

基于特征重聚焦网络的多尺度近岸舰船检测

刘荻¹, 张焱^{1*}, 赵琰², 石志广¹, 张景华¹, 张宇¹¹国防科技大学电子科学学院 ATR 重点实验室, 湖南 长沙 410073;²国防科技大学电子科学学院电子信息系统与复杂电磁环境效应国家重点实验室, 湖南 长沙 410073

摘要 针对监控视频中的多尺度近岸舰船检测问题,提出了一种基于特征重聚焦网络的舰船目标检测算法,设计了由多维特征聚合模块(MFAM)与注意力特征重构模块(AFRM)组成的特征重聚焦策略。其中,MFAM 基于输入的特征金字塔构建特征聚合块,进一步融合多尺度舰船不同层次特征的语义信息。AFRM 基于多分支空洞卷积以及通道与空间注意力机制提升网络对目标非局部信息的表征和对背景干扰的抑制,并构建了用于目标检测的特征重聚焦金字塔。在 Seaships7000 舰船公开数据集上的实验结果表明,相比其他算法,本算法对监控视频中多尺度近岸舰船的检测效果更好。

关键词 机器视觉; 卷积神经网络; 舰船目标检测; 特征重聚焦网络; 注意力机制

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/AOS202141.2215001

Multi-Scale Inshore Ship Detection Based on Feature Re-Focusing Network

Liu Di¹, Zhang Yan^{1*}, Zhao Yan², Shi Zhiguang¹, Zhang Jinghua¹, Zhang Yu¹¹National Key Laboratory of Science and Technology on Automatic Target Recognition, College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China;²State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China

Abstract Aiming at the problems of multi-scale inshore ship detection in surveillance videos, this paper proposes a ship target detection algorithm based on feature re-focusing network, and designs a feature re-focusing strategy, which consists of a multi-scale feature aggregation module (MFAM) and attention feature re-assignment module (AFRM). Specifically, MFAM fuses the semantic information of different levels of features of multi-scale ships by constructing a feature aggregation block based on the input feature pyramid. AFRM is composed of multi-branch dilated convolutions as well as channel and spatial attention mechanisms, which can improve the network's representation of target non-local information and suppressing interference of background, and a feature re-focusing pyramid is established for target detection. The experimental results on the Seaships7000 ship public data set show that compared with other algorithms, the algorithm has a better detection effect on multi-scale inshore ships in surveillance videos.

Key words machine vision; convolutional neural network; ship inshore detection; feature re-focusing network; attention mechanism

OCIS codes 150.0155; 150.1135; 100.4996

收稿日期: 2021-04-28; 修回日期: 2021-05-16; 录用日期: 2021-06-03

基金项目: 国家自然科学基金(62075239,61302145)

通信作者: *atrthreefire@sina.com

1 引言

对近岸舰船进行动态监测与准确识别,可以有效提升航道的利用率与港口的管理效率,具有重要的研究意义与应用价值^[1-5]。视频监控系统具有成本低廉、部署灵活及可实时成像的特点,在生产生活中得到了广泛应用,是社会治理、流量管控的重要手段^[6]。对监控视频中的近岸舰船进行检测可及时、准确地获取目标的细粒度位置信息,进一步提升海事监测、航道管理、港口调度的智能化水平。

目前,对监控视频中舰船目标的检测可分为基于特征建模的传统算法^[7-8]和基于语义信息的机器学习算法^[9-11]。传统算法主要依赖对背景与目标特征的准确建模。在背景建模方面,Wei 等^[12]提出了一种 AIVS3(Automated intelligent video surveillance system for ships),该系统基于图像的色彩信息,通过水平线检测、水域建模等算法从输入图像中提取可能存在舰船目标的区域。Arshad 等^[13]基于边缘纹理信息与形态学特征结合的算法提取舰船目标特征。在目标建模方面,Sullivan 等^[14]构建了特定类别舰船模板,通过分析模板与目标在频域的相关性对目标进行判别。Bao 等^[15]采用梯度直方图(HoG)特征对舰船目标进行判别。传统算法虽然可以通过构建与目标和背景相关的模型对舰船目标进行判别,但易受目标复杂结构、多样化外形尺寸以及周围环境、天气状况变化的影响;且算法复杂的模型构建与特定的参数约束条件,限制了其在复杂场景下的准确性与泛化能力。相比传统算法,基于卷积神经网络(CNN)的算法在近岸舰船目标检测任务中的性能更好。Shao 等^[16]基于 YOLOv2(You only look once, v2)提出了一种显著性感知卷积神经网络(SACNN),对深度语义、显著性图及海岸线等多种特征进行融合,构建了由粗到精的舰船目标检测机制。聂鑫等^[17]基于 YOLOv3 引入了预测框不确定性回归,并对损失函数进行改进,提升了算法对监控视频近岸舰船的检测能力。Nie 等^[18]根据大气散射模型及视觉成像理论,合成了近岸舰船在多种天气状况下的仿真影像,提升了算法在不同天气条件下对舰船目标的检测能力。基于 CNN 的近岸舰船检测算法^[16-19]虽然提升了对目标的检测能力,但其网络结构的设计未充分考虑目标自身的多尺度特性,对不同尺度近岸舰船目标的特征利用与判别能力有限。

针对上述问题,本文提出了一种基于特征重聚

焦网络(FRN)的监控视频多尺度近岸舰船检测算法。FRN 的核心为特征重聚焦策略(FRS),通过设计多维特征聚合模块(MFAM)与注意力特征重构模块(AFRM),实现对舰船语义特征的融合与再分配,进而自适应强化特征图对不同尺度目标语义信息的表征,提升网络对多尺度近岸舰船的检测能力。其中,MFAM 由多个卷积、池化、上采样操作组成,可将输入特征图不同层次的语义信息融合为特征聚合块(FAB),实现舰船浅层纹理与高层抽象信息的互补。AFRM 由多组具有不同膨胀率的空洞卷积^[20]、通道注意力模块(CAM)与空间注意力模块(SAM)组成,可提取目标的局部与全局信息,突出目标的判别性特征并弱化背景干扰。此外,FRN 还引入了 RefineDet^[21]由粗到精的锚点框回归机制,以准确判别目标。在 Seaships7000 监控视频舰船数据集上的测试结果表明,相比其他基于 CNN 的目标检测算法,如 Faster 区域卷积神经网络(Faster RCNN)^[22]、Libra RCNN^[23]、单步多框检测(SSD)^[24]、YOLOv3^[25]、RefineDet^[21]算法,FRN 算法对监控视频中近岸舰船的检测精度更高。

2 特征重聚焦网络

由于舰船目标种类不一且自身结构、所处位置复杂多样,监控视频中的近岸舰船存在不同目标尺度差异大、目标在视野中被截断、目标与目标相互遮挡、目标与周围背景混淆等情况,如图 1 所示。因此,充分融合并高效利用目标的多维度语义信息是对复杂场景中近岸舰船准确检测的关键。

2.1 整体结构

FRN 的整体结构如图 2 所示,该网络以 VGG-16(Visual geometry group)网络为基础,将原始全连接层(FC6、FC7)分别替换为卷积层(Conv6 与 Conv7),并在 Conv7 后新增由 Extra1 与 Extra2 构成的 Extra 卷积层,以提取近岸舰船的多尺度特征,其参数配置如表 1 所示。选取 VGG-16 网络第 3、第 4 个卷积组的第 3 层(Conv4_3、Conv5_3)、Conv7 以及 Extra2 的输出作为基础特征,构建包含特征图 P_2 、 P_3 、 P_4 、 P_5 (对应尺寸分别为 80 pixel×80 pixel、40 pixel×40 pixel、20 pixel×20 pixel、10 pixel×10 pixel)的特征金字塔网络(FPN),以实现为目标基础特征的初步提取与融合。然后,采用 FRS 对目标在 FPN 中的语义信息进行融合与提取。其中,来自 FPN 的不同尺度特征图经过 FRS 的 MFAM 构建 FAB,AFRM 以 FAB 为输入构建用于最终目标检

