



研究与开发

复杂场景下多尺度船舶实时检测方法

周薇娜, 刘露

(上海海事大学信息与工程学院, 上海 201306)

摘要: 船舶检测在军事侦察、海上目标跟踪、海上交通管制等任务中发挥着重要作用。然而, 受船舶外形尺度多变和复杂海面背景的影响, 在复杂海面上检测多尺度船舶仍然是一个挑战。针对此难题, 提出了一种基于多层信息交互融合和注意力机制的 YOLOv4 改进方法。该方法主要通过多层信息交互融合 (multi-layer information interactive fusion, MLIF) 模块和多注意感受野 (multi-attention receptive field, MARF) 模块构建一个双向细粒度特征金字塔。其中, MLIF 模块用于融合不同尺度的特征, 不仅能将深层的高级语义特征串联在一起, 而且将较浅层的丰富特征进行重塑; MARF 由感受野模块 (receptive field block, RFB) 与注意力机制模块组成, 能有效地强调重要特征并抑制冗余特征。此外, 为了进一步评估提出方法的性能, 在新加坡海事数据集 (Singapore maritime dataset, SMD) 上进行了实验。实验结果表明, 所提方法能有效地解决复杂海洋环境下多尺度船舶检测的难题, 且同时满足了实时需求。

关键词: 多尺度船舶检测; 多层信息交互融合; 多注意感受野; 双向细粒度特征金字塔

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022258

A real-time detection method for multi-scale ships in complex scenes

ZHOU Weina, LIU Lu

School of information Engineering, Shanghai Maritime University, Shanghai 201306, China

Abstract: Ship detection plays an important role in tasks such as military reconnaissance, maritime target tracking, and maritime traffic control. However, due to the influence of variable sizes of ships and complex background of sea surface, detecting multi-scale ships remains a challenge in complex sea surfaces. To solve this problem, an improved YOLOv4 method based on multi-layers information interactive fusion and attention mechanism was proposed. Multi-layers information interactive fusion (MLIF) and multi-attention receptive field (MARF) were applied and combined reasonably to build a bidirectional fine-grained feature pyramid. MLIF was used to fuse feature of different scales, which not only concatenated high-level semantic features from deep layers, but also reshaped richer features from shallower layers. MARF consisted of receptive field block (RFB) and attention mechanism module, which ef-

收稿日期: 2022-02-11; 修回日期: 2022-08-15

基金项目: 国家自然科学基金资助项目 (No.52071200, No.61404083); 专用集成电路与系统国家重点实验室开放研究课题基金资助项目 (No.2021KF010)

Foundation Items: The National Natural Science Foundation of China (No. 52071200, No.61404083), The State Key Laboratory of ASIC and System (No. 2021KF010)



fectively emphasized the important features and suppressed unnecessary ones. In addition, to further evaluate the performance of the proposed method, experiments were carried out on Singapore maritime dataset (SMD). The results illustrate that the method proposed can effectively solve the problem of difficult detection of multi-scale ships in complex marine environment, and meet the real-time requirements at the same time.

Key words: multi-scale ship detection, multi-layers information interactive fusion, multi-attention receptive field, bidirectional fine-grained feature pyramid

0 引言

海洋环境中的视觉目标检测是计算机视觉领域的重要研究课题之一。准确、快速的多尺度船舶检测方法不仅能为码头管理、港口监控和安全航行提供非常重要的信息,而且对走私船舶的检测和海上救援也起着重要的作用。

传统的船舶检测算法大多基于船舶结构和形状进行手动特征设计。虽然这些算法^[1-3]取得了一些好的结果,但即使使用当时最好的非线性分类器对这些手动设计的特征进行分类,其检测精度仍不能满足实际需要^[4]。

近年来,随着深度学习技术的发展,卷积神经网络(convolutional neural network, CNN)在目标检测方面取得了越来越显著的性能。目前,基于 CNN 的目标检测算法可分为两类。

(1) 基于区域推荐的算法,形成了以 Faster R-CNN^[5-7]为代表的两阶段算法。该类算法由于具有较高的检测精度,在多尺度船舶检测中得到了广泛的应用,但检测速度不足以满足实时检测的要求。

(2) 基于回归的算法,形成了以 YOLO^[8-10]和 SSD^[11]为代表的单阶段算法。该类算法将检测问题转化为回归问题,大大提高了检测速度,在实时船舶检测中具有突出的优势。

目前应用于海上船舶检测算法大多基于合成孔径雷达(synthetic aperture radar, SAR)遥感图像,但随着 SAR 成像速度不断地提高,算法能够实时快速地检测显得尤为重要。如张佳欣等^[12]提出了一种基于改进 YOLOv3 的 SAR 图像船舶检测

方法,通过改进算法的检测策略提升了算法对复杂场景下不同船舶目标的适应性。Peng 等^[13]提出了一种由多个上采样注意单元组成的密集注意方法,该方法采用上采样空间和上采样通道注意机制,通过空间上下文信息捕获目标的变化特征。张筱晗等^[14]提出精简网络结构数据驱动的目标分布聚类算法,结合双向高低层特征融合机制,减少了训练和测试时间,但检测精度较低,同时模型复杂度较高。然而,由于卫星获取的 SAR 遥感图像中,场景种类比较单一,获取的船舶特征稳定性不足。

为了克服这个问题,研究者们提出了其他新的多尺度船舶检测方法^[15-17]。Shao 等^[15]构建了一个新的大型船舶数据集,其中包含 6 种常见的船舶类型,用于训练和评估船舶目标检测算法。Guo 等^[16]提出了一种新的基于海面无人水面车辆(unmanned surface vessel, USV)图像的目标检测算法,主要通过融合 DenseNet 和 YOLOv3 网络来实时检测真实海洋环境中的多尺度船舶。该方法在一定程度上能够提高检测的稳定性,以减少特征损失。Shao 等^[17]首次提出利用陆地监控摄像机网络捕获的视觉图像来进行实时船舶检测,并设计了一个重要的感知 CNN 框架来预测船舶的类别与位置。

