

基于改进YOLOv8算法的复杂场景船舶目标检测

车晓辰, 马淑华, 郭泽旭, 沙晓鹏

(东北大学秦皇岛分校 控制工程学院, 河北 秦皇岛 066004)

摘 要: 为了提高复杂场景下船舶目标检测的精度和鲁棒性, 基于YOLOv8算法进行模型改进. 在骨干层引入CD3模块并嵌入无参注意力SimAM模块; 在颈部层引入注意力尺度序列融合ASF(attention-based scale sequence fusion)模块; 在头部层增加了检测头用于预测输出; 采用剪枝降低模型计算量; 最后使用蒸馏进一步提升模型性能. 实验在阿里天池官网提供的复杂场景船舶检测数据集上进行验证, 结果表明, 改进后模型的 AP_{50} 和AP相比于YOLOv8分别提升4.7%和2.9%, 召回率和精确率分别提升3.2%和4.2%, 而模型参数量减少56.1%, 计算量降低30.5%. 改进后的模型在减少参数的同时, 提升了总体性能.

关键词: YOLOv8算法; CD3; ASF-YOLO; SimAM; 剪枝蒸馏

中图分类号: TP 391.4 **文献标志码:** A **文章编号:** 1005-3026(2025)12-0038-10

Ship Target Detection in Complex Scenarios Based on Improved YOLOv8 Algorithm

CHE Xiao-chen, MA Shu-hua, GUO Ze-xu, SHA Xiao-peng

(School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China. Corresponding author: CHE Xiao-chen, E-mail: 353469229@qq.com)

Abstract: To improve the accuracy and robustness of ship target detection in complex scenarios, modifications were implemented based on the YOLOv8 algorithm. The CD3 module was introduced in the backbone layer with the parameter-free attention SimAM module embedded. The attention-based scale sequence fusion (ASF) module was incorporated in the neck layer, and an additional detection head was added to the head layer for prediction output. Pruning was adopted to reduce the computations of the model, followed by distillation to further improve model performance. The experiment was conducted on the complex scenario ship detection dataset from Alibaba Tianchi for verification. The results demonstrate that compared with YOLOv8, the improved model achieves increases of 4.7% in AP_{50} and 2.9% in AP, respectively. Recall and precision are improved by 3.2% and 4.2%, while model parameters and computations are reduced by 56.1% and 30.5%. The optimized model thus improves overall performance while reducing parameters.

Key words: YOLOv8 algorithm; CD3; ASF-YOLO; SimAM; pruning and distillation

随着水路交通的发展, 海上运输不断增加, 导致海上船舶种类和数量不断增多, 某些水域变得十分拥挤, 船舶在运行过程中发生事故的概率大大增加^[1]. 随着计算机视觉的发展, 基于卷积神经网络的算法在船舶检测领域得到广泛应用. 周旗开等^[2]提出一种基于改进YOLOv5s的轻量级船舶目标检测方法, 构建了融合ECA (efficient channel

attention) 模块与BiFPN (bidirectional feature pyramid network) 模块的目标检测模型. Li等^[3]提出了一种基于改进Faster-RCNN的船舶检测算法, 设计了一种连接结构实现特征放大和多尺度特征跳跃, 该结构提升了船舶检测精度但降低了检测速度. Jiang等^[4]提出了一种基于改进YOLOv7-tiny模型的轻量级目标检测方法, 在主干网络中引入

坐标注意力机制,将 ODConv(omni-dimensional dynamic convolution)添加到网络中;另外,在模型中引入了 CARFE(content-aware reassembly of features)和 SIoU(smoothed-IoU)损失函数以提高网络对小目标的收敛速度和检测精度.Zhao 等^[5]提出了一种基于 EYOLOX-VFL 的船舶目标检测方法,在 YOLOX 骨干网络中加入 ECA 模块,将损失函数替换为 EFIoU(efficient force-IoU),以增强模型对不同尺度物体信息的提取能力,提高算法的回归性能.Lyu 等^[6]提出了一种基于 YOLOv5s-lite-CBAM 模型的检测系统,将 YOLOv5s 中的 C3 模块替换为 ShuffleNetV2 和 CBAM(convolutional block attention module)注意力机制的组合模块,减轻了模型体量,提高了检测精度.Zhao 等^[7]提出了一种基于 Swin-Transformer 的 YOLO 模型,在骨干网络中嵌入 Swin-Transformer 网络架构和坐标注意力机制;其次,利用具有残差结构的 PANet(path aggregation network)构建特征金字塔,提高全局特征提取能力。

YOLOv8^[8]是一种以实时性能和优异准确率著称的目标检测模型.然而,其在复杂海洋环境中的船舶检测效果仍然有限.为解决这一挑战,本文针对复杂场景下的船舶目标检测任务,提出了改进的 YOLOv8 算法,主要贡献如下:

1) 由于船舶位置和拍摄角度不同,类别众多且形状复杂,因此在全局的 C2f(cross stage partial network bottleneck with 2 convolutions and a faster implementation)结构中增加了 DCNv3^[9].该模块允许卷积核动态调整输入特征图的空间采样位置,从而能够更好地捕捉目标物体在尺度、角度及形变等方面的多样性。

2) 为了捕获不同尺度的船舶信息,将颈部层使用 ASF-YOLO^[10]进行改进,这对于大小、形状各异以及可能出现遮挡的密集船舶情况下的检测尤为重要.多尺度融合对于各种尺寸的船舶具有优秀的识别能力,能同时兼顾细节与全局信息。

3) 为了解决小型船舶及因远距离拍摄导致的小目标识别困难问题,在主干网络中引入 SimAM^[11]以增强特征提取的能力。

4) 进一步减少了计算量,同时大幅降低了模型参数量.本文使用剪枝算法使计算量降低至原来的 1/2,但带来了性能的略微下降;因此结合知识蒸馏技术,使模型性能进一步提升并超越剪枝前的水平,显著提高了模型的鲁棒性和检测

