

改进 YOLOv8 的水面红外目标识别方法研究

刘建华, 朱梦琳[†], 尹泉贺

(华北水利水电大学信息工程学院, 河南 郑州 450000)

摘 要:针对红外船舶图像目标识别任务中,因存在不同尺度大小的红外图像目标且噪声干扰大导致特征信息丢失漏检误检率高的挑战,提出了一种基于 YOLOv8 的船舶目标识别算法 CL_YOLOv8。首先,针对输入的船舶数据经过逐层的特征提取和空间变换后丢失大量信息的问题,设计可编程梯度信息结构 CbPGI,使模型获得可靠的梯度信息来更新网络权重,提高检测精度。其次,针对 CbPGI 模块使模型计算复杂度、参数量增加的问题,引入高速推理卷积 PartialConv,构建 LWPC 结构。该结构只在输入通道的一部分上应用常规卷积进行空间特征提取并融合,减轻模型存储需求。构建 LWPC 结构使算法计算复杂度下降了 31.3%,模型参数量也下降了 22.8%。相同实验条件下,CL_YOLOv8 算法在自建的船舶数据集上检测精度 mAP50 达到 91.0%,相较于基线模型提升了 3.05%。

关键词:水面红外目标;船舶识别;多尺度融合;YOLOv8

中图分类号:TP273

文献标识码:A

Research on Water Surface Infrared Target Recognition Method with Improved YOLOv8

LIU Jianhua, ZHU Menglin[†], YIN Quanhe

(School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, Henan 450000, China)

Abstract: During the task of targets identifying infrared ship images, due to the presence of targets of various scales in the infrared images and noise interference that leads to loss of feature information, the resulting algorithm in high missed and false detection, propose a YOLOv8-based ship target recognition algorithm named CL_YOLOv8. Firstly, for the input ship data, a large amount of information is lost during feature extraction and spatial transformation layer by layer, a programmable gradient information structure CbPGI (Combined Programmable Gradient Information) is designed to enable the model to obtain reliable gradient information to update network weights and improve detection accuracy. Secondly, to address the problem of increased model computational complexity and parameters caused by the CbPGI module, we introduced the high-speed inference convolution PartialConv. We constructed the LWPC (Lightweight PartialConv) structure. This structure applies conventional convolution on only a portion of the input channels for spatial feature extraction and fusion, reducing the model storage requirements. The algorithm computational complexity of constructing the LWPC structure decreased by 31.3%, and the model parameter decreased by 22.8%. Under the same experimental conditions, the CL_YOLOv8 algorithm achieved a detection accuracy of 91.0% in mAP50 on the self-built ship dataset, which is 3.05% higher than the baseline model.

Key words: surface infrared targets; ship recognition; multi-scale fusion; YOLOv8

海上船舶目标识别可以提高海上交通的安全性和效率。基于船载摄像头的实时监控技术通过在水上航行的船舶上搭载光电平台^[1],利用计算机视觉算法对摄像机返回的视频帧进行船舶目标检测,一旦发现可疑目标立即发出预警,从而更好地规划船舶航线和进行调度。红外遥感通过“测量”物体向外辐射的热量获得图像,具有穿透性强的特点,不受大雾、下雨、夜间或能见度低等恶劣环境的影响,可以提供清晰的图像全天候监控海上目标,因此本文选择红外船舶图像进行研究。

传统的红外图像检测方法有以下几种,基于均值滤波^[2]、中值滤波^[3]、高斯滤波^[4]、HVS、稀疏表示等方法。基于深度学习的检测方法是运用卷积神经网络对大量数据进行学习以获取目标特征,通过直接回归并进行非极大值抑制后输出结果,检测方法分为二阶段算法和单阶段算法。二阶段算法的代表有 RCNN^[5]、SPPNet^[6]、Faster RCNN^[7]等,单阶段的检测算法有 SSD^[8]、YOLO^[9]等。目前深度学习在红外目标检测领域取得了一些重大进展,但仍存在一些挑战和需要改进的地方。例如,一些较大的模型在计算资源有限的场景中的应用不广泛,同时,在噪声干扰下,某些模型识别精度将会受到挑战。针对上述问题,本文的主要工作如下:

(1)构建双主干特征提取 CbPGI 结构,CbPGI 结构包括主分支、辅助可逆分支和多级辅助结构,解决了船舶目标经过逐层的特征提取和空间变换时存在信息瓶颈和可逆函数问题。

(2)针对上述结构参数量、计算复杂度变大的问题,引入高速推理卷积 PartialConv 构建 LWPC 结构,在输入通道的一部分上应用常规卷积进行空间特征提取并融合,减少计算复杂度和降低内存访问同时保持特征信息流。