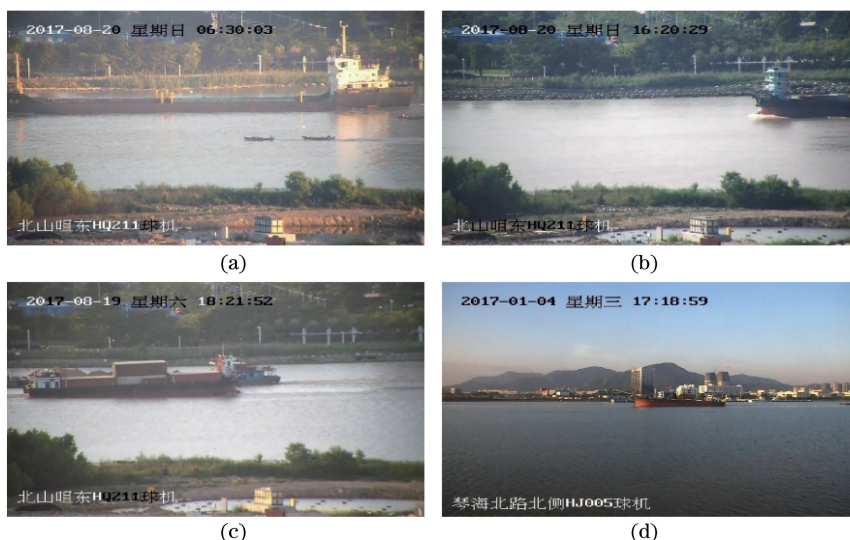


图 1 监控视频中的近岸舰船目标。(a)尺度差异较大;(b)舰船被截断;(c)舰船相互遮挡;(d)舰船与周围背景混淆
 Fig. 1 Inshore ship target in the surveillance video. (a) Size difference is large; (b) ship is cut off; (c) ships block each other; (d) ship is confused with the surrounding background

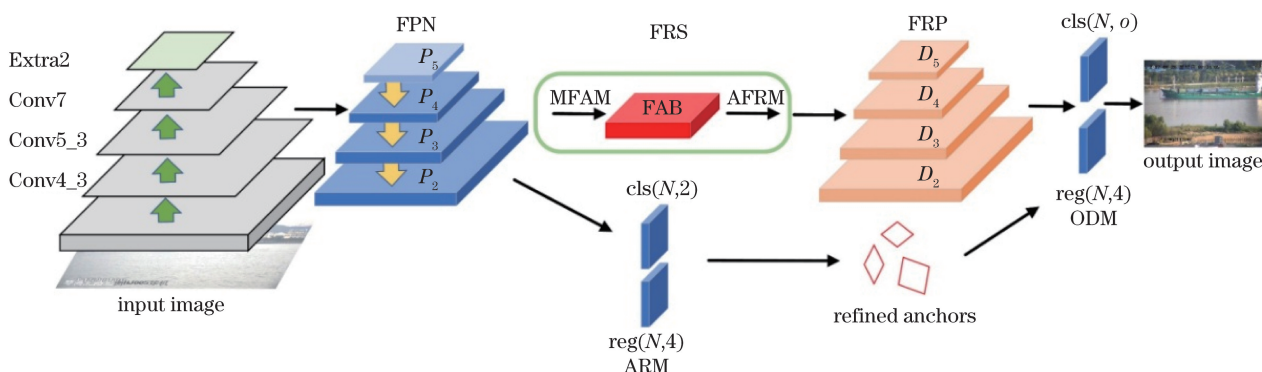


图 2 FRN 的结构

Fig. 2 Structure of the FRN

测的重构特征金字塔(包含特征层 D_2 、 D_3 、 D_4 、 D_5)。为进一步提升算法的性能,FRN 借鉴 RefineDet 构建了由锚点细化模块(ARM)与目标检测模块(ODM)组成的目标多级检测机制。区别于 RefineDet 将 ARM 应用于基础特征图,FRN 将 ARM 应用于 FPN 的特征图 P_2 、 P_3 、 P_4 、 P_5 ,以更精细的特征对锚点框进行初步分类(cls)与回归(reg),并生成精细化的锚点框(Refined anchors)。其中, N 为锚点框的数量, o 为输出的类别数。用 ODM 在重构特征金字塔上对该锚点框进行回归与判别,并通过非极大值抑制(NMS)、置信度约束等后处理操作,得到近岸舰船目标的最终检测结果。

2.2 特征重聚焦策略

2.2.1 多维特征聚合模块

FRS 由 MFAM 与 AFRM 组成。近岸舰船的多尺度特性导致目标语义信息在不同特征图上的分

布比较分散,文献[26]虽然通过融合基础特征图 C_{i+1} 与高维特征图 P_i 构建了精细化特征图 P_{i+1} ,从而增强高层特征图对目标浅层语义信息的表征,但其迭代化构建策略使特征图更偏向对相邻上层语义信息的提取,对相隔较远特征图中目标语义信息的表征能力有限,如图 3 所示。其中, k 、 s 、 p 分别为卷积核的尺寸、滑动步长及填充率,DeConv 为反卷积层,ReLU 为修正线性单元激活函数。此外,FPN 自顶向下的信息流动方式不能有效增强高层特征图中包含的目标结构化、局部化语义信息,一定程度限制了算法对较大尺寸目标的特征提取与判别能力。为充分融合目标在不同层的语义信息,FRS 设计了 MFAM,其特征融合方式如图 4 所示。该模块以输入 FPN 的特征图 P_3 作为基准,分别采用上采样倍数 $r=2$ 、 4 的上采样插值扩充特征图 P_5 、 P_4 的尺寸,用步长 $s=2$ 、卷积核尺寸 $k=3$ 、填充率 $p=2$ 的

表 1 基础特征提取网络的结构及参数
Table 1 Structure and parameter of basic feature extraction network

Name	Layer	Layer group	Filters number	Kernel size	Stride size	Padding size
VGG-16	Conv1_1	1	64	3	1	1
	Conv1_2					
	Conv2_1	2	128	3	1	1
	Conv2_2					
	Conv3_1	3	256	3	1	1
	Conv3_2					
	Conv3_3					
	Conv4_1	4	512	3	1	1
	Conv4_2					
	Conv4_3					
	Conv5_1	5	512	3	1	1
	Conv5_2					
	Conv5_3					
	Conv6	6	1024	3	1	1
	Conv7	7	1024	1	1	0
Extra	Extra1	8	256	1	1	0
	Extra2		512	3	2	1

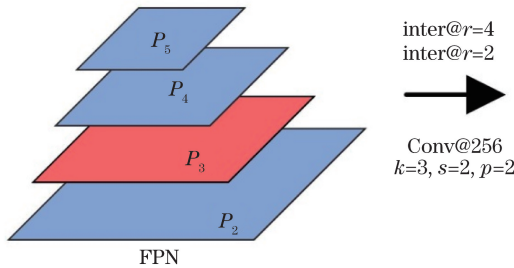


图 4 MFAM 的特征融合方式
Fig. 4 Feature fusion strategy of the MFAM

2.2.2 注意力特征重构模块

MFAM 构建的 FAB 虽然能对目标不同维度的语义信息进行自适应融合与表征,但缺乏有效的信息过滤机制,导致 FAB 不同通道、空间区域特征可能存在冗余。因此,FRS 设计了 AFRM,进一步对目标信息进行提取与筛选,并构建了用于检测的多尺度重构特征图 D_i , $i=2,3,4,5$,如图 5 所示。其中,长、宽、通道数分别为 c, h, w 的输入特征图 X_i 经膨胀率 R 为 1、2、3,卷积核尺寸为 3 的三路空洞卷积分支沿通道维度拼接,进一步得到表征目标局部与全局信息的特征图 S_i 。由于舰船周围背景复杂多样,特征融合也会引入较多的背景干扰信息。为突出目标特征,提升信息流动效率,AFRM 设计了并行 CAM 与 SAM,对 S_i 的特定通道及空间位

