华为 Atlas 200 低功耗加速模块, 集成海思 Ascend 310 AI 处理器, 内置 2 个 AI Core, 8 个 A55 Arm Core(max 1.6 GHz), 可支持 128 位宽的 LPDDR4x, 可实现最大 22 TOPS(INT8)的计算能力。

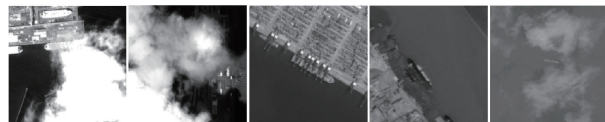
2.2 数据集

实验数据集 JS25k 主要以“吉林一号”卫星标注的船舶数据为主体, 同时挑选 DOTA 数据集、HRRSD 数据集和 NWPU VHR-10 数据集等公开数据集内的船舶类数据, 并进行下采样处理。原始共收集整理了 17 691 张图片, 尺寸范围集中在 400×400 pixels~ 800×800 pixels, 分辨率在 0.5~4 m, 主要以 2 m 为主。图 5(a)为常见的目标误检数据示意图(背景干扰、纹理丢失、云雾遮挡、目标重叠等造成图像的误检), 图 5(b)为常见困难检测数据示意图, 图 5(c)为普通检测数据示意图。

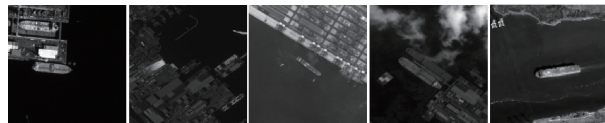
本文针对错误标注等现象进行了数据清洗, 剔除了错标、锚框偏移等错误数据; 同时针对样本尺度分布不均衡, 实际影像存在背景噪声、雨雪遮挡、大视角拍摄等情况, 采用传统的数据增强如裁剪、旋转、拼接、噪声干扰和空间坐标仿射变换等方法, 对小样本数据和高价值目标数据进行了扩充, 减小了数据分布对模型训练的干扰。目前实验数



(a) 误检数据样例
(a) False positives data examples



(b) 困难数据样例
(b) Difficult data examples



(c) 普通数据样例
(c) Common data examples

图 5 数据集样例

Fig. 5 Dataset examples

据集整体规模为 25 326 张, 并按照 6:2:2 随机划分为训练集、验证集及测试集。

2.3 基准测试

Jiang 等^[20]的研究表明, 对于某些特定领域的数据集, 高版本的 YOLO 模型的检测性能并不一定更好, 且 YOLO 系列模型在 v5 之后由各个科研团队开发, 版本整体的继承性并不连续。为了选取最优的基准模型, 本文首先使用自建数据集 JS25k 作为训练样本, 对当前工程上广泛应用的 YOLO 系列模型如 YOLOv5 和 YOLOv8 等, 以及思路相近的 YOLOX 和 YOLOR 模型进行了基准对照训练测试, 过程中考虑到模型复杂度(GFLOPs)和参数量(Params)需要满足在轨部署限制, 最终选取如下模型作为代表进行测试, 测试过程中严格控制训练流程及参数配置, 测试结果如表 1 所示。

表 1 YOLO 系列算法基准测试

Table 1 Performance evaluation of the YOLO algorithms

模型	mAP(0.5)/%	Params/M	FLOPs/G
YOLOv3	91.3	61.9	156.4
YOLOv4	91.2	64.4	142.8
YOLOv5l	97.1	44.0	107.3
YOLOv6l	96.5	59.6	150.7
YOLOv7l	95.7	36.9	104.7
YOLOv8l	96.9	43.7	165.2
YOLOXm	89.7	25.3	73.8
YOLOR_csp	89.3	52.9	120.4

通过检测结果可以看出, YOLOXm 的模型复杂度最低, 参数量最少, 但其检测性能较差, 检测精度仅为

89.7%。在同等参数规模和运算量下, YOLOv5l 准确率高于 YOLOv6l 和 YOLOv7l, 与 YOLOv8l 相近, 但 YOLOv8l 的模型复杂度较 YOLOv5l 提升了约 54%, 难以满足在轨测试的时效性需求, 因此最终选择 YOLOv5l 作为基准模型进行改进。

2.4 消融实验

本文在基准模型上进行了多项优化, 为了验证各个模块对模型性能的提升, 同步进行了大量消融实验。过程中严格统一控制参数配置, 逐一增加控制变量, 最终实验结果如表 2 所示。

