

DOI: 10.19650/j.cnki.cjsi.J2210128

融合注意力和多尺度特征的典型水面小目标检测^{*}

童小钟,魏俊宇,苏绍璟,孙 备,左 震

(国防科技大学智能科学学院 长沙 410072)

摘 要:为解决多场景复杂海况背景水面小目标检测存在的可利用特征少、纹理信息弱等问题,提升无人艇的环境感知能力,本文提出一种融合注意力和多尺度特征的典型水面小目标检测算法。首先,在网络的深层使用空洞空间金字塔池化模块融合目标的全局先验信息。其次,通过注意融合模块自适应地增强目标浅层空间位置和深层语义信息特征,提高网络的特征表示能力。最后,通过多尺度特征融合实现高性能的目标检测。本文构建了典型水面小目标数据集,并基于无人艇开展了真实海况下水面小目标检测的算法验证。实验结果表明,该算法在无人艇 NVIDIA 平台检测速率达到 17 FPS,能准确识别水面小目标, mIoU 比原始特征金字塔网络算法提升 7.58%,平均检测精度提升 11.41%,达到 82.36%。

关键词:小目标检测;多尺度特征;注意力融合;无人艇

中图分类号: TP391.4 TH701 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Typical small target detection on water surfaces fusing attention and multi-scale features

Tong Xiaozhong, Wei Junyu, Su Shaojing, Sun Bei, Zuo Zhen

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410072, China)

Abstract: To address the problems of small targets detection with few available features and weak texture information in the context of complex sea conditions in multiple scenarios, and to improve the environmental perception capability of unmanned surface vehicles (USV), we propose a typical small targets detection method using attention mechanism and multi-scale features. Firstly, the global prior information of the target is fused in the deep layers of the network using atrous spatial pyramid pooling module. Secondly, the shallow spatial location and deep semantic information features of the target are adaptively enhanced by the attention fusion module to improve the feature representation capability of the network. Finally, the high performance target detection is achieved through multi-scale feature fusion. We construct a typical surface small target dataset, and the method is evaluated by experiments of surface small target detection under real sea conditions based on USV. Experimental results show that the proposed method in the NVIDIA platform reaches 17 FPS, which can accurately identify small target on the water surface. Compared with the original FPN algorithm, the mIoU is improved by 7.58%, and the average detection accuracy is improved by 11.41% to 82.36%.

Keywords: small target detection; multi-scale features; attention fusion; unmanned surface vehicles

0 引 言

水面无人艇作为海上无人系统和新型作战力量的代表性装备,在海上情报侦察、领海海域监视、巡逻搜救和反潜排雷等军事领域发挥巨大的应用价值^[1]。无人艇可以执行人无法执行或不愿执行的多种危险、枯燥的任务,

在海面巡查监管、灾害事故搜救、水文信息获取、水文环境监测、海底地形测绘和气象预报等民用领域扮演越来越重要的角色。随着无人艇作业任务的多样性、活动半径的扩大以及复杂海况的影响,提高无人艇的环境感知和任务规划能力十分迫切^[2]。

准确高效的目标检测、具备良好的环境感知能力是无人艇自主化、智能化最重要的一环。无人艇平台通常

收稿日期:2022-07-13 Received Date: 2022-07-13

^{*} 基金项目:国家自然科学基金(52101377)、湖南省自然科学基金(2020JJ5672)、湖南省研究生科研创新(CX20220015)项目资助

搭载可见光相机、红外热像仪、激光雷达和航海雷达等多种类型的传感器,通过利用不同类型的传感器信息感知水面的障碍、目标的位置、类别和运动状态信息。由于红外热像仪缺少纹理信息和细节信息^[3]、激光雷达作用距离有限、航海雷达容易受海杂波影响且只能得到目标的运动状态信息,因此可见光相机是无人艇对目标进行检测最常用的方式。然而,由于多场景复杂海况背景下海浪、海雾、水面光线反射等环境因素的影响,海面目标的特点主要包括在图像中所占像素数少、目标尺度变化大、纹理信息弱、可利用特征少等^[4]。

近年来,深度学习极大推动了计算机视觉的发展,卷积神经网络(convolutional neural networks, CNN)通过构建多个网络层,以监督或非监督的方式训练 CNN 模型,对弱小目标之间的特征信息进行学习和建模表示,使模型具有更加精细准确的描述能力,从而实现抽象化和层次化的目标特征表示。基于 CNN 的方法因实时性强,检测精度高,可应用场景广,成为水面小目标检测领域的研究热点。范云生等^[5]基于电子海图与雷达图像融合进行环境建模实现无人艇的动态路径规划。彭艳等^[6]结合多了尺度的空间上下文和相关滤波信息实现了实时精确的目标跟踪。吴鹏等^[7]通过航海雷达和光电吊舱的协同提升了无人艇的海面环境感知能力。李尚君等^[8]基于卡尔曼滤波对航海雷达得到的目标航迹进行预测,并通过光电吊舱及其姿态补偿算法有效提升了目标跟踪性能。程亮等^[9]针对水上目标的特点提出了基于改进 YOLOv3 的增强型轻量级水上目标检测网络。孙备等^[10]通过设计空间注意力和多尺度特征融合的改进型 SSD 实现了水面渔船、快艇、桥墩等目标的识别。柳晨光等^[11]将三维激光点云投影至二维栅格表示,利用八邻域算法对障碍物进行聚类,通过激光雷达对无人艇周围 100 m 范围内的障碍目标进行检测。Bovcon 等^[12]构建了海洋目标数据集作为海面目标检测的基准。Bovcon 等^[13]设计一种新的解码器网络,融合 IMU 的惯性信息和编码器的视觉特征,提高水面障碍物检测的准确率。Cheng 等^[14]构建内陆水域漂浮废物检测的毫米波雷达数据和图像,并设计基于图像和雷达传感器数据融合的漂浮物检测策略。Zust 等^[15]设计新的支架式学习机制,依靠弱标签训练障碍物分割网络,从而显著降低标记成本。

基于此,本文首先构建了水面目标数据集用于网络的训练和测试。然后设计了专门用于典型水面小目标检测的融合注意力和多尺度特征的算法,通过设计的注意力融合模块关注小目标的重要空间位置细节和通道语义信息,获取和利用图像全局上下文先验信息,增强小目标的特征表示,在不增加模型复杂度的情况下,该算法能够提高检测性能。最后,将算法部署在无人艇 NVIDIA Xavier 嵌入式计算平台实现了真实海况的水面小目标检

