

DOI: 10.13382/j.jemi.B2407917

改进 YOLOv8n 的遥感图像目标检测算法

王海群¹ 武泽锴¹ 晁 帅²

(1. 华北理工大学电气工程学院 唐山 063000; 2. 华北理工大学招生就业处 唐山 063000)

摘要:针对遥感图像中目标尺度差异较大、类别多样且分布不均等因素造成检测时目标定位不精准、漏检和误检等问题,提出一种改进 YOLOv8n 的遥感图像目标检测算法。首先,构建 SC_C2F 模块作为主干网络的特征提取模块,通过在 Bottlececk 结构中引入空间通道重建卷积,增强不同尺度通道和空间的特征提取能力;其次,设计 ESPPM 模块替换原金字塔池化模块,引入自适应平均池化层与大可分离核残差注意力机制,丰富上下文信息,提高模型多尺度特征聚合能力;再次,结合 GSConv 轻量化卷积与 VoVGSCSP 结构,引入 Slim-PAN 结构到颈部网络,在减少模型计算量的同时保持检测精度;最后,引入增加参数表示法的旋转框作为角度坐标回归,并设计 RBCL 损失函数计算旋转框损失,使检测过程更加贴合目标形状,提高对小目标和密集目标的检测精度。将改进的 YOLOV8n 算法在 DOTA 数据集上进行实验,相较原算法 mAP@0.5 提高 5.1%,计算量降低 0.4 GFLOPs。

关键词: 遥感图像; YOLOv8n; 特征提取模块; 损失函数; 旋转框; 注意力机制

中图分类号: TP391.4; TN957.52 **文献标识码:** A **国家标准学科分类代码:** 520.6020

Improve the YOLOv8n object detection algorithm for remote sensing images

Wang Haiqun¹ Wu Zekai¹ Chao Shuai²

(1. Faculty of Electrical Engineering, North China University of Science and Technology, Tangshan 063000, China;

2. Admissions and Employment Office, North China University of Science and Technology, Tangshan 063000, China)

Abstract: To address the issues of inaccurate target localization, missed detections, and false detections caused by large scale differences, diverse categories, and uneven distribution of targets in remote sensing images, an improved YOLOv8n remote sensing image target detection algorithm is proposed. Firstly, the SC_C2F module is constructed as the feature extraction module of the backbone network. By introducing spatial channel reconstruction convolution into the Bottlececk structure, the feature extraction ability of different scale channels and spaces is enhanced; Secondly, design an ESPPM module to replace the original pyramid pooling module, introduce an adaptive average pooling layer and a large separable kernel residual attention mechanism, enrich contextual information, and improve the model's multi-scale feature aggregation ability; Again, by combining GSConv lightweight convolution with VoVGSCSP structure, the Slim PAN structure is introduced into the neck network to reduce model computation while maintaining detection accuracy; Finally, a rotation box with added parameter representation is introduced as the angle coordinate regression, and an RBCL loss function is designed to calculate the rotation box loss, making the detection process more in line with the target shape and improving the detection accuracy for small and dense targets. The improved YOLOV8n algorithm will be tested on the DOTA dataset and compared to the original algorithm mAP@0.5 Increase by 5.1% and reduce computational load by 0.4 GFLOPs.

Keywords: remotely sensed imagery; YOLOv8n; feature extraction module; loss function; rotating boxes; attention mechanisms

0 引言

随着遥感技术的迅猛发展,获取高分辨率遥感图像已成为国家在环境监测和城市规划等多个领域做出决策的关键手段为现代城市管理和发展提供宝贵的数据支持。

在分析遥感图像时,目标检测技术^[1]负责从遥感图像中精确识别和定位图像中的目标物体。基于深度学习的目标检测算法主要分为两阶段检测器和单阶段检测器两大类。两阶段检测器以 R-CNN^[2] 系列为代表,典型算法包括 Fast R-CNN^[3]、Faster R-CNN^[4] 和 Mask R-CNN^[5] 等。该方法先生成候选区域,在候选区域上进行目标分类和定位。单阶段检测器如 YOLO 系列(YOLOv3^[6]、YOLOv5^[7]、YOLOv7^[8])、SSD^[9]、Centernet^[10] 等,该方法直接从输入图像中生成目标的类别和位置,不需要额外生成候选区域。

近年国内外在遥感图像目标检测领域的研究取得诸多进展。为解决微小物体特征尺度不匹配及遥感图像中环境干扰的问题,Chen 等^[11]提出跨尺度空间感知引导网络(CSPGNet),优化微小物体的特征表示。唐建宇^[12]等结合注意力机制与旋转框,增强图像特征的表达效果,提升目标检测精度。闫钧华等^[13]提出一种弱小目标检测算法 CC-YOLO,通过端到端的目标检测,提升小目标检测性能。雷帮军等^[14]通过引入上下文空间感知模块,增强网络对多尺度目标特征的提取能力。马梁等^[15]提出一种动态选择机制的轻量化特征提取模块,增强模型对小目标的检测能力。朱圣博等^[16]通过改进 YOLOv8s,利

用通道分组和跨通道信息交互,实现小目标的高精度检测。尽管上述研究在一定程度上解决了遥感图像目标检测中的部分问题,但仍存在以下不足:1)现有方法在处理多尺度和旋转目标时,检测精度仍有待提升;2)遥感图像中复杂背景和任意方向的目标易导致定位不准确、漏检和误检等问题;3)现有模型在兼顾检测速度和精度方面仍有优化空间。特别是 YOLOv8n 模型,虽然具有检测速度快、模型小的优势,但在多尺度和旋转目标上的检测精度仍需进一步提高。

针对上述问题,本文在 YOLOv8n 模型的基础上,优化算法结构,提出了一种改进 YOLOv8n 的遥感图像目标检测算法。具体研究如下:1)设计重建跨阶段特征提取模块 SC_C2F,增强模型的特征提取能力;2)设计自适应大可分离残差空间金字塔模块 ESPPM 和 RBCL 损失函数,增强模型的多尺度特征聚合能力和对任意方向目标的定位精度;3)在颈部网络引入轻量化特征金字塔模块 Slim_PAN,在保证检测速度的同时提高对遥感目标的检测精度。与现有研究相比,所研究方法在复杂背景和任意方向下对多尺度遥感目标的检测能力有所提升。

1 YOLOv8n 算法介绍

YOLOv8 是在 2023 年初开发和发布,是一种先进的实时性目标检测算法。YOLOv8n 主要由主干网络(backbone)、特征金字塔网络(neck)和输出网络(head)3个关键部分构成,具体结构如图1所示。

