

俞成海, 谭勇, 叶泽支, 等. 基于改进 YOLOv5 的铁道辙叉损伤检测研究[J]. 智能计算机与应用, 2025, 15(3): 207-212.
DOI: 10.20169/j.issn.2095-2163.250331

基于改进 YOLOv5 的铁道辙叉损伤检测研究

俞成海, 谭勇, 叶泽支, 卢智龙

(浙江理工大学 计算机科学与技术学院(人工智能学院), 杭州 310018)

摘要: 铁道辙叉表面损伤如裂纹和掉块会对铁路运输安全造成严重威胁。虽然现有的深度学习模型可有效检测铁道表面损伤,但对结构更复杂的辙叉损伤检测效果欠佳。为此,本文构建了一个包含不同天气环境的铁道辙叉损伤数据集,并基于 YOLOv5 提出了一种新的 YOLOv5-CDC 网络模型。模型在主干网络中引入了通道注意力机制,提高模型对辙叉缺陷的定位能力;在 Neck 模块中采用了改进的密集连接结构,增强了不同抽象层特征的融合;此外,在 Neck 和检测头间加入了通道块注意力模块,降低了复杂光照对检测性能的不利影响。实验结果表明,本文提出的改进模型相较原 YOLOv5 模型在铁道辙叉损伤检测任务上平均精度提升 3.4%。经实验证明,本文方法可以有效提高铁道辙叉缺陷的检测性能,具有一定的实用价值。

关键词: 铁道辙叉; 缺陷检测; YOLOv5; 注意力机制

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2025)03-0207-06

Research on damage detection of railway crossing based on improved YOLOv5

YU Chenghai, TAN Yong, YE Zezhi, LU Zhilong

(School of Computer Science and Technology(School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Surface defects on rail frogs such as cracks and spalling pose a serious threat to railway transportation safety. Although existing deep learning models are effective in detecting rail surface damage, their performance on the more complex structure of frog damage detection is poor. To address this problem, this paper constructs a rail frog damage dataset containing different weather environments, and proposes a new YOLOv5-CDC network model based on YOLOv5. The model introduces channel attention mechanisms in the backbone network to improve the model's localization ability for frog defects. An improved dense connection structure is adopted in the Neck module to enhance the fusion of features from different abstraction levels. In addition, channel-spatial attention modules are added between the Neck and detection head to reduce the adverse effects of complex lighting on detection performance. The experimental results show that the improved model proposed in this paper achieves an average precision improvement of 3.4% over the original YOLOv5 model on rail frog damage detection. It is validated that the proposed method can effectively improve the detection performance on rail frog defects and holds certain practical value.

Key words: railway frog; defect detection; YOLOv5; attention mechanism

0 引言

随着高铁网络的不断扩张,铁路运输仍是国内的主要交通方式之一。铁道辙叉是实现列车线路转换的关键设备,由于辙叉各部件为铸造产品,结构又较为复杂,不可避免地会存在铸造缺陷,长期承载负荷会导致裂纹和脱落。辙叉损伤将严重危及行车安全,必须及时检测和更换。目前,国内对辙叉的检查主要依赖人工巡检,效率低下,误检率高,难以满足检测需求。因此,开发辙叉表面损伤检测技术具有

重要意义。

随着深度学习的进展,目标检测技术已广泛应用于多种领域。将其应用于辙叉损伤检测,可以有效降低成本、提高效率,减少不必要损失。当前主流检测算法分为 2 类,分别是:基于回归的单阶段检测,如 YOLO^[1]、SSD^[2]等直接预测目标信息;基于区域建议的两阶段检测,如 RCNN^[3]系列,就是先进行区域建议,生成多个候选区域,再进行分类回归,输出检测结果。针对铁轨表面损伤检测任务,已有研究者提出多种算法以增强检测性能。例如,罗晖