(3)提出一种新的船舶目标识别算法 CL_YOLOv8,该算法在恶劣环境中目标多尺度情况下具有更高的检测精度。

1 改进的 CL_YOLOv8 算法

1.1 CbPGI 特征提取模块

CbPGI 模块通过辅助可逆分支和多级辅助信息两条路径来提供可靠的梯度,并指导主分支学习,形成一个互补和增强的网络结构(图 1)。CbPGI 模块主要由三个部分组成。

(1)主分支:用于推理的架构,沿用基模型的

结构。

(2)辅助可逆分支:将可逆分支视为深度监督分支的扩展,设计辅助可逆分支。在目标检测网络的特征学习过程中,主分支深层易受信息瓶颈效应影响,导致关键特征信息丢失,进而削弱其对目标任务的适配性。针对这一痛点,所设计的辅助可逆分支能够向主分支深层传递可靠的梯度信息,通过梯度反向传播驱动主分支参数的优化学习,引导模型聚焦并提取与目标检测任务高度相关的核心特征。辅助可逆分支包括 CBFuse 模块、c2f 卷积和普通卷积。在训练阶段,CBFuse 将来自不同层级的特征图通过上采样到同一尺寸,然后进行融合,以增强模型对目标的识别能力。

(3)多级辅助信息:多级辅助信息的概念是在辅助监督的特征金字塔层之间和主分支之间插入一个集成网络,它结合不同预测头返回梯度。多层次辅助信息是为了汇总包含所有目标对象的梯度信息,并将其传递给主分支,然后更新参数。此时,主分支的特征金字塔层次的特性不会被某些特定对象的信息主导,可以缓解深度监督中的断裂信息问题。

深层信息能够提供更高级别的抽象特征和上下文信息,这对于处理不同尺度的问题至关重要。这些特征可以帮助模型更准确地识别和定位对象。

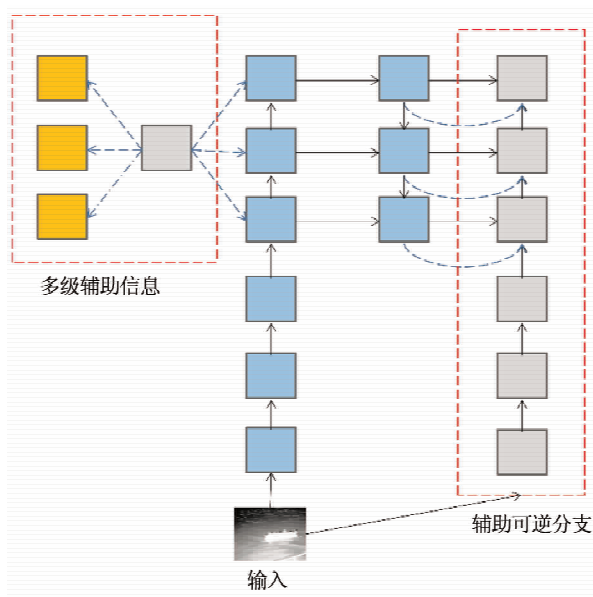


图 1 CbPGI 模块网络结构图

1.2 轻量化 LWPC 结构

LWPC 结构由 ConvModule、Split、PartialConv^[10]卷积组成。其结构如图 2 所示。其中,PartialConv 仅对有效区域卷积,在降低模型通道

数、减少计算冗余与内存访问的同时,仍能有效提取空间特征。其输出分为特征输出和掩码更新两部分。特征输出计算中,设 W 为卷积核的权重, b 为对应的偏置; X 代表当前卷积(滑动)窗口内的特征值(像素值), M 为对应的二值掩码。经过局部卷积操作后的特征数据可以表示为:

$$x' = \begin{cases} WT(X \times M) \frac{\text{sum}(1)}{\text{sum}(M)} + b & \text{if } \text{sum}(M) > 0 \\ 0, & \text{其他} \end{cases} \quad (1)$$

式中, $\text{sum}(1)$ 表示滑动窗口对应的像素点数

量,即卷积核大小。如式(1)所示,经过局部卷积得到的特征数据只与未屏蔽的输入数据有关,即只与掩模图 M 上数据 1 所对应位置的像素点有关。

局部卷积操作后,按如下方式更新掩码:若卷积操作的输出至少依赖了一个有效输入值,则将该位置标记为有效。表达式为:

$$m' = \begin{cases} 1, & \text{if } \text{sum}(M) > 0 \\ 0, & \text{其他} \end{cases} \quad (2)$$