然而,这些方法应用在真实海面环境下多尺度船舶检测时仍然存在一些难题,尤其当图像中背景信息复杂、包含大量不同尺度船舶、复杂环境中的尾流等噪声类似于小尺度船舶等。所有这些情况都会对多尺度船舶的准确检测造成很大的干扰。因此,增强上下文语义信息,并突出船舶

的显著特征,对于提高多尺度船舶的检测性能至关重要。

为了能同时兼顾检测精度和速度,本文提出了一种基于多层信息交互融合和注意力机制的网络结构。首先,采用 CSPDarkNet53^[18]作为主干网络进行特征提取;然后,在主干网络之后,采用改进的细粒度特征金字塔结构。特征金字塔主要包含两个关键模块,即多层信息交互融合(multi-layer information interactive fusion, MLIF)模块和多注意感受野(multi-attention receptive field, MARF)模块。MLIF 模块分别通过级联和重塑操作将较深层和较浅层的特征融合到当前层,它可以充分增强上下文语义特征。MARF 模块合理地结合了感受野模块(receptive field block, RFB)^[19]、空间注意(spatial attention, SA)模块和通道注意(channel attention, CA)模块^[20]。其中, RFB 旨在进一步扩展不同区域的感受野,获得更多的语义特征。SA 和 CA 模块用于突出船舶的重要特征,可有效抑制周围环境的干扰,例如船舶的波浪和尾迹。最后,在特征金字塔的不同尺度特征上使用 YOLOv4 检测模块生成最终的检测结果。该模块沿用了 YOLOv3 检测模块,因此在文中用 YOLOv3-Head 表示。此外,为了评估本文所提方法的性能,本文在新加坡海事数据集(Singapore maritime dataset, SMD)上进行实验, SMD 是视觉船舶检测的代表数据集之一。检测结果表明,与现有的 Faster-RCNN、SSD、YOLOv3、YOLOv4^[21]、RetinaNet^[22]、YOLOx^[23]和 CenterNet^[24]方法相比,所提方法对于复杂背景下的多尺度船舶检测更为有效。本文的主要贡献总结如下。

(1) 提出充分利用注意力机制,并结合网络中多层次信息的交互融合方法解决复杂环境下多尺度船舶的同步检测问题。

(2) 合理结合 MLIF 和 MARF 模块,引入细粒度特征金字塔方法以细化不同尺度特征图的语

义信息,增强非局部特征之间的关系。

(3) 构建了以 SMD 中多个船舶类别为基础的沿海、近海邻域背景的船舶数据集,并通过多角度实验比较,论证了所提方法的有效性和先进性。

1 算法原理

1.1 YOLOv4

目标检测网络通常使用一个非常深的主干网络沿自底向上的路径进行特征提取,但检测小尺度物体时,深的网络会造成物体语义信息或者位置信息的丢失。因此,一些检测方法通过设计一个性能良好的特征金字塔结构来提高多尺度物体检测的准确率,如 YOLOv4 和 RetinaNet。与之前的 YOLO 系列网络框架相比, YOLOv4 主干网络采用的 CSPDarkNet53 是 DarkNet53 的改进版本。YOLOv4 的特征金字塔部分是由特征金字塔网络(feature pyramid network, FPN)^[25]和路径聚合网络(path aggregation network, PAN)^[26]通过自顶向下路径将高级特征融合到低级特征来生成的。所采用的多尺度特征融合主要通过将 13×13、26×26、52×52 共 3 种不同尺寸的特征图构建特征金字塔,分别用于检测大、中、小目标。这种多尺度预测使得 YOLOv4 对弱目标更加敏感,显著提高了其检测能力。YOLOv4 的检测模块仍沿用 YOLOv3 的检测模块,但在损失函数部分进行了改进^[27]。所采用的损失函数具体如下。

$$L_{\text{total}} = L_{\text{conf}} + L_{\text{cls}} + L_{\text{ciou}} \quad (1)$$

其中, L_{conf} 、 L_{cls} 和 L_{ciou} 分别表示置信损失、分类损失和完全交并比(complete intersection over union, CIOU)损失。 L_{conf} 和 L_{cls} 由以下计算式表示。

$$L_{\text{conf}} = - \sum_{i=0}^{S^2} \sum_{j=0}^M I_{ij}^{\text{obj}} [\hat{C}_i \ln(C_i) + (1 - \hat{C}_i) \ln(1 - C_i)] - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^M I_{ij}^{\text{noobj}} [\hat{C}_i \ln(C_i) + (1 - \hat{C}_i) \ln(1 - C_i)] \quad (2)$$



$$L_{\text{cls}} = -\lambda_{\text{coord}} \sum_{i=0}^{s^2} I_i^{\text{obj}} \sum_{c \in \text{classes}} [\hat{p}_i(C) \ln(p_i(C)) + (1 - \hat{p}_i(C)) \ln(1 - \hat{p}_i(C))] \quad (3)$$

其中, s 表示网格的大小, M 表示先验框的数量, 变量 λ_{coord} 用于提高有对象的边界框的预测概率, λ_{noobj} 用于降低没有对象的边界框的预测概率。

L_{ciou} 的计算式如下。

$$L_{\text{ciou}} = 1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v \quad (4)$$

其中, $\rho^2(b, b^{\text{gt}})$ 代表 b 和 b^{gt} 之间的欧氏距离, c 表示可以同时包含预测框和对象边界框的最小封闭区域的对角线距离, v 用于度量纵横比的一致性, 并且 α 为正数。

1.2 改进方法的网络结构

YOLOv4 的特征金字塔结构主要是通过融合不同特征层来增强网络中上下文语义特征, 但这些来自不同特征映射的特征仍然不平衡。具体来说, 高级特征层中包含丰富的语义信息, 但缺乏准确的位置信息。而低级特征层中包含丰富的位置信息, 但语义信息较少。尤其是在复杂的海洋环境中, 不仅包含了不同尺度的船舶, 而且周围环境也非常复杂, 这就使得 YOLOv4 的检测效果不理想。此外, 随着检测网络层数加深, 小尺度的船舶特征就会消失, 那么沿自上而下路径融合