作者简介: 谭勇(1999—),男,硕士研究生,主要研究方向:目标识别。

通信作者: 俞成海(1975—),男,副教授,主要研究方向:移动应用,人工智能。Email: ych@zstu.edu.cn。

收稿日期: 2023-08-15

等学者^[4]采用 Gabor 滤波器进行图像去噪,以降低噪声对缺陷检测的干扰。韩强等学者^[5]设计了基于多尺度特征融合的检测网络,提高了对小面积钢轨损伤区域的检测精度。刘浩翰等学者^[6]面对与背景高度相似的工业部件缺陷检测问题时,在 YOLOv3 基础上使用注意力机制,获得更高的召回率。尽管相关文献对轨道损伤检测进行了大量研究并取得显著成效^[7],但对于结构更为复杂的轨道辙叉,现有方法的检测精度仍然欠佳。针对目前问题,本文在 YOLOv5 模型基础上进行改进;首先,在主干网络加入 CA 注意力机制,增强特征表达;其次,Neck 模块采用改进的密集连接,增强特征融合;最后,在检测头前加入 CBAM 模块。

1 数据集采集与制作

由于目前现有数据集无法满足本文实验要求,因此通过某铁路公司辙叉巡检人员采集数据,数据

采集时间段为 08:00~22:00,格式为 JPEG,分辨率最小为 1 200 像素×1 600 像素,最大为 3 000 像素×4 000 像素。拍摄方式为拍摄者俯视辙叉,相距辙叉 0.5~1.5 m 拍摄。

为保证数据样本多样性,经过数据筛选,所选图像拍摄时间包含清晨、中午、傍晚、夜晚,天气环境包括晴、多云、雨等,光线条件包括顺光和逆光,拍摄角度包括 90°俯视,45°俯视。

使用 LabelImg 对数据集中的图像进行了标注。为满足深度学习对大量训练样本需求,增加模型泛化能力,本文采用随机缩放、翻转、亮度增强、对比度增强、颜色增强的方式进行数据增强,每种增强方式增强概率 50%,至少采用一种增强。数据集图像增加到 1 145 张,数据集有以下几个特点:图像背景复杂,包含各种天气环境;损伤尺度变化大,同一图像上的不同损伤尺寸差距较大;包含多角度图片。数据集部分图片如图 1 所示。



图 1 数据集部分图片

Fig. 1 Part of the data set

其中,训练集、验证集、测试集的比例为 8 : 1 : 1 (训练集 915 张、验证集 115 张、测试集 115 张)。

2 总体网络设计

本文以 YOLOv5 模型为基础,结合 C3_CA 模块、CBAM 注意力机制和改进 Dense 模块构建了一种新的

铁道辙叉损伤检测模型 YOLOv5-CDC,模型结构如图 2 所示。输入端部分包括 Mosaic 数据增强、使用 K-means 聚类算法进行锚框计算、自适应图片缩放,其中, Mosaic 数据增强是随机使用 12 张图片,随机缩放,再随机分布进行拼接,丰富检测数据集,可以有效地解决单张图片中目标较少的问题,使网络的鲁棒性更好。