卷积操作对特征图 P_2 进行降采样,并将变换后的特征图沿通道维度叠加,得到通道数为 1024、尺寸为 $40 \text{ pixel} \times 40 \text{ pixel}$ 的叠加特征图 (C_{concat})。为提升叠加特征图的通道关联性,强化对多尺度目标语义信息的表征,该特征图依次经四个通道数为 256、卷积核尺寸 $k=3$ 的特征转换层 T_i ,得到对目标进行高维抽象与浅层纹理自适应表征的特征聚合块 X_i^{FAB} 。其中, $i=2,3,4,5$ 为与输入 FPN 特征图对应的特征聚合块编号。

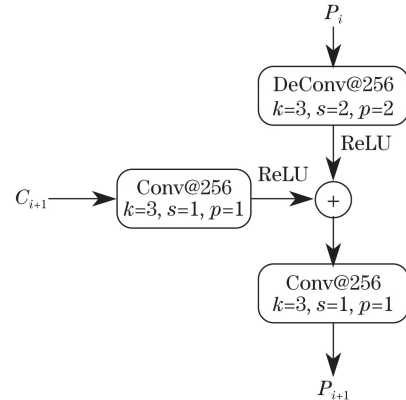


图 3 FPN 的特征融合方式
Fig. 3 Feature fusion strategy of the FPN

置的关键特征进行筛选,得到注意力特征图 A_i 。借鉴残差网络(ResNet)的思想,对特征图 A_i 与 FPN 中相同尺度特征图 P_i 构建残差连接(图 5 中的符号 \oplus),得到特征重聚焦后的精细化特征图 D_i ,以提升算法的优化效率。

1) 输入图像经过网络的前向传播后,舰船及周围背景的多维度特征被存储在输出特征图的不同通道。为提升特征图不同通道中关键信息的影响力,AFRM 设计了 CAM,其结构如图 6 所示。其中,输入特征 S_i 沿空间维分别经最大 $\text{Max}(\cdot)$ 与均值 $\text{Avg}(\cdot)$ 池化操作,得到与 S_i 通道数一致、空间尺度为 1×1 的池化特征,从而反映不同通道特征影响力的均值与极值水平。均值与极值特征经逐元素相加后由三层全连接层组成的多层感知器(MLP)进行

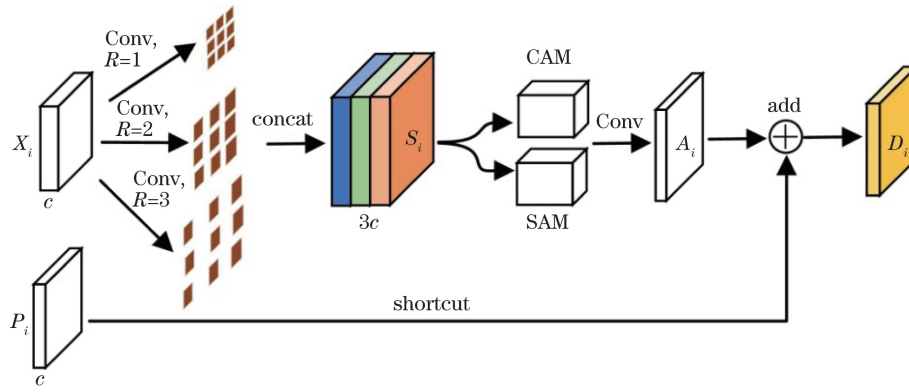


图 5 AFRM 的结构

Fig. 5 Structure of the AFRM

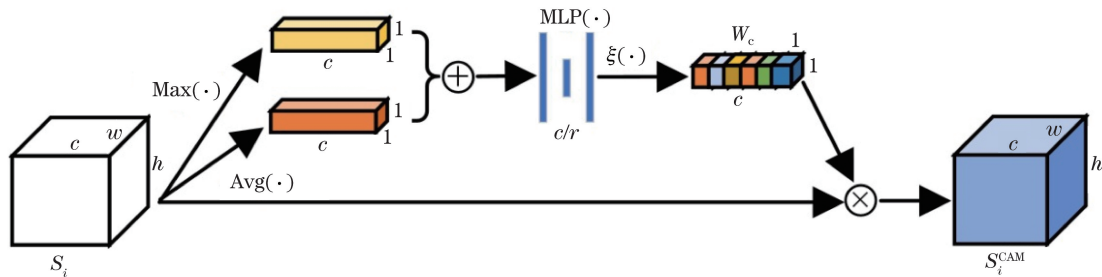


图 6 CAM 的结构

Fig. 6 Structure of the CAM

特征变换及 Sigmoid 激活函数 $\xi(\cdot)$ 的拉伸, 得到 $(0,1)$ 范围内的通道加权值 W_c 。将 W_c 与输入特征图 S_i 相乘, 即可得到加权后的特征图 S_i^{CAM} 。

2) 由于不同类别舰船的部件组成、轮廓结构存在一定的特异性, 对特征图空间维度语义信息进行筛选, 可凸显目标特定区域的识别性特征, 增强目标关键部位、结构的影响力。图 7 为 AFRM 设计的空间注意力模块。输入特征 S_i 沿通道维度分别经最

大池化 $\text{Max}(\cdot)$ 与均值池化 $\text{Avg}(\cdot)$ 操作并进行拼接, 获得通道数为 2, 长宽尺寸与输入特征一致的通道池化特征图, 从而反映网络对特征图不同区域的响应程度。然后, 经卷积核尺寸为 7×7 , 输出通道数为 1 的卷积 (Conv) 操作及 Sigmoid 激活函数 $\xi(\cdot)$ 的拉伸, 得到输出范围为 $(0,1)$ 的空间加权特征图 W_s 。用 W_s 对输入特征 S_i 沿空间维度进行加权, 得到空间注意力加权特征 S_i^{SAM} 。

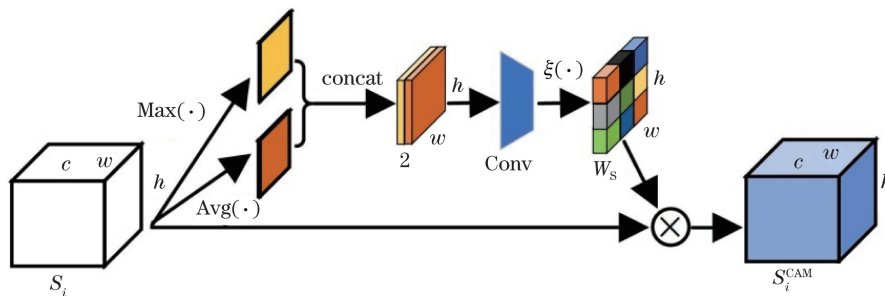


图 7 SAM 的结构

Fig. 7 Structure of the SAM

2.3 其他模块

监控视频中近岸舰船的分布比较稀疏, 仅有少量预定义锚点框与目标存在较多重叠, 导致正负样本不均衡, 且近岸舰船的多尺度、大长宽比特性也增加了算法对目标的预测难度。为提升算法对目标预

测的精确度, FRN 借鉴 RefineDet 算法引入了由 ARM 与 ODM 组成的目标精细化检测策略。为充分利用目标的多维度语义特征, FRN 以 FPN 作为 ARM 的输入对预定义锚点框进行初步回归, 并滤除低于预设阈值的锚点框。ODM 则基于特征重聚

焦金字塔进一步对预筛选的锚点框进行回归,得到目标最终的位置及类别预测结果。

2.4 损失函数

FRN 的损失函数 L 由 ARM 损失 L_{ARM} 与 ODM 损失 L_{ODM} 两部分组成,可表示为

$$L = L_{\text{ARM}} + L_{\text{ODM}}, \quad (1)$$

$$L(x, c_{\text{ls}}, l, g) = [L_{\text{conf}}(x, c_{\text{ls}}) + \alpha L_{\text{loc}}(x, l, g)] / N_c, \quad (2)$$

式中, L_{loc} 为目标预测位置的损失, L_{conf} 为目标置信度损失, x 为网络的预测值, c_{ls} 、 l 分别为算法预测的目标类别与位置, g 为目标的真值参数, N_c 为真值目标与预设锚点框满足交叠率阈值时正样本锚点的数目, α 为不同损失值的权重调节因子, 实验中设置为 1.0。