船舶特征是不可能的。因此, 为了增强上下文语义特征, 突出船舶的重要特征, 建立一个细粒度特征金字塔是非常有必要的。本文方法的网络结构如图 1 所示, 可以看出, 本文所提方法主要通过融合 MLIF 和 MARF 模块构造一个双向融合的特征金字塔。该特征金字塔结构主要包括空间金字塔池化 (spatial pyramid pooling, SPP) 模块^[28]、特征提取 (feature extraction, FE) 模块^[29]、MLIF 模块和 MARF 模块。

1.3 SPP 和 FE 模块

YOLOv4 的主干特征提取网络的最深层特征会经过 SPP 模块进行特征提取并扩大感受野, 然后经过 FE 模块进一步深层次地提取特征。SPP 和 FE 模块的具体结构如图 2 所示。SPP 模块由 1 个瓶颈层、3 个内核大小为 (5×5)、(9×9) 和 (13×13) 的最大池化层以及 1 个级联操作组成。它能够显著地增加感受野, 分离得到最重要的上下文特征, 几乎不会降低网络运行速度。FE 模块主要由两个瓶颈层和两个卷积层组成, 其中瓶颈层用于减少特征通道的数量, 卷积层用于提取特征。FE 模块主要用于生成更多上下文和语义特征。

1.4 MLIF 模块

Cao 等^[30]提出, 当主干网络非常深时, 可以通过融合更大规模的特征层来扩展小型船舶的语

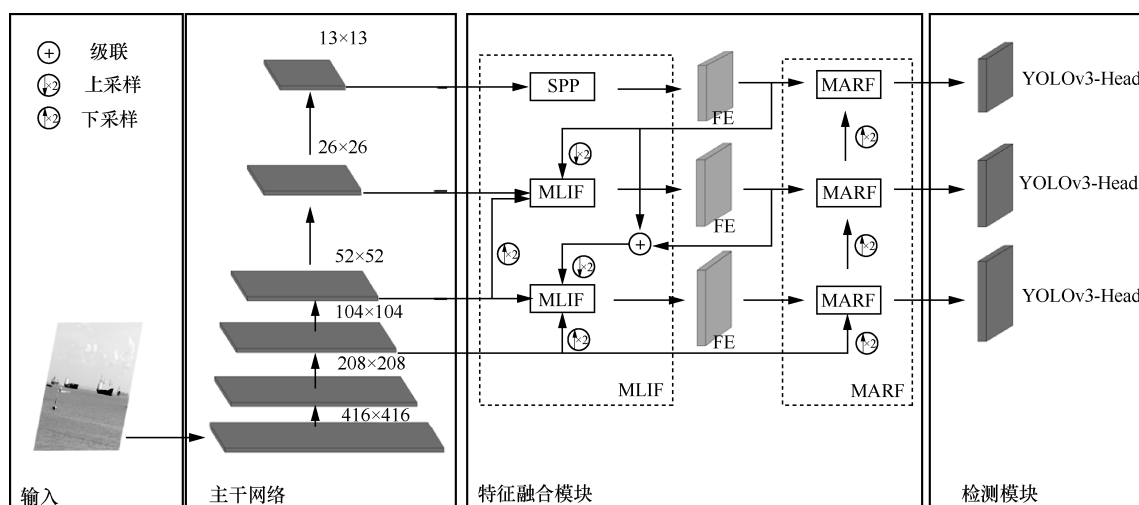


图 1 本文方法的网络结构

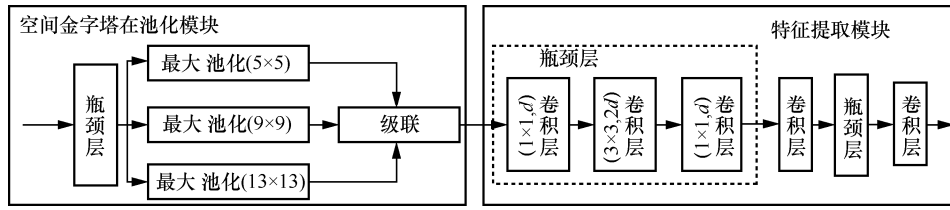


图2 SPP 和 FE 模块的具体结构

义信息。此外,低级特征中包含丰富的位置信息有利于物体的定位。因此,本文将主干网络中的4层不同尺度的特征层融入 MLIF 模块,从而将深层特征的语义信息与浅层特征的边缘特征信息结合起来。这种操作可以使算法感知更多的小尺寸船舶的特征,有利于海洋环境中多尺度船舶的检测。

MLIF 模块的具体结构如图 3 所示。从图 3 可以看到 MLIF 和 MLIF-s 模块分别用于主干特征网络的 $f_{26 \times 26}$ 和 $f_{52 \times 52}$ 特征层,它们主要通过两个级联操作融合不同层的特征。在 MLIF 模块中,级联 1 连接当前特征 $f_{26 \times 26}$ 和深层特征 $f_{13, FE}$, 级联 2 连接当前特征 $f'_{26 \times 26}$ 和浅层特征 $f_{52 \times 52}$ 。在 MLIF-s 模块中,级联 1 连接当前特征 $f_{52 \times 52}$ 和深层特征 $f_{26, FE}$, 级联 2 连接当前特征 $f'_{52 \times 52}$ 、深层特征 $f_{13, FE}$ 和浅层特征 $f_{104 \times 104}$ 。具体实现过程由以下计算式描述。

$$f_{13, FE} = F_{FE} (F_{SPP} (f_{13 \times 13})) \quad (5)$$

$$f'_{26 \times 26} = \text{Conv}_{1 \times 1} (\text{Concat} (f_{13, FE}, f_{26 \times 26})) \quad (6)$$