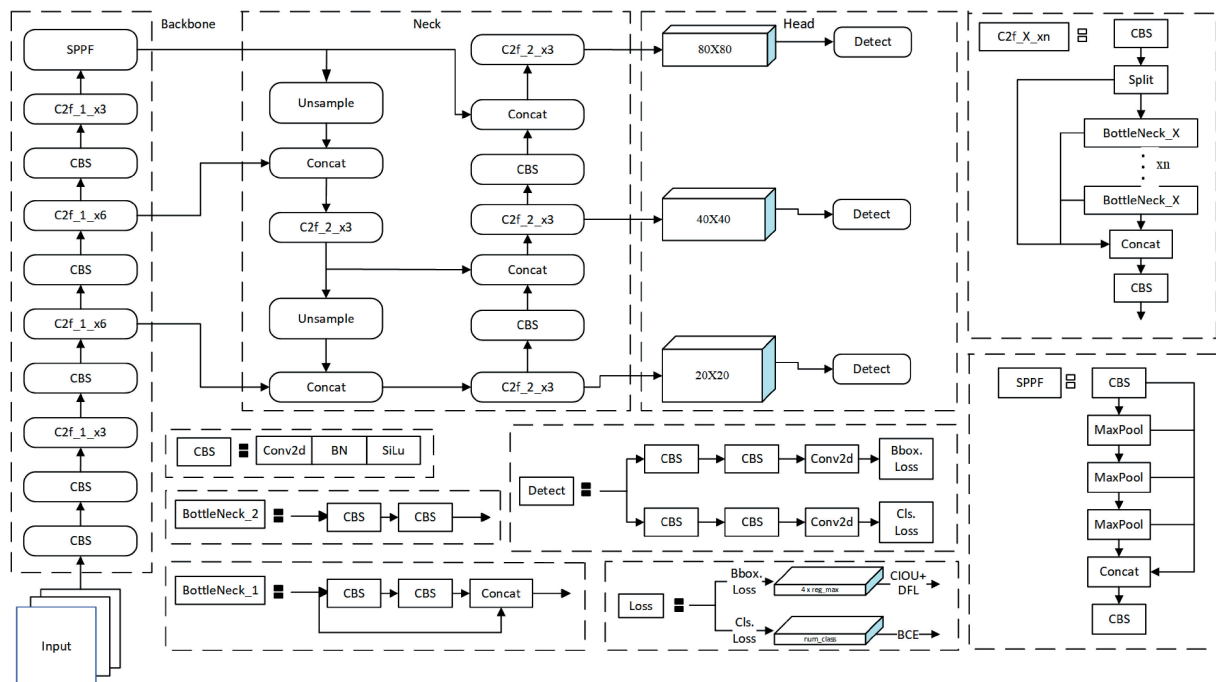


图1 YOLOv8n 算法结构

Fig. 1 Structure of the YOLOv8n algorithm

主干网络采用跨阶段部分 Darknet53 (CSPDarknet53) 结构, 该结构具有高效的特性提取能力, 能从输入图像中提取多层次、多尺度的特征。特征提取模块继承跨阶段部分网络 (CSPNet) 的设计思想进行结构优化, 该模块融合跨阶段信息具有更高效的特征细节保留能力。

特征金字塔网络采用金字塔网络 (FPN) + 路径聚合网络 (PAN) 结构, FPN 通过自顶向下和横向连接的方式进行特征融合。PAN 通过自底向上的路径实现多尺度特征聚合。两者结合将来自主干网络的多层特征进一步融合, 共同处理多尺度信息。输出网络采用解耦头结构, 采用无锚框取代传统锚框设计, 包含多个输出层, 每层处理不同尺度的特征, 从融合特征中进行目标检测, 预测不同尺度的目标类别、位置和置信度。

2 YOLOv8n 算法改进

针对 YOLOv8n 基线模型对遥感图像目标检测目标

定位不准、漏检和误检等问题, 本文提出一种改进 YOLOv8n 的遥感图像目标检测模型。

1) 设计 SC_C2F 模块, 通过引入空间通道重建卷积到 Bottleneck 结构, 减少网络参数冗余, 增强模型对不同尺度空间和跨通道的特征提取能力。

2) 设计 ESPPM 模块, 通过引入自适应平均池化层和大可分离核残差注意力机制到 SPPF 模块, 保留图像细粒度特征, 实现高效多尺度的特征融合, 增强模型特征表示能力。

3) 通过结合分组混洗卷积 GSCov 和跨阶段模块 VoVGSCSP, 在颈部网络处引入 Slim-PAN 结构, 保持训练速度和模型精度。

4) 增加参数表示法作为旋转框的角度坐标回归, 并提出 RIoU 损失函数, 采用具有角度周期性与边长一致性惩罚项的 RBCL 损失函数, 计算旋转框检测损失, 提高目标检测精度。

