

针对多尺度和小目标的道路缺陷检测模型 *

李咏然^{1,2} 增兆祥^{3,4} 唐庭龙^{1,2}

(1. 三峡大学水电工程智能视觉监测湖北省重点实验室 宜昌 443002; 2. 三峡大学计算机与信息学院 宜昌 443002;
3. 昆明理工大学信息工程与自动化学院 昆明 650504; 4. 昆明理工大学云南省人工智能重点实验室 昆明 650504)

摘要:针对复杂道路场景中多尺度和形变道路缺陷检测的难题,提出一种改进的YOLOv8n道路缺陷检测模型DMS-YOLO (dynamic multi-scale YOLO)。首先,设计自适应上下文感知特征金字塔网络,实现多尺度特征的全局融合与动态加权,显著提升了模型对复杂缺陷的感知与表达能力,与现有主流特征金字塔网络相比,在精度和计算效率上表现出一定优势。其次,提出自适应多尺度动态检测头,采用可变形卷积(DCNv3)提升模型对复杂形状特征的捕捉能力,并设计协同注意力机制融合尺度和任务注意力,增强模型对多尺度信息的理解。最后,利用Focaler-IoU思想改进CIoU损失函数,提高对小目标的检测能力。实验结果表明,在减少计算量的基础上,DMS-YOLO模型在RDD2022数据集上mAP@0.5达到了87.9%,较原来的基准模型提高了3%,同时参数量为 3.67×10^6 ,计算量为8 GFLOPs,模型体积仅有7.3 MB,具备轻量化特性和易部署性。同时,在SVRDD数据集上,DMS-YOLO在各项性能指标上均有提升,进一步验证了所提模型具有较好的泛化性和鲁棒性。与其他主流模型和最新检测算法相比,DMS-YOLO的综合指标均表现优异,对道路缺陷检测具有实际应用意义。

关键词:道路缺陷检测;YOLOv8n;特征金字塔网络;可变形卷积;损失函数

中图分类号: TP391.4; TN911.73 文献标识码: A 国家标准学科分类代码: 520.2060

Road defect detection model for multi-scale and small targets

Li Yongran^{1,2} Zang Zhaoxiang^{3,4} Tang Tinglong^{1,2}

(1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002, China; 2. School of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; 3. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China; 4. Key Laboratory of Artificial Intelligence in Yunnan Province, Kunming University of Science and Technology, Kunming 650504, China)

Abstract: To address the challenges of detecting multi-scale and deformed road defects in complex road scenarios, an improved YOLOv8n model for road defect detection, named DMS-YOLO, is proposed. First, an adaptive context-aware feature pyramid network is designed to achieve global fusion and dynamic weighting of multi-scale features, significantly enhancing the model's ability to perceive and express complex defects. Compared to existing mainstream feature pyramid networks, this approach demonstrates clear advantages in both accuracy and computational efficiency. Second, an adaptive multi-scale dynamic detection head is introduced, leveraging deformable convolution (DCNv3) to improve the model's capability in capturing complex shape features, and a Collaborative Attention Mechanism is designed to integrate scale and task attention, enhancing the model's understanding of multi-scale information. Finally, the CIoU loss function is improved using the Focaler-IoU idea to enhance the detection of small targets. Experimental results show that, with reduced computational cost, the DMS-YOLO model achieves a mAP@0.5 of 87.9% on the RDD2022 dataset, a 3% improvement over the baseline model. The model has 3.67×10^6 parameters, 8 GFLOPs of computational cost, and a model size of only 7.3 MB, demonstrating its lightweight nature and ease of deployment. Additionally, on the SVRDD dataset, DMS-YOLO improves on all

performance metrics, further validating the model's generalization and robustness. Compared to other mainstream models and state-of-the-art detection algorithms, DMS-YOLO shows superior overall performance, demonstrating its practical application value in road defect detection.

Keywords: road defect detection; YOLOv8n; characteristic pyramid network; deformable convolution; loss function

0 引言

随着城市化进程的加速和交通网络的不断拓展,我国道路总长已达到 $535 \times 10^4 \text{ km}$ ^[1]。然而,随着交通量的不断增加以及自然环境因素的侵蚀,道路损坏问题日益凸显,这些道路缺陷不仅增加了车辆的运行成本,还可能引发交通事故,对人们的生命财产安全构成威胁^[2]。同时,长期存在的道路缺陷还可能对社会经济发展造成深远的负面影响。因此,及时准确地检测出道路缺陷,对于保障道路安全和提升城市运行效率具有重要意义^[3]。

早期道路缺陷检测主要依赖人工检查,配合测量尺、照相机等简单工具记录裂缝、坑洞等缺陷。这种方法存在效率低、主观性强且覆盖范围有限等问题,难以满足现代道路网络大规模和高精度的监测需求^[4]。

随着图像处理技术的发展,自动化检测技术开始被应用于道路缺陷检测。郭全民等^[5]提出基于张正友标定法、阈值分割法和投影法相结合的混凝土路面裂缝检测技术。Nnolim 等^[6]提出一种基于偏微分方程的自适应预处理算法,通过图像增强和梯度匹配实现自动检测裂缝。马文涛等^[7]采用 Zhang-Suen 算法结合形态学运算改进裂缝骨架提取,解决光照不均和噪点等问题,提高了裂缝检测的准确性。然而,这些传统图像处理方法对光照、噪声等环境变化较为敏感,在处理复杂路面纹理图像时鲁棒性较差。此外,这些方法依赖人工设定的规则和参数,难以有效适应复杂的缺陷形态和背景条件^[8]。

近年来,随着深度学习的发展,其强大的特征学习能力可以自动从图像中提取判别特征,为道路缺陷的自动检测提供了一种新方法。晏班夫等^[9]提出了一种结合 Faster R-CNN 和形态学的道路缺陷检测方法,利用 Faster R-CNN 进行自动特征学习和目标区域建议,并通过形态学操作细化裂缝的形态特征。廖延娜等^[10]提出了一种基于热力图的 YOLOv4 桥梁裂缝检测算法,通过改进检测头和引入 Dice 系数损失,有效解决了检测框重叠问题,提升了检测精度和速度。He 等^[11]在 YOLOX-S 的基础上,设计了主辅双路模块 (main and auxiliary dual path module, MADPM) 以增强裂缝特征提取,提出了非均匀融合结构 (uneven fusion structure with transpose and inception convolutions, TI-UFS) 以优化特征融合,并引入 EIOU 损失函数加速网络训练。陈建瑜等^[12]提出了一种基于 YOLOv5 的路面缺陷快速检测模型,通过优化锚框

参数和引入卷积块注意力模块 (CBAM) 机制,提高了检测精度和速度。于天河等^[13]提出了一种基于 SW-Net 的道路裂缝分割算法,通过结合跳跃级往返多尺度融合模块和注意力门机制,可以有效提取各种复杂裂缝的特征。Pei 等^[14]提出了 DSC-C2f 特征提取模块,以适应裂缝的细长连续形态,并在网络各阶段之间集成了一个考虑通道关系和长距离位置依赖性的坐标注意模块,有效促进了深层和浅层语义信息的融合,从而提升了检测精度。