精度。

1 YOLOv8 算法简介

YOLOv8 是 Ultralytics 公司开源的一个 SOAT(state of the art)模型,它建立在先前成功版本的基础上,进一步提升了算法的性能和灵活性,设计了新的骨干网络,改进了新的 Anchor-Free 结构检测头以及新的损失函数.YOLOv8 算法包含输入层、骨干层(Backbone)、颈部层(Neck)和头部层(Head)4 个主要部分.骨干层用于提取图像的特征信息,在使用卷积的同时还采用了独有的模块 C2f 来提高网络的计算速度并降低了内存消耗;颈部层进一步增强骨干层提取的特征信息,采用 PANet 和 FPN(feature pyramid network)结构,既增强了多个尺度上的语义表达,又提升了多个尺度上的定位能力.头部层使用的是 Decoupled Head,包含 3 个输出分支,每个输出分支又分为两部分,分别进行目标分类和边界框回归。

2 改进的 YOLOv8 算法

本文改进了 YOLOv8 算法的内部模块和结构,引入了多尺度信息提取、特征融合和注意力机制等技术,改进后的网络结构如图 1 所示。

首先将拼接 P2 层和 ASF-YOLO 模块用于提取多尺度特征,之后改进 C2f 模块和 SimAM 模块以增强特征融合,从而优化模型在复杂海洋环境中的船舶目标检测性能.改进后的模块增大了计算量,因此采用剪枝操作对模型进行轻量化处理,随后使用知识蒸馏技术在不改变模型参数和计算量的前提下进一步提升模型性能.通过这些改进,期望提高模型对数据集中各类船舶的识别精确率,从而更好地满足实际应用中的船舶检测需求.本文的研究目标是船舶识别,实现对不同种类和大小的船舶进行边界框回归及二分类.由于数据集中包含各种类型和尺寸的船舶,若将它们统一归为一类会导致类别不平衡问题,进而影响模型的性能和泛化能力.因此,在进行二分类时,需采取相应策略以解决该问题。

2.1 CD3 模块构建

本文构建的 CD3(C2f DCNv3)模块通过将 DCNv3 模块融入 C2f 模块实现.该设计将 C2f 模块的主干替换为 DCNv3 模块.由于 DCNv3 对输入和输出进行等量分组处理,且组间互不干扰,

每组仅采用单一卷积核进行特征提取,该结构既降低了参数量,又引入了可变形卷积操作.这种操作允许网络根据目标对象的非均匀形状和空

间变形自适应调整卷积核采样位置,使模型能更有效地捕捉目标细节特征,尤其在复杂场景下可显著提升目标边界建模精度.

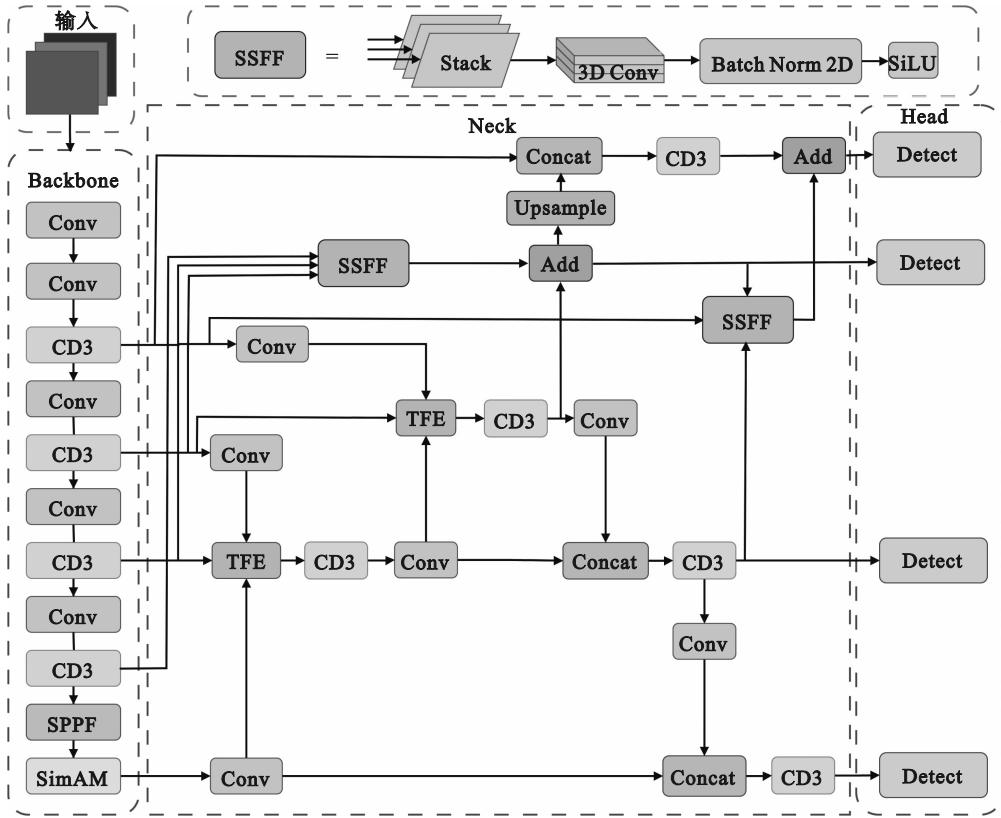


图 1 改进的 YOLOv8 网络结构

Fig. 1 Improved YOLOv8 network architecture

可变形卷积的原理^[12]是通过引入偏移量(offset)来调整卷积核在输入特征图上的采样位置.传统的卷积操作是固定的,每个卷积核在输入特征图上的采样点位置是固定的,对特征图上的每个位置 P_0 操作如式(1)所示:

$$y(P_0) = \sum_{P_n \in R} \omega(P_n) \cdot x(P_0 + P_n). \quad (1)$$

其中: $\omega(P_n)$ 为第 n 个位置 P_n 的卷积核; R 是卷积核的所有区域的集合; $x(P_0 + P_n)$ 为 $(P_0 + P_n)$ 位置特征图的值.

可变形卷积允许卷积核在每个位置上进行局部的微小偏移量 ΔP_n ,以适应目标对象的非均匀形状和空间变形,如式(2)所示:

$$y(P_0) = \sum_{P_n \in R} \omega(P_n) \cdot x(P_0 + P_n + \Delta P_n). \quad (2)$$

其中, $\{\Delta P_n | n=1, \dots, N\}$, $N=|R|$, 偏移量不会超过卷积核尺寸.这种偏移量是通过学习得到的,可以根据目标对象的特征来自适应地调整,从而更好地捕捉目标对象的细节特征.

