

基于表观细粒度辨别网络的近海船舶目标检测方法

闵令通, 范子满, 窦飞阳, 吕勤毅, 李鑫
(西北工业大学电子信息学院 西安 710072)

摘要: 近海船舶目标检测是一项非常具有挑战性的任务, 受到学者专家广泛关注。基于卷积神经网络 (CNN) 和注意力机制的检测器在近海船舶目标检测方面的应用取得了显著成就。然而, 船舶目标检测存在着表观相似和背景干扰导致检测过程中出现误检的问题。为此, 本文提出了一种用于 Faster RCNN (更快的基于区域的卷积神经网络) 的表观细粒度辨别的检测头模块。该模块包括类别细粒度分支和高效全维动态卷积定位分支。其中类别细粒度分支通过全局特征建模和灵活的感知范围来挖掘和利用类别细粒度辨别特征, 高效全维动态卷积定位分支通过高效灵活的感知船舶边界信息来区分目标与背景, 从而减少误检漏检问题。通过在近海船舶公开数据集 Seaships7000 上进行实验验证, 本文算法减少了误检漏检, 提升了检测器性能。

关键词: 船舶目标检测; 类别细粒度; 表观判别; 全维动态卷积; 自注意力

中图分类号: TP75; TP183 **文献标志码:** A **文章编号:** 2095-1000(2024)02-0001-09

DOI: 10.12347/j.ycyk.20231114001

引用格式: 闵令通, 范子满, 窦飞阳, 等. 基于表观细粒度辨别网络的近海船舶目标检测方法[J]. 遥测遥控, 2024, 45(2): 1-9.

Nearshore Ship Object Detection Method Based on Appearance Fine-grained Discrimination Network

MIN Lingtong, FAN Ziman, DOU Feiyang, LYU Qinyi, LI Xin

(School of Electronic Information, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Offshore ship object detection is a very challenging task and has received widespread attention from scholars and experts. Detectors based on Convolutional Neural Networks (CNN) and attention mechanisms have made significant progress in offshore ship object detection. However, the problem of false detection in the detection process is caused by the apparent similarity and background interference of ship targets. In order to solve this problem, this paper proposes a detection head module for fine-grained appearance discrimination implemented with Faster RCNN. This module includes a category fine-grained branch and an efficient full-dimensional dynamic convolution localization branch. The category fine-grained branch mines and utilizes category fine-grained identification features through global feature modeling and flexible perception range. The efficient omni-dimensional dynamic convolution positioning branch distinguishes objects and backgrounds through the efficient and flexible perception of ship boundary information, thereby reducing false and missed detections. Through experimental verification on the offshore ship public dataset Seaships7000, the proposed algorithm reduces false detections and missed detections and improves detector performance.

Keywords: Ship object detection; Similar feature extraction; Apparent discrimination; Dynamic convolution; Self-attention

Citation: MIN Lingtong, FAN Ziman, DOU Feiyang, et al. Nearshore Ship Object Detection Method Based on Appearance Fine-grained Discrimination Network[J]. Journal of Telemetry, Tracking and Command, 2024, 45(2): 1-9.

0 引言

随着我国在海洋强国建设方面的不断推进, 智慧航运和智慧海洋工程领域迅速蓬勃发展。在

这一背景下, 借助海事图像或视频进行高效的近海舰船目标检测, 以确保航运和海洋工程的安全, 已经成为一项急切的需求^[1]。近海船舶目标检测在海域监控^[2]、近海风电平台^[3]和海上交通控制^[4]等多

个领域中均有重要应用。

深度学习的不断进步为目标检测领域开辟了崭新的路径。一系列基于卷积神经网络(Convolutional Neural Network, CNN)的检测器^[5-7]持续提高了自然图像的检测准确性和推理速度。例如, Liu 等人^[8-11]提出了单阶段目标检测网络, 极大地提升了推理速度。Ren 等人^[12]提出了生成高效候选提议的 Faster RCNN 双阶段目标检测网络, 提升了检测器的准确性。针对近海船舶图像, 基于 Faster RCNN 的双阶段目标检测网络为准确检测奠定了基础。Wei 等人^[13]通过改进 Faster RCNN 的卷积层提

升了舰船目标的检测精度。Guo 等人^[14]通过设计旋转不变的候选框提议网络提升了舰船目标的检测性能。Dong 等人^[15]基于 Faster RCNN 设计了视觉注意力机制, 提升了舰船目标的准确率。Dong 等人^[16]通过改进候选提议生成网络提升了船舶目标检测性能。这些工作有效地推动了舰船目标检测的发展, 但是忽视了舰船目标存在表观相似和背景干扰的问题。如图 1 所示, 普通货船与渔船之间存在外观一致, 尺寸相似等特点。Faster RCNN 检测结果显示图中的渔船被检测为普通货船, 岸边也出现了将背景误检为普通货船的情况。