改进的 YOLOv8n 算法结构如图 2 所示。

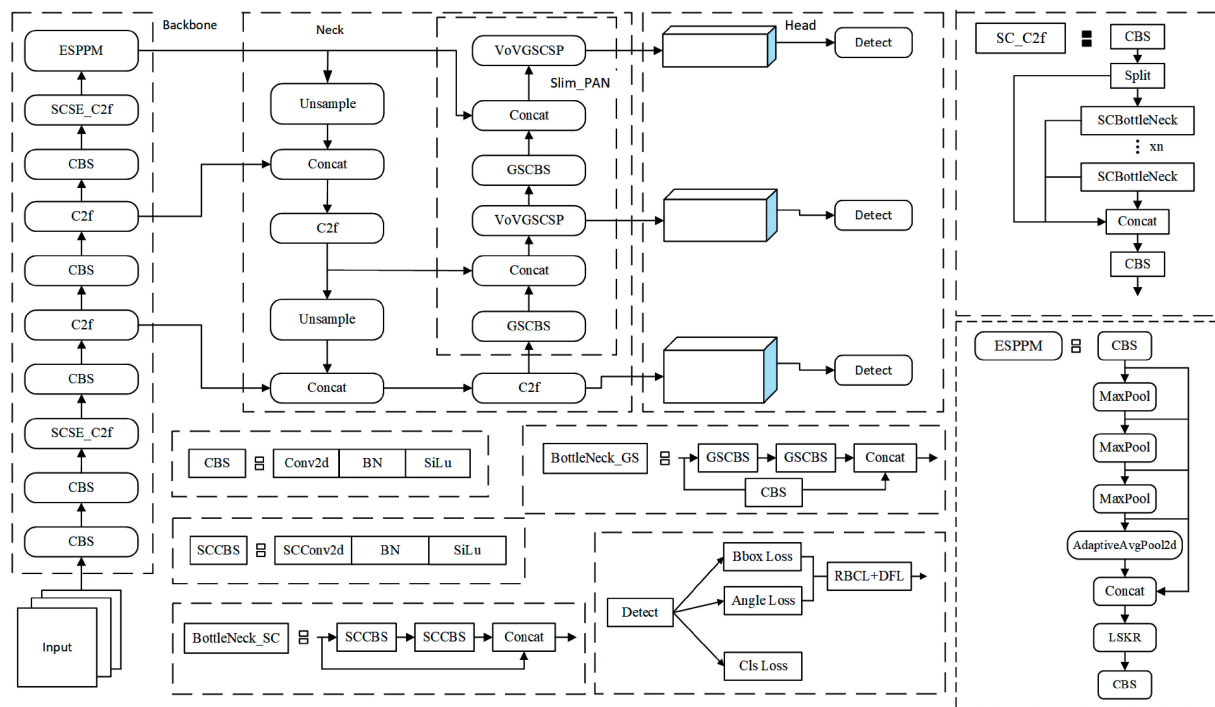


图 2 改进后的 YOLOv8n 算法结构

Fig. 2 Improved YOLOv8n algorithm structure

2.1 SC_C2F 模块

卷积模块在计算机视觉任务中常伴随大量计算参数冗余, 为解决计算冗余和提取特征不完善的问题, 构建 SC_C2F 模块。该模块通过在 Bottleneck 结构中引入 SCConv 替代传统卷积操作, 构建 Bottleneck_SC 模块, 并利用瓶颈操作限制特征提取过程参数冗余, 增强网络特

征表示能力。

SCConv^[17] 模块由空间重建单元 (SRU) 和通道重建单元 (CRU) 组成, 其目的是对特征图进行空间和通道重构, 提高特征提取能力。其中, SRU 通过重新排列特征图的空间结构, 捕捉更为丰富的空间信息, 而 CRU 则通过对特征图的通道进行动态调整, 使得每个通道更有效地

表示特征,其结构如图3所示。特征图 X 通过 SRU 获得空间细化特征 X_w ,再通过 CRU 获得通道细化特征 Y ,最终输出增强后的关键特征,具体实现流程如下。

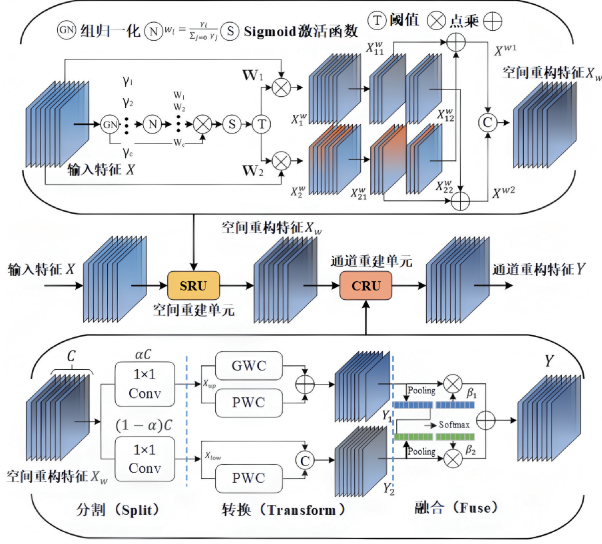


图3 SCConv 结构

Fig. 3 SCConv structure diagram

假设上一层输入特征图 X 如式(1)所示。

$$X \in R^{C \times H \times W} \quad (1)$$

式中: C 为特征图通道数; H 为特征图高度; W 为特征图宽度。首先,通过减去平均值 μ 并除以标准差 σ 来标准化输入特征 X ,如式(2)所示。

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (2)$$

式中: μ 和 σ 分别代表 X 的均值和标准差; ε 是小量常数; γ 和 β 则是仿射变换参数; GN 为组归一化层。归一化相关权重 $W_\gamma \in R^c$ 用于表示各特征映射的重要性,由式(3)得到。

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^c \gamma_j} \quad i, j = 1, 2, \dots, C \quad (3)$$

将经 W_γ 重新加权的特征映射的权值通过 Sigmoid 函数映射到 $(0,1)$ 范围,并通过阈值进行门控。将阈值以上的权重设置为 1,得到信息权重 W_1 ,将其设置为 0,得到非信息权重 W_2 , W 的获取过程如式(4)所示。

$$W = Gate(Sigmoid(W_\gamma(GN(X)))) \quad (4)$$

最后,将输入特征 X 分别乘以 W_1 和 W_2 ,得到两个加权后的输入特征:

$$\begin{aligned} X_1^w &= W_1 \otimes X \\ X_2^w &= W_2 \otimes X \end{aligned} \quad (5)$$

式中: X_1^w 是具有较大信息量和表达性空间内容的特征; X_2^w 是信息量较小的冗余特征,在采用交叉重构运算,将

加权后的特征 X_1^w 和 X_2^w 进行拼接,得到空间精细特征映射 X^w ,过程表示如式(6)所示。

$$\begin{aligned} X_{11}^w \oplus X_{22}^w &= X^{w1} \\ X_{21}^w \oplus X_{12}^w &= X^{w2} \\ X^{w1} \cup X^{w2} &= X^w \end{aligned} \quad (6)$$

式中: $\otimes$ 是逐元素乘法; $\oplus$ 是逐元素求和; $\cup$ 是串联。将 SRU 应用于中间输入特征 X 后,抑制空间维度的冗余特征,CRU 则抑制通道冗余特征。其中,转换操作将 X_{up} 作为“富特征提取”的输入,分别进行 GWC 和 PWC,再相加得到输出 Y_1 ;将 X_{low} 作为“富特征提取”的补充进行 PWC,并与原输入取并集得到 Y_2 。融合操作采用简化的 SKNet 方法合并 Y_1 和 Y_2 ,再通过全局平均池化结合全局空间和通道信息,得到池化后的 $S1$ 和 $S2$ 应用 Softmax 操作,计算特征权重向量 β_1 和 β_2 。最后,利用 β_1 和 β_2 计算的输出 Y 如式(7)所示。

$$Y = \beta_1 \times Y_1 + \beta_2 \times Y_2 \quad (7)$$

输出特征图 Y 包含经过空间和通道重构增强的特征信息。

2.2 ESPPM 模块

1) 大可分离核残差注意力机制 LSKR

大可分离核注意力机制 LSKA^[18]是通过分离大核卷积操作实现高效特征捕捉和长程依赖关系建模的技术,将传统卷积分解为小核卷积的同时,利用大核卷积在捕捉全局特征方面的优势,使其对序列数据处理有出色表现。