为了弥补卷积与多头注意力之间的差距,在传统卷积运算中融入长距离依赖,如式(3)所示:

$$y(P_0) = \sum_{k=1}^K w_k m_k x(P_0 + P_k + \Delta P_k). \quad (3)$$

其中: K 表示采样点总数; w_k 表示第 k 个采样点对应的投影权重; m_k 表示第 k 个采样点的调制尺度,使用Sigmoid函数进行归一化.

DCNv3还在神经元之间共享了权值,将原始卷积中的 w_k 参照可分离卷积的思想分离成深度和点两部分,深度部分由 m_k 来调制,点的部分使用共享投影权重 w_k .同时引入了多组机制,将聚合空间分成 G 个组,每一组均配备专属的采样偏移量 ΔP_{gk} 与调制尺度 m_{gk} .同一卷积层内各组提供不同的空间聚合方式,增加了多样性.使用Softmax函数对沿着样本点的元素进行归一化,使得调制标量之和限制为1.如式(4)所示:

$$y(P_0) = \sum_{g=1}^G \sum_{k=1}^K w_g m_{gk} x_g(P_0 + P_k + \Delta P_{gk}). \quad (4)$$

其中, w_g 表示第 g 个采样点对应的投影权重.将DCNv3融合到C2f中并应用于全局,有效解决了标准卷积在网络中处理长距离依赖关系,显著提升了数据集中小目标的检测性能.该模块继承了卷积固有的归纳偏置属性,使得模型在有限训练

数据条件下能更高效地学习,降低对大规模数据的依赖,同时缩短训练周期,并减少了资源消耗.

2.2 嵌入 SimAM 注意力机制

SimAM 是一种轻量级、无参数的注意力机制.基于图像的局部自相似性,该机制通过计算特征图中每个像素与其相邻像素之间的相似性来生成注意力权重.

信息丰富的神经元往往以独特的放电模式区别于相邻神经元,并展现出空域抑制特性,即在激活时抑制周围神经元活动.因此,对于具备显著空域抑制效应的神经元应赋予较高重要性.

SimAM 定义了能量函数来度量目标神经元与相同通道内其他神经元之间的线性可分性.每层输入的参数包括特征图的输入通道数 C ,特征图的宽度 w ,高度 h ,对于每个通道会产生 $M=h \times w$ 个能量函数.通过采用二值标签并添加 L_2 正则项,最小能量的解析解如式(5)所示:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}. \quad (5)$$

其中: $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$, $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$ 分别为每个通道神经元的均值与方差; t 为该二维通道内空间位置为 (i, j) 的目标神经元的值; λ 为 L_2 正则化系数.每个神经元 t 与其“周围神经元”的关系主要体现在与均值 $\hat{\mu}$ 的差异上,式(5)中的核心项 $(t - \hat{\mu})^2$ 直接度量了 t 偏离该通道所有神经元均值 $\hat{\mu}$ 的程度. $|t - \hat{\mu}|$ 越大, t 距离 $\hat{\mu}$ 越远,式(5)计算出的能量值 e_i^* 就越小,这表明目标神经元 t 与该通道整体分布特性差异越显著,其重要性越高.这种设计使得 SimAM 能够有效捕捉在激活值空间上偏离常态的、具有区分性的信息.通过 $1/e_i^*$ 得到每个神经元的重要性之后,对特征进行增强处理,如式(6)所示:

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X. \quad (6)$$

其中: E 为所有通道内每个神经元 e_i^* 的均值; $\text{Sigmoid}(\cdot)$ 是一个单调递增函数,值域为 $0 \sim 1$,用于约束重要性 $1/e_i^*$ 的大小; \odot 表示哈达玛积; X 表示特征图的值.

SimAM 作为一种轻量级、无参数的注意力机制,通过利用图像的局部自相似性和空域抑制效应,保持了模型的轻量性和高效性.基于以上优点,本文在骨干网络的最后一层加入了 SimAM 模块,进一步强化了对信息丰富、模式独特区域的关注,为卷积神经网络带来了显著的性能

提升.

2.3 改进的 ASF-YOLO

ASF-YOLO 是一个基于 YOLO 的模型,专门用于解决不同类型细胞的多尺度问题和小细胞检测与分割难题.改进后的 ASF-YOLO 模块由尺度序列特征融合模块 SSFF(scale sequence feature fusion module)和三重特征编码器 TFE(triple feature encoder)两个主要模块组成,融合到颈部层中.该模型将初始模块中的通道和位置注意力机制替换为对应位置加和运算,并将原始的 3 个检测头增加到 4 个检测头输出,如图 1 所示.

SSFF 旨在通过创新的融合策略和结构,更高效地融合深层特征图携带的高阶抽象信息与浅层特征图蕴含的精细化细节.在这一过程中,尽管图像经下采样后尺寸有所减小,但所保持的尺度不变特征得以保留,确保了多尺度信息的有效传递.如式(7)所示:

$$\left. \begin{aligned} F_o(w, h) &= G_o(w, h) \times f(w, h), \\ G_o(w, h) &= \frac{1}{2\pi\sigma^2} e^{-\frac{(w^2+h^2)}{2\sigma^2}}. \end{aligned} \right\} \quad (7)$$

