DOI: 10.13382/j.jemi.B2407982

融合双重观察与注意力机制的灰度图像检测算法*

朱 硕[1] 张绪康[2] 宾 杰[1] 汪宗洋[3] 江 蕊[1]

(1. 无锡学院江苏省通感融合光子器件及系统集成工程研究中心 无锡 214105;2. 南京信息工程大学电子与信息工程学院 南京 210044;3. 无锡汐沅科技有限公司 无锡 214000)

摘 要:灰度图像由于其单通道构成的限制,导致图像中目标对比度低、特征信息模糊以及缺少颜色信息,因此检测精度低、且检测难度较大。为提升灰度图像检测的准确率,降低误检和漏检率,提出一种融合双重观察与注意力机制的目标检测算法 SAC-YOLO。首先,在主干网络中引入变换空洞卷积,将标准卷积层转换为空洞卷积层,并结合全局上下文模块,提升模型在处理不同尺度和复杂度信息的准确性;其次,特征融合部分采用高效多尺度注意力机制,通过编码全局信息来重新校准各通道权重,跨纬度交互捕捉灰度图像中的像素级关系;最后,添加超分辨率重构检测头,内置感受野注意力模块和卷积模块,关注感受野内空间信息,为大尺寸卷积核提供有效注意力权重,使得模型能够更加精确地适应和表达灰度图像中的小目标信息的特征。在 NEU-DET 数据集中进行对比实验,改进后的 YOLOv8 算法对于灰度图像信息的识别精度达到 79.3%,相较于 YOLOv8 原始网络提升了 3.1%,由可视化实验可以看出,误检漏检问题得到改善。以上实验结果表明,SAC-YOLO 检测效果良好,能够实现在灰度图像场景下的高质量检测。

关键词: 目标检测;YOLOv8;灰度图像检测;感受野注意力;空洞卷积

中图分类号: TP391.4; TN919 **文献标识码**: A **国家标准学科分类代码**: 520.2060

Gray image detection algorithm integrating double observation and attention mechanism

Zhu Shuo[1] Zhang Xukang[2] Bin Jie[1] Wang Zongyang[3] Jiang Rui[1]

(1. Jiangsu Province Engineering Research Center of Photonic Devices and System Integration for Communication Sensing Convergence, Wuxi University, Wuxi 214105, China; 2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 3. Wuxi Xiyuan Technology Co., Ltd., Wuxi 214000, China)

Abstract: Owing to the constraint of the single-channel structure of gray images, the target contrast within the image is low, the feature information is indistinct, and the color information is lacking. Hence, the detection accuracy is low and the detection process is arduous. To enhance the accuracy of gray image detection and reduce the rates of false detection and missed detection, an object detection algorithm, SAC-YOLO, combining dual observation and attention mechanism was proposed. Firstly, transform atrous convolution was integrated into the backbone network to convert the standard convolution layer into an atrous convolution layer, and the global context module was combined to enhance the model's accuracy in processing information of different scales and complexities. Secondly, the feature fusion part employs an efficient multi-scale attention mechanism to recalibrate the weight of each channel by encoding global information and interactively captures the pixel-level relationship in gray images across latitudes. Finally, a super-resolution reconstruction detection head was added, and a receptive field attention module and a convolution module were constructed to focus on the spatial information within the receptive field and provide effective attention weights for the large-size convolution kernel, enabling the model to adapt and represent the characteristics of small target information in gray images more precisely. The comparison experiment in the NEU-DET dataset reveals that the recognition accuracy of the improved YOLOv8 algorithm for gray image information attains 79.3%, which is 3.1% higher than that of the original YOLOv8 network. It can be observed from the visualization experiment that the issue of

收稿日期: 2024-11-23 Received Date: 2024-11-23

*基金项目:国家自然科学基金资助项目(52075520)、江苏双创博士基金(JJSSCBS20210871)项目资助

false detection and missed detection has been alleviated. The above experimental results indicate that SAC-YOLO has an excellent detection effect and can achieve high-quality detection in grayscale image scenarios.

Keywords: object detection; YOLOv8; gray image detection; receptive field attention; void convolution

0　引　言

目标检测是计算机视觉领域的重要技术之一,其主要目的是模拟人类的视觉和认知能力,探索统一的框架来实现对不同类型目标的识别和定位。这项技术在无人驾驶、智能家居、智能电网和智慧医疗等多个领域发挥着关键作用[1]。目前,绝大多数目标检测技术以彩色图像为数据集图像,检测效果良好,但容易受到天气、光照强度等环境因素的影响。相比之下,灰度图像以其单通道灰度信息的特性,简化目标特征,降低环境噪声干扰,使得模型在进行目标检测任务中能够大大降低烟雾、雾霾、雨雪等恶劣环境造成的影响。但灰度图像由于缺少颜色信息的描述,图像中会出现纹理细节较少,对比度和信噪比较低以及模糊成像等问题,使得现阶段目标检测算法难以提取到图像的深层语义特征,检测效果较差。因此,研究并设计高效准确的灰度图像目标检测算法具有重要意义。

灰度图像检测技术的发展主要分为基于人工特征的检测和基于深度学习的检测[2]两个阶段。基于人工特征的检测依靠传统的图像处理技术和算法开发人员的专业知识来手动设计特征。通过设计特征和使用支持向量机对于分类,识别各种类型的对象,但是手动特征设计复杂,并且在性能和适应性方面具有局限性。随着深度学习技术的快速发展,图像处理和计算机视觉领域取得了巨大的进步,目前基于深度学习的目标检测算法分为双阶段算法和单阶段算法。双阶段目标检测算法,如区域卷积神经网络(region-based convolutional neural network, R-CNN)[3]、快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)[4]和掩码区域卷积神经网络(mask region-based convolutional neural networks, Mask R-CNN)[5],首先生成一组可能包含待检测目标的候选框,然后通过特征提取来确定这些候选目标的位置和类别。这类算法在精确度上表现出色,但由于其复杂的算法结构,训练速度较慢且资源消耗较大。相比之下,单阶段目标检测算法,如单步骤多框检测器(single shot multibox detector, SSD)[6]、YOLO 系列(you only look once, YOLO)[7]和实时 Transformer 检测器(real-time detection transformer, RT-DETR)[8],直接利用卷积神经网络从图像中提取特征,并同时预测目标的位置和类别。由于省去了候选框生成的步骤,单阶段算法在检测速度上更为高效,且网络结构相对简单。基于深度学习的检测技术通过训练神经网络来提取目标特征,虽然其计算量相对较大,但随着计算机性能的提升和神经网络的不断优化,该检测方法逐渐成为灰度图像检测领域的研究重点。其中,Ghose 等[9]以 Faster R-CNN 作为基础网络,引入 PiCA-Net 作为检测网络的注意力机制,专注于像素级特征信息,辅助以 R3-Net,重建和整合由 PiCA-Net 提取的大量信息,检测性能显著提升,但引入的网络并未集成到 Faster R-CNN 中,训练过程较为复杂。Liu 等[10]利用坐标注意力和特征融合技术构建了名为 DG-Light-NLDF 的灰度图像目标检测模型,该模型通过扩张线性瓶颈结构代替卷积结构提取目标纹理和语义特征,避免误检,并利用简化全局模块进一步强调目标的位置特征,抑制背景干扰。Li 等[11]提出了以 YOLOv5 为内核的检测算法,在特征提取阶段扩展和迭代跨级部分连接结构,最大化浅层特征的使用,增加多尺度检测头,提升目标检测准确性,但是这种方法的高计算需求使其不适合部署在边缘设备上。Hu 等[12]提出了一种基于 YOLOv7 的检测方法,将红外和可见光的双通道特征提取与注意力模块相结合,以提高检测精度。然而,对于细节特征较少的图像,网络的检测能力仍有不足。Zhou 等[13]提出 YOLO-CIR 模型,它通过红外图像增强算法提升图像质量,然后使用 ConvNext 提取特征,并集成分裂注意力模块以优化特征融合。虽然这种方法提高了准确性,但它需要对红外图像进行预处理,增加了检测任务的复杂性。