测。实验结果表明,该算法具有优异的小目标检测性能,在真实海况的数据集上对小目标的检测精度相比原始的 FPN 网络提升了 11.41%。

1 无人艇环境感知平台总体架构

本文在实验中搭建了如图 1 所示的无人艇平台,用于典型海面小目标数据的采集及算法的部署验证。无人艇平台搭载了光电吊舱和航海雷达传感器用于获取水面光电图像和雷达回波的实时采集以实现水面环境的感知。此外,无人艇还搭载了惯导模块及 GPS 获取无人艇自身的姿态和位置信息,然后通过天线模块建立与岸基之间的通信,实现数据的传输和信息的交互。



图1 无人艇平台

Fig. 1 USV platform

无人艇搭载的光电吊舱具有可见光相机、非制冷型热成像仪和激光测距仪 3 种传感器。本文研究的重点是通过光电吊舱获取可见光图像对典型水面小目标进行检测,并研究在不同海况和多种类型场景下的无人艇环境感知技术。

2 数据集构建

本文的实验分为数据采集和无人艇算法部署验证两个阶段。在实验数据采集阶段,本文通过无人艇平台搭载 4K 高清摄像机拍摄典型海面场景的小目标,构建了 Seaship 数据集,然后通过人工去重、删除低质量图像、手工标注等步骤建立水面目标数据集共 1 200 张图片。数据集场景信息丰富,包含雨雾天气、大浪等多种场景,覆盖真实海况信息。在本文中,自己制作的 Seaship 数据集按照 DUT-OMRON^[16]数据集的格式进行构建,并按照 6:2:2 的比例划分为训练集、验证集和测试集。Seaship 数据集中目标最小尺寸仅为 10 pixels,部分样本如图 2

所示,图3总结了 Seaship 数据集目标像素数占图像总像素数比例的分布情况。



图2 多场景典型水面小目标图像

Fig. 2 Multi-scene images of typical small targets on the sea surface

3 融合注意力和多尺度特征的典型水面小目标检测方法

本文融合注意力和多尺度特征的水面小目标检测算法 (cross attention feature fusion pyramid network,

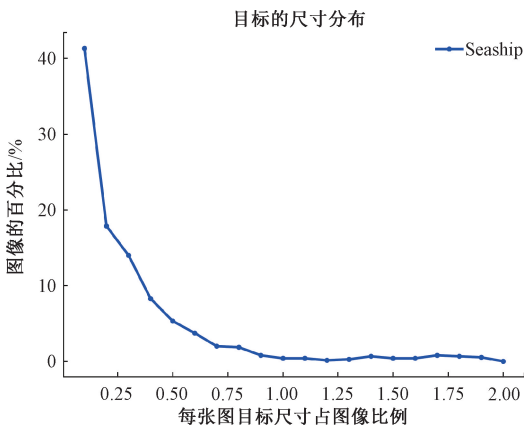


图3 Seaship 数据集目标占图像总像素数分布情况

Fig. 3 Distribution of targets in the Seaship dataset in terms of the total number of pixels in the image

CAFFPN)整体框架如图4所示。CAFFPN 由特征提取和特征融合两部分组成,特征提取阶段设计了关注小目标全局上下文信息的空洞空间金字塔池化模块和注意力融合模块;特征融合阶段采用不同尺度进行特征融合的方式以增强算法对目标的表示能力,提高水面小目标的检测性能。

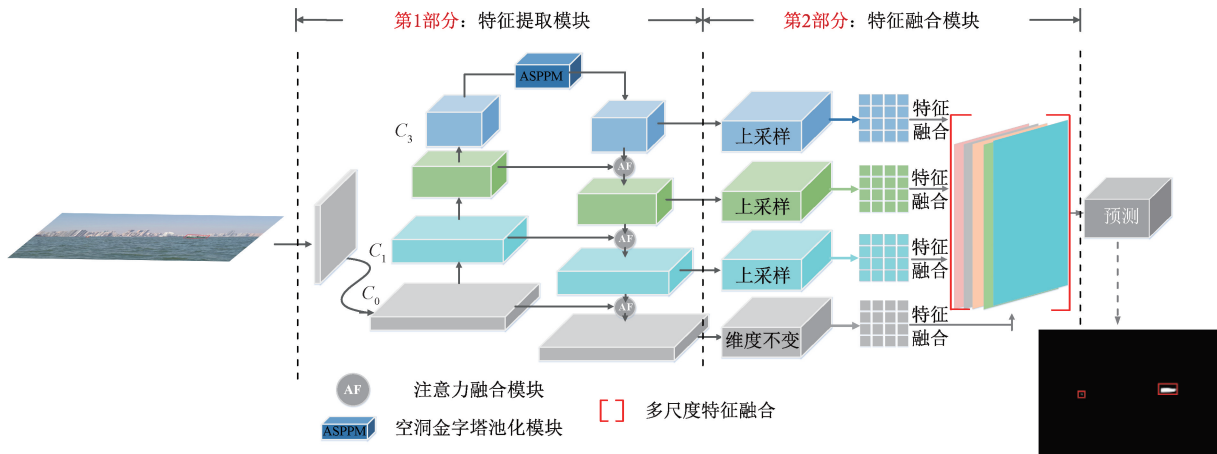


图4 本文提出的 CAFFPN 网络结构

Fig. 4 Architecture of the proposed CAFFPN

3.1 网络结构

如图4所示,给定典型水面图像作为输入,CAFFPN 依次进行特征提取和特征融合,然后预测模块对每个像素进行分类并输出水面小目标的分割结果。3.2 节详细介绍了特征提取模块的细节,该模块由两部分组成:获取全局上下文信息的空洞空间金字塔池化模块 (atrous spatial pyramid-pooling module, ASPPM) 和融合不同特征层的注意力融合模块 (attention fusion module, AFM)。输入图像经过预处理和下采样,提取丰富的小目标特征,