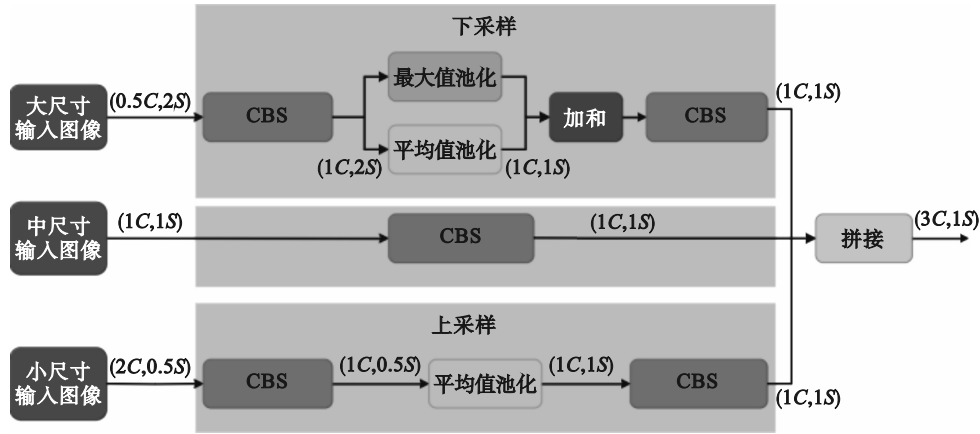
其中, $f(w, h)$ 是输入宽为 w 、高为 h 的图像数据,通过二维高斯滤波器 $G_o(w, h)$ 对图像平滑处理后,得到新的二维图像数据 $F_o(w, h)$.这些图像尽管具有相同的像素分辨率,但其内在的尺度特性各异.这样的差异使得各个特征图实质上构成了一个尺度谱系,其中每个特征图对应着图像的不同尺度表现.为了便于后续处理,将这些具有不同内在尺度特征但分辨率一致的特征图,通过适当的调整统一至相同分辨率,继而进行拼接操作.将不同尺度信息的特征图沿水平方向堆叠,运用三维卷积算法对这一堆叠结构进行深度学习,旨在从中提取蕴含尺度序列信息的综合特征.

TFE 模块设计的主要目的在于有效融合来自不同尺度的特征图,提升对尺寸差异显著目标的识别与定位精度,从而增强网络对小目标的检测能力.

图 2 展示了 TFE 的结构,其中 C 为通道数, S 为特征图的尺寸.以当前中尺寸特征图输入的通道数和尺寸为基准,对于大尺寸特征图,将其通道数缩减至 $1C$ (即与中尺寸输入的通道数基数相同),随后运用最大池化与平均池化相结合的混合下采样策略进行处理,该策略在降低分辨率的同时,仍保留高分辨率特征中的关键图像细节与多样性特征.对于小尺寸特征图,同样采用卷积模块对其通道数进行适应性调整(保持 $1C$ 的通

道数),之后使用最近邻插值方法进行上采样.最后对大、中、小3个尺寸的特征图各进行一次卷积操作,得到3幅具备不同尺度信息但尺寸和通道

数操作一致的特征图.最终在通道维度上进行拼接,形成融合多尺度信息的综合性特征图.



CBS表示一个由卷积层(convolution)、批归一化层(batch normalization)和SiLU激活函数组成的序列模块.

图2 TFE结构图

Fig. 2 TFE structure diagram

2.4 剪枝

神经网络剪枝技术旨在对网络架构进行精简优化,识别并剔除那些对模型性能贡献有限却占据大量计算资源的冗余元素.通过这一过程,能够在保持或近似保持模型原有预测精度的同时,显著降低其在计算、存储等方面的开销,从而使复杂模型适应更广泛的硬件环境,提升实际部署与运行的效率.

通过采用LAMP(layer-wise adaptive magnitude-based pruning)^[13]评分策略,基于逐层自适应幅度的剪枝方法实现更为精细化与高效的资源优化.幅度为权重绝对值的大小^[14-15].在神经网络架构中,相邻卷积层之间的连接均附带有权重参数,该参数本质上量化了输入信号对最终输出结果的影响程度.在实施剪枝操作时,对这些权重值进行评分以决定每个连接结构是否保留.当前广泛应用的一种评判标准是直接比较权重绝对值的大小.依据这一标准,被标记为剪枝目标的往往是那些权重绝对值较小的连接,因为它们通常被视为对网络整体功能贡献甚微的冗余成分^[16-17].LAMP作为一种先进的全局剪枝方法,克服了传统全局剪枝策略易引发的层崩溃问题^[18].其设计思路可概括如下:

将二维的权重张量展平为一维数组权重集合 W ,设 u, v 为该一维数组的两个索引, $w[u]$ 表示由索引 u 所对应的权重项.依据假设条件,权重项的数值大小与其索引位置呈有序关系,即当 $u < v$ 时,

$w[u] < w[v]$.在此前提下,LAMP评分如式(8)所示:

$$\text{score}(u; W) = \frac{(w[u])^2}{\sum_{v \geq u} (w[v])^2}. \quad (8)$$

其中: $(w[u])^2$ 表示目标连接对应权重的平方; $\sum_{v \geq u} (w[v])^2$ 表示该层在索引 u 之后的所有权重平方之和.当 $(w[u])^2 > (w[v])^2$ 时, $\text{score}(u; W) > \text{score}(v; W)$,该得分用于衡量目标权重 $w(u)$ 相对于其他权重的重要性.权重项的数值越大,其对应的LAMP评分值亦随之增大.评分较低的权重项被视作相对不重要,适合作为剪枝对象.此外,LAMP评分机制确保每层至少包含一个评分值为1的最优通道,从而有效规避层崩溃现象的发生,实现了全局剪枝与局部剪枝优势的有机整合.在设定期望的总体剪枝比例后,依据各连接的LAMP分数优先剪除分数最低的连接.

剪枝前后每层参数量对比如图3所示.剪枝后每层的参数量覆盖于基础模型之上.LAMP算法成功融合全局剪枝与局部剪枝的优势,通过动态、精细的评分系统指导剪枝过程,在满足资源约束条件的同时,确保模型保持优异的预测性能与结构合理性.

2.5 蒸馏

知识蒸馏是一种迁移学习算法,主要应用于深度学习模型的压缩与优化.通过将大型、复杂

且性能优异的教师模型的知识传递给小型、简洁且计算效率更高的学生模型,使学生模型在保持或接近教师模型预测能力的同时,具有更低的计算成本、更小的模型体积和更快的推断速度.根

据迁移策略的不同,知识蒸馏可分为两大类:一是基于输出概率分布的蒸馏(Logits 蒸馏法);二是基于中间层特征传递的蒸馏算法.本文同时采用这两种方法进行蒸馏,结构如图 4 所示.

