

融合多尺度和上下文的无人机救援小目标检测算法

刘园¹, 赵静¹, 蒋国平¹, 徐丰羽¹, 陆宁云²

(1. 南京邮电大学自动化学院、人工智能学院, 江苏 南京 210023;

2. 南京航空航天大学航空航天结构力学及控制全国重点实验室, 江苏 南京 210016)

摘要: 针对无人机 (UAV, unmanned aerial vehicle) 图像中小目标所包含的特征信息少, 导致模型检测精度不足的问题, 面向无人机海面救援任务提出了一种融合多尺度和上下文信息的图像小目标检测算法。首先, 针对小目标特征信息设计上下文增强模块, 通过增强特征层的上下文信息, 有效地增加了模型对小目标的处理能力。其次, 为提高模型的鲁棒性, 设计了空间注意力模块加强对重要特征的学习。最后, 使用平衡 L1 损失函数优化基线算法的损失函数, 加强了模型检测时的稳定性。基于 Tiny-Person 数据集, 与基准算法进行大量实验对比, 所提算法在 AP50_tiny 上提高了 2.06%, 一定程度上提高了对海面小目标的检测性能, 对救援行动具有积极影响。

关键词: 海面救援; 无人机图像; 小目标检测; 注意力机制; 特征融合

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2024.00390

Fusion of multi-scale and context for small target detection algorithm of unmanned aerial vehicle rescue

LIU Yuan¹, ZHAO Jing¹, JIANG Guoping¹, XU Fengyu¹, LU Ningyun²

1. College of Automation & College of Artificial Intelligence, Nanjing University of Post and Telecommunication, Nanjing 210023, China

2. State Key Laboratory of Mechanics and Control for Aerospace Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract: Aiming at the problem of insufficient feature information contained in small targets under unmanned aerial vehicle (UAV) images that led to insufficient detection accuracy of the model, a small target detection algorithm for UAV sea rescue images that integrated multi-scale and contextual information was proposed. Firstly, context enhancement module was designed for small target feature information, which effectively enhanced the ability of the model to process small targets by enhancing the contextual information of the feature layer. Secondly, to improve the robustness of the model, spatial attention module was designed to enhance the learning of important features. Finally, balance L1 loss was used to optimize the loss function of the baseline algorithm and enhance the stability of the model during the process of detection. Based on the Tiny-Person dataset, through extensive experimental comparison with the benchmark algorithm, the proposed algorithm improves the detection performance of small targets on the sea surface by 2.06% on AP50_tiny, which has a positive impact on rescue operations.

收稿日期: 2023-12-15; 修回日期: 2024-05-31

通信作者: 赵静, zhaojing@njupt.edu.cn

基金项目: 国家自然科学基金项目(No.51775284); 直升机动力学全国重点实验室开放课题(No.2024-ZSJ-LB-02-05); 航空航天结构力学及控制国家重点实验室开放课题(No.MCMS-E-0123G04); 工业控制技术国家重点实验室开放课题(No.ICT2023B21); 南京邮电大学校级自然科学基金(No.NY223119)

Foundation Items: The National Natural Science Foundation of China (No.51775284), The Open Research Project of the National Key Laboratory of Helicopter Aeromechanics (No.2024-ZSJ-LB-02-05), The Open Research Project of the State Key Laboratory of Aerospace Structural Mechanics and Control, Nanjing University of Aeronautics and Astronautics, China (No.MCMS-E-0123G04), The Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No.ICT2023B21), The Natural Science Foundation of Nanjing University of Posts and Telecommunications (No. NY223119)

Key words: sea rescue, unmanned aerial vehicle image, small object detection, attention mechanism, feature fusion

0 引言

海面救援是一项充满挑战的任务,面临着多重技术难点。首先,由海浪和海流等因素引起的海洋环境复杂多变,使得救援行动充满风险;其次,迅速定位失踪目标也是一项艰巨的任务,海域广阔,搜索范围庞大,这增加了救援难度;最后,通信障碍和远离基地的救援区域进一步加大了危机应对的难度。因此,有效的海面救援战略和技术实现变得尤为迫切,本文旨在通过对小目标进行迅速而精准地识别与定位,以提高救援整体效率和成功率。与传统的人力和船只搜索相比,无人机体积小、飞行速度快,不受地理限制,可以在海面上空进行搜索和监视,覆盖范围广泛,可迅速调度至事发地点,结合计算机视觉和深度强化学习的智能算法,利用多种传感器,如高分辨率摄像头和红外热像仪等,提供全方位的环境感知和数据采集,为海面救援提供实时数据,提高了搜索效率。以马来西亚航空公司MH370客机失联事件为例,无人机在海面救援中发挥了独特的作用,能够快速准确地识别海上的紧急信号、漂浮物体或沉船残骸,实现对庞大海域的高效搜索,为及时救援提供了智能、迅速、科学的解决方案,凸显了其在海面救援任务中的不可替代性。

然而,无人机图像中一般是大量低分辨率的小目标。小目标的定义通常涉及相对尺寸和绝对尺寸两个方面。其中,相对尺寸是目标与图像的相对比例。Chen等^[1]将小目标定义为同一类别中,目标边界框面积与图像面积之比的中位数在0.08%~0.58%之间。然而,这种定义方式容易受到图像处理的影响,从而对模型的检测性能不能令人信服。绝对尺寸是根据目标的绝对像素大小来定义小目标,以COCO^[2]数据集为例,小目标被定义为分辨率小于32×32像素的目标。而在航空图像的小行人数据集Tiny-Person^[3]中,小目标的像素值范围被定义为区间[20, 32],并且进一步将微小目标的像素值范围限定为区间[2, 20]。这种统一标准有助于确保模型在处理小目标时一定程度上具有一致性和普适性。