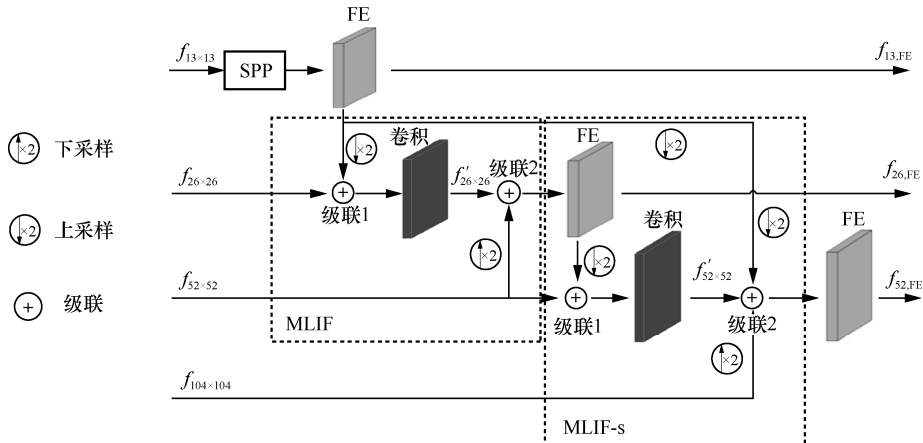


图3 MLIF 模块的具体结构

$$f_{26, FE} = F_{FE} (\text{Concat} (f'_{26 \times 26}, f_{52 \times 52})) \quad (7)$$

$$f'_{52 \times 52} = \text{Conv}_{1 \times 1} (\text{Concat} (f_{26, FE}, f_{52 \times 52})) \quad (8)$$

$$f_{52, FE} = F_{FE} (\text{Concat} (f'_{52 \times 52}, f_{13, FE}, f_{104 \times 104})) \quad (9)$$

1.5 MARF 模块

MARF 模块的结构和参数如图 4 所示。特征层 f_n 和 f_{2n} ($n = 13, 26, 52$) 主要来自 CSPDarkNet53 主干特征网络。这些特征首先经过 RFB 扩大特征接收的范围,以进一步细化小尺度船舶的语义特征。然后将注意力机制中 SA 模块和 CA 模块以并行的方式结合在一起进行特征融合,最后输出特征层。MARF 模块的输出可表示为:

$$f_{out} = \text{Concat} (F_{SA} (f_{2n}), F_{CA} (\text{Conv} (F_{RFB} (f_n)))) \quad (10)$$

其中, RFB 主要通过 1×1 或 3×3 卷积运算减少通道数,进而减少网络的计算量。然后采用膨胀卷积并行捕获多尺度船舶的上下文语义信息。此外, RFB 有较大的感受野和较小的网络参数。RFB 主要由 3 个卷积分支和 1 个级联操作组成。每个分支包含 3 个卷积层,最后一层是膨胀率分别为 1、

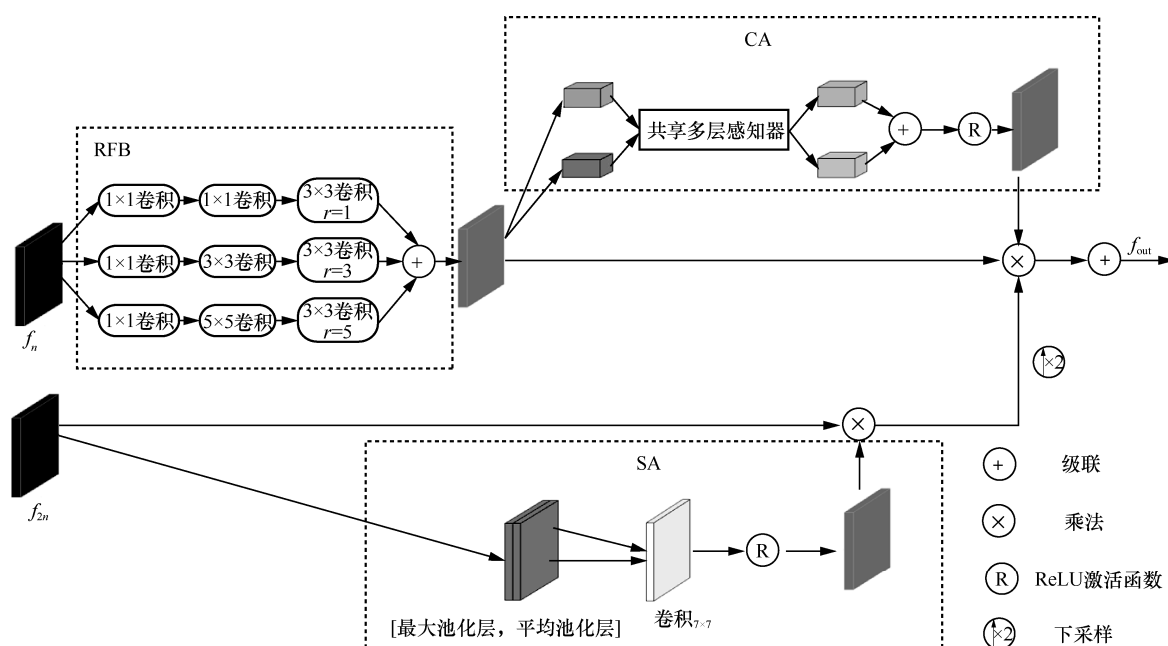


图4 MARF 模块的结构和参数

3、5的膨胀卷积，用于捕获全局特征。由于大膨胀率卷积可以获得船舶的非局部特征，如轮廓、边缘等，因此适合于多尺度船舶的检测。通过级联操作，在相同的特征映射上增强船舶的局部特征及其全局依赖性。最终的输出特征在校正线性单元函数后获得，用于改善 RFB 的非线性。具体可由以下计算式表示。

$$F_{RFB} = \text{CONCAT} \left\{ \text{Conv}_{3 \times 3, r=1} \left(\text{Conv}_{1 \times 1} \left(\text{Conv}_{1 \times 1} (f_n) \right) \right), \right. \\ \left. \text{Conv}_{3 \times 3, r=3} \left(\text{Conv}_{3 \times 3} \left(\text{Conv}_{1 \times 1} (f_n) \right) \right), \right. \\ \left. \text{Conv}_{3 \times 3, r=5} \left(\text{Conv}_{5 \times 5} \left(\text{Conv}_{1 \times 1} (f_n) \right) \right) \right\} \quad (11)$$