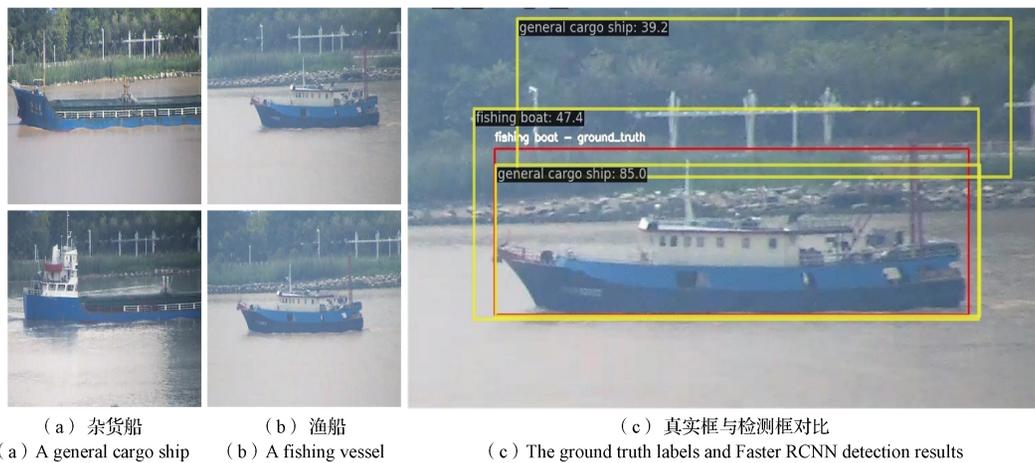


图 1 真实标签(红色)和 Faster RCNN 检测结果(黄色)

Fig.1 The ground truth labels (in red) and Faster RCNN detection results (in yellow)

注意力机制是解决表观相似和背景干扰的近海船舶目标检测的有效方法, 可以提升船舶目标的检测准确率。注意力机制中尤其是 Transformer^[17-19]和动态卷积^[20-22]在全局信息建模方面和灵活感知范围方面展现出卓越的性能, 专家学者做出了大量研究工作。例如, Carion 等人^[23]基于 Transformer 提出了端到端的目标检测器, 有效推动了 Transformer 在目标检测的应用。Cheng 等人^[24]基于 Transformer 设计了检测头部, 全局特征的注意力机制能够捕获类别信息显著特征, 提升检测器的准确性。Wei 等人^[25]基于动态卷积设计了检测头部, 灵活感知的注意力机制能够提升复杂检测环境下的检测器准确性。与上述使用注意力机制的方法不同, 本文聚焦于全局特征和灵活感知注意力机制的联合设计。

近海船舶图像存在表观相似和背景干扰导致的误检漏检问题。为了解决上述问题, 本文采用

联合全局特征和灵活感知注意力机制的方法, 重新设计了 Faster RCNN 的头部网络。该网络包括两部分: 类别细粒度分支(Class Fine-Grained Branch, CFGB)和高效全维动态卷积定位分支(Location Branch with Dynamic Convolution, LBDC), 能够增强对检测头输入特征对细粒度辨别能力和定位能力。其中, CFGB 模块旨在学习到表观相似目标的细粒度差异, 增强类别特征的判别能力, 降低表观相似导致的误检。LBDC 模块旨在灵活捕获目标边界特征信息以区分目标与背景, 增强回归框的定位能力, 降低背景干扰造成误检的可能。最后, 本文采用较为常用的 ResNet50^[26]和 FPN^[27]的 Faster RCNN 构建了细粒度辨别网络(Fine-Grained Discrimination Network, FNet), 并在公开数据集 Seaships7000^[28]上验证了 FNet 网络的有效性。

1 方法

FDNet网络框架如图2所示,包括骨干网络、特征金字塔网络(Feature Pyramid Network, FPN)、区域建议网络(Region Proposal Network, RPN)、感兴趣区域(Region of Interest, RoI)Align^[29](对齐)模块与检测头网络。工作流程如下:首先,输入图像依次经过骨干网络、FPN、RPN和RoI相关的操作,包括提取RoIs和RoI Align后获得的特征向量 $F \in \mathbb{R}^{C \times H \times W}$;其次,检测器头部的输入特征 $F \in \mathbb{R}^{C \times H \times W}$ 分别送入到高效全维动态卷积定位分支LBDC模块和类别细粒度分支CFGB模块,得到模块输出目标定位框特征向量 Y_o 和目标类别特征向量 Y_c 。最后 Y_c 和 Y_o 分别经过全连接层(Fully Connected Layer, FCL)得到船舶目标的类别信息和定位框信息。训练损失计算与Faster R-CNN保持一致,其中回归损失采用Smooth L1,类别损失采用交叉熵损失(Cross Entropy Loss, CEL)。接下来分别对LBDC模块和CFGB模块进行介绍。