尽管有许多道路缺陷检测模型被提出,但是目前道路缺陷检测仍存在一些难题,如道路缺陷形状和大小差异较大,模型难以准确提取病害特征^[15];模型在处理不同尺度的特征时,难以同时保留全局信息和细节特征,导致对复杂背景下的目标检测能力有限,容易出现误检、漏检的情况^[16];模型无法兼顾低计算量和高准确性之间的平衡,难以在实际应用中实现高效的部署^[17]。因此,本文在 YOLOv8 的基础上,提出了一种改进 YOLOv8 的轻量级道路缺陷检测模型,命名为 DMS-YOLO (dynamic multi-scale YOLO),模型代码公开在 <https://github.com/woshiLiyr/DMS-YOLO>。

1) 重构颈部网络,本文设计金字塔上下文提取模块 (pyramid context extraction, PCE) 整合不同层级的特征信息,提升模型对全局信息的感知能力,同时强调关键特征。此外,设计层级加权融合增强模块 (hierarchical weighted fusion enhancement module, HWFEM),动态调整特征图的强度,进一步增强模型在复杂背景下多尺度特征的表示能力。

2) 设计自适应多尺度动态检测头 (adaptive multi-scale dynamic head, AMS-Dyhead),通过引入可变形卷积增强模型对复杂形状和不规则边界的捕捉能力,并设计协同注意力机制结合尺度注意力与任务注意力,提高模型对目标形状和姿态变化的适应性。

3) 应用 Focal-IoU 思想改进 CIoU 损失函数,改善类别不平衡问题,提高模型对小目标的检测能力,进而提升整体检测性能。

1 YOLOv8 算法

YOLOv8 被广泛应用于目标检测、物体跟踪等应用领域,主要由主干网络 (backbone)、颈部网络 (neck) 以及预测头 (head) 3 个部分构成。YOLOv8 网络框架如图 1 所示。

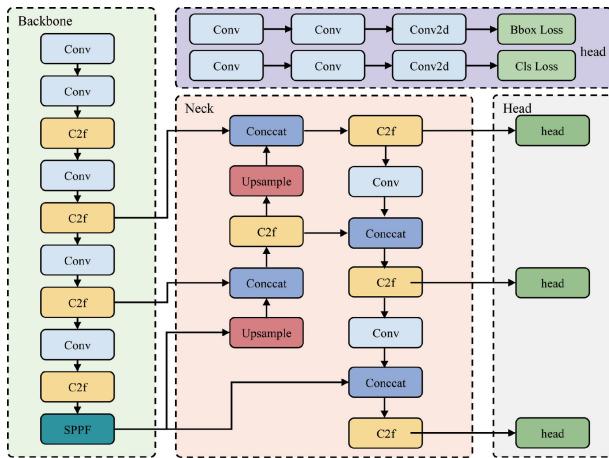


图 1 YOLOv8 网络框架

Fig. 1 YOLOv8 network structure

YOLOv8 的主干网络基于 CSPDarknet53 的思想, 通过多层卷积提取多尺度特征图。颈部网络部分采用了路径聚合网络 (path aggregation network, PAN) 结构^[18], 这是一个双向特征融合网络, 包含自上而下和自下而上的路径。自上而下路径负责传递高层语义信息, 自下而上路径负责传递低层细节信息, 这种结合可以使模型能够更好地融合低层次的细节信息与高层次的语义信息。检测头部分采用的是解耦合头结构, 将分类分支和边界框回归分支分离开来。分类任务使用二分类交叉熵损失 (BCE Loss), 而边界框回归任务采用分布式聚焦损失 (DFL Loss) 和完全交并比损失 (CIoU Loss), 这样的解耦设计可以更好地处理分类和定位任务。此外, YOLOv8 采用的是无锚点结构 (anchor-free)。相比于传统的锚点机制, 无锚结构减少了超参数调整的需求, 简化了模型设计。

YOLOv8 凭借其高效的特征提取能力和较轻的计算负担, 在多个任务中取得了显著效果。然而, 在处理小目标或多尺度目标的复杂道路缺陷检测任务中, 仍存在不足。为此, 提出 DMS-YOLO 模型, 以提升道路缺陷检测的性能。

2 本文方法

针对 YOLOv8 模型, 本文提出了一系列改进策略以提高模型在道路缺陷检测中的性能。具体来说, 针对原始模型在多尺度特征融合和目标形变适应方面的不足, 设计了自适应上下文感知特征金字塔网络 (adaptive context-aware feature pyramid network, ACA-FPN), ACA-FPN 通过改进传统的特征融合方法, 实现了多尺度特征的全局化整合, 并动态调整特征权重, 以更精准地捕捉复杂场景中的关键特征。其次, 设计了 AMS-Dyhead, AMS-

Dyhead 在检测头中引入可变形卷积 DCNv3 (deformable convolution V3), 并设计协同注意力机制融合尺度与任务注意力, 以提升模型对目标形状变化的适应能力。此外, 为了改善现有回归损失函数对中等难度样本的处理效果, 将 ClIoU 替换为 Focaler-IoU 损失函数, 从而进一步增强了模型对小目标和复杂场景的检测性能。

2.1 ACA-FPN 网络

在 YOLOv8 特征融合模块中, 使用的是传统的 PAN 结构进行多尺度特征融合。虽然该结构通过构建双路径金字塔有效实现了不同尺度特征的交互, 但在处理道路缺陷检测任务时仍存在以下 3 个主要问题: 1) 该结构通过逐层上采样将相邻层级的特征图进行融合, 然而在这一过程中, 简单的插值操作会带来信息稀释, 导致模型无法充分利用不同分辨率特征图中的所有信息^[19]。2) 道路缺陷检测任务中的目标物体通常具有较大的尺度, 逐步上采样低分辨率特征图会削弱全局语义信息的表达, 从而对大尺度目标及复杂缺陷的精准识别带来一定限制。3) 在高层特征与低层特征的融合过程中, 缺乏对特定语义信息的重点关注, 无法有效区分不同语义层次的重要性, 从而限制了模型的特征表达能力。

为了解决上述问题, 提出了 ACA-FPN 网络, 结构如图 2 所示。ACA-FPN 主要由两个部分组成, 即 PCE 模块和 HWFEM 模块。首先, 通过 PCE 模块整合多尺度特征, 提取全局上下文信息, 并通过自适应加权不同通道的特征, 增强模型对关键特征的捕捉能力, 同时利用自注意力机制建立长程依赖, 进一步提升对全局与局部信息的捕捉能力。其次, HWFEM 模块在融合各层特征时自适应分配权重, 实现层级加权融合, 确保特征在融合过程中能有效捕捉和突出重要的语义信息。此外, 模型还设计了动态插值融合模块 (dynamic interpolation fusion, DIF), 该模块通过对低分辨率特征图上采样后与高分辨率特征图逐元素相加, 进一步促进了多尺度特征间的信息交互。