LSKA 模块和 LSKR 模块如图4所示,其中,LSKR 模块通过大核分离技术和残差链接相结合,缓解长序列中关键信息消失的问题。

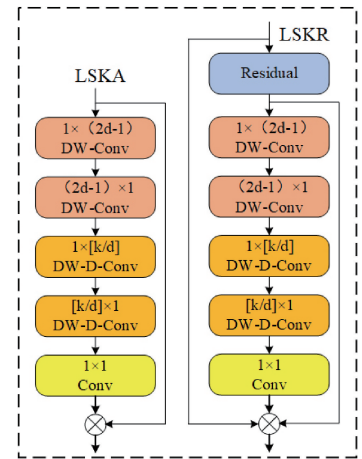


图4 两种模块对比

Fig. 4 Comparison of the two modules

尽管 LSKA 在捕捉长程依赖关系方面表现出色,但在处理复杂场景时,其特征表示能力仍有提升空间,而残

差链接已被证明在深度神经网络中能有效缓解梯度消失问题。基于此,构建大可分离核残差注意力机制 LSKR,在 LSKA 的基础上引入残差结构^[19],使原始输入和输出之间保留一定联系,解决网络模型在处理长序列数据的局限性。

2) ESPPM 模块

空间金字塔快速池化模块(SPPF)是一种通过在特征图上并行多尺度最大池化操作(例如 1×1 、 3×3 、 5×5 和 7×7 池化窗口)捕捉多尺度空间信息的技术,通过并行池化操作产生的特征图在池化后进行融合,生成不同尺度信息的特征图。

针对遥感图像高分辨率、复杂背景纹理和多尺度目标特征等特点,本文在 SPPF 中引入 LSKR,并将其与自适应平均池化(adaptive avg pool)操作结合,构建 ESPPM 模块,结构如图 2 所示。ESPPM 利用 LSKR 长序列建模的优势,通过分离大核卷积操作,保持高效特征捕捉。其中,大核卷积具有较大的感受野,能够捕获特征图中的长程依赖关系和全局信息,处理遥感图像中的大范围背景纹理和多尺度目标;小核卷积则能够高效地捕捉局部特征,减少计算复杂度,通过将大核卷积分解为小核卷积,使 LSKR 能够在保持高精度的同时,提高计算效率,残差结构则用于增强模型在处理长序列数据和多尺度特征聚合时的稳定性;自适应平均池化操作能够在不同尺度上对特征图进行池化,生成不同尺度的信息,确保特征图在不同尺度上的信息得到充分表示。

在 ESPPM 模块中,自适应平均池化操作生成的多尺度特征图与 LSKR 模块生成的长程依赖关系特征图进行融合,该融合增强了特征图的多尺度特征聚合能力,并通过 LSKR 的长程依赖关系建模,确保了不同尺度特征之间有效的信息交互。

2.3 Slim_PAN 结构

传统目标检测模型在处理高分辨率、多尺度目标等复杂背景的遥感图像时需要较大的计算资源和参数量。为解决此问题,在颈部网络引入轻量化的 Slim-PAN^[20]结构。

Slim-PAN 结构的实现主要包括 GSConv^[20]轻量化卷积、VoV-GSCSP^[20]注意力机制和高效特征融合策略 3 个关键部分。GSConv 的计算成本约为标准卷积的 $1/2$,但对模型学习能力的贡献相当。因此,在颈部网络中用 GSConv 替换部分卷积模块。

如图 5 所示,将 Bottleneck_GS 轻量化结构替换颈部特征提取模块中的 Bottleneck 结构,并通过一次性聚合构建跨级部分网络模块 VoV-GSCSP。结合 GSConv 和 VoV-GSCSP 结构,Slim-PAN 在降低模型参数和计算量的同时,实现高效特征融合策略。

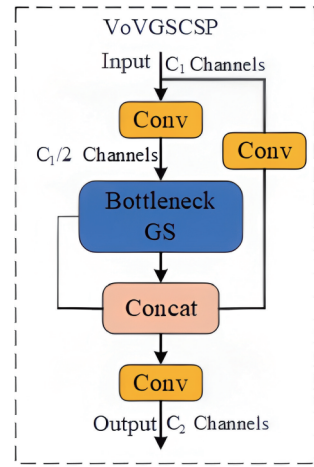


图 5 VoV-GSCSP 结构

Fig. 5 VoV-GSCSP structure diagram

2.4 优化损失函数

遥感图像中的目标(如建筑物、车辆等)通常具有不规则形状,并且可能以任意角度分布。传统的矩形边界框(bounding box)无法紧密地贴合这些目标,并且原模型使用标准的边界框参数 (x, y, w, h) 来表示目标的位置和大小,这在处理常规图像中的目标检测任务时是有效的,但对于遥感图像,这种表示方法不足以准确描述复杂的目标形状和方位,导致原 YOLOv8n 模型对遥感图像进行目标检测时存在局限性。

为解决目标方向任意带来的检测局限性,本文在原 YOLOv8n 模型边界框参数 (x, y, w, h) 的基础上,增加角度参数 θ (图 2 中 Detect 结构)确定边界框方位,并且采用目前主流的五参数表示法,通过结合 RBCL 损失函数,进一步提升检测精度。

1) 旋转框参数表示法

五参数旋转框表示法引入了旋转角度信息 θ ,边界框参数可由 5 维向量 (x, y, w, h, θ) 确定。其中, x, y 表示目标最小外接矩形的中心坐标; w 和 h 分别为外接矩形的宽和高; θ 为外接矩形相对于水平轴 x 的旋转角度,正值表示逆时针旋转,负值表示顺时针旋转。

图 6(a)所示为旋转框理想状态,五参数表示法可分为两种分别如图 6(b)和(c)所示。图 6(b)为 OpenCV 定义法,其中五参数表示法具有独特的周期性,角度 θ 的取值范围为 $[-\pi/2, 0)$ 。其 w 和 h 总是满足 $w \geq h$,在接近 0 度或 180 度时,预测值和真实值的小误差可能会导致较大的回归损失,直接计算损失会导致角度不连续性。图 6(c)为长边定义法,其中五参数表示法的角度周期性取决于长边方向。角度 θ 的取值范围为 $[-\pi, 0)$,长边定义法的特点是无论外接矩形是水平、垂直或倾斜放置, w 始终代表外接矩形的长边,用固定比例或尺度进行回归