1.1 高效全维动态卷积定位分支LBDC模块

面对背景干扰下船舶目标的定位任务,近海船舶目标跟与沿岸环境融为一体,使得目标边界难以清晰地被检测器捕捉,如图1所示,导致误检现象的发生。受到CSPDarknet53^[30](一种深度卷积网络)的启发,采用了多层残差结构和全维动态卷积,联

合了全局特征信息和灵活感知注意力机制,构建了船舶目标边界信息高效提取的动态卷积定位分支。其中,全维动态卷积能够对目标边界进行灵活建模,多层残差结构有助于特征的重复利用,从而对边界信息多次建模,有助于目标精准定位。

首先,LBDC模块对输入特征 F 采用 1×1 卷积进行处理;然后,对处理之后的特征沿通道维度划分,即 $T_1 \in \mathbb{R}^{C' \times H \times W}, T_2 \in \mathbb{R}^{C' \times H \times W}, C' = \frac{1}{2}C$ 。为了灵活感知目标边界信息,采用了基于全维动态卷积的Bottleneck(瓶颈)模块(同时为了减少参数规模,本文只将 T_2 作为Bottleneck模块的输入,得到该模块的输出特征 B ,而 T_1 保持不变);最后,为了增加对特征信息的多重利用和保持特征维度不变,将 B, T_1, T_2 在通道维度进行拼接之后,经过 1×1 卷积处理得到输出 Y_o ,可表示为:

$$Y_o = f_{cl}(\text{Cat}(B, T_1, T_2)) \quad (1)$$

其中, $f_{cl}(\cdot), \text{Cat}(\cdot)$ 分别表示 1×1 卷积运算和通道维度拼接。

在Bottleneck模块中,如图2所示,采用了两个 3×3 的全维动态卷积和残差结构。全维动态卷积可以灵活地感知船舶目标边界信息,根据不同目标自适应的学习细微差异信息。如图3(e)所示,全维动态卷积会将作用于输入的卷积核进行四个维度加权求和,可用公式(2)表示:

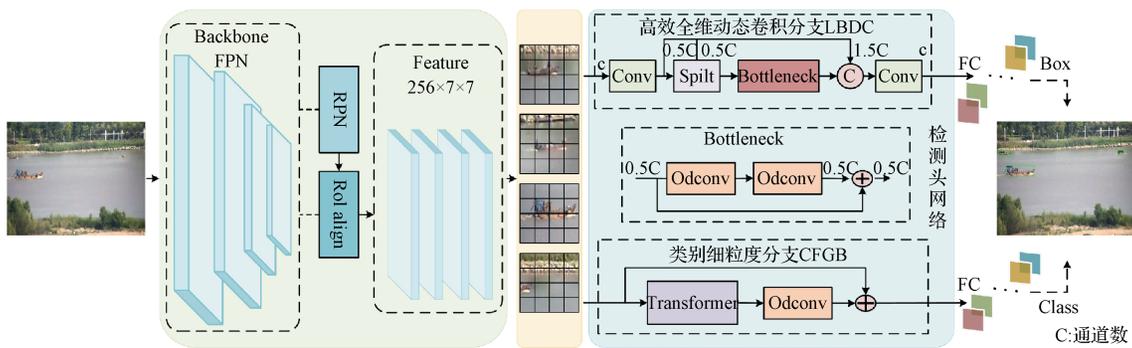


图2 FDNet的总体检测框架

Fig. 2 The overview detection framework of FDNet

$y = (\alpha_{w_1} \cdot \alpha_{f_1} \cdot \alpha_{c_1} \cdot \alpha_{s_1} \cdot W_1 + \dots + \alpha_{w_n} \cdot \alpha_{f_n} \cdot \alpha_{c_n} \cdot \alpha_{s_n} \cdot W_n) * x \quad (2)$

其中, x, y 分别为全维动态卷积的输入输出, W_i 为卷积核权重。四个维度权重对于 W_i 的作用如图3中(a)、(b)、(c)、(d)所示, α_{f_i} 表示卷积核的注意力标量, $\alpha_{c_i}, \alpha_{s_i}, \alpha_{w_i}$ 分别表示沿着卷积核的空间维度、输入通道维度和输出通道维度的三种注意力机制。

$\alpha_{f_i}, \alpha_{c_i}, \alpha_{s_i}, \alpha_{w_i}$ 的计算流程如下:首先全维动态卷积的输入向量 x 统一经过全局平均池化(Global Average Pooling, GAP)、全连接层(FCL)和线性激活函数(Rectified Linear Unit, ReLU)得到中间向量,随后该中间向量分别通过四个FC和Sigmoid函数计算得到。

LBDC 模块可以提升目标边界信息的判别能力, 减少背景干扰导致的误检。另外, Bottleneck 中的残差结构的设计有助于深层网络更好地收敛。最后, 与固定卷积核不同, 全维动态卷积的引入, 让网络可以灵活地捕获船舶目标边界信息的细微差异。

