

基于轻量级 CDM-YOLO 的多尺度桥梁 混凝土缺陷检测^{*}

范崇俊 董绍江 罗家元 张 霞 闫凯波

(重庆交通大学机电与车辆工程学院 重庆 400074)

摘 要: 针对爬壁机器人进行桥梁检测时存在的对尺度多变,密集混杂混凝土表观缺陷识别精度低,易错检漏检问题,提出了一种基于 YOLOv12n 的轻量级 CDM-YOLO 算法。首先,针对多尺度缺陷难识别问题,在骨干网络浅层引入 MFC 多尺度特征融合网络,改善其提取多样化和细粒度特征能力,丰富特征信息流,使模型适应多尺度缺陷。其次,针对相似缺陷易混淆问题,在骨干网络深层和颈部使用动态 tanh 机制强化模型特征融合和聚焦能力,提升其对不同缺陷的分辨力,降低漏检错检。最后,针对密集混杂缺陷,在模型颈部采用 CARAFE 上采样算法,加强深度语义信息流动,优化模型对密集缺陷的识别能力。上述方法在保持网络实时轻量的前提下提升了模型的检测精度。CDM-YOLO 的 mAP(IoU=0.5)和召回率相对于 YOLOv12n 分别提升了 2.43% 和 3.25%,表明其能更好应对多尺度和密集缺陷,错检漏检发生率更低,并支持有限算力的爬壁机器人及现场设备。

关键词: 混凝土表观缺陷;YOLOv12n;缺陷检测;桥梁混凝土;爬壁机器人;桥梁检测

中图分类号: TN06; TP183 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi-scale concrete defect detection in bridges based on lightweight CDM-YOLO

Fan Chongjun Dong Shaojiang Luo Jiayuan Zhang Xia Yan Kaibo

(School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: To address the challenges faced by wall-climbing robots during bridge inspections—namely low accuracy in identifying densely clustered, visually complex concrete defects across varying scales, leading to false positives and false negatives—we propose a lightweight CDM-YOLO algorithm based on YOLOv12n. First, to tackle the difficulty in recognizing multi-scale defects, we introduce a multi-scale feature fusion network into the backbone. This enhances the backbone's ability to extract diverse and fine-grained features, enabling the model to adapt to defects of varying scales. Second, to address the confusion between dissimilar defects, a dynamic tanh mechanism is employed in the neck to enhance feature focus, clearly distinguishing different defects and reducing false positives and negatives. Finally, for densely clustered defects, the CARAFE algorithm is applied in the neck to strengthen deep semantic information flow, optimizing the model's ability to identify dense defects. These methods improve detection accuracy while maintaining real-time performance and lightweight characteristics. Compared to YOLOv12n, CDM-YOLO achieves a 2.43% improvement in mAP (IoU=0.5) and a 3.25% increase in recall. This demonstrates its superior handling of multi-scale and dense defects with lower false positive and false negative rates, making it suitable for wall-climbing robots and field equipment with limited computational resources.

Keywords: concrete surface defects; YOLOv12n; defect detection; bridge concrete; wall-climbing robot; bridge inspection

0 引 言

截至 2024 年末,我国建成桥梁已达 107.93 万座,长度

约 8 576.49 万延米^[1],发展态势蓬勃。但随着早期建成桥梁服役时间的增加,其混凝土结构在自然老化和复杂外因,如干湿循环、温度应力和复杂荷载作用下会不可避免的出

收稿日期:2025-09-28

^{*} 基金项目:重庆市自然科学基金项目(CSTB2024NSCQ-LZX0024)、重庆市技术创新与应用发展专项重大项目(CSTB2025TIAD-STX0023)、重庆市技术创新与应用发展重点研发项目(CSTB2024TIAD-KPX0081)、重庆市教委科学技术研究项目(KJZD-K202300711)资助

现各种缺陷,为桥梁安全服役埋下隐患,甚至在未来造成重大安全事故,定期进行桥梁检测和评估就具有了重要意义。传统桥梁检测依赖人工检测,但存在检测效率低、检测周期长且经验依赖严重等问题^[2]。为应对上述问题,通过无人机或机器人搭载图像采集设备和深度学习算法代替人工,进行智能化检测的方式在近年来逐渐获得青睐。该方法具有检测效率高、经验要求弱和操作简便等优点,适用于工程现场和人工难检场景,如水上桥墩和斜拉索桥梁高耸立面等缺陷壁面检测。而针对不同应用场景,大量优化算法也被提出。倪彤元等^[3]在 SSD(single shot multiBox detector)网络深层中引入了一种局部特征敏感的卷积块注意力机制(convolutional block attention module, CBAM),改善微小裂缝缺陷感知能力,提高了识别精度,但不能适应多种类别缺陷检测;Yang 等^[4]提出了一种图卷积网络(graph convolutional network, GCN)与多轴视觉变换器结合的网络,改善了标签关联特征,实现高精度多种类缺陷分类,但不能报告目标准确位置;梁胤杰等^[5]设计了泊松混合损失方法对于干扰背景缺陷进行增广,并引入 SPD-Conv 增强复杂背景和低分辨率图像的位置检测能力,而 Wang 等^[6]则引入了 SENet(squeeze-and-excitation network)降低背景噪音,并通过不对称处理正负样本损失以平衡损失信息,提高位置检测精度,但上述两种方法模型规模均不够轻量;苗力恒等^[7]针对水下资源受限环境基于路径聚合网络(path aggregation network, PAN)重构了 YOLOv8 颈部(neck),在降低参数量的前提下提升模型多尺度特征融合能力,但推理速度略低;韩建峰等^[8]基于 U-Net(U-Net)提出了一种能够解决无人机场路面裂缝检测断裂问题的分割算法,其泛化能力强,但难以分辨密集混凝土缺陷特征;Huang 等^[9]针对一般模型语义表示不足及高阶特征关联弱的问题融合了跨位置表达的超图网络,从而实现精确分辨缺陷位置及形态,但欠缺实验验证;丁威等^[10]基于爬壁机器人及稳定的视觉采集设备,结合一种裂缝分割网络(feature pyramid transformer, FPT)完成多尺度裂缝准确测量,兼顾实际应用和检测精度,但又仅局限于裂缝缺陷。目前,各种缺陷检测算法存在如对尺度多变,密集混杂缺陷的错检漏检,模型规模、精度和适应缺陷种类不能兼顾,多缺陷检测应用较少等问题。