1) PCE 模块

为了解决 PAN 在多尺度特征信息融合中的信息稀释问题, 并增强模型对全局和局部信息的表达能力, 设计了 PCE 模块。与传统逐层融合不同, PCE 模块采用全局化的多尺度融合策略, 直接将来自不同尺度的特征图同时输入, 整体处理, 以保留更多的上下文信息。PCE 模块由金字塔池化部分和注意力增强部分组成, 金字塔池化部分通过整合不同尺度的空间信息以捕捉全局语义, 注意力增强部分则通过通道加权和自注意力机制强化关键信息的表达, 如图 3 所示。

在金字塔池化部分, 首先对来自主干网络的多尺度特征图进行自适应池化操作, 将它们缩小至统一分辨率。接着, 应用深度可分离卷积将空间卷积和通道卷积分开处理, 从而更有效的提取局部空间信息。这种融合方式

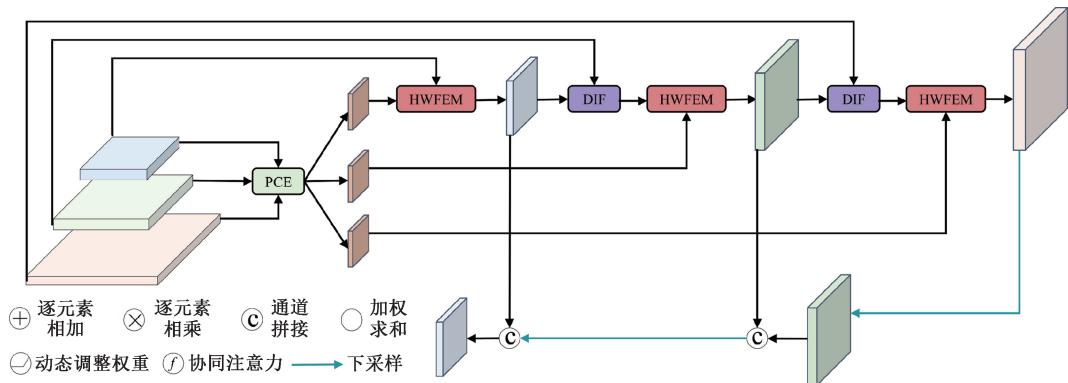


图 2 ACA-FPN 结构

Fig. 2 ACA-FPN structure

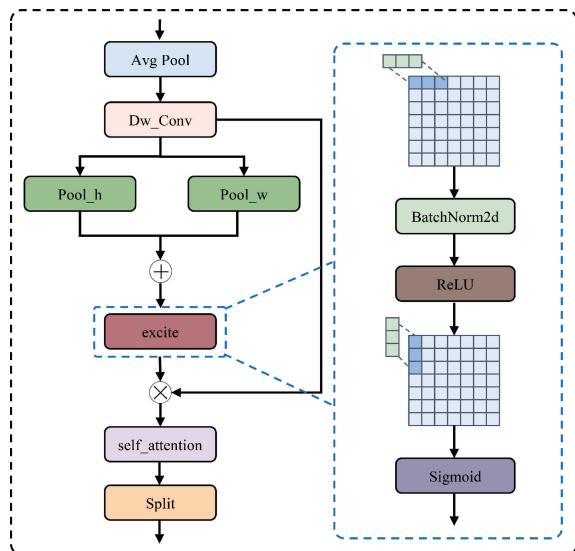


图 3 金字塔上下文提取模块结构

Fig. 3 Pyramid context extraction module structure

一方面保持了不同尺度特征的全局信息,另一方面增强了对局部细节的表达能力,提升模型对道路缺陷中小尺度目标(如坑洞)的捕捉能力。

在注意力增强部分,模型结合通道注意力和自注意力机制,以增强对局部和全局特征的捕捉。在通道注意力中,首先通过分组卷积在宽度和高度方向分别进行卷积,逐步捕捉水平和垂直方向的上下文信息。随后,通过批量归一化和 ReLU 激活函数提升模型的非线性表达能力,并通过 Sigmoid 激活函数生成权重图,将其与原特征逐通道相乘,从而赋予不同通道不同的重要性,使模型能够更加精准的捕捉到关键特征。

加权后的特征图进一步通过自注意力机制捕捉特征图中长距离的依赖关系。自注意力机制通过生成查询、键和值特征映射,计算特征图各位置间的依赖关系,使模型在处理大尺度目标时能够充分捕捉全局语义信息。此

外,通过残差连接保持原始特征的完整性,确保融合后的特征在捕捉长程上下文信息的同时不丢失细节,进一步增强了模型对全局和细节信息的捕捉能力。

2) HWFEM 模块

在计算机视觉任务中,有效融合不同分辨率的特征图对于提升模型性能至关重要。然而,传统特征融合过程中通常只是在相邻层之间进行逐层融合,信息传递往往依赖于多次上采样和相邻层级的累积,这可能导致低分辨率特征在逐层传递过程中信息丢失问题。为了解决这一问题,提出 HWFEM 模块。该模块采用加权引导和逐层叠加的策略,通过自适应加权和高效信息传递,增强了低分辨率的全局语义信息在多尺度特征中的表达,确保特征在融合过程中的完整性和区分度。

在 HWFEM 模块中,首先对来自低分辨率的特征生成权重图,权重图经过 Sigmoid 激活自适应生成加权系数,并通过上采样与对应的高分辨率特征逐元素相乘,实现加权引导融合。这可以动态调整不同通道的权重,使模型聚焦于高分辨率特征图的关键区域,同时抑制冗余信息的影响,加权引导融合部分如图 4 所示。

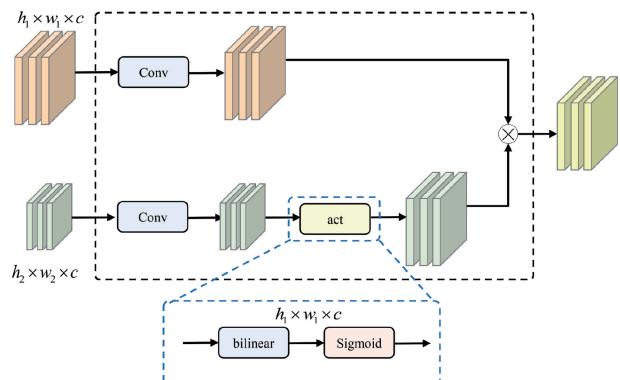


图 4 加权引导融合部分

Fig. 4 Weighted guided fusion section

此外,在每一级融合过程中通过 DIF 模块,低分辨率特征被上采样至目标分辨率,并直接与当前层级的高分辨率特征逐元素相加。这种逐层叠加融合确保了多尺度信息的全局传递,相比传统的逐层传递方式,逐层叠加不仅减少了信息稀释,还使得低分辨率的全局语义信息能够直接作用于所有高分辨率特征图上,从而更好地整合全局与局部信息。