ASPPM 充分利用小目标的全局上下文信息。不同特征层之间的语义信息存在差距,本算法使用注意力融合模块来自适应地增强深层和浅层特征,以获得更准确的通道和目标的空间位置信息。为了确保浅层的空间位置信息传递给深层的网络,减少深层的水面小目标在浅层的细节信息丢失,CAFFPN 使用 ResNet-20^[17] 作为特征提取的主干架构。金字塔特征融合模块在 3.3 节进行了详细的阐述。之后,将不同尺度的特征层上采样到相同的大小,然后通过级联操作将不同网络层的特征融合起来,生

成能够显著提升小目标特征表示鲁棒性的特征图。预测模块最终输出水面小目标的检测结果。

3.2 特征提取模块

1) 空洞空间金字塔池化模块

Zhao 等^[18]证明了金字塔池化模块可以构建有效的全局上下文先验,减少上下文信息的损失。如图 4 所示,本算法在 FPN 基础框架的最深特征层(C_3)中构建了分层上下文金字塔池化模块,以充分利用全局信息。单一的 3×3 或 1×1 卷积不能充分混合所有多尺度的上下文信息。本算法首先通过 3 个不同扩张率的空洞卷积来增加深层网络的感受野,使用 3 个不同大小的池化模块,划分为不同的子区域,形成不同位置的池化表示。然后将不同尺度的特征图进行上采样后与不同大小的空洞卷积特征图融合,通过级联的方式获得丰富的感受野,形成最终的目标特征表示。

包含不同尺度和子区域的分层全局先验可以减少子区域之间的信息丢失^[19]。本算法提出了新颖的 ASPPM,其结构如图 5 所示。以更好地捕捉水面小目标图像的全局先验信息。对于层数为 N 的金字塔,首先通过不同扩张率的卷积将特征图上采样到与输入特征(C_3)相同的尺寸。最后将不同层次的特征串联起来,形成最终的全局上下文先验。值得注意的是,非线性卷积扩张率的大小和全局池化的大小是可以改变的。本算法根据小目标的像素分布特性,将 ASPPM 扩张率的大小分别设置为 3, 6, 9, bin 大小分别设置为 1×1 、 2×2 和 4×4 。全局上下文先验的 ASPPM 计算公式为:

$$y_j = \begin{cases} \delta(\mathcal{B}(\text{Conv}_{3 \times 3}(x, \text{dilation} = i))), & i = 3, 6, 9 \quad j = 1, 2, 3 \\ U(\text{Conv}_{1 \times 1}(\mathcal{B}(\delta(\text{AdapPool}(x, i))))), & i = 1, 2, 4 \quad j = 4, 5, 6 \end{cases}$$

$$L(X) = \text{Conv}_{1 \times 1}(\text{Concat}[y_j(x), x]) \quad (1)$$

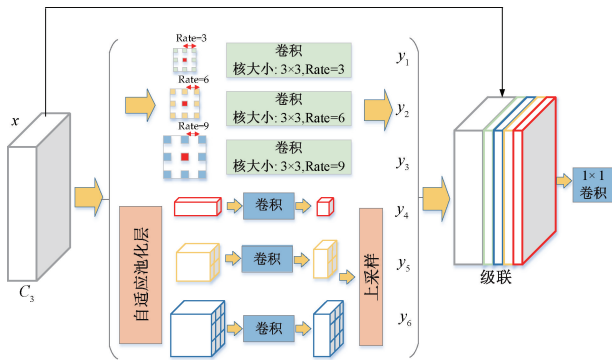


图 5 ASPPM 的结构

Fig. 5 Architecture of the ASPPM

其中, $\text{Conv}_{1 \times 1}$ 是核大小为 1×1 的逐点卷积^[20], $\text{Conv}_{3 \times 3}$ 是核大小为 3×3 的不同扩张率空洞卷积, U 、 $\mathcal{B}$ 和 δ 分别是上采样、批量归一化和 ReLU 激活函数。Concat 是不同尺度的级联特征操作, AdapPool 是不同特征大小的全局池化操作。

2) 注意力融合模块

深层特征网络能够提取小目标的语义信息,但可能会丢失小目标的空间细节信息。浅层特征网络能够有效保留目标的空间位置信息,但缺乏对目标的高级语义理解。从图 3 中可以看出,典型水面小目标占图像总像素数少,随着网络的加深,目标可能在深层网络中丢失。此外,当网络不能充分利用不同特征层的信息时,目标的特征表示就很困难。

如图 6 所示,为了确保在保持小目标的深层特征表示的同时提取丰富的语义信息,本文提出了注意力融合模块。此外,该模块还用于解决目标细节丢失、不同层特征融合时存在的信息冗余和特征融合不充分的问题。图 6 中 X 和 Y 分别指浅层细节信息和深层语义信息。深层语义特征图被上采样到相应的浅层特征图相同的大小,然后通过注意力融合模块获得特征图输出 Z 。注意力融合模块由关注目标浅层空间位置信息的空间注意力 (spatial attention, SA) 和关注目标深层语义信息的通道注意力 (channel attention, CA) 组成。输入的浅层特征 X 和深层语义特征 Y 分别由 SA 和 CA 处理,然后将处理后的特征进行融合。

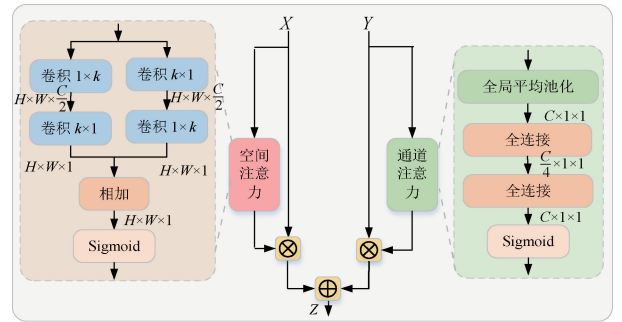


图 6 注意力融合模块的结构

Fig. 6 Architecture of the attention-fusion module

空间注意力:本文设计了 SA 来学习小目标在浅层特征图包含的大量细节信息^[21],通过 SA 帮助网络关注目标详细的空间位置信息,有助于提高小目标有效的可检测特征。本算法将浅层特征表示为 $X \in \mathbb{R}^{C \times H \times W}$, 而空间位置信息的集合表示为 $\mathbb{R} = \{(x, y) | x = 1, 2, \dots, H; y = 1, 2, \dots, W\}$, 其中, (x, y) 代表特征的位置坐标。输入的浅层特征图采用 $1 \times k$ 和 $k \times 1$ 的卷积核大小进行卷积,分别提取水面小目标的空间位置和浅层特征的细节信息。最后,使用 sigmoid 函数将空间特征图归一化。关