为解决上述精度-规模-应用矛盾,本文在轻量级 YOLOv12n 模型^[11]的基础上构建了 CDM-YOLO 模型,并基于自研爬壁机器人和检测软件针对实际缺陷混凝土壁面开展验证。该模型通过引入 MFC 多尺度特征融合网络、DyACf 模块和 CARAFE 上采样算法来改善网络模型对多样化特征提取、关联特征融合和深度语义信息保留能力。实验结果表明,CDM-YOLO 的 mAP(IoU = 0.5)上升到 87.53%,召回率上升到 81.10%,精确率也达到 91.17%,相较于原始 YOLOv12n 模型的 mAP(IoU = 0.5)为 85.10%,召回率为 77.85%,精确率为 91.56%,模型预测准确性取得了一定提升。关于模型规模及计算复杂度,

CDM-YOLO 的计算量和参数量仅为 7.6 GFLOPs 和 2.86 M,保持了网络的轻量化。综上,CDM-YOLO 在保持了轻便高效的前提下增强了多尺度混凝土缺陷的识别能力,为桥梁混凝土表观缺陷检测提供了一种轻量级多尺度缺陷检测方法。

1 混凝土缺陷检测算法设计

1.1 YOLOv12 模型概述

YOLOv12 模型首次引入了一种区域划分简单,感受野较大的局部注意力机制,称之为区域注意力机制(area attention, AAttn)。AAttn 省略了一般注意力机制中窗口划分和额外位置编码的操作,并利用 flash attention 算法优化缓存和显存混合读写的 I/O 开销,从而在不降低卷积神经网络(convolutional neural networks, CNN)速度的前提下,利用丰富上下文关联提高检测精度。这一创新架构及其特性正有助于模型在复杂工程环境下分辨近邻的具有相似特征的不同缺陷,提升检测精度。由于现场进行桥梁检测需要采用机器人检测系统作为承载,算力有限。故基于规模最小的 YOLOv12n 为基准模型(baseline)进行优化。

1.2 CDM-YOLO 构建

考虑实际工程场景桥梁混凝土表观缺陷类别多,缺陷尺度多变且不同类别和尺度缺陷密集混杂,本文针对此对 YOLOv12n 进行如下改进优化:

1) 针对实际缺陷尺度变化大,难以兼顾问题,在模型骨干网络(backbone)浅层用多尺度特征融合网络(MFC)代替传统 C3k2 和 Conv 模块的堆叠,以改善 backbone 难以提取多层次、多样化和细粒度特征问题,丰富模型特征信息流,从而使模型适应尺度多变的混凝土缺陷。

2) 针对现场壁面存在相似缺陷多,易错检漏检问题,对模型 backbone 深层和 neck 的原始 A2C2f 进行改进,引入动态 tanh(dynamic tanh, DyT)机制,简化 AAttn 的计算量和参数量,稳定训练梯度流,同时强化 neck 的特征融合能力,从而提升模型对相似类别缺陷的识别能力,降低漏检和错检。

3) 针对密集混杂缺陷,在 neck 中引入内容感知特征重组(content-aware reassembly of features, CARAFE)上采样算法代替最临近插值算法,实现局部特征自感知,加强深度语义信息流动性,从而不损失密集缺陷边界特征和混杂缺陷的局部典型特征,提高模型检测精度。

联合上述改进,经实验证明,可以综合提升模型对相似和多尺度缺陷检测能力,检测精度有较好的提升。将改进后的模型命名为 CDM-YOLO,其结构如图 1 所示。

1.3 多尺度特征融合网络 MFC

为应对工程环境存在的多变尺度缺陷,设计了一种 MFC 多尺度特征融合网络,总体结构如图 2 所示。

MFC 网络具有 3 条^[12]供不同缺陷信息流动的通路,包括一条 1×1 卷积通路,一条深度可分离卷积(depthwise

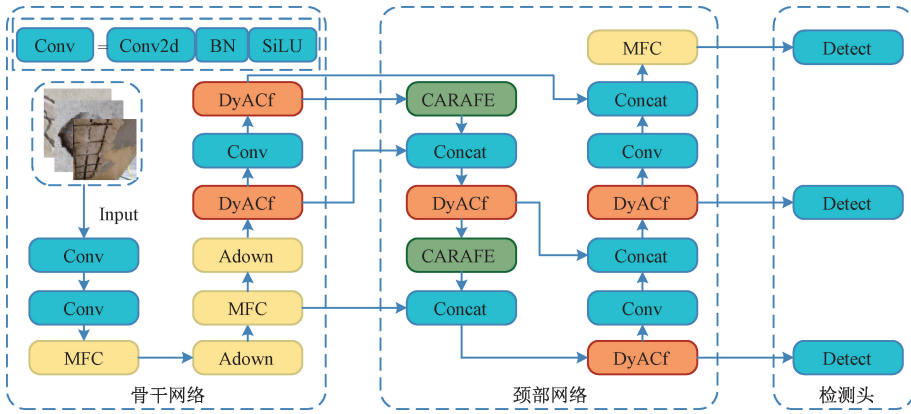


图 1 CDM-YOLO 模型总体结构

Fig. 1 CDM-YOLO overall structure

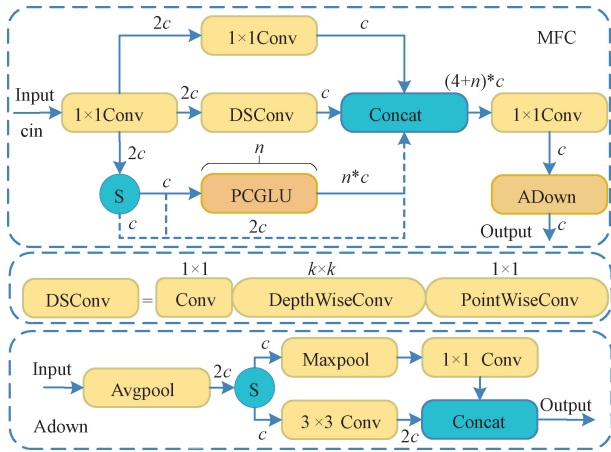


图 2 MFC 网络结构

Fig. 2 MFC network structure

不同通路的通道整合、层次表示和细粒度特征,从而获得了兼具多样性与语义深度的丰富特征信息流。为进一步简化计算量与参数量,引入了基于下池化和 1×1 卷积的 Adown 模块^[16]以降低特征图尺寸。