2.2 AMS-Dyhead 模块

道路缺陷图像中,缺陷种类繁多、形状各异,且场景复杂。在 YOLOv8 检测头中,检测小物体或形状不规则物体时,往往难以精准捕捉目标的形状和位置特征。为了提升模型在干扰背景下对不同形状、大小的道路缺陷的检测能力,本文通过改进动态检测头 (dynamic head)^[20],设计了 AMS-Dyhead,如图 5 所示。

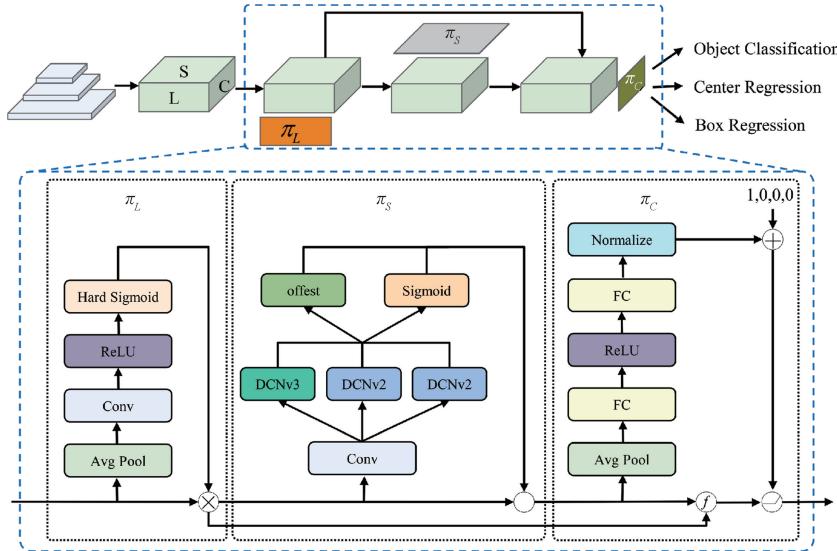


图 5 AMS-Dyhead 框架
Fig. 5 AMS-Dyhead structure

动态检测头通过结合尺度感知、空间感知和任务感知 3 种注意力机制,提升了算法对多尺度和多任务特征的表达能力。AMS-Dyhead 首先对输入的高、中、低分辨率的特征图进行初步提取,然后通过动态卷积进一步增强特征的空间感知能力。在这一过程中,AMS-Dyhead 引入了 DCNv3^[21],DCNv3 能够根据特征图的局部变化自适应调整感受野,使卷积操作更适应复杂的空间结构和多尺度变化,可以提升模型捕捉复杂形状和不规则边界特征的能力,适用于检测形状不规则的道路缺陷。

其次,AMS-Dyhead 将原本独立使用的尺度注意力和任务注意力进行了融合。尺度注意力用于捕捉多尺度特征图中的显著性信息,确保关注到不同尺度的关键区域;任务注意力则用于调节特征对分类和定位等任务的响应。在 AMS-Dyhead 中,通过在空间与任务特征维度上逐级融合,生成综合注意力图,使得模型在融合多尺度特征的同时,能够动态适配任务需求,实现更细粒度和多层次的特征表达。这种协同注意力机制不仅增强了特征图的表达能力,有效减弱了不同任务关注点的割裂现象,还使模型可以更灵活地适应复杂视觉任务需求。

2.3 损失函数改进

预测框与真实框之间的重叠度,以提高定位精度。YOLOv8 采用的是 CIoU 损失函数来优化边界框回归,其损失函数的定义如下:

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha\nu \quad (1)$$

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \quad (2)$$

$$\nu = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (3)$$

式中: b, b^{gt} 分别表示预测框和真实框的中心点坐标; $\rho^2(b, b^{gt})$ 为预测框和真实框的欧氏距离; c^2 为最小外接矩形的对角线距离; w^{gt}, h^{gt} 分别为真实框的宽高; ν 用于衡量长宽比的一致性; α 是一个正权衡参数。通过式(1)~(3)可知,CIoU 主要考虑中心点距离、宽高比以及重叠面积来优化框回归,但未考虑到不同样本难易程度对回归结果的影响。

为了解决这一问题,本文提出使用 Focal-IoU 损失函数^[22]来替换 CIoU 损失函数。Focal-IoU 损失函数定义如式(4)、(5)所示。

在目标检测任务中,边界框回归函数通常用于优化

$$IoU_{focaler} = \begin{cases} 0, & IoU \leq d \\ \frac{IoU - d}{u - d}, & d < IoU < u \\ 1, & IoU \geq u \end{cases} \quad (4)$$

$$L_{Focaler-IoU} = 1 - IoU_{focaler} \quad (5)$$

当 IoU 低于下限阈值 d 时, 损失为 0; 当 IoU 超过上限阈值 u 时, 损失为 1; 而当 IoU 处于这两个阈值之间时, 损失会随着 IoU 值的增加而线性递增。这种设计使损失函数在特定区间对 IoU 值变化更为敏感, 从而更加关注那些边界框重叠度适中的样本, 有助于模型更有效地从中等难度样本中提取特征, 而不仅仅关注于易于或难以处理的样本。

2.4 DMS-YOLO 模型

综上所述, 本文提出了一种改进后的 YOLOv8 网络

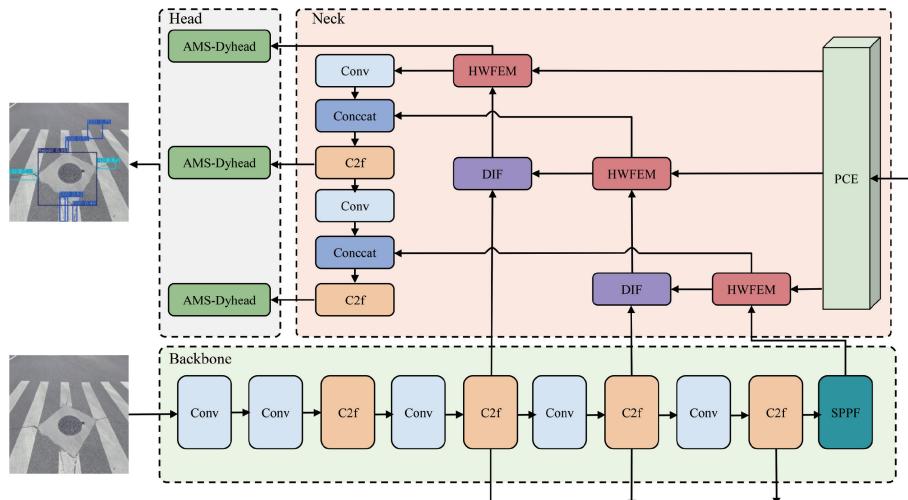


图 6 DMS-YOLO 网络结构

Fig. 6 DMS-YOLO network structure

3 实验与结果分析

3.1 实验数据集与实验环境

本文选用 RDD2022 数据集中来自中国的无人机和车载摄像头拍摄的图像, 该数据集是唯一同时涵盖无人机与车载摄像头两种拍摄模式的数据来源^[23]。这种多样化的采集方式能够更贴近实际道路缺陷检测的应用场景, 有助于模型在不同采集条件下的性能验证与适应性提升。所选图像涵盖 5 种类型的道路缺陷, 分别是纵向裂缝 (D00)、横向裂缝 (D10)、网状裂缝 (D20)、坑洼 (D40) 以及修补 (Repair), 总计 4 378 张图像, 共 8 764 个实例。此外, 数据集中包含多种光照条件 (如晴天、阴天以及湿滑路面), 并且图像质量存在差异, 部分图像存在模糊现象, 这些因素可以增强模型在实际应用中的鲁