近年来,得益于深度学习算法中卷积神经网络强大的特征提取能力,无人机图像的目标检测技术

面临着新的发展机遇^[4]。针对目标对象的深度强化学习检测算法可以分为单阶段和双阶段两类,目前常见的单阶段目标检测算法有SSD (single shot multibox detector)^[5]、YOLO (you only look once)^[6]系列等。Liu等^[5]提出的SSD算法率先引入多尺度的思想,对小目标的检测性能相比于YOLO算法有了较好的提升,但由于缺乏深层语义信息,无法得到充分的训练,在实际应用中检测性能显示不佳^[7]。双阶段目标检测算法是基于卷积神经网络(CNN, convolutional neural network)改进的,典型的有Fast R-CNN (fast region-based convolutional neural network)^[8]和Faster R-CNN^[9]等算法网络,其精度较单阶段更高。2014年, Girshick等^[10]提出了R-CNN算法,开创了将深度学习应用到目标检测领域的先河。后期, Fast R-CNN和Faster R-CNN在保证检测精度的同时,加快了模型的运算速度。2017年, He等^[11]提出了Mask R-CNN,这一网络模型在目标检测的基础上生成目标的精确分割掩模,使得算法模型的功能更加多样化和全面化。2022年, 阎海庆等^[12]针对识别目标的定位精度和稳定性较差的问题,提出了改进Libra R-CNN,改善了目标检测识别的平滑性能。Lin等^[13]针对目标检测器中存在的正负样本极度不平衡的问题,设计了RetinaNet结构,实现了对不平衡样本更好的处理,但在处理小尺寸目标时,由于RetinaNet特征分辨率不足从而导致检测性能下降。基于小目标在无人机图像中尺寸较小、对定位精度的要求更高、以及会在更高分辨率的特征图上失去细节等局限性,极易导致难以捕捉小目标的精细特征,典型的检测算法的检测效果性能不佳。

针对小目标检测的问题,在单阶段和双阶段深度学习算法的基础上,相继出现了很多的优化算法。2015年, He等^[14]提出了空间金字塔池化网络(SPPNet, spatial pyramid pooling network)结构,解决了R-CNN结构无法接收不同尺寸输入的问题,但SPPNet是基于固定的网格划分,在特征提取过程中易丢失小目标的细节信息。PANet^[15]结构利用自适应特征池来提升对小目标的检测能力,但在一些小目标密集的数据集上,需要经过多次实验寻找适当的超参数配置,耗费大量的时间和计算资源。

Cascade R-CNN^[16]模型在 Faster R-CNN 的基础上, 提出不同的正负样本交并比 (IoU, intersection over union) 阈值, 从而提高目标检测精度, 但对于小目标, 级联结构可能无法提供足够的精确性和鲁棒性。Sang 等^[17]利用注意力机制选择性地强调特征信息, 为网络模型特征提取带来更有用的细节信息。Wang 等^[18]提出了 MSCANet 模型, 通过与不同的感受野紧密连接来聚合不同尺度的上下文特征, 从而捕捉多尺度上下文信息, 解决了密集采样下尺度变化的问题。此外, 文献[3]提出了一种新的微小行人研究基线, 采用尺度匹配 (SM, scale match) 算法, 对微小行人进行检测, 开创了一个新的研究方向。

在进行海面救援行动时, 这些新方法为实际应用提供了更多可能性。通过研究分析各种算法的模型结构, 针对海面救援小目标特征信息少, 对定位精度的要求高等挑战, 综合考虑检测识别的精度和稳定性, 选用模型需对微小目标有一定的检测能力^[19]。基于 SM 算法, Hong 等^[20]提出了尺度选择金字塔网络 (SSPNet, scale selection pyramid network), 在处理任意输入尺寸的图像基础上, 能够满足检测速度和稳定性要求, 在实际应用中具有较大优势。因此, 本文选取 SSPNet 模型作为基线算法进行改进, 以得到更加丰富的特征信息, 提高检测性能, 更准确地找到待救援人员的位置。其基本结构主要包括骨干网络模块、区域候选网络 (RPN, region proposal network) 模块、区域池化 (ROI pooling, region of interest pooling) 模块和分类回归模块 4 个部分。根据卷积网络性能、训练样本集、小目标的像素值等因素, 本文采用 ResNet50^[21]作为特征提取的主干网络, 进行特征值的处理与传递, 使得模型能够在较少层数下获得较好的训练效果^[22]。此外, 本文提出了结合上下文增强模块的复合特征金字塔 (FPN, feature pyramid network) 结构, 以及一种新的 Head 网络结构, 该网络结构中主要包括空间注意力模块和平衡损失函数。本文的主要贡献体现在以下 3 个方面:

1) 设计融合多尺度空洞卷积模块 (MACM, multi-scale atrous convolution module) 和上下文增强模块 (CEM, context enhancement module) 的复合 FPN 结构, 通过对传统的 FPN 结构各个特征层进行深入的特征提取, 针对不同的感受野进行进一

步优化, 随后将各层的特征进行融合, 从而更好地捕捉输入图像中的上下文信息, 减少特征提取过程中小目标关键特征的丢失;

2) 设计空间注意力模块 (SAM, spatial attention module), 将经过注意力分配提取的特征图与前一模块处理后的特征图进行融合, 实现对注意力权重的再分配, 以便更好地学习关键目标的特征信息, 提高网络模型的抗干扰能力;

3) 引入平衡损失函数优化基线算法, 从而解决小目标检测时回归误差较大的问题, 增强模型检测器的稳定性, 平衡训练过程中对小目标和大目标的关注度, 以确保模型更有效地学习和适应小目标的特征, 从而改善整体检测性能。

1 无人机海面救援小目标检测算法设计

为提高海面救援工作时对小目标的检测性能, 本文提出了一种基于上下文特征信息、多尺度、注意力机制等知识的算法网络。其中, 空间注意力模块的关键在于确定算法模型对输入特征信息不同位置或区域的关注程度, 这一过程通常包括计算一个权重矩阵, 而 SAM 的自适应性使其能够灵活地适应不同任务、场景和数据集的要求^[23]。SAM 的功能可以通过不同的方式实现, 包括基于卷积神经网络、循环神经网络 (RNN, recurrent neural network)^[24]和自注意力机制^[25]等。SAM 在提升模型准确性、增强泛化能力、提高目标定位精度以及降低对无关信息敏感性等方面展现出显著的优势。