在优化过程中, 采用交叉熵函数对算法的目标类别预测流程进行优化, 可表示为

$$L_{\text{conf}}(x, c) = - \sum_{t \in X_{\text{Pos}}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{t \in X_{\text{Neg}}} \log(\hat{c}_i^0), \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}, \quad (3)$$

式中, X_{Pos} 为正样本的数目, X_{Neg} 为负样本的数目。当第 t 个目标预测值与第 j 个真值匹配时, x_{ij}^p 为 1;

否则, x_{ij}^p 为 0。 \hat{c}_i^p 、 \hat{c}_i^0 分别为第 t 个目标为类别 p 和背景的概率。用 Smooth L1 函数对算法预测的目标位置进行优化, 可表示为

$$L_{\text{loc}}(x, l, g) = \sum_{t \in X_{\text{Pos}}} \sum_{m \in \{c_x, c_y, w, h\}} x_{ij}^k L_1(l_t^m - \hat{g}_j^m), \quad (4)$$

式中, l_t^m 、 \hat{g}_j^m 分别为算法预测值与真值的参数化坐标, c_x 、 c_y 、 w 、 h 分别为参数化的中心点二维坐标及预测框的宽、高。

3 实验结果与分析

3.1 数据集及超参数设置

采用 Seaships7000 公开数据集^[27]验证 FRN 对监控视频中近岸舰船的检测性能。Seaships7000 数据集包含 7000 幅分辨率为 1920 pixel × 1080 pixel 的可见光图像, 均由部署在珠海横琴岛周围多个地点的监控摄像头采集。图 8 为数据集中标注的 6 类舰船, 包括矿砂船 (Ore carrier)、散装货船 (Bulk cargo carrier)、杂货船 (General cargo carrier)、集装箱船 (Container ship)、渔船 (Fishing ship) 及客船 (Passenger ship)。



图 8 Seaships7000 数据集中的舰船目标。(a) 矿砂船; (b) 散装货船; (c) 杂货船; (d) 集装箱船; (e) 渔船; (f) 客船
Fig. 8 Ship target in the Seaships7000 data set. (a) Ore carrier; (b) bulk cargo carrier; (c) general cargo carrier; (d) container ship; (e) fishing ship; (f) passenger ship

不同舰船的尺寸分布如图 9(a) 所示, 其中, 渔船与集装箱船的尺寸最小, 杂货船和客船次之, 矿砂船与散装客船的尺寸最大。不同类别舰船的数量占比如图 9(b) 所示。为充分验证本算法的性能, 随机选取 1750、1750 及 3500 幅不同的图像作为训练集、验证集以及测试集, 如图 9(c) 所示。

实验设置的超参数如表 2 所示, 其中, SGD 为随

机梯度下降法。在训练过程中, 采用随机镜像翻转、区域裁剪、缩放、对比度与亮度变换等策略对训练集样本进行扩充, 每隔 5 个 Epoch 对模型进行保存并通过验证集评估模型性能, 选取在验证集上性能最优的模型进行最终测试。在测试阶段, 设置的目标置信度阈值及非极大值抑制阈值分别为 0.5 与 0.45。算法采用 Pytorch 1.7.1 框架搭建, 在装有 Ubuntu18.04

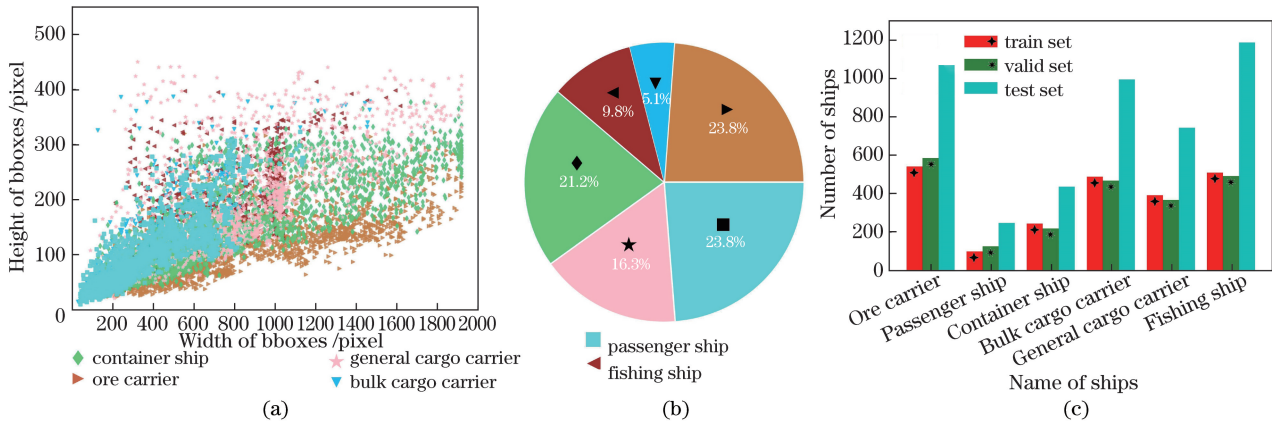


图 9 Seaships7000 数据集的统计。(a) 尺寸分布;(b) 数量占比;(c) 数据集的划分

Fig. 9 Statistics of the Seaships7000 data set. (a) Size distribution; (b) percentage of quantity; (c) division of the data set

表 2 实验设置的超参数

Table 2 Hyperparameters of experimental settings

Hyperparameter	Value
Anchor box sizes	32,64,128,256
Anchor box scales	1,1.414,0.707
Epoch	100
Optimizer	SGD
Learning rate	0.001
Learning rate decay steps	50,75
Learning rate decay rate	0.1
Weight decay	5×10^{-4}
Mini-batch	6

操作系统的 GPU 工作站上完成测试。

3.2 评价指标

为充分验证模型对近岸舰船的检测能力,选取 MS COCO(Microsoft common objects in context) 的平均精度(AP)对不同算法进行评测,统计并绘制了检测结果与真值在特定交并比(IoU)阈值时,算法对不同类别舰船检测的 AP 及精确率-召回率(PR)曲线,并用帧率(FPS)对算法的测试速度进行评估。AP 可表示为

$$X_{AP} = \frac{1}{N} \frac{1}{10} \sum_{c=1}^N \sum_{X_{IoU}=0.5}^{0.95} p_c^{IoU}, \quad (5)$$

表 3 FRS 内部模块对 FRN 检测性能的影响

Table 3 Effect of FRS internal modules on FRN detection performance

Algorithm	AP	AP ^{0.5}	AP ^{0.75}	AP ^s	AP ^m	AP ^l	FPS
FRN	0.692	0.948	0.830	0.122	0.525	0.703	25
FRN (w/o MFAM)	0.684	0.945	0.823	0.096	0.550	0.695	31
FRN (w/o AFRM)	0.686	0.952	0.826	0.092	0.536	0.699	37
FRN (w/o CAM)	0.681	0.942	0.818	0.110	0.525	0.691	29
FRN (w/o SAM)	0.682	0.946	0.824	0.153	0.508	0.693	27
FRN (w/o SAM&CAM)	0.678	0.944	0.824	0.097	0.517	0.688	34
FRN(w/o ARM)	0.683	0.945	0.824	0.101	0.487	0.695	27
RefineDet	0.674	0.946	0.815	0.134	0.505	0.685	47

