

内河船舶检测算法的轻量化改进^{*}董健^{1,2} 赵欣² 毋路遥^{1,2} 吴凯丽^{1,2} 司福祺²

(1. 中国科学技术大学研究生院科学岛分院 合肥 230026; 2. 中国科学院合肥物质科学研究院 合肥 230031)

摘要: 为了解决现有的船舶检测算法参数量与计算量较大,以及尺度和视角变化导致检测结果波动的问题,提出一种改进 YOLOv8n 的轻量化内河船舶检测算法 YOLO-LISD。首先,设计融合细节增强卷积的高效特征共享检测头替换原检测头,提升检测一致性;其次,引入 slim-neck 方法改进颈部网络,在保证检测性能前提下缩小模型体积;再次,提出全局通道自适应幅度剪枝算法深度压缩模型,提升检测效率;最后,设计基于空间和通道相关性的特征知识蒸馏,提高剪枝后模型的检测精度。实验结果表明,YOLO-LISD 相比 YOLOv8n 模型参数量与计算量分别减少 68.4% 与 56.8%,在 SeaShips 数据集上检测的准确率与 mAP_{50,95} 分别提高 1.1% 和 2.1%。实际应用中,在低算力设备上检测速度达到 55 fps,满足实时性要求。与其他算法对比展现明显优势,验证了该方法的优越性。

关键词: 船舶检测;YOLOv8n;轻量化;通道剪枝;知识蒸馏

中图分类号: TP391;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Lightweight improvement of inland ship detection algorithm

Dong Jian^{1,2} Zhao Xin² Wu Luyao^{1,2} Wu Kaili^{1,2} Si Fuqi²

(1. Science Island Branch, Graduate School of University of Science and Technology of China, Hefei 230026, China;

2. Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China)

Abstract: To address the challenges of large parameter sizes and computational demands in existing ship detection algorithms, as well as the fluctuations in detection results caused by scale and perspective variations, we propose an improved lightweight inland ship detection algorithm, YOLO-LISD, based on YOLOv8n. First, an efficient feature-sharing detection head, incorporating detail-enhanced convolution, is introduced to replace the original detection head, improving detection consistency. Second, a slim-neck method is incorporated to optimize the neck network, reducing the model size while maintaining detection performance. Third, a global channel-adaptive magnitude-based pruning algorithm is proposed for depth compression, enhancing detection efficiency. Finally, a feature knowledge distillation approach, leveraging spatial and channel correlations, is designed to improve the detection accuracy of the pruned model. Experimental results demonstrate that, compared to YOLOv8n, YOLO-LISD reduces the number of parameters and computational complexity by 68.4% and 56.8%, respectively, while improving detection accuracy and mAP_{50,95} on the SeaShips dataset by 1.1% and 2.1%, respectively. In practical applications, the detection speed of low computing power equipment reaches 55 fps, meeting real-time requirements. Compared to other algorithms, it demonstrates significant advantages, validating the superiority of the proposed method.

Keywords: ship detection; YOLOv8n; lightweight; channel pruning; knowledge distillation

0 引言

内河航运凭借其单位运输成本低、运载能力大等技术优势,已成为现代物流体系中的重要运输方式。然而,随着航运需求的指数级增长,船舶密度持续攀升导致航道事故频发,同时部分船舶的非规范排放对流域生态环境造成显

著威胁。基于计算机视觉的内河船舶智能检测技术,对实现航行安全监管、污染排放监控以及航运系统智能化转型具有重要战略价值。

当前船舶检测研究主要基于卷积神经网络架构展开。刘芬等^[1]通过中位数优化 K-means 聚类算法改进 YOLOv5 锚框生成机制,结合 MIoU 损失函数优化边界框回归,有

效缓解复杂海天背景下的误检漏检问题;罗芳等^[2]提出 SAF-YOLOX 算法,利用双向特征金字塔与多路并行检测头增强多尺度目标检测能力,但引入额外分支导致计算量增加;Yang 等^[3]提出了一种分裂双向特征金字塔网络的特征融合模块,以减轻复杂背景对 SAR 船舶图像的干扰。上述改进方案虽提升检测性能,但均以增加模型复杂度为代价。

在船舶检测算法轻量化技术研究领域,谢友晨等^[4]基于 YOLOv5s 开发 YOLO-WIT 算法,通过复合距离损失函数提升弱小目标定位精度,优化削减检测头使得模型参数量减少 31.6%。张炳焱等^[5]采用 FasterNext 模块替换 YOLOv5 的 C3 模块,结合 NAM 注意力机制实现参数量减少 9.74%。张桑等^[6]以 MobileNetV3 重构 YOLOv7 主干网络,引入 WiseIoU 损失函数,在保证精度的同时降低 38.4% 参数量。Zhao 等^[7]基于 ShuffleNetv2 与 Ghost 卷积优化 YOLOv5,移除冗余检测头后模型尺寸减小 70.3%。上述改进方案因基础网络架构的特征表达能力局限,改进策略虽能实现部分性能指标优化,但在综合检测性能方面仍与前沿模型存在显著差距。

YOLOv8 凭借其卓越的检测精度与推理效率,自发布后迅速成为目标检测领域的研究热点。研究者通过直接轻量化设计与模型压缩等方法,推动模型轻量化研究取得显著进展。王海群等^[8]通过引入多尺度通道注意力 MCA 机制强化主干网络特征提取能力,结合检测头重参数化设计与双向特征金字塔网络 BiFPN 优化颈部结构,在模型参数量压缩 32.6% 的同时实现 mAP 值 2.5% 的提升。朱圣博等^[9]提出的融合多尺度注意力机制的 C2f-FE 模块,通过动态检测头实现多层级特征自适应融合,在参数减少 42.3% 的情况下增强小目标检测灵敏度。杨志渊等^[10]采用非对称检测头与 slim-FPN 协同优化策略,设计快速卷积模块 FasterConv 重构 Bottleneck 结构,实现 33% 参数量削减与检测精度 1.4% 提升的同时,将推理速度提高 8%,验证了结构简化对实时性的正向作用。

尽管当前 YOLOv8 轻量化研究虽取得显著进展,系统性分析仍揭示出三方面共性技术挑战:首先,多尺度特征交互不足制约目标检测性能;其次,注意力机制的引入虽能提升特征提取与表达能力,但导致推理延迟增加,形成特征增强与计算效率间的矛盾;再者,针对特定场景设计的轻量化方案严重削弱模型泛化能力。针对上述关键技术挑战,本文提出轻量化算法 YOLO-LISD (YOLO-lightweight inland ships detector),聚焦内河船舶检测场景,目标构建兼顾精度、效率与泛化性的轻量化框架。本研究工作包含以下创新:

1) 提出一种轻量化的高效特征共享检测头 EFSDH,通过构建参数共享的跨尺度交互通路,摒弃传统注意力机制,而是结合细节增强卷积强化特征提取能力,实现多尺度特征的语义一致性表达与计算效率的协同优化。

2) 引入 slim-neck 方法,使用 GSConv 与 VoV-GSCSP 模块分别替换颈部网络中的标准卷积与 c2f 模块,减轻网络的计算负担,并提高检测准确率。

3) 提出一种基于 GCAMBP 分数的通道剪枝算法,全局动态量化通道的重要性来精确移除冗余通道,在确保关键语义信息完整性的前提下,构建高压缩比的轻量化模型架构,进一步提升了推理速度。

4) 提出一种基于空间-通道间相关性损失函数的 SCKD 知识蒸馏方法,指导学生模型学习教师模型特征图中空间信息和通道信息的完整性关系,提升剪枝后模型检测性能。先剪枝后蒸馏的协同范式具有架构普适性,为其他任务场景的模型轻量化提供通用解决方案。

1 YOLO-LISD 算法设计

轻量化设计是在保证模型检测精度的前提下,减少模型参数的数量和计算资源的消耗。本文采用多种轻量化策略优化模型,旨在提高检测性能并降低计算复杂度。轻量化研究主要分为两个方向^[11]: 1) 直接设计轻量化的网络^[12]; 2) 对训练好的模型进行压缩。为了获得总体性能更优的模型,本文基于 YOLOv8n 网络模型进行改进,融合上述两种轻量化方法,设计了轻量级内河船舶检测算法 YOLO-LISD,其网络结构如图 1 所示。

1.1 直接轻量化网络设计

1) 轻量化检测头 EFSDH

YOLOv8n 原始解耦检测头设计存在一些局限性。首先,模型对不同尺度的特征采用单独的检测头处理,未能高效共享多尺度特征信息。其次,解耦检测头使用独立的分支处理分类和回归任务,增加了计算负荷。针对上述技术局限,近期研究提出若干改进方案:杨文等^[13]提出动态注意力检测机制,通过嵌套式注意力模块实现多尺度特征的自适应融合,但引入的结构复杂性导致模型复杂度提升。胡玮等^[14]设计多尺度融合检测头,采用跨层级特征重组路径与轻量化卷积优化特征交互过程,然而多路径架构引发额外计算冗余。赵露强等^[15]在检测头前引入自适应特征融合模块,通过动态权重分配平衡语义抽象与细节特征表达,但模块堆叠造成参数与计算量倍增。

为了改进这些不足,本文设计了一种融合细节增强卷积^[16] (detail-enhanced convolution, DEConv) 的高效特征共享检测头 (efficient feature sharing detection head, EFSDH),其结构如图 2 所示。其核心思想是将细节增强卷积作为共享特征卷积,使不同检测头共用相同的卷积核参数。这一设计灵感源于观察到在不同特征层次上检测到的物体在比例尺度上具有一定的相似性,且细节增强卷积有效地捕捉特征图中不同方向和尺度的细节信息。