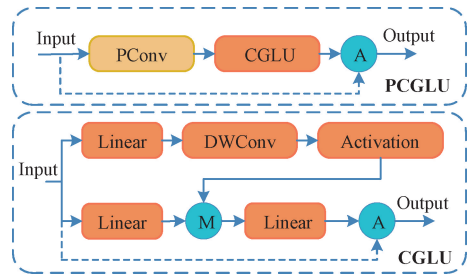


图 3 PCGLU 通路内部结构

Fig. 3 Internal structure of PCGLU channel

separable convolution, DSConv) 通路^[13], 和一条带 PCGLU 模块的通路。经 Conv 粗提取的缺陷特征先经过一个 1×1 卷积压缩以降低计算量,然后同时通过 3 条通路。其中 1×1 卷积通路对全部特征进行维度压缩变换以进行跨通道信息交互,获得通道整合特征信息。DSConv 通路对缺陷特征执行逐深度卷积(DWConv)和逐点卷积(PWConv),提取了输入的空间特征并进行升维,从而获得强化的层次特征表示。由于计算阶段的分离,这一过程的计算量大大降低,使 MFC 网络更加轻量。

对于带 PCGLU 模块的通路(如图 3 所示),内部则采用主流残差连接方式堆叠。初始粗粒度特征首先经过部分卷积(partial conv, PConv)操作^[14],在少量增加计算量的前提下留存其显著部分。接着进行门控线性卷积(convolutional GLU, CGLU)^[15],即使用门控分支选择性的“开关”值分支以控制值分支内的信息流。门控分支中采用 DWConv 以实现通道级注意力位置编码,从而使得 CGLU 能够从 PConv 结果中再次选优基于最近邻特征的局部空间显著特征,获取足量细粒度特征,并提升较小尺度缺陷的识别精度。最后,MFC 采用一个 1×1 卷积融合

MFC 网络的主要创新在于其 3 条并行通路可以产生丰富多样的细粒度特征信息流,显著增强了 backbone 特征分离和流动能力,产生并留存不同规模缺陷的特征表达,从而使模型获得适应多种尺度混凝土缺陷的能力。

1.4 特征聚焦 DyACf

归一化层起到了稳定模型输入,提升网络深度和学习效率的作用。然而对于引入了 AAttn 的 YOLOv12n 模型,批归一化(batch norm, BN)存在对序列化输入适应性差和对批处理超参数敏感的缺点;层归一化(layer norm, LN)虽然能够适应序列数据,但又具有计算量巨大,对不同类别输入特征分离度差的问题。为改进上述问题,在 AAttn 前后引入了动态 DyT 机制^[17]。DyT 机制采用形如统计上 LN 层“输入数据-输出数据”S 型曲线的带可学习参数的 tanh 函数来替代 YOLOv12n 中复杂的归一化层,其仅对输入元素进行变换而不计算复杂统计量,兼具简单性和功能性。公式为:

$$\text{DyT}(x) = \gamma \tanh(\alpha x) + \beta \quad (1)$$

式中: α 为缩放可学习参数, γ 和 β 为通道可学习参数。

DyT 对输入特征进行元素级 α 倍缩放来适应不同特征,然后进行 $\tanh$ 运算平滑压缩极端值完成归一化,最后进行 γ 倍缩放和 β 偏置适应不同通道。这些参数均随模型训练过程学习更新,以动态跟随网络不同深度和通道的阶段性特征,保证特征实时最优。学习情况如图 4 所示。

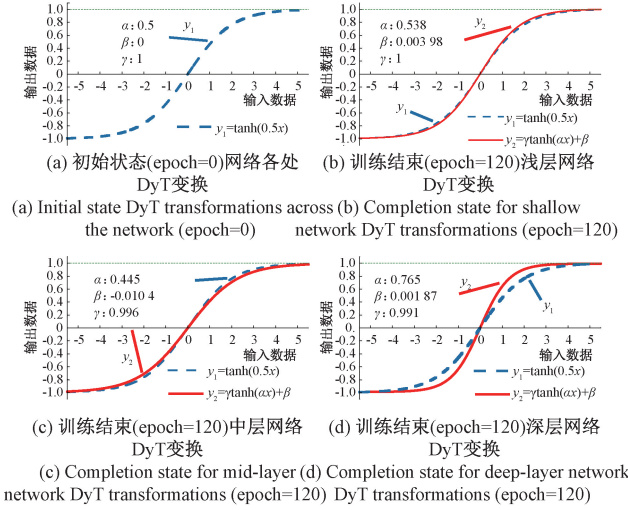


图 4 DyT 机制学习情况可视化

Fig. 4 Visualization of DyT mechanism learning progress

由图 4 可知,随着网络深度增加,训练结束时 DyT 参数距离初始参数差值有变大趋势,说明其对输入特征压缩能力逐渐增强,起到了稳定训练梯度的作用。DyT 的引入避免了 BN 层不适应序列数据的缺点,并利用形式简单的公式规避了 LN 层的复杂统计量计算。其利用输入数据,尤其是极端值的非线性压缩稳定了训练梯度流,降低了模型对非典型缺陷特征的学习难度,加快收敛速度。DyT 的动态可学习参数还降低了模型对初始化参数的敏感性,提高了模型检测的鲁棒性。基于 DyT 和 AAttn 设计了 DyACf 模块,具体结构如图 5 所示。

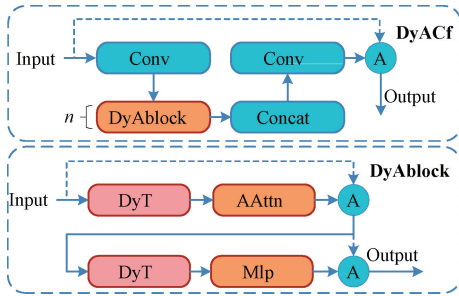


图 5 DyACf 模块结构

Fig. 5 DyACf module structure

由图 5 可知,DyACf 模块将前级多样化特征进行通道混合,寻找特征内在联系,接着通过 AAttn 注意力关注内在联系的有效部分,并用 DyT 进一步简化关联映射,从而有效分离易混淆的相似特征。DyACf 因 AAttn 而具有较

强区域缺陷特征聚焦能力,又因 DyT 而具有高效分离和序列化能力,从而能够有效分离相似类别缺陷区域,较好聚焦于不同特征类别,提高网络识别易混淆的相似类别桥梁混凝土表观缺陷的能力,降低模型错检和误检率。