经过局部卷积层的多次连续运算后,掩码最终会全部变为 1,即提取输入中的有效像素。

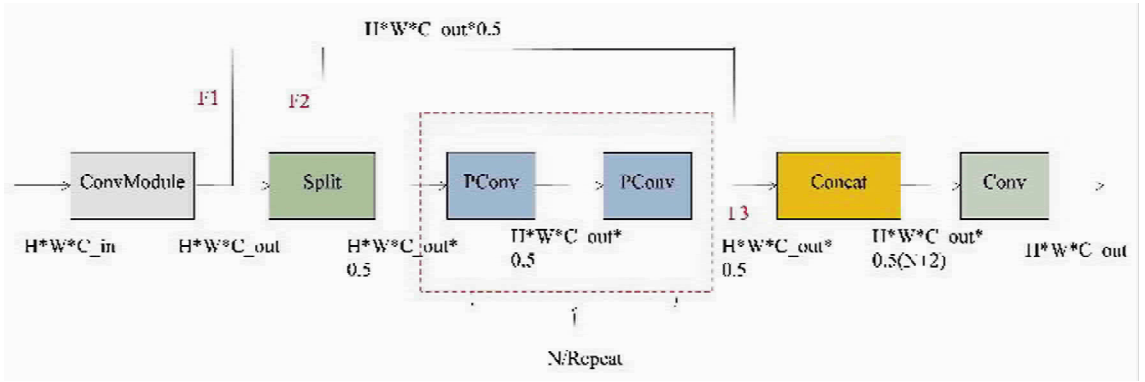


图 2 LWPC 模块网络结构

1.3 CL_YOLOv8 网络结构

CL_YOLOv8 算法针对随着网络层数的不断加深带来的信息瓶颈和可逆函数问题,设计新的特征提取结构 CbPGI。该结构包括主分支、多级辅助结构和辅助可逆分支,使网络获得更加可靠的梯度信息,提高模型检测精度,减轻计算复杂度,但同时牺牲掉了一些模型参数量。其次针对该模型的参数量和计算复杂度变大的问题,引入高速推理卷积 PartialConv 构建 LWPC 结构,只在输入通道的一部分上应用卷积进行空间特征提取,减轻模型参数量,提升船舶目标检测精度同时也减少了存储需求。

2 实验结果与分析

2.1 实验环境和评估指标

实验采用 win10 系统,显卡型号 RTX 3080,软件配置 Anaconda、Pycharm,python 版本 3.8,torch 版本 1.11.0+cu113。经过使用预训练权重和不使用预训练权重结果对比,选择了不使用预训练权重的方法,训练轮数为 100 个 epoch。评估指标采用平均精度 mAP50、GFLOPs(每秒十亿次浮点运算)和参数量(Parameters)作为衡量检测精度的指标。GFLOPs 用来描述模型的计算复杂度;参数量则表示模型的参数个数,通常用于描述模型

的大小和存储需求。

$$AP50 = \frac{1}{n} \sum_{i=1}^n P_j^{IoU=0.5} (R_j^{IoU=0.5}) \quad (3)$$

式中, P 为 Precision 精确度,表示正确预测的正样本数(True Positives, TP)与所有预测为正样本的样本数之比表示准确率; R 为召回率,表示正确预测的正样本数与所有实际正样本的样本数之比; AP50 是指当并集交集(IoU)阈值设置为 50% 时,所有类别的平均精度。

2.2 数据集选取

本数据集分为两部分,一部分采用艾睿光电红外开源平台^[11]提供的 8402 张船舶照片,另一部分来自网络收集,共 998 张船舶图片,并用开源图形图像注释工具 labelImg 中的矩形框对船舶进行标注(图 3)。以图片的左上角为坐标原点(0,0),使用 $(x_1, y_1), (x_2, y_2)$ 的形式记录矩形框的位置, x_1 表示矩形框的左上角横坐标, y_1 表示矩形框的左上角纵坐标, x_2 表示矩形框的右下角横坐标, y_2 表示矩形框的右下角纵坐标。所有标签信息以 txt 文件的形式进行保存。本数据库包含了不同场景、不同时段、不同分辨率的海上和港口的邮轮、散货船、军舰、帆船、皮划艇、集装箱船、渔船目标共计 9400 张照片。