EFSDH 首先通过一个独立的 1×1 卷积完成通道对齐,然后使用两次共享的 3×3 细节增强卷积实现不同尺度特征图共享相同的权值。特征共享机制保持一致的特征提

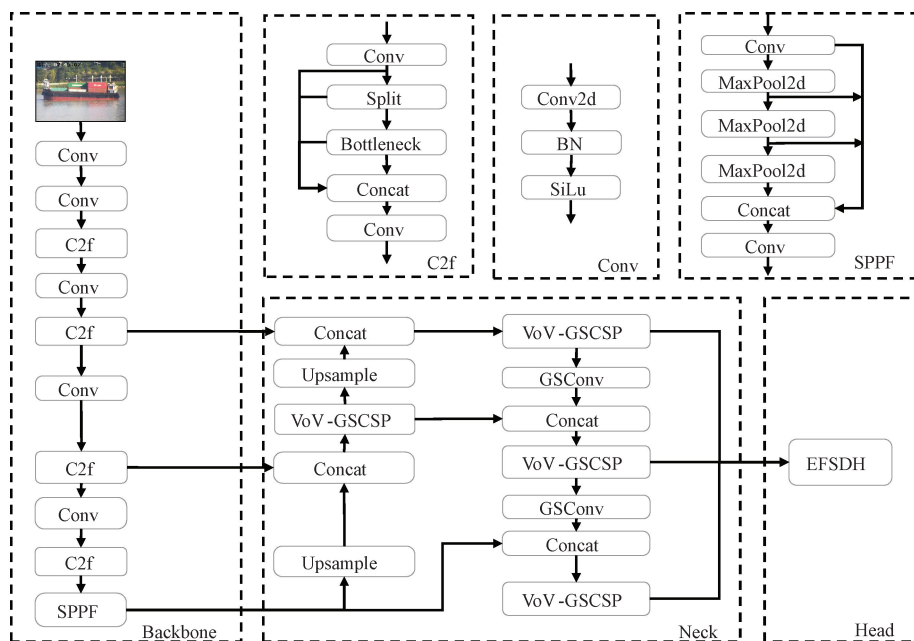


图 1 YOLO-LISD 网络结构

Fig. 1 YOLO-LISD network architecture

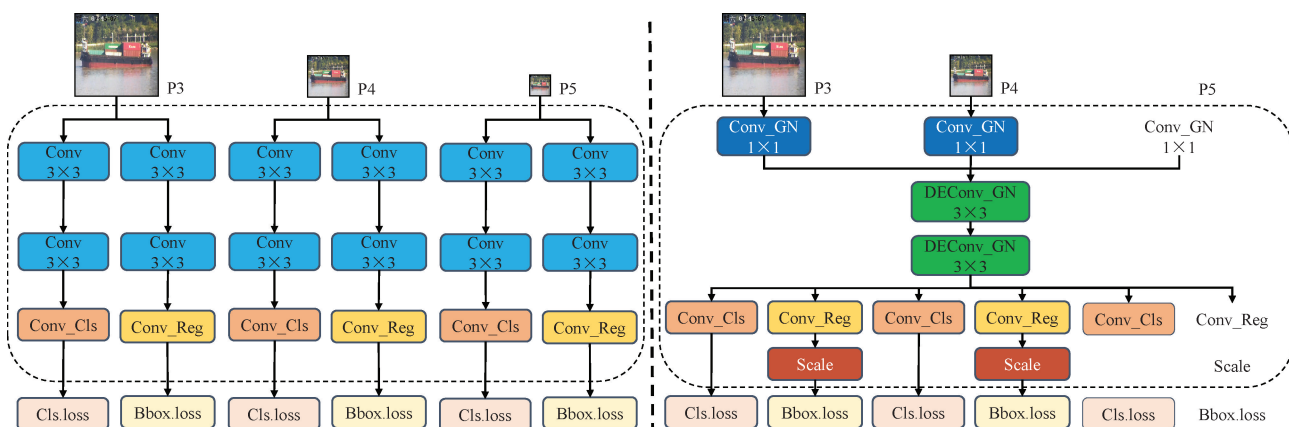


图 2 YOLOv8 解耦检测头(左)与 EFSDH 模块(右)结构

Fig. 2 YOLOv8 decoupled detection head (left) and EFSDH module (right) structure

取模式,不仅提升对不同大小和视角的船舶目标的鲁棒性和检测精度,而且显著减少参数量。此外,为适配不同大小目标,在回归任务中增加了 Scale 层动态调整特征图尺度,进一步增强对各种尺寸目标的检测能力。与 YOLOv8n 检测头相比,EFSDH 将参数量减少至原始解耦检测头的 51%,使模型更容易在计算资源受限的系统上应用。

DEConv 作为 EFSDH 的核心模块,进一步提升了特征表达能力。DEConv 将中心差分卷积、水平差分卷积、垂直差分卷积和方向增强卷积运算集成于一体,采用并行运算捕获特征图中不同方向的细节信息,其结构图如图 3 所示。由于不同特征层的分布统计特性存在显著差异,DEConv 中的归一化层若采用批归一化(batch normalization, BN),可能因分布不一致引发归一化偏差,尤其在小批量训练或

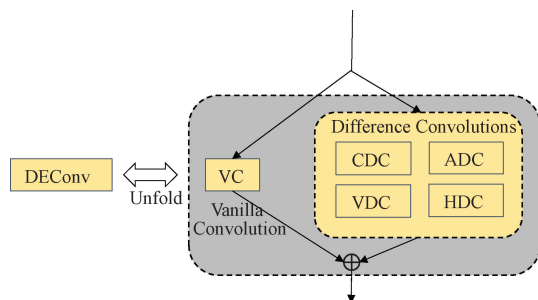


图 3 DEConv 模块结构

Fig. 3 DEConv module structure

推理阶段表现不稳定。为解决这一问题,DEConv 采用组归一化(group normalization, GN)替代 BN。

2) 颈部网络轻量化 slim-neck

slim-neck^[17]是一种轻量化目标检测模型颈部网络的解决方案,最初提出时应用在自动驾驶领域。组可分离卷积(grouped and shuffle convolution, GSConv)作为 slim-neck 方法的核心模块,其通过结构优化实现高效卷积操作,实现在提升了特征提取能力的同时,显著减少参数量与计算量。本文使用 GSConv 及以其为基础构建的 GSBottleneck 和 VoV-GSCSP 模块,分别替换颈部网络中的 Conv、Bottleneck 和 C2f 模块。这 3 种模块协同作用,构建了一个高效的特征提取与融合框架。GSConv、GSBottleneck 和 VoV-GSCSP 如图 4 所示。

GSConv 首先使用标准卷积对输入进行提取特征,然

后使用深度可分离卷积对标准卷积输出进行细粒度特征提取,接着将这两种不同类型卷积的输出拼接,最终使用通道混洗操作得到输出特征。由于深度可分离卷积在处理时无需对每个输出通道执行大量跨通道卷积,相比标准卷积显著减少了参数量和计算量,尤其在输入和输出通道都较多的颈部网络中使用,优势更为明显。在模型逐步将输入图片的空间信息转移至通道时,通道混洗均匀地交换不同通道上的局部特征信息,最大限度减少了空间维度压缩和通道维度扩展导致的语义信息部分丢失。值得注意的是,GSConv 仅使用标准卷积完成一半输出通道数量的特征提取,另一半由深度可分离卷积完成。这种设计保留了丰富特征信息,计算成本也只需标准卷积的 50% 左右。

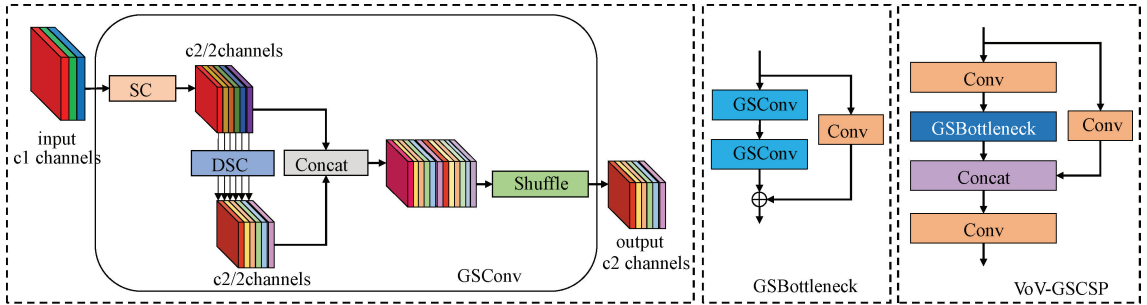


图 4 GSConv、GSBottleneck 和 VoV-GSCSP 模块
Fig. 4 GSConv, GSBottleneck, and VoV-GSCSP modules

GSConv 模块和残差连接策略结合得到 GSBottleneck 模块,然后 GSBottleneck 融合跨阶段部分网络结构设计策略得到 VoV-GSCSP 模块。VoV-GSCSP 将输入特征分成两条并行分支处理,并在后续阶段进行拼接,最终通过标准卷积进一步优化特征表示。这种设计使模型在减少计算资源需求的同时,增强了特征的流动性,能够保持出色的特征提取与融合能力。

1.2 模型压缩

1) 全局通道自适应幅度剪枝

模型剪枝旨在尽可能保持模型精度的条件下,通过减少模型中不重要的部分降低计算复杂度和参数量,实现模型轻量化。文献[18]提出一种层级自适应幅度剪枝算法,通过计算每层单个权重平方值并归一化衡量重要性,然后对分数全局排序,剪除低分的单个权重,实现非结构化剪枝。由于其剪枝单位是单个权重,权重矩阵将变得稀疏,对硬件的适配性变差。文献[19]使用的 slim 剪枝方法逐层独立剪枝,没有充分考虑层之间的依赖关系,导致部分关键层被过度剪枝,影响模型性能。