会导致模型在不同尺度的目标上表现不一致,甚至在小目标上完全失效,直接计算损失会导致参数模糊性问题。

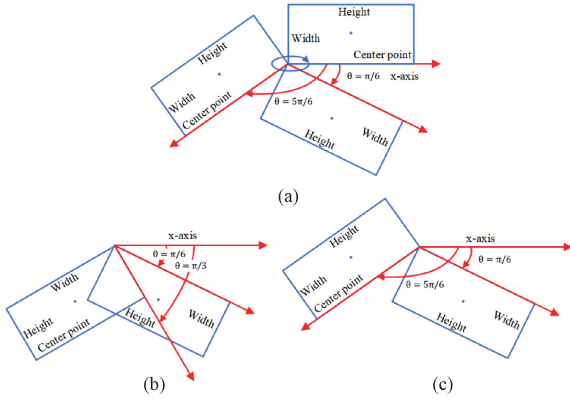


图6 旋转框表示法存在的问题

Fig. 6 Problems with the rotation box representation

2) RBCL 损失函数

YOLOv8n 采用损失函数为 $\text{CIoU}^{[21]}$, 损失函数的公式如式(8)~(11)所示。

$$\text{CIoU} = \text{IoU} - \left(\frac{\rho^2(b, b^{gt})}{c^2} \right) + \alpha v \quad (8)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (9)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (10)$$

$$\text{Loss}_{\text{CIoU}} = 1 - \text{IoU} + \left(\frac{\rho^2(b, b^{gt})}{c^2} \right) + \alpha v \quad (11)$$

式中: IoU 表示预测框和真实框的交并比; $\rho^2(b, b^{gt})$ 表示预测框中心点 b 和真实框中心点 b^{gt} 之间的欧氏距离; c 为包围预测框和真实框最小闭包区域的对角线长度; $\rho^2(b, b^{gt})/c^2$ 为距离惩罚项; v 衡量预测框和真实框的长宽比一致性; α 为宽高比损失系数; w^{gt} 和 h^{gt} 为真实框的宽高; w 和 h 为预测框的宽高。

虽然 CIoU 综合考虑了预测框与真实框的重叠面积、中心距离及长宽比,但在计算两个边界框之间的差异时,模型主要关注目标的重叠区域、中心点距离和宽高比等因素,忽略了预测框与真实框的角度差异,即目标可能存在的旋转角度,对形状和旋转信息造成丢失。

为解决上述问题,提出 RIoU 损失函数。在 CIoU 基础上引入角度周期性和边长一致性惩罚项,解决参数表示法的问题以及旋转信息的丢失,改进损失函数公式如式(12)所示。

$$\text{Loss}_{\text{RIoU}} = 1 - \text{IoU} + \left(\frac{\rho^2(b, b^{gt})}{c^2} \right) + \alpha v + \lambda \theta + f l \quad (12)$$

式中: λ 和 f 为权重系数,调节角度周期性和边长一致性

损失对总损失的影响; θ 为预测框角度与真实框角度之差, l 为预测边界框长度与真实边界框长度之差,公式如式(13)、(14)所示。

$$\theta = \sin^2 \left(\frac{\theta_{\text{pred}} - \theta_{\text{true}}}{2} \right) \quad (13)$$

$$l = (l_{\text{pred}} - l_{\text{true}})^2 + (s_{\text{pred}} - s_{\text{true}})^2 \quad (14)$$

式中: θ_{pred} 为预测角度值; θ_{true} 为真实的角度值;通过引入 $\sin^2$ 函数,可以使该损失函数具有周期性,以解决角度不连续性问题。 $\sin^2$ 函数确保了两种情况。1) 当 $\theta_{\text{pred}} = \theta_{\text{true}}$ 时,损失为 0,表示预测完全准确时没有惩罚;2) $\sin$ 函数在 $[0^\circ, 180^\circ]$ 是单调的,而 $\sin^2$ 则确保损失总是非负,即使角度相隔 180° 也不超出其周期性范围 $[0, 1]$ 。因此,对于 θ_{pred} 和 θ_{true} 之间的任何差异,即使两者相距 180° ,损失也不会无限增大,使该惩罚项有效处理预测角度与真实角度间的周期性不连续问题,确保模型在学习过程中正确处理角度周期性。

l_{pred} 、 l_{true} 分别为预测和真实边框长边长度, s_{pred} 、 s_{true} 分别为预测和真实边框短边长度, $(l_{\text{pred}} - l_{\text{true}})^2$ 、 $(s_{\text{pred}} - s_{\text{true}})^2$ 分别为预测和真实长边长度、预测和真实短边长度之间的平方误差损失,通过惩罚预测与真实值在长边和短边两个维度上的偏差,确保预测边框的尺寸与真实目标尽可能接近。该惩罚项确保了尺寸的一致性,尤其是当目标形变较大时,通过分别对短边和长边计算损失,确保模型在预测时考虑到目标的长宽比,而不仅仅是总尺寸;该惩罚项对长边和短边计算损失,而不是对宽度和高度或者对边长进行直接计算,可以帮助模型减少参数间的耦合和依赖,解决了参数模糊性问题。

RBCL 综合考虑旋转角度信息和边长参数信息,解决旋转框表示法角度不连续和参数模糊性以及 CIoU 形状和旋转信息造成丢失的问题,分类损失为交叉熵损失;置信度损失为二元交叉熵损失;边界框和角度回归损失均为 L1 损失:

$$L_{\text{cls}} = - \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (15)$$

$$L_{\text{bbox}} = \sum_{i=1}^M \text{SmoothL1}(\hat{b}_i, b_i) \quad (16)$$

$$L_{\text{angle}} = \sum_{i=1}^M \text{SmoothL1}(\hat{\theta}_i, \theta_i) \quad (17)$$

式中: y_i 是第 i 个预测框的真实类别标签; $\hat{y}_i$ 是第 i 个预测框的预测类别概率; $\hat{b}_i$ 是第 i 个预测框的预测边界框参数; b_i 是第 i 个预测框的真实边界框参数; $\hat{\theta}_i$ 是第 i 个预测框的预测角度; θ_i 是第 i 个预测框的真实角度。

结合以上各项损失,得到总损失函数 RBCL 如式(18)所示。

$$\text{Loss}_{\text{RBCL}} = \text{Loss}_{\text{RIoU}} + \text{Loss}_{\text{bbox}} + \text{Loss}_{\text{cls}} + \text{Loss}_{\text{angle}} \quad (18)$$

3 实验结果与数据分析

3.1 实验环境及参数设置

为评估改进算法在遥感图像目标检测的表现,本实验在 Ubuntu 20.04 系统上进行操作,采用深度学习框架为 Pytorch 1.11.0,Python 版本为 3.8,CUDA 版本为 11.3。模型训练采用 GPU 为 NVIDIA Geforce RTX 3070ti。实验训练关键参数如表 1 所示。