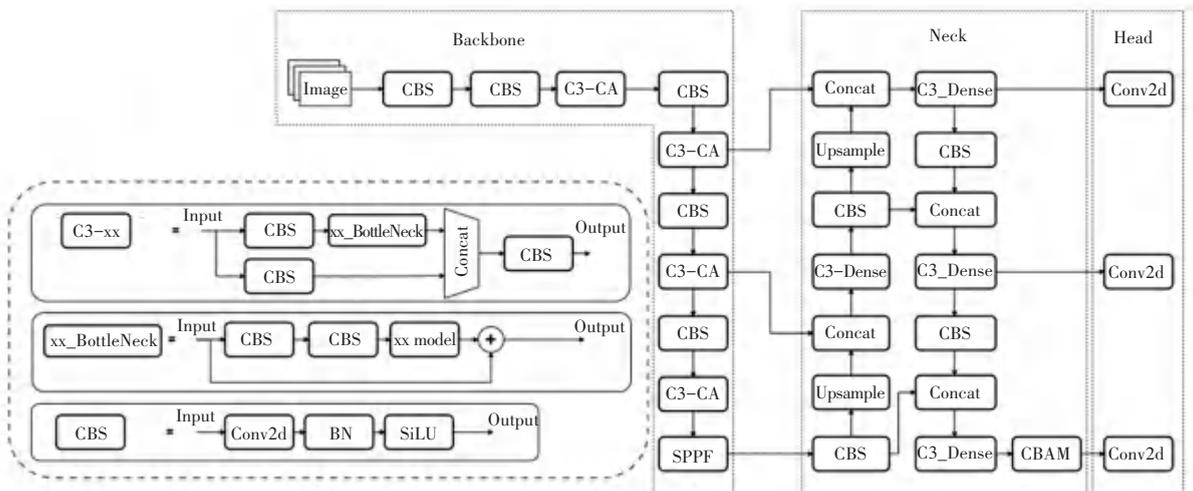


图 2 YOLOv5-CDC 网络图

Fig. 2 YOLOv5-CDC network diagram

在训练前,采用K-means聚类算法对训练集自适应锚框计算,能够加快网络收敛速度,提高训练效率。自适应图片缩放是将原始图片统一缩放到一个标准尺寸,再送入检测网络中。在Backbone模块中,本文将C3结构融合CA注意力机制提出了一种新型的通道注意力结构C3_CA。该结构首先将基础层的特征图分为2部分,其中一部分经过CA模块处理,然后通过跨层级连接,从而增强了Backbone的特征表达能力。在Neck模块中,本文改进了原有的C3模块,加入了Dense连接,以提高不同抽象层级特征之间的融合。在Neck模块和检测头之间,本文插入了通道注意力机制(CBAM)结构,显著增强了Neck层特征图的表达能力,并减少了漏检现象。

3 改进方法

3.1 CA注意力机制

由于辙叉图像背景复杂,包含各种天气、光照条件,导致掉块在图像中成像的纹理不相同,原模型算

法在使用卷积提取特征时,容易丢失特征信息,对掉块的检测效果不佳。针对卷积神经网络(CNN)难以从全局特征中提取特征的问题,通道注意力机制可以明显改善模型性能。

CA模块通过精确的位置信息对通道之间的关系和长程依赖进行编码,编码分为2个步骤:坐标信息嵌入(Coordinate Information Embedding)和坐标注意力生成(Coordinate Attention Generation),具体结构如图3所示。为了引入通道注意力机制,首先对输入特征图 x 使用尺寸为 $(H,1)$ 和 $(1,W)$ 的池化核,将每个通道内的特征进行全局平均池化。这样可以将高宽维度上的信息降维为通道维度上的一个特征向量,该向量表示了每个通道的整体特征重要性。计算公式为:

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_c(h, i) \quad (1)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

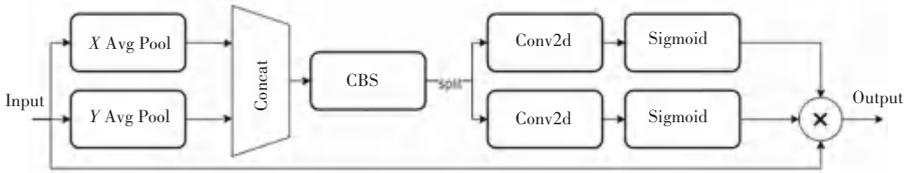


图3 CA注意力机制

Fig. 3 CA attention mechanism

在坐标信息嵌入阶段,将全局平均池化得到的特征向量与原始特征图连接在一起,然后通过共享的 1×1 卷积对其进行转换,具体公式为:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

生成结果是对空间信息在水平方向和垂直方向的中间特征图,接着按照空间维度将 f 切分为2个单独的张量 f^h 和 f^w ,通过2个 1×1 卷积和 F_h 及 F_w 将特征图 f^h 和 f^w 转换到和输入 X 相同通道数的形式。然后,利用Sigmoid激活函数运算获得特征图垂直方向的注意力权重 g^h 和水平方向的注意力权重 g^w ,权重表示了每个通道在垂直和水平方向上的重要性。研究推得的公式为:

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

最后,将学习到的通道注意力权重与原始特征图相乘,从而加权地强调或抑制每个通道的特征。数学公式具体如下:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

与SE,ECA等注意力机制相比,CA注意力机制更关注于强调位置信息以捕捉目标结构,针对辙叉形状颜色与背景区别明显的特点,其特征强调能力作用较强,能使网络更为有效地关注于辙叉结构本身。因此,本文选择使用CA注意力机制进行YOLOv5网络模型的改进和优化。通过该机制,改进后的模型能够自适应学习每个通道的重要性,并增强网络提取相关特征的能力,从而更准确地检测辙叉图像中的损伤。