式中, N 为目标的类别数, p_c^{IoU} 为算法对第 c 类目标在特定 IoU 阈值时的 AP, X_{IoU} 为 IoU 阈值。若 AP 表示以 0.05 为步长, IoU 取 0.5~0.95 之间的特定阈值时,算法对所有类别目标 AP 的均值,其值越高,表明算法对所有类目标的综合检测识别性能越优, AP^{0.5}、AP^{0.75} 分别表示 IoU 为 0.5、0.75 时算法对所有类目标检测的平均精度均值。AP^{0.75} 的值越高,表明算法在更严格的交叠率约束条件下,对目标的判别与定位准确度越高。AP^s、AP^m、AP^l 分别为算法对输入图像中边长尺寸小于 32 pixel、介于 32~96 pixel 之间、大于 96 pixel 的目标在不同 IoU 阈值条件下的 AP 均值。FPS 可表示为

$$X_{FPS} = 1/t, \quad (6)$$

式中, t 为算法处理一幅图像的平均耗时。

3.3 实验结果与分析

3.3.1 特征重聚焦策略探究

FRS 内部模块对算法所有类别和特定类别舰船目标检测性能的影响如表 3、表 4 所示。其中, w/o 表示不包含对应模块。由于不含 FRS 的 FRN 与 RefineDet 在网络结构、数据流处理等方面相似,因此,实验用 RefineDet 作为对比基准。

表 4 FRS 内部模块对不同类别舰船检测性能的影响

Table 4 Effects of FRS internal modules on the detection performance of different types of ships

Algorithm	Passenger ship	General cargo carrier	Fishing ship	Bulk cargo carrier	Ore carrier	Container ship
FRN	0.654	0.705	0.646	0.696	0.683	0.768
FRN (w/o MFAM)	0.642	0.718	0.633	0.697	0.661	0.752
FRN (w/o AFRM)	0.656	0.701	0.640	0.687	0.684	0.749
FRN (w/o CAM)	0.632	0.693	0.644	0.681	0.678	0.758
FRN (w/o SAM)	0.642	0.703	0.640	0.680	0.680	0.746
FRN (w/o SAM&CAM)	0.637	0.691	0.638	0.680	0.671	0.748
FRN (w/o ARM)	0.641	0.698	0.645	0.696	0.658	0.761
RefineDet	0.622	0.702	0.631	0.678	0.658	0.749

从表 3 中可以发现, MFAM 与 AFRM 均能提升 FRN 对舰船目标的综合检测能力, 相比基准算法, 两者单独作用于 FRN 时, 均能提升算法对舰船检测的定位精准度及对中等尺寸舰船的检测能力。由于中等尺寸舰船目标在不同层次特征图上的分布比较广泛, MFAM 可对目标特征进行深度汇聚, AFRM 引入的多分支空洞及注意力模块可充分提取并过滤目标不同区域的特征, 从而提升算法对中等尺寸舰船语义信息的提取与表征, 强化局部、浅层纹理信息对目标定位的贡献。此外, CAM 对算法目标定位精准度及对大尺寸目标检测性能的影响比较突出, SAM 则对算法定位中等尺寸目标精准度的影响更明显。仅有多尺度空洞卷积时, 算法对中等尺寸目标的检测能力虽然有所提升, 但其内部较多的冗余特征弱化了对较小尺寸目标的检测能力。采用 ARM 在 FPN 特征图上对基础锚点框进行回归时, 可充分利用浅层语义特征图中的高维抽象特征, 提升算法对中小尺寸目标的检测能力。从表 4 可以发现, 相比基准算法, 采用 FRS 后, FRN 对不同类别舰船目标的检测性能均有普遍提升, 对客船与矿砂船的检测性能提升最显著, 分别提升了 0.032 和 0.025, 这表明 FRS 可有效提升 FRN 对不同类舰船的特征提取与判别能力。MFAM 与 AFRM 均能保证算法对舰船目标检测的稳定性, 未引入 MFAM 或 AFRM 时, 算法对特定舰船(如矿砂船、集装箱船)的检测能力略有降低, 对其他类舰船依然能保持较高的检测精度。此外, 通道及空间注意力机制均能有效提升算法对舰船目标的整体检测能力, 未引入 CAM 或 SAM 时, 算法对特定类别舰船(如客船、散装货船、矿砂船以及集装箱船)的检测性能有明显下降, 这表明对特征在空间以及通道维度的筛选与优化, 可有效提升算法对目标的检测能力。

为进一步探究 FRS 应用前后算法对目标的表

征差异, 分别绘制了不同 IoU 及多目标条件下, FRS 输入特征图 P_2, P_3, P_4, P_5 与输出特征图 D_2, D_3, D_4, D_5 的可视化效果, 并将其与基准算法的检测结果进行对比, 如图 10 所示。特征图可视化的灰度值可表示网络对区域内语义信息的关注程度, 由于不同特征图与目标几何尺寸的匹配程度和对目标语义信息的表征存在差异, 特征图的可视化效果也各不相同。可以发现, 相比其他特征图, 本算法中 P_3, D_3 特征图对目标语义信息的表征更清晰。对比发现: 1) FRN 对重叠且存在截断的舰船目标判别能力更强。如舰船语义信息在基准网络 P_3 特征图中分布较为发散且不同舰船特征混叠严重(两艘舰船所在区域的灰度值较低且区分度较弱), 而 FRS 输出的 D_3 特征图对目标主体区域的响应值更高且响应区域更集中(D_3 中箭头所指的区域)。原因是 FRS 对目标多维度、局部与全局特征的充分融合, 使不同维度特征间的相关性与互补性更强, 从而提升了算法对目标信息的综合表征能力。2) FRN 受背景信息的干扰程度更低。如图中背景区域(区域 A、B、C、D)的灰度值在 FRS 输出的 D_3 特征图中比基准网络 P_3 更低。原因是 FRS 内部注意力机制(CAM 与 SAM)的存在, 进一步抑制了特征图不同通道与不同区域的背景信息, 降低了背景信息对算法的干扰。

3.3.2 不同算法的对比

表 5 为 FRN 与基于 CNN 的经典目标检测算法对舰船目标的检测结果。可以发现, 相比 Faster RCNN、FPN、Libra RCNN 等两阶段算法, FRN 的 AP 分别提升了 0.039、0.019、0.009, $AP^{0.75}$ 分别提升了 0.069、0.027、0.017, AP^s 分别提升了 0.020、0.040、0.050, AP^m 分别提升了 0.102、0.053、0.075。得益于强大的特征聚合能力, 相比 Faster RCNN、FPN, FRN 可准确感知目标不同层次的语义信息,