针对传统通道剪枝方法的硬件兼容性与全局相关性建模的不足,本文通过建立通道重要性评估的全局视角,提出一种全局通道自适应幅度剪枝算法(global channel-adaptive magnitude-based pruning, GCAMPB),能够适配当前硬件设备的加速机制,其流程如图 5 所示。

GCAMPB 是一种由通道重要性分数指导的结构化剪

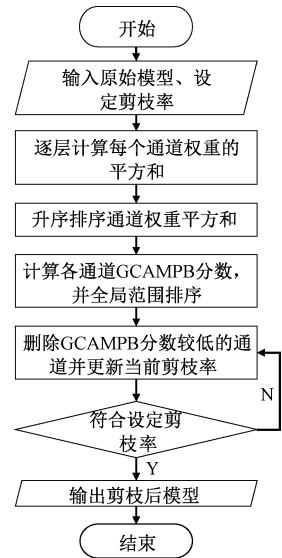


图 5 模型剪枝流程图
Fig. 5 Model pruning flow chart

枝算法,其核心思想是通过动态归一化机制计算通道重要性分数并参与全局排序,达到综合平衡不同层的剪枝力度,并通过剪除全局排序中较低分数的通道实现模型压缩。

假设当前卷积层的权重矩阵为 $W_l \in \mathbf{R}^{C_{out} \times C_{in} \times k \times k}$,其中 C_{in} 和 C_{out} 分别表示输入通道数和输出通道数, k 为卷积核大小。GCAMPB 首先计算每个通道的权重平方和:

$$S_{l,j} = \sum_{c=1}^{C_{in}} \sum_{u=1}^k \sum_{v=1}^k |W_l[j, c, u, v]|^2 \quad (1)$$

式中: $S_{l,j}$ 表示第 l 层第 j 个输出通道的权重平方和。为了更准确地反映通道在层内的相对重要性, 首先将通道权重平方和升序排序, 然后对每个通道 k 定义其 GCAMPB 分数:

$$score(k, S_l) = \frac{S_l[k]}{\sum_{j \geq k} S_l[j]} \quad (2)$$

式中: S_l 表示通道权重平方和张量, $S_l[k]$ 表示索引 k 对应的通道权重平方和。假设权重平方和根据给定的索引按升序排序, 即当 $u \leq v$ 时, $S_l[u] \leq S_l[v]$ 。动态归一化机制的核心在于, 当前通道的分母仅包含比其重要性更高或相等的通道, 使分数更加精确地刻画了通道的重要性排名。

完成所有层的层内通道重要性分数计算并汇总, 剪除全局分数中得分最低的通道, 直至满足预设的剪枝率, 并同步调整网络后续层的输入通道。剪枝完成后, 通过微调恢复模型性能。

2) 空间-通道相关性知识蒸馏

知识蒸馏作为一种高效的模型压缩技术, 其核心在于将大型教师模型的知识传递给轻量的学生模型, 使得学生模型在结构较为简单时, 依然能够达到较高水准的检测精度。文献[20]提出保持特征间通道相关性的重要性, 并将其应用于特征知识蒸馏。然而仅考虑特征图在通道维度上的相关性, 这在需要空间感知的视觉任务中可能存在局限性。

本文在通道相关性特征知识蒸馏的基础上, 提出一种空间-通道相关性知识蒸馏方法 (spatial-channel correlation knowledge distillation, SCCKD), 通过通道与空间的双重建模机制, 使得学生模型学习到教师模型特征图中空间和通道的完整信息。具体而言, 该方法通过计算师生模型的空间和通道间相关性矩阵, 并利用这些矩阵提出空间损失函数和通道损失函数。具体流程如图6所示。

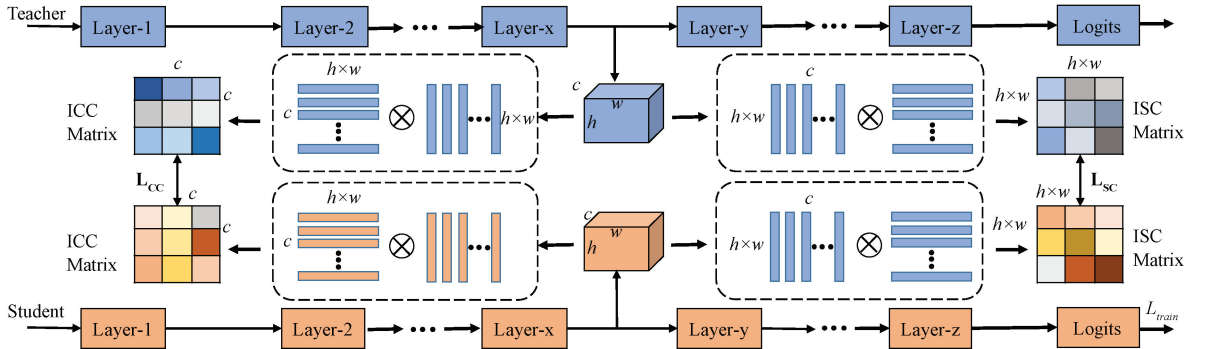


图6 SCCKD示意图

Fig. 6 Schematic diagram of SCCKD

2 实验与分析

2.1 实验环境

为评估改进的轻量化YOLO模型对船舶检测任务的

在SCCKD知识蒸馏中, 教师模型与学生模型的特征图分别记为 $F_T \in \mathbf{R}^{C \times H \times W}$ 和 $F_S \in \mathbf{R}^{C' \times H \times W}$, 其中 C 和 C' 分别表示教师和学生模型的通道数, H 和 W 为特征图的空间维度。为了捕捉特征图的空间位置关系和通道间特征协作关系, 本文通过构建空间相关性矩阵、通道相关性矩阵对特征图的多维关系进行建模。首先, 沿通道维度展平特征图 F_T 和 F_S , 得到空间相关性矩阵 S_T 和 S_S :

$$S_T = f_c(F_T) \cdot f_c(F_T)^T \quad (3)$$

$$S_S = f_c(F_S) \cdot f_c(F_S)^T \quad (4)$$

式中: $f_c(F_T)$ 将特征图展平为 $[H \times W, C]$ 的矩阵, S_T 和 S_S 分别描述了教师模型和学生模型特征图中空间位置之间的关系。对 S_T 和 S_S 这两个矩阵进行归一化处理, 使得它们的空间维度上具有相同的尺度。为了让学生模型获得相似的特征多样性, 空间相关性损失 L_{sc} 通过最小化学生和教师空间相关性矩阵的 L_2 距离来实现:

$$L_{sc} = \frac{1}{(H \times W)^2} \|S_S - S_T\|_2^2 \quad (5)$$

同理, 沿空间维度展平特征图, 得到通道相关性矩阵 C_T 和 C_S 。归一化处理后, 通道间相关性损失 L_{cc} 通过最小化学生和教师通道相关性矩阵的 L_2 距离来实现:

$$L_{cc} = \frac{1}{C^2} \|C_S - C_T\|_2^2 \quad (6)$$

本文提出的空间相关性损失和通道间相关性损失, 以提高学生模型对教师模型特征表达的全面学习能力。最终的损失函数定义为:

$$L_{SCCKD} = L_{train} + \alpha L_{sc} + \beta L_{cc} \quad (7)$$

式中: L_{train} 是正常训练的损失, L_{cc} 是通道间相关性损失, L_{sc} 是空间相关性损失, α 和 β 是权重因子。本文对颈部网络输出的P3、P4和P5层特征应用SCCKD知识蒸馏, 引导学生模型学习教师模型的特征关系。

检测效果, 实验平台运行 Windows 10 操作系统, 处理器为 Intel Core i5-10400F @ 2.9 GHz, 搭载 32 GB 内存和 Nvidia GeForce RTX 3090 Ti 显卡, 显存容量为 24 GB。使用 Python 3.8 作为编程语言, 深度学习框架为

Pytorch2. 2. 2+cu118。模型训练的关键参数如表 1 所示。

表 1 超参数设置
Table 1 Hyper parameter setting

参数名称	数值
Learning rate	0. 01
Momentum	0. 937
Weight decay	0. 000 5
Batch size	32
Epoch	100

2. 2 实验数据集

本文采用了公开数据集 SeaShips^[21] 来训练和评估目标检测模型,以保证实验结果的可重复性和可靠性。



图 7 常见船舶类型
Fig. 7 The common types of ships



图 8 数据增强
Fig. 8 Data enhancement

2. 3 评价指标

本文主要研究的是基于 YOLOv8n 算法进行改进实现内河船舶检测算法的轻量化与精度提升,因此实验选取准

SeaShips 数据集涵盖了多种场景和目标类型,在船舶目标检测领域被广泛应用。所有图像由部署在海岸线的视频监控摄像头采集而来,涵盖了各种成像条件的变化,如船体部分、比例、光照、视角、背景和遮挡情况。SeaShips 数据集总共包含 7 000 张分辨率为 1 920×1 080 图片样本,包含 6 种常见的船舶类型,如图 7 所示,分别是散货船、集装箱船、渔船、杂货船、矿砂船和客船。

为了进一步提升模型的泛化能力,首先对数据集进行质量筛选,剔除了一些标注不准确或图像质量较差的样本,从剩余数据集中随机选择 1 000 张作为验证集,1 000 张作为测试集。然后在剩余的样本中,随机使用噪声添加、尺度变换、随机擦除、随机亮度调整和水平翻转多种数据增强操作,将数据集扩充至 8 000 张作为训练集,形成最终的 10 000 张完整数据集。本文使用的数据增强方式如图 8 所示。

确率(precision, P)、召回率(recall, R)和平均均值精度(mean average precision, mAP)这 3 项将作为性能评价指标。在模型轻量化任务中,模型的参数量、计算量和模型

体积也应当作为评价指标。考虑到模型部署到终端设备上,采用模型推理每秒帧数(frames per second, FPS)作为模型实时性能评价指标。因此本实验评价指标共有 7 项,分别是 P、R、 $mAP_{50:95}$ 、参数量、计算量、模型体积和 FPS。其中 P、R、 mAP 的计算公式分别为:

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P(r) dr \quad (10)$$

$$mAP = \frac{\sum_{i=0}^N AP_i}{N} \quad (11)$$

式中: TP 表示正样本预测为正的例子; FP 表示负样本中预测为正的例子; FN 表示正样本被错误预测为负的例子; TN 表示负样本中预测为负的例子。 N 表示检测种类数。在计算 FPS 的过程中,输入图片大小为 640×640 ,为保障准确性并且发挥显卡并行计算能力, $batch-size$ 设置为 32,并且进行 200 次热身训练使其达到最佳状态,最后进行 1 000 次测试取平均值。

2.4 消融实验

为了验证本文提出的 YOLO-LISD 算法的有效性,本文以 YOLOv8n 作为基线,在经数据增强后的 SeaShips 数据集上设计多组消融实验,以评估各项改进对模型性能的影响。消融实验结果如表 2 所示。

表 2 消融实验

Table 2 Ablation experiments

序号	YOLOv8n +				P/%	$mAP_{50:95}$ /%	参数量 / 10^6	计算量 /G	FPS
	EFS DH	slim-neck	GCAMBP	SCCKD					
1	×	×	×	×	96.8	78.9	3.01	8.1	1 129.9
2	√	×	×	×	96.9	79.7	2.36	6.5	1 174.3
3	×	√	×	×	97.6	78.8	2.80	7.3	1 094.7
4	√	√	×	×	97.7	79.7	2.22	5.8	1 135.1
5	√	√	√	×	97.0	79.5	0.95	3.5	1 377.7
6	√	√	√	√	97.9	81.0	0.95	3.5	1 377.4

消融实验结果表明,本文提出的方法在内河船舶目标检测算法的轻量化和性能提升方面均发挥了显著作用。实验 1 展示了未经任何改进的 YOLOv8n 实验数据,作为后续改进措施实验结果的对照基准。实验 2 将原解耦检测头替换为高效特征共享检测头,多尺度特征共享卷积核参数,实现参数量与计算量同步降低,平均检测精度提升 0.8%。实验 3 采用轻量化的 GSConv 及其衍生模块优化颈部网络,在进一步压缩计算成本的同时提升检测准确率,虽平均检测精度下降 0.1%但在容忍范围内。FPS 相比基线网络下降了 35,这是由于深度可分离卷积导致内存访问量的增加,从而影响模型的推理速度,未来将着重优化这一问题。实验 4 表明 EFS DH 检测头的高效计算特性有效对冲 slim-neck 的延迟效应,使 FPS 回升至 1 135.1,验证了模块间的计算效率互补性。实验 5 表明基于 GCAMBP 分数的全局通道自适应幅度剪枝策略通过剪除贡献度较低的通道,实现参数量减少 57.2%、计算量降低 39.7%,该策略虽然导致检测性能出现可控性衰减,但其在模型压缩与精度保持间实现了有效权衡。实验 6 采用了 SCCKD 特征蒸馏,通过最小化学生模型与教师模型 P3、P4 和 P5 层输出之间的差异进行训练,从而学习到教师模型的特征提取方式。通过模仿教师模型的行为,学生模型能够在保持较少参数的同时提升性能,有效弥补了