在通道注意力模块中，经过 RFB 的特征图谱首先沿着其宽度和高度平行采用最大池化层和平均池化层；然后，使用共享多层感知器沿特征图谱的通道维度输出权重；接着，将共享 MLP 应用于每个描述符之后，使用元素求和来合并输出特征向量；最后经过 ReLU 激活函数生成通道注意特征图谱。在空间注意力模块中，浅层特征首先沿着其通道轴应用最大池化和平均池化操作，并将它们连接起来，以生成有效的特征图谱；然后

在级联的特征图谱上应用卷积层进行进一步特征提取；最后同样经过 ReLU 激活函数生成空间注意特征图谱。经过 CA 和 SA 模块生成的特征都经过与当前特征进行相乘后，浅层特征再经过下采样与通道注意特征进行元素求和合并输出最终的特征向量。通道注意特征和空间注意特征的计算式如下。

$$F_{CA} = F_{\text{ReLU}} \left(\text{MLP} \left(\text{Maxpool}(F) \right) + \text{MLP} \left(\text{Avgpool}(F) \right) \right) \quad (12)$$

$$F_{SA} = F_{\text{ReLU}} \left(\text{Conv}_{7 \times 7} \left(\text{Maxpool}(F) \right) + \text{Avgpool}(F) \right) \quad (13)$$

2 实验结果分析

2.1 数据集

本文数据集中的图像主要来自新加坡海事数据集 (Singapore maritime dataset, SMD)。SMD 提供由分别固定在岸上和船上的佳能 70D 相机拍摄的新加坡水域周围的高清格式 (1 080×1 920 像素) 的可见光 (visual-optical, VIS) 和近红外 (near infrared, NIR) 视频。它包含 81 个视频文件，包

括 9 个不同类别的 240 842 个目标标签。SMD 中的图像包含大量不同的船舶，如浮标、快艇、皮划艇和轮船，且这些船舶在图中的尺度变化较大。这对于检测算法来说是非常具有挑战性的。本文将 SMD 随机分为 3 个部分，即训练集、验证集和测试集，其比例为 7:1:2，并从 SMD 中提取 4 470 张图像作为训练集，635 张图像作为验证集，1 245 张图像作为测试集。SMD 的部分样本和标签如图 5 所示。



图 5 SMD 的部分样本和标签

2.2 评估指标

本文采用的算法评估指标为：准确率 (precision, P)、召回率 (recall, R)、平均准确率 (average precision, AP) 与 F 分数 (F1)。精度均值 (mean average precision, mAP) 是多个类别的 AP 的平均值，用来衡量模型在所有类别上训练效果的好坏。F1 是一个综合指标，用于结合准确率和召回率来判断不同网络的性能。此外，为了评估目标检测网络的检测速度，本文使用每秒传输帧数 (frames per second, FPS) 作为网络检测速度评估标准，其计算式如下。

$$\text{FPS} = \frac{1}{T_{\text{per-img}}} \quad (14)$$

其中， $T_{\text{per-img}}$ 是处理图像时方法的推断时间成本。

2.3 参数设置

本实验使用的配备是 Intel i5-8 400HQ (2.5 GHz)

CPU 和 Nvidia GeForce GTX 1 070 GPU，均采用 Python 语言编写，且均使用 PyTorch 平台。在网络训练中，采用初始学习率为 0.01 的多项式衰减学习率调度策略，动量为 0.9 的标准 Adam 优化器，同时标签平滑设置为 0.01。每批次输入网络的图像个数则根据网络框架的深度和 GPU 内存限制设置为 16，并设置分割系数为 16，让网络能够最大限度地提高 GPU 的内存利用率，同时实现网络的快速训练。此外，通过采用提前停止的策略，使网络能够在 100 个 Epoch 内终止训练过程，从而加快网络的训练速度。

2.4 消融实验

本文通过提出双向“细粒度”融合特征金字塔模块对多尺度船舶进行了更为精确的检测，其中 MLIF 和 MARF 模块的改进正是细粒度融合特征金字塔模块的主要部分。因此，为了评估本文算法对 SMD 图像中多尺度船舶检测的有效性，以 CSPDarkNet53 主干特征网络为基础，在 SMD 数据集上进行 MLIF 和 MARF 模块的消融实验，定量分析了两个模块的检测性能，具体见表 1。

表 1 不同模块精度对比

FPN	PAN	MLIF	MARF		mAP	FPS
			RFB	SA,CA		
√	√				0.648	27.2
	√	√			0.696	27.1
√			√		0.684	27.5
√				√	0.632	27.7
√			√	√	0.701	26.5
		√	√		0.733	26.1
		√		√	0.704	26.3
		√	√	√	0.769	26.0

表 1 显示了船舶检测网络应用各个模块的检测结果，从表 1 可以看出，在 FPS 基本一致的情况下，MLIF-PAN 的 mAP 值比 FPN-PAN (YOLOv4) 高 4.8%，说明 MLIF 模块能够提高网络的检测能力。FPN-RFB 的 mAP 值比 FPN-PAN 提高了 3.6%，MLIF-RFB 的 mAP 值比 MLIF-PAN 提高了 3.7%。这是因为在复杂环境中