综上,为权衡检测精度以及运行效率,研究选取 YOLOv8 作为基础网络,结合当前灰度图像检测的不足,提出一种融合双重观察与注意力机制的目标检测算法 SAC-YOLO(self-adaptation and atrous convolution-based YOLO)。

1) 构建变换空洞卷积,用空洞卷积层代替标准卷积层,引入双重观察机制,使得每个特征信息结合两种空洞率进行卷积,再由两个全局上下文模块关注图像整体内容,共享信息间权重,使得网络更全面和灵活的理解特征信息。

2) 在特征融合部分采用高效多尺度注意力模块,重新组织通道维度和批次维度,并进行信息的跨维度交互,提升网络对于灰度图像信息的细节捕捉。

3) 引入超分辨率重构检测头,其感受野注意力模块使网络重点关注感受野的空间特征,大尺度卷积核辅助注意力模块更加有效地处理重要空间特征。

1 YOLOv8 目标检测算法原理

YOLO 系列神经网络开放用于目标检测和图像分割任务，基于先前 YOLO 版本的基础，YOLOv8 拥有很好的可拓展性，不仅支持之前版本的切换运行，还可以调整参数生成 5 种不同比例的模型 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x，便于在各种应用场景的硬件设备上灵活部署[14]。YOLOv8 网络结构主要由主干网络(Backbone)，颈部网络(Neck)和预测头(Head)3 个部分构成。

YOLOv8 通过训练数据集自动学习锚点框，沿用 YOLOv5 中的 mosaic 数据增强，随机选取 4 张图像拼接其中的部分区域，改善目标分布不均匀的情况[15]。Backbone 主干网络主要由跨阶段部分模块(cross stage partial network，C2f)、标准卷积(convolutional layer，Conv)和快速空间金字塔池化(spatial pyramid pooling fast，SPPF)模块组成。C2f 模块借鉴了 YOLOv7 中的增强局部特征提取网络(enhanced local feature extraction network，ELEN)模块，通过多分支跨层链接高效聚合网络结构，丰富了模型的梯度流，从而实现更好的参数利用率并增加网络深度[16]。Conv 模块则包含 2 D 卷积块，用于对输入的特征信息进行下采样，同时进行归一化和激活处理。最终，Backbone 的下采样部分通过 SPPF 模块完成。SPPF 模块将池化核大小分别为 13×13、9×9、5×5 和 1×1 的最大池化层串行结合，并引入并行结构，以分别处理不同像素大小的特征图，完成特征融合。

YOLOv8 的 Neck 部分采用了 PAFPN 结构，结合了特征图金字塔网络(feature pyramid networks，FPN)和路径聚合网络(path aggregation network，PAN)。特征金字塔结构通过上采样将深层特征图中的强语义信息传递到浅层，从而实现多尺度特征的融合；而路径聚合网络则通过下采样将浅层的位置信息传递到深层，进一步提升多尺度特征图的表示能力和检测精度。

相较于 YOLOv7，YOLOv8 采用解耦头结构，将分类和回归任务分支处理，并以 DFL Loss 和 CIoU Loss 作为损失函数，在模型训练和推理中获取更高性能的收益，为了提高训练的准确率，YOLOv8 还引入了 YOLOX 中的最后 10 epoch 关闭 Mosiac 增强的操作[17]。

2 本文方法

2.1 网络整体结构

灰度图像中目标特征信息少，细节纹理不明显等问题，导致目标检测难度较大，检测效果不佳。研究以 YOLOv8 作为基础网络，提出基于双重观察与注意力机制的目标检测算法 SAC-YOLO，整体结构如图 1 所示。