表 1 训练模型关键参数

Table 1 Key parameters of the training model	
参数	设置值
优化器	SGD
输入图像大小	640×640
学习率	0.001
epochs	200
批处理次数	16
Mosaic	10

3.2 数据集

本文实验采用公开遥感图像 DOTA 数据集,该数据集包含 2 806 张高分辨率遥感图像和 188 282 个实例,涵盖飞机、船舶、储油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、大型车辆、小型车辆、直升机、环形道、足球场、游泳池 15 个不同目标,图像大小为 800×800~4 000×4 000。

针对原始图像尺寸不一、难以直接输入到网络中进行训练的问题,本文对原始图像进行切割处理,最终获得 21 046 张分辨率为 1 024×1 024 的子图像。将最终得到的图像按照训练集、验证集、测试集为 7:2:1 进行随机分配。其中训练集为 14 732 张图像,验证集为 4 208 张图像,测试集为 2 106 张图像。

3.3 评价指标

本文采用准确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mAP)、浮点运算次数 (GFLOPs) 来评估模型性能。准确率、召回率、平均精度均值的公式如式 (19)~(21) 所示。

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

(19)

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

(20)

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i$$

(21)

式中:准确率衡量所有被预测为正类的样本中,实际为正类的比例;TP 指预测为正类且真实情况也是正类的样本数;FP 则指预测为正类但实际为负类的样本数。召回率衡量在所有实际为正类的样本中,被正确识别为正类

的比例;FN 为预测为负类但实际为正类的样本数。平均精度均值是所有类别平均精度的平均值;AP_i 表示第 i 类的平均精度值;k 为所有类别总数。

3.4 引入 ESPPM 模块的对比实验

为探究在 SPPF 模块中引入不同注意力机制的效果,使用性能优异的注意力机制 (如 SE、CA、ECA、LSKA),并将其分别与在 SPPF 相同位置引入的 LSKA 和 LSKR 进行对比。此外,将引入自适应平均池化层和 LSKR 结合的 SPPF 与分别引入 LSKA 和 LSKR 的 SPPF 进行对比,实验结果如表 2 所示。

表 2 引入不同注意力机制的 SPPF 模块对比实验

Table 2 Comparative experiments of ESPPM modules with different attention mechanisms			
方法	mAP@0.5	计算量/GFLOPs	Params/(×10 ⁶)
无	0.706	8.3	3.15
AdaptiveAvg	0.726	8.3	3.15
CA ^[22]	0.730	8.4	3.32
ECA ^[23]	0.741	8.5	3.23
SE ^[24]	0.734	8.3	3.32
LSKA	0.743	8.6	3.45
LSKR	0.745	8.6	3.45
本文	0.747	8.7	3.53

由表 2 可见,在 SPPF 模块相同位置引入注意力机制的 SE、CA、ECA、LSKA 普遍提高了模型的检测性能,其中 LSKA 注意力机制表现最佳,mAP@0.5 较基线提升 3.7%,引入自适应平均池化层后,mAP@0.5 较基线提升 2.0%,展示了其多尺度特征聚合和细节保留方面的优势;与仅引入 LSKA 相比,LSKR 在 mAP@0.5 较基线提升 3.9%,展示其引入残差链接后性能的改进。

ESPPM 模块是在引入 LSKR 的基础上嵌入自适应平均池化层,实验数据显示,ESPPM 与分别引入 LSKR 和 LSKA 相比,尽管计算量和参数量略微增加,但 ESPPM 在 mAP@0.5 上分别提升 0.2% 和 0.4%,相较基线则实现 4.1% 的提升,验证了自适应平均池化层与 LSKR 结合后增强了模型的多尺度特征聚合能力,证明本文设计模块的有效性。

3.5 消融实验

为系统性解析各改进模块对 YOLOv8n 整体性能的贡献,本文在 DOTA 数据集进行消融实验。通过设计 9 组实验,验证所提改进策略的有效性,实验结果如表 3 所示。

前 4 组实验为在 YOLOv8n 模型添加角度参数并引入 L1 角度回归的基础上,分别引入 SC_C2f 模块和 ESPPM 模块,以及同时引入两者,观察算法性能变化。后 4 组实验是在 YOLOv8n 添加角度参数基础上引入

RBCL 损失函数,在此基础上分别引入 SC_C2f 模块和 ESPPM 模块,以及同时引入两者,观察算法性能变化,最后一组在第八组实验基础上引入 Slim_PAN 结构,验证其可行性。

表 3 消融实验

Table 3 Ablation experiments									
组别	SC_C2f	ESPPM	Slim_PAN	RBCL	P	R	mAP@0.5	计算量/GFLOPs	Params/($\times 10^6$)
1	×	×	×	×	0.785	0.668	0.706	8.3	3.08
2	✓	×	×	×	0.777	0.716	0.745	8.2	2.96
3	×	✓	×	×	0.778	0.713	0.747	8.7	3.53
4	✓	✓	×	×	0.797	0.695	0.749	8.6	3.42
5	×	×	×	✓	0.801	0.688	0.741	8.3	3.08
6	✓	×	×	✓	0.779	0.713	0.750	8.2	2.96
7	×	✓	×	✓	0.797	0.703	0.749	8.7	3.53
8	✓	✓	×	✓	0.786	0.712	0.755	8.6	3.42
9	✓	✓	✓	✓	0.805	0.698	0.757	7.9	3.04

通过表 3 可知,各改进部分均有效果。在第 2 组实验中,将 SC_C2f 模块引入主干网络,计算量和参数量分别减少 0.1 GFLOPs 和 0.12×10^6 。第 3 组实验引入 ESPPM 模块,mAP50 和 R 分别增加 4.1% 和 4.5%。相较于第 2 和第 3 组实验,同时引入 SC_C2f 和 ESPPM 模块,mAP@0.5 分别提高 0.4% 和 0.2%。

第 5 组到第 8 组实验在结合 RBCL 损失函数的基础上引入改进模块。第 5 组实验,RBCL 损失函数使 P 和 mAP@0.5 分别提升 1.6% 和 3.5%,第 6 组实验引入 SC_C2f 模块, P 和 mAP@0.5 分别提升 0.2% 和 0.5%。第 7 组实验引入 ESPPM 模块,mAP@0.5 增加 0.2%;相较于第

6 和第 7 组,第 8 组实验同时引入 SC_C2f 和 ESPPM 模块,mAP@0.5 分别提升 0.5% 和 0.6%, P 分别提高 0.7% 和 0.8%;第 9 组实验在第 8 组基础上引入 Slim_PAN 结构,较第 8 组实验,mAP@0.5 和 P 分别提升 0.2% 和 1.9%,计算量和参数量分别降低 0.7 GFLOPs 和 0.38×10^6 。最终改进整体相较于基线 P 、 R 和 mAP@0.5 分别提升 2.0%、3.0% 和 5.1%,计算量减少 0.4 GFLOPs。