注浅层空间位置信息的 SA 权重计算如下:

$$\begin{aligned} C_1 &= \text{Conv}_2(\text{Conv}_1(X, W_1^1), W_1^2) \\ C_2 &= \text{Conv}_1(\text{Conv}_2(X, W_1^1), W_2^2) \\ SA &= S(X, W) = \sigma(C_1 + C_2) \end{aligned} \quad (2)$$

式中: Conv_1 和 Conv_2 表示核大小分别为 $1 \times k$ 和 $k \times 1$ 的卷积运算, W 和 sigmoid 函数是 SA 的参数。最后, 通过对 SA 的加权得到浅层特征输出:

$$X' = SA(X) \otimes X \quad (3)$$

其中, $\otimes$ 和 $SA(\cdot)$ 分别表示元素相乘和空间注意力。

通道注意力: 深层特征包含高度抽象的语义信息^[17], 但缺乏关于目标的细节信息。本文使用全局 CA^[22] 对深层特征进行加权, 使用多个感受野来捕捉高辨别力的通道特征。首先由式(4)生成通道特征信息, 得到大小为 $1 \times 1 \times C$ 的全局感受野的特征图, 聚集全局上下文信息; 然后通过两个全连接层捕捉通道之间的依赖关系, 如图 6 所示。深层通道信息 CA 的注意力权重通过式(5)得到:

$$y = \frac{1}{H \times W} \sum_{i=1}^{H,W} Y[:, i, j] \quad (4)$$

$$CA = \sigma(\mathcal{B}(W_2 \delta(\mathcal{B}(W_1 y)))) \quad (5)$$

式中: H , W , σ , $\mathcal{B}$ 和 δ 分别表示高度、宽度、sigmoid 函数、批量归一化和 ReLU 函数。此外, W_1 和 W_2 代表 2 个完全连接的操作, 其中, $W_1 \in \mathbb{R}^{C \times C}$ 和 $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, r 表示通道衰减率。最后, 根据 CA 权重得到深层特征输出。

$$Y' = CA(Y) \otimes Y \quad (6)$$

其中, $\otimes$ 和 $CA(\cdot)$ 分别表示元素相乘和通道注意力。浅层和深层特征分别通过 SA 和 CA 后得到注意特征融合图的最终输出,

$$Z = SA(X) \otimes X + CA(Y) \otimes Y \quad (7)$$

3.3 特征融合模块

在特征提取模块之后, 单尺度的特征图无法充分表征水面小目标的特征。如图 4 第 2 部分所示, 本文设计了专门用于聚合深层和浅层特征的金字塔特征融合模块, 实现水面小目标的多层特征融合。首先将不同层的特征图上采样到相同大小, 并将含有语义信息的深层特征和含有细节信息的浅层特征连接起来, 生成含有丰富信息的特征图。通过多尺度特征融合减少小目标的特征丢失和深层网络中浅层空间位置信息的利用不足, 保证小目标特征的鲁棒性。多特征融合的特征图生成通过下式得到:

$$P'_i = U^{2^i}(\delta(\text{GN}(\text{Conv}_{3 \times 3}(P_i)))) \quad \{i = 1, 2, 3\} \quad (8)$$

其中, $U(\cdot)$ 表示使用双线性插值的上采样, 2^i 是上采样的倍数。 $\text{Conv}_{3 \times 3}$, GN 和 δ 分别表示核大小为 3×3 的卷积运算, 组归一化和 ReLU 函数。

3.4 损失函数

由于真实海况下多场景的图像背景和典型海面小目标之间存在严重的类别不平衡问题, 本文在网络训练过程中采用了 soft-IoU^[23] 损失函数, 其定义如下:

$$\ell_{\text{soft-IoU}}(x, s) = \frac{\sum_{i,j} x_{i,j} \cdot s_{i,j}}{\sum_{i,j} s_{i,j} + x_{i,j} - x_{i,j} \cdot s_{i,j}} \quad (9)$$

其中, $s \in \mathbb{R}^{H \times W}$ 是给定输入的可见光图像预测的分数图, $x \in \mathbb{R}^{H \times W}$ 是给定的标签真值。

4 实验结果与分析

为证明本文所提出的算法用于典型水面小目标检测的性能, 采用自己构建的多场景复杂海况的 Seaship 数据集对算法进行训练和测试, 并与其他方法进行了对比, 验证本文提出的算法在复杂海况下对水面小目标表现出良好的检测性能。

4.1 实验环境

本文所有实验都是基于 PyTorch 深度学习框架实现。输入的图像被随机裁剪成 480×480 的大小, 采用随机旋转、随机裁剪等方式对训练数据进行增强。所有图像都被归一化, 以加速网络收敛。使用 AdaGrad 优化器^[24], 批量大小为 8, 初始学习率和权重衰减分别设置为 0.05 和 1×10^{-4} , 网络模型总共训练了 300 个 epoch。实验的硬件环境为配备英特尔 i9-10900X @ 3.7 GHz CPU 和单个 TITAN RTX GPU 的计算机用于训练和测试。

4.2 评价指标

为了验证本文提出算法的有效性, 本文使用平均交并比 (mean intersection over union, mIoU), F-measure、精度与召回率关系 (precision-recall, PR) 曲线和接收器操作特性 (receiver operating characteristic, ROC) 曲线 4 个指标对算法的性能进行评估。mIoU 是经典的语义分割评价指标, 用于表征算法的轮廓描述能力。它被定义为预测和标签之间的交集和并集面积的比, 计算公式为:

$$mIoU = \frac{\# \text{Area of Overlap}}{\# \text{Area of Union}} \quad (10)$$

F-measure 用于衡量精度和召回率之间的关系。精度、召回率和 F-measure 定义如下:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F_{\text{measure}} &= \frac{(\beta^2 + 1) Precision \times Recall}{\beta^2 Precision + Recall} \end{aligned} \quad (11)$$

其中, FP 和 FN 分别表示假阳性和假阴性的数量。

PR 曲线用于描述精度和召回率之间的动态变化;曲线越接近右上方,性能就越好。平均精度 (average precision, AP) 用于准确评估 PR 曲线,定义如下:

$$AP = \int P(R) dR$$
 (12)