3.2 改进密集网络模块

在铁道辙叉损伤检测中,损伤块的尺度大小不一,为进一步增强模型的特征融合能力,本文在模型的Neck层加入改进的密集连接模块(DenseNet)。DenseNet的关键创新点在于特征图之间的密集连接(Dense Connectivity)。该网络中的每个层都会接收前面所有层的特征图作为输入,并将自己的特征图输出传递给后面的所有层。这种密集的连接机制导

致网络层数的线性增长,但参数数量的增长则大大减缓。

由于 DenseNet 的直接连接机制,特征图数量快

速增长,使得模型的计算资源消耗变大,且训练时间过长,本文在 DenseNet 的基础上,减少网络连接,具体改进如图 4 所示。

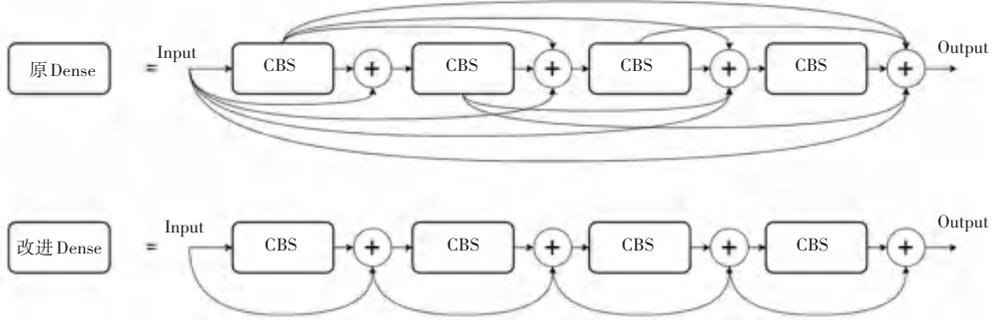


图 4 密集连接模块

Fig. 4 Dense connection module

3.3 CBAM 模块

CBAM 是一种轻量级的注意力机制,旨在提高卷积神经网络在图像处理任务中的性能。CBAM 注意

力机制如图 5 所示。由图 5 可看到, CBAM 由 2 部分组成:通道注意力模块(Channel Attention Module)和空间注意力模块(Spatial Attention Module)。