结构, 命名为 DMS-YOLO, 其网络结构如图 6 所示。首先, 构建 ACA-FPN 网络代替原来的 PAN 结构, 从而实现多尺度特征的有效融合, 并利用上下文信息动态调整不同特征的权重, 增强了模型对多尺度道路缺陷的检测性能。其次, 提出了 AMS-Dyhead 检测头, 通过引入 DCNv3 替换空间感知注意力中的卷积网络以更好的捕捉不同形状和尺度的目标, 同时设计协同注意力机制使得模型在分类和定位任务之间实现更细粒度的特征表达。最后, 引入 Focaler-IoU 替代传统 IoU 损失函数, 通过焦点损失机制使模型更加关注中等困难的样本, 避免过度聚焦于最容易或最难的样本, 同时缓解正负样本不平衡问题, 提升了模型对小目标或模糊边界目标的检测能力。

棒性。数据集按照 8 : 1 : 1 随机划分为训练集、验证集和测试集。

为了验证所提出算法的有效性, 本文建立了相应的实验平台, 所有消融实验均在相同实验环境与同样实验配置参数下进行。在实验过程中, 使用 Ubuntu20.04 作为操作系统, GPU 型号为 NVIDIA GeForce RTX 3090, CPU 型号为 Intel(R) Xeon(R) Silver 4310 CPU @ 2.10 GHz。运行库版本为 CUDA11.8, 深度学习框架采用 PyTorch2.1.1, Python 版本为 3.11.9, 并选择 YOLOv8n 作为基准网络模型。实验参数设置如表 1 所示。

3.2 评价指标

实验采用精度 (precision, P)、召回率 (recall, R)、平均精度 (mean average precision, mAP)、F1-score、参数量 (param)、计算量 (GFLOPs) 等指标评估 DMS-YOLO 模型的性能。各指标的具体计算公式如下:

表 1 实验参数表

Table 1 Experimental parameter table

参数名	value
迭代次数	300
批大小	32
输入图像尺寸	640×640
工作线程数	8
优化器	SGD
学习率	0.01
动量	0.937
权重衰减系数	0.000 5

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (9)$$

$$F1-score = \frac{2(precision \times recall)}{precision + recall} \quad (10)$$

式中: TP 代表模型正确检测到目标对象的数量; FN 代表模型漏检目标对象的数量; FP 代表误检目标对象的数量。各个类别的平均精度 (AP) 通常被视为精确率-召回率曲线下的面积, mAP 是所有类别 AP 的平均值。 $mAP@0.5$ 是指当预测置信度为 0.5 时, 模型所能到达的精确率, $mAP@0.5:0.95$ 则是指当预测置信度在 0.5 和 0.95 之间时模型所能达到的精确率。 $F1-score$ 是精确率和召回率的加权平均, 用以衡量模型的稳定性。

3.3 模块设计有效性分析

1) ACA-FPN 网络有效性分析

为了进一步验证 ACA-FPN 特征金字塔网络的有效性, 本文在 YOLOv8n 基准模型上添加了目前比较流行的几种特征融合网络的实验, 包括双向特征金字塔 (BiFPN)^[24]、SlimNeck^[25]。实验结果如表 2 所示。

表 2 不同特征融合网络实验对比

Table 2 Comparison of experiments with different feature fusion networks

模型	mAP@0.5/%	Precision/%	Param/(×10 ⁶)	计算量/GFLOPs
YOLOv8n	84.9	83.0	3.01	8.1
BiFPN	85.3	82.6	2.0	7.2
SlimNeck	85.6	80.9	2.8	7.3
ACA-FPN	86.3	85.7	3.83	7.7

从表 2 的实验结果可以看出, 在 YOLOv8n 的基础上, 虽然 BiFPN、SlimNeck 能够减少参数量和计算量, 但其在精度提升方面效果有限。BiFPN、SlimNeck 分别在 $mAP@0.5$ 上提升了 0.4% 和 0.7%, 但 Precision 均有所下

降, 分别为 82.6% 和 80.9%。相比之下, ACA-FPN 在 $mAP@0.5$ 和 Precision 上均取得了显著提升, $mAP@0.5$ 提升至 86.3%, 相较于 YOLOv8n 提升了 1.4%, 同时 Precision 显著提高至 85.7%。此外, 尽管 ACA-FPN 的参数量略高于其他模型, 但计算量仅为 7.7 GFLOPs, 低于基准模型 YOLOv8n 的计算量。综上所述, ACA-FPN 不仅在检测精度上远超其他特征融合网络, 还在保持计算效率的同时, 实现了对多尺度特征的更有效融合。

2) AMS-Dyhead 有效性分析

AMS-Dyhead 通过将可变形卷积 DCNv3 引入动态检测头, 增强了模型的空间感知能力, 并提高了其对目标形变的适应能力。同时通过融合尺度注意力和任务注意力提升了模型对不同尺度特征的响应能力和对多任务的特征表达精度。为了评估 DCNv3 在 AMS-Dyhead 中不同分辨率位置对模型性能的影响, 并确定最佳组合方式, 设计有效性实验如表 3 所示。其中, YOLOv8n 为基准模型, Dyhead 为未替换卷积的动态检测头, all 表示将全部卷积替换为 DCNv3, high、mid、low 分别表示仅将高、中、低分辨率卷积替换为 DCNv3, high&mid、high&low 和 mid&low 则分别表示同时替换高中分辨率、高低分辨率和中低分辨率的卷积为 DCNv3。

由表 3 可知, 引入 AMS-Dyhead 模块后, 模型参数数量有所降低, 且在道路缺陷检测中精度大幅提升。与 YOLOv8n 相比, AMS-Dyhead 在 $mAP@0.5$ 上提高了 1.6%, Precision 提高了 2.9%。与原始 Dyhead 对比时, AMS-Dyhead 在参数量下降的基础上, Precision 提升了 5%, $mAP@0.5$ 提升了 0.3%, 验证了设计的优化效果。特别是当仅替换低分辨率特征图的卷积为 DCNv3 时, 模型能够在最大程度提升精度的同时参数量降至 2.84×10^6 。