1.5 富语义上采样 CARAFE

YOLOv12n 采用最近邻插值算法完成输入图像上采样,融合深层抽象特征和细节信息,保留丰富层次特征。但该方法忽略了采样像素点邻域灰度及其变化信息,造成上采样结果信息变化不连续,关联语义信息损失严重。这一问题使得原始 YOLOv12n 难以分辨密集混凝土缺陷特征,并降低其对混杂缺陷特征检测的准确率。若换用反卷积法,可以改善关联语义信息损失问题,但该方法又有局部特征丢失和算力依赖严重的问题。综合考虑上述问题,采用轻便的 CARAFE 上采样算法^[18]来替换传统算法。

CARAFE 上采样过程分两过程,即卷积核预测 (kernel prediction, KP) 过程和内容感知重组 (content-aware reassembly, CAR) 过程。其中 KP 过程如图 6 所示。

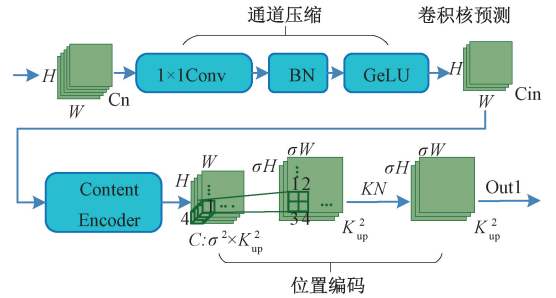


图 6 CARAFE 的卷积核预测过程

Fig. 6 CARAFE convolutional kernel prediction process

KP 过程首先对输入图像进行卷积核预测:将输入通道通过 1×1 卷积压缩以降低计算量,再使用具有中等感受野的位置信息编码核与输入图像进行卷积,从而获得语义信息关联充分的局部特征。最后进行核归一化,完成语义信息重组。该过程公式描述如式(2)~(3)所示。

$$k_{\text{encoder}} = k_{\text{up}} - 2 \quad (2)$$

$$D'_l = KP(\delta(D_l, k_{\text{encoder}})) \quad (3)$$

式中: k_{encoder} 表示位置信息编码核尺寸, k_{up} 表示内容感知核尺寸, $\delta(D_l, k_{\text{encoder}})$ 表示以位置 l 为中心的特征图 D_l 的 $k_{\text{encoder}} \times k_{\text{encoder}}$ 临近区域。

接着进行 CAR 过程,公式如式(4)所示。

$$B_l = CAR(\delta(D_l, k_{\text{up}}), D'_l) \quad (4)$$

式中: $\delta(D_l, k_{\text{up}})$ 表示以位置 l 为中心的特征图 D_l 的 $k_{\text{up}} \times k_{\text{up}}$ 临近区域。

该过程将上阶段获得的富语义特征按原始输入图像特征分布进行特征再编码,从而实现输入图像局部内容自感知和增强,生成具有明确语义信息和较强局部特征的高分辨率上采样图像。CAR 具体操作流程如图 7 所示。

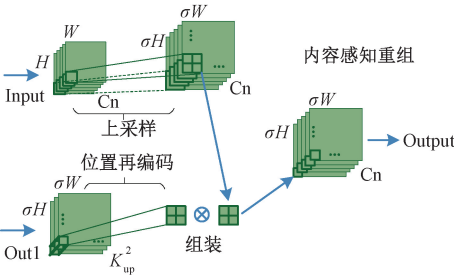


图 7 CARAFE 的内容感知重组过程

Fig. 7 CARAFE content-aware reorganization process

相比于传统上采样, CARAFE 上采样算法的 KP 过程通过连续卷积获取了充分关联的邻域细节特征, 克服了局部特征丢失问题, 并增强密集缺陷边界特征; CAR 过程通过原始图像位置自编码和组装实现语义增强, 重组出上下文关联紧密的深度语义信息, 克服了关联语义信息损失严重问题, 并自然区分混杂缺陷特征。两过程协同改善上采样输出图像质量, 增强密集混杂缺陷的检测能力。图 8 为 CARAFE 与 baseline 上采样的特征图效果对比, 由图可见 CARAFE 采样结果更细腻, 分辨率更高, 边界特征明显, 更适应密集缺陷检测。

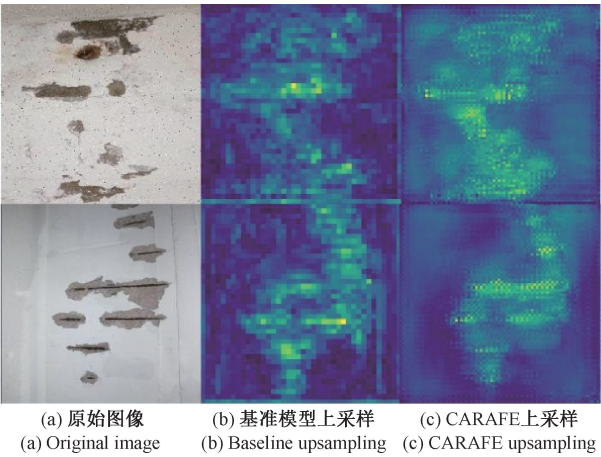


图 8 模型上采样结果对比

Fig. 8 Upsampling results comparison

2 实验设计与结果分析

2.1 数据集构建

本文使用的混凝土表面缺陷图像部分搜集于互联网, 部分来源于南洋理工大学开源数据集 Public Crack^[19], 一共获取到了含裂缝、剥落和露筋 3 种类型缺陷的 7 861 张图片。为降低深度学习模型学习难度, 丰富数据特征, 减轻数据“长尾效应”, 将图像尺寸统一处理为 640×640, 并进行图片模糊、加高斯噪声、对比度变化、旋转和翻转等数据增强处理, 得到了 12 119 张图片。按 7:2:1 的比例划分训练集、验证集和测试集, 用于后续实验。数据集图片及处理示例如图 9 所示。



图 9 原始缺陷图像和增强缺陷图像示例图

Fig. 9 Example of original and enhanced defect images

2.2 实验环境及超参数配置

本文实验平台为 Ubuntu 11. 4. 0, 模型均采用 Python 语言并依托 Pytorch 框架实现。详细计算机硬件及环境配置如表 1 所示。

表 1 实验环境主要配置

Table 1 Main configurations of experimental environment

计算机配置	参数信息
CPU 型号	16 核 AMD EPYC 9354
内存	60 GB
GPU 型号	RTX 4090 24 GB
Pytorch 版本	2. 2. 2
Python 版本	3. 11
CUDA 版本	11. 8