其中, P 为精度, R 为召回率。

真阳性率 (true positive rate, TPR) 和假阳性率 (fals positive rate, FPR) 之间的动态关系由 ROC 描述。TPR 和 FPR 定义如下,其中 FN 表示假阴性的数量。并使用曲线下面积 (area under the curve, AUC) 定量评估 ROC。

$$TPR = \frac{TP}{TP + FN}$$
 (13)

$$FPR = \frac{FP}{FP + TN}$$

4.3 典型水面小目标 Seaship 数据集实验

为了证明本文所提出小目标检测算法的有效性,将融合注意力和多尺度特征的典型水面小目标检测算法与几种先进的语义分割算法 FPN^[25]、UNet^[26]、ACM-FPN^[27]、ACM-UNet^[27]、BASNet^[28] 和 CorrNet^[29] 在自己构建的多种场景真实海况的 Seaship 数据集上进行检测效果和性能对比实验。为了保证实验的可对比性,上述几种对比算法参照作者已公开代码的原始参数设置,总共训练 300 epoch 的模型在测试集上对算法性能进行对比。

表 1 详细列出了 5 种不同方法的 mIoU、F-measure、AP 和 AUC 评价指标。每一列中最好的结果用粗体突出显示。从表中可以看出,本文算法在大多数评价指标上都取得了最好的结果。所提出的算法在典型水面场景小目标检测方面取得了显著的性能提升。本文构建的 Seaship 数据集包含雨雾、大浪等充满挑战的复杂场景。本文算法首先为典型水面小目标检测重新设计了骨干网络,该网络在水面小目标特征提取阶段构建了有效的全

局先验,以减少深层特征中的上下文信息损失,并更好地捕捉水面小目标的全局上下文先验信息。之后,利用设计的注意力融合模块自适应地增强 CNN 深层和浅层特征的语义信息和空间位置细节信息,关注不同网络层的小目标特征表示,并通过多尺度特征融合模块实现不同网络层的渐进特征融合。通过这些有针对性的网络结构和模块设计,典型水面小目标的固有特征被保留在网络的深层,增强了网络的小目标特征表示能力,从而显著提高了检测性能。

表 1 不同方法的评价指标对比
Table 1 Comparison of evaluation indicators for different methods

方法	mIoU/ ($\times 10^{-2}$)	F-measure/ ($\times 10^{-2}$)	AP/ ($\times 10^{-2}$)	AUC/ ($\times 10^{-2}$)
FPN	59.73	76.43	70.95	89.20
UNet	59.72	79.52	75.45	91.76
ACM-FPN	63.94	80.19	76.99	92.48
ACM-UNet	63.19	78.63	72.73	90.16
BASNet	66.79	83.31	82.19	95.63
CorrNet	65.17	82.10	77.56	93.70
本文	67.31	84.60	82.36	94.27

为了证明本文设计的注意力融合模块对水面小目标检测的效果,图 7 对注意力融合的作用效果进行了可视化。从图 7 中可以看出,注意力融合模块显著提高了海面小目标的识别能力,可以更准确地捕获目标的形状,并且它们提供了较强的小目标特征信号线索。如图 7(c) 所示,注意力融合更加聚焦图像中目标的信号区域,捕获更清晰的特征,识别出更多细节的边界信息。

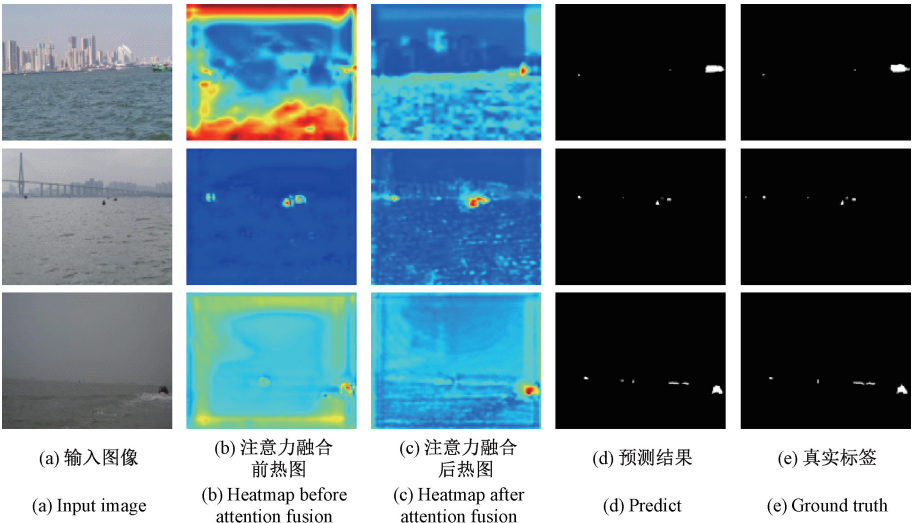


图 7 CAFFPN 通过注意力融合模块前/后的可视化热图
Fig. 7 Visualization map of CAFFPN and CAFFPN w/o attention fusion

图8 比较了7种不同方法的PR和ROC曲线结果。本文提出的算法在所有指标上都优于对比的方法,这表明它有能力在分割目标的同时大幅抑制背景并准确地检测目标。

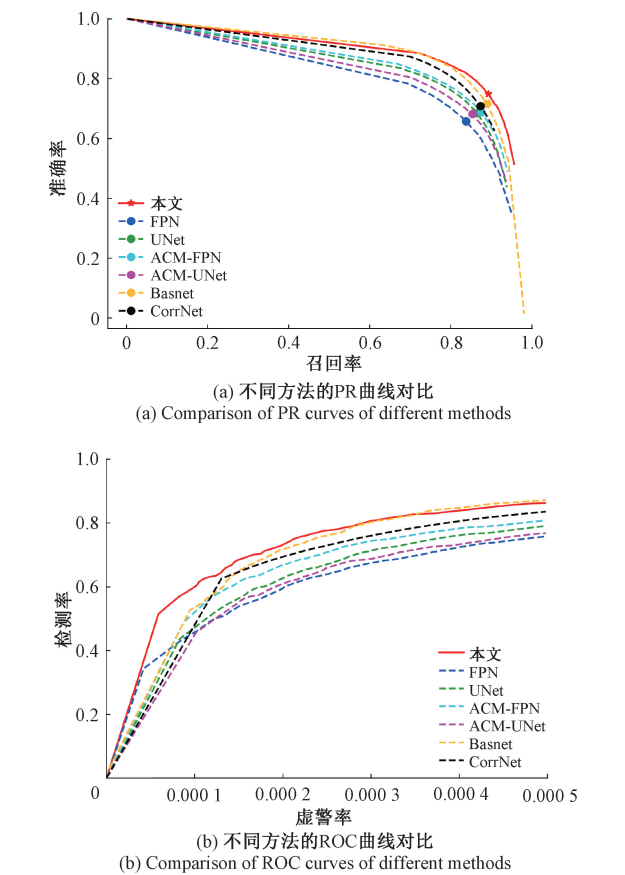


图8 本文方法与其他方法的PR和ROC结果对比
Fig. 8 Comparison of PR and ROC results of different methods