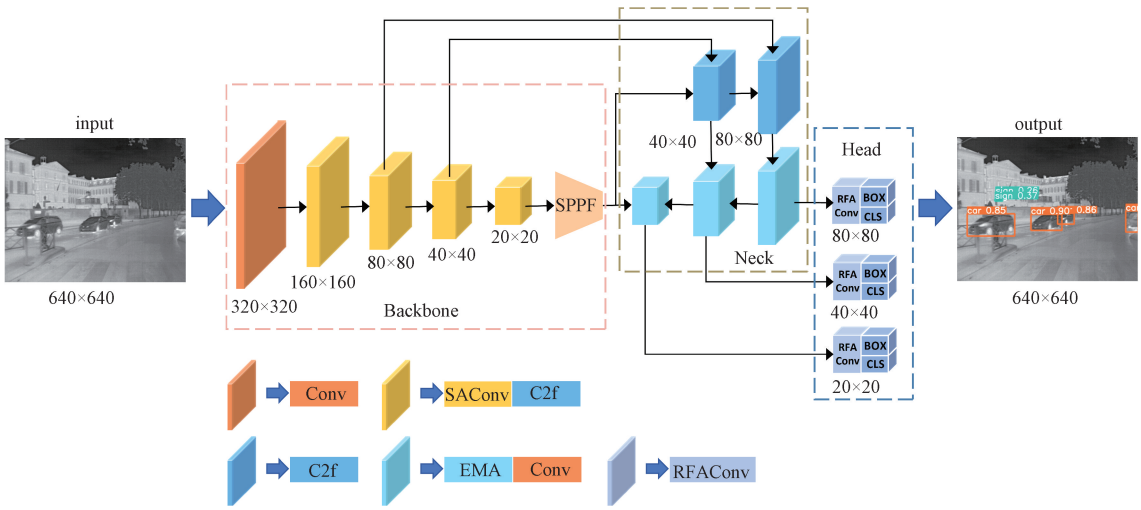


图 1 SAC-YOLO 结构

Fig. 1 Structure diagram of SAC-YOLO

在 Backbone 网络的 C2f 结构中添加变换空洞卷积，灵活适应不同尺度的特征，有效提升网络对于模糊信息的理解能力，在 Neck 网络中引入高效多尺度注意力模块，提升模型灰度图像中纹理细节的关注，Head 部分使用超分辨率重构检测头，结合空间注意力和感受野特征，强化对小目标区域的关注，避免漏检现象。

2.2 变换空洞卷积模块

一般卷积因卷积核大小固定，对于上下文信息捕捉有限，在提取灰度图像中模糊特征信息时出现困难，其主要表现为较小的目标可能会被大感受野忽略，而较大的

目标可能会受背景或局部信息影响,导致检测不准确。变换空洞卷积模块(switchable atrous convolution, SAConv)[18]可以使卷积层捕获更大范围特征,增强主干网络的特征提取能力,更好的理解复杂度较高的特征信息,SAConv 模块结构如图 2 所示。

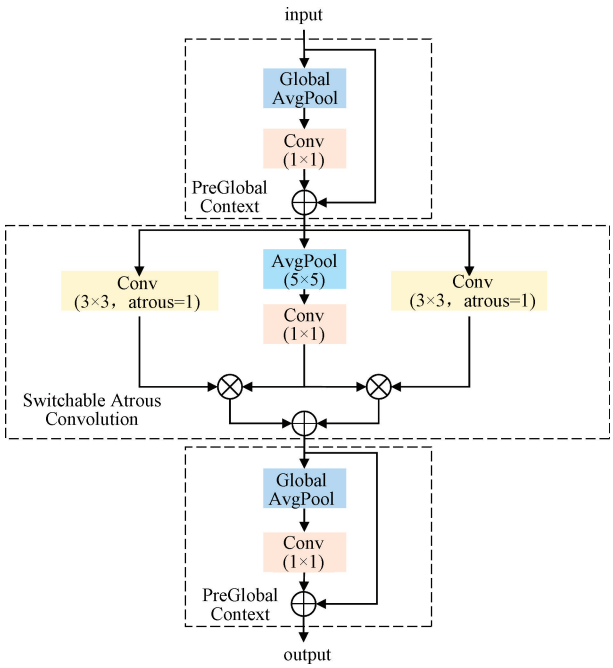


图 2　SAConv 模块结构

Fig. 2　SAConv module structure diagram

由图 2 可以看出,在 SAConv 构架中,输入图像首先通过一个全局上下文模块,提取整个图像特征,获得全局信息,并将全局信息送入空洞卷积层中,空洞卷积层由两个并行的空洞卷积构成,对输入信息进行两次观察,即为输入信息赋予两种空洞率,进行两次不同的空洞卷积。为避免较大空洞率导致的权重缺失问题,在卷积后设置开关函数,在开关函数中存入初始化权重,将并行卷积结果引入开关函数中由开关函数判定卷积融合后的输出权重和特征,然后将输出信息导入进第 2 个全局上下文模块,归纳并筛选空洞卷积层中的重要特征,形成更全面的图像信息,降低模型检测难度,SAConv 模块计算如式(1)所示。

$$Output = S(x) \times Conv(x,y,1) + (1 - S(x)) \times Conv(x,y,\Delta y,r) \tag{1}$$

式中:x 为输入信息;$Output$ 为输出信息;$Conv(x,y,r)$ 是权重为 y 的卷积运算;r 是 SAConv 的超参数;Δy 为可训练值;$S(x)$ 为开关函数。

在主干网络中 C2f 中加入 SAConv 结构,为输入特征应用不同的空洞率计算空洞卷积,将卷积结果进行权重共享,结合全局上下文模块统筹全局信息,提升网络的灵活性以及对于模糊信息的适应性。

2.3　高效多尺度注意力模块

由于灰度图像中色彩的单一性,导致图像中重要信息与背景对比度较低,细微特征和边缘信息的丢失,所以网络对于一些局部区域内的信息检测效果不佳。注意力机制则可以指导模型动态的关注于重要目标区域,忽略无关的背景信息,增强模型对于灰度图像中细节信息的理解能力,因此,在网络的特征融合部分引入高效多尺度注意力模块(efficient multi-scale attention, EMA)[19],EMA 结构如图 3 所示。

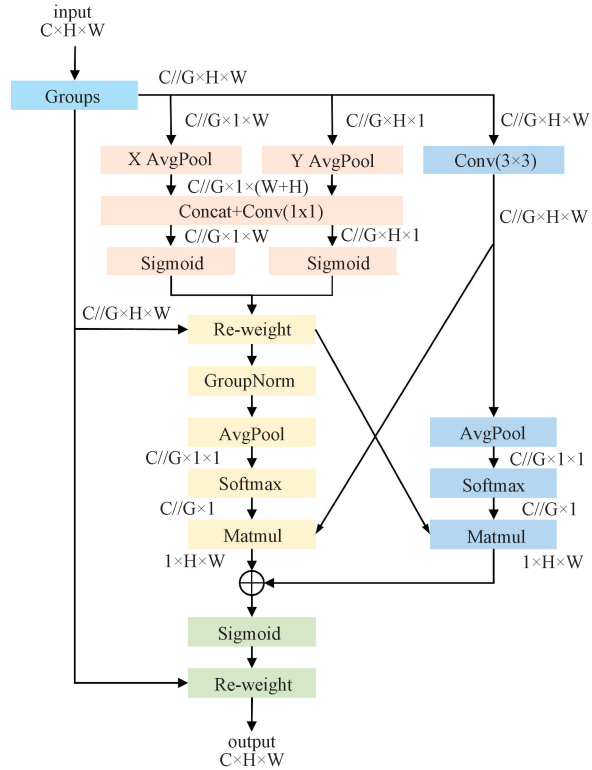


图 3　EMA 结构

Fig. 3　Structure diagram of EMA

EMA 先按照通道注意力机制的处理方法,将输入为 $G \times H \times W$ 的特征图 X 按通道维数方向划分成为 G 个子特征图,划分公式如式(2)所示。

$$X = [X_0, X_1, \cdots, X_{G-1}], X_i \in R^{C//G \times H \times W}, G \ll C \tag{2}$$