RFB 模块可以通过增强局部特征及其全局依赖性,如波浪、相机抖动等,获取丰富的语义信息进一步准确地检测多尺度船舶。在另一方面,仅采用 SA 和 CA 模块来进行特征融合时,FPN-SA、CA 和 MLIF-SA、CA 的 mAP 值会有所降低,这主要是由于该方法缺乏 RFB 模块来扩大船舶感受野。此外,MLIF-MARF 的 mAP 值比 MLIF-RFB 高,FPN-MARF 的 mAP 值比 FPN-RFB 高。这主要是因为注意力机制有效抑制了来自 RFB 的冗余特征,网络可以更加突出船舶的显著特征并对其进行区分,证明了 MARF 模块在提高多尺度船舶检测精度方面具有泛化能力。

为了进一步验证所提方法对多尺度船舶的检测能力,本文分别比较了 FPN-PAN、MLIF-PAN 和本文方法(MLIF-MARF)在检测不同尺寸船舶方面的性能。不同模块对多尺度船舶检测精度对比见表 2,可见本文方法 mAP 指标均优于 FPN-PAN(YOLOv4),在检测小尺寸船舶方面,比 FPN-PAN 提升高达 40.9%的准确率。这主要是因为本文方法能够有效地增强非局部特征之间的关系,充分整合不同特征图上的语义信息。

表 2 不同模块对多尺度船舶检测精度对比

对比项	mAP			FPS
	大尺寸	中尺寸	小尺寸	
FPN-PAN	0.675	0.375	0.192	27.2
MLIF-PAN	0.749	0.665	0.511	25.1
本文算法 (MLIF-MARF)	0.781	0.679	0.601	26.0

图 6 比较了 FPN-PAN、MLIF-PAN 和本文方法的直观检测效果。从 SMD 中选择了 3 组不同场景下的船舶图像,在船上拍摄的画面,船舶比例小、排列稀疏的可见光图片如图 6(a)所示;在岸上拍摄的船舶比较密集,且包含大量不同尺度的船舶可见光图片如图 6(b)所示;光线比较昏暗的近红外光船舶图片如图 6(c)所示。根据 3 组不同场景的检测结果,以及和真值(ground truth)图的对比,本文可以得出如下结论。

(1) MLIF 和 MARF 两个模块能够有效地提高网络对周围环境和船舶的检测性能。因为在图 6 中,3 组场景下本文算法(MLIF-MARF)的检测结果均不存在误报,而 MLIF-PAN 和 FPN-PAN 的检测结果都会出现漏检的情况。

(2) MARF 模块能够增强具有全局依赖性的

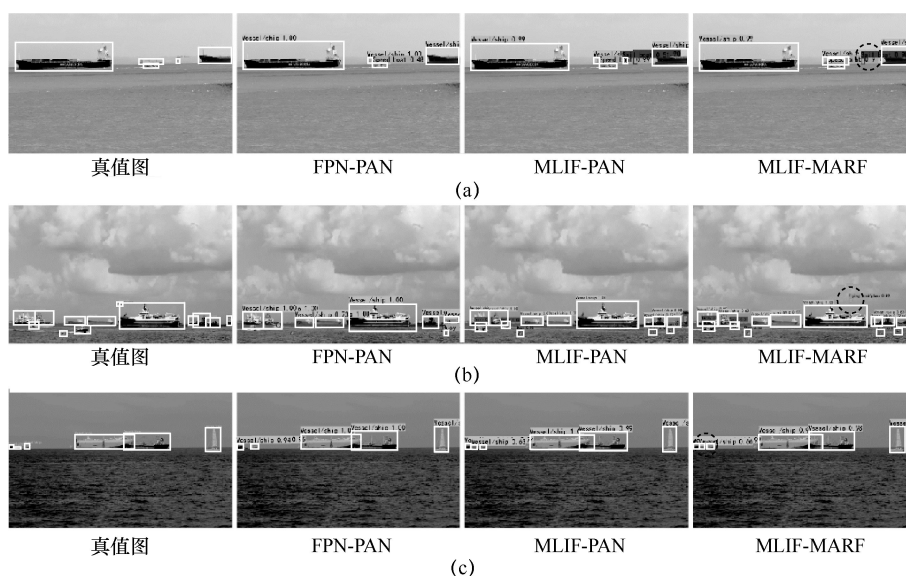


图 6 FPN-PAN、MLIF-PAN 与本文算法(MLIF-MARF)在不同环境下的结果对比

局部特征,并且进一步细化小尺度物体的显著特征。这点可由图 6(b)和图 6(c)中看出,本文算法(MLIF-MARF)能够准确检测虚线圆圈框中的小物体,而 FPN-PAN、MLIF-PAN 结构却不能。

此外,为了验证算法对背景特征的抑制作用,各个模块特征激活图如图 7 所示,分别显示了 FPN-PAN、MLIF-PAN 和本文算法(MLIF-MARF)3 种方法在不同环境下的特征激活图。从图 7 中可以清晰看出 YOLOv4(FPN-PAN)算法受背景干扰,检测存在很多漏检和误检现象;MLIF-PAN 算法虽然使得船舶检测率有所提升,但周边的环境对算法仍然造成了很大的影响。而本文方法(MLIF-MARF)在增加了注意力机制模块之后,对周边环境的冗余特征抑制作用较大,能够准确地识别船舶特征并将周边环境对它的影响降到最低。

2.5 对比最先进的技术

为了进一步验证本文算法在检测多尺度船舶方面的性能,将本文所提算法与 7 种具有代表性的目标检测方法 Faster-RCNN、SSD、RetinaNet、CenterNet、YOLOx、YOLOv3 和 YOLOv4 进行对比实验,这里均采用相同的数据划分和优化参数方法进行训练,与其他目标检测方法的对比实验结果见表 3。