由PR曲线结果(图8(a))可知,所提出的方法实现了最好的精度和召回率,这意味着它能够在目标的大小、形状和位置不同的挑战性场景中保证整体的目标定位和检测精度。由ROC曲线结果(图8(b))可知,所提出的方法实现了最佳性能,概率检测对假阳性率的变化反应迅速。实验结果证明本文提出的算法能够有效应对真实海况条件下雨雾天气、大浪等复杂场景的变化,表现出真实海况场景下典型水面小目标检测的鲁棒性和环境适应性。

4.4 消融实验

为证明本文所设计的注意融合模块新颖性和有效性,构建不同变体进行详细的消融研究。

1)CAFFPN-跨层特征融合:考虑不同特征层之间的跨层特征调制,改变CA和SA关注的特征层。具体来说,SA对浅层特征和深层特征动态加权调制,CA对深层特征和浅层特征加权调制。最后,通过相加的方式进行特征融合。

2)CAFFPN w/o AF:CAFFPN的这一变体去掉了CA和SA模块,使用常见的逐元素相加代替AF模块来实现不同层的特征融合。其目的是探索注意力融合模块的有效性。

3)CAFFPN w/o SA:在这个变体中只考虑CA,去掉SA以研究其贡献。

4)CAFFPN w/o CA:在这个变体中只考虑SA,去掉CA以评估其优势。

表2列出了不同变体的消融实验结果,每一列的最佳结果用粗体标出。本文为跨层特征融合设置一个消融模块,以探索浅层细节特征和深层语义特征之间信息交互的效果。表2中的结果显示,CAFFPN-跨层特征融合的mIoU、F-measure和AP分别下降了5.05%、5.92%和7.24%。实验结果表明,最大限度发挥不同特征层的优势是提高网络对水面小目标检测特征表示能力的关键。如果去掉AF模块,mIoU、F-measure和AP分别减少了4.71%、5.5%和7.09%,说明SA和CA融合的重要性,AF模块充分关注不同层的空间位置和语义特征信息,并融合不同子特征的信息,增强CNN对水面小目标的特征表示,从而提高水面目标分割的性能。为进一步说明本文提出的注意力融合模块的有效性,将注意力融合前后的热图可视化,如图7所示,在注意力融合模块的帮助下,CAFFPN深层特征图的形状分割准确,能够灵敏反应信息线索。此外,本文对缺乏SA和CA的变体分别进行了消融实验,以探索两种不同类型注意力的贡献。如表2所示,没有SA的CAFFPN和没有CA的CAFFPN的各项指标都有明显下降。这是因为SA帮助网络专注于浅层细节特征和位置信息,而CA更好地利用深层语义信息来增强网络表征能力,从而产生更好的小目标检测结果。

表2 不同注意力融合模块在Seaship数据集上的检测结果
Table 2 Different attention fusion modules and their detection results on the Seaship dataset

方法	mIoU/ ($\times 10^{-2}$)	F-measure/ ($\times 10^{-2}$)	AP/ ($\times 10^{-2}$)	AUC/ ($\times 10^{-2}$)
CAFFPN-跨层特征融合	62.26	78.68	75.12	91.44
CAFFPN w/o AF	62.60	79.10	75.27	91.47
CAFFPN w/o SA	62.80	79.35	75.28	91.64
CAFFPN w/o CA	63.19	79.20	76.22	92.26
本文	67.31	84.60	82.36	94.27

4.5 SIRST公共数据集结果与分析

为了进行更加公正的比较,本文算法在SIRST^[30]公开数据集与最先进的几种方法进行对比实验。训练集和测试集按照作者的原始设置,采用Seaship数据集相同的评价指标。

表 3 所示为 7 种不同方法在 SIRST 公共数据集上的定量测试结果。与其他方法相比,本文算法的提升十分明显,各项指标均取得最好的性能,进一步证明本文为小目标检测专门设计的融合注意力和多尺度特征网络的有效性,能够显著提高小目标检测性能。

表 3 在 SIRST 数据集上与最先进方法的定量比较
Table 3 Quantitative comparison with state-of-the-art methods on the SIRST dataset

方法	mIoU/ ($\times 10^{-2}$)	F-measure/ ($\times 10^{-2}$)	AP/ ($\times 10^{-2}$)	AUC/ ($\times 10^{-2}$)
Top-hat	28.75	69.29	58.49	84.40
FPN	72.18	80.39	75.90	93.10
U-Net	73.64	80.81	76.11	94.01
TBC-Net	73.40	—	—	—
ACM-FPN	73.65	81.60	78.33	93.79
ACM-U-Net	74.45	81.68	78.08	93.63
本文	78.14	83.53	80.61	94.52

为直观表示本文算法对小目标的检测效果,图 9 所示为部分 SIRST 数据集检测结果,在每张图像的左上角都标注了检测方法。为了更直观地显示精细分割的结果,目标区域在右下角被放大。虚线和实线圆圈分别表示正确检测的目标和虚警。