3.6 对比实验

为了进一步验证改进算法的有效性,将改进后的算法在处理后的 DOTA 数据集进行实验验证并进行对比分析,结果如表 4 所示。

表 4 不同模型在 DOTA 数据集的对比实验

Table 4 Comparative experiments of different models on DOTA datasets				
模型	mAP@0.5	计算量/GFLOPs	Params/($\times 10^6$)	帧率/fps
Faster-RCNN	0.630	206.7	41.2	15
Mask-RCNN	0.653	258.5	43.8	13
SSD	0.595	115.0	26.5	60
TOOD ^[25]	0.672	150.4	45.0	40
YOLOv5n	0.694	21.7	9.12	125
YOLOv7-tiny	0.703	6.5	6.01	149
YOLOv8n	0.706	8.3	3.08	164
YOLOv8-LGA ^[26]	0.736	8.5	3.21	134
FEW-YOLOv8 ^[27]	0.693	29.5	11.19	98
本文	0.757	7.9	3.04	155

实验结果表明,相较于典型的双阶段目标检测算法 Faster-RCNN 和 Mask-RCNN,SSD 在计算量和参数量方面表现出优势,但其检测精度稍有不足,本文改进算法在检测精度、计算量和参数量的综合方面表现出优势。

典型的单阶段目标检测算法 TOOD 和 SSD 在检测精度、计算量和参数量方面各有优势,其中 TOOD 较 SSD 精度更高达到 67.2%,但 SSD 计算量和参数量较 TOOD 更低,两者与本文改进算法相比,整体性能仍存在一定差距。虽然 YOLOv5n 精度仅与 YOLOv8n 相差 1.2%,但其

计算量和参数量分别高 13.4 GFLOPs 和 6.04×10^6 ,不足以弥补在精度上的劣势,证明本文所选基线算法的优越性。YOLOv7-tiny 在计算量方面比 YOLOv8n 低,但其检测精度相较 YOLOv8 低 0.3%,参数量多 3×10^6 ,并且其帧率要比本文改进后的算法还要低 6 fps,相较于原算法则低 15 fps,证明其检测速度的不足,即使其计算量较低也不足以弥补差距。

YOLOv8-LGA 和 FEW-YOLOv8 与改进后的算法在同一 DOTA 数据集上进行对比实验。实验数据表明,

YOLOv8-LGA 的 mAP@ 0.5 比改进后算法低 2.1%, FEW-YOLOv8 的 mAP@ 0.5 低 6.4%。此外, YOLOv8-LGA 和 FEW-YOLOv8 的计算量和参数量均高于改进后的算法, YOLOv8-LGA 的帧率比改进后算法低 21 fps, FEW-YOLOv8 的帧率低 57 fps。基于以上实验结果, 本文算法提出的改进策略在多个维度均体现出优势, SC_C2F 模块有效提升了模型的特征提取能力, ESPPM 模块则强化模型的多尺度特征聚合效果, Slim_PAN 模块在降低参数量的同时保持了检测精度, 而 RBCL 损失函数则增强了目标定位精度。

综上所述, 与原算法及各种目标检测模型相比, 本文改进算法在各项性能指标方面表现最优, 在保持 155 fps 实时检测速度和 3.04×10^6 模型参数量的同时, mAP@ 0.5 达到 75.7%。

3.7 实验结果与可视化分析

为直观分析改进算法的效果, 实验部分的检测对比结果如图 7(a)~(f) 所示。图 7A~D 分别为近岸多尺度目标、停机场多种类目标、停车场多尺度目标和复杂背景

的多尺度目标。

本文对主干网络、颈部网络及损失函数进行改进: 构建 SC_C2F 模块、ESPPM 模块、RBCL 损失函数和引入 Slim_PAN 结构来提高模型的检测精度。由图 7(c) 可见, 尽管 SC_C2F 模块提高了模型的特征提取能力, 使得误检和漏检较原算法有所减少, 但其减少的计算冗余导致其在定位精度处仍有不足;

由图 7(d) 可见, ESPPM 模块因其出色的长序列处理能力以及全局和局部信息的关注, 使得其在处理遥感小目标和复杂背景时检测能力较原算法有所提升, 但在多尺度目标处理上仍有误检和漏检。由图 7(e) 可见, RBCL 损失函数通过加入惩罚项规范模型检测, 进一步提升检测精度。

由图 7(f) 可见, 相比原算法, 改进后的算法在复杂背景和多尺度目标检测方面表现优异, 有效减少了密集目标以及小目标的误检和漏检现象, 整体检测效果提升明显。

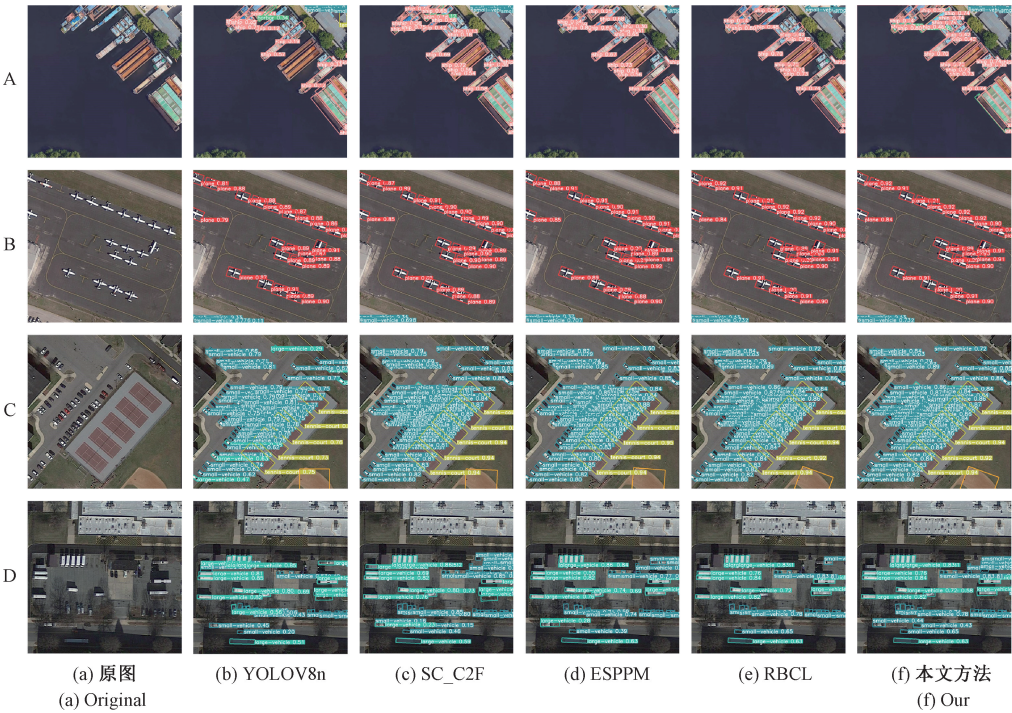


图 7 改进模块检测对比结果

Fig. 7 Comparison results of improved module detection

4 结 论

针对遥感图像中目标尺度差异大、密度不均及类别多样性等问题, 提出了一种改进 YOLOv8n 的遥感目标检测算法。通过构建 SC_C2F 和 ESPPM 模块并引入主干