1.1 融合多尺度和上下文特征的 SSPNet 模型

为满足海面救援行动对小目标检测的精度要求, 本文融合多尺度和上下文特征, 结合尺度选择金字塔网络进行改进。

所提算法整体网络结构如图 1 所示。首先, 该模型使用 ResNet50 作为骨干网络, 相比于 ResNet101, 该网络结构对小目标进行特征提取的效果更好。然后, 在基线模型的基础上, 在特征金字塔模块中加入上下文增强和多尺度空洞卷积结构, 联系上下文特征信息, 强化模型分辨小目标对象和背景的能力, 允许算法模型在不同尺度上进行卷积操作, 更全面地感知图像的语境信息, 提升算法模型对小目标周围环境的理解, 从而增强小目标检测的精度和鲁棒性。为进一步优化模型的检测性能, 在区域建议网络模块加入空间注意力机制, 针对基线算法检

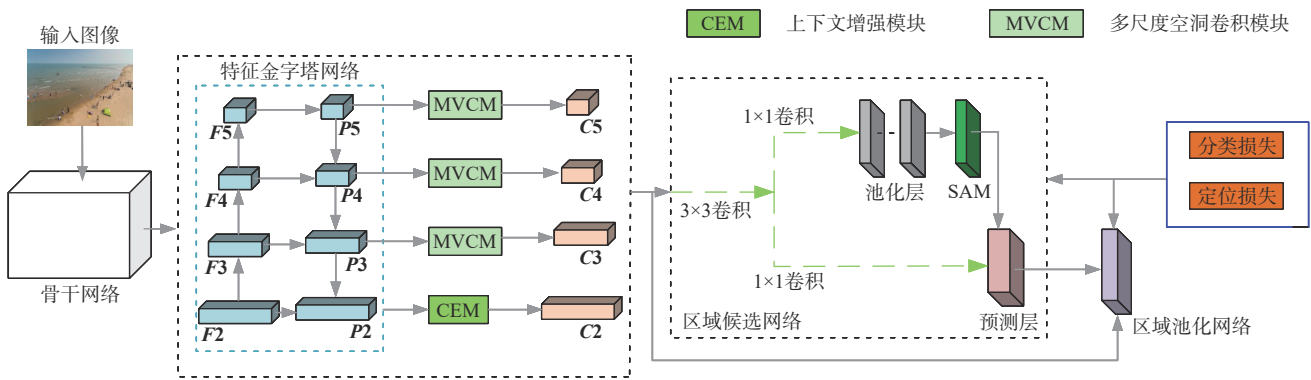


图1 所提算法整体网络结构

测小目标时注意力权重分配不均的问题，加强对关键特征的关注学习，减少背景干扰，提高小目标检测的精度。最后，引入平衡L1损失函数^[26]，用作整体网络的定位损失，使得算法模型在面对噪声或异常数据时，具有一定的鲁棒性和平滑性，从而使模型更易收敛且更快地学习到有效的特征，有助于提高模型的训练速度和检测效率。

1.2 上下文增强模块

进行小目标检测时，由于无人机拍摄的图像通常具有广阔的范围，此时检测任务的目标在网络模型训练的过程中，通常会丢失小目标部分特征信息，从而影响模型最终的检测性能。为解决这一问题，本节采取了针对特征信息最丰富的特征层进行数据增强的策略，通过利用上下文特征信息和多尺度特征提取的方法，捕捉不同尺度下各个特征层的关键信息，以减少小目标特征的丢失，并加强网络模型的学习能力，从而增强模型的检测效果。

上下文增强模块网络结构如图2所示。将该模块嵌入模型特征金字塔模块部分，当模型的骨干网

络提取的特征图输入特征金字塔模块时，经过一系列尺度选择、特征提取操作，根据初步处理之后的特征图上下文信息，在该部分网络中选择性增强某一层特征值，最终通过256个输出通道输送增强后的特征图到下一个模块，进行更进一步处理。

骨干网络模块输出图像经过尺度选择金字塔网络的初步处理之后，输出自顶而下的特征层图像，分别是P5、P4、P3、P2，P_glb是P5层应用全局平均池化而得到的全局上下文特征向量，输出只有单个像素大小的特征图。进一步，利用卷积核大小为3×3的空洞卷积，将输出的P2、P3、P4、P5特征层分别输入MACM模块，保证卷积前后输入输出图像的特征维度不变，再利用Concat特征融合方法，将多尺度卷积化之后的特征层拼接融合，输出与输入MACM模块特征图像尺度相同的特征图图像层。最后，将输出的特征图像层分别进行下采样操作，再经过卷积计算之后，输出4个尺度相同的特征图C2_lat、C3_lat、C4_lat和C5_lat。在进行下采样的操作过程中，P_glb在网络结构中

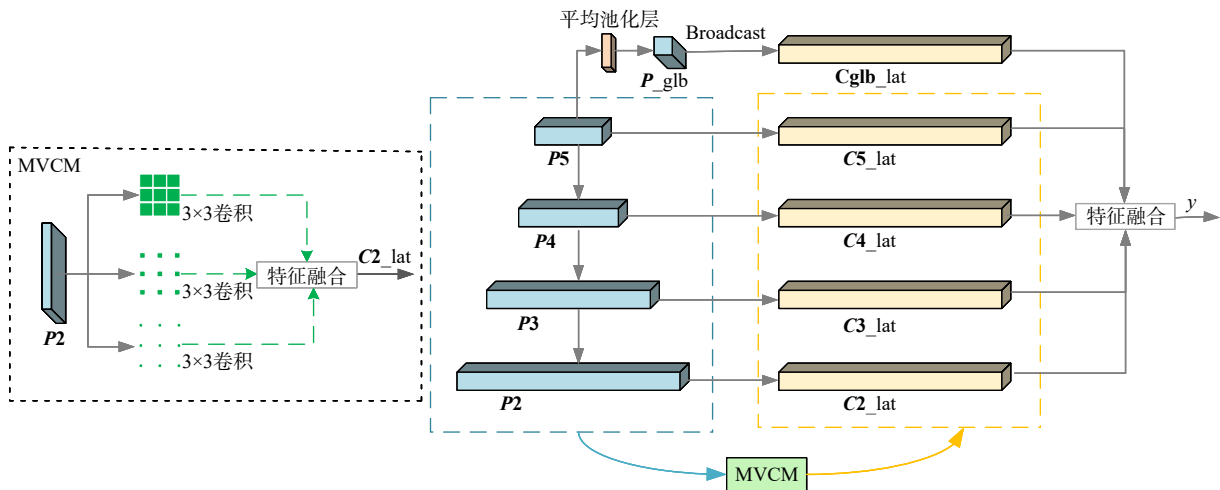


图2 上下文增强模块网络结构

通过 Broadcast 可以自适应匹配 $C5_lat$ 的特征尺度, 生成 $Cg1b_lat$ 特征图像, 将 $C2_lat$ 、 $C3_lat$ 、 $C4_lat$ 、 $C5_lat$ 和 $Cg1b_lat$ 这 5 个特征图像进行融合, 生成最终经过上下文信息增强的特征图 y , 与经过 MACM 模块输出的其他 3 层特征图像层 $C3$ 、 $C4$ 和 $C5$ 相结合, 使维度相同但尺度不同的 4 个特征图构成上下文金字塔结构。CEM 模块的输出结果如式(1)所示