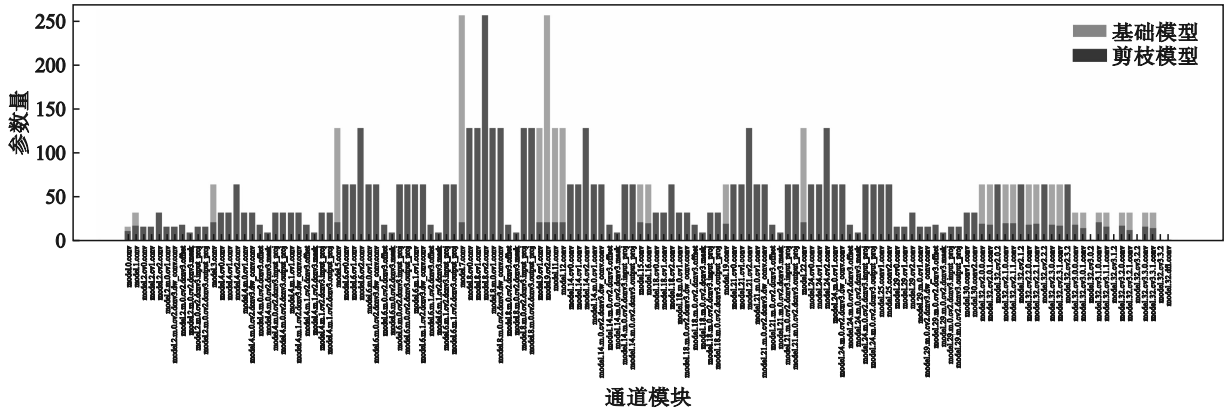


图 3 剪枝通道对比

Fig. 3 Comparison of pruning channels

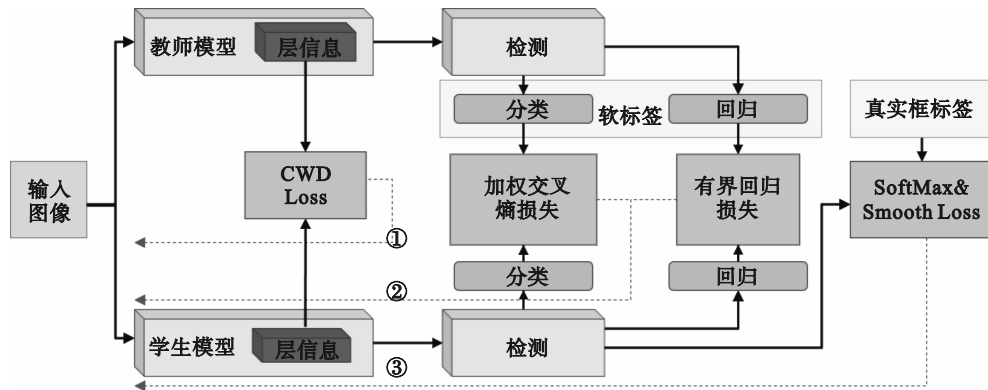


图 4 知识蒸馏结构图

Fig. 4 Structural diagram of knowledge distillation

在蒸馏过程中,学生模型一方面学习真实数据的硬标签,另一方面学习教师模型的软概率分布,从而获取教师模型对各类别置信度及其关系的细致知识.通过最小化学生模型与教师模型输出分布的KL(Kullback-Leibler)散度,精准复制教师模型的决策边界与类别区分能力.本文采用BCKD^[19]方法进行最终的训练学习,损失函数为

$$\mathcal{L}_{\text{BCKD}} = \alpha * \mathcal{L}_{\text{CE}}(s, y) + \beta * \mathcal{L}_{\text{KD}}(s, t). \quad (9)$$

其中: $\mathcal{L}_{\text{CE}}(s, y)$ 为针对分类任务的交叉熵损失函数,它衡量的是学生模型输出的预测对数概率分布与真实标签之间的差距,在图4中的③处, s 和 y 分别代表学生模型的原始输出和真实框标签; $\mathcal{L}_{\text{KD}}(s, t)$ 为基本知识蒸馏任务中的损失函数,在图4中的②处, t 表示教师模型的原始输出; α, β 为损失函数的权重.

基于中间层特征的蒸馏方法使学生模型能

够深入学习教师模型内部网络结构所蕴含的中间层特征表达.该方法强制要求学生模型的部分中间层产生与教师模型相应中间层输出尽可能接近的网络响应.本实验的中间层均选自头部层,采用CWD(channel-wise distillation)^[20]策略实现特征知识迁移,如式(10)所示:

$$\varphi(\phi(y_T), \phi(y_S)) = \varphi(\phi(y_{CT}), \phi(y_{CS})); \quad (10)$$

$$\phi(y_C) = \frac{\exp\left(\frac{y_{C,i}}{\mathcal{T}}\right)}{\sum_{i=1}^{w-h} \exp\left(\frac{y_{C,i}}{\mathcal{T}}\right)}. \quad (11)$$

其中: y_T 是教师模型的激活映射; y_S 是学生模型的激活映射; y_C 为特征图上中心点的特征向量; y_{CT} 和 y_{CS} 分别代表教师模型和学生模型在特征图上中心点的特征向量; $\phi(\cdot)$ 为使用Softmax归一化将激活映射值转换成概率分布,如式(11)所示; \mathcal{T} 为

超参数 $\phi(\cdot)$ 用于衡量教师模型和学生模型在每一个通道间的概率分布差异,采用 KL 散度这种非对称方式进行衡量,如式(12)所示:

$$\varphi(y_T, y_S) = \frac{\mathcal{T}^2}{C} \sum_{c=1}^C \sum_{i=1}^{w \cdot h} \phi(y_{T,c,i}) \cdot \log \left[\frac{\phi(y_{T,c,i})}{\phi(y_{S,c,i})} \right]. \quad (12)$$

式中: C 表示通道(channel); i 表示该通道中第 i 个像素的位置. \mathcal{T} 越大,每个通道关注的区域就会越大.当 $\phi(y_{T,c,i})$ 很大时,要增大 $\phi(y_{S,c,i})$ 才能最小化 KL 散度.

当同时采用 BCKD 和 CWD 两种蒸馏方法时,学生模型不仅能掌握教师模型在高层决策逻辑上的知识,还能学习其底层的特征提取机制,实现从局部到全局、从细节到整体知识的全面迁移,从而高效且精准地压缩和继承教师模型的知识.