图 3 labelImg 标注页面

2.3 不同模型对比实验

本实验用 YOLOv5s、YOLOv7、YOLOv8 这几个网络模型进行实验评估,实验结果如表 1 所示。YOLOv5s 检测精度 map @0.5 达到 92.3,但参数量、计算复杂度都偏大,对边缘设备的部署是一个挑战。YOLOv7 的检测精度 map @0.5 为 84.4,精度偏低。YOLOv8 检测精度 map @0.5 为 88.3,参数量、计算复杂度在可接受的范围内。CL_YOLOv8 (本文算法)的检测精度 map @0.5 为 91.0,参数量和计算复杂度分别为 4.4 Params/M 和 12.7 GFLOPs。

表 1 实验结果分析

模型	Map@0.5	Map@0.5-0.9	参数量/M	GFLOPs
YOLOv5s	92.3	65.6	7.0	15.8
YOLOv7	84.8	60.2	37.3	105.2
YOLOv8	88.3	62.0	3.0	8.2
CL_YOLOv8	91.0	64.4	4.4	12.7

2.4 消融实验

为验证本文算法的特征提取能力和目标识别性能,对算法进行消融实验。实验选取 YOLOv8 为基模型,基模型的检测精度 map @0.5 为 88.3,参数量和计算复杂度分别为 3.0M 和 8.2GFLOPs。加入 CbPGI 结构之后模型的检测精度 map @0.5 为 91.4,参数量和计算复杂度分别为 5.7M 和 18.5GFLOPs。加入 LWPC 结构之后检测精度 map @0.5 为 91.0,参数量和计算复杂度分别为 4.4M 和 12.7GFLOPs。消融实验结果如表 2 所示。

表 2 消融实验

模型	Map@0.5	Map@0.5-0.9	参数量/M	GFLOPs
Base	88.3	62.0	3.0	8.2
+CbPGI	91.4	64.9	5.7	18.5
+CbPGI+LWPC	91.0	64.4	4.4	12.7

2.5 实验结果可视化

为了更加直观地呈现图像中的船舶识别情况,将实验结果放在图中显示。图 4 是基模型的检测效果,可以看出 liner、canoe、fishing boat 的检测效果分别为 84.6%、75.7%和 73.3%。图 5 是改进后的 CL_YOLOv8 算法,可以看到改进后的算法对 liner、canoe、fishing boat 的检测效果分别达到 87.3%、83.3%和 79.9%,同时 bulk carrier 和 sailboat 也有所提升。图 6 是两个算法的准确率、召回率、mAP50 以及 mAP50: 95 的结果对比,可以看出 CL_YOLOv8 算法具有更好的检测效果。

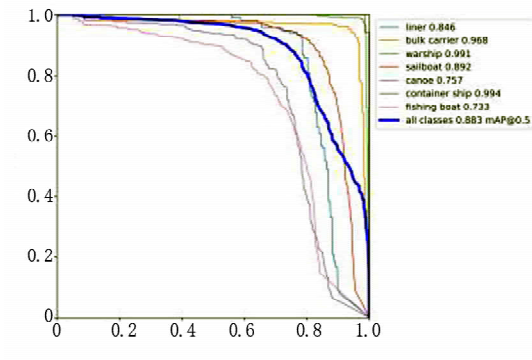


图 4 基模型检测效果

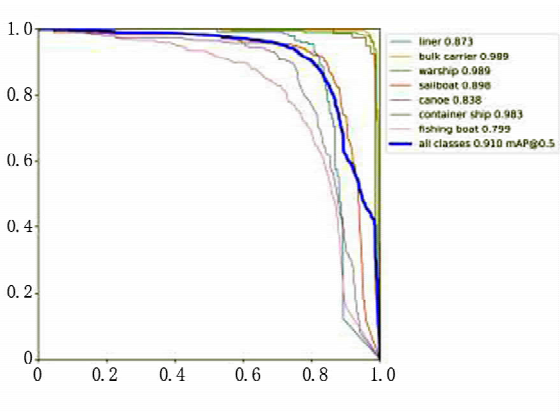


图 5 CL_YOLOv8 算法检测效果

3 结 论

使用 YOLOv8 作为基模型,针对红外船舶图像目标识别任务中因船舶多尺度且噪声干扰导致特征信息丢失漏检误检率高的挑战,设计双主干特征提取网络 CbPGI 结构。该结构能获得更加可靠的梯度信息和上下文信息,提高检测精度。其次针对 CbPGI 双主干网络参数量、计算复杂度变大的问题,引入轻量化 PartialConv 卷积,构建 LWPC 结构。构建 LWPC 结构的算法计算复杂度下降了