关于模型的超参数, 综合考虑模型收敛速度及最终精度, 选用 AdamW 优化器, 设定模型训练轮次 (epoch) 为 120, 批处理大小 (batch size) 为 64, 初始学习率 (lr0) 为 0. 000 6。为增强模型泛化性能, 提高训练速度和训练稳定性, 启用马赛克 (mosaic) 增强、训练热身 (warm-up)、AMP 算法和 flash attention 算法进行训练。详细超参数配置如表 2 所示。

表 2 模型主要超参数配置

Table 2 Model main hyperparameter configuration

超参数名称	超参数值	超参数名称	超参数值
epoch	120	IoU	0. 680 0
batch size	64	momentum	0. 940 0
lr0	0. 000 6	mosaic rate	0. 937 0
lrf	0. 100 0	close-mosaic	60
weight decay	0. 000 5	热身动量	0. 842 0
optimizer	AdamW	热身轮次	3

2.3 评价指标

本文采用了 6 个常用技术指标以合理评估模型不同方面性能,依次为:精确率(precision, P)、召回率(recall, R)、类别平均精度(mean average precision, mAP)、帧率(frames per second, FPS)、模型参数量(parameters)和模型浮点计算量(GFLOPs)。其中,精确率衡量模型预测正样本的准确性,召回率衡量模型正确搜索正样本的能力,类别平均精度综合评价模型预测多个类别的平均准确性,帧率衡量模型实时计算速度,参数量和浮点计算量综合衡量模型规模及计算复杂度。指标详细计算公式为:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

式中: P 为精确率, R 为召回率, TP 为模型预测正确的正样本, FP 为预测错误的正样本, FN 为未预测到的正样本。

$$mAP = \frac{1}{N} \sum_{i=0}^N P_i dR \tag{7}$$

式中: N 为样本类别数, P_i 为第 i 类别的准确率。

$$FPS = \frac{1\ 000}{pre + infer + post} \tag{8}$$

式中: FPS 为实时帧率, pre 为数据前处理时间, $infer$ 为模型推理时间, $post$ 为后处理时间,单位均为 ms。

2.4 消融实验

为评估 CDM-YOLO 的改进是否有效,在保证硬件配置和运行环境均相同的前提下设计了 8 组消融实验,实验结果如表 3 所示。

表 3 CDM-YOLO 模型模块消融实验结果

Table 3 Results of CDM-YOLO ablation experiment

YOLOv12n	MFC	DyACf	CARAfE	计算量/ GFLOPs	参数量/ M	精确率/ %	召回率/ %	mAP(IoU=0.5)/ %
✓				6.2	2.55	91.56	77.85	85.10
✓	✓			7.3	2.71	90.68	80.33	87.30
✓		✓		5.8	2.51	88.05	81.30	85.89
✓			✓	6.1	2.64	90.45	79.83	86.48
✓	✓	✓		7.3	2.71	89.11	82.24	87.21
✓	✓		✓	7.6	2.85	93.08	80.13	87.36
✓		✓	✓	6.1	2.65	90.06	80.07	87.41
✓	✓	✓	✓	7.6	2.86	91.17	81.10	87.53

由表 3 可知,在 baseline 分别引入 MFC、DyACf 和 CARAFE 模块后,mAP(IoU=0.5)均有不同程度增长,分别增长了 2.20%、0.79%和 1.38%,表明替换网络对多尺度缺陷均具有不同程度缓解能力。其中单独引入 MFC 网络指标增长最高,说明该模块的多样化特征提取和深度语义信息流动性有效增强了模型应对尺度多变缺陷的能力,并适应相对复杂的背景。而 DyACf 模块的改进在提升了 mAP(IoU=0.5)的前提下仍使模型计算量和参数量由 baseline 的 6.2 GFLOPs 和 2.55 M 分别下降到 5.8 GFLOPs 和 2.51 M,说明该模块不仅提高了网络识别混合类别缺陷的能力,还简化了 AAttn 的复杂计算,助力网络轻量化。对于 CARAFE 算法,其计算量和参数量分别为 6.1 GFLOPs 和 2.64 M,在略微提升参数量及降低计算量的基础上提升了 mAP(IoU=0.5)值,说明其上采样过程对算力依赖低,但语义信息明确,局部特征较强,能够改善模型对密集和混杂缺陷特征的检测能力。进一步,若将模块两两组合,则模块改善网络能力更强。其中效果最好的组合是 DyACf 和 CARAFE,其 mAP(IoU=0.5)达到了 87.41%,相较于 baseline 提升了 2.31%,这是由于两

模块存在协同作用,即 DyACf 对 CARAFE 提供的缺陷关联语义信息进行了有效的区域关注,从而协同提高了模型检测的准确性。

最后,同时改进 3 处的 CDM-YOLO 表现最优。其 mAP(IoU=0.5)较 baseline 提升 2.43%,recall 提升 3.25%,precision 仅下降 0.39%,但仍保持在 91%以上。其参数量和计算量保持在 2.86 M 和 7.6 GFLOPs,仅比 baseline 高 1.4 GFLOPs 和 0.31 M,保持了网络的轻量化。综上,消融实验表明了 CDM-YOLO 各改进模块均有效,而多个模块组合也有效,且模块越多效果越优,说明模块间协同性好。CDM-YOLO 有效提升了 baseline 的性能,并兼顾了网络的轻量化。

2.5 主流模型性能对比

为进一步评估改进后 CDM-YOLO 相对其他模型的性能和优势,选取 YOLO(you only look once)系列和 RT-DETR(real-time detection transformer)系列主流模型及 Faster R-CNN(fast region-based convolutional network)和 SSD 等经典模型进行常用指标对比分析,对比实验结果如表 4 所示。

表 4 主流模型性能指标对比实验结果

Table 4 Comparative experimental results of mainstream models

模型名称	不同类别的 AP(IoU=0.5)/%			mAP (IoU=0.5)/%	精确率/	召回率/	计算量/	参数量/	帧率/
	裂缝	剥落	露筋		%	%	GFLOPs	M	fps
Faster R-CNN	91.10	76.39	90.07	85.85	63.28	80.75	180.1	28.29	50.15
SSD	88.84	67.10	88.13	81.36	87.44	74.15	60.9	23.88	58.95
YOLOv8n	90.69	70.52	90.15	83.79	86.22	78.51	8.1	3.01	96.79
YOLOv11n	92.00	77.49	92.64	87.38	90.23	81.69	6.3	2.58	87.39
RT-DETR-HGNetv2-l	88.76	76.57	89.33	84.89	92.39	79.86	103.4	31.99	26.33
RT-DETRv2-R18vd	29.45	85.70	93.00	69.36	69.61	65.84	60.0	20.00	34.12
YOLOv5n	91.17	77.54	90.11	86.27	89.07	80.66	7.1	2.50	89.11
YOLOv10n	91.20	78.48	91.05	86.89	89.86	79.72	8.2	3.69	75.47
YOLOv13n	91.21	75.17	90.66	85.68	92.73	77.74	6.2	2.45	44.54
YOLOv12n	91.72	73.34	90.68	85.25	91.56	77.85	6.0	2.55	60.67
CDM-YOLO	92.22	78.12	92.25	87.53	91.17	81.10	7.6	2.86	56.24