$$y = \text{Concat}(C2_lat, C3_lat, C4_lat, C5_lat, Cg1b_lat) \quad (1)$$

通过利用特征图上下文信息, CEM 模块有效地增强了本文算法模型对特征图深层次语义信息的获取能力。MACM 模块的设计使得特征层的特征提取和感知能力得到改善, 在引入不同大小的空洞之后, 可以捕获不同尺度的上下文信息, 从而更全面地理解图像或特征图中的内容, 也可以在不丧失分辨率的情况下增加感受野, 有助于保留图像或特征图的详细信息。MACM 模块的计算过程如式(2)和式(3)所示。其中, P_i 为第 i 个特征层的特征图像, r 为空洞率, 分别可以取 1、3 和 5, Conv 为卷积操作, Concat 为特征融合操作, Ci_lat 为第 i 个特征层图像的卷积输出结果

$$P_i = \begin{cases} \text{Conv}_r = 1(x) \\ \text{Conv}_r = 3(x) \\ \text{Conv}_r = 5(x) \end{cases} \quad (2)$$

$$Ci_lat = \text{Concat}(\text{Conv}_1, \text{Conv}_3, \text{Conv}_5) \quad (3)$$

1.3 空间注意力模块

基于无人机图像的海面救援任务中, 原始图像含有大量特征信息, 其中包括目标物体的形状、颜色、纹理等, 但相比较广阔的背景特征信息而言, 检测任务的目标尺度小, 训练时易与背景的特征信息混淆, 因此, 本文采用 SAM 机制, 加强对小目

标重要特征的关注学习, 有效利用特征图的重要特征, 抑制无效的特征, 减少背景、噪声等特征信息的干扰, 从而提高网络模型的学习能力^[27]。SAM 嵌入网络结构之后, 根据接收到的特征图, 利用区域候选网络本身的特征信息, 过滤由于特征金字塔网络中特征层对不同尺度进行融合, 产生的大量冗余信息和冲突信息, 防止微小物体的特征信息直接被淹没在冲突信息中, 从而降低了多尺度表达能力^[28]。空间注意力模块网络结构如图 3 所示。

该模块首先在空间维度上分别进行自适应最大池化 (MaxPool, adaptive maximum pooling) 和自适应平均池化 (AvgPool, adaptive average pooling), 其次将池化后的两张特征图网络加入激活函数, 能够在优化和反向传播过程中更容易处理对应的特征信息, 有助于优化算法找到局部最小值, 最后进行归一化操作便于将两个不同的特征层图像拼接融合, 生成统一的特征图信息, 重新加权分配注意力, 获得空间权重, 对不同的特征层自适应地赋予不同的权重, 并获得最终的输出特征图像。

$F1$ 是来自特征金字塔模块的输出图像特征值信息, $F2$ 是区域建议网络模块本身含有的特征信息。对于小目标, 在训练过程中由于注意力感受野的分布不同, 得到的权重也有所区别, 为了使模型将更多的注意力用来关注学习有效的特征信息, 增强模型的性能, 本文在区域建议网络结构中, 嵌入了一个改进的 SAM 机制, 利用 RPN 特征信息 $F2$ 细化特征映射的特征分布, 稳定该结构的训练模型, 减少信息的丢失。其中, SAM 卷积层的卷积核大小为 2×2 , 使得模型可以获得更大的感受野, 加强前景特征和抑制背景特征从而细化特征分布, 此外, 增加了注意力卷积层, 从而提高网络表示能力

$$S(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

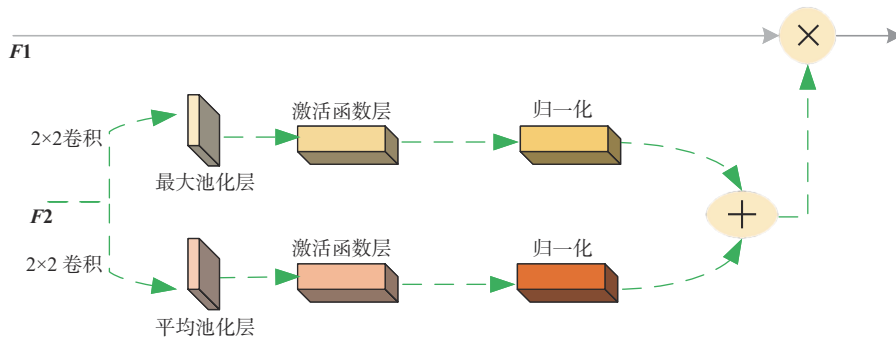


图 3 空间注意力模块网络结构

SAM选用Sigmoid函数作为激活函数,其函数计算如式(4)所示,其中, x 是输入图像特征像素值,输入映射区间为 $[0, 1]$,便于明确指定算法模型关注每个位置或区域的程度,从而使算法模型在优化和反向传播过程中可以更精确地处理输入信息。

1.4 损失函数

海面救援任务中,针对小目标的检测过程极易被其他因素干扰,导致模型无法精确定位小目标的位置^[29],以及在实际应用的过程中,小目标匹配性能通常不如中、大目标性能优越,造成检测时有很多正样本被遗漏^[30]等问题,本文引用了平衡L1损失函数作为定位损失函数,从而改善整体网络模型的检测性能。

在目标检测模型中,总损失函数Loss一般由定位损失函数 L_{loc} 和分类损失函数 L_{cls} 组成,如式(5)所示

$$\text{Loss} = L_{loc} + L_{cls} \quad (5)$$

基线算法使用交叉熵损失函数^[31](CE Loss, cross entropy loss)作为分类损失,Smooth L1损失函数^[8]作为定位损失,计算式分别如式(6)和式(7)所示

$$\text{CELoss} = -\sum_{cls} p(x) \log q(x) \quad (6)$$

$$S_{L1}(h) = \begin{cases} 0.5h^2, & |h| < 1 \\ |h| - 0.5, & |h| \geq 1 \end{cases} \quad (7)$$