31.3%，模型参数量也下降了 22.8%。与基模型相比，CL_YOLOv8 算法在船舶数据集上将 mAP50 提高了 3.05%，检测精度达到 91.0%。

由于环境因素如烟雾或光照条件，船舶的边界

变得模糊、部分被遮挡，未来将进一步关注更加优秀的损失函数、注意力机制等改进方案，提升算法处理边界框的能力。

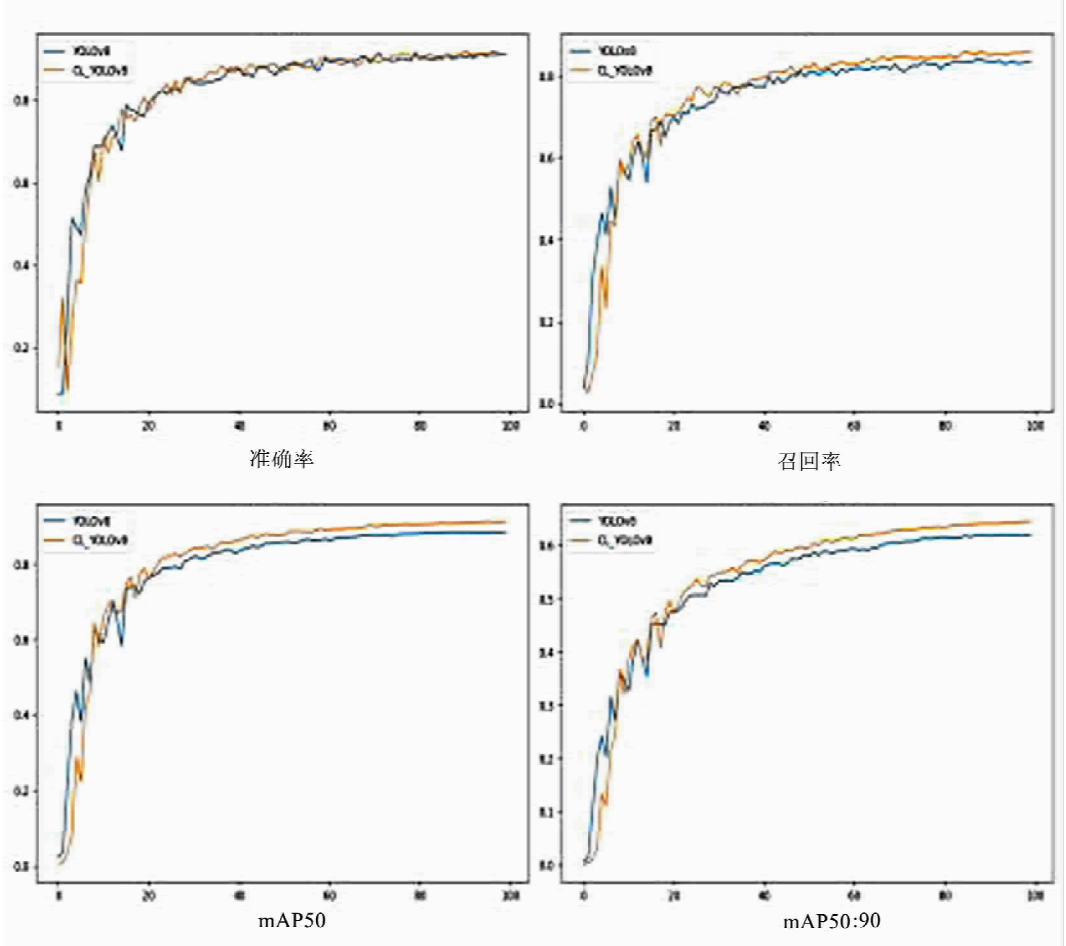


图 6 两种算法对比

参考文献

[1] 金久才,张家林,刘德庆,等. 基于改进 SiamRPN 的高速无人船海上目标视觉跟踪研究 [J]. 海洋科学进展,2024,42(3): 545—554.

[2] RADHIKA R, MAHAJAN R. An adaptive optimum weighted mean filter and bilateral filter for noise removal in cardiac MRI images[J]. Measurement: Sensors, 2023, 29: 100880.

[3] ZHANG Y C, YUAN Y X. ResNet-based surface normal estimator with multilevel fusion approach with adaptive median filter region growth algorithm for road scene segmentation [J]. International Journal of Computational Vision and Robotics, 2024, 14(1): 99—117.

[4] ZHANG Y X, SHI Z X, WANG H Q, et al. An efficient autofocus method for microscope based on the improved first-order derivative Gaussian filtering operator[J]. Optics & Laser Technology, 2024, 170: 110225.

[5] 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.

[6] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904—1916.

[7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28.

[8] 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.

[9] 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.

[10] CHEN J R, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021—12031.