由表 4 可知, CDM-YOLO 对比 YOLOv8n、RT-DETRv2-R18vd 和 SSD 模型具有性能上的显著优势, 其计算量和参数量分别低于上述模型 0.5、52.4、53.3 GFLOPs 和 0.15、17.14、21.02 M, 而 mAP(IoU=0.5) 和召回率则比上述模型高 3.74%、18.17%、6.17% 和 2.59%、15.26%、6.95%。同时, CDM-YOLO 相较于 YOLOv5n、YOLOv10n、YOLOv13n 和 RT-DETR-HGNetv2-l 的 mAP(IoU=0.5) 和召回率指标分别高 1.26%、0.64%、1.85%、2.64% 和 0.44%、1.38%、3.36%、1.24%, 而其计算量和参数量相差并不大(较 YOLOv5n 高 0.5 GFLOPs 和 0.36 M、YOLOv13n 高 1.4 GFLOPs 和 0.41 M; 较 YOLOv10n 低 0.6 GFLOPs 和 0.83 M、RT-DETR-HGNetv2-l 低 95.8 GFLOPs 和 29.13 M), 体现出较好的综合优势。对于 FPS, CDM-YOLO 相对 baseline 仅下降 4.43 fps, 仍保持在 56 fps 以上, 具备检测的实时性。细分缺陷中, CDM-YOLO 对裂缝的识别精度最佳, 为 92.22%, RT-DETRv2-R18vd 对剥落最敏感, AP 达到 85.70%。对于露筋缺陷, CDM-YOLO、RT-DETRv2-R18vd 和 YOLOv11n 均有不错的表现, AP 均达到 92% 以上, 识别能力较强。综上, CDM-YOLO 在各类缺陷检测上相对于大部分对比模型在检测精度、错检率和漏检率上均具备一定优势, 并能兼顾网络的轻量性和实时性, 能够更好的应对桥梁混凝土表观缺陷检测任务。

2.6 泛化性实验

为验证 CDM-YOLO 的泛化能力, 另选取 Roboflow 滑铁卢大学开源混凝土缺陷数据集(含裂缝, 剥落, 露筋和泛碱缺陷, 其中泛碱缺陷未使用)^[20]进行模型泛化验证。由表 5 可知, 由于不同数据集特征分布有所差异, 跨数据集验证时各模型指标均有所下降, 但对比各模型, CDM-YOLO 的 mAP 和 P 仍高于 baseline、YOLOv13n 和 RT-DETR-HGNetv2-l 模型 3.18%、10.13% 和 4.73%、6.06%

和 1.20%、0.18%, 表明其泛化性能更优。召回率虽然低于 YOLOv10n 和 RT-DETR-HGNetv2-l, 但计算量和参数量却优于上述二模型 0.6 GFLOPs、0.83 M 和 95.8 GFLOPs、29.13 M(表 4), 表明其速度更快, 更适合有限算力设备。

表 5 泛化实验结果

Table 5 Generalization experiment results %

模型名称	精确率	召回率	mAP(IoU=0.5)
SSD	75.34	44.46	50.28
YOLOv8n	70.87	47.23	50.82
RT-DETR-HGNetv2-l	76.49	59.59	52.23
YOLOv5n	69.25	47.47	50.90
YOLOv10n	74.38	49.26	53.14
YOLOv13n	70.61	45.58	48.70
YOLOv12n	66.54	48.03	50.25
CDM-YOLO	76.67	48.25	53.43

2.7 可视化分析

为观察 CDM-YOLO 特征提取和区域关注的效果, 对图像进行了 grad-CAM 热力图分析, 结果如图 10 所示。热力图中颜色相对较暖的区域代表模型更关注的区域。由图 10 可知, CDM-YOLO 对分布较广的网状裂缝(第 1 行)既能关注到缺陷中心, 也不丢弃分布较广的周围裂缝信息, 而对较小尺度的剥落缺陷(第 2 行)外环境特征几乎无错误关注。最后, CDM-YOLO 对大面积露筋(第 3 行)和纤细裂缝(第 4 行)两种差异情形, 也能达成对不同缺陷特征的兼顾, 并相对于 baseline 保留更多边缘信息。热力图说明, CDM-YOLO 对多种类别或尺度缺陷均能进行较好关注, 缺陷和背景分离度高, 边缘信息丢失少, 特征关注具有一定优势。为更加直观的展示检测效果, 对检测结果进行可视化处理, 结果如图 11 所示。

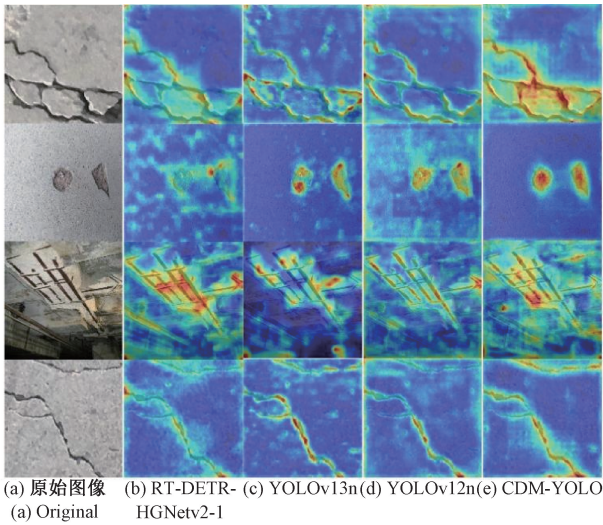


图 10 不同模型的特征关注热力图
Fig. 10 Thermal maps of different models

由图 11 可知,在密集剥落和复杂背景小露筋缺陷(前 2 行)检测上,baseline 和其他模型明显存在遗漏和错误,而 CDM-YOLO 检测结果与原始标注最为接近;对于纤细裂缝与粗大裂缝混合缺陷,CDM-YOLO 在全检测的基础上其框准性相对其他主流模型具有较大优势。综上,可视化分析表明,CDM-YOLO 在各种不同尺度缺陷情形下,其对缺陷的正确关注能力及目标检测精度均取得了不错的效果,并较好抑制了错检漏检。