其中, $p(x)$ 和 $q(x)$ 分别代表真实值和RPN网络最终输出的预测值, h 是预测值和真实值的差值。

Smooth L1定位损失函数对离群值具有较好的鲁棒性,能够提高模型的检测性能,但在目标尺度相差较大的情况下,不同的样本所需的权重分布不同,因此,本文引入平衡L1定位损失函数,定义如式(8)所示, $L_b(x)$ 为定位函数的具体值大小。平衡L1损失函数兼具Smooth L1损失函数的优点,通过平衡样本和任务,有助于模型在目标定位任务中更好地处理重要特征信息,提高模型训练过程的鲁棒性,从而提高小目标检测精度

$$L_b(x) = \begin{cases} \frac{m}{n}(n|x|+1)\ln(n|x|+1) - m|x|, & |x| < 1 \\ t|x| + C, & |x| \geq 1 \end{cases} \quad (8)$$

其中, m 是控制损失函数的梯度大小, m 值越小,梯度增加越大;当像素值 $x = 1$ 时, n 可以确保式(8)的两个计算式有相同的值,使得函数图像能够连续; t 为整体提升放大率来调整回归误差上界的控

制量,帮助目标函数更好地平衡相关任务。控制 m 和 n 两个因素相互加强,可以达到更平衡的训练,其中, m 、 n 和 t 这3个参数受到式(9)的约束,本文实验中,设置 $m = 0.5$, $t = 1.5$

$$m \ln(n + 1) = t \quad (9)$$

2 实验分析

2.1 实验数据集

本文实验中,采用高质量标注的小目标数据集Tiny-Person,该数据集以海域为背景,其图像场景大多为无人机高空拍摄而来,是极具代表性的海域人物检测数据集,Tiny-Person数据集示意图如图4所示,共包含1 610张图片,其中794张图片用于训练,816张图片用于测试,并包含了72 651个人类目标注释^[32]。图片中人物目标最小分辨率为 2×3 像素和 3×2 像素。



图4 Tiny-Person数据集示意图

2.2 评价指标

为了评估模型的有效性,本文使用精确率(P , precision)、召回率(R , recall)和平均精度(AP, average precision)作为评价指标。AP值为精确率-召回率(PR, precision-recall)曲线下的面积,分类器越好,AP值越高,代表该模型的检测性能越好,式(10)为AP值的计算式

$$I_{AP} = \int P dR \quad (10)$$

其中, P 和 R 的计算如式(11)和式(12)所示,其中,TP(true positive)为检测结果中正确的目标数量,FP(false positive)为检测结果中错误的目标数量,FN(false negative)为正确的目标中未被检测出的数量

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

计算 Tiny-Person 数据集的 AP 值时, 考虑目标尺度过小的情况, 本文选择将数据集划分为3个主要尺度区间: tiny [2, 20]、small [20, 32]、all [2, inf], 其中, inf 指数值无穷大。进一步, 将 tiny 尺度细分为3个子尺度区间: tiny1 [2, 8]、tiny2 [8, 12]、tiny3 [12, 20]。这种尺度划分的目的是更精细地考虑不同大小目标的性质, 以更准确地评估模型性能, 特别是当数据集中存在极小目标时, 也可以在评估过程中提供更多关于模型性能的信息。

海面人物目标检测中, 通常使用 IoU 作为评价指标之一。IoU 通过计算预测框与真实目标框的交集面积与它们的并集面积之比来衡量检测精度, 常见的 IoU 阈值为 0.25、0.50 和 0.75, 代表了不同的重叠程度。一般情况下, 较高的 IoU 值表示预测出的目标与真实目标的交集较大, 表明模型的预测更加贴合真实目标, 检测精度相对较高。然而, 海面人物目标检测中, 由于小目标只占据很小的像素点, 如果使用较大的 IoU 阈值, 会导致检测精度降低。为了更好地衡量小目标检测算法的效果, 选择 IoU 为 0.50 时对应的 AP 值作为评价指标, 这是因为 0.50 被广泛认为是 IoU 阈值的一个折中值, 能够在考虑检测精度的同时, 对小目标有更好的鲁棒性。

2.3 实验设置

本文实验使用 Linux 操作系统, 单个 NVIDIA GeForce RTX 3090 显卡, 显存大小为 24 GB, 实验基于 PyTorch1.11 框架对算法模型进行搭建、训练和测试, 利用 Python3.8、CUDA 11.3 环境进行仿真训练, 实验数据集为 Tiny-Person。整个模型训练周期为 12 轮, 初始学习率设置为 0.002, 在训练过程中学习率衰减使用线性策略, 训练批次设置为 2, 在训练到第 8 轮和第 11 轮时, 分别将学习率减少为原来的 0.1 倍, 采用随机梯度下降算法优化网络的权重。实验配置信息见表 1, 除必要的改进外, 本实验中所有模型的超参数都设置为默认。

2.4 实验结果分析

为了验证改进模型的有效性, 本文分别进行了模块消融实验、算法模型精度对比实验和改进前后算法的检测效果对比, 以此评估改进模型对检测结果的影响。

表 1 实验配置信息

参数名称	参数设置
运行系统	Linux
GPU	NVIDIA GeForce RTX 3090
显存	24 GB
Python	3.8
CUDA	11.3
训练周期	12 轮
初始学习率	0.002
学习率衰减策略	线性策略