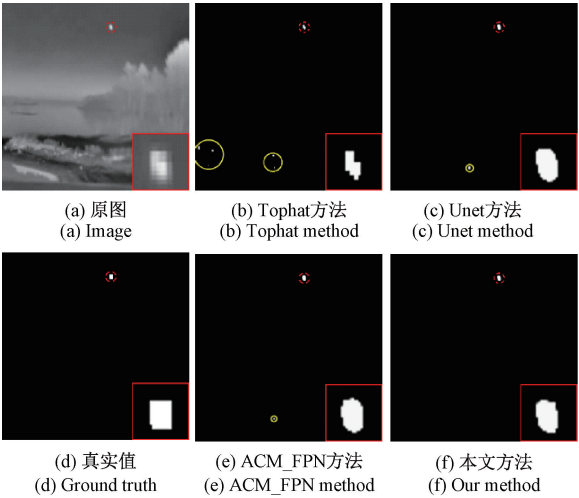


图 9 SIRST 数据集检测结果

Fig.9 Detection results based on SIRST dataset

上述结果表明,本文提出的算法对小尺寸目标的检测能力更加有效。其他方式都出现不同程度的虚警现象,本文提出的算法能够对小目标进行准确的定位、形状和轮廓的细节分割,进一步证明本文设计的融合通道空间注意力和多尺度特征能够有效保留小目标在

网络的深层代表性特征,增强网络对小目标的特征表示能力。

4.6 推理性能分析

推理性能是无人艇实际部署和应用的关键,本文提出的算法在固定的高性能计算平台上进行网络模型的训练和测试。在计算推理性能时,忽略数据准备的时间,仅考虑算法处理图像的时间。表 4 中列出了不同方法对单个图像帧的平均推理时间。此外,还计算了本文算法在单个 TITAN RTX GPU 的计算机和 NVIDIA Xavier 开发板上的推理速度,并在测试环境中分别使用 G 和 B 来区分它们。从表 4 中可以看出,本文提出的算法检测速度和检测精度做到了很好的权衡,在单个 TITAN RTX GPU 达到了 125 FPS 的检测速度,能够满足实时性的要求。

表 4 不同算法的推理性能
Table 4 Inference times for different algorithms

方法	Params/M	FLOPs/G	FPS/FPS
FPN	8.08	27.66	160
UNet	8.25	27.59	145
ACM-UNet	8.27	27.66	101
ACM-FPN	6.16	15.48	130
Basnet	87.06	447.75	72
本文方法(G)	7.40	30.91	125
本文方法(B)	7.40	8.79	17

4.7 真实海域小目标检测应用

本文将算法搭载在无人艇平台进行真实海况条件下的实验验证。提出的算法部署和验证均在 NVIDIA Xavier 计算平台上进行。该嵌入式计算平台具有性能优异的 GPU 和 CPU 的计算资源,并且具备较好的散热性能和较低的功耗,能够快速传输数据资源,可广泛部署在机器学习和高效实时的处理对象检测等计算机视觉任务。图 10 展示了部署在无人艇平台的 NVIDIA Xavier 图像处理平台的实物图。

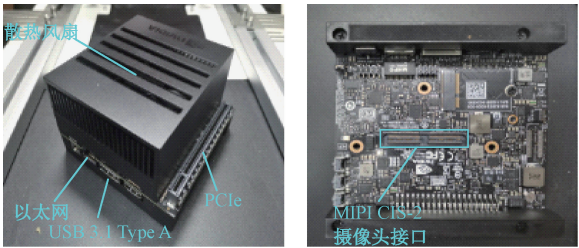


图 10 NVIDIA Xavier 计算平台实物

Fig.10 NVIDIA Xavier computing platform

得益于 NVIDIA Xavier 的强劲性能,无人艇平台能够实时高效接收多传感器的环境感知信息,对获取的目标图像进行检测,并将检测结果传输到控制终端,推理速度达到了 17 FPS,实现典型水面小目标的实时检测。图 11 所示为无人艇对典型水面小目标检测的实验结果。从图 11 中可以看出,本文提出的算法能够在多种场景下对水面小目标准确的检测,基本满足无人艇环境感知的需求。

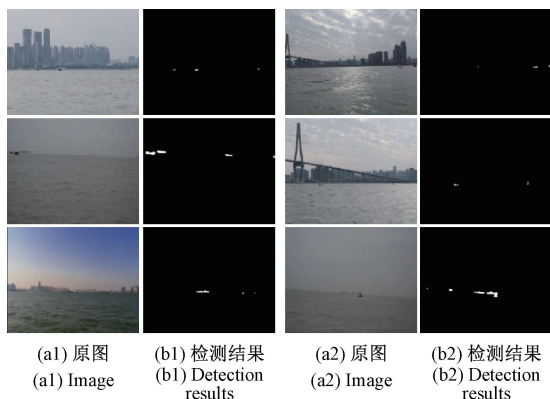


图 11 真实海况无人艇水面小目标检测结果

Fig. 11 Real-sea state USV surface small target detection results

5 结 论

本文提出了融合注意力和多尺度特征的典型水面小目标检测算法。该网络包括特征提取和特征融合模块。在特征提取阶段,使用 ASPPM 来减少深层网络的上下文信息丢失,更好地利用小目标的全局上下文先验信息。为了关注不同层的目标特征,设计了空间注意力和通道注意力,分别对深层和浅层目标的语义和空间位置信息进行自适应增强,显著提高了小目标固有信息的利用。最后,提出了多尺度特征融合策略实现了高性能检测。本文构建了复杂真实海况下的水面目标数据集,解决了算法实验中图像数据匮乏的问题。

实验结果表明,本文算法对真实海况的典型水面小目标有良好的检测性能和推理速度,与主流的深度学习方法相比,具有更好的检测精度。在本文构建的 Seaship 数据集上,算法 mIoU 比原始 FPN 提升 7.58%,平均检测精度 82.36%,提升 11.41%,具有一定的鲁棒性和实时性,基本达到无人艇水面小目标检测的要求。本文提出的算法在典型场景取得了较为理想的水面小目标检测性能,未来的研究应该更加聚焦在恶劣海况和复杂背景下的水面小目标检测、更加关注降低模型