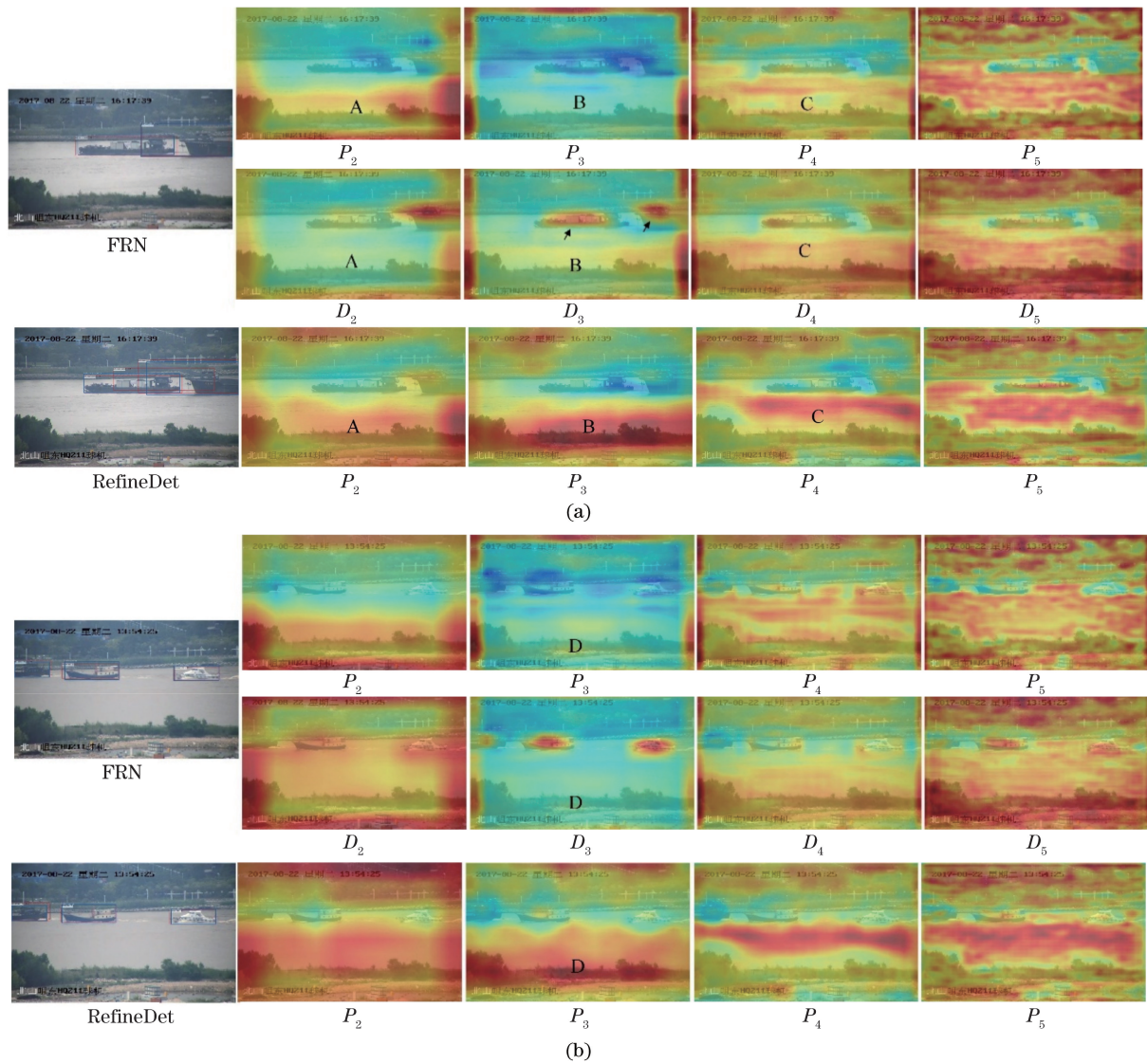


图 10 FRN 与基准算法的特征图可视化。(a)舰船相互重叠;(b)多舰船

Fig. 10 Visualization of feature maps of FRN and benchmark algorithms. (a) Overlapping ships; (b) multiple ships

表 5 不同算法对近岸舰船目标的检测结果

Table 5 Detection results of different algorithms on inshore ship targets

Algorithm	Type	AP	AP ^{0.5}	AP ^{0.75}	AP ^s	AP ^m	AP ^l	FPS
Faster RCNN	two-stage algorithm	0.653	0.919	0.761	0.102	0.423	0.670	6
FPN		0.673	0.941	0.803	0.082	0.472	0.688	16
Libra RCNN		0.683	0.940	0.813	0.072	0.450	0.698	15
YOLOv3	one-stage algorithm	0.509	0.782	0.601	0.122	0.310	0.517	52
RetinaNet		0.659	0.899	0.759	0.201	0.403	0.676	17
FCOS		0.573	0.830	0.673	0.093	0.317	0.585	18
SSD		0.663	0.906	0.793	0.003	0.422	0.687	90
RefineDet		0.674	0.946	0.815	0.134	0.505	0.685	47
Ours		0.692	0.948	0.830	0.122	0.525	0.703	25

提升了算法对中小尺寸舰船的检测能力。Libra RCNN 虽然也实现了对特征的聚合与再分配,但缺乏高效的特征筛选机制,导致目标与目标、目标与背

景的语义信息混淆度较高,限制了其对中小尺寸舰船的检测能力。此外,由于未引入区域建议机制,相比两阶段算法,FRN 的检测速度优势显著,PFS 达

到 25; 相比 YOLOv3、FCOS (Fully convolutional one-stage detector)、RetinaNet、SSD、RefineDet 等单阶段检测算法, FRN 对近岸多尺度舰船的检测优势显著, AP 分别提升了 0.183、0.119、0.033、0.029、0.018。基于简单的特征金字塔结构和有限的多尺度舰船目标特征甄别能力, YOLOv3 和 RetinaNet 仅提升了对小尺寸舰船的检测能力, 对大中型舰船的检测能力提升有限。SSD 的多层次特征图分散了不同尺寸舰船的特征, 弱化了因特征融合引入的背景干扰, 但也削弱了浅层特征图对目标抽象信息的表征能力, 使算法对小尺寸舰船的检测能力有所降低。FRN 可对提取的目标层次化语义信息进一步融合并筛选, 提升算法对目标显著性特征的表征

并弱化背景干扰, 从而提升对多尺度舰船目标的检测能力。在检测速度方面, 额外设计的模块增加了 FRN 的参数量与计算量, 使其检测速度相比 SSD、YOLOv3 等算法有显著降低。

在 IoU 阈值为 0.5~0.95、0.5 以及 0.75 时, 不同算法检测结果的 PR 曲线如图 11 所示。可以发现, 除 FCOS 与 YOLOv3 算法外, FRN 与其他算法的 PR 曲线在召回率低于 0.5 时存在较多重叠, 不同算法对目标的检测能力差异较小。但随着召回率的提升, 除 Libra RCNN 与 FRN 的曲线存在一定重叠外, 其他算法的曲线均有明显下降趋势, 对目标的检测精确率有大幅降低, 这表明本算法在较高召回率条件下对近岸舰船的综合检测能力更优。

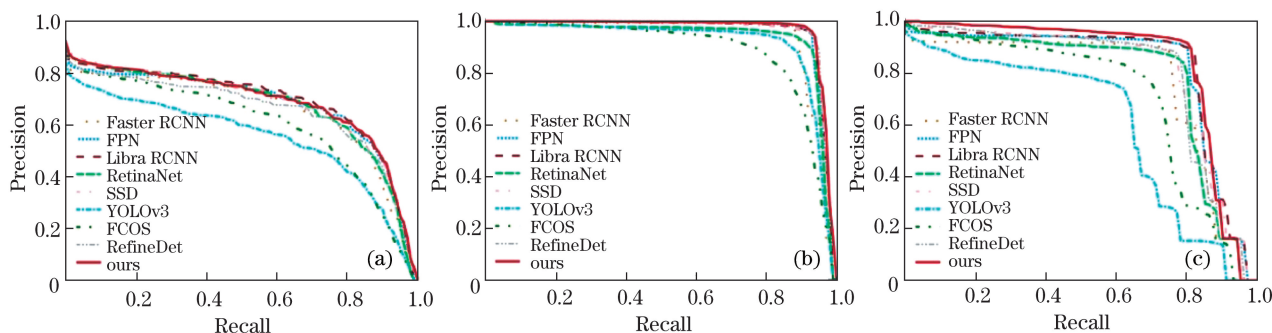


图 11 不同 IoU 时算法的 PR 曲线。(a) IoU 的取值范围为 0.5~0.95; (b) IoU 为 0.5; (c) IoU 为 0.75

Fig. 11 PR curves of the algorithm with different IoU. (a) Value range of IoU is 0.5~0.95; (b) IoU is 0.5; (c) IoU is 0.75

本算法与其他算法对 6 类舰船检测的 AP 值及 PR 曲线分别如表 6 与图 12 所示。可以发现, FRN 算法对类别样本不平衡、尺寸差异显著的目标适应性更强。Faster RCNN 算法对集装箱船检测的 AP 值为 0.771, 对客船、渔船检测的 AP 值仅为 0.591 与 0.582; FCOS 算法对集装箱船检测的 AP 值为

0.731, 对客船检测结果的 AP 值仅有 0.463; 相比其他算法, FRN 算法对 6 类舰船目标的整体检测性能更优; 且对渔船、散装货船以及矿砂船的检测性能更突出, AP 值分别为 0.646、0.696 以及 0.683。对舰船目标特异性结构的准确提取与判别是评价算法检测性能的关键, 由于杂货船与集装箱船的几何形