表 2 算法模块消融实验
Table 2 Ablation experiments of algorithm module

EIoU	NWD	AT1	AT2	AT3	DFF	mAP(0.5)/%	Params/M	FLOPs/G
						97.1	43.99	107.3
✓						97.3	43.99	107.3
	✓					97.4	43.99	107.3
✓	✓					97.6	43.99	107.3
		✓				95.8	43.99	107.3
			✓			95.5	43.99	107.3
				✓		97.6	43.99	107.3
					✓	97.9	47.41	110.1
✓	✓			✓	✓	98.2	47.41	110.1

表 2 中依次为损失函数、特征交互模块以及特征融合模块对模型检测性能提升的效果, 其中 AT1~AT3 分别表示特征交互模块嵌入到模型的不同位置的性能对比, 结果表明 AT3 下, 将特征交互模块嵌入到网络的特征连接层内效果最佳。分析测试结果, 可以看出各部分改进均提升了模型性能, 最终的 MS-YOLO 准确率达到 98.2%, 较基准模型提高了 1.1%。此外本文还绘制了各个对照实验训练过程中的 P - R 曲线, 如图 6 所示。对比各个改进模块对应 P - R 曲线局部放大时的包含关系以及均值平均精度, 可以直观看出各改进模块的对模型性能的提升。

重分配得到了加强, 更加关注局部细节信息, 对密集重叠目标的检测能力有了进一步提升。

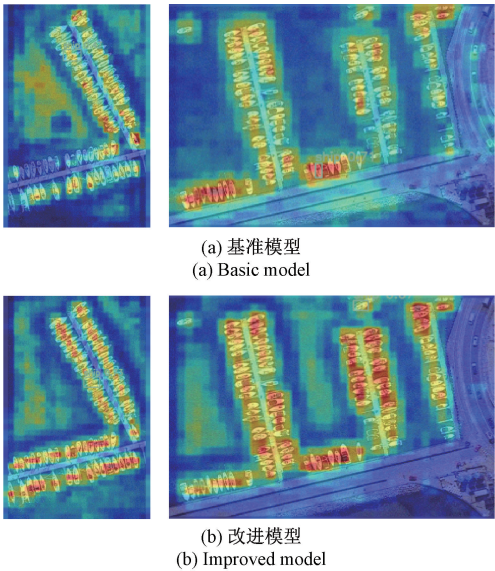


图 7 密集重叠目标检测及热力图对比
Fig. 7 Dense overlapping object detection and heat map comparison

此外, 本文在地面 PC 端对不同场景下真实的遥感影像进行了对照测试。分别使用基准模型和改进的 MS-YOLO 模型对卫星在轨实拍图像进行识别检测, 部分检测效果如图 8 所示。图 8(a)为基准模型检测结果, 图 8(b)为改进模型检测结果, 可以看出 MS-YOLO 在地物遮挡、背景干扰以及目标重叠等问题上均有不同程度的改善, 总体上看, 在参数量、运算复杂度轻微增加的情况下, 模型的检测性能得到了显著提升。

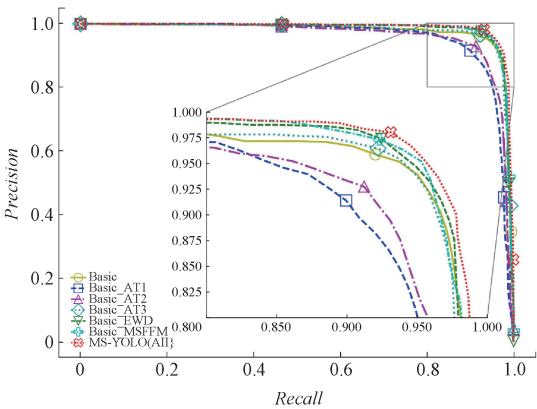


图 6 算法模块 P - R 曲线
Fig. 6 P - R curve of algorithm model

为了进一步展现模型改进的有效性, 本文选取了典型的密集重叠目标图像进行检测并绘制了对应的热力图, 如图 7 所示。对比图 7(a)基准模型与图 7(b)改进模型的热力图, 可以看出改进 MS-YOLO 模型对目标显著区域的权