然后把各通道信息分解为为两个 1×1 的特征编码分支和一个 3×3 的特征编码分支。在 1×1 的两个特征编码分支上,分别沿着水平和垂直方向对通道进行一维全局平均池化操作,计算公式如式(3)和(4)所示。

$$Z_C^H = \frac{1}{W}\sum_{i=0}^{W} x_c(H,i) \tag{3}$$

$$Z_C^w = \frac{1}{H}\sum_{j=0}^{H} x_c(j,W) \tag{4}$$

式(3)为水平方向一维全局平均池化，Z_C^H 表示高度为 H 的第 C 个通道的输出；式(4)为垂直方向一维全局平均池化，Z_C^W 表示宽度为 W 的第 C 个通道的输出。

再将两个支路的特征编码沿高度方向特征拼接，经 1×1 卷积运算分解为两个向量，利用非线性 Sigmoid 函数线性拟合。然后将 1×1 分支的中特征信息用乘法聚合在一起，形成一条全新的支路，以实现跨通道特征信息交互。同时，3×3 的特征编码分支通过 3×3 的卷积获得局部区域内的语义信息。两条支路同时进入跨空间信息部分，1×1 分支先进行二维全局平均池化操作和 Softmax 函数线性变换，二维全局平均池化操作如式(5)所示。

$$Z_C = \frac{1}{H \times W} \sum_{j=0}^{H} \sum_{i=0}^{W} x_c(i,j) \tag{5}$$

随后将 1×1 分支和 3×3 分支的输出特征进行矩阵点积运算相乘，在通道特征联合激活机制之前将输出维度转化为 $R_1^{1\times1\times C//G} \times R_3^{C//G\times H\times W}$。3×3 分支在跨空间信息部分进行同样的操作，即先通过二维全局平均池化，将池化结果用 Softmax 线性变换，和 1×1 分支输出特征进行矩阵点积运算，并在联合激活机制之前将输出维度转化为 $R_3^{1\times1\times C//G} \times R_1^{C//G\times H\times W}$。最后将两条分支的输出特征结合，组成一个包含双重空间注意力权重的特征图，帮助模型捕捉图像中的像素级关系，增强特征表示的能力。

2.4　超分辨率重构检测头

YOLOv8 模型有大、中、小 3 种不同尺寸的检测头，在彩色图像的场景中均产生优异效果。但在灰度图像中，网络的特征提取和特征融合会导致特征图尺寸不断变小，小目标信息很可能会随着整体像素的压缩而被模型的检测头忽略。为解决上述问题，引入区域特征自适应卷积(receptive-field attention convolution，RFAConv)[20]构成超分辨率重构检测头，RFAConv 结构如图 4 所示。

超分辨率重构检测头是在解耦头的基础上，采用卷积和空间注意力机制相结合的方式，将标准卷积替换为 RFAConv，增强检测头部分的细节恢复能力和空间感知能力，从而提高模型对多尺度信息识别的鲁棒性。从图 4 可以看出，RFAConv 结构分为两个分支，输入为 $C\times H\times W$ 的特征图 X 通过感受野自适应卷积进入右边支路，感受野自适应卷积以一个 3×3 的大尺度卷积核为核心，引入一个同样尺寸的滑动窗口动态关注图像中重要信息，然后将卷积核信息与窗口信息相乘得到特征信息。左边支路通过平均池化聚合感受野特征的全局信息，然后运用 1×1 卷积进行信息交互，归一化操作关注其中的重要信息权重。采取参数共享策略，将注意力图中的权重与感受野卷积相结合提取特征，调整特征尺寸，输出感受野空间注意力特征信息。RFAConv 计算过程如式(6)所示。

$$F = Softmax(g^{1\times1}(AvgPool(X))) \times ReLU(Norm(g^{k\times k}(X))) = Arf \times Frf \tag{6}$$

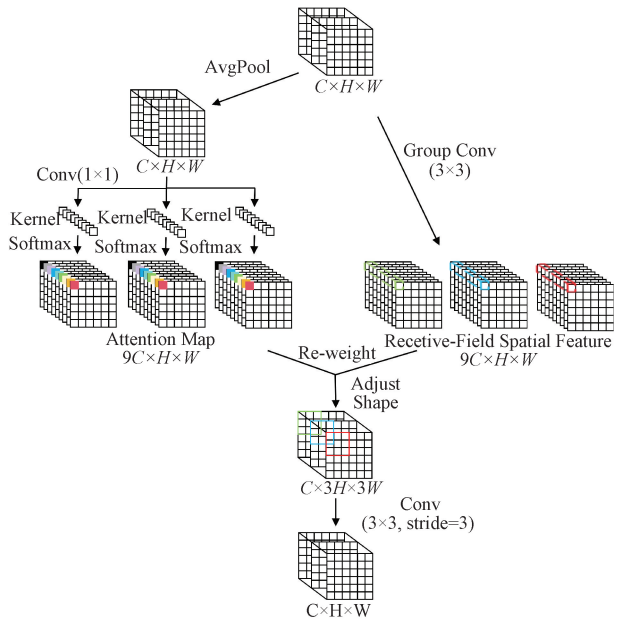


图 4　RFAConv 结构

Fig. 4　Structure diagram of RFAConv

式中：F 是特征输出；X 为特征输入；Arf 为注意力图的权重信息；Frf 为感受野空间特征信息；$g^{1\times1}$ 表示 1×1 的分组卷积；$AvgPool$ 代表平均池化；$Norm$ 是归一化操作；$ReLU$ 是激活函数。

3　实验与结果分析

3.1　实验环境与数据集

为确保实验的公平性，本文所有实验均在相同的实验设置和训练参数下进行，具体的实验配置如表 1 所示，训练参数如表 2 所示。

表 1　实验配置

Table 1　Experimental configuration

配置名称	配置信息
操作系统	Windows10
CPU	Intel Core i7 10750H
GPU	NVIDIA GeForce RTX 2070
编程语言	Python3. 9
算法框架	Pytorch1. 12
加速环境	CUDA11. 3

表 2　训练参数

Table 2　Training parameters

参数名称	参数信息
学习率	0. 01
权重衰减系数	0. 000 5
图像尺寸	640×640
批量大小	8
训练周期	300

为评估改进算法在灰度图像下的有效性和可行性，实验采用开源灰度数据集 NEU-DET 数据集进行消融实验和对比实验。NEU-DET 数据集是一个钢铁表面缺陷检测数据集，由总共 1 800 张图像组成，带有 6 类标注对象即 crazing、patches、inclusion、pitted_surface、rolled-in_scale 和 scratches，按照 7∶2∶1 的比例将数据集随机分为训练集，测试集和验证集[21]。