3 实验及结果分析

3.1 数据集准备

本文基于阿里天池提供的复杂场景下的船舶检测数据集进行实验.数据集的收集涵盖了各种复杂的海上场景,包括船只密集的渔业区域,交通繁忙的港口以及船舶与岸边交通工具混合的场景.不仅包含船舶与海洋背景的交互,还囊括了岸线建筑、浮标、渔网、波浪、光照条件及天气等自然与人工干扰因素.船舶尺度具有显著变化,从小型渔船到大型邮轮,目标尺寸及拍摄距离差异悬殊;此外,部分图像中船舶密集分布,形成复杂的遮挡关系.与传统的船舶数据集相比,该数据集中存在大量船舶出现在图片的背景中,而非作为图像的主体部分存在,这给识别增加了巨大困难;同时考虑到实际中船舶并行行驶的情况,数据集还增加了船舶部分被遮挡和完全被遮挡的比例.此外,由于将不同种类的船视为同一类别,而不同类型船舶间的像素差异较大,进一步增加了检测难度.因此,本文数据集的复杂度较高,与一般数据集相比,这种多样化的环境使得识别任务困难许多.

在实验阶段,遵循了严格的在线数据增强策略以提高模型的泛化性能与适应性.具体手段包括随机像素内容变换、随机翻转、随机裁剪、随机填充等.这些操作能够在不增加额外数据收集成本的前提下,模拟出更为多样化的视觉输入,有效防止模型过拟合.特别地,采用了多尺度训练策略,将图像尺寸调整为 320 像素 \times 320 像素,

352 像素 \times 352 像素, 384 像素 \times 384 像素, 416 像素 \times 416 像素, 448 像素 \times 448 像素, 480 像素 \times 480 像素, 512 像素 \times 512 像素, 544 像素 \times 544 像素和 576 像素 \times 576 像素,确保模型能够应对不同分辨率下的船舶检测任务.

在验证与推理阶段,所有图像统一调整为 480 像素 \times 480 像素,作为算法输入的标准尺寸,以保证评估的公平性和一致性.在对比实验中,对于其他先进的船舶检测算法,其数据增强策略均与本文所采用的方法保持一致,以确保不同方法间性能比较的公正性.

3.2 评价指标

本文实验的数据集仅包含单一类别,因此采用 AP_{50} 、AP、精确率 P 、召回率 R 、参数量及计算量 (FLOPs) 作为模型评估指标, AP_{50} 是指交并比 (IoU) 阈值为 0.5 时计算的平均精度,该阈值通常被视为相对宽松的匹配标准,用于衡量模型在目标检测中是否达到基本定位准确度.而 AP 值的计算涵盖了 IoU 阈值从 0.50 到 0.95 的变化范围 (以 0.05 为间隔) 的多个阈值下的平均精度. AP 值越高,说明模型在各种 IoU 阈值下的检测精度都较高,具有更强的泛化能力.召回率指模型在真实存在的正类样本中成功找出并正确分类的比例.精确率指模型预测为正类且真实标签也是正类的比例,它衡量了模型在所有被标记为正类的预测结果中是正确结果的比例.此外,还对比了计算量和模型参数量,以体现在实际部署中的可行性.

3.3 消融实验

本文在复杂场景下的船舶检测数据集中以 YOLOv8 为基准模型进行消融实验,并在此基础上依次添加 CD3 模块、改进的 ASF-YOLO 结构、SimAM 注意力机制,随后依次进行剪枝和蒸馏处理.改进后的最终模型在测试集上计算得到的 AP_{50} 和 AP 值相比于基准模型 YOLOv8 分别提升了 4.7% 和 2.9%,召回率和精确率分别提升了 3.2% 和 4.2%,同时模型参数量减少了 56.1%,计算量降低了 30.5%.表 1 展示了各改进阶段在测试集上的性能指标.

模型 B 将基准模型 A 中的所有 C2f 模块均替换为 CD3 模块,在减少参数量的同时降低了计算量.深度可变形卷积通过将标准卷积分解为深度卷积和逐点卷积两部分,允许卷积核在常规网格采样基础上进行灵活偏移,从而更好地适应输入特征图中的非刚性形变.相较于传统卷积操作,

可变形卷积在保持对目标形状变化敏感性的同时,通过动态调整滤波器的采样位置,减少了对大量参数的依赖.该改进不仅显著降低了参数量与计算量,而且在保持甚至提升模型检测性能的

基础上,增强了模型对复杂场景及非规则形状目标的适应性.这一改进有效平衡了模型效率与精度之间的关系.

表 1 消融实验结果
Table 1 Results of ablation experiment

模型	基准模型及其改进	AP ₅₀	AP	精确率	召回率	计算量/GFLOPs	参数量×10 ⁻⁶
A	YOLOv8	0.575 14	0.335 09	0.682 93	0.509 46	8.2	3.011 1
B	A+CD3	0.582 01	0.337 91	0.661 61	0.517 03	7.4	2.668 9
C	B+(ASF+检测头)	0.603 01	0.349 79	0.685 91	0.538 99	11.4	2.234 9
D	C+SimAM	0.609 97	0.362 16	0.720 73	0.523 11	11.4	2.234 9
E	D+剪枝	0.604 83	0.349 49	0.708 05	0.508 71	5.7	1.320 9
F	E+蒸馏	0.622 36	0.363 86	0.725 06	0.541 26	5.7	1.320 9

模型 C 在模型 B 的基础上,将颈部层替换为 ASF 结构,并额外增加了一个检测头.该改进增加了计算量但减少了参数量,同时使其指标均得到提升,其中召回率和 AP₅₀ 性能的提升尤为显著.召回率作为衡量模型检测所有正类样本能力的指标,其提升表明模型在复杂场景下能更有效地识别和捕获目标对象,从而减少漏检现象.ASF 结构通过其独特的注意力机制使模型能更精准地聚焦于目标区域的关键特征,从而使预测边界框更贴近真实标注值,即提高 IoU 值.新增的检测头进一步增强了模型对不同尺度和不同姿态目标的捕捉能力,特别是对小目标或遮挡严重的物体,其检测效果得到显著改善.这些改进共同推动了整体召回率的大幅提升.

模型 D 在模型 C 基础上仅增加了一个 SimAM 模块.由于该注意力机制不含可训练参数,且计算复杂度低,其计算开销与基准模型相比几乎保持不变,各项性能指标均获得显著提升,其中 AP 和分类精确率的改善尤为突出.这表明,SimAM 能有效增强模型对目标位置信息的敏感性以及类别特征的判别能力,使其在不同尺度、姿态和遮挡条件下的目标检测任务中均能保持较高的精度,从而全面提升模型的平均检测性能.