复杂度的同时提升检测效率和检测精度、网络在不同场景的鲁棒性和泛化性等。

参考文献

- [1] 金霄,郑开原,王得朝,等. 国外军用无人水面艇发展综述[J]. 中国造船, 2020, 61 (S1): 221-227.
JIN X, ZHENG K Y, WANG D CH, et al. Overview of development of foreign military unmanned surface vehicle[J]. Ship Building of China, 2020, 61 (S1): 221-227.
- [2] 侯瑞超,唐智诚,王博,等. 水面无人艇智能化技术的发展现状和趋势[J]. 中国造船, 2020, 61 (S1): 211-220.
HOU R CH, TANG ZH CH, WANG B, et al. Development status and trend of intelligent technology for unmanned surface vehicles[J]. Ship Building of China, 2020, 61 (S1): 211-220.
- [3] 陈卓,方明,柴旭,等. 红外与可见光图像融合的 U-GAN 模型[J]. 西北工业大学学报, 2020, 38(4): 904-912.
CHEN ZH, FANG M, CHAI X, et al. U-GAN model for infrared and visible images fusion[J]. Journal of Northwestern Polytechnical University, 2020, 38 (4): 904-912.
- [4] 刘俊琦,李智,张学阳. 可见光遥感图像海面目标检测技术综述[J]. 计算机科学, 2020, 47(3): 116-123.
LIU J Q, LI ZH, ZHANG X Y. Review of maritime target detection in visible bands of optical remote sensing images[J]. Computer Science, 2020, 47(3): 116-123.
- [5] 范云生,柳健,王国峰,等. 基于异源信息融合的无人水面艇动态路径规划[J]. 大连海事大学学报, 2018, 44(1): 9-16.
FAN Y SH, LIU J, WANG G F, et al. Dynamic path planning for unmanned surface vehicle based on heterologous information fusion[J]. Journal of Dalian Maritime University, 2018, 44(1): 9-16.
- [6] 彭艳,陈加宏,李小毛,等. 时空上下文融合的无人艇海面目标跟踪[J]. 中国科学(技术科学), 2018, 48(12): 1357-1372.
PENG Y, CHEN J H, LI X M, et al. Sea surface object tracking for USV with spatio-temporal context fusion[J]. Scientia Sinica (Technologica), 2018, 48(12): 1357-1372.

- [7] 吴鹏,孙备,苏绍璟,等. 面向无人艇的航海雷达与光电吊舱协同环境感知方法[J]. 仪器仪表学报, 2021, 41(8): 154-163.
- WU P, SUN B, SU SH J, et al. Cooperative environment perception method for nautical radar and photoelectric pod facing to unmanned surface vehicles[J]. Chinese Journal of Scientific Instrument, 2021, 41(8): 154-163.
- [8] 李尚君,岳林,李哲,等. 一种面向异构传感器的无人水面艇海面目标识别跟踪系统[J]. 中国舰船研究, 2021, 16(S1): 131-137.
- LI SH J, YUE L, LI ZH, et al. Sea surface target recognition and tracking system for unmanned surface vehicle with heterogeneous sensors[J]. Chinese Journal of Ship Research, 2021, 16(S1): 131-137.
- [9] 程亮,杨渊,张云飞,等. 面向无人艇智能感知的水上目标识别算法研究[J]. 电子测量与仪器学报, 2021, 35(9): 99-104.
- CHENG L, YANG Y, ZHANG Y F, et al. Research on water target recognition algorithm for unmanned surface vessel[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(9): 99-104.
- [10] 孙备,左震,吴鹏,等. 面向无人艇环境感知的改进型SSD目标检测方法[J]. 仪器仪表学报, 2021, 42(9): 52-61.
- SUN B, ZUO ZH, WU P, et al. Object detection for environment perception of unmanned surface vehicles based on the improved SSD[J]. Chinese Journal of Scientific Instrument, 2021, 42(9): 52-61.
- [11] 柳晨光,郭珏菡,吴勇,等. 无人水面艇三维激光雷达目标实时识别系统[J]. 机械工程学报, 2022, 58(4): 202-211.
- LIU CH G, GUO J H, WU Y, et al. 3D LiDAR based real-time object recognition system for unmanned surface vehicles[J]. Journal of Mechanical Engineering, 2022, 58(4): 202-211.
- [12] BOVCON B, KRISTAN M. A water-obstacle separation and refinement network for unmanned surface vehicles[J]. 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020: 9470-9476.
- [13] BOVCON B, MUHOVIC J, VRANAC D, et al. MODS—A USV-oriented object detection and obstacle segmentation benchmark[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23: 13403-13418.
- [14] CHENG Y, ZHU J, JIANG M, et al. FloW: A dataset and benchmark for floating waste detection in inland waters[J]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 10933-10942.
- [15] ZUST L, KRISTAN M. Learning maritime obstacle detection from weak annotations by scaffolding[J]. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022: 1888-1897.
- [16] YANG C, ZHANG L, LU H, et al. Saliency detection via graph-based manifold ranking[J]. 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3166-3173.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [18] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230-6239.
- [19] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [20] FANG H, XIA M, ZHOU G, et al. Infrared small UAV target detection based on residual image prediction via global and local dilated residual networks[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [21] ZHAO T, WU X. Pyramid feature attention network for saliency detection[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 3080-3089.
- [22] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42: 2011-2023.
- [23] DENG H, SUN X, LIU M, et al. Small infrared target detection based on weighted local difference measure[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54: 4204-4214.

- [24] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 257-269.
- [25] LIN T Y, DOLLÁR P, GIRSHICK R B, et al. Feature pyramid networks for object detection [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936-944.
- [26] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [27] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection [J]. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 949-58.
- [28] QIN X, ZHANG Z V, HUANG C, et al. BASNet: Boundary-aware salient object detection [J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7471-81.
- [29] LI G, LIU Z, BAI Z, et al. Lightweight salient object detection in optical remote sensing images via feature correlation [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-12.
- [30] DAI Y, WU Y, ZHOU F, et al. Attentional local

contrast networks for infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59: 9813-24.

作者简介



童小钟, 分别于 2018 年和 2021 年于国防科技大学获得学士学位和硕士学位, 正在国防科技大学攻读博士学位, 主要研究方向为智能探测和目标感知。

E-mail: tongxiaozhong@nudt.edu.cn

Tong Xiaozhong received his B. Sc. and M. Sc. degree both from National University of Defense Technology in 2018 and 2021, respectively. He is currently a Ph. D. candidate at National University of Defense Technology. His main research interest includes intelligent detection and target perception.



魏俊宇(通信作者), 分别于 2011 年和 2016 年于国防科技大学获得硕士和博士学位, 现为国防科技大学助理研究员, 主要研究方向为智能探测和目标感知。

E-mail: yujy@nudt.edu.cn

Wei Junyu (Corresponding author) received his M. Sc. degree and Ph. D. degree both from National University of Defense Technology in 2011 and 2016, respectively. He is currently a lecturer at National University of Defense Technology. His main research interests include intelligent detection and target perception.