还使用两种不同领域的数据集来验证改进网络在灰度图像下的通用性，其中，FLIR-ADAS 数据集是由 FLIR 公司采集到的行人和车辆的灰度图像数据集，共 14 000 张图像，有 11 种不同尺度的标注对象。另一个数据集是 InfiRay 公司开源的红外航拍人车检测数据集，该数据集共 8 402 张图像，有 7 种标注对象，小目标信息较多。选择这两个数据集作为模型泛化能力实验数据集，同样按照 7∶2∶1 的比例进行划分[22-23]。

3.2　评价指标

采用准确率(precision，P)，召回率(recall，R)和平均精度均值(mean average precision，mAP)来评估模型的检测能力，选择模型参数量(Params)和浮点计算量(GFLOPs)来评估模型的复杂程度，以验证改进后模型对灰度图像的检测性能和效率。选择帧数(frames per second，FPS)来评估模型的计算速度。

准确率表示在识别出的物体中，检测到的正确目标在全部目标中的占比。召回率表示所有正确目标中，模型能够预测出来的比例。指标计算公式如式(7)和(8)所示。

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

式中：TP 表示为正确检测到目标信息的数量，即模型判断为正例的数；FP 表示为模型误判为目标信息的数量；FN 表示漏检的目标信息的数量。

平均精度(average precision，AP)表示模型对特定类型目标的检测准确性，以该目标 Precision 和 Recall 分别作为横纵坐标绘制而成的函数图像，对数据集中的全部类型目标的 AP 值进行平均，可得平均精度均值，用于网络模型的整体性能评估，其公式如式(9)和(10)所示。

$$AP = \int_0^1 P(R)\,\mathrm{d}R \tag{9}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP \tag{10}$$

式中：n 为类别个数。

参数量反应网络结构的复杂程度，结构越复杂，内存占用的参数量就越多，浮点计算量反应模型训练或检测时的计算性能，浮点计算量越大，说明模型具有更高的计算能力。FPS 表示网络每秒处理的图像的帧数，帧率越高说明网络的计算效率越高。

3.3　消融实验

为证明 SAC-YOLO 中变换空洞卷积，高效多尺度注意力机制和超分辨率重构检测头的有效性，以 YOLOv8 网络作为基础，对基础模型逐步添加改进模块，在 BCCD 数据集下训练，在 NEU-DET 数据集进行模块的消融实验，以分析以上改进点对整体算法的影响。

由消融实验结果表 3 可以得出，在灰度图像检测中 YOLOv8 的 mAP@0.5 为 76.2%，mAP@0.5∶0.95 为 44.7%，准确率和召回率分别是 74.8%和 66.7%，参数量是 3×10^6，浮点数是 8.2 GFLOPs，帧数为 65 fps。首先在主干网络中引入 SAConv 模块，与基线模型相比 mAP@0.5 提升了 0.9%，召回率达到 73.1%，浮点数下降了 0.8 GFLOPs，帧数提升至 78 fps，但是准确率有所下降，SAConv 使用开关函数组合两种不同的空洞率，为输入信息提供多样化空洞卷积，获得更为丰富的上下文信息，但削弱了网络对于局部信息的关注度。

表 3　消融实验结果

Table 3　Results of ablation experiments

模型	mAP@0.5/%	mAP@0.5:0.95/%	P/%	R/%	Params/($\times10^6$)	浮点数/GFLOPs	帧率/fps
YOLOv8	76.2	44.7	74.8	66.7	3	8.2	65
YOLOv8+SAConv	77.1	46	67.9	73.1	3.3	7.4	78
YOLOv8+EMA	78.1	45.8	68.9	75.6	3	8.3	62
YOLOv8+RFAHead	77.6	45.5	73.9	70.8	3.9	8.4	58
YOLOv8+SAConv+RFAHead	77	45.2	73.7	68.7	4.2	7.8	70
YOLOv8+SAConv+EMA	78.4	46.4	74.9	70.7	3.3	7.5	75
YOLOv8+SAConv+EMA+RFAHead	79.3	46.4	75.3	71.2	4.2	7.9	72

在颈部网络中添加 EMA 注意力机制，通过对输入信息进行通道维数重组以及跨维度信息交互，将网络的特征提取能力放大至像素级，在引入少量参数量和浮点数的同时，网络的 mAP@0.5 提升到 78.1%，但由于检测头对于灰度图像整体像素的压缩，仍会有少量重要信息被忽略，导致模型准确率只有 68.9%。所以在检测头添加

感受野注意力模块,根据输入信息的复杂性和重要性动态调整感受野,提升大尺寸卷积核效率,有效捕捉重要空间信息,使得网络 mAP@0.5 达到 77.6%,召回率提升了 5.9%,准确率与基础网络基本相同,参数量和浮点数分别是 3.9×10^6 和 8.4 GFLOPs,检测效率略微下降。引入 SAConv 的同时替换检测头为 RFAHead,使得网络的主干部分和检测头部分同时实现参数共享,提升网络在灰度图像检测中的灵活性和自适应性,相较于 YOLOv8 网络,mAP@0.5 提升了 0.5%,准确率为 73.7%,召回率上涨到 68.7%。为兼顾网络的全局信息识别能力与局部特征捕捉能力,在 YOLOv8 的基础上结合 SAConv 与 EMA 模块,特征提取部分运用 SAConv 获得更为广泛的灰度图像信息,特征融合部分引入 EMA 注意力机制,整合并关注细节信息,网络的 mAP@0.5 相较于先前的网络提升至 78.4%,准确率得到有效改善,达到 74.9%,召回率为 70.7%。在第 5 组实验的网络基础上,将检测头改进为 RFAHead,mAP@0.5 进一步提升至 79.3%,准确率和召回率将较于先前实验也有所提升,分别为 75.3% 和 71.2%,参数量对比基础网络有略微提升,帧数为 72 fps,检测效率相较于基础网络有所提升。综上所述,所采用的改进点有效提高了模型对于灰度图像的目标检测性能,充分发挥了各改进点的作用,证明了 SAC-YOLO 的有效性和可行性。