模型 E 是在模型 D 基础上进行剪枝操作,计算量减少至原来的 1/2,参数量变为剪枝前的 59.1%.在 Finetune 时不更改任何训练策略,其他各项指标仅小幅度下降,充分体现了本实验采用的剪枝方法的有效性和稳健性,能够在大幅降低模型复杂度的同时,最大限度地保留模型的核心学习能力和预测精度.未来的研究可进一步探索优化 Finetuning 策略或其他模型压缩技术,以期在保证性能的前提下,实现模型 E 的进一步精简与优化.

模型 F 是对模型 E 进行蒸馏得到的,教师模型仅将模型 D 的基准模型 YOLOv8n 替换为 YOLOv8s,教师模型的 AP₅₀ 为 0.626 53,AP 为 0.375 23,召回率为 0.569 27,精确率为 0.666 26,计算量达到 32.7 GFLOPs,参数量为 7.970 5×10⁶,而模型 F 在计算量仅需 5.7 GFLOPs 和参数量为 1.320 9×10⁶ 的情况下,取得了接近教师模型的性能表现,其中精确率甚至超过教师模型.与模型 E 相比,模型 F 各项指标均获得显著提升.

图 5 和图 6 分别展示了模型 A 与模型 F 的部分检测效果,图片均来自本文所用数据集的测试集.在轻微干扰情况下,模型 A 存在较明显的错检问题(如图 5a 所示),而模型 F 较好地克服了这一缺点(图 6a).在有遮挡情况下,无论是图 5b 所示的非船舶轻微遮挡,还是图 5d 所示的同类型船舶混淆干扰且遮挡面积较大的情况,模型 A 均未识别出目标,而模型 F 表现出较优的性能.在远距离、小目标情况下,模型 A 存在较明显的漏检问题(如图 5c 所示).在背景复杂、干扰较多的情况下(如图 5e、图 6e 所示),模型 F 能更好地检测出船舶位置,而模型 A 错检较多.在远距离且船舶密集排列的情况下(如图 5f、图 6f 所示),模型 A 存在明显漏检问题,而模型 F 能较好地检测出船舶位置.

综上,说明改进后的模型鲁棒性更强,检测效果更佳,充分表明了改进方法的有效性.

3.4 与其他 SOTA 检测器对比实验

为了更严谨且全面地验证本文所提算法的优越性,将该算法与目前最先进的若干目标检测算法进行了详尽的对比分析.数据集采用阿里天池复杂场景下的船舶检测数据集,训练和推理环境均为 Ubuntu20.04 系统、NVIDIA GeForce

RTX3060 GPU、CUDA 11.7、cuDNN 8.9.4、Python 3.8.18 和 PyTorch 1.13.1. 实验结果如表 2 所示.

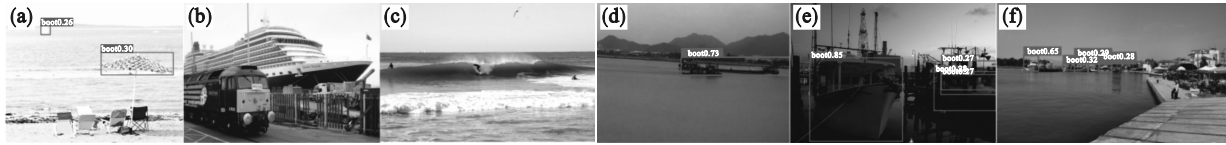


图 5 模型 A 的部分检测结果

Fig. 5 Partial detection results of model A

(a)一轻微干扰;(b)一非船舶轻微遮挡;(c)一远距离、小目标;(d)一同类型船舶混淆干扰且遮挡面积较大;
(e)一背景复杂、干扰较多;(f)一远距离且船舶密集排列.

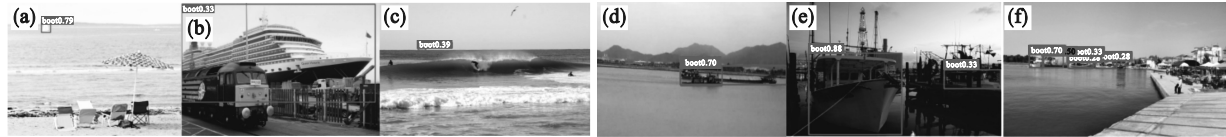


图 6 模型 F 的部分检测结果

Fig. 6 Partial detection results of model F

(a)一轻微干扰;(b)一非船舶轻微遮挡;(c)一远距离、小目标;(d)一同类型船舶混淆干扰且遮挡面积较大;
(e)一背景复杂、干扰较多;(f)一远距离且船舶密集排列.

表 2 与其他 SOTA 检测器对比实验结果

Table 2 Comparison of experimental results with other SOTA detectors

模型	AP ₅₀	AP	精确率	召回率	计算量/GFLOPs	参数量×10 ⁻⁶
YOLOv5-n	0.565	0.324	0.663	0.504	7.2	2.51
YOLOv5-m	0.592	0.342	0.669	0.533	24.0	9.12
YOLOv6-n	0.577	0.337	0.708	0.481	11.9	4.24
YOLOv7-tiny	0.605	0.324	0.595	0.583	13.2	6.01
gold_yolo-n	0.587	0.335	0.699	0.518	12.1	5.60
ResNet50-FasterRCNN	0.548	0.298	0.543	0.387	251.4	41.35
SSD	0.521	0.281	0.643	0.558	122.3	48.31
CCAT-YOLOv8n	0.581	0.338	0.651	0.518	15.9	5.72
YOLOv8n-Bi-FPN-EMA	0.587	0.336	0.687	0.528	7.8	2.52
模型 F(本文模型)	0.622	0.364	0.725	0.541	5.7	1.32

从对比实验中可以看出,本文改进的最终模型 F 的 AP₅₀, AP 和精确率均明显高于其他先进模型,并且召回率也保持较高的水平,而计算量和参数量明显低于其他模型.对于 AP₅₀这一指标,YOLOv7-tiny 最接近本文模型,其召回率在对比实验中最高,但计算量和参数量远高于本文模型.对于 AP 和精确率这 2 个指标,表现较好的是 YOLOv6-n 和 gold_yolo-n,但这 2 个模型的召回率表现较差,且计算量和参数量较大.而在计算量和参数量这 2 个指标中,只有 YOLOv5-n 最接近本文模型,但该模型其余各项指标均表现较差,其中 AP₅₀ 和召回率比本文模型分别低了 5.7% 和 6.2%. CCAT-YOLOv8n^[21] 和 YOLOv8n-Bi-FPN-EMA^[22] 都是在船舶检测领域对 YOLOv8n 的改进算法,按照文中结构进行复现并在本文所使用的环境中进行训练.综上,在复杂船舶目标

检测中,与表 2 中其他 SOTA 检测器相比,本文改进的模型计算量和参数量最低(即模型体积最小),对设备的负载最低;AP₅₀ 与 AP 指标最高,同时精确率最高,减少了错检的发生.