1.2 类别细粒度分支 CFGB 模块

面对船舶目标表观相似模块的分类任务, 只关注目标的显著区域特征^[31]是无法全面捕捉到目标的特定细微差异, 导致表观相似的船舶目标出现漏检误检。因此, 本文设计了类别细粒度分支 (CFGB) 模块, 如图 3 所示, 采用 Transformer^[19]、全维动态卷积^[22]和残差结构, 旨在同时考虑显著特征和目标细微差异, 联合全局建模注意力机制和灵活感知注意力机制。具体而言, 本文选择使用 Transformer 作为模块的一部分, 以便能够有效地提取图像中的显著特征。同时, 为了更细致地关

注船舶目标的特定细微差异, 本文还引入了全维动态卷积。全维动态卷积在处理目标之间的微小差异时具有优势, 通过动态调整卷积核的权重, 能够灵活地捕捉目标的局部特征, 尤其是在目标表观相似性较高的情况下。首先, 对输入特征 $F \in \mathbb{R}^{C \times H \times W}$ 的空间维度展平得到向量 $F' \in \mathbb{R}^{C \times HW}$; 其次, 向量 F' 经过 Transformer, 从全局信息建模来捕获显著特征信息的长远依赖关系得到输出向量 $Y_T \in \mathbb{R}^{C \times HW}$; 最后, 为了灵活地感知以提取类别细微特征差异, 对向量 Y_T 进行空间维度还原操作后再经过公式(2)的全维动态卷积, 并与输入特征 F 经过残差结构, 得到类别细粒度分支的输出 Y_C :

$$Y_C = F + f(Y_T) \quad (3)$$

其中, $f(\cdot)$ 表示公式(1)中全维动态卷积操作, 省略了空间维度还原操作。

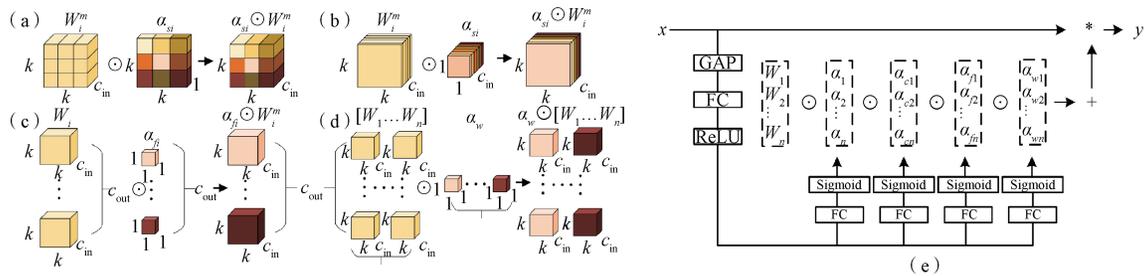


图 3 全维动态卷积

Fig. 3 Omni-dimensional Dynamic Convolution (ODConv)

在 Transformer 结构中, 经过空间维度展平的输入特征 F' 通过第一个残差模块, 即层归一化操作 (LayerNorm)、自注意力操作 (Self-Attention)、DropPath 得到中间特征 $\widehat{M} \in \mathbb{R}^{C \times (H \times W)}$ 。其中, Self-Attention 操作会将输入沿着通道维度分为 N 部分, 每个部分会被复制为三份, 如图 4 所示, 即查询 Q 、键 K 、值 V 。 Q, K, V 通过简单的线性投影, 然后经过缩放的点积运算得到特征矩阵, 可用公式(4)表示为:

$$\widehat{M}_i = \text{Softmax} \left(\frac{Q_i K_i^T}{\sqrt{d}} \right) V_i \quad (4)$$

其中, \widehat{M}_i 是由第 i 个部分导出的特征矩阵, d 为缩放参数, 一般取为 K 的维度。为了方便表示, 公式中省略了 LayerNorm 和 DropPath。 \widehat{M}_i 和输入特征 F' 构成了 Transformer 第一个残差模块的输出向量 P 。第二个残差模块由 LayerNorm、多层感知机

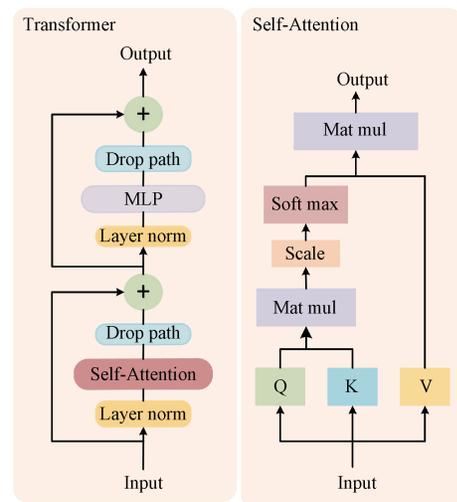


图 4 Transformer 和自我注意力机制

Fig. 4 Transformer and Self-attention mechanism

(Multilayer Perceptron, MLP) 和 DropPath 构成, MLP 包含两个线性层和一个非线性激活函数, 可

以用如下公式(5)表示:

$$\hat{P}_i = f_i((P_i w_1 + b_1) \times w_2 + b_2) \quad (5)$$

其中, \hat{P}_i 表示第 i 个元素经由 MLP 导出的输出值, w_1, b_1 表示第一个线性变换层的权重矩阵和偏置向量, w_2, b_2 表示第二个线性变换层的权重矩阵和偏置向量, $f_i(\cdot)$ 表示非线性激活函数。MLP 模块的输出向量 \hat{P} 和第一个残差模块的输出向量 P 构成了第二个残差模块的输出向量 Y_T 。