表 3 AMS-Dyhead 模块有效性实验

Table 3 AMS-Dyhead module effectiveness experiment

模型	mAP@0.5/%	Precision/%	Param/(10 ⁶)	计算量/GFLOPs
YOLOv8n	84.9	83.0	3.01	8.1
Dyhead	86.2	80.9	2.88	8.1
all	84.7	83.3	2.78	8.5
high	86.3	85.9	2.84	8.4
mid	85.0	81.6	2.84	8.4
high&mid	84.6	84.2	2.81	8.5
high&low	85.9	84.0	2.81	8.4
mid&low	85.5	81.6	2.81	8.5
low	86.5	85.9	2.84	8.4

3.4 消融实验

为了验证 DMS-YOLO 中各项改进点在道路缺陷检测中的有效性, 本文以 YOLOv8n 模型作为基础网络, 在 RDD2022 数据集上进行消融实验, 实验结果如表 4 所示。

表4 消融实验结果对比

Table 4 Comparison of results of ablation experiments

序号	ACA-FPN	AMS-Dyhead	Focaler-IoU	mAP@0.5/%	mAP@0.5;0.95/%	Precision/%	Recall/%	Param/($\times 10^6$)	计算量/GFLOPs
1				84.9	58.6	83.2	80.6	3.01	8.1
2	✓			86.3	59.2	85.7	76.7	3.83	7.7
3		✓		86.5	59.9	84.8	79.2	2.84	8.4
4			✓	85.6	58.6	85.1	79.5	3.01	8.1
5	✓	✓		87.4	58.9	85.3	81.8	3.67	8.0
6	✓	✓	✓	87.9	59.8	85.5	82.7	3.66	8.0

由表4可知,与原始的YOLOv8n相比,使用ACA-FPN网络,尽管参数量略有增加,但mAP@0.5提升了1.4%,同时计算量减少5%;改进的检测头AMS-Dyhead能够大幅提高模型检测精度,mAP@0.5提高了1.6%,且参数量减少了6%;将原YOLOv8网络中的边界框回归损失函数重新设计为Focaler-IoU,在保持模型参数量、计算量不变的同时,mAP@0.5提升0.7%;融合ACA-FPN及AMS-Dyhead两个改进模块,mAP@0.5提升了2.5%,参数量略有提升;最后,同时引入ACA-FPN、AMS-Dyhead

和Focaler-IoU3种改进策略,构建DMS-YOLO模型,该模型的平均精度达到87.9%,在提升3%的同时计算量有所下降。图7为通过每个训练轮次得到的YOLOv8n和DMS-YOLO模型的召回率和mAP@0.5曲线变化对比。从图7可以看出,DMS-YOLO模型的召回率、mAP@0.5均高于原算法。此外,DMS-YOLO的参数量和计算量均符合轻量化设计的要求,确保了模型在实际应用中易于部署并具备实时性。

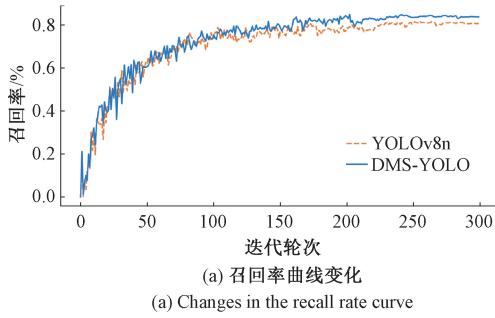
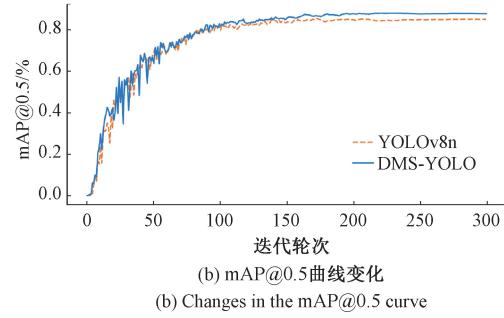
(a) 召回率曲线变化
(a) Changes in the recall rate curve(b) mAP@0.5曲线变化
(b) Changes in the mAP@0.5 curve

图7 曲线对比

Fig. 7 Comparison of curve variation

3.5 对比实验

为了评估DMS-YOLO改进算法的有效性,本文将DMS-YOLO模型与其他主流的目标检测模型在RDD2022数据集上进行了对比实验,实验包括二阶段有锚框模型Faster R-CNN,一阶段有锚框模型SSD、

YOLOv5n,一阶段无锚框模型YOLOv8、YOLOv9t、YOLOv10n、YOLOv11n和目前最新版本YOLOv12n,以及基于Transformer的RT-detr模型。本文在相同的数据集及实验条件下进行了实验,且模型均已收敛,实验结果如表5所示。

表5 主流模型在RDD2022数据集上实验对比

Table 5 Experimental results of RDD2022 dataset in mainstream models

模型	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5;0.95/%	Param/($\times 10^6$)	计算量/GFLOPs	F1-score/%	模型大小/MB
Faster R-CNN	52.0	79.8	71.5	43.3	41.45	94.4	63	108.0
SSD	75.4	51.2	70.9	39.1	26.77	60.7	61	33.7
YOLOv5n	82.9	78.5	83.1	55.7	2.51	7.2	80	5.1
YOLOv8n	83.0	80.6	84.9	58.6	3.01	8.1	81	6.0
YOLOv8s	84.6	81.6	85.7	60.2	11.1	28.7	83	21.5
RT-detr	85.5	79.2	84.0	57.2	29.3	105.2	82	56.4
YOLOv9t	83.4	76.5	84.6	55.6	2.62	10.7	79	17.1
YOLOv10n	85.2	75.2	84.3	56.5	2.71	8.4	80	5.5
YOLOv11n	85.4	77.3	84.4	58.3	2.90	7.6	81	5.8
YOLOv12n	83.5	78.9	85.2	57.1	2.56	6.3	81	5.3
DMS-YOLO	85.5	82.7	87.9	59.8	3.66	8.0	84	7.3

由表 5 可知, DMS-YOLO 模型在 RDD2022 数据集上的检测性能优于其他主流检测模型。DMS-YOLO 的 mAP@0.5 和 mAP@0.5:0.95 分别达到了 87.9% 和 59.8%, 较 YOLOv8n 分别提高了 3% 和 1.2%, 在精度和召回率上, DMS-YOLO 同样表现出色, Precision 达到 85.5%, Recall 达到 82.7%, 相较于基准模型提升明显。同时, 与当前最新版本 YOLOv12n 相比, 尽管 DMS-YOLO 在轻量化指标上略有不足, 但在精确率、召回率、mAP@0.5 和 mAP@0.5:0.95 等指标上均表现出明显优势, 分别提高了 2%、3.8%、2.7% 和 2.6%。此外, DMS-YOLO 的 F1-score 为 84%, 表明其在精度和召回率之间取得了更好的平衡, 具有较好的稳定性。在保持高精度的同时, DMS-YOLO 的参数量和计算量也表现合理。尽管 DMS-YOLO 的参数量为 3.66×10^6 , 略高于 YOLOv8n 的 3.01×10^6 , 但其模型大小仅为 7.3 MB, 与 YOLOv8n 相差不大, 而较于 Faster-RCNN、SSD、RT-detr 及 YOLOv8s 等模型, 其计算量、参数量和模型大小明显更少, 表现出较高的计算效率和部署优势。因此, DMS-YOLO 模型在道路缺陷检测任务中相较于其他模型具有更高的检测精度和稳定性, 同时在精度与模型轻量化之间实现了良好的平衡, 具