剪枝过程中表征能力的损失。最终获得的 YOLO-LISD 检测器较原始 YOLOv8n,参数量压缩 68.4%(3.01 M→0.95 M),计算量降低 56.8%(8.1 GFlops→3.5 GFlops),关键检测指标显著提升:检测准确率提升 1.1%(96.8%→97.9%), $mAP_{50:95}$ 提高 2.1%(78.9%→81.0%)。本文的改进方法在参数量压缩与计算密度降低的基础上,同步提升检测性能,实现计算效率与检测精度平衡。

2.5 对比实验

1) 不同模型对比实验

为了更好地验证 YOLO-LISD 的优势,本文将 YOLOv4 发布至今的 YOLO 系列算法作为对比对象,在经数据增强后的 SeaShips 数据集进行了实验对比。此外,本文还与最先进的船舶检测模型进行了性能对比。

不同模型对比实验结果如表 3 所示, YOLO-LISD 的准确率、精度、参数量、计算复杂度和模型大小都优于其他 YOLO 系列算法及当前先进模型,更满足内河船舶检测算法在算力有限设备部署的实际需求。

2) 不同剪枝率性能对比

对直接轻量化设计后的模型按照不同比例进行剪枝,实验结果如表 4 所示。由表 4 可知,随着剪枝率的逐步提高,模型的参数量和计算量逐渐减少,但模型性能并非总是下降。当在剪枝率为 33.3%时,模型 $mAP_{50:95}$ 值反而

得到提升,这表明模型存在重复特征学习,导致参数和计算资源浪费。当剪枝率为 40%时,P 和 mAP_{50:95} 值略有下降但都在容忍范围之内,因此将该模型作为学生模型进行后续的知识蒸馏操作。

表 3 不同检测算法对比

Table 3 Comparison of different detection algorithms					
模型	P	mAP _{50:95}	参数量	计算量	模型体积
	/%	/%	/10 ⁶	/G	/MB
YOLOv4-tiny	91.8	69.2	5.89	16.2	22.5
YOLOv5s	95.7	77.4	7.03	15.8	13.7
YOLOv7-tiny	95.3	73.2	6.02	13.1	11.7
YOLOv8n	96.8	78.9	3.01	8.1	6.0
YOLOv9-t	96.9	78.8	2.80	11.7	6.3
YOLOv10n	94.3	76.8	2.70	8.2	5.5
YOLO11n	95.9	77.6	2.58	6.3	5.2
文献[22]	97.6	—	2.65	7.5	—
文献[23]	95.4	73.0	2.32	6.5	—
文献[24]	—	76.8	8.8	23.5	—
文献[25]	97.4	75.7	3.69	7.4	—
YOLO-LISD(ours)	97.9	81.0	0.95	3.5	2.4

表 4 不同剪枝率性能对比

Table 4 Performance comparison of different pruning rates					
剪枝率	P	mAP _{50:95}	参数量	计算量	FPS
/%	/%	/%	/10 ⁶	/G	
base	97.7	79.7	2.22	5.8	1 135.1
33.3	97.7	80.3	1.11	4.3	1 248.3
40	97.0	79.5	0.95	3.5	1 377.7
50	96.8	77.1	0.89	3.0	1 423.1
66.6	96.4	75.1	0.86	2.7	1 582.8

3) 不同蒸馏损失权重性能对比

将剪枝后的模型作为学生模型,YOLOv8m 作为教师模型,使用 SCCKD 方法对颈部网络 P3、P4 和 P5 输出层使

用 SCCKD 方法进行知识蒸馏。该方法采用 α 和 β 两个超参数来平衡空间相关性损失和通道间相关性损失,实验结果如表 5 和表 6 所示。

表 5 不同 α 值对 SCCKD 效果的影响

Table 5 Effects of different α values on SCCKD performance				
α	P/%	R/%	mAP ₅₀ /%	mAP _{50:95} /%
0	97.0	94.0	98.1	79.5
0.5	96.8	95.8	98.7	80.3
1.0	97.9	94.3	98.8	80.3
2.0	97.7	93.9	98.6	80.0
4.0	98.1	94.0	98.8	80.6
5.0	96.4	96.4	98.7	80.5

表 6 不同 β 值对 SCCKD 效果的影响

Table 6 Effects of different β values on SCCKD performance				
β	P/%	R/%	mAP ₅₀ /%	mAP _{50:95} /%
0	97.0	94.0	98.1	79.5
0.25	97.3	94.9	98.5	80.6
1.0	97.4	95.4	98.8	80.6
1.5	98.1	95.0	98.8	80.4
2.0	96.7	94.8	98.3	80.2
3.0	96.5	95.4	98.5	80.4

实验结果表明,SCCKD 在合理选择超参数 α 和 β 时,能够有效提升模型的性能,尤其在 mAP50 和 mAP50-95 指标上表现优异。然而,P 或 R 指标可能会因超参数变化而出现波动,这是由于空间相关性损失和通道相关性损失的权重变化可能对特定样本的特征学习产生较大影响,从而影响了指标的局部波动。当 $\alpha=4.0,\beta=1.0$ 时,定量结果最好。

当 $\alpha=4.0,\beta=1.0$ 时,使用 SCCKD 对学生模型进行知识蒸馏,P、R、mAP₅₀ 和 mAP_{50:95} 分别提高了 0.9%、1.7%、0.9%和 1.5%。实验结果如表 7 所示。

表 7 SCCKD 对模型性能的影响

Table 7 Effect of SCCKD on model performance							
模型	P/%	R/%	mAP50/%	mAP _{50:95} /%	参数量/10 ⁶	计算量/G	模型体积/MB
教师模型	98.4	97.8	99.1	85.5	25.84	78.7	49.6
学生模型	97.0	94.0	98.1	79.5	0.95	3.5	2.4
YOLO-LISD	97.9	95.7	99.0	81.0	0.95	3.5	2.4

2.6 系统应用设计与分析

本文基于 PyQt5 设计并实现了一套内河船舶检测系统,如图 9 所示。系统支持模型权重加载、模型推理、结果可视化操作,并在检测结果统计面板实时显示目标种类、位置和置信度信息。