CFGB 模块基于 Transformer 的结构, 能够对整个图像进行全局上下文建模, 而不仅仅是在局部感受野内操作, 这有助于模型更好地理解不同船舶目标的显著信息与其周围背景环境的关系。另外, 采用了全维动态卷积, 能够灵活地增加感知范围, 捕

捉船舶相似外观特征之间微弱的差异性, 有助于模型增强对表观相似类别之间的判别能力。

2 实验结果

为了验证 FNet 算法的有效性, 本文在公开数据集 Seaships7000 上进行训练和测试, 如图 5 所示。首先本文对比了 YOLOv3^[8]、FCOS^[9]、YOLOv7^[10]、SSD^[11]、Libra-RCNN^[32] 和 Faster-RCNN^[12] 七个模型与本文模型 FNet, 从而验证了 FNet 的有效性。之后分别对 CFGB 和 LBDC 模块进行了消融试验, 验证了本文的类别细粒度分支对类别细粒度特征提取的有效性和高效全维动态卷积对边界信息的挖掘优势。



图 5 Seaship7000 数据集示意图

Fig. 5 Illustration of the Seaship7000 dataset

2.1 数据集和实验设置

近海船舶检测公开数据集 Seaship7000 总共包含 7 000 张图像 6 个类别(矿砂船、散货船、杂货船、集装箱船、渔船和客船), 分辨率为 1 920×1 080。本文使用 Python3.9、Pytorch2.0、Cuda11.8 构建实验环境, 采用一块显存为 24 GB 的 GeForce RTX 4090 GPU, 所有实验在目标检测开源工具 mmdetection^[33] 上运行。评价标准采用 COCO2014^[34], 分别为 mAP、AP50、AP75、APS、APM 和 APL。

2.2 模型对比

表 1 展示了本文模型 FNet 与其他模型在 Seaship7000 的对比结果。FNet 相较于基准模型

Faster-RCNN 来说, 在 mAP 上提升了 3%, 在 AP50 上提升了 0.6%, 在 AP75 上提升了 5.1%, 说明整体表现优秀。值得注意的是, 针对小目标的指标 APS 也提升了 3.8%, 说明 FNet 模型对小目标的检测也有良好的效果。对比实验结果, 说明了 FNet 模型有助于特征的提取与挖掘, 对表观相似的细粒度判别有效。

本文针对算法效率问题分析了不同模型推理速度 FPS, 参数量 Parameters(M) 和 FLOPs(G) 的比较。其中, FPS 是使用空间分辨率大小为 1 024×1 024 的图像, 在 GeForce RTX 4090 GPU 的实验平台上测试。具体对比结果见表 1, 可以看出本文的模型相较于基准模型 Faster-RCNN, 推理速度 FPS

下降了 3.3 img/s、Parameters 和 FLOPs 分别增加了 23.5 M 和 286.8 G, 这说明引入类别细粒度分支和动态卷积定位分支牺牲了部分推理速度。

表 1 不同模型在 Seaship7000 上的对比
Table 1 Comparison of different models on the Seaship7000 dataset

Models	mAP	AP50	AP75	APS	APM	APL	FPS	Parameters	FLOPs
YOLOv3	58.8	93.6	67.3	4.2	34.9	60.3	49.4	3.7	1.7
FCOS	57.5	90.6	65.3	5.1	37.5	58.4	26.1	32.1	202.8
YOLOv7	58.4	93.3	64.8	17.0	40.2	59.4	83.3	36.5	103.2
SSD	64.8	94.9	75.2	3.1	40.8	66.4	61.2	24.4	30.7
Libra-RCNN	59.2	92.5	68.1	5.1	35.4	60.6	31.0	41.2	226.3
Faster-RCNN	64.4	95.3	75.7	11.2	45.4	65.6	17.5	41.6	210.9
FDNet(Ours)	67.4	95.9	80.8	15.0	49.0	68.3	14.2	65.1	497.7

2.3 定性分析

本文给出了 Faster-RCNN 和 FDNet 检测结果的对比图, 如图 6 所示。Faster-RCNN 在面对较小目标和岸边出现干扰时容易出现误检, 在面对表观