2.8 算法和机器人的联合应用

若将机器人与深度学习算法结合,则可用于各类工程现场和人工难检情况,比如水上桥墩、桥墩内部和斜拉索桥梁高耸立面及顶端等缺陷壁面的自动化检测。轮式爬壁机器人通过在本体吸附腔上方搭载高速径流风机定速旋转排气,从而在本体吸附腔内形成小于大气压的负压,利用大气-吸附腔的压差产生的吸附力来完成壁面吸附。为减小转弯半径,提升灵活性,采用电机直驱行走轮的方

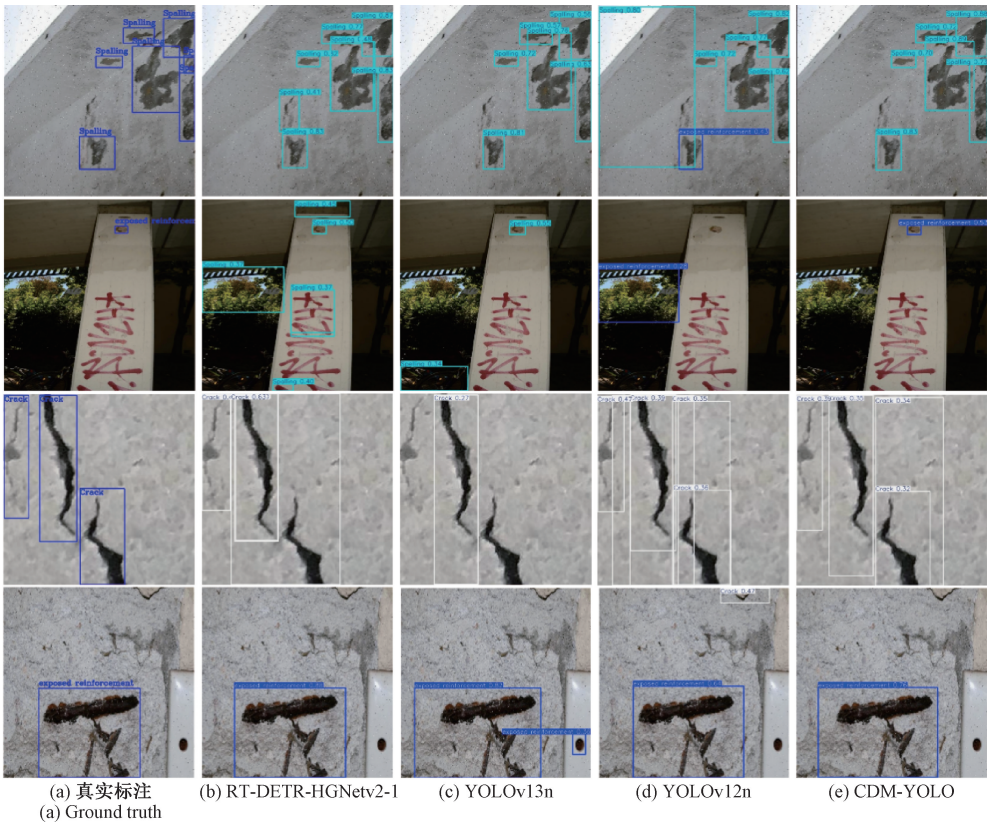


图 11 不同模型检测效果可视化示例图
Fig. 11 Visualization of different model testing images

式完成机器人壁面行走和原地转向。机器人应用场景和基本吸附模块本体如图 12 所示。
为使得检测效果更佳,采用支持彩色 RGB 图像采集的高清相机 HP60C 作为图像采集设备。该摄像头最大支持 1 080P 分辨率,且安装方便。高清摄像头及其安装如图 13 所示。

2.9 现场测试

为方便现场人员操作,基于 Python 实现了现场电脑的检测软件,该软件可以检测图片和视频两种输入类型。为进一步验证 CDM-YOLO 实际性能,使用该软件与机器人联动进行现场测试。软件界面和实际混凝土壁面检测效果则如图 14 所示。



(a) 待检桥墩 (b) 机器人本体壁面吸附
(a) Inspecting bridge piers (b) Robot adhered to the wall surface

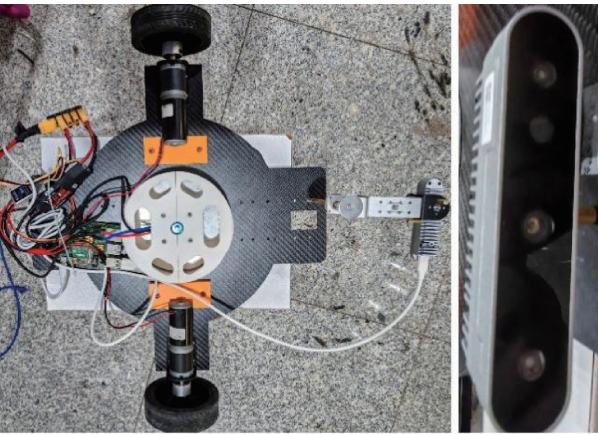
图 12 机器人应用场景及壁面吸附

Fig. 12 Robot application scenarios and wall adsorption



图 14 检测软件及实际壁面检测效果

Fig. 14 Testing software and testing effect



(a) 搭载摄像头的机器人模块 (b) 摄像头
(a) Camera-equipped robot module (b) Camera

图 13 图像采集设备及其安装

Fig. 13 Image acquisition equipment and installation

图 14 中,第 1 行检测到壁面存在宽度变化的裂缝缺陷,置信度达 0.81;第 2 行检测到小剥落缺陷,置信度达 0.78;最后 1 行检测到中等大小剥落缺陷,置信度达 0.83,且没有将青苔误检。由图 14 可知,在实际工程环境下,CDM-YOLO 检测置信度均较高,且能够分辨不同尺度缺陷,可以较好完成检测任务,具有一定泛化性能并取得较好的实践效果。