3.4 对比实验

在保证相同实验环境及数据集的前提下,将 SAC-YOLO 与现阶段主流算法进行综合对比实验,选择的算法包括 Faster R-CNN、Cascade R-CNN[24]、RetinaNet[25]、SSD、YOLOv3[26]、YOLOv5、YOLOv7、YOLOX、YOLOv8、RT-DETR。在 NEU-DET 数据集上以 mAP@0.5、mAP@0.5∶0.95、Params、浮点数和帧数的实验参数做对比,验证改进后网络模型的检测性能,实验结果如表 4 所示。

表 4 NEU-DET 数据集对比实验

Table 4 Comparative experiments on NEU-DET dataset

Method	mAP@0.5/%	mAP@0.5:0.95/%	Params/($\times10^6$)	浮点数/GFLOPs	帧率/fps
SSD	62.1	33.8	41.1	145.3	40.7
YOLOv3-Tiny	61.8	34.9	8.2	12.9	27.8
YOLOv5n	74	37.9	1.7	9.5	30.9
YOLOv5s	75.2	40.4	7	16.4	26.8
YOLOv7-Tiny	75.8	36	6.1	13.1	33.4
YOLOX-s	69.1	31.3	8	21.6	40.3
RetinaNet	74.9	-	36.2	124.4	25.5
Faster R-CNN	72.2	-	72	167.3	16.9
Cascade R-CNN	73.2	-	84.6	26.64	19.3
YOLOv8	76.2	44.7	3	8.2	65
RT-DETR	71.2	39.8	29.2	105.2	42
SAC-YOLO	79.3	46.4	4.2	7.9	72

从表 4 的实验结果可以得出,YOLOv3-Tiny 和 SSD 算法的检测精度最差,YOLOv3-Tiny 的 mAP@0.5 只有 61.8%,SSD 算法在 41.1×10^6 高参数量的情况下,mAP@0.5 仅比 YOLOv3-Tiny 高 0.3%,效果不佳。Faster R-CNN 和 Cascade R-CNN 的平均精度均值分别达到 72.2%和 73.2%,检测效果较好,但作为二阶段算法,其参数量和浮点数过于庞大,帧数只有 16.9 和 19.3 fps,实时检测效率低。RetinaNet 网络的 mAP@0.5 为 74.9%,参数量和浮点数分别是 36.2×10^6 和 124.4 GFLOPs,检测性能优于先前列出的一阶段和二阶段算法,但是它的浮点数较大,在实际应用中部署较为困难。同时选择以 Transformer 网络为基础,进行特征提取和关联学习的 RT-DETR 目标检测算法进行对比实验,通过结果可以看出,相较于 YOLO 系列,该算法计算复杂度仍然略高,参数量和浮点数达到了 29.9×10^6 和 105.2 GFLOPs,且识别精度要低于 YOLOv8。在 YOLO 系列中,YOLOv5 和 YOLOv7 算法开发程度较高,YOLOv5 选择的是 YOLOv5n 和 YOLOv5s 版本,识别精度为 74%和 75.2%,YOLOv7 算法选择的则是复杂度相对较低的 YOLOv7-Tiny 版本,其识别精度达到 75.8%。但相较于以上算法,YOLOv8 算法 mAP@0.5 是 76.2%,检测效果最好,因此选择以 YOLOv8 作为基础模型,提出 SAC-YOLO 算法,改进算法的 mAP@0.5 为 79.3%,mAP@0.5∶0.95 为 46.4%,参数量和浮点数是 4.2×10^6 和 7.9 GFLOPs,在引入少量参数量的前提下,识别精度相比 YOLOv8 网络上涨 3.1%,帧数为 72 fps,可以满足在灰度图像检测中检测精度和检测效率的需求。

3.5 泛化性实验

为体现改进后模型的泛化能力,展现模型在不同类型灰度图像中仍具有良好的适应能力,在 FLIR-ADAS 和红外航拍人车检测数据集下进行泛化性实验,以测试 SAC-YOLO 对灰度图像检测的鲁棒性,对比实验结果如图 5 和 6 所示,横坐标表示训练轮数,纵坐标表示平均精度均值。

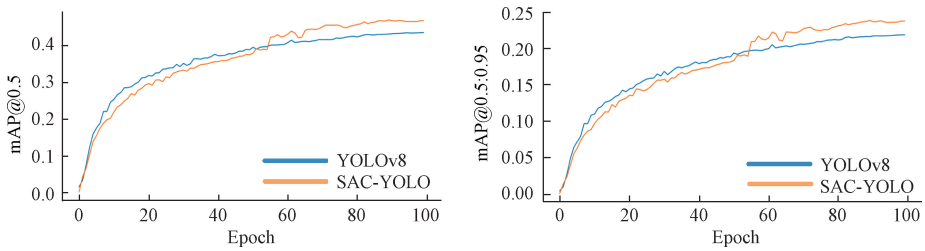


图 5　FLIR-ADAS 数据集上的泛化性实验结果

Fig. 5　Experimental results of generalization on FLIR-ADAS dataset

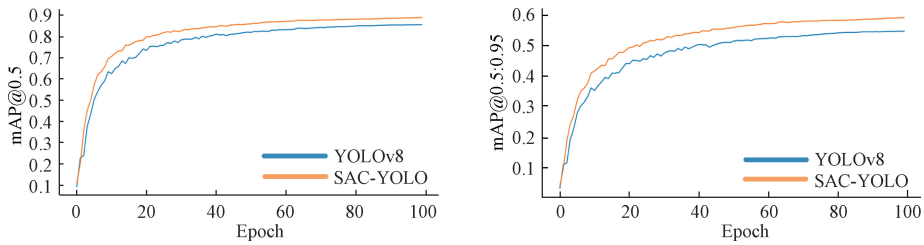


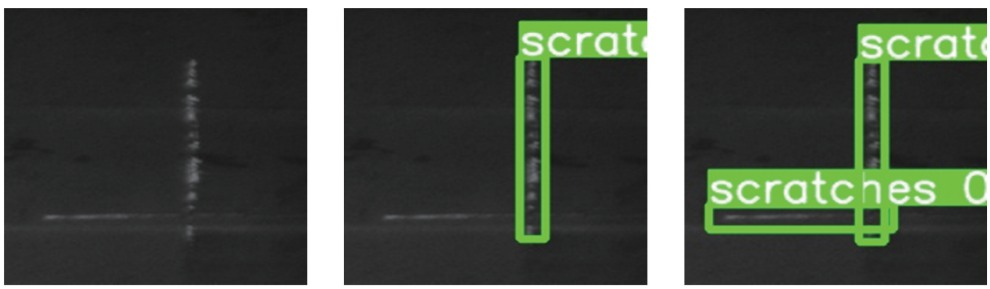
图 6　红外航拍人车检测数据集上的泛化性实验结果

Fig. 6　Experimental results of generalization on infrared aerial person-car detection dataset