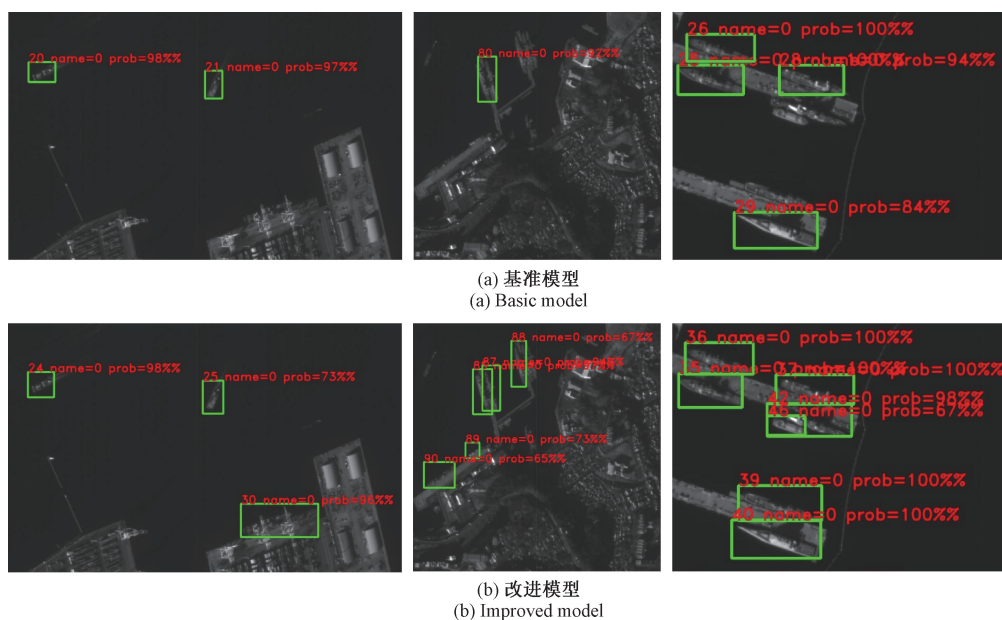


图 8 遥感影像复杂场景下检测结果对比

Fig. 8 Comparison of detection results under complex conditions of remote sensing images

2.5 对比实验

为了更加客观的评估 MS-YOLO 的性能,同时考虑模型参数量和运算复杂度,本文选取了当前代表性的 YOLO 系列模型^[21]及 RT-DETR 系列模型^[22]进行了对比训练测试,训练过程严格控制参数配置,最终实验结果如表 3 所示。

表 3 模型对比实验

Table 3 Comparative experiments on different models

模型	mAP(0.5)/ %	AR	Params/ M	FLOPs/ G
YOLOV9m	94.9	89.1	20.1	76.8
YOLOV9c	95.4	90.4	25.5	102.8
YOLOV9e	96.9	93.5	58.1	192.5
YOLOV10l	96.4	92.3	24.4	120.3
YOLOV10x	97.1	93.8	29.5	160.4
RT-DETR-l	97.8	—	42.0	108.0
RT-DETRv2-l	97.7	—	42.0	108.0
YOLOV11l	97.4	93.0	25.3	86.9
YOLOV11x	97.7	93.8	56.9	194.9
YOLOV12l	97.0	92.8	26.4	88.9
YOLOV12x	97.6	93.9	59.1	199.0
YOLOV13l	97.6	93.2	27.6	88.4
MS YOLO(本研究)	98.2	94.8	47.4	110.1

可以看出,同版本 YOLO 模型内,参数规模及运算复杂度的增大确实提高了检测性能,但跨版本间对比, YOLO_{v11} 在数据集 JS25k 上的检测效果优于 YOLO_{v12},

而 YOLOv13l 的效果与规模更大的 YOLOv12x 相当。对比各项指标,本研究提出的 MS-YOLO 模型检测精度优于当前 SOTA 方法,虽然在参数量和运算复杂度上较最新的 YOLOv13l 和 RT-DETRv2 等有一定的差距,但在平均检测精度和召回率上分别高于 YOLOv13l 模型 0.6% 和 1.6%,综合性能满足工程需求。