网络, 减少参数冗余, 丰富上下文信息, 增强模型多尺度和跨通道特征提取能力。颈部网络采用 Slim-PAN 结构, 降低计算量的同时提高检测精度。最后, 结合 RBCL 损失函数, 解决目标方向任意的回归问题。改进后的算法在 DOTA 数据集上进行实验, mAP@ 0.5 较 YOLOv8n 提升 5.1%, 计算量减少 0.4 GFLOPs, 说明改进后的算法对

提高遥感目标整体检测效果的有效性,在减少漏检和误检的同时,降低模型的计算复杂度。但本文模型在轻量化方面仍然需改进,后续可以通过剪枝、量化等技术构建更加轻量化的模型,以适应资源受限的设备或环境。对于需要实时处理的应用场景,优化模型模块以满足低延迟需求,确保在保持高精度的同时能够快速响应。

参考文献

- [1] ZOU Z, CHEN K, SHI Z, et al. Object detection in 20 years: A survey [J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [3] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [6] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [7] QIAO S, CHEN L C, YUILLE A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10213-10224.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [10] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.
- [11] CHEN P, WANG J, ZHANG Z, et al. CSPGNet: Cross-scale spatial perception guided network for tiny object detection in remote sensing images [J]. Digital Signal Processing, 2024: 154104674-104674.
- [12] 唐建宇, 唐春晖. 基于旋转框和注意力机制的遥感图像目标检测算法 [J]. 电子测量技术, 2021, 44(13): 114-120.
TANG J Y, TANG CH H. Object detection algorithm for remote sensing image based on rotating box and attention mechanism [J]. Electronic Measurement Technology, 2021, 44(13): 114-120.
- [13] 闫钧华, 张琨, 施天俊, 等. 融合多层次特征的遥感图像地面弱小目标检测 [J]. 仪器仪表学报, 2022, 43(3): 221-229.
YAN J H, ZHANG K, SHI T J, et al. Ground weak target detection in remote sensing images fused with multi-level features [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.
- [14] 雷帮军, 朱涵. 基于上下文空间感知的遥感图像旋转目标检测 [J]. 光电与控制, 2025, 32(3): 69-75.
LEI B J, ZHU H. Rotating target detection in remote sensing images based on contextual spatial perception [J]. Electro-Optics & Control, 2025, 32(3): 69-75.
- [15] 马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标检测 [J]. 光电工程, 2022, 49(4): 49-65.
MA L, GOU Y T, LEI Y, et al. Small target detection in remote sensing images based on multi-scale feature fusion [J]. Optoelectronic Engineering, 2022, 49(4): 49-65.
- [16] 朱圣博, 魏利胜, 高港, 等. 基于改进 YOLOv8s 的光学遥感小型船舶检测算法 [J]. 电子测量与仪器学报, 2024, 38(10): 48-57.
ZHU SH B, WEI L SH, GAO G, et al. Optical remote sensing small ship detection algorithm based on improved YOLOv8s [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 48-57.
- [17] LI J, WEN Y, HE L. Sconv: Spatial and channel reconstruction convolution for feature redundancy [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.
- [18] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.
- [19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [20] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A

- better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv preprint arXiv: 2206.02424, 2022.
- [21] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [23] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [25] FENG C, ZHONG Y, GAO Y, et al. Tood: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2021: 3490-3499.
- [26] 刘勇, 吕丰顺, 李学琨, 等. 基于 YOLOv8-LGA 的遥感图像目标检测算法 [J/OL]. 光电子·激光, 1-12 [2025-02-25]. <http://kns.cnki.net/kcms/detail/12.1182.04.20240709.1518.016.html>.
LIU Y, LYU F SH, LI X K, et al. Object detection algorithm for remote sensing image based on YOLOv8-LGA [J/OL]. Optoelectronics & Laser, 1-12 [2025-02-25]. <http://kns.cnki.net/kcms/detail/12.1182.04.20240709.1518.016.html>.
- [27] 席阳丽, 屈丹, 王芳芳, 等. 基于 FEW-YOLOv8 遥感图像目标检测算法 [J/OL]. 郑州大学学报(工学版), 1-8 [2025-02-25]. <https://doi.org/10.13705/j.issn.1671-6833.2025.04.007>.
XI Y L, QU D, WANG F F, et al. Target detection algorithm

based on FEW-YOLOv8 remote sensing image [J/OL]. Journal of Zhengzhou University (Engineering Science), 1-8 [2025-02-25]. <https://doi.org/10.13705/j.issn.1671-6833.2025.04.007>.

作者简介



王海群, 1996 年于河北科技大学获得学士学位, 2003 年于河北工业大学获得硕士学位, 现为华北理工大学副教授, 主要研究方向为智能控制与应用领域。

E-mail: wanghq0604@yeah.net

Wang Haiqun received her B. Sc. degree from Hebei University of Science and Technology in 1996 and M. Sc. degree from Hebei University of Technology in 2003. She is now an associate professor at North China University of Science and Technology. Her main research interests include intelligent control and application.



武泽锴 (通信作者), 2023 年于河北建筑工程学院获得学士学位, 现为华北理工大学硕士研究生, 主要研究方向为图像处理与模式识别。

E-mail: 1270518284@qq.com

Wu Zekai (Corresponding author) received his B. Sc. degree from Hebei Institute of Civil Engineering and Architecture in 2023. He is now a M. Sc. candidate at North China University of Science and Technology. His main research interests include image processing and pattern recognition.



晁帅, 现为华北理工大学博士研究生, 招生就业处工作, 主要研究方向为目标检测。

E-mail: 1920807580@qq.com

Chao Shuai now is a Ph. D. candidate in North China University of Science and Technology. He is working in the Admissions and Employment department of North China University of Science and Technology. His main research interest includes target detection.