由图 5 可以看出，相较于 YOLOv8 基础网络，SAC-YOLO 模型能够提取更为丰富的全局特征和局部特征，在 60 轮往后检测性能得到有效提高，说明改进网络不仅适用于处理图像模糊和目标边缘化的问题，在 FLIR-ADAS 这种多尺度信息数据集仍有较好表现。InfiRay 公司开源的红外航拍人车检测数据集图像是由航拍无人机捕捉到的，所以其图像中目标信息大多为密集的小目标信息，如图 6 所示，SAC-YOLO 在此数据集上，mAP@0.5 和 mAP@0.5∶0.95 的值始终比 YOLOv8 有着更加显著的提升，且收敛速度更快。综上，SAC-YOLO 模型在 NEU-DET 数据集上具备优异性能，且适用于其他灰度图像场景下的目标检测任务，具备良好的泛化性。

3.6　可视化实验

为直观展示改进方法在不同灰度图像场景下的表现，在数据集中随机选取灰度图像进行测试，测试结果如图 7 所示。图 7(a)~(c)分别代表原始图像，YOLOv8 检测结果图像以及 SAC-YOLO 检测结果图像。



(a) 原始图像
(a) Original image

(b) YOLOv8检测结果
(b) Results of YOLOv8 detection

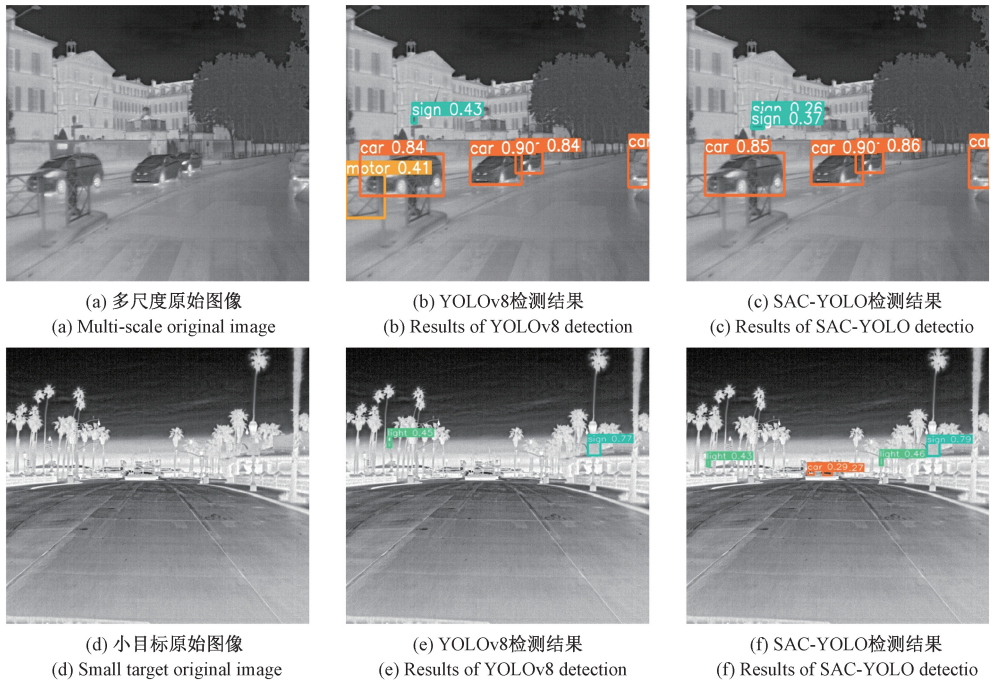
(c) SAC-YOLO检测结果
(c) Results of SAC-YOLO detection

图 7　钢铁表面缺陷检测效果对比

Fig. 7　Comparison of steel surface defect detection effects

从图 7 可以看出，灰度图像中目标信息模糊，对比度较低，YOLOv8 网络很难准确识别图像中的缺陷信息，导致出现漏检的问题，而 SAC-YOLO 采用双重观察机制，充分把握图像特征，有效区分背景信息和缺陷信息，增强网络的特征提取能力。在图 8 的 Multi-scale image 图中，展现了不同尺度的灰度图像信息对基础网络造成的影响，如对于图中的大尺度背景信息，YOLOv8 网络误检为 motor，对于小目标的 sign 目标则是出现了漏检的问题，在 Small target image 图中，则进一步展现了小目标的灰度图像信息对基础网络造成的影响，如对于图片右侧路

灯以及远距离车辆，YOLOv8网络出现出现了漏检的问题，SAC-YOLO以变换空洞卷积把握多元化信息特征，同时引入EMA注意力机制关注灰度图像的局部信息，大大增强网络对于多尺度目标特征的提取能力，因此能够更好的检测出多尺度的目标信息。图9的Original image为航拍下的灰度车辆信息，图像中车辆目标密集，SAC-YOLO在检测头部分的感受野中引入空间注意力，有效还原在特征提取和特征融合阶段被压缩的特征信息，进一步增强模型的理解能力，相较于基础网络，改进网络在该场景中也表现出良好的性能。



(a) 多尺度原始图像
(a) Multi-scale original image

(b) YOLOv8检测结果
(b) Results of YOLOv8 detection

(c) SAC-YOLO检测结果
(c) Results of SAC-YOLO detectio

(d) 小目标原始图像
(d) Small target original image

(e) YOLOv8检测结果
(e) Results of YOLOv8 detection

(f) SAC-YOLO检测结果
(f) Results of SAC-YOLO detectio

图8　道路场景检测效果对比

Fig. 8　Comparison of road scene detection effects

4　结　论

针对灰度图像对比度低、识别精度差、检测难度高等问题，提出一种融合双重观察与注意力机制的目标检测算法SAC-YOLO。为解决灰度图像中因特征模糊而导致的误检漏检的问题，基于变换空洞卷积重新构建特征提取网络，采用双重观察机制获取全局信息，提升模型对于复杂特征信息的理解能力为解决特征融合阶段小目标特征信息难以捕捉的问题，引入高效多尺度注意力模块，重点关注局部信息特征，以较小的计算成本有效提升特征融合能力，增强模型对边缘和细节信息的捕捉能力；通过添加超分辨率重构检测头，获得感受野的空间信息特征，兼顾多层次特征信息，确保模型在灰度图像场景下的检测稳定性更高。实验结果表明，所提出的SAC-YOLO模型在灰度表面缺陷检测中识别精度达到79.3%，帧率达到72 fps，检测性能以及检测效率均优于现阶段目标检测模型。通过泛化性实验，证明方法不仅适用于灰度缺陷检测，在多场景灰度图像检测中同样拥有良好的鲁棒性。为适应更加复杂和多样化的灰度图像检测，在未来研究中将继续优化算法结构，以轻量化网络为前提，进一步提升网络的检测精度和检测效率。

参考文献

[1] 刘洪江，王懋，刘丽华，等. 基于深度学习的小目标检测综述[J]. 计算机工程与科学，2021，43(8)：1429-1442.
LIU H J，WANG M，LIU L H，et al. Survey of small object detection based on deep learning [J]. Computer Engineering and Science，2021，43(8)：1429-1442.