重合时也容易发生漏检。与之不同的是, 在面对上述情况时 FDNet 模型在一定程度上缓解了误检漏检的发生。

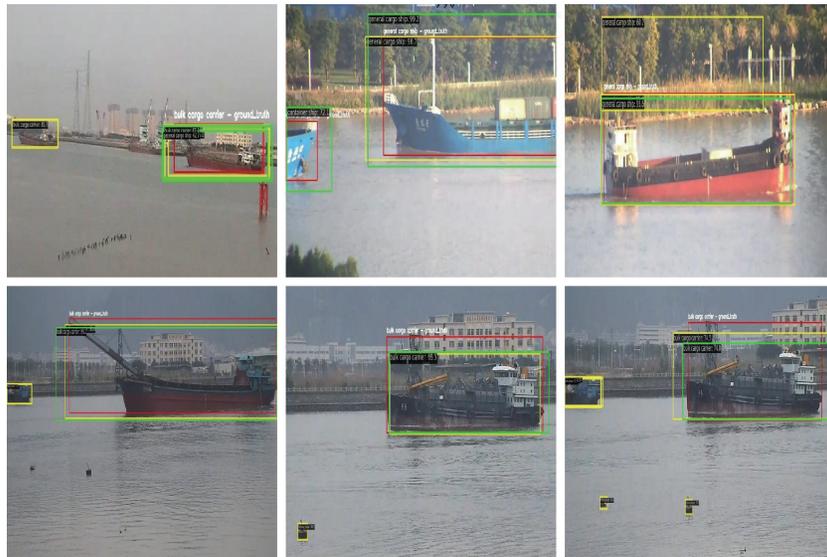


图 6 Seaship7000 数据集检测结果对比, 红色为真实标签, 绿色为本文的模型 FDNet, 黄色为 Faster-RCNN

Fig. 6 Comparison of detection results on the Seaship7000 dataset, with red indicating true labels, green representing our FDNet model, and yellow denoting Faster-RCNN

为了验证本文模型的泛化能力, 本文从互联网上选取了一些真实图片进行实验, 置信度阈值为 0.6, 检测结果如图 7 所示。在真实图像上, SSD 在多目标的复杂背景下会出现漏检(如图 7 中第 4 张图片)。Libra-RCNN^[32]、Faster-RCNN^[12]在面对船舶表观不清晰时会出现不同程度的漏检和误检(如图 7 中第 1 张、第 3 张和第 5 张图)。本文的模型相较于其他三个模型降低了误检漏检的发生概率, 提升了检测的准确率。在一定程度上解决了

表观相似和背景干扰导致的误检漏检问题。

2.4 消融试验

为了验证 CFGB 和 LBDC 模块的有效性, 本文在基准模型上单独加入 CFGB 模块和 LBDC 模块, 并选择 ResNet 50 作为骨干网络。表 2 展示了在 Seaship7000 数据集上的实验结果。

基准模型在单独引入 LBDC 模块后, mAP、AP50 和 AP75 分别增长了 3.2%、1.0% 和 4.0%。这表明 LBDC 在提取船舶目标的边界信息上效果显