有较高的实际应用价值。

3.6 可视化分析

为了更直观评估改进后的 DMS-YOLO 模型在道路缺陷检测中的检测效果, 本文在 RDD2022 数据集上挑选了多组复杂场景图片进行测试, 如图 8 所示。从图 8 可以观察到, YOLOv8n 模型存在误检和冗余框的问题。例如, 将图 8(b1) 中的落叶检测为坑洞, 在图 8(b3) 中对同一目标生成了多个目标框, 图 8(b4) 中则将斑马线边缘检测为裂缝。而改进后的模型可以在图像复杂背景的情况下获得较好的检测效果。因此, DMS-YOLO 成功改进了原模型, 在道路缺陷检测过程中, 增强了特征提取和融合能力, 提高了算法面对复杂样本的处理能力, 降低了误检率和漏检率。

然而, 尽管 DMS-YOLO 在处理复杂样本方面取得了一定进步, 它在质量较差的图像上仍然面临挑战。例如, 模型能够在背景相对简单的模糊图像中维持较好的性能, 如图 8(c2) 所示。而当面对图片模糊且伴随复杂背景的图像时, 例如图 8(c4) 中, 模型将路面标记线的油漆掉落误判为道路裂缝。

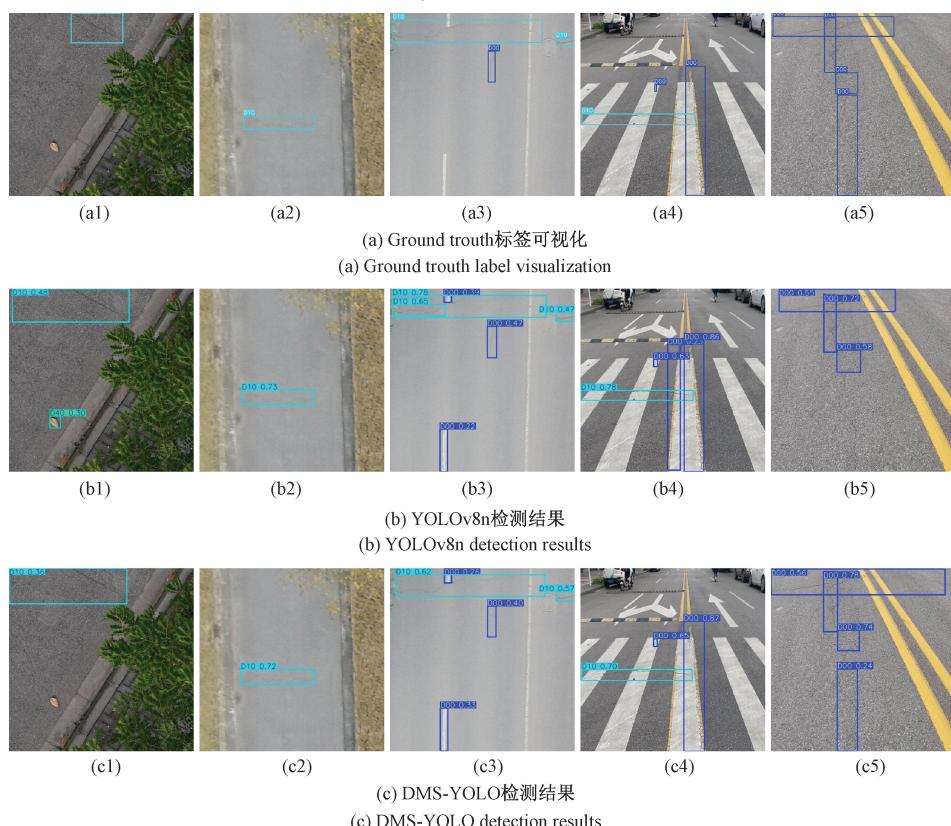


图 8 检测效果对比

Fig. 8 Comparison chart of detection effect

3.7 泛化性分析

为了进一步验证 DMS-YOLO 模型的泛化能力及鲁棒性,本研究选取公开数据集 SVRDD^[26]进行测试。该数据集涵盖了北京市 5 个区的道路破损图像,共计 8 000 张,涵盖了不同环境干扰及各种光照条件。数据集中共 7 个类别的道路损伤类型,分别为纵向裂缝、横向裂缝、龟裂、坑槽、纵向修补、横向修补和井盖。表 6 为 DMS-YOLO 在该数据集上的实验结果。

表 6 改进模型在 SVRDD 数据集上的实验结果

Table 6 Improve the experimental results of the model on the SVRDD dataset

指标	YOLOv8n	DMS-YOLO
mAP@ 0.5/%	59.6	61.3
mAP@ 0.5:0.95/%	35.1	35.9
Recall	54.0	56.7
F1-score	60.0	61.0
Param/($\times 10^6$)	3.01	3.67
计算量/GFLOPs	8.1	8.0

由表 6 可以看出,改进后的模型在计算量下降的前提下,在 SVRDD 数据集中,其 mAP@ 0.5、mAP@ 0.5:0.95、Recall 及 F1-score 分别提升了 1.7%、0.8%、2.7% 和 1%,提升效果明显,验证了改进模型具有普适性和鲁棒性。

4 结 论

为解决复杂道路场景中多尺度、形变目标检测出现误检、漏检的问题,本文设计了基于 YOLOv8 模型改进的检测模型 DMS-YOLO。首先,设计了 ACA-FPN,采用全局化多尺度特征融合和上下文加权处理,提高了模型在背景复杂、光照变化及缺陷多样性等复杂场景中的特征表达能力和目标识别精度。其次,提出 AMS-Dyhead,通过引入 DCNv3 卷积并结合尺度注意力与任务注意力,提升了模型对形态变化、边界不规则的道路缺陷的检测能力。最后,引入 Focaler-IoU 损失函数使模型更关注中等难度样本,有效提升了对小目标的检测精度。实验结果表明,本文提出的 DMS-YOLO 模型在 RDD2022 和 SVRDD 数据集上的多项指标均高于基准模型,同时具备易部署性和实用性。

在未来的研究中,计划扩充和丰富道路缺陷数据集,涵盖更广泛的缺陷类型,进一步提升检测算法的精度与泛化能力。此外,为了增强模型的适应性和鲁棒性,将采集更多低质量的实际道路图像,包括模糊、低光照以及部分缺陷被遮挡等情况,从而全面提高模型在实际应用中的性能。同时,也将考虑采用更轻量化的方法来优化模