表 3 与其他目标检测方法的对比实验结果

方法	R	P	F1	mAP	FPS
Faster-RCNN ^[7]	0.741	0.851	0.786	0.775	7.3
RetinaNet ^[22]	0.696	0.784	0.731	0.703	22.9
SSD ^[11]	0.306	0.768	0.403	0.473	40.0
CenterNet ^[24]	0.673	0.794	0.712	0.688	52.3
YOLOx ^[23]	0.644	0.775	0.689	0.670	59.9
YOLOv3 ^[10]	0.512	0.780	0.593	0.579	10.4
YOLOv4 ^[21]	0.618	0.747	0.661	0.648	27.2
本文算法	0.678	0.865	0.744	0.769	26.0

由表 3 可以看出,两阶段算法 Faster-RCNN 的准确率要高于单阶段的算法,这是因为两阶段算法比单阶段算法更注重准确性,但其检测速度很慢、达不到实时检测效果。相比于其他单阶段的检测算法,本文提出的检测算法在检测精度方面表现最好,远高于 YOLOv4 算法,主要是因为 YOLOv4 通过融合 3 层特征图来获取船舶的细粒度特征信息,这使其不能够充分地提取小尺度的船舶语义信息。此外,RetinaNet 的 mAP 值相比于其他算法也是比较高的,这主要是因为该算法同样具有一个融合多层特征信息的金字塔结构,但该算法的船舶检测速度要低于本文所提算法。另一方面,从 F1 角度来看,本文所提算法比其他

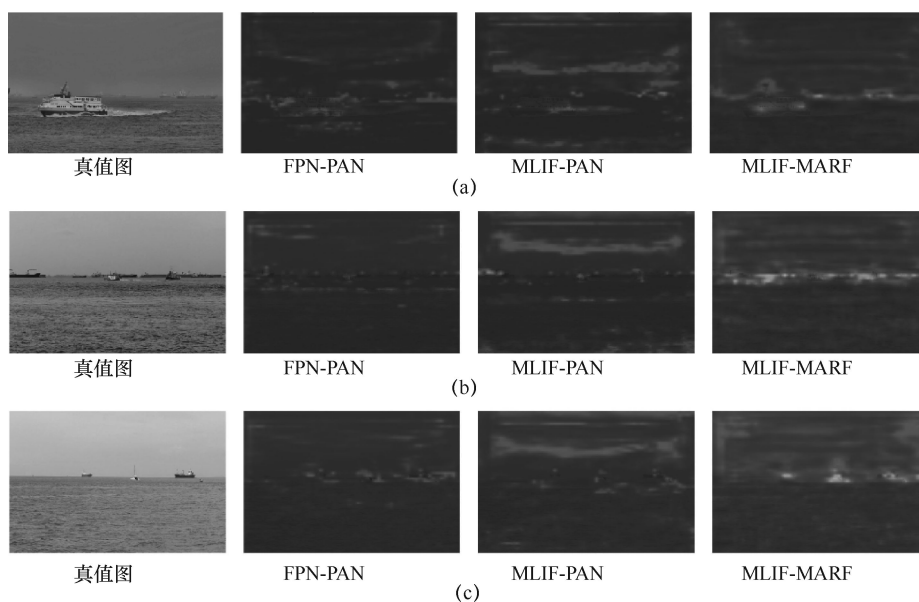


图 7 FPN-PAN、MLIF-PAN 与本文算法(MLIF-MARF)在不同环境下的特征激活图



算法提高了 34.1%，进一步证明本文算法的综合性能较好。最后，根据表 3 最后一列所示的 FPS 值可见，本文所提算法的 FPS 比 Faster-RCNN 快得多。其中 YOLOx 的运行速度比其他方法快，这可能是因为它在骨干网络中采用了深度可分离卷积，且属于无锚框算法。但本文所提算法仍能满足实时性要求。

图 8 中展现了 5 种检测算法在 SMD 上的检测结果。图 8 (a) 为相机固定在船舶上采集的，这类图片的主要特点是包含较少的船舶且占据整张图片的比例较小，这会严重干扰每种算法的准确率。从图 8 (a) 可以看出，本文的算法准确地检测到所有船舶目标，而对于 SSD 则没有检测出船舶 A。图 8 (b) 中每张图片都包含了许多不同尺寸的船舶目标且它们都密集分布。从图 8 (b) 可

以看出，SSD、Faster-RCNN、YOLOv4 和 CenterNet 都存在漏检（飞机 B 或小船 C）。而对于本文的算法由于采用了细粒度特征金字塔结构，因此可以准确地检测到每种船舶目标。图 8 (c) 中的船舶图片则主要来自近红外视频，这类图片的整体画面光线比较昏暗。从图 8 (c) 可以看出，在 YOLOv4、SSD 和 CenterNet 的检测结果中，一些检测目标或多或少地被漏检了（飞机 D、船舶 E、船舶 F 和船舶 G）。而本文算法同样能够准确检测到各种船舶目标。以上 3 种场景的船舶检测结果充分证明了本文所提算法能够通过细粒度特征金字塔模块，增加网络自下向上的信息融合路径，提供更准确的网络低层特征细节信息，从而进一步提高多尺度船舶的检测准确率。