4 结 论

1) 本文在 YOLOv8 的基础上进行改进,在主干网络中引入的 DCNv3 卷积能更好地适应复杂海况(如波浪遮挡、姿态变化)以及不同船型(如游艇、货轮、帆船等)带来的目标几何形变和局部视角变化.

2) 嵌入骨干层的 SimAM 无参数注意力机制能自主学习并强化与船舶目标强相关的空间/通道特征,有效抑制岸线/岛屿背景混淆等高噪声信息的干扰.

3) 颈部集成的ASF模块通过自适应的多尺度特征融合与注意力加权机制,显著提升了模型对部分遮挡船舶(如船身被其他船只遮挡)以及远距离小尺度目标(如海上漂浮的小型救生艇或渔船)的识别能力。

4) 结合剪枝和蒸馏的优化策略,在压缩模型规模和计算量的同时,确保模型在计算资源受限的实时检测场景中仍能维持高性能的复杂背景判别能力。

参考文献:

- [1] 鄢然,王帅安,周煜圣.区块链技术在航运业的应用综述[J].交通工程与信息学报,2022,20(3):1-14.
(Yan Ran, Wang Shuai-an, Zhou Yu-sheng. Application of blockchain technology in the shipping industry[J]. *Journal of Transportation Engineering and Information*, 2022, 20(3):1-14.)
- [2] 周旗开,张伟,李东锦,等.基于改进YOLOv5s的光学遥感图像舰船分类检测方法[J].激光与光电子学进展,2022,59(16):1628008.
(Zhou Qi-kai, Zhang Wei, Li Dong-jin, et al. Ship detection and classification method for optical remote sensing images based on improved YOLOv5s[J]. *Laser & Optoelectronics Progress*, 2022, 59(16): 1628008.)
- [3] Li Y D, Zhang S S, Wang W Q. A lightweight faster R-CNN for ship detection in SAR images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 4006105.
- [4] Jiang Z K, Su L, Sun Y X. YOLOv7-ship: a lightweight algorithm for ship object detection in complex marine environments [J]. *Journal of Marine Science and Engineering*, 2024, 12(1): 179-190.
- [5] Zhao Q C, Wu Y Q, Yuan Y B. Ship target detection in optical remote sensing images based on E2YOLOX-VFL [J]. *Remote Sensing*, 2024, 16(2): 318-340.
- [6] Lyu Z L, Wang C Y, Sun X J, et al. Real-time ship detection system for wave glider based on YOLOv5s-lite-CBAM model [J]. *Applied Ocean Research*, 2024, 144: 103833.
- [7] Zhao K, Liu R T, Wang S Y, et al. ST-YOLOA: a Swin-Transformer-based YOLO model with an attention mechanism for SAR ship detection under complex background [J]. *Frontiers in Neurorobotics*, 2023, 17: 1170163.
- [8] Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO (version 8.1.9) [EB/OL]. (2023-01-10) [2024-02-24]. <https://github.com/ultralytics/ultralytics>.
- [9] Wang W H, Dai J F, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 14408-14419.
- [10] Kang M, Ting C M, Ting F F, et al. ASF-YOLO: a novel YOLO model with attentional scale sequence fusion for cell instance segmentation [J]. *Image and Vision Computing*, 2024, 147: 105057-105096.
- [11] 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. Online: PMLR, 2021: 11863-11874.
- [12] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNetsv2: more deformable, better results [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 9300-9308.
- [13] Lee J, Park S, Mo S, et al. Layer-adaptive sparsity for the magnitude-based pruning [EB/OL]. (2021-05-09) [2024-02-26]. <https://arxiv.org/abs/2010.07611>.
- [14] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient ConvNets [EB/OL]. (2017-03-17) [2024-02-26]. <https://arxiv.org/abs/1608.08710>.
- [15] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 1314-1324.
- [16] Blalock D, Ortiz J J G, Frankle J, et al. What is the state of neural network pruning? [EB/OL]. (2020-03-06) [2024-02-26]. <https://arxiv.org/abs/2003.03033>.
- [17] Molchanov P, Tyree S, Karras T, et al. Pruning convolutional neural networks for resource efficient inference [EB/OL]. (2017-06-08) [2024-02-26]. <https://arxiv.org/abs/1611.06440>.
- [18] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2024-02-26]. <https://arxiv.org/abs/1704.04861>.
- [19] Wang Q, Liu L, Yu W X, et al. BCKD: block-correlation knowledge distillation [C]//2023 IEEE International Conference on Image Processing (ICIP). Kuala Lumpur: IEEE, 2023: 3225-3229.
- [20] Shu C Y, Liu Y F, Gao J F, et al. Channel-wise knowledge distillation for dense prediction [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 5291-5300.
- [21] 李斌,雷钧涵,郭毅.基于CCA和Transformer的YOLOv8船舶目标检测算法[J].控制工程,2024,31(5):901-911.
(Li Bin, Lei Jun-han, Guo Yi. YOLOv8 ship target detection algorithm based on CCA and Transformer [J]. *Control Engineering of China*, 2024, 31(5): 901-911.)
- [22] 惠卓凡,李鹏龙,沈烈,等.基于改进YOLOv8的渔港船舶进出港目标检测与统计方法[J].大连海洋大学学报,2024,39(3):498-505.
(Hui Zhuo-fan, Li Peng-long, Shen Lie, et al. Detection and statistics method of ship entry and exit in a fishing port based on improved YOLOv8 [J]. *Journal of Dalian Ocean University*, 2024, 39(3):498-505.)