3 结 论

本研究针对一般模型对尺度多变,密集混杂桥梁混凝土缺陷检测精度低,易错检漏检等问题,提出了一种轻量级 CDM-YOLO 模型,并应用自研轮式爬壁机器人和检测软件进行验证。由于引入了能够提取细粒度特征的网络 MFC、进行有效区域聚焦的模块 DyACf 和改善深度语义信息流动性的模块 CARAFE,CDM-YOLO 的特征提取和融合,深度语义关联和流动能力较强。与原始模型相比,其 mAP(IoU=0.5)达到 87.53%,R 达到 81.10%,P 也达到 91.17%,而参数量保持在 7.6 GFLOPs,计算量保持在 2.86 M,具备一定检测精度和模型规模上的优势。总的来说,CDM-YOLO 为桥梁混凝土表面缺陷检测提供了一种轻量级多尺度缺陷检测方法,为采用机器人和现场设备进行检测提供便利。

参考文献

[1] 《中国公路学报》编辑部. 中国桥梁工程学术研究综述 2024[J]. 中国公路学报, 2024, 37(12): 1-160.
Editorial Department of China Journal of Highway and Transport. Review on China's bridge engineering research: 2024 [J]. China Journal of Highway and Transport, 2024, 37(12): 1-160.

[2] 王煜, 齐宏拓, 杨整涛, 等. 基于点云分层融合架构的混凝土气孔缺陷量化评估方法[J]. 仪器仪表学报, 2024, 45(7): 86-98.
WANG Y, QI H T, YANG ZH T, et al. Quantitative assessment method of concrete airhole defects based on point cloud hierarchical fusion architecture [J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 86-98.

[3] 倪彤元, 南晓鹤, 刘强, 等. 基于改进 SSD 识别混凝土表面裂缝研究[J]. 浙江工业大学学报, 2025, 53(3): 237-243.
NI T Y, NAN X H, LIU Q, et al. Research on the identification of cracks on concrete surface based on improved SSD [J]. Journal of Zhejiang University of

- Technology, 2025, 53(3): 237-243.
- [4] YANG J X, LI H, ZHANG L, et al. Multilabel concrete bridge damage classification using graph convolution[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-14.
- [5] 梁胤杰, 南新元, 蔡鑫, 等. 基于数据增广与改进 YOLOv8 的桥梁缺陷检测[J]. 广西师范大学学报(自然科学版), 2025, 43(3): 84-97.
- LIANG Y J, NAN X Y, CAI X, et al. Bridge defect detection based on data augmentation and improved YOLOv8[J]. Journal of Guangxi Normal University (Natural Science Edition), 2025, 43(3):84-97.
- [6] WANG J J, BI L, MA X, et al. An efficient YOLOX-based method for photovoltaic cell defect detection[J]. Instrumentation, 2024, 11(2): 83-95.
- [7] 苗力恒, 田莹. 改进 YOLOv8n 的多尺度轻量化水下目标检测[J]. 电子测量与仪器学报, 2025, 39(4): 141-151.
- MIAO L H, TIAN Y. Improved YOLOv8n multi-scale and lightweight underwater target detection[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(4): 141-151.
- [8] 韩建峰, 张静, 宋丽丽, 等. 改进 U-Net 的无人机航拍路面破损检测方法[J]. 电子测量技术, 2024, 47(19): 137-145.
- HAN J F, ZHANG J, SONG L L, et al. Improved pavement damage detection method of UAV based on U-Net [J]. Electronic Measurement Technology, 2024, 47(19): 137-145.
- [9] HUANG Y L, YUAN R, YANG J X, et al. Hypergraph-based network for accurate concrete bridge surface damage instance segmentation [C]. 2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM), 2024: 230-233.
- [10] 丁威, 夏哲, 舒江鹏, 等. 基于负压吸附爬壁机器人和 Transformer 的混凝土桥塔裂缝识别检测[J]. 中国公路学报, 2024, 37(2): 53-64.
- DING W, XIA ZH, SHU J P, et al. Recognition and detection of concrete bridge tower cracks using a negative pressure adhesion wall-climbing robot and Transformer [J]. China Journal of Highway and Transport, 2024, 37(2): 53-64.
- [11] RADZI S F M, ABD RAHMAN M A, YUSOF M K A M. YOLOv12-ECA: An efficient attention-enhanced detector for real-time UAV-based pothole detection[C]. 2025 14th International Conference on Information Technology in Asia (CITA), 2025: 102-107.
- [12] FENG Y F, HUANG J G, DU SH Y, et al. Hyper-YOLO: When visual object detection meets hypergraph computation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4):2388-2401.
- [13] CHOLLET F. XCEPTION: Deep learning with depthwise separable convolutions [C]. 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 1800-1807.
- [14] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [15] SHI D. TransNeXt: Robust foveal visual perception for vision transformers [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17773-17783.
- [16] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision, 2024: 1-21.
- [17] ZHU J CH, CHEN X L, HE K M, et al. Transformers without normalization [C]. 2025 Computer Vision and Pattern Recognition Conference, 2025: 14901-14911.
- [18] WANG J Q, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features [C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [19] 李金沛, 孟晓林, 胡亮亮, 等. 基于改进 YOLOv8 的桥梁小目标裂缝检测[J]. 清华大学学报(自然科学版), 2025, 65(7): 1260-1271.
- LI J P, MENG X L, HU L L, et al. Bridge small-target crack detection based on improved YOLOv8[J]. Journal of Tsinghua University(Natural Science Edition), 2025, 65(7):1260-1271.
- [20] 周颖, 孟诗乔, 孔庆钊, 等. 建筑结构损伤智能检测与响应智能预测研究综述[J]. 建筑结构学报, 2024, 45(6): 107-132.
- ZHOU Y, MENG SH Q, KONG Q ZH, et al. State-of-the-art of intelligent damage detection and response prediction of building structures [J]. Journal of Building Structures, 2024, 45(6): 107-132.

作者简介

范崇俊, 硕士研究生, 主要研究方向为视觉目标检测。

E-mail: pathfinder6319@163.com

董绍江(通信作者), 教授, 博士, 主要研究方向为特种机器人系统研发及工程应用、图像识别技术及人工智能技术。

E-mail: dongshaojiang100@163.com

罗家元, 教授, 博士, 主要研究方向为航空结构件及工业机器人结构优化及疲劳损伤预测等。

E-mail: jiayuan_luo@126.com

张霞, 教授, 博士, 主要研究方向为人形机器人、新型柔性驱动器、人机共融和肌电信号解码等。

E-mail: zx512@126.com

闫凯波, 讲师, 博士, 主要研究方向为结构冲击动力学与保护等。

E-mail: kaibo_yan@yeah.net