为验证在算力受限环境中的使用性,在无 GPU 加速的测试平台(Intel Core i5-12400F @ 2.5 GHz)上导入 YOLO-LISD 模型,并对本地存储的内河船舶视频进行测试。其中测试视频共 600 帧,分辨率为 1 920×1 080,选择视频推理的平均帧率 FPS 作为检测实时性能评估指标。

实验结果显示, YOLO-LISD 推理时延降低至 18.1 ms, 达到 55.1 fps 的处理速度, 较基准模型 YOLOv8n 的 42.3 fps 提升 30%。实验数据表明, YOLO-LISD 满足算力受限环境中实时检测需求。



图 9 YOLO-LISD 内河船舶检测系统

Fig. 9 YOLO-LISD inland ship detection system

为检验 YOLO-LISD 在现实环境中的检测效果, 对铜陵长江大桥下游、芜湖滨江公园处江面行驶的船舶进行实地检测。当相机跟随船舶转动使其位于视野中央, 船

舶在不同角度下的形状和大小有所差异, 此时 YOLOv8n 会出现检测框偏移、误检情况, 并且置信度波动较大; 而 YOLO-LISD 能够保持稳定的检测效果, 该场景的可视化实验结果如图 10(a) 所示。当固定相机不动, 船舶在进入或离开视野的过程中, YOLOv8n 会发生误检或漏检现象, 且检测结果不稳定; 而 YOLO-LISD 依然能够准确识别目标, 并保持稳定的置信度, 该场景的可视化实验结果如图 10(b) 所示。YOLO-LISD 的优越性得益于改进的检测头设计, 实现多尺度特征高效共享和一致性处理。

YOLO-LISD 在其他多个方面表现仍优于 YOLOv8n, 检测结果如图 11 所示。首先, 只有部分船体进入视角时, YOLO-LISD 能够更准确地检测出目标。其次, 集装箱船与散货船在某些角度下具有相似特征, YOLO-LISD 凭借其高效的特征提取能力避免误检。再次, YOLO-LISD 正确识别到距离较远的对岸船舶, 表现出模型在小目标检测中性能依旧稳定。最后, 在大多数场景中, YOLO-LISD 识别到物体的置信度均高于 YOLOv8n, 检测结果更加可靠。YOLO-LISD 相较于 YOLOv8n 是改进有效的, 不仅实现了模型的轻量化, 而且在提升检测精度和鲁棒性方面表现出显著优势。



图 10 YOLOv8n 与 YOLO-LISD 动态检测效果对比

Fig. 10 Comparison of dynamic detection effect between YOLOv8n and YOLO-LISD

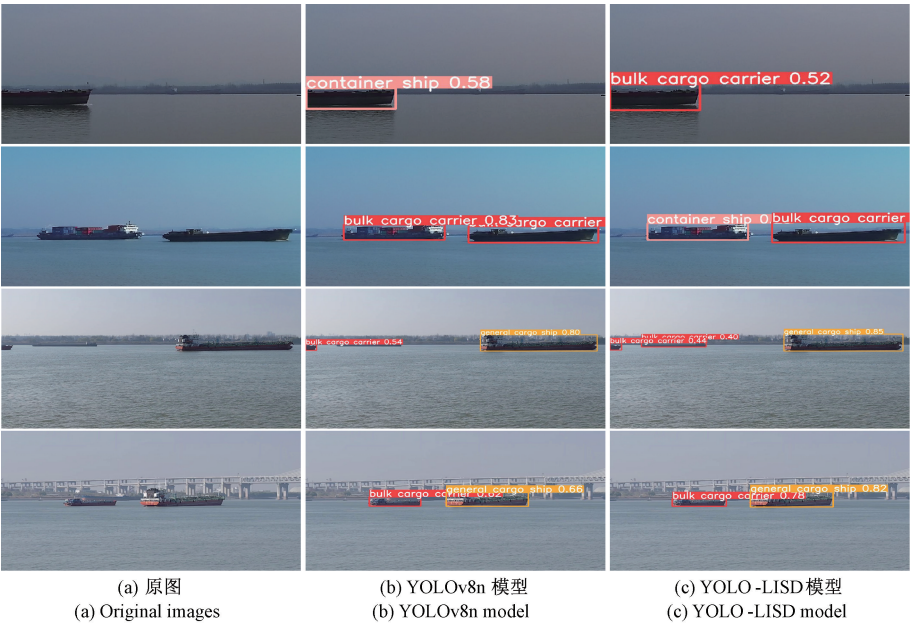


图 11 部分检测效果对比

Fig. 11 Comparison of some detection results

3 结 论

本文基于 YOLOv8n 做了一系列改进,提出了一种针对内河船舶检测的轻量化模型 YOLO-LISD,改善了检测过程中因视角、距离变化导致的结果波动问题,同时实现了轻量化,满足在低算力设备上实时检测需求。具体来说,首先设计融合细节增强卷积的特征共享检测头替代原有检测头,显著降低了模型的参数数量与计算成本,并且提高模型动态检测的一致性。其次引入 slim-neck 方法对颈部网络进行轻量化设计,在轻量化网络的同时提升检测准确率。然后设计基于 GCAMBP 分数的通道剪枝技术进一步压缩模型冗余,提高模型检测效率。最后设计空间-通道相关性知识蒸馏算法,对剪枝后的模型进行优化,提升检测精度。实验结果表明,YOLO-LISD 相比基线模型参数数量与计算量分别降低了 68.4% 与 56.8%,同时 P 与 mAP_{50,95} 分别提高了 1.1% 与 2.1%,达到了 97.9% 与 81.0%。该模型在降低计算成本的同时实现了更好的检测性能。未来将致力于在嵌入式系统中部署该模型,并优化其在低功耗设备上的性能和应用效果,进一步拓展其在更多领域的应用。

参考文献

[1] 刘芬,孙杰,张帅,等. 基于 YOLOv5 的红外船舶目标检测算法[J]. 红外与激光工程, 2023, 52(10): 222-233.

LIU F, SUN J, ZHANG SH, et al. Infrared ship target detection algorithm based on YOLOv5 [J]. Infrared and Laser Engineering, 2023, 52(10):

222-233.

[2] 罗芳,李家威,何道森. 尺度适应性感受野的船舶目标检测方法[J]. 计算机应用研究, 2024, 41(8): 2521-2527.

LUO F, LI J W, HE D S, Ship object detection based on scale-adaptive receptive field [J]. Application Research of Computers, 2024, 41(8): 2521-2527.

[3] YANG X, ZHANG J N, CHEN CH Z, et al. An efficient and lightweight CNN model with soft quantification for ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.