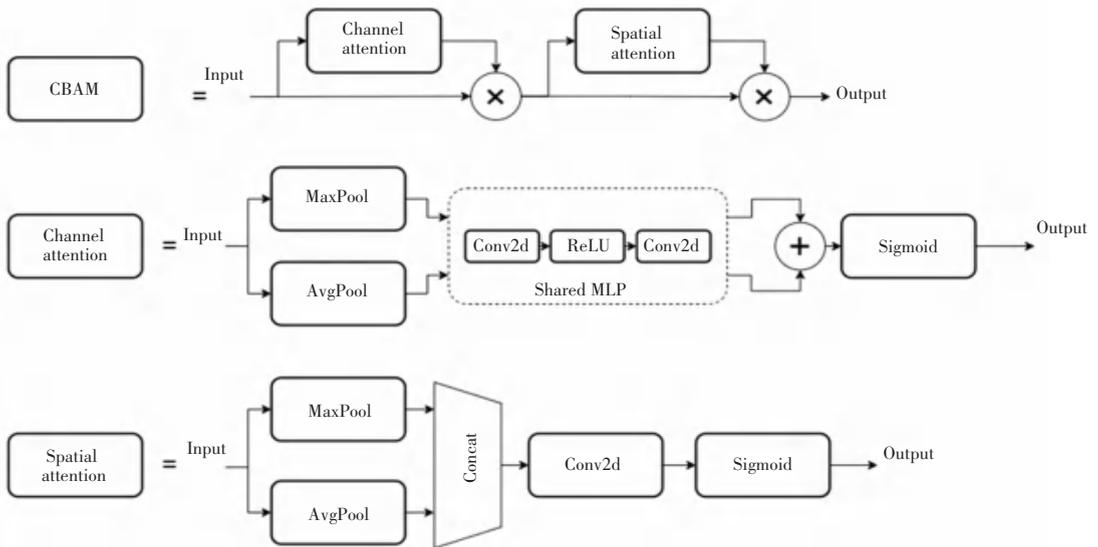


图 5 CBAM 注意力机制

Fig. 5 CBAM attention mechanism

通道注意力模块通过全局平均池化和全局最大池化,获得输入特征图在通道维度上的全局信息,并利用该信息赋予不同通道不同的权重,从而突出对当前任务更重要的特征通道。空间注意力模块则是通过平均池化和最大池化在空间维度上获取输入特征图的全局信息,并基于该信息为不同的空间位置赋予权重,以关注输入图像的更重要区域。

CBAM 机制结合了空间注意力和通道注意力, 辙叉损失存在损伤区域密集的情况,融入 CBAM 注意力机制,在网络学习时能增强辙叉区域权重,使其更加关注辙叉本身,减少背景的权重。同时,通过通

道注意力机制,能减少不同光照条件带来的不利影响。本文在 Neck 层结果送入检测头前加入 CBAM 模块,能有效聚焦特征,增强检测效果。

4 实验分析

4.1 实验环境

本次实验使用 Ubuntu 操作系统,使用的 GPU 为 RTX 3080 10 GB,CPU 为 AMD EPYC 7601,内存 64 GB。运行环境使用 Python3. 8, Pytorch 2. 0. 1, Cuda12. 1。所用数据集为自制数据集,实验所需超参数见表 1。

表1 实验使用超参数

Table 1 The experimental hyperparameters

参数	值
Batch_size	16
IoU 阈值	0.5
优化器	SGD
initial learning rate	0.01
Epoch	200
图像大小	640×640

4.2 特征提取注意力改进分析

为验证不同注意力机制对YOLOv5 Backbone网络特征提取的提升,本文构建了嵌入CBAM、ECA、CA、NAM和Triplet等注意力机制的C3模块,并分别替换基础YOLOv5中的原始C3模块,形成多个集成注意力的模型变体。在铁道辙叉损伤检测任务上对这些模型进行性能分析对比。实验结果见表2。

表2 注意力机制改进实验

Table 2 Attention mechanism improvement experiment

网络	P/ %	R/ %	mAP @ 0.5/ %	F1
YOLOv5	73.1	59.1	65.6	65.4
YOLOv5-CBAM	82.5	57.5	67.8	67.8
YOLOv5-ECA	81.1	59.5	67.4	68.6
YOLOv5-NAM	78.0	57.2	65.4	66.0
YOLOv5-Triplet	83.0	59.1	66.1	69.0
YOLOv5-CA(ours)	81.6	59.9	68.3	69.1

除了NAM(Normalization-based Attention Module)模块外,其他4种注意力机制均不同程度地提高了YOLOv5在检测精度、召回率和F1指标上的性能表现。具体来说,与原YOLOv5相比,集成CA

模块的变体在mAP指标上获得2.7%的绝对提升,F1提升3.7%;相对于其他注意力模块,CA模块的优势也较为明显,错检率和漏检率都有所下降。网络的稳定性较强。实验表明在YOLOv5的Backbone中引入CA注意力机制在铁道辙叉损伤检测上能有效提高模型的特征提取能力,改善模型性能。

4.3 消融实验

YOLOv5-CDC对YOLOv5模型的主干网络、Neck进行了改进。为评估不同模块改动和不同模块组合对于算法性能优化的程度,设计了消融实验,共设计了5个实验,分别为基础网络YOLOv5, YOLOv5-CA, YOLOv5-Dense, YOLOv5-CBAM和YOLOv5-CDC。消融实验结果见表3。

本文在YOLOv5检测模型中引入了多个注意力机制模块,以提高其特征表达和融合能力。具体来说,在主干网络中加入了CA注意力模块,实验结果表明,相比原YOLOv5模型,平均精度提升了2.7%,F1值提升了3.7%。这说明通道注意力机制可以有效增强主干网络的特征提取能力。在Neck模块中,本文采用了改进的Dense连接方式,也显著提升了模型性能,平均精度和F1值分别提升了1.9%和2.2%。这主要是由于Dense连接可以增强不同抽象级别特征之间的融合。而在Neck和检测头之间,本文加入了通道块注意力(CBAM)模块,平均精度获得了0.9%的提升。分析发现,这是因为CBAM显著增强了Neck层特征图的表达和融合能力,并减少了漏检现象的发生。综合以上改进后,本文提出的YOLOv5-CDC模型相比原YOLOv5模型,平均精度提升了3.4%,F1值提升了5.3%。总体来说,提高了模型的表达能力和鲁棒性。