1) 上下文增强模块

为测试上下文增强模块对特征层的影响, 本实验针对不同特征层进行增强训练, 上下文增强模块实验对比见表 2。表 2 中 AP50_all 代表当 IoU 阈值为 0.5 时, 所有尺度样本同时进行评估的平均精度值; AP50_tiny、AP25_tiny 和 AP75_tiny 分别代表了当 IoU 阈值为 0.50、0.25 和 0.75 时, 尺度大小为 tiny 对应的样本进行评估的平均精度值。通过逐层增强和所有特征层增强分别进行训练, 其中, 增强 P3、P4 特征层时对网络性能的提升效果较小, 增强 P2 特征层时对网络模型的性能贡献最佳, AP50_all 提升了 0.71%, AP50_tiny 提升了 0.58%, 其原因是 P2 特征层保留了原始图像中更为丰富和细致的信息, 具有较高的分辨率, 可以使得模型更好地理解图像, 提高检测的准确性和鲁棒性。增强所有特征层后检测精度在各项指标上都有所降低, 相较于增强单个特征层, 此时特征信息冗余, 明显抑制了算法网络的检测效果。不同 IoU 阈值情况下的增强效果差异可能与训练样本中小目标的边界框标注有关, 在低 IoU 阈值 (如 0.25) 的情况下, 一些稍微与目标轮廓有重叠的边界框也可能被模型认为是正样本, 从而导致模型在增强 P3 特征层时, AP25_tiny 的检测效果较好。本文实验中, 综合考虑小目标的鲁棒性和图像检测效果, 通常选用 IOU 为 0.50 时的实验结果作为参考, 选用增强 P2 特征层的网络作为该模块的最终结构。

2) 损失函数

为了提高小目标的检测性能, 本文实验使用平衡 L1 损失函数替换基线算法的损失函数。在数据集 Tiny-Person 上训练过程的损失下降曲线如图 5 所示, 一轮训练有 5 100 次迭代, 一共训练 12 轮。由于平衡 L1 损失函数引入动态平衡参数, 可以根

表2 上下文增强模块实验对比

方法	AP50_all	AP50_tiny	AP25_tiny	AP75_tiny
基线算法	60.11	56.07	76.10	7.82
增强P2层	60.82	56.65	76.74	7.98
增强P3层	60.42	56.56	76.96	7.78
增强P4层	60.52	55.83	76.11	7.77
增强所有特征层	58.48	55.32	75.49	6.89

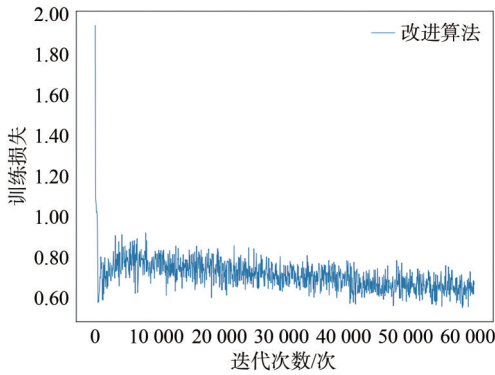


图5 在数据集Tiny-Person上训练过程的损失下降曲线

据目标的大小和难易程度来调整损失函数的权重，使得模型更好地适应不同尺寸的目标，具有抗噪能力强、平衡精度、定位准确及减少模型训练过程中不稳定性的优势。从图5的曲线变化可知，优化后的损失函数值总体下降趋势平缓，可以达到理想效果。

3) 消融实验

通过逐步应用各个模块来分析网络模型中每个模块的效果。消融实验结果见表3，“√”表示模型含有该模块，从表3可知，各模块对网络模型的检测能力有一定的影响。首先，加入空间注意力改进模块，由于检测目标特征像素值相对于原始图像而言过小，添加该模块后，网络模型在训练过程中可以进一步分配注意力权重，能够有效抑制冗余干扰，从而使得AP50_tiny的检测精度提升了0.24%；其次，加入上下文增强模块后，AP50_tiny的检测

精度为56.65%，与基线算法相比提高了0.58%，主要针对包含特征信息较多的特征层进行增强，可以感受到更多的特征信息，说明增大模型感受野后可有效获取到检测目标的全局上下文信息，减少小目标特征的丢失，从而增强模型的学习能力；最后，由于平衡损失函数相较于原函数具有良好的鲁棒性和平滑性，适应性更强，因此添加平衡L1损失函数，使得AP50_tiny的检测精度与基线算法相比提高了0.77%。两两模块融合之后，虽然抑制了部分指标的检测效果，但模型的总体检测精度是提升的，并且当3个模块均存在时，能够缓解抑制情况。相对于基线算法，本文方法的AP50_all、AP50_tiny、AP50_tiny1、AP50_tiny2、AP50_tiny3和AP50_small分别改善了1.20%、2.06%、2.07%、2.84%、1.70%和1.38%，由此可知，本文算法在[2, 20]的像素尺度区间范围内精度提升较高，改善了小目标的检测性能。

4) 对比实验

本文比较了近年来一些基于深度学习网络的目标检测算法和改进前后的算法，展现了本文所提检测模型在海域人物小目标检测中的良好性能。首先在COCO数据集上预训练骨干网络，利用其参数初始化网络模型，之后基于本文设置的实验参数，在Tiny-Person数据集上，将基于锚框的单阶段算法RetinaNet、两阶段算法Faster R-CNN FPN，与改进后算法进行对比，在Tiny-Person数据集上与其他算法的比较见表4。一些在COCO数据集上表现极佳的算法，如Libra R-CNN^[24]、SSD^[5]算法等，并没有在小目标数据集上取得良好的成绩，原因可能是Tiny-Person数据集中的目标尺寸过小，在训练过程中丢失大量特征，导致网络模型在检测时无法定位小目标的位置信息。而相对于基线算法，由于本文网络模型对小目标特征信息进行关键性特征

表3 消融实验结果

基准算法	平衡L1损失函数	SAM	CEM	AP50_all	AP50_tiny	AP50_tiny1	AP50_tiny2	AP50_tiny3	AP50_small	AP25_tiny	AP75_tiny
√				60.11	56.07	43.33	59.02	64.68	71.15	76.10	7.82
√	√			60.89	56.84	44.25	59.73	65.21	70.43	75.84	8.00
√		√		60.91	56.31	43.65	57.94	65.36	72.31	76.19	8.20
√			√	60.82	56.65	44.14	59.35	64.81	71.60	76.74	7.98
√		√	√	61.21	57.46	44.50	59.92	66.23	72.45	77.34	7.89
√	√	√		61.06	56.41	43.96	57.93	65.29	71.92	75.31	8.40
√	√		√	60.96	56.67	43.77	59.23	65.36	72.03	76.08	8.03
√	√	√	√	61.31	58.13	45.40	61.86	66.38	72.53	76.51	9.40