[4] 谢友晨,徐天阳,汤张泳,等. YOLO-WIT: 红外遥感弱小船舶目标检测算法[J/OL]. 北京航空航天大学学报, 1-17 [2025-01-21]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0834>.

XIE Y CH, XU T Y, TANG ZH Y, et al. YOLO-WIT: Infrared remote sensing weak and small ship target detection algorithm[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 1-17 [2025-01-21]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0834>.

[5] 张炳焱,张闯,石振男,等. 基于 YOLO-FNC 模型的轻量化船舶检测方法[J]. 中国舰船研究, 2024, 19(5): 180-187.

ZHANG B Y, ZHANG CH, SHI ZH N, et al. Lightweight ship detection method based on YOLO-FNC model[J]. Chinese Journal of Ship Research,

- 2024,19(5):180-187.
- [6] 张燊,胡林,孙祥娥,等.基于注意力机制及多尺度融合的红外船舶检测[J].激光与光电子学进展,2023,60(22):256-262.
- ZHANG SH, HU L, SUN X E, et al. Infrared ship detection using attention mechanism and multiscale fusion[J]. Laser & Optoelectronics Progress, 2023, 60(22):256-262.
- [7] ZHAO X M, SONG Y L, SHI S X, et al. Research on lightweight ship target detection algorithm based on improved YOLOv5 [C]. IEEE 11th International Conference on Information, Communication and Networks(ICICN), 2023: 728-732.
- [8] 王海群,魏培旭,解浩龙,等.基于改进 YOLOv8 的红外船舶检测[J].电光与控制,2025,32(1):61-67.
- WANG H Q, WEI P X, XIE H L, et al. Infrared ship detection based on improved YOLOv8 [J]. Electronics Optics & Control, 2025,32(1):61-67.
- [9] 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 SH B, WEI L SH, 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.
- [10] 杨志渊,罗亮,吴天阳,等.改进 YOLOv8 的轻量级光学遥感图像船舶目标检测算法[J].计算机工程与应用,2024,60(16):248-257.
- YANG ZH Y, LUO L, WU T Y, et al. Improved lightweight ship target detection algorithm for optical remote sensing images with YOLOv8[J]. Computer Engineering and Applications, 2024,60(16):248-257.
- [11] 胡霞,仲林林.基于改进 R3det 的无人机电力杆塔倾斜程度检测[J].仪器仪表学报,2023,44(10):189-200.
- HU X, ZHONG L L. Incline detection of power towers from UAV images based on the improved R3det[J]. Chinese Journal of Scientific Instrument, 2023, 44(10): 189-200.
- [12] 杨瑞君,张浩,叶璟.改进 YOLOv8n 的轻量级遥感图像军用飞机检测算法[J].电子测量技术,2025,48(1):154-165.
- YANG R J, ZHANG H, YE J. Improved lightweight military aircraft detection algorithm for remote sensing images with YOLOv8n[J]. Electronic Measurement Technology, 2025,48(1):154-165.
- [13] 杨文,胡昊,李凌志,等.基于机器视觉的铁路限界入侵检测方法[J].铁道科学与工程学报,2025,22(3):1328-1343.
- YANG W, HU H, LI L ZH, et al. Railway boundary foreign object intrusion detection method based on machine vision[J]. Journal of Railway Science and Engineering, 2025,22(3):1328-1343.
- [14] 胡玮,王满利,张长森,等.基于退钻过程检测的矿井下钻杆计数方法研究[J/OL].煤炭学报,1-11[2025-02-26]. <https://doi.org/10.13225/j.cnki.jccs.2024.0880>.
- HU W, WANG M L, ZHANG CH S, et al. Research on drill rod counting method in mines based on drill withdrawal process detection[J/OL]. Journal of China Coal Society, 1-11 [2025-02-26]. <https://doi.org/10.13225/j.cnki.jccs.2024.0880>.
- [15] 赵露强,彭强吉,兰玉彬,等. CottonBud-YOLOv5s 轻量型棉花顶芽检测算法[J].农业工程学报,2025,41(4):176-185.
- ZHAO L Q, PENG Q J, LAN Y B, et al. CottonBud-YOLOv5s lightweight cotton bud detection algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(4): 176-185.
- [16] CHEN Z X, HE Z W, LU ZH M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.
- [17] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A lightweight design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21(3), DOI:10.1007/s11554-024-01436-6.
- [18] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning[J]. ArXiv preprint arXiv: 2010.07611,2020.
- [19] 杨森,张鹏超,王磊,等.集成改进 YOLOv8n 与通道剪枝的轻量化番茄叶片病虫害识别方法[J].农业工程学报,2025,41(2):206-214.
- YANG S, ZHANG P CH, WANG L, et al. Identifying tomato leaf diseases and pests using lightweight improved YOLOv8n and channel pruning[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(2): 206-214.
- [20] LIU L, HUANG Q L, LIN S H, et al. Exploring inter-channel correlation for diversity-preserved knowledge distillation [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 8251-8260.
- [21] SHAO ZH F, WU W J, WANG ZH Y, et al. SeaShips: A large-scale precisely annotated dataset for ship detection[J]. IEEE Transactions on Multimedia,

2018, 20(10): 2593-2604.

[22] CAO X T, SHEN J Y, WANG T, et al. Ship detection based on improved YOLOv8 algorithm[C]. 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control(RAIIIC), 2024:20-23.

[23] CHENG K Y, SONG M M, XU Y ZH, et al. YOLO-FE: A lightweight ship detection algorithm based on improved YOLOv8[C]. 2023 5th International Conference on Robotics, Intelligent Control and Artificial Intelligence(RICAI), 2023:1082-1087.

[24] 邢汇源,崔亚奇,王子玲,等. 复杂海况下的海上船舶目标检测算法[J]. 现代防御技术, 2024, 52(6):88-96.

XING H Y, CUI Y Q, WANG Z L, et al. Target detection algorithm for ships at sea under complex sea conditions[J]. Modern Defense Technology, 2024, 52(6): 88-96.

[25] LANG C, YU X Y, RONG X W. LSDNet: A lightweight ship detection network with improved YOLOv7[J]. Real-Time Image Processing, 2024, 21(60), DOI: 10.1007/s11554-024-01441-9.

作者简介

董健, 硕士研究生, 主要研究方向为深度学习、目标检测。

赵欣(通信作者), 博士, 副研究员, 主要研究方向为环境光学监测仪器、光学遥感技术。

E-mail: xzhao@aiofm.ac.cn