为了全面直观的对比各个模型性能,对测试结果进行了可视化处理,如图 9 所示。

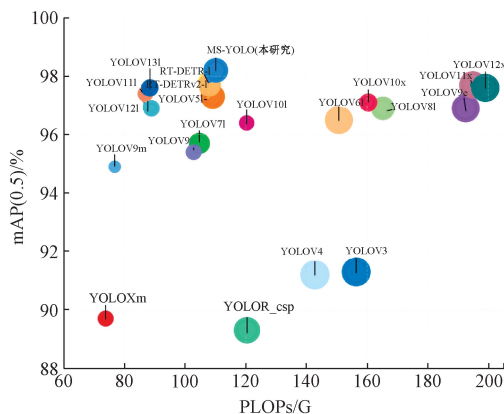


图 9 模型算法性能对比

Fig. 9 Performance comparison of model algorithms

图 9 中圆形半径代表模型参数量, 横纵坐标代表模型的推理速度和检测精度, 可以看出本文提出的 MS-YOLO 模型的模型复杂度远小于同等性能的 YOLOv11-x 和 YOLOv12-x 等模型, 检测精度高于同规模的 YOLOv13l 和 RT-DETR-l 等模型, 整体表现优异。

综上所述,MS-YOLO 具备良好的检测性能和便利的部

署方式,目前该套模型已经通过量化、模型转换等操作部署到了项目组自研的在轨 AI 载荷中,载荷如图 10 所示。边缘端进行多组不同场景遥感影像的测试,在满载算力 22 TOPS 的情况下,针对 416×416 pixels 大小的瓦片数据,模型的处理速度可以达到 92.3 fps,整体满足业务时效性需求。

在边缘端对在轨拍摄的真实船舶影像进行检测、下传和解译等处理,并在底图上绘制实际标注结果。如图 11 所示。在港口、海面等典型场景下不同尺寸的船舶,模型都具有良好的检测结果。

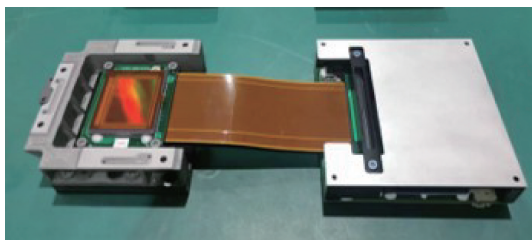


图 10 自研在轨 AI 载荷

Fig. 10 Self-developed on-orbit AI payload

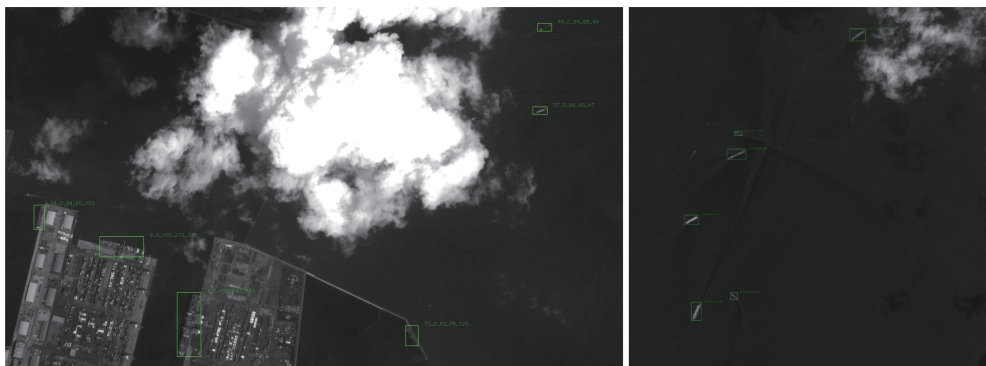


图 11 模型边缘端部署检测结果

Fig. 11 Detection results of edge deployment mode

3 结 论

本研究提出了一种基于多尺度特征融合的检测模型 MS-YOLO。首先,拓展多层次感受野,交叉融合不同维度特征,提高特征提取能力;之后,在连接层嵌入注意力机制,计算最小能量函数,提高显著区域细节特征的表达能力;最后,设计损失函数 EWD,优化不同尺度锚框间的度量方式,降低合并误检概率。此外,构建了一套场景丰富的遥感船舶数据集 JS25k。所提模型在数据集上的检测精度达到了 98.2%,综合性能优于当前 YOLO 系列模型和 RT-DETR 系列模型,在边缘端部署后能够满足业务时效性需求。目前模型在特定角度和某些复杂场景下,仍存在一定误检的现象,未来将考虑扩充困难样本,改进模型检测头旋转框结构,进一步提高模型的检测性能和泛化能力。

参考文献

- [1] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [2] 聂光涛, 黄华. 光学遥感图像目标检测算法综述[J]. 自动化学报, 2021, 47(8): 1749-1768.
- NIE G T, HUANG H. A survey of object detection in optical remote sensing images[J]. Acta Automatica Sinica, 2021, 47(8): 1749-1768.