表4 在 Tiny-Person 数据集上与其他算法的比较

算法	AP50_tiny	AP50_tiny1	AP50_tiny2	AP50_tiny3	AP50_small	AP50_all	AP25_tiny	AP75_tiny
SSD ^[5]	34.00	13.54	35.16	48.78	57.14	—	61.21	2.52
RetinaNet ^[33]	52.56	33.90	58.00	63.72	65.69	—	73.09	6.64
YOLOv5s ^[34]	49.61	32.21	52.11	60.95	64.23	—	71.51	6.63
Libra R-CNN ^[24]	44.68	27.08	49.27	55.21	62.65	—	—	—
SSPNet ^[35]	50.86	34.85	54.80	60.35	67.28	55.99	—	—
Faster R-CNN FPN ^[9]	47.35	30.25	51.58	58.95	63.18	—	68.43	5.83
Faster R-CNN FPN+SM ^[3]	51.33	33.91	55.16	62.58	66.96	—	71.55	6.46
Faster R-CNN FPN*	51.98	39.62	55.18	62.03	67.70	55.89	70.53	7.42
RetinaNet*	54.00	42.64	58.17	61.30	69.90	58.21	77.12	6.13
SSPNet*	56.07	43.33	59.02	64.68	71.15	60.11	76.10	7.82
改进算法	58.13	45.40	61.86	66.38	72.53	61.31	76.51	9.40

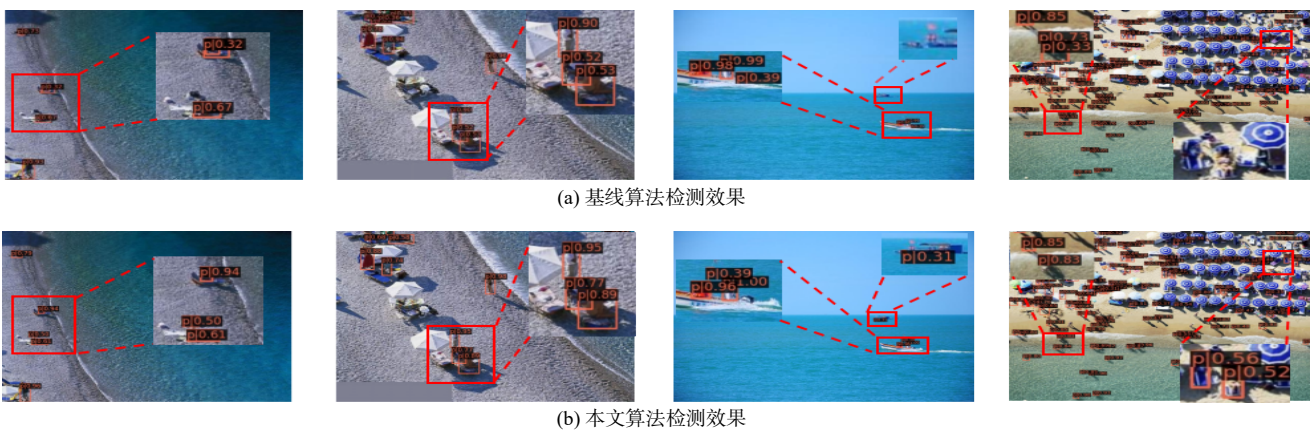


图6 算法结构改进前后检测效果对比

增强，使得 AP50_tiny 的检测精度提升了 2.06%。作为检测任务的评价标准，AP 值越高，检测器性能越好，“*”表示在相同的实验设置下，本文实验重新进行公平比较。

为了直观地展示本文算法对海面小目标的检测效果，我们使用 Tiny-Person 数据集中具有代表性的图片验证，由于原图过大而截取部分区域进行检测，算法结构改进前后检测效果对比如图 6 所示。从图 6 可以看出本文算法可以有效地提高小目标的检测效果，其中部分目标的检测精度有所提升，而对于错检、漏检情况也得到了改善。

3 结束语

为了改善小目标在图像区域中所占区域小、分辨率低，容易被错检、漏检等情况，提高算法网络对无人机图像小目标的检测性能，使得海面救援等行动能够更加快速便捷，本文基于基线算法进行改进，嵌入注意力模块和上下文增强模块，通过输入图像传输的上下文特征信息，加强对重要特征信息

的学习，改善检测精度，使得改进后模型相比于基线模型的 AP50_tiny、AP50_tiny1、AP50_tiny2 和 AP50_tiny3 分别提高了 2.06%、2.07%、2.84% 和 1.70%。使用 Tiny-Person 数据集对本文所提算法进行实验验证，实验结果表明，改进后算法的检测效果较为优异，验证了该算法的有效性。后续的研究工作中，将进一步思考如何将算法网络与更多先进的方法相结合，增加网络模型对不同尺寸目标的适应性，提高检测性能，以更好地对海域等目标任务进行检测。

参考文献:

- [1] CHEN C Y, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Proceedings of the Asian Conference on Computer Vision. Cham: Springer, 2017: 214-230.
- [2] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2014: 740-755.
- [3] YU X H, GONG Y Q, JIANG N, et al. Scale match for tiny person detection[C]//Proceedings of the 2020 IEEE Winter Conference