表3 消融实验

Table 3 Ablation experiment

网络	CA	Dense_A	CBAM	P/ %	R/ %	mAP @ 50/ %	F1
YOLOv5	-	-	-	73.1	59.1	65.6	65.4
YOLOv5-CA	√	-	-	81.6	59.9	68.3	69.1
YOLOv5-Dense	-	√	-	79.1	59.1	67.5	67.6
YOLOv5-CBAM	-	-	√	79.3	62.1	66.5	69.7
YOLOv5-CDC	√	√	√	81.8	62.2	69.0	70.7

4.4 对比实验

为了客观评估本文提出的改进YOLOv5模型的检测效果,本文采用了多模型对比实验。本文在铁道辙叉损伤这个目标检测数据集上,分别测试了Faster R-CNN、SSD、YOLOv5、YOLOv8等经典检测

算法的性能,见表4。本文的改进模型相较原始YOLOv5的平均精度(mAP)提升了3.4%,这证明本文提出的改进密集模块和注意力机制对检测性能提升的有效性。与其他经典检测器如Faster R-CNN、SSD和YOLOv8相比,本文的模型也展现出了

更优秀的检测能力。通过严格的对比实验和评价指标,可以得出本文所提出的改进 YOLOv5 模型在检测精度和泛化能力上都获得了显著提升。

表 4 对比实验
Table 4 Comparison experiment %

网络	<i>P</i>	<i>R</i>	<i>mAP@ 50</i>
SSD 300	69.4	54.3	55.4
Faster R-CNN	70.9	58.6	64.6
YOLOv8	74.9	60.4	65.9
YOLOv5	73.1	59.1	65.6
YOLOv5-CDC(ours)	81.8	62.2	69.0

5 结束语

本研究以铁道辙叉损伤检测为应用背景,针对当前目标检测模型在该任务上的检测精度不高和误检率问题,构建了一个包含复杂场景的铁道辙叉损伤图像数据集。基于经典的 YOLOv5 模型,本文通过在不同模块嵌入注意力机制进行模型优化:在 Backbone 中集成了 CA 模块提升低层特征表达;Neck 部分采用改进的 Dense 连接增强多尺度特征融合;检测头前接入 CBAM 模块聚焦相关特征。实验表明,相较原 YOLOv5,优化后的 YOLOv5-CDC 模型在平均精度和 *F1* 指标上分别获得 3.4% 和 5.3% 的提升,与主流检测器比较也展现出良好的泛化性。研究结果证明了在目标检测模型中融合注意力机制

的有效性。但是考虑到当前数据集的目标类别单一,后续工作将继续扩充数据集并探索更复杂场景下的应用。

参考文献

- [1] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster RCNN: Towards realtime object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [2] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, realtime object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 779-788.
- [4] 罗晖, 徐广隆. 基于图像增强与深度学习的钢轨表面缺陷检测 [J]. 铁道科学与工程学报, 2021, 18(3): 623-629.
- [5] 韩强, 刘俊博, 冯其波, 等. 基于多层次特征融合的钢轨表面损伤检测方法 [J]. 中国铁道科学, 2021, 42(5): 41-49.
- [6] 刘浩翰, 孙铖, 贺怀清, 等. 基于改进 YOLOv3 的金属表面缺陷检测 [J]. 计算机工程与科学, 2023, 45(7): 1226-1235.
- [7] 张辉, 宋雅男, 王耀南, 等. 钢轨缺陷无损检测与评估技术综述 [J]. 仪器仪表学报, 2019, 40(2): 11-25.
- [8] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 13708-13717.
- [9] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]// FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision—ECCV. Lecture Notes in Computer Science. Cham: Springer, 2018: 3-19.