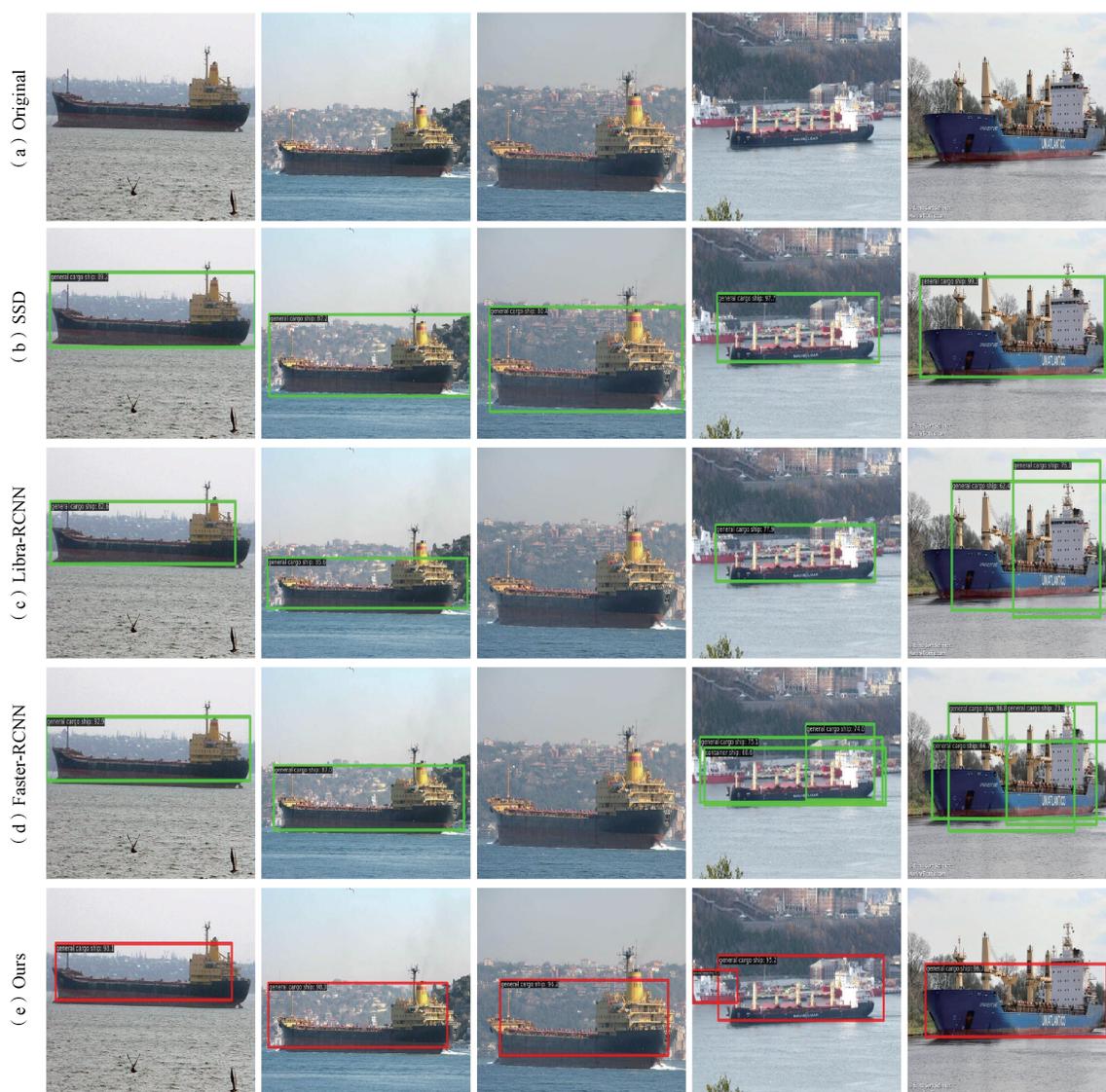


图7 真实图像检测对比结果

Fig. 7 Comparison results of real image detection

表2 消融实验

Table 2 Ablation experiment

Models	Backbone	分类分支(CFGB)	定位分支(LBDC)	mAP	AP50	AP75	APS	APM	APL
Baseline	ResNet50			64.4	95.3	75.7	11.2	45.4	65.6
	ResNet50		√	67.6	96.3	79.7	12.3	49.8	68.6
FDNet(Ours)	ResNet50	√		67.3	95.8	80.1	14.0	50.0	68.3
	ResNet50	√	√	67.8	95.8	80.6	24.2	47.8	68.8

著，有助于船舶目标定位。LBDC模块在APS、APM和APL上分别增长了1.1%、4.4%和3.0%，表明在面对尺度不同时定位依然准确有效。

基准模型单独引入CFGB后，mAP、AP50和AP75分别增长了2.9%、0.5%和4.4%，这表明CFGB模块在提取船舶目标的显著性类别信息上效

果优秀，有助于船舶表观细粒度的提取和挖掘。在APS、APM和APL上分别增长了2.8%、4.6%和2.7%，证明了CFGB模块在不同尺度上都可以对显著性特征和表观进行高效的提取和判别。

为了直观地看出CFGB和LBDC模块对各个类别船舶的检测效果，本文在表3中展示了散货船、

集装箱船、渔船、杂货船、矿砂船和客船六个类别的 mAP 值。可以看到, 对渔船和杂货船的 mAP 分别增长了 5.2% 和 1.2%, 说明 CFGB 和 LBDC 模块在一定程度上缓解了这两个类别中表观重合导致的漏检误检问题。

表 3 各个类别 mAP

Table 3 mAP for each category

Models	Backbone	散货船	集装箱船	渔船	杂货船	矿砂船	客船
Baseline	ResNet50	63.4	75.7	60.3	67.2	60.0	59.9
+ LBDC	ResNet50	67.8	77.7	64.0	68.6	62.9	64.4
+ CFGB	ResNet50	66.9	77.0	64.4	68.7	63.0	63.9
Ours	ResNet50	67.5	78.5	65.5	68.4	63.4	63.6

3 结束语

针对近海船舶目标中表观相似导致的误检问题, 本文提出了一种 FNet 网络。该网络包括类别细粒度分支和高效全维动态卷积定位分支。类别细粒度分支对表观细粒度差异进行精细学习, 挖掘和利用船舶目标的类别信息。高效全维动态卷积定位分支灵活的感知目标的边界信息, 增强定位框的准确性。最后, 本文采用 ResNet50 和 FPN 的 Faster R-CNN 构建了表观相似细粒度辨别网络, 并在公开数据集 Seaships7000 验证了 FNet 网络的有效性。值得注意的是, 近海船舶目标还存在着容易受湖面和天气影响的问题, 将在未来的研究中对此问题进行深入的探讨。另一方面, 本文的模型不能对光学视频进行实时处理, 这是因为本文的模型基于经典两阶段检测网络 Faster RCNN, 采用候选框的形式牺牲了推理速度, 这是目前二阶段检测算法的普遍问题。未来会致力于提高二阶段检测方法的推理速度, 使其满足实时性的要求。