- on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2020: 1246-1254.
- [4] 魏泽发, 崔华. 基于SqueezeNet卷积神经网络的车辆检测[J]. 物联网学报, 2020, 4(3): 120-125.
WEI Z F, CUI H. Vehicle detection based on SqueezeNet convolutional neural network[J]. Chinese Journal on Internet of Things, 2020, 4(3): 120-125.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot Multi-Box detector[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 779-788.
- [7] 刘洋, 战荫伟. 基于深度学习的小目标检测算法综述[J]. 计算机工程与应用, 2021, 57(2): 37-48.
LIU Y, ZHAN Y W. Survey of small object detection algorithms based on deep learning[J]. Computer Engineering and Applications, 2021, 57(2): 37-48.
- [8] GIRSHICK R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 1440-1448.
- [9] 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.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [11] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2980-2988.
- [12] 闫海庆, 雷远华, 王静, 等. 基于改进Libra-RCNN的输电线路绝缘子识别[J]. 湖南电力, 2022, 42(2): 44-49.
LYU H Q, LEI Y H, WANG J, et al. Transmission line insulator identification based on improved libra-RCNN[J]. Hunan Electric Power, 2022, 42(2): 44-49.
- [13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [14] 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.
- [15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.
- [16] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.
- [17] SANG H W, WANG Q H, ZHAO Y. Multi-scale context attention network for stereo matching[J]. IEEE Access, 2019, 7: 15152-15161.
- [18] WANG X, LV R R, ZHAO Y, et al. Multi-scale context aggregation network with attention-guided for crowd counting[C]//Proceedings of the 2020 15th IEEE International Conference on Signal Processing (ICSP). Piscataway: IEEE Press, 2020: 240-245.
- [19] 吕晓华, 魏铭辰, 刘立波. 基于位置可学习视觉中心机制的零售商品检测方法[J]. 物联网学报, 2023, 7(4): 142-152.
LYU X H, WEI M C, LIU L B. Retail commodity detection method based on location learnable visual center mechanism[J]. Chinese Journal on Internet of Things, 2023, 7(4): 142-152.
- [20] HONG M B, LI S W, YANG Y C, et al. SSPNet: scale selection pyramid network for tiny person detection from UAV images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8018505.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [22] 林椿珉, 曾烈康, 陈旭. 边缘智能驱动的高能效无人机自主导航算法研究[J]. 物联网学报, 2021, 5(2): 87-96.
LIN C M, ZENG L K, CHEN X. Research on power efficient autonomous UAV navigation algorithm: an edge intelligence driven approach[J]. Chinese Journal on Internet of Things, 2021, 5(2): 87-96.
- [23] 王正文, 宋慧慧, 樊佳庆, 等. 基于语义引导特征聚合的显著性目标检测网络[J]. 自动化学报, 2023, 49(11): 2386-2395.
WANG Z W, SONG H H, FAN J Q, et al. Semantic guided feature aggregation network for salient object detection[J]. Acta Automatica Sinica, 2023, 49(11): 2386-2395.
- [24] 姚红革, 张玮, 杨浩琪, 等. 深度强化学习联合回归目标定位[J]. 自动化学报, 2023, 49(5): 1089-1098.
YAO H G, ZHANG W, YANG H Q, et al. Union regression object localization based on deep reinforcement learning[J]. Acta Automatica Sinica, 2023, 49(5): 1089-1098.
- [25] 杜鹏, 宋永红, 张鑫瑶. 基于自注意力模态融合网络的跨模态行人再识别方法研究[J]. 自动化学报, 2022, 48(6): 1457-1468.
DU P, SONG Y H, ZHANG X Y. Self-attention cross-modality fusion network for cross-modality person re-identification[J]. Acta Automatica Sinica, 2022, 48(6): 1457-1468.
- [26] PANG J M, CHEN K, SHI J P, et al. Libra R-CNN: towards balanced learning for object detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 821-830.
- [27] 潘翔, 陈前斌, 黄昂, 等. 基于改进YOLOX的无人机航拍图像小目标检测算法[J]. 南京邮电大学学报(自然科学版), 2024, 44(1): 90-100
PAN X, CHEN Q B, HUANG A, et al. A small target detection

algorithm of UAV aerial photography images based on improved YOLOX[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2024, 44(1): 90-100.

- [28] 陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. 光电工程, 2022, 49(3): 69-81.
CHEN X, PENG D L, GU Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. Opto-Electronic Engineering, 2022, 49(3): 69-81.
- [29] 宁欣, 田伟娟, 于丽娜, 等. 面向小目标和遮挡目标检测的脑启发CIRA-DETR全推理方法[J]. 计算机学报, 2022, 45(10): 2080-2092.
NING X, TIAN W J, YU L N, et al. Brain-inspired CIRA-DETR full inference model for small and occluded object detection[J]. Chinese Journal of Computers, 2022, 45(10): 2080-2092.
- [30] 廖龙杰, 吕文涛, 叶冬, 等. 基于深度学习的小目标检测算法研究进展[J]. 浙江理工大学学报(自然科学), 2023, 49(3): 331-343.
LIAO L J, LYU W T, YE D, et al. Research progress of small target detection based on deep learning[J]. Journal of Zhejiang Sci-Tech University (Natural Sciences), 2023, 49(3): 331-343.
- [31] CAO J, SU Z, YU L Y, et al. Softmax cross entropy loss with unbiased decision boundary for image classification[C]//Proceedings of the 2018 Chinese Automation Congress (CAC). Xi'an, China, 2018: 2028-2032.
- [32] 邵香迎, 郭颖, 王友伟. AF-RetinaNet: 一种基于自适应融合与特征细化的微小行人检测算法[J]. 控制与决策, 2024, 39(3): 939-946.
SHAO X Y, GUO Y, WANG Y W. AF-RetinaNet: a tiny person detection algorithm based on adaptive fusion and feature refinement[J]. Control and Decision, 2024, 39(3): 939-946.
- [33] GONG Y Q, YU X H, DING Y, et al. Effective fusion factor in FPN for tiny object detection[C]//Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2021: 1159-1167.
- [34] 张源川. 深度卷积神经网络下的小目标检测方法[D]. 重庆: 重庆邮电大学, 2022.
ZHANG Y C. Small object detection method based on deep convolutional neural network[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2022.
- [35] GUO G Q, CHEN P F, YU X H, et al. Save the tiny, save the all: hierarchical activation network for tiny object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(1): 221-234.

[作者简介]



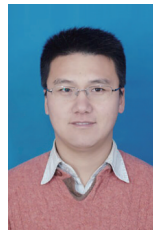
刘园 (2000–), 女, 南京邮电大学自动化学院、人工智能学院硕士生, 主要研究方向为无人系统、目标检测、深度学习。



赵静 (1983–), 女, 博士, 南京邮电大学自动化学院、人工智能学院副教授, 主要研究方向为无人系统、容错控制、智能控制与应用。



蒋国平 (1966–), 男, 博士, 南京邮电大学自动化学院、人工智能学院教授, 主要研究方向为网络建模、传播与控制、混沌控制、混沌信息处理与通信系统设计。



徐丰羽 (1979–), 男, 博士, 南京邮电大学自动化学院、人工智能学院教授, 主要研究方向为机器人与自动化、机电一体化技术、智能制造装备与控制技术。



陆宁云 (1978–), 女, 博士, 南京航空航天大学航空航天结构力学及控制全国重点实验室教授, 主要研究方向为无人系统、过程建模与监测、故障诊断。