[2] 张传聪，李范鸣，饶俊民. 基于特征显著性融合的红外小目标检测[J]. 半导体光电，2022，43(4)：828-834.
ZHANG CH C，LI F M，RAO J M. Based on characteristics of significant fusion of infrared small target detection [J]. Journal of Semiconductor Optoelectronic，Lancet 2022，43(4)：828-834.

[3] 叶飞，骆星智，宋永春，等. 基于双特征融合的改进R-CNN电力小金具缺陷检测方法研究[J]. 电子测量与仪器学报，2023，37(7)：213-220.
YE F，LUO X ZH，SONG Y CH，et al. Research on

defect detection method of electric small metal tools based on improved R-CNN with double feature fusion [J]. Journal of Electronic Measurement and Instrument, 2023, 37(7): 213-220.

[4] REN SH 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, 2016, 39(6): 1137-1149.

[5] 蔡彪,沈宽,付金磊,等. 基于 Mask R-CNN 的铸件 X 射线 DR 图像缺陷检测研究[J]. 仪器仪表学报,2020, 41(3):61-69.
CAI B, SHEN K, FU J L, et al. Research on defect detection of casting X-ray DR image based on mask R-CNN [J]. Chinese Journal of Scientific Instrument, 2020, 41(3): 61-69.

[6] CHU H CH, WANG T H, MIAO Q N, et al. Ship detection based on improved SDD algorithm [J]. Instrumentation, 2024, 11(4): 35-43.

[7] WANG CH Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. 2023IEEE/CVF Conference on Computer Vision and Parttern Recognition (CVPR). New York: IEEE, 2023: 7464-7475.

[8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection With Transformers[M]. 2020.

[9] GHOSE D, DESAI S M, BHATTACHARYA S, et al. Pedestrian detection in thermal images using saliency maps[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[10] LIU ZH Y, ZHANG X S, JIANG T P, et al. Infrared salient object detection based on global guided lightweight non-local deep features [J]. Infrared Physics & Technology, 2021, 115: 103672.

[11] LI SH SH, LI Y J, LI Y, et al. YOLO-firi: Improved yolov5 for infrared image object detection [J]. IEEE Access, 2021, 9: 141861-141875.

[12] HU SH M, ZHAO F, LU H ZH, et al. Improving YOLOv7-tiny for infrared and visible light image object detection on drones [J]. Remote Sensing, 2023, 15(13): 3214.

[13] ZHOU J J, ZHANG B H, YUAN X L, et al. YOLO-CIR: The network based on YOLO and ConvNeXt for infrared object detection [J]. Infrared Physics & Technology, 2023, 131: 104703.

[14] TERVEN J R, CORDOVA-ESPARAZA D M, et al. Ultralytics/YOLOv8: YOLOv8 docs[EB/OL]. (2023-01-10) [2024-05-21].

[15] JOCHER G, STOKEN A, BOROVEC J, et al. Ultralytics/YOLOv5: v3.1-bug fixes and performance improvements [EB/OL]. (2020-10-29) [2024-05-21].

[16] JIANG K L, XIE T Y, YAN R, et al. An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation[J]. Agriculture, 2022, 12(10): 1659.

[17] PANBOONYUEN T, THONGBAI S, WONGWEERANIMIT W, et al. Object detection of road assets using transformer-based YOLOX with feature pyramid decoder on thai highway panorama [J]. Information, 2021, 13(1): 5.

[18] QIAO S Y, CHEN L CH, YUILLE A. DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]. Computer Vision and Pattern Recognition. IEEE, 2021.

[19] OUYANG D L, HE S, ZHAN G, et al. Efficient multi-scale attention module with cross-spatial learning [C]. ICASSP2023-2023IEEEInternational Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023: 1-5.

[20] ZHANG X, LIU CH, YANG D G, et al. RFAConv: Innovating spatial attention and standard convolutional operation[J]. ArXiv preprint arXiv: 2304.03198, 2023.

[21] HE Y, SONG K CH, MENG Q G, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493-1504.

[22] WAN D H, LU R SH, HU B T, et al. YOLO-MIF: Improved YOLOv8 with multi-Information fusion for object detection in gray-scale images [J]. Advanced Engineering Informatics, 2024, 62: 102709.

[23] 周进,裴晓芳. 基于注意力与量化感知的航拍红外目标检测[J]. 计算机系统应用, 2024, 33(11): 111-120.
ZHOU J, PEI X F. Aerial infrared target detection based on attention and quantitative perception [J]. Computer System Applications, 2024, 33(11): 111-120.

[24] LI F. Bag of tricks for fabric defect detection based on Cascade R-CNN[J]. Textile Research Journal, 2021, 91(5-6): 599-612.

[25] MIAO T, ZENG H CH, YANG W, et al. An improved

lightweight RetinaNet for ship detection in SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 4667-4679.

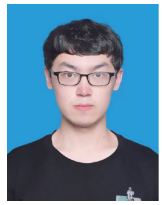
[26] ZHAO L Q, LI SH Y. Object detection algorithm based on improved YOLOv3[J]. Electronics, 2020, 9(3): 537.

作者简介

朱硕(通信作者),2014 年于中国科学院大学获得博士学位。现为无锡学院副教授,主要研究方向为计算机视觉、智能感知等。

E-mail: zshuo2011@ 163. com

Zhu Shuo (Corresponding author) received her Ph. D. from the University of Chinese Academy of Sciences in 2014. She is currently an associate professor at Wuxi University. Her research interests include computer vision and intelligent perception.



张绪康,2023 年于山东航空学院获得学士学位,现为南京信息工程大学硕士研究生,主要研究方向为深度学习和目标检测。

E-mail: aa3235539536@ 163. com

Zhang Xukang received his B. Sc. degree from Shandong Aeronautical University in 2023. He is currently a M. Sc. candidate at Nanjing University of Information Science and Technology. His main research interests include deep learning and object detection.