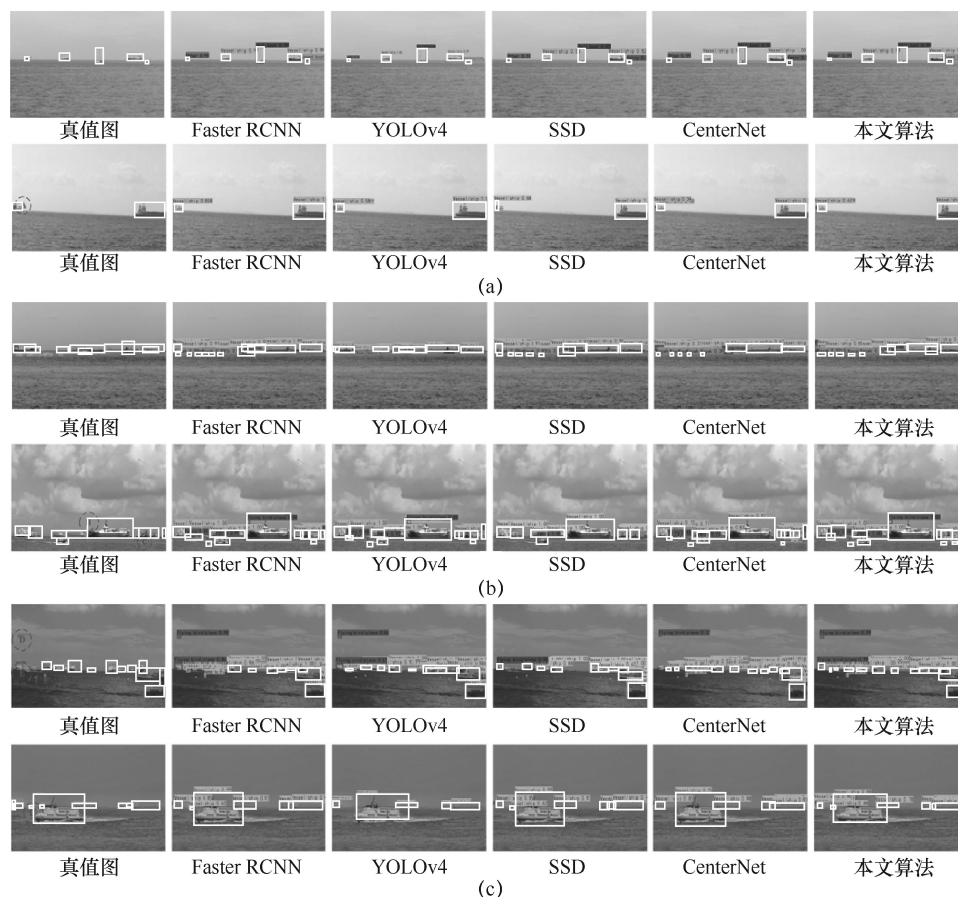


图 8 不同算法检测结果对比

3 结束语

为解决多尺度船舶检测难的问题, 本文提出了一种基于 YOLOv4 的高效多层注意力接收融合网络方法。该方法主要在特征金字塔部分融合了 MLIF 和 MARF 模块, 使网络能够在多尺度上融合特征, 充分增强上下文语义特征, 扩展感知范围, 提高了多尺度舰船的检测精度。实验结果表明, 改进后的方法在复杂的海洋环境下对多尺度船舶具有良好的检测效果, 与 YOLOv4 相比, 检测的准确度有大幅提升。

参考文献:

- [1] HUANG J, JIANG Z G, ZHANG H P, et al. Region proposal for ship detection based on structured forests edge method[C]//Proceedings of 2017 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2017: 1856-1859.
- [2] ZHU Q Y, JIANG Y L, CHEN B. Design and implementation of video-based detection system for WHARF ship[C]//Proceedings of IET International Conference on Smart and Sustainable City 2013 (ICSSC 2013). IET, 2013: 493-496.
- [3] LI S, ZHOU Z Q, WANG B, et al. A novel inshore ship detection via ship head classification and body boundary determination[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1920-1924.
- [4] LIU L, WANG X G, CHEN J, et al. Deep learning for generic object detection: a survey[J]. International Journal of Computer Vision, 2020, 128(2): 261-318.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 27th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 1440-1448.
- [7] 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.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 779-788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 7263-7271.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB]. 2018: arXiv.1804.02767.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
- [12] 张佳欣, 王华力. 改进 YOLOv3 的 SAR 图像舰船目标检测[J]. 信号处理, 2021, 37(9): 1623-1632.
ZHANG J X, WANG H L. Ship target detection in SAR image based on improved YOLOv3[J]. Journal of Signal Processing, 2021, 37(9): 1623-1632.
- [13] PENG X L, ZHONG R F, LI Z, et al. Optical remote sensing image change detection based on attention mechanism and image difference[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(9): 7296-7307.
- [14] 张筱晗, 姚力波, 吕亚飞, 等. 双向特征融合的数据自适应 SAR 图像舰船目标检测模型[J]. 中国图象图形学报, 2020, 25(9): 1943-1952.
ZHANG X H, YAO L B, LYU Y F, et al. Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images[J]. Journal of Image and Graphics, 2020, 25(9): 1943-1952.
- [15] 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.
- [16] LI Y, GUO J, GUO X, et al. A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved YOLOv3[J]. Sensors, 2020, 20(17): 4885.
- [17] SHAO Z, WANG L, WANG Z, et al. Saliency-aware convolution neural network for ship detection in surveillance video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(3): 781-794.
- [18] WANG C Y, LIAO H, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 390-391.
- [19] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 385-400.
- [20] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [21] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[EB]. 2020: arXiv. 2004.10934.
- [22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International



- Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2999-3007.
- [23] GE Z, LIU S, WANG F, et al. YOLOx: exceeding yolo series in 2021[EB]. 2021.
- [24] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6569-6578.
- [25] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2117-2125.
- [26] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.
- [27] 康帅, 章坚武, 朱尊杰, 等. 改进 YOLOv4 算法的复杂视觉场景行人检测方法[J]. 电信科学, 2021, 37(8): 46-56.
- KANG S, ZHANG J W, ZHU Z J, et al. An improved YOLOv4 algorithm for pedestrian detection in complex visual scenes[J]. Telecommunications Science, 2021, 37(8): 46-56.
- [28] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [29] CHEN P Y, HSIEH J W, WANG C Y, et al. Recursive hybrid fusion pyramid network for real-time small object detection on embedded devices[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 402-403.
- [30] CAO C, WU J, ZENG X, et al. Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network[J]. Sensors, 2020, 20(17): 4696.

[作者简介]



周薇娜(1982-), 女, 博士, 上海海事大学信息工程学院副教授、硕士生导师, 主要研究方向为图像处理、目标检测算法和 ASIC 设计。



刘露(1996-), 女, 上海海事大学信息工程学院硕士生, 主要研究方向为计算机视觉、目标检测与人工智能等。