- [3] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [4] MING Q, MIAO L J, ZHOU Z Q, et al. CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.
- [5] 徐丹青, 吴一全. 光学遥感图像目标检测的深度学习算法研究进展[J]. 遥感学报, 2024, 28(12): 3045-3073.
- XU D Q, WU Y Q. Progress of research on deep learning algorithms for object detection in optical remote sensing images[J]. National Remote Sensing Bulletin, 2024, 28(12): 3045-3073.
- [6] SAPKOTA R, CHEPPALLY R H, SHARDA A, et al. RF-DETR object detection vs YOLOv12: A study of transformer-based and CNN-based architectures for single-class and multi-class greenfruit detection in complex orchard environments under label ambiguity[J]. ArXiv preprint arXiv:2504.13099, 2025.
- [7] ZHAI H Z, DU J W, AI Y H, et al. Edge deployment of deep networks for visual object detection: A review[C]// IEEE Sensors Journal, 2024.
- [8] LI P, CHE C. SeMo-YOLO: A multiscale object detection network in satellite remote sensing images [C]//2021

- International Joint Conference on Neural Networks (IJCNN), 2021: 1-8.
- [9] YIN R T, XU Q Z, DING Y F. Ship detection from optical remote sensing imagery based on scene classification and saliency-tuned retinanet [C]//2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2021: 3553-3556.
- [10] HU R Z, TING R. An object detection method for remote sensing images based on DA-YOLO [C]//International Conference on Image, Vision and Intelligent Systems, 2022: 119-130.
- [11] 王迎龙, 孙备, 丁冰, 等. BG-YOLO:复杂大视场下低慢小无人机目标检测方法[J]. 仪器仪表学报, 2025, 46(2): 255-266.
- WANG Y L, SUN B, DING B, et al. BG-YOLO: A low-altitude slow-moving small UAV target detection method in complex large field of view[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 255-266.
- [12] 李泽胤, 李栋, 房建东, 等. 改进的 YOLOv8n 遥感图像轻量化检测模型[J]. 电子测量技术, 2025, 48(6): 130-142.
- LI Z Y, LI D, FANG J D, et al. Improved YOLOv8n lightweight detection model for remote sensing images[J]. Electronic Measurement Technology, 2025, 48(6): 130-142.
- [13] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [14] WEI X T, LI Z S, WANG Y T. SED-YOLO based multi-scale attention for small object detection in remote sensing [J]. Scientific Reports, 2025, 15(1): 3125.
- [15] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [16] 朱圣博, 魏利胜, 高港, 等. 基于改进 YOLOv8s 的光学遥感小型船舶检测算法[J]. 电子测量与仪器学报, 2024, 38(10): 48-57.
- ZHU S B, WEI L S, GAO G, et al. Optical remote sensing small ship detection algorithm based on improved YOLOv8s [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 48-57.
- [17] YANG L X, ZHANG R Y, LI L D, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]//International Conference on Machine Learning, 2021: 11863-11874.
- [18] WANG J W, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. ArXiv preprint arXiv:2110.13389, 2021.
- [19] ZHANG Q L, YANG Y B. SA-Net: Shuffle attention for deep convolutional neural networks [C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021: 2235-2239.
- [20] JIANG T Y, ZHONG Y. ODverse33: Is the new YOLO version always better? A multi domain benchmark from YOLO v5 to v11[J]. ArXiv preprint arXiv:2502.14314, 2025.
- [21] KHANAM R, HUSSAIN M. A review of YOLOv12: Attention-based enhancements vs. previous versions[J]. ArXiv preprint arXiv:2504.11995, 2025.
- [22] LYU W Y, ZHAO Y, CHANG Q Y, et al. RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer [J]. ArXiv preprint arXiv:2407.17140, 2024.

作者简介

霍东升(通信作者), 硕士, 工程师, 主要研究方向为嵌入式 AI、遥感数据处理。

E-mail: huodongsheng@jll.cn

王栋, 博士, 研究员, 主要研究方向为卫星设计、遥感数据处理。

E-mail: dshuo08@163.com

刘云贺, 硕士, 工程师, 主要研究方向为遥感数据处理。

E-mail: dshuo08@163.com