参考文献

[1] 叶晨, 逯天洋, 肖涵灏, 等. 海事监控视频舰船目标检测研究现状与展望[J]. 中国图象图形学报, 2022, 27(7): 2078-2093.
YE C, LU T Y, XIAO Y H, et al. Maritime surveillance videos based ships detection algorithms: a survey[J]. Journal of Image and Graphics, 2022, 27(7): 2078-2093.

[2] 文莉莉, 孙苗, 邬满. 基于 Faster R-CNN 的海域监管预警方法[J]. 吉林大学学报(信息科学版), 2021, 39(4): 421-429.

WEN L, SUN M, WU M. Early warning method of sea area supervision based on faster R-CNN[J]. Journal of Jilin University (Information Science Edition), 2021, 39(4): 421-429.

[3] 王小虎. 海上船舶融合识别算法研究与改进[J]. 南方能源建设, 2023, 10(4): 131-137.
WANG X. Research and improvement of offshore ship fusion recognition algorithm[J]. Southern Energy Construction, 2023, 10(4): 131-137.

[4] 周薇娜, 刘露. 复杂场景下多尺度船舶实时检测方法[J]. 电信科学, 2022, 38(10): 67-78.
ZHOU W N, LIU L. A real-time detection method for multi-scale ships in complex scenes[J]. Telecommunications Science, 2022, 38(10): 67-78.

[5] MIN L, FAN Z, LV Q, et al. YOLO-DCTI: Small object detection in remote sensing base on contextual transformer enhancement[J]. Remote Sensing, 2023, 15(16): 3970.

[6] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(4): 1452-1459.

[7] 谢星星, 程焱, 姚艳清, 等. 动态特征融合的遥感图像目标检测[J]. 计算机学报, 2022, 45(4): 735-747.

[8] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv: 2018: 1804.02767.

[9] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.

[10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[11] 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.

[12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[13] WEI S, CHEN H, ZHU X, et al. Ship detection in remote sensing image based on faster r-cnn with dilated convolution[C]//2020 39th Chinese Control Conference

- (CCC). IEEE, 2020: 7148-7153.
- [14] GUO H, YANG X, WANG N, et al. A rotational libra R-CNN method for ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(8): 5772-5781.
- [15] DONG Y, CHEN F, HAN S, et al. Ship object detection of remote sensing image based on visual attention[J]. Remote Sensing, 2021, 13(16): 3192.
- [16] DONG C, LIU J, XU F, et al. Ship detection from optical remote sensing images using multi-scale analysis and Fourier HOG descriptor[J]. Remote Sensing, 2019, 11(13): 1529.
- [17] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]//International Conference on Learning Representations, 2020.
- [18] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
- [20] YANG B, BENDER G, LE Q V, et al. CondConv: conditionally parameterized convolutions for efficient inference[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019: 1307-1318.
- [21] CHEN Y, DAI X, LIU M, et al. Dynamic convolution: Attention over convolution kernels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11030-11039.
- [22] LI C, ZHOU A, YAO A. Omni-dimensional dynamic convolution[C]//International Conference on Learning Representations. 2021.
- [23] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Springer International Publishing, 2020: 213-229.
- [24] CHENG G, LI Q, WANG G, et al. SFRNet: Fine-grained oriented object recognition via separate feature refinement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-10.
- [25] WEI Q, HU X, HOU Q. Dynamic-YOLOv5: An improved aerial small object detector based on YOLOv5[C]//2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), IEEE, 2023: 679-683.
- [26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [27] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [28] SHAO Z, WU W, WANG Z, et al. Seaships: A large-scale precisely annotated dataset for ship detection[J]. IEEE Transactions on Multimedia, 2018, 20(10): 2593-2604.
- [29] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.
- [30] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2020: 2004.10934.
- [31] SONG G, LIU Y, WANG X. Revisiting the sibling head in object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11563-11572.
- [32] PANG J, CHEN K, SHI J, et al. Libra r-cnn: Towards balanced learning for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 821-830.
- [33] CHEN K, WANG J, PANG J, et al. MMDetection: Open mmlab detection toolbox and benchmark[J]. arXiv preprint arXiv: 2019: 1906.07155.
- [34] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.

[作者简介]

- 闵令通 1987年生, 副教授, 硕士生导师。
范子满 2000年生, 硕士研究生。
窦飞阳 2001年生, 硕士研究生。
吕勤毅 1990年生, 副教授, 硕士生导师。
李鑫 1984年生, 副教授, 硕士生导师。

(本文编辑: 潘三英)