型,在保证检测精度的前提下,尽可能减少模型的体积和计算需求,使其更方便部署在内存和计算能力受限的嵌入式设备中。

参 考 文 献

- [1] ZHANG Y C, ZUO Z W, XU X B, et al. Road damage detection using UAV images based on multi-level attention mechanism [J]. Automation in Construction, 2022, 144: 104613.
- [2] AI D, JIANG G, LAM S K, et al. Computer vision framework for crack detection of civil infrastructure—A review [J]. Engineering Applications of Artificial Intelligence, 2023, 117: 105478.
- [3] JI AN K, XUE X L, WANG Y N, et al. An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement [J]. Automation in Construction, 2020, 114: 103176.
- [4] YU L, HE S, LIU X, et al. Intelligent crack detection and quantification in the concrete bridge: A deep learning-assisted image processing approach [J]. Advances in Civil Engineering, 2022, 2022(1): 1813821.
- [5] 郭全民, 刘才臻. 路面病害巡检评估系统中的裂缝检测技术 [J]. 国外电子测量技术, 2015, 34 (7): 47-50.
- GUO Q M, LIU C ZH. Pavement crack detection method in concrete pavement disease detection and estimation system [J]. Foreign Electronic Measurement Technology, 2015, 34 (7): 47-50.
- [6] NNOLIM U A. Partial differential equation - Based enhancement and crack detection [J]. Mathematical Problems in Engineering, 2019(1): 8157205.
- [7] 马文涛, 樊春玲. 基于计算机视觉路面裂缝的识别与测量 [J]. 电子测量技术, 2020, 43 (19): 85-88.
- MA W T, FAN CH L. Recognition and measurement of pavement cracks based on computer vision [J]. Electronic Measurement Technology, 2020, 43(19): 85-88.
- [8] MOHAN A, POOBAL S. Crack detection using image processing: A critical review and analysis [J]. Alexandria Engineering Journal, 2018, 57(2): 787-798.
- [9] 晏班夫, 徐观亚, 栾健, 等. 基于 Faster R-CNN 与形态法的路面病害识别 [J]. 中国公路学报, 2021, 34 (9): 181-193.
- YAN B F, XU G Y, LUAN J, et al. Pavement disease detection based on Faster R-CNN and morphological

- operations [J]. China Journal of Highway and Transport, 2021, 34 (9): 181-193.
- [10] 廖延娜, 宋超. 基于深度学习的桥梁裂缝定位算法研究[J]. 国外电子测量技术, 2022, 41 (4): 112-118. LIAO Y N, SONG CH. Research on bridge crack location algorithm based on deep learning [J]. Foreign Electronic Measurement Technology, 2022, 41(4): 112-118.
- [11] HE X Y, TANG ZH W, DENG Y B, et al. UAV-based road crack object-detection algorithm[J]. Automation in Construction, 2023, 154: 105014.
- [12] 陈建瑜, 邹春龙, 王生怀, 等. 改进 YOLOv5 的路面缺陷快速检测方法研究[J]. 电子测量技术, 2023, 46 (10): 129-135. CHEN J Y, ZOU CH L, WANG SH H, et al. Research on rapid detection method of pavement defects by improving YOLOv5 [J]. Electronic Measurement Technology, 2023, 46 (10): 129-135.
- [13] 于天河, 徐博超, 侯善冲, 等. 基于图像的道路裂缝分割及量化方法研究[J]. 仪器仪表学报, 2024, 45 (9): 77-91. YU T H, XU B CH, HOU SH CH, et al. Research on image-based segmentation and quantification of road cracks [J]. Chinese Journal of Scientific Instrument, 2024, 45 (9): 77-91.
- [14] PEI J, WU X, LIU X. YOLO-RDD: A road defect detection algorithm based on YOLO [C]. 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2024: 1695-1703.
- [15] WANG R, LIU Z, LIU H, et al. FDSC-YOLOv8: Advancements in automated crack identification for enhanced safety in underground engineering [J]. Computer Modeling in Engineering & Sciences, 2024, 140(3): 1-15.
- [16] LI G, LIU Q, REN W, et al. Automatic recognition and analysis system of asphalt pavement cracks using interleaved low-rank group convolution hybrid deep network and SegNet fusing dense condition random field [J]. Measurement, 2021, 170: 108693.
- [17] LIU Q, LIU Z. Research on pavement crack detection based on improved YOLOv4 [C]. 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE). IEEE, 2024: 1212-1217.
- [18] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [19] WANG C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [J]. Advances in Neural Information Processing Systems, 2024, 36: 1-19.
- [20] DAI X, CHEN Y, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7373-7382.
- [21] LI H, ZHANG Y, ZHANG Y, et al. DCNv3: Towards next generation deep cross network for CTR prediction [J]. ArXiv preprint arXiv: 2407.13349, 2024.
- [22] ZHANG H, ZHANG S. Focaler-IoU: More focused intersection over union loss [J]. ArXiv preprint arXiv: 2401.1052, 2024.
- [23] ARYA D, MAEDA H, CHOSH S K, et al. RDD2022: A multi-national image dataset for automatic road damage detection [J]. Geoscience Data Journal, 2024, 11(4): 846-862.
- [24] 王宸, 杨帅, 周林, 等. 基于自适应多尺度特征融合网络的金属齿轮端面缺陷检测方法研究[J]. 电子测量与仪器学报, 2023, 37 (10): 153-163. WANG CH, YANG SH, ZHOU L, et al. Research on metal gear end-face defect detection method based on adaptive multi-scale feature fusion network [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37 (10): 153-163.
- [25] LI H, LI J, WEI H, et al. Slim-neck by CSConv: A lightweight-design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21 (3): 62-62.
- [26] REN M, ZHANG X, ZHI X, et al. An annotated street view image dataset for automated road damage detection [J]. Scientific Data, 2024, 11(1): 407-407.

作者简介



李咏然, 2019 年于三峡大学获得学士学位, 现为三峡大学硕士研究生, 主要研究方向为深度学习、机器学习。

E-mail: lliyr@qq.com

Li Yongran received her B. Sc. degree from China Three Gorges University in 2019. Now she is a M. Sc. candidate at China Three Gorges University. Her main research interests include deep learning and machine learning.

臧兆祥(通信作者),2013 年于华中科技大学获得博士学位,现为昆明理工大学副教授,主要研究方向为机器学习、决策智能。
E-mail: zhaoxiang_zang@kust.edu.cn

Zang Zhaoxiang (Corresponding author)

received his Ph. D. degree from Huazhong University of Science and Technology in 2013. Now he is an associate professor at Kunming University of Science and Technology School. His main research interests include machine learning and decision-making intelligence.



唐庭龙,2018 年于浙江工业大学获得博士学位,现为三峡大学副教授,主要研究方向为从事机器学习、深度学习、计算机视觉。

E-mail: tangtinglong@ctgu.edu.cn

Tang Tinglong received his Ph. D. degree from Zhejiang University of Technology in 2018. Now he is an associate professor at China Three Gorges University. His main research interests include machine learning, deep learning and computer vision.