表 6 不同算法对近岸舰船目标的 AP

Table 6 AP of different algorithms on inshore ship targets

Algorithm	Type	Passenger	General cargo	Fishing	Bulk cargo	Ore	Container
		ship	carrier	ship	carrier	carrier	ship
Faster RCNN	two-stage algorithm	0.591	0.690	0.582	0.644	0.638	0.771
FPN		0.637	0.699	0.614	0.677	0.643	0.769
Libra RCNN		0.638	0.716	0.623	0.686	0.661	0.775
YOLOv3	one-stage algorithm	0.444	0.543	0.443	0.467	0.486	0.674
RetinaNet		0.665	0.685	0.617	0.618	0.594	0.773
FCOS		0.463	0.625	0.533	0.549	0.537	0.731
SSD		0.620	0.712	0.582	0.667	0.628	0.771
RefineDet		0.622	0.702	0.631	0.678	0.658	0.749
Ours (FRN)		0.654	0.705	0.646	0.696	0.683	0.768

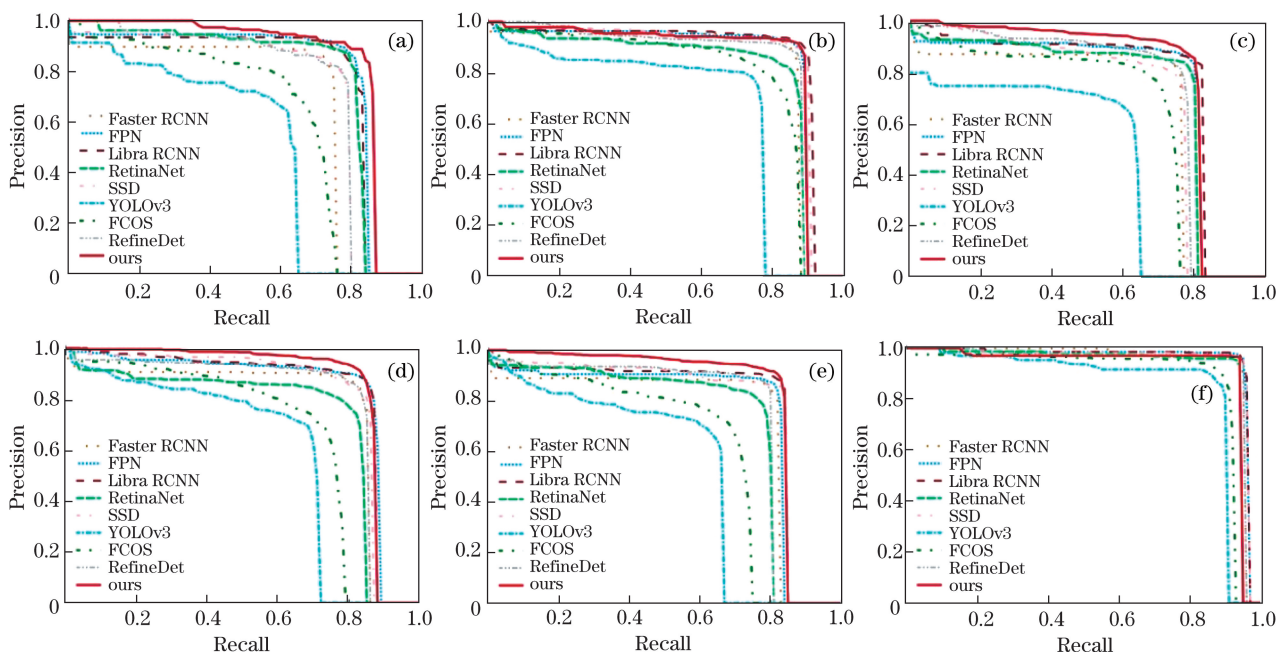


图 12 不同算法对不同类别舰船的 PR 曲线。(a)客船;(b)杂货船;(c)渔船;(d)散装货船;
(e)矿砂船;(f)集装箱船

Fig. 12 PR curves of different algorithms for different types of ship. (a) Passenger ship; (b) general cargo carrier;
(c) fishing ship; (d) bulk cargo carrier; (e) ore carrier; (f) container ship

状比较规整,不同算法对该类目标的特征判别能力整体更优,而散装货船和矿砂船的船首均带有尖锐且细长的吊机配件,船体长宽比较大且目标与背景混淆度较高,进一步增加了算法对目标检测的难度。图 12 中不同算法除对集装箱船检测结果的 PR 曲线差异较小外,对其他类别舰船检测结果的 PR 曲线均有明显差异。而本算法对不同类别舰船的检测结果均具有一定优势,这充分验证了本算法对多尺度、多类别近岸舰船优良的检测能力。

不同算法对多种复杂环境下舰船目标的检测结果如图 13 所示。可以发现,图 13(a)中多艘舰船相互重叠,Faster RCNN、Libra RCNN、RetinaNet 算法对舰船定位存在偏差,检测结果存在虚警与漏检等情况,FRN 算法对密集航迹的舰船目标定位更准确。图 13(b)中岸边矿砂对矿砂船检测造成干扰,FPN 和 RefineDet 算法均将 A 区域的砂石连同河岸检测为舰船,Faster RCNN、Libra RCNN、YOLOv3 算法对目标定位不准确,FRN 算法可准确判别背景干扰与舰船目标。图 13(c)中的舰船尺度较小,Faster RCNN、FPN、RetinaNet、FCOS 以及 SSD 算法均将 A 处船只漏检,YOLOv3 算法在 B 处存在虚警,RefineDet 算法对 C 处舰船定位存

在偏差,FRN 算法则强化了低维特征图对目标高维语义信息的表征,从而提升了对小目标渔船的检测能力。图 13(d)中光照不足导致舰船目标结构模糊度高、目标背景区分度低,FPN、RetinaNet、Libra RCNN、YOLOv3、FCOS 以及 RefineDet 算法均存在检测虚警,而 FRN 算法高效的特征过滤机制,提升了其对目标纹理、语义信息的表征,实现了对低照度条件下近岸舰船的准确检测。

4 结 论

针对监控视频中的多尺度近岸舰船检测问题,提出了一种以 FRS 为核心的 FRN。FRS 由 MFAM 与 AFRM 组成,其中, MFAM 以多尺度特征图构建 FAB,实现了对近岸舰船不同维度语义信息的自适应聚合;AFRM 则设计了多组空洞卷积并结合通道与空间注意力机制,实现了对 FAB 中目标非局部特征的精细化提取与过滤并抑制了背景干扰。在 Seaships7000 数据集上的实验结果表明,相比其他基于 CNN 的目标检测算法,本算法对监控视频中近岸舰船的检测精度更高。下一步还将研究监控视频中近岸舰船目标的成像特性,对网络结构进行优化升级,提升算法对监控视频中近岸舰船的检测能力。

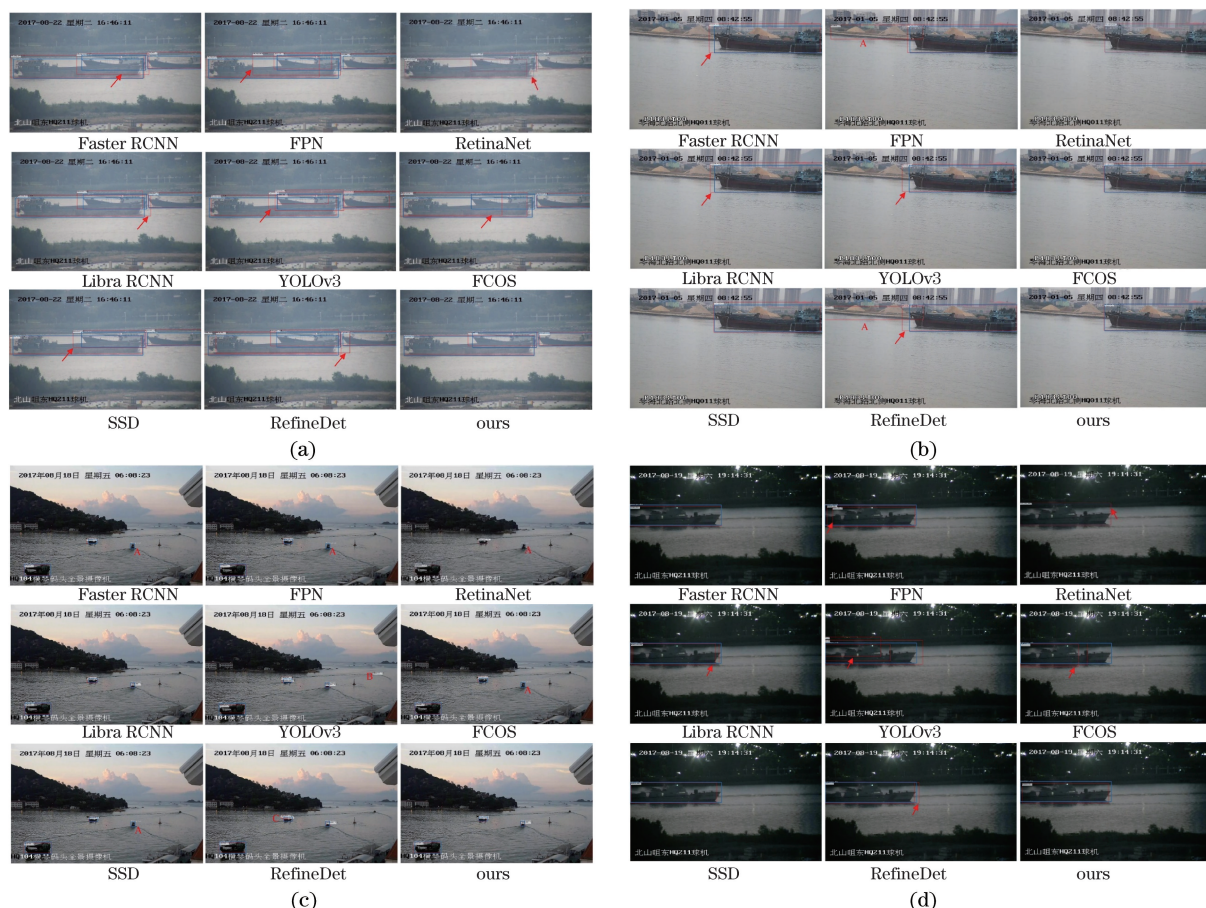


图 13 不同算法对近岸舰船目标的检测结果。(a)多目标重叠场景;(b)目标与背景混淆场景;
(c)小尺寸目标场景;(d)光照不足场景

Fig. 13 Detection results of different algorithms on inshore ship targets. (a) Multi-target overlapping scene;
(b) target and background confuse the scene; (c) small size target scene; (d) underlit scene

参 考 文 献

- [1] Wu Z. Research on detection, recognition and tracking of channel ships under dynamic background based on deep learning [D]. Yichang: China Three Gorges University, 2019.
吴喆. 基于深度学习的动态背景下航道船舶检测识别与跟踪研究[D]. 宜昌: 三峡大学, 2019.
- [2] Song M Z, Qu H S, Jin G. Weak ship target detection of noisy optical remote sensing image on sea surface [J]. Acta Optica Sinica, 2017, 37 (10): 1011004.
宋明珠, 曲宏松, 金光. 含噪光学遥感图像海面弱小舰船目标检测 [J]. 光学学报, 2017, 37 (10): 1011004.
- [3] Zhang Y, Zhang Y, Shi Z G, et al. Rotationally unconstrained region proposals for ship target segmentation in optical remote sensing [J]. IEEE Access, 2019, 7: 87049-87058.
- [4] Li J Y, Li X R, Zhao L Y. Docked ship detection based on edge line analysis and aggregation channel features [J]. Acta Optica Sinica, 2019, 39(8): 0815004.
- [5] 黎经文, 厉小润, 赵辽英. 基于边缘线分析与聚合通道特征的港口舰船检测 [J]. 光学学报, 2019, 39 (8): 0815004.
- [5] Wang W S, Huang M, Li T J, et al. Saliency detection for ship targets on four-band multi-spectral remote sensing images [J]. Acta Optica Sinica, 2020, 40(17): 1728001.
王文胜, 黄民, 李天剑, 等. 四波段多光谱遥感图像的船舶目标显著性检测 [J]. 光学学报, 2020, 40 (17): 1728001.
- [6] Popik A, Zaniewicz G, Wawrzyniak N. On-water video surveillance: data management for a ship identification system [J]. Zeszyty Naukowe Akademii Morskiej W Szczecinie, 2019, 60(132): 56-63.
- [7] Prasad D K, Rajan D, Rachmawati L, et al. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(8): 1993-2016.
- [8] Chen Y T, Li Y Y, Yao T T. Ship detection from remote sensing image under complex sea conditions [J].

- Laser & Optoelectronics Progress, 2019, 56(18): 181007.
陈彦彤, 李雨阳, 姚婷婷. 复杂海况下遥感图像舰船目标检测方法研究[J]. 激光与光电子学进展, 2019, 56(18): 181007.
- [9] Spraul R, Sommer L, Schumann A. A comprehensive analysis of modern object detection methods for maritime vessel detection[C]//Artificial Intelligence and Machine Learning in Defense Applications II. International Society for Optics and Photonics, 2020, 11543: 1154305.
- [10] Xu Z J, Ding Y. Ship object detection of remote sensing images based on adaptive rotation region proposal network[J]. Laser & Optoelectronics Progress, 2020, 57(24): 242805.
徐志京, 丁莹. 自适应旋转区域生成网络的遥感图像舰船目标检测[J]. 激光与光电子学进展, 2020, 57(24): 242805.
- [11] Liu Y Q, Feng J K, Xing B W, et al. Water surface object detection based on deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181502.
刘雨青, 冯俊凯, 邢博闻, 等. 基于深度学习的水面目标检测[J]. 激光与光电子学进展, 2020, 57(18): 181502.
- [12] Wei H, Nguyen H, Ramu P, et al. Automated intelligent video surveillance system for ships [J]. Proceedings of SPIE, 2009, 7306: 73061N.
- [13] Arshad N, Moon K S, Kim J N. An adaptive moving ship detection and tracking based on edge information and morphological operations [J]. Proceedings of SPIE, 2011, 8285: 82851X.
- [14] Sullivan M D R, Shah M. Visual surveillance in maritime port facilities [J]. Proceedings of SPIE, 2008, 6978: 697811.
- [15] Bao X F, Zinger S, Wijnhoven R, et al. Ship detection in port surveillance based on context and motion saliency analysis [J]. Proceedings of SPIE, 2013, 8663: 86630D.
- [16] Shao Z F, Wang L G, Wang Z Y, et al. Saliency-aware convolution neural network for ship detection in surveillance video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(3): 781-794.
- [17] Nie X, Liu W, Wu W. Ship detection based on enhanced YOLOv3 under complex environments[J]. Journal of Computer Applications, 2020, 40(9): 2561-2570.
聂鑫, 刘文, 吴巍. 复杂场景下基于增强 YOLOv3 的船舶目标检测[J]. 计算机应用, 2020, 40(9): 2561-2570.
- [18] Nie X, Yang M F, Liu R W. Deep neural network-based robust ship detection under different weather conditions[C]//2019 IEEE Intelligent Transportation Systems Conference (ITSC), October 27-30, 2019, Auckland, New Zealand. New York: IEEE Press, 2019: 47-52.
- [19] Nalamati M, Sharma N, Saqib M, et al. Automated monitoring in maritime video surveillance system[C]//2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), November 25-27, 2020, Wellington, New Zealand. New York: IEEE Press, 2020: 1-6.
- [20] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2015-11-23) [2021-04-20]. <https://arxiv.org/abs/1511.07122>.
- [21] Zhang S F, Wen L Y, Bian X, et al. Single-shot refinement neural network for object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4203-4212.
- [22] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [23] Pang J M, Chen K, Shi J P, et al. Libra R-CNN: towards balanced learning for object detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 821-830.
- [24] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [25] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-04-20]. <https://arxiv.org/abs/1804.02767>.
- [26] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//IEEE conference on Computer Vision and Pattern Recognition(CVPR), July 21-26, 2017, Honolulu, USA. New York: IEEE Press, 2017: 2117-2125.
- [27] Shao Z F, Wu W J, Wang Z Y, et al. SeaShips: a large-scale precisely annotated dataset for ship detection [J]. IEEE Transactions on Multimedia, 2018